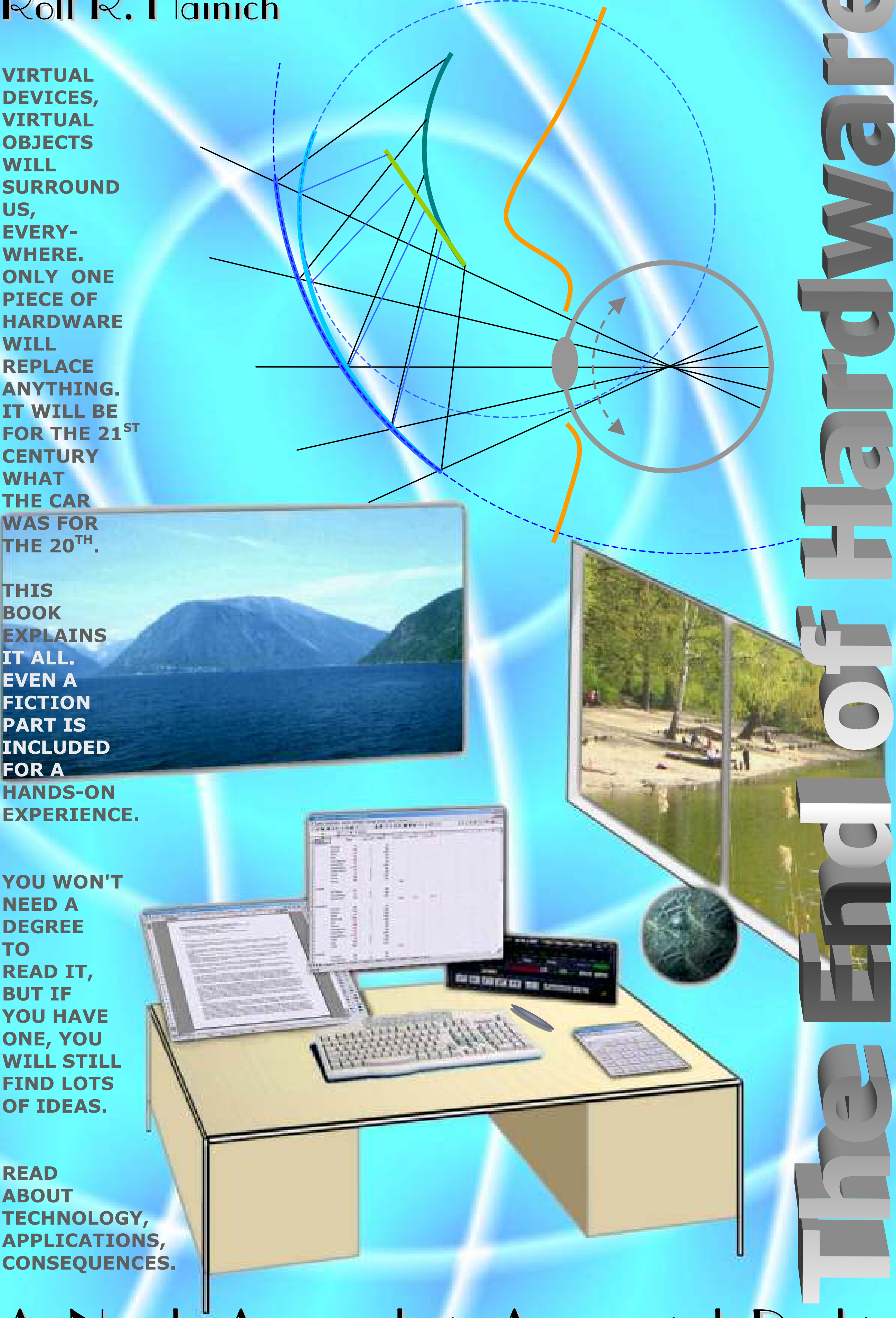
Rolf R. Hainich

The End of Hardware



VIRTUAL DEVICES, VIRTUAL OBJECTS WILL SURROUND US, EVERY-WHERE. ONLY ONE PIECE OF HARDWARE WILL REPLACE ANYTHING. IT WILL BE FOR THE 21ST CENTURY WHAT THE CAR WAS FOR THE 20TH.

THIS BOOK EXPLAINS IT ALL. EVEN A FICTION PART IS INCLUDED FOR A HANDS-ON EXPERIENCE.

YOU WON'T NEED A DEGREE TO READ IT, BUT IF YOU HAVE ONE, YOU WILL STILL FIND LOTS OF IDEAS.

READ ABOUT TECHNOLOGY, APPLICATIONS, CONSEQUENCES.

A Novel Approach to Augmented Reality

to Sigrid

The End of Hardware

A Novel Approach to Augmented Reality

2nd edition, 2nd print

Copyright © 2006,2007 Rolf R. Hainich
All rights reserved

ISBN: 1-4196-5218-4
ISBN13: 978-1-4196-5218-9
BookSurge,LLC
An Amazon.com Company

E-Book for participants of ISMAR 2008
- not for distribution -
This is the 2006 print edition, with a few minor updates of 2007.

Latest materials you will find at
http://www.theendofhardware.com
and in the ISMAR2008 keynote lecture

Contains bibliographical references and index

Computer Science
Keywords: Augmented Reality; Virtual Reality; 3D;
Mobile Computing; Media Technology; Holography

Disclaimer: All proprietary names and product names mentioned are trademarks or registered trademarks of their respective owners. We do not imply that any of the technologies or ideas described or mentioned herein are free of patent or other rights of ourselves or others. We do also not take any responsibility or guarantee for the correctness or legal status of any information in this book or any documents or links mentioned herein and do not encourage or recommend any use of it. You may use the information presented herein at your own risk and responsibility only. To the best of our knowledge and belief, no trademark or copyright infringement exists in these materials. In the fiction part, the sketches, and anything printed in special typefaces, names, companies, cities, and countries are used fictitiously for the purpose of illustrating examples, and any resemblance to actual persons, living or dead, organizations, business establishments, events, or locales is entirely coincidental. If you have any questions or objections, please contact us immediately. 'We' in all above terms comprises the publisher as well as the author. If you intend to use any of the ideas mentioned in this book, please do your own research and patent research and contact the author.

Foreword

to the 2nd edition

About a year go, the first edition of this book had been completed. Meanwhile, ideas emerged how to present some things even better, some new technical concepts surfaced, new references were added to the book's website. By the end of July this new edition had been compiled. It should be in stores in November.

In the first edition, I suggested holographic encoding instead of the object oriented MPEG extensions usually tried to represent three-dimensional images. Back then I thought somebody certainly had already come up with this long before, as it appeared to be a very natural conclusion, but finding something in myriads of holography papers is pretty difficult. Yet a very recent article [83] indicated that the idea of holographic encoding when I published it may have been entirely new. So I was finally justified in having written several more pages of explanation on it, originally more driven by the fact that most literature I found presents the basics of holography in a quite unsatisfactory way, for this purpose at least.

There were more reasons for this new edition. Many interesting little items have been added to the applications and especially to the fiction chapter. In the media chapter, a new technique is presented that will make camera arrays a lot more economical.

The concept of the mask display - as simple as it may be - is still ahead of time and deserved some little additions. The orientation by interlinked meaningless image detail I think I managed to explain even better now - albeit the 'End of Hardware' is more about hardware, the final hardware so to say, than about software. What we need real soon now is a vision interface hardware, small and versatile enough for personal use, an equivalent of the PC, to create everyday applications on it. Without the PC, there wouldn't be any noteworthy software industry by now, or would it.

Rolf R. Hainich

The End of Hardware

a novel approach to augmented reality

The End of Hardware

Content

The End of Hardware

a novel approach to augmented reality

Preface

Mobile phones will soon have harddisks and replace a camcorder, a PC, anything. But how to use all this ?

Travelling with 2 pieces of luggage, looking for my connecting flight, I just imagine fumbling around with this thing, trying to see any informations on its dwarfish display. Taking a pen, trying to stab little blots on the display, doesn't help either.

This technology, getting as near to the *personal communicator* as anything so far, will surely hit the wall in the near future, simply because there is no acceptable display, and rollout displays or projectors won't do.

Long ago I realized this problem, and now technology is ready to solve it. We need to eliminate the screen in favor of a near-eye projector, glasses with a tiny add-on that could finally weigh less than 20g. Forget those ancient virtual reality goggles. Forget anything hampering direct sight. It can be done way better.

Virtual objects, virtual devices will surround us, everywhere. They will soon replace most of today's user interface hardware, screens, keypads, entire installations, and they will do a lot more. Nevertheless, these virtual things won't clutter up our view at all, because other than with some classical approaches to augmented reality, here they will be seamlessly integrated and fixed to the real environment, rather than sticking in front of our eyes.

We will have to envision entirely new applications and usage habits. It's a new world to explore.

No more desktops mapped to a computer screen, but operating systems mapped to the real world. Available anywhere. Boosting productivity beyond any expectations.

When I started my activities back in 1993, the generation of steady, high quality images merged into natural sight was already resolved, and the only major obstacles were computing power and small high resolution displays.

Ten years later, these obstacles were gone, but very little had happened in all the other fields. Somewhere, programs and joint projects were launched whose aims looked ten years old. Basic parts of augmented reality technology still deliver themes for theses all over the world (fortunately at least this happens), but it's usually a matter of chance if somebody, maybe from institutions not generically involved in the thematic at all, picks up this or that problem because of necessity. Very few researchers around the world are dedicatedly concerned with the thematic, and a few small or medium enterprises are developing in this field. The funding of this research is still microscopic compared to that of large screen displays, for just one example.

In the beginning, only Japanese companies built camcorders, i.e. had the full spectrum of technologies at hand for tiny display glasses. Meanwhile, mobile phone technology delivers the best starting point and it's available in many countries. This is just good luck, not the farsightedness of managements.

Only a very small number of books cover augmented reality, and even fewer really deliver a strategic view. Some of the most important issues, visual orientation and mask displays for example, are hardly treated or not even at all. So I literally had no choice but to write this book. It has two main objectives:

- showing as many applications as possible, because this already induces a lot of thought and inspirations about realization;
- demonstrating the feasibility, i.e. naming the technological approaches as well as calculating the results achievable. Delivering the recipes. Otherwise, the usual skeptics could just too easily do anything away as mere speculations.

It should be readable to the (interested) public, but it can't be entirely popular science, as it has to treat several issues in a way that professionals can profit. Yet this isn't such a contradiction, as professionals from different disciplines also usually don't know a lot of each other's field.

Many augmented reality projects now deal with screens all over the place, projected items, lots of additional installations. In the context considered here, this would be sort of installing phone booths everywhere, rather than using a mobile phone. It may be the right thing for certain applications, but generally shouldn't we try to reduce hardware and costs, not to multiply them? Recent studies say copper will get rare even in spite of recycling. So will other resources. It will therefore be wise anyway to push forward technologies that reduce large appliances, office space, wiring.

We will see that the actual IT hardware of the ultimate perceptional interface could fit in a sugar cube and take less than 10 milliwatts of power. Micro and nano technologies will be very important here. Some tasks will turn out easier than anticipated. We don't need tactile feedback all the time: nobody wants gloves. We don't need gestures: too complicated. Pressing virtual keys is so simple to implement, you wouldn't believe it, but even this won't be the end. We'll finally operate anything by just looking at it. It will be like magic.

Current near eye displays won't do. We need something that dynamically adapts to eye motion, incorrectly sitting glasses, focus and aperture effects, geometry changes. Also of course something that does not impair direct sight at all, has very high resolution, a very large field of view, and a *mask display* to make virtual objects non transparent.

It can be built, and for the basic functions somebody will already have to reinvent the wheel, as many of the problems have long been solved in some military projects. Yet the entire technology still doesn't have a real supporting structure, no homogenous academic or industrial basis. It is not done with research papers covering this or that mathematical or programming problem on still inadequate interface hardware. It is not done with fancy demonstrations or utterly expensive industrial installations on the same basis.

What is necessary in this field is an entire, ambitious, multi threaded development program that involves, or better, is initiated by major industrial companies, has adequate funding and a sufficient time scale, and aims for nothing less than perfection.

It's not only technology that we have to deal with. First of all, applications have to be envisioned, fantasy has to be stimulated, the entire way of thinking has to be fully comprehended.

I remember when I had my first programming course back in the 70's. Nobody could afford a private computer back then, but I had heard a lot about it. I knew digital circuits, but I wondered about all those silly articles describing 'thinking' machines.

This couldn't work, so what? After writing my first lines of code, I knew. This thing wasn't meant to think. Within minutes I figured out how to build one. It was all simple and clear.

The important issue was not how to do it, but first to know what it should do. Just as with all innovations:

The most difficult and important step is not to find the answers, but to ask the right questions. So it is here. We know there is something that could do amazing tings. We have to do a lot of thinking about what it could realistically do and what not. Once the goals are defined, we will be able to build it. Hence, this book is meant to ponder the technology as well as its applications. The entire scope.

I went as far as to write a little SciFi prose to make this more colorful. Only by exercising our fantasy and delving into the scenario will we be able to find new applications, and there are many. Utterly important here it is to sort out the impossible and the unacceptable, or this would really be nothing but fiction.

Just "can't be done" is not an option. I've seen quite some examples where would-be experts said something wouldn't work and a short time later it had been accomplished.

Applications dictate the construction. I try to define the necessities and the technologies that could do it. You'll see that it will finally be working. We will be able to build those ultra light plastic glasses that don't even look as if they got any super technology in them. We will have to face all those problems arising from the everyday use of virtual objects. And a lot more.

Displays and especially optics have been widely neglected. Maybe nobody realized the potential of a really good solution? So I'll address this extensively, including new ideas like holographic optics and image generators that may change things entirely.

The main chapters in this book I started to write in 2004 and 2005 (even 1993 in some detail, as nearly everything in the earlier papers is still current and true).

First is an introduction, outlining the technological and economical basics and a lot of applications. Then a piece of fiction, that in my opinion will say more than any of these brittle tech ideas and that is, as said, crucial for the comprehension of the entire thematic. The applications list is also continued in part 4, with emphasis on virtual media. There I'll also address themes like 3D and surround cinema and holographic encoding.

Part 3 goes into technical detail, exploring possible solutions. Even this is meant to be understandable to the general public as well as to experts from a wide variety of fields.

It may also help anyone interested in actual construction work in figuring out what to do and how. Sensitize technologists and managers to new possibilities. Motivate people to become involved in this new development, or at least to keep it in mind with their everyday decisions.

As soon as a first - even a simple - implementation succeeds, drawing competition and accelerating the development, it will start an avalanche. An eye operated mobile phone integrated into a pair of glasses, for example. A killer application.

John Doenuts just discovered these surgical telescope magnifiers to be an ideal mobile phone accessory

The End of Hardware

part 1: Introduction

Why virtual devices?

Quite some time ago, new technologies in image processing and presentation were promising the next step in the evolution of user interfaces. Virtual reality (VR) went beyond commonly used graphics by linking computer generated images to real world parameters such as point and angle of view, location and movement in space, and physical interaction.

In [1] and especially in [2], I had suggested several improvements in this field, introducing virtual devices and other things now called augmented reality. This chapter is partly based on the original papers, of course updated, and enhanced with many more comments, pictures and examples.

In the beginning of Virtual Reality, the user was totally shielded from reality and interacted in a totally virtual environment. Bulky and low resolution displays caused more headache than fun. Alas, even nowadays the available devices could need lots of improvements. Another variety of VR is a near eye display that inserts a virtual computer screen into the natural view of the user, but fixed to the user's perspective, not to real objects [3]. This type has yet become very current in military applications.

A real merging of reality and virtuality had first been accomplished with military flight simulators, combining a real cockpit with a virtual outside view (example: CAE FOHMD, see below).

Another current type of object insertion displays maintenance manuals and drawings into real machines, an application long in use in aircraft and even sometimes in automobile maintenance, yet they're mostly just projecting the blueprints onto the machine.

For practically all these interface solutions that are trying to bring virtual objects into the real world, the term Augmented Reality is meanwhile firmly established.

Yet under this label a wide range of approaches are summoned, some that create virtual objects by projectors and screens, some that refer to the 'augmented' experience from ubiquitous computing, chips and devices strayed all over the place, some that draw augmentation from the experience with 'intelligent', i.e. computer equipped clothing.

In this book we won't deal so much with these varieties. The most interesting aspect of this technology in my view is the minimization of hardware, not its maximization.

I also think that the known applications so far have accomplished but a small part of the imaginable. They do not efficiently utilize all our capabilities.

Human beings can handle a large number of different things simultaneously by arranging them in space. We have a very strong dedicated processing unit to do this (our brain), and we make extensive use of it in any natural environment.

These capabilities should be fully exploited in the user interface of an information system. Hence, an operating system should show data and applications in 3D (3-dimensional) spaces, and even much better, in the real 3D space. All of which quite obviously can't efficiently be done on a 2D computer screen, but could be done with a stereo vision simulator, i.e. transparent display glasses.

What is essential, that we do not only use passive display objects, but let the objects become interactive, behave like real machines.

This is very similar to a windowed operating system, and not by accident, because the metaphor that made these OS successful was the imitation of devices and operation modes that we were used to.

The next logical step now is, to pull these devices and objects from the computer screen back into the real world and use our entire environment as a giant 3D 'screen'.

Virtual hardware
The Virtualware* concept

In a perfectly simulated 3D space, we could replace real hardware with virtual objects - that is, software - saving lots of material and energy, and the equipment doing this would, for the first time, fulfil the paradigm of the really universal communicator.

Indeed any software application is just very similar to a hardware device, except for the fact that it is not cast in hardware and has no real buttons to operate but virtual keys etc. instead.

A motorcycle or a chainsaw of course, are not ideally suited for a virtual implementation. We have to concentrate our effort on information machines, like PCs, screens, cellphones, notepads, and - generally - any keypad or panel of any machine there is.

Now this represents a huge advancement already: with some advanced display glasses (I call this a vision simulator** until somebody comes up with a catchier name), we could replace all of these machines or partial machines with just one piece of hardware and – obviously – an appropriate piece of software.

What needs to be developed are display glasses that are able to produce a perfect, high resolution image under all circumstances, that are not dependent on a special fixture to the head, that can totally compensate for any kind of head movements and perspectives and thereby generate the illusion of really *immobile* objects.

Another necessity are ways of seamless interaction with it by hand or finger pointing, eye pointing, speech, etc..

All this seems very difficult at a first glance, as we have to guarantee that a device essentially not bigger than ordinary plastic glasses, should incorporate perfect position sensing, eye tracking, its own cameras, and so on. Not to mention high resolution displays, optics, and of course some audio components.

---

* This term was introduced for the concept of virtual hardware in conference papers from 1993 on [1], [2]. Some people meanwhile also used it as a name for some entirely different things in software technology, that shouldn't be mixed up with our theme. **Introduced in 1994 [2].

The concept of such an ideal device has meanwhile been dreamt of by many researchers in this or another way since quite some years of course, but the difficulty proved larger than first anticipated, technology wasn't yet powerful enough, and most projects at least in the academic area had far too little resources to approach the task in a comprehensive and continuous way. The disappointment inevitably arising, led to the general conclusion that it couldn't be done in a decent timeframe, and most researchers turned to other things like intelligent clothing, ubiquitous computing and so on.

Yet there are ways to solve the problem. We shall see that all this in the end will be easier than it appears, and I will touch upon many ways to address the most important issues. Technology could already deliver within a short time.

The augmented reality concept presented here is open, communicative, and inherently mobile, as virtual objects are no longer bound to any specific installation but to real locations and objects that could be anywhere. Different people may work with the same virtual objects, and objects may be made available to anyone anywhere, by wirelessly distributing their appearance or contents.

Beneath virtual objects, an advanced vision simulator would also offer novel and yet unmatched capabilities to provide the perfect 3D movie display, something not practically possible with holography or conventional stereoscopy. With conventional high definition TV, even the novel flat panel displays are quite bulky and expensive things that could advantageously be replaced by vision simulators, even much for the better, because even a cinema sized screen could be simulated in economy sized rooms, or the screen paradigm could just be abandoned at all. The vision simulator is the equivalent of many large screens at once, as it can simulate them all concurrently, in different directions.

Within this first chapter, I will now systematically develop the paradigm of virtual devices and objects, review the necessary technologies, outline many possible applications and finally try to address the effects on economy, environment and society.

A classification of virtual objects

Virtual objects and devices come in several varieties that I will now systematize, also in comparison with traditional screen objects like in window based operating systems.

Virtual objects : Images *

Devices

Images: 2D - Image,
- Cross-Section

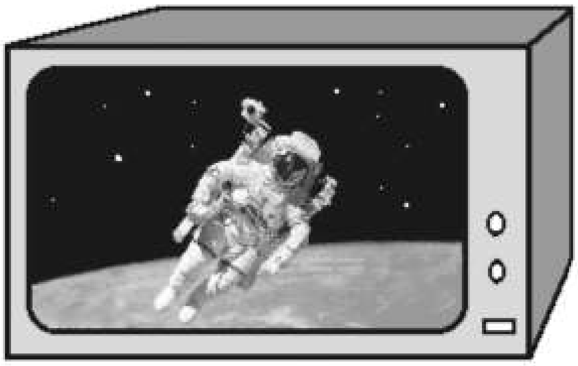
3D - Volume
- Surface
- Solid

---

[*] New Bight international airport, Cat Island, Bahamas

Devices = images with controls

2D - Typical Windows screen object
- Virtual 2D display device
- Virtual paper



3D - Virtual device with depth/volume
- Virtual 3D display window

We see that most applications for virtual devices will only require 2D or 2½D presentation in the real 3D space.
Some new types of devices will also be designed, that make full use of 3D representation.

What will be very different from a screen based system, that we can arrange such objects and controls in the real environment where we have many more degrees of freedom, and more ergonomical ways to operate them.

A virtual workspace

This is an illustration how virtual devices would replace current computer screens, and more:

The only real items in this picture are the desk and the keyboard. There is no computer monitor, not even a simulated one. Instead, several program windows are freely positioned in 3D space. They are operated with fingers or by just looking at their controls or fields. The usual window controls and the surfaces are activated as virtual buttons. Instead of a mouse, a mouse pen is used for more difficult inputs.

Everything virtual you see is just simulated in the display glasses. We have a spreadsheet up front, a text processor (tilt backwards) to the left, a virtual stereo in the right corner of the desk, and another window (calculator or notepad) on the desk itself.

Program windows alone however, are a bit boring. So I also inserted a real 3D object (could be a globe). It could also be operated (turned etc.) by hand. 3D models from computer aided design or art applications are other good candidates for these kinds of objects.
In addition to the office stuff, a virtual TV decorates the facing wall. Last but not least, the window: here we see the (maybe a bit ugly) view replaced with a better one.
We could even consider to simulate the 'light' falling in from the 'window', by overlaying processed images from cameras mounted on the vision simulator to the real and virtual scene (we will see that we need these cameras anyway, as position sensors).

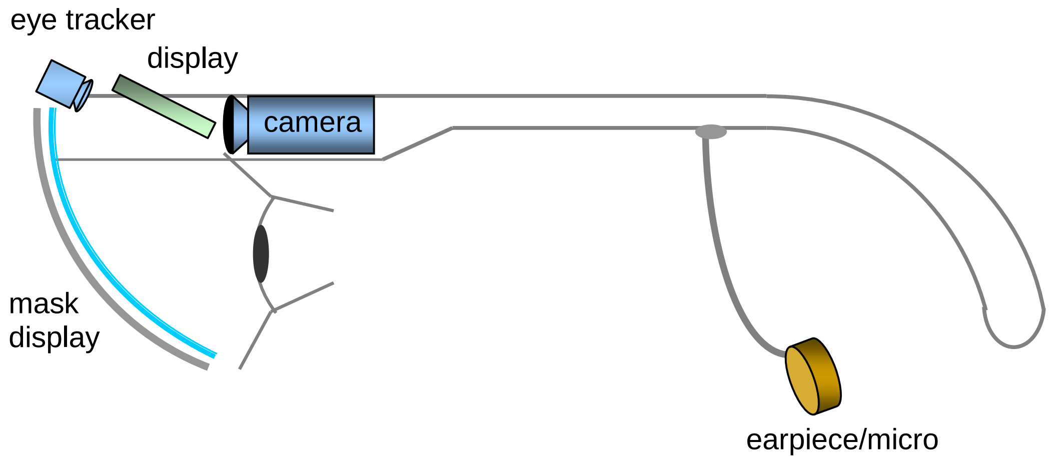
What is also visualized is the 'shadow' around virtual objects, that results from the inevitable fuzziness of the mask display (we will turn to this later).

We see that a single vision simulator can replace an entire host of monitors. It can lead to a much better utilization of computing power, a more natural and powerful way of interaction, and a better oversight.
As we use the same display to show all these monitors in all directions, we get a virtual resolution for our virtual workspace that is many times larger than that of our actual hardware. We would need several giant screens with ultra high resolution to achieve this the conventional way.

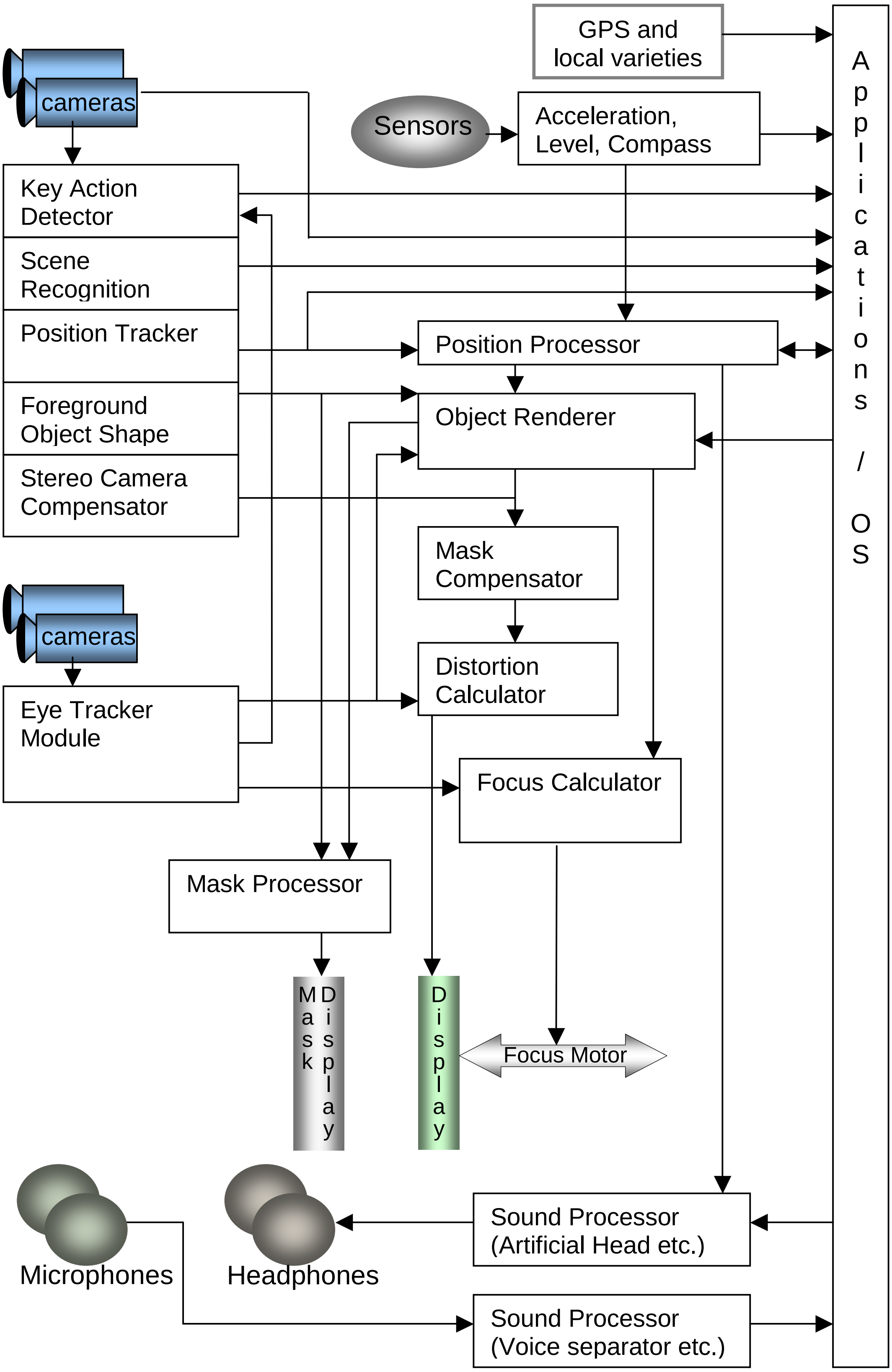
This is but a tiny first glimpse on the new possibilities, still very close to the old 'windows' paradigm of known operating systems.
A lot of imagination will be needed to unleash the full power of such a system. A whole new class of operating systems will have to be written to implement all the new features.

Let's now have a look at the one device that will make it happen. There are many shapes it could take or technologies it could use, including fancy things like holographic displays and mirrors. I will also address this extensively later on in the design chapter.

The vision simulator



Object display	→	light, high resolution glasses, with display (chip, scanner, holographic)
Object positioning	→	Head position sensors: Conventional, Optical triangulation Ultrasonic, Acceleration sensors GPS, New methods (Camera)
	→	Scene/Viewpoint recognition
Object insertion	→	Real scene acquisition (Camera)
	→	Masking techniques Additional Mask Display (LCD or other)
Interaction	→	Finger, key pressing detection (Camera)
Adaptive Display/Rendering strategies (inlays etc.)		
		Dynamic focusing (for natural stereo)
Fine positioning	→	Eye tracking
Audio		Microphone/Earpiece for playback, recording, noise cancellation, commands
Misc.		Keys at the vision simulator Headlamps (e.g. infrared for night vision)



Principal block scheme of the vision simulator and computing device (hard-and software)

Displays

The largest difficulty on the way to an acceptable virtual reality device lies in available display technology and optical designs.

With displays, it is difficult to get them small, light and crisp enough. One problem are the small pixel sizes, that are difficult to achieve with technologies usually developed for large screens.
The maximum currently is at 1280x1024 pixels in applications and first 1920x1200 display chips are announced. If we want to simulate a computer screen of just this resolution, as is common today, we would like to have about 2 times as much, as we need not only to show this item, but some surroundings as well.

This implies that with display chips, some more development is still necessary, even though a lot has happened during the last years. Several technology and design problems with micro displays have been resolved or will hopefully soon be resolved by pioneering enterprises ([33], [51]). The issue has been neglected by most of the major manufacturers, because hardly anybody seems to have envisioned a market large enough.
Useful display technologies for example are LCD, LED and DMD displays*. Another solution is the laser scanner, directly writing a laser beam image into the eye. Holographic display chips, currently in prototype versions, may also become an interesting option in the future.

What has not yet been realized anywhere, is a mask display that can cut out parts of the real sight in order to achieve non transparent virtual objects. Other than LCD, far from being optimal, there is nothing in sight so far to solve this. Nevertheless this is utterly important and I'm sure that we will see a solution.
In the applications area, there were and are many exciting military and scientific projects, like flight simulators or helmet displays for pilots, but these are scarcely published.

---

* We will further discuss display technology in the design chapter.

Scientific and commercial implementations are also heavily used in some fields, but they are also not very convenient and pretty expensive.
As stated already, display glasses should be as transparent and lightweight as possible, serving the only purpose to project pictures into the user's otherwise totally unobscured viewing area.

To illustrate this, let's have a look at one of the first projects that already implemented many of our desired features, in a display helmet for a flight simulator. The FOHMD (Fiber Optic Helmet Mounted Display) development was started as early as 1981 (first prototype 1984), and other than most other military applications that are simple data displays even today, it was the first really earnest approach to virtual environments.

Head Mounted Display System (FOHMD) for out-of-cockpit view of the Tornado Low Level Flight Simulator (photo courtesy of CAE)*

This project is especially important for our thematic, because it already involved advanced eye and motion tracking technologies to produce a stable virtual display, something not to be found in almost any other display or application even today.

---

* From 'Proceedings of the Workshop on integrative 3D visualization' [44]

Acceleration sensors and motion prediction to stabilize the images against head movements and compensate for trivialities like the one frame delay of any video system, are working perfectly here. I say this because these issues are still reappearing as problems in scientific papers. Apparently with all new technologies, many things tend to be reinvented many times.

The FOHMD uses a very high resolution display with external projectors from where the pictures are transported to the helmet by (expensive and fragile) glass fiber bundles. A high resolution, eye tracker operated picture inlay boosts resolution at the center of view to an equivalent of about 3000x3000 pixels. The field of view is 67^0 vertical, 127^0 horizontal (both eyes). This is still far more than anything you could buy in a shop now.

The computer installation to support it filled rooms of course, at that time. For our objective, we will need something many times lighter and less expensive. Even now, more than a decade later, there remains a lot more to develop until we get to a mass market product. While computers in general have taken giant leaps, several technologies necessary in vision simulator glasses, including optics, are way behind.

Several end user products with more or less limited capabilities have been built meanwhile. The available resolution in this price class is normally 600x800. A few professional or semi professional see-through glasses go up to 1280x1024, (Saab, NVIS, Cybermind) but are still a bit heavy, between 700 and 1300g. Some interesting developments have been announced by Microvision [15] (no details yet).

An overview of current products you'll find at page 125, and in [29] or [50]. For military technology, an interesting source is [37].

As stated, we would need about 1600x1200 pixels at least (better twice as much) and the device should weigh less than 100g. I am now confident that even less than 20g could finally be achieved after about one more decade of development.

There can be little doubt that Display devices for vision simulation glasses will be an important key technology with a huge market, especially as they will replace many traditional screen displays.

Optics

The human eye is quite good at seeing, isn't it? Optically, it's a mess. All the rest is done in the brain. I think this is very smart after all, and if nature had tried to give the eye a 'perfect' optical design, our ancestors would probably have long been eaten.
We should learn from this. The major problem with today's display goggles is that they are attempts to perfect optical design. This can't work well. It gets heavy, the field of view is small and the optics have literally to be screwed to the head.

As we can only see sharp in a little area around the center of view, a display with dynamic focusing and eye trackers does not require an optical design delivering an image that is crisp everywhere. It could be dynamically focused to the area being looked at. Optical distortion should as well be compensated electronically, resulting in simpler, lighter, and cheaper optics.
The system also has to compensate for image size changes due to dynamic focusing and for image position changes due to eye movements and position changes of the glasses towards the head. No fixtures should be necessary any more.

Dynamic focusing is indispensable anyway. Most 3D systems cause headache because the apparent stereoscopic distance seen, does not match the required eye focus adaptation. A dynamic focusing system should solve this problem by changing focus according to distances of objects seen in a virtual scene.
With laser scanners, focus is always (almost) crisp. It could be an advantage or a disadvantage, as it may also be irritating.

Finally, display glasses should not disturb normal sight at all, yet offer the capability to project virtual images over the entire field of view. This sounds ambitious but has already almost been achieved with quite conventional optics by placing the displays to the side or over the eyebrows, out of view, and mirroring the picture before the eyes. What's obvious, that a large viewing area definitely needs a curved mirror, unlike most current products.

Eye tracking

An accurate and artifact free positioning of virtual objects will be essential for the acceptance of the technology.

Parallactical errors, like angle differences that result from looking through the display more straight or more sideways, should also be eliminated.

As the display screens of our glasses are positioned near the eye, just simply looking about will move the pupilla relative to the display center and hence shift the entire image relative to the user's virtual position. This is absolutely not desirable, as it destroys the virtual steadiness of the picture and may be one cause of vertigo (dizziness).

Varying positions of the device towards the eye are to be expected as well, at least if we do not want to use these special fixations that are part of almost any current 'VR goggles' and are making these devices so unbearable.

So we will definitely need an eye tracking system, that determines the position of the pupil relative to the display. Such systems are state of the art. They simply use a camera and some image processing software to track the user's pupil.

They are still expensive in the professional area, but cheap versions have already been built even into cameras and camcorders to guide the autofocus (Canon EOS5, 1992, and Movieboy E1, 1994). What we need here, would have to be much faster and more accurate, but this is simply a matter of chip complexity.

Some applications of an eye-tracking system would be:

- Dynamic focusing in order to keep the eye adaptation aligned with the virtual distance of virtual objects or details of 3D scenes displayed.
- Creating picture inlays of higher resolution in the center of vision, as with the FOHMD flight simulator display.
- Exploiting eye-pointing in an advanced user interface.
- Automatic adaptation of the display system for best sight in any constellation.

Position sensors

In the field of relative position sensing, there are many companies with excellent products. The most common solutions use electromagnetic fields (Polhemus-sensor) or camera recognition of light spots connected to the target object. The latter are more accurate, but have difficulties with obstacles in the lightpath.

An intelligent visual orientation system based on image processing would promise a much better solution and could induce a radical change in the ergonomic constraints of virtual reality.

It will be absolutely important that virtual objects can be defined to appear at a certain place in real space, to stay there and only there, which means they have to disappear when the place or room is left, but also have to reappear when the same place is entered again.

This requires the recognition of environments, which can also be achieved by cameras attached to the Display, in conjunction with appropriate image processing.

Such a system would indeed also be useful as a very precise and direct position sensor, simply by exploiting the geometrical object data gained from the recognition process. While this is inherently a bit slow, things are greatly improved by adding some cheap and accurate acceleration sensors to keep track of fast movements, a method already used in aircraft simulators in order to improve the efficiency of triangulation and eye tracking systems.

It is important to note that visual orientation would not actually have to separate seen objects, or to identify their nature. The entire difficulty is reduced to the acquisition of 3-dimensional basic structures and their comparison and alignment with stored references. What we need is a system that can swiftly compare stored 3D structures, regardless of perspective.

Basic orientation can be simplified if we include a GPS system. Since the first introduction of the proposed ideas in 1993, GPS* receivers have become unexpectedly small and cheap.

---

* Global Positioning System. What your car navigator uses. Exploits timing differences of signals from a number of earth satellites.

With a differential GPS transmitter nearby, an accuracy of about 1m can be achieved. Yet we have to keep in mind that GPS reception cannot be expected everywhere, especially not indoors.

We could also envision systems that work like GPS but are confined to local environments, like buildings.

Illumination

If we want to generate natural looking objects, we have to show them in the same brightness and light color as the surrounding scene. We also have to simulate the same directional lighting according to light sources that the vision simulator cameras and software are able to identify. In order to do this, image analysis has to find the light sources either from directly seeing them, or, more difficult, from shadows that real objects are casting.

Then we have to decide whether virtual objects should cast shadows. This is not really simple, because we could only do this appropriately with the mask display, hence not sharp.

It may also be wise not to try this much realism, as it would always be safer if we could identify virtual objects as such.

Object merging – the mask display

Image presentation would be significantly improved, if we could cover up light from real world objects at those directions where virtual objects are located. This would simply keep the virtual objects from appearing translucent.

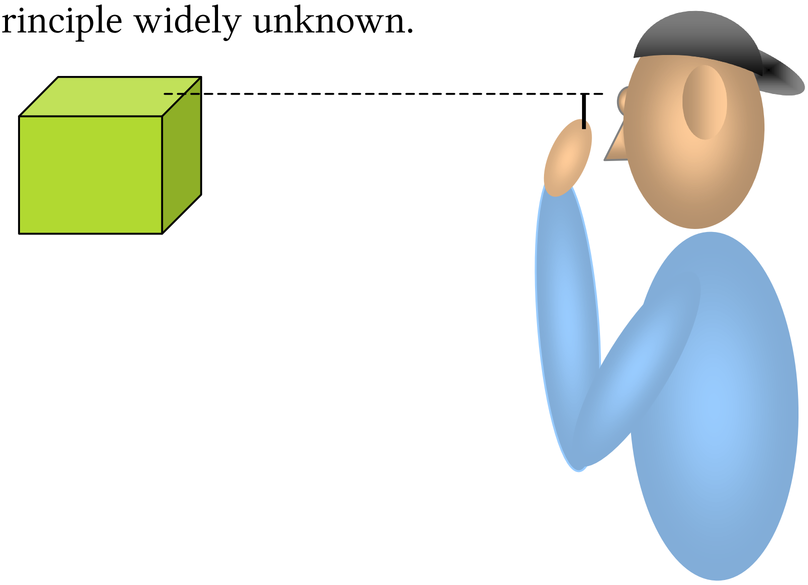
With constructions intercepting direct sight with optics, or even replacing it entirely with camera pictures [25],[77], it would be easy, but these approaches aren't acceptable for everyday use.

Another method would be 'subtracting' the brightness/color values of the real scene from the corresponding pixels of the virtual objects, an approach that works well to a certain extent and has yet been thoroughly investigated with projectors [19]. In our context, there would however be a frequent problem to cover exceedingly bright lights or objects in the background.

Real masking is, in principle, not so difficult to achieve. A cheap black-and-white transparent LCD display panel could dim direct sight at locations of virtual objects. This would however incur some light attenuation due to the polarizer filters necessary with current LCD technology, not optimal for night applications.

There are some other display technologies perhaps better suited for this, especially considered that a curved device would be nice to have. We'll discuss this later on, in the design chapter.

The fact that the mask display is always out of focus, e.g. appears with blurred edges, is not so much of a problem, as you'll see in some sample pictures below. I already proposed this in [2] but it's still a principle widely unknown.



Demonstrating the principle of the mask display is simple: just use a black chip (your thumb would also do for a first try) and move it, 2 cm before the eye, so that it just covers up exactly the edge of some distant object. It may be surprising that this results in a pretty sharp edge of total obstruction, even though an area of half transparent shadow remains beneath it. Use some thin masking object to verify that even tiny distant objects could be cut out quite selectively. The proper size of the mask also depends on your pupil diameter. A real mask display would have to be corrected for this, and also for edge diffraction effects.

Generating the correct masking shapes of course also requires an entire knowledge of the real scenery, hence cameras, taking pictures from approximately the locations of the user's eyes.

Simulated insertion of virtual images into a real scene

Original view of a perfectly white wall with a rectangular section cut out (covered by mask display with typical unsharp borders). This is from an original digital photo taken with a camera with an opening similar to the diameter of a human eye's lens (2.5 mm). The distance to the mask was like in a real display assembly (2.5cm).

A Virtual image has been inserted into the cutout section. This simulation was done with an image processing software, taking the image as 100% transparent. The image is slightly larger than the 100% black area, therefore a little of the background light shines through at the edges. Of course, if one chose a naturally darker background, this effect would be smaller and the image insertion would be less obvious.

In this example, the masking is applied on a more natural background

Obviously in this case, the masking process is hardly visible any more.
If the background texture is known (by the position cameras for example) we could also modify the edges of the displayed image, in order to compensate for the blur of the mask.

Foreground object overlay

Another issue we must address for a really hassle free augmented reality representation, is the appearance of real objects in front of a virtual object.

Otherwise, the resulting scenery could be very irritating. For example, a virtual keyboard should not hide the user's hands, or it would be unusable.

This was recognized a dozen years ago (e.g.[2]), but is still subject to basic research (e.g.[62]) as many of the themes we're discussing here, because augmented reality has been neglected so much during all this time.

For the virtual object to behave correctly, we have 2 possibilities:

1) to cut out the shape of the real object from the virtual object and the mask display, in order to uncover it, or
2) to overlay a picture of the real object, as taken from the position sensor cameras, over the picture of the virtual object.

The method of choice would usually be 1), at least as long as the image resolution and processing capabilities of the VR display are limited.

We could use a stereo camera to recognize objects with and cut out an appropriate shape from any virtual image supposed to stay in the background.

This is not really so difficult. No object would have to be really 'recognized', it would be entirely sufficient to spot structures that are positioned before the virtual object, starting from a cross correlation of two stereoscopic camera pictures.

In case of general orientation, there are some situations that may not so easily be dealt with. For example, we may consider an entirely monochrome, featureless surface: correlation would be useless here, the distance could at best be constructed from the borders, and some heuristic approaches will also be necessary for reasonable results.

With the foreground overlay of a hand, for example, we wouldn't have this kind of problems. There we always have enough features to exploit for a reliable object recognition.

Virtual keys and Icons

There will be many kinds of interaction possible with virtual objects. We will delve deeper into this issue in the design chapter. One essential difference to most of the current approaches, is that we don't rely on gestures, nor on tactile feedback. Gestures are complicated to detect. and tactile feedback needs gloves (who wants that).

A frequent type of interaction, in my opinion, will just be touching part of an object with a finger. This may be a handle or a virtual key. Feedback will not be tactile but visual and/or with sound.
The most important fact here: as only virtual objects or parts of them can react, we already know where to look for a finger to appear. We know the object coordinates, and we know exactly which pixels to watch in the left or right stereo picture of the orientation cameras. This facilitates the task entirely.

We could also use the eye trackers to determine if the user looks at a key, making it quite likely that he means to operate it . Having the object show a reaction like a key pressed in, or like sticking to the finger, is then quite simple.

Eye pointing alone is even more elegant. 'Staring' at an object can be a quite distinct action, safe enough for error free general use. The major advantages are that it works hands free, which makes it perfectly fit for all mobile applications at least, and that it allows for unambiguous pointing even at distant objects.

Eye pointing could be accompanied by blinking for example, allowing to implement very ambitious operation modes with sort of 'click' or 'double click' actions or drag and drop operations, entirely without using the hands.

Essentially, I think that eye operated modes, together with speech, mouse pen or finger inputs for text, will lead to a very convenient and efficient user interface.

Virtual start menus and task bars

Displaying any items fixed to the user's view rather than the environment, should normally be avoided. Exceptions are alphanumerical vector displays of data that cannot be brought into context with real objects in view, as well as start menus, virtual taskbars or the like, that will probably remain a method for primary access to applications.

Virtual bars and menus would usually carry one or more virtual keys. They would, contrary to 'normal' virtual objects, often be transparent, and would appear in areas not primarily inflicting with visual orientation. Above the center of view, for example.

Virtual taskbars could be activated in various ways, for example by touching a real key sitting right on the vision simulator, on the handles or at the edges of the glasses.

Another method would be voice commands, with all the problems involved using these in public places. The same applies for gestures.

A single virtual key staying at the edge of sight could also serve as a sort of a global 'start' button, as long as the user doesn't become allergic to it.

As stated, I regard virtual objects stuck before to the user's eyes as absolutely untypical for a viable augmented reality interface. In office, application windows should stay in (real) place. In maintenance, they should stick to the items repaired. For orientation, directions should be painted on the floor. And so on.

Only if the user is in motion and needs to see anything not related to physical objects in reach, it would be a necessity to fix a display object to his own field of view. Such an object should be a transparent vector graphics like overlay in the same virtual distance as the real scene, in order to distract as little as possible. There is no technical difficulty at all to build and program this, such items have been tried out and used since a long time, and they could easily be integrated into a full featured vision simulator as well. I'm mentioning them here for completeness.

Realization

As we have seen, a display for the virtual devices concept will be a fairly complex system, involving up to four microcameras with image processing (two scenic cameras and two eye trackers) and up to four displays (two high resolution color screens and two mask displays), not to mention acceleration sensors, possible acoustical interfaces, communication links etc. The miniaturization necessary won't be a problem, since there already are hosts of affordable microcameras and other very small devices available. The processing power necessary will still be high however, and will most likely have to be provided outside the interface itself.

An intermediate step, prior to the full implementation of all features, could be a system for stationary professional use like in medicine, where the display might still be acceptable at weights up to 300 grams, while the computing equipment could be placed in an extra unit. Yet with chips already used in mobile phones, one could probably build a battery operated wireless glasses assembly real soon now, weighing less than 100g.

Necessary Software for Vision Simulators

Object Presentation	Rendering Correction of display optics Adaptive resolution Adaptive dynamic focusing
Insertion	Image adaptation Masking
Positioning	Image/scene recognition Fast structural matching/search algorithm Acceleration or other sensor interface Anticipative positioning
Interaction	Finger or Hand recognition Eye tracking Voice Others
Operating System	Standardized software interface for object presentation and interaction (3D-Windows)

Hardware add-ons

Earpieces (headphones) and microphones are very important add-ons to any vision simulator.

The device could then be used as a versatile 'personal communicator', replacing handheld phones, notepads, music players, and, with the anyway included cameras, a camcorder.

It would probably be a good idea to integrate microphones into the earpieces, which would also result in an 'artificial head' microphone assembly (something already available under the name 'Soundman', actually working with the real head of the user instead of a dummy), yielding a perfect directional reproduction when listened to with headphones later on, and the possibility to derive normal surround sound channels with the help of some advanced audio processing.

The two cameras of the vision simulator would naturally provide for 3D pictures, something not yet very easy to use for video production, but maybe it could become of interest with future developments in image processing and display technology. The vision simulator technology will promote 3D video anyway.

Really useful would be the already mentioned buttons right at the vision simulator, that could be used to trigger some actions: just touch your glasses and the options menu for the home appliances appears in front of you, for example. Or, even better, use your eyes to point at something and start an action on it by pressing one of the buttons.

It's trivial that we will need some sophisticated radio, cellphone and network connections. The vision simulator has to talk to its computing unit, to local networks and to other vision simulators.

With RFID and active tag technologies, a dedicated radio transmission unit for these could also be helpful. Very simple RFID chips of course require far too much RF power to react at all. With smarter units, solar powered active price tags in supermarkets for example, that could carry their own radio or infrared transmitters and receivers, establishing a communication would be no problem. It would depend on the provider of these units, who could communicate with them and to what extent.

Applications

As stated, the technology outlined could replace many of the information technology devices currently available, the PC screen being but the most obvious.

TVs are another very obvious candidate, of course, then all kinds of control panels, and many more.

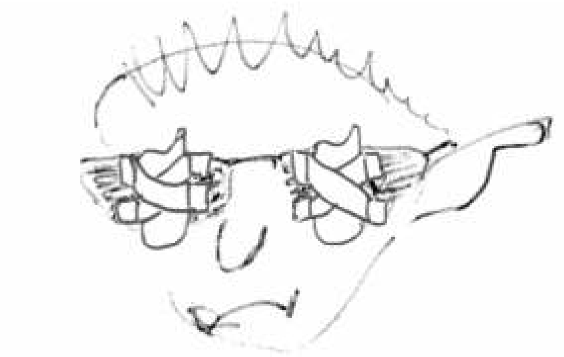
In the following I will illustrate several types of example applications. Some more will follow in the media chapter.

The most vivid overview may be the fiction part that follows this chapter. There are many remarks and sidekicks in it scratching themes that might inspire your fantasy. I hope you like it.

The "laptop back" (humpbackus aeroterminalus),
a civilisatory disease

Goggle phones

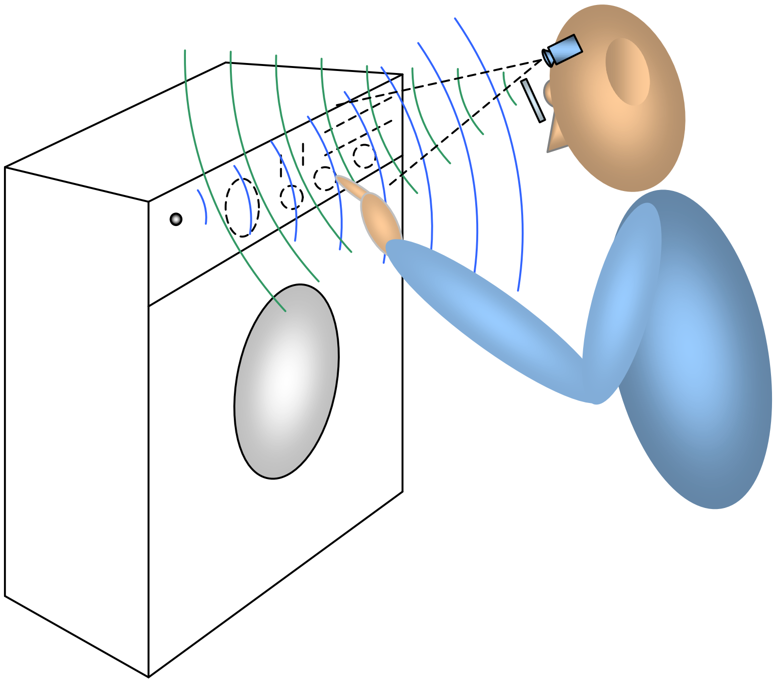
Let's start with an application implementing only part of the functionality, as this is easier for a first product and can be a technology test bed that is sophisticated but still manageable. Just have a look at these very small cellphones that are now equipped with cameras and a lot more things. Their hardware would already be light enough to fit into some (not yet perfect) glasses, it's only built the wrong way.



During his sabbatical, John successfully demonstrated the feasibility of light weight display glasses

No joke at all: it is absolutely possible with available technology right now, to integrate a complete mobile phone into a pair of glasses at fairly convenient size and weight (≈50 grams). A one-eyed, fixed transparent display would be sufficient. Considering the increasing habit of wearing micro headsets even despite of the inconvenience with the cable, it's quite conceivable that such a device would encounter a great demand. Iris recognition could replace passwords, eye pointing could be used for menus and dialing, and the virtual screen size possible would enable full featured web surfing, where users could click on links by just looking at them, just to mention a few of the possibilities.
Such an application obviously lacks about all of the environmental integration that we are talking about here, yet it allows to develop, test and improve a lot of the underlying technologies in the context of a real product, that would already pave the way and cause the demand for more sophisticated ones.

Virtual control panels



This is one example of the hardware savings potential of vision simulators. The washing machine has no classical control panel. Instead it has an empty surface, possibly with some markers to help the position sensor cameras. An infrared or radio emitter and receiver element are the only features required.
The machine sends an image of its virtual control panel to the vision simulator, which then generates this picture, in the correct perspective and position.

A very important advantage of virtual control panels: they can be customized, can be much larger than the actual hardware.
Using a virtual control panel follows straightforward the already depicted virtual key paradigm. Here, the real (empty) panel provides for tactile feedback, so it's feeling real some way.

Key pressing or other actions are reported to the washing machine by transmitting infrared or other wireless signals. So this is a two-way communication.

The control panel does not have to stay in place. Once activated, we could take it off and operate it from a more convenient position.
Radio operated panels could be made to appear anywhere desired, so one could operate any equipment in the house without leaving the chair.

Remote remotes

TVs today hardly have any control buttons any more. Everything went into the remote.

That's for a reason, obviously, it saves money and nobody wants to run to the TV to switch stations anyway.

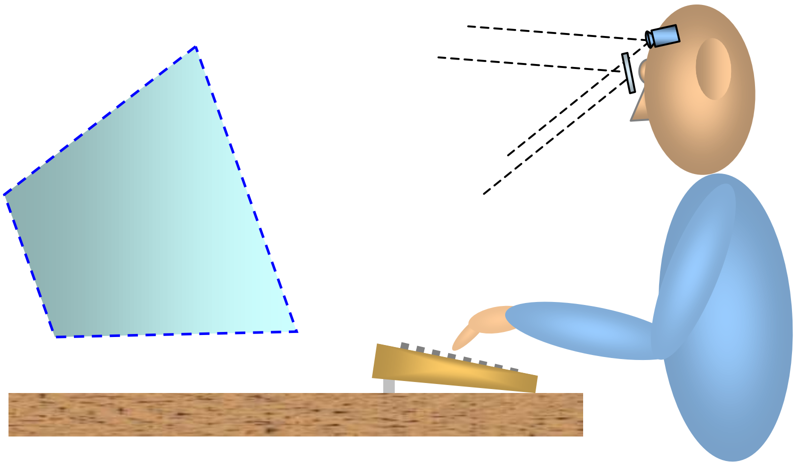
Now would we equip a virtual TV with a remote control ? - I guess not really, because these remotes tend to clutter up tables. The better way would be to place some virtual controls beneath the display window, like with an ancient TV set, but to operate them remotely.

Eye pointing would be perfect for this purpose. Eye trackers could determine the user's viewing direction quite accurately. When looking at a key it could get highlighted, saying "don't stare at me or I'll get depressed". So it may be.

Our vision simulator could also serve as a universal remote control for all kinds of classical devices. A lot better than any we have now. This may even be one of the most frequent applications for the first time.

All these virtual control functions could be implemented in the 'goggle phones' mentioned above, by just replacing the eye operated virtual control panel of the phone with that of another device and adding an infrared and maybe a radio transmitter.

Mixed 'virtuality'



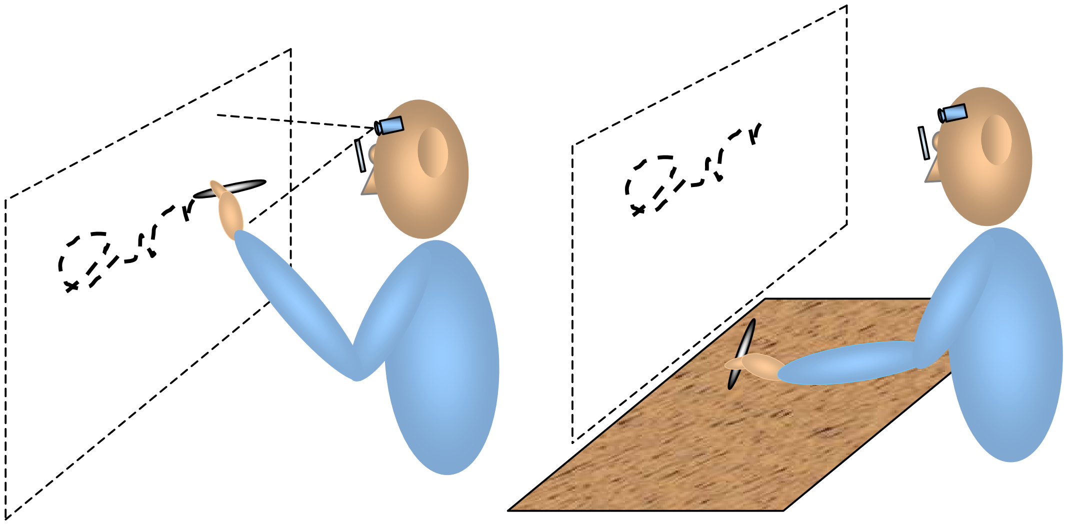
Combining a virtual monitor with a real keyboard

Virtual devices are not a dogma. It would often be smart to combine real parts with virtual ones. A keyboard, for example, can be clumsy to operate if only virtual. With a foldable keyboard, we would still have a very portable solution.
Pens or mice operated on a real surface are other examples for this approach. Game applications need some real joysticks, grips for virtual swords, bats, golf clubs. Certainly, the gadget industry won't die out.

No Force

Some people think that omnipresent tactile feedback would be indispensable for augmented reality. Here I strongly disagree. Not only don't we need this for the vast majority of applications, it would also require the use of gloves, something absolutely unacceptable for everyday use. There are several professional applications that use force feedback and sometimes also hand movement recognition built into special gloves, but these are mostly virtual – not augmented – reality, and there's already enough literature about them. The really revolutionary feature of augmented reality will be eye pointing, hands-free.

Virtual writing

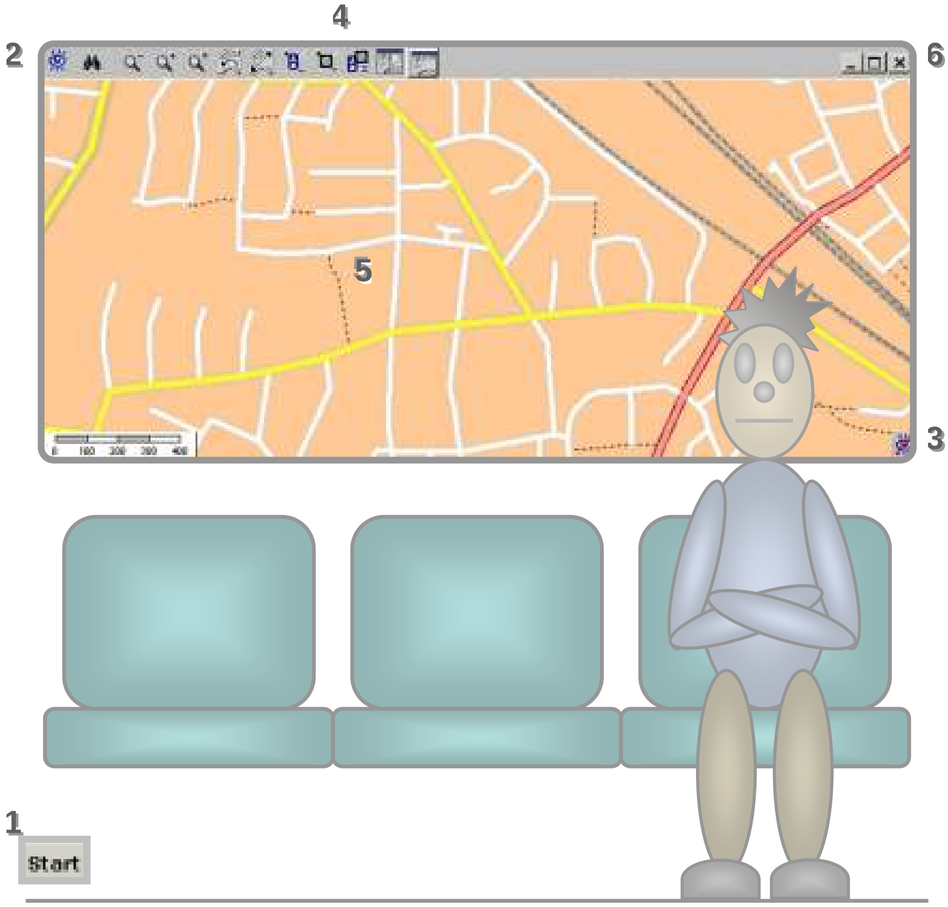


This is one of the most obvious application examples :

A special pen (with acceleration sensors and perhaps a flashing LED or other markers) could be used to ensure easy detection of the user's hand movements to the position sensor cameras. Virtual lines are drawn when the pen is moved inside the virtual paper plane.
This writing plane would normally better be chosen to fit to a real desktop's surface, or at least to a drawing board, to provide a tactile surface.
Another possibility is to write on a desk with a mouse like pen or a mouse, but to view the results in the virtual window. This would work exactly like a normal mouse with a PC. It would be intuitive, hardly need any processing power and be very reliable. The mouse or pen could also be switched between multiple virtual PC application windows, by eye pointing for example.

We could also think about mixed mode operation, moving the cursor with the eyes, yet clicking and writing with the pen.

Under way

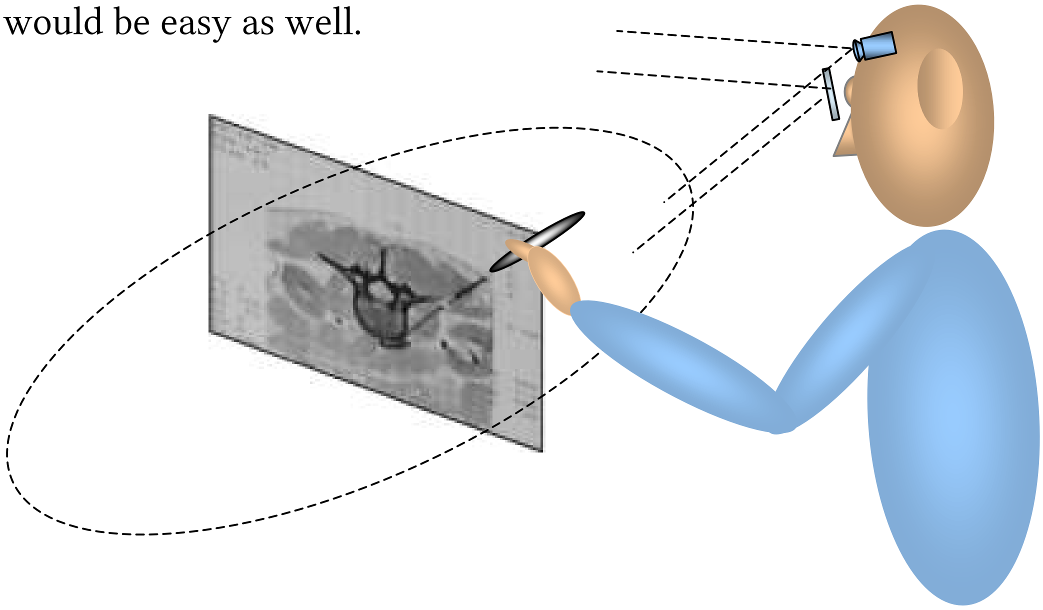


This is a nice example to illustrate the use of simple eye operated applications: suppose you're riding the subway. Wondering where to go next, a city map would come in handy. So you shortly peer at that start icon at the edge of the glasses' display range (1), select the map from the transparent menu popping up (not shown here), place the program window by shortly peering at its desired corners (2,3), and the program starts. The guy facing you is still in front as the window has been 'looked' onto the wall.

Now you may select functions from the icon menu (4) or drag and zoom the map (5). Eye lock can be engaged or quit by blinking. Having looked up your target, you close the application by peering at the exit icon (6), or you just leave the train and let others wipe up after you (a joke, that's the virtue of virtual devices, they don't litter). You think the guy there may wonder about your strange looking about? Don't worry, the mask displayed for the program window will shield your eyes from him.

A medical pilot application

A quite complete implementation of the technology, yet with lower requirements at least for image resolution and orientation, can be accomplished with medical applications. Displaying images from online diagnostics at their very location of origin results in substantial ergonomic advantages with surgery under NMR (magnetic resonance), CT (computer tomography) and ultrasound imaging.
- Open NMR, Ultrasound and fast CT deliver black&white images of typically 128x128...512x512 pixels. So relatively low resolution, even black and white displays are sufficient.
- The objects are flat (simply cross sections). So we need but simple algorithms for rendering. Showing multiple cross sections would be easy as well.

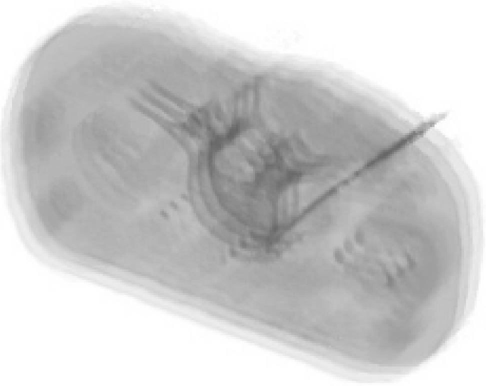


The picture shows a cross section of the human torso, taken with an NMR, that has been set to a perspective similar to a virtual look 'inside' a real patient lying in front of the surgeon. The bone structure of the vertebrae appears in black. The NMR image is from a real minimally invasive operation, where the tiny instrument used is also visible (slanted black line to the right).
The operation shown was carried out under continuous NMR control. A virtual display as visualized here, would have largely facilitated the operation, as it would have eliminated the necessity to repeatedly look up to the monitor and back to the instrument.

Several more or less similar appearing applications have meanwhile been implemented. Nevertheless, dynamic head position and eye tracking have been thoroughly addressed but very recently in [4] (an application with offline data but nevertheless very close to our general concept of virtual devices).

In such an application with a single surgeon carrying out minimally invasive therapy, a simple 2D or 3D flatscreen floating just above the operation field could also be used, as in this case no cameras need to take up the operation field or the surgeon's hands and there may not be much to see that this screen could cover up. As a simple solution, this would at least deliver a convenient viewing position for the surgeon, but in case we want a realistic perspective, we would still have to measure the surgeon's head position and calculate the image accordingly.

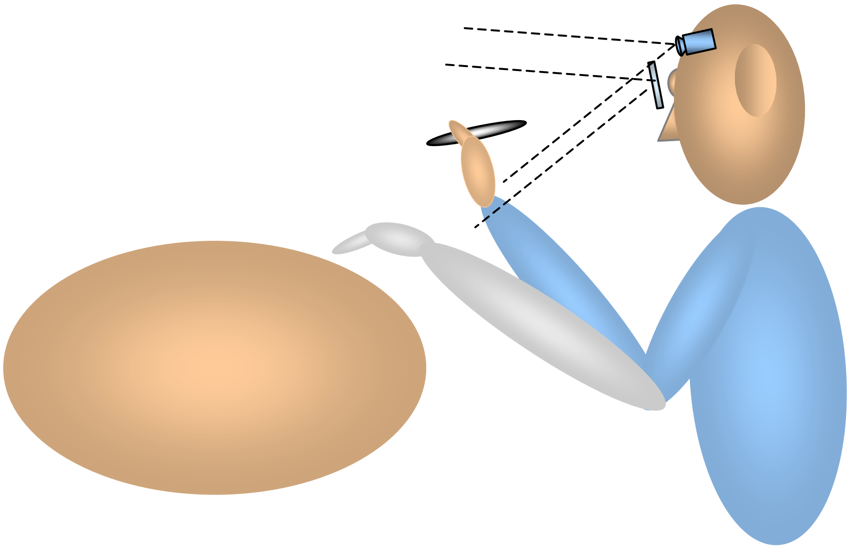
Another option would be to show several adjacent layers simultaneously in a transparent way, giving the surgeon real 3D information with very little computing effort. This only works with a display like ours, that allows for dynamic looking around by head movements.



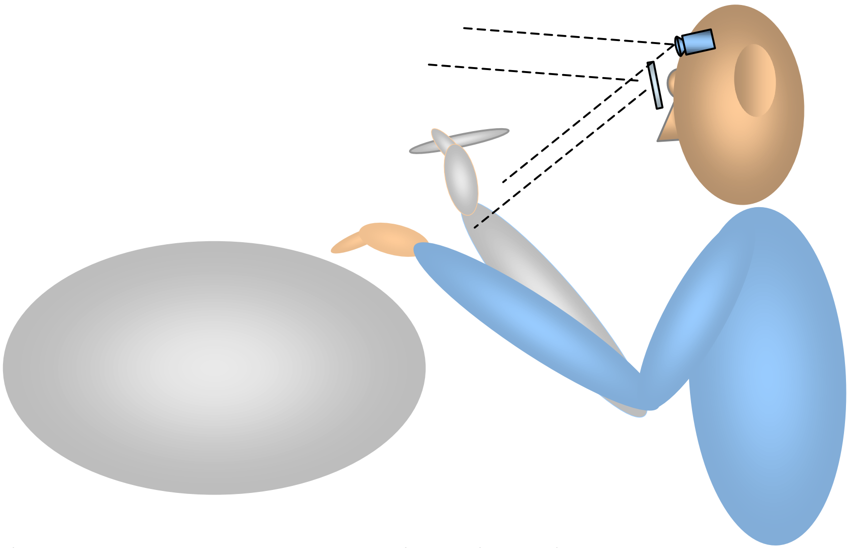
It is a very important fact indeed that dynamic image presentation can be used to exploit the vast 3D processing abilities of our brains. For example, if we would use a normal ultra sound device and project its echo image just into the body at originating location and in real time, the beam 'flowing' around inner structures would just appear to depict them three dimensional, if it can be moved rapidly enough to let these images appear timely related in our visual system.

It is also possible to use raw data from tomographic scans, have them perspectively displayed in a transparent way, and have this entire 'pixel cloud' either move itself or being viewed like a virtual object that the viewer can look at from different angles. Without rendering any surfaces at all, a perfect 3D impression will emerge (Dornier medical once experimented with this).

Cooperative telemanipulation
(remote surgery)

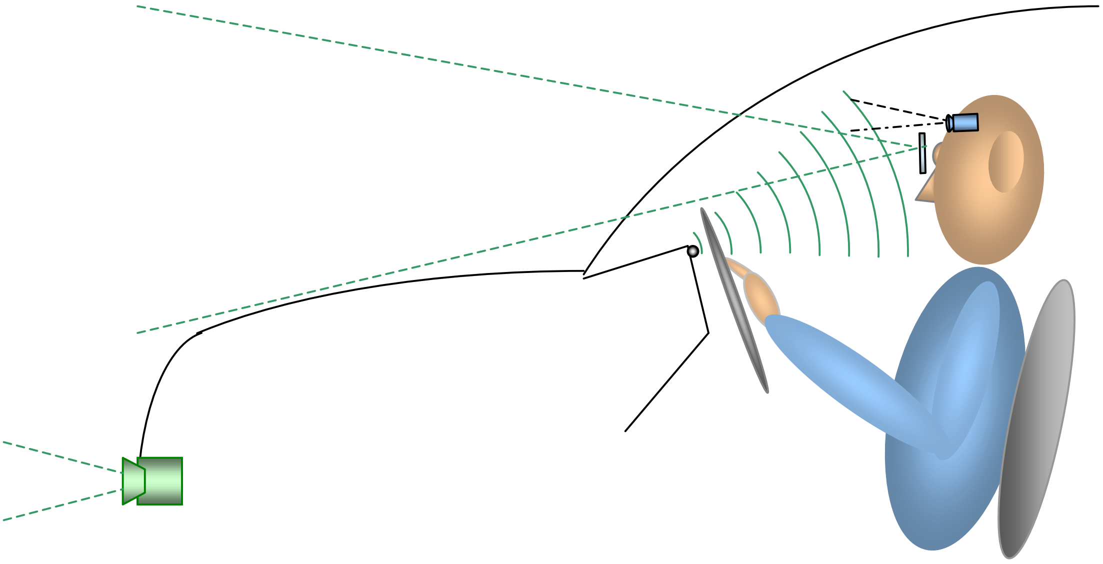


This is a quite complicated application, in that both participants here see their own hands as well as the other's. The operating surgeon sees the patient (a dinosaur egg) directly, and he can also see the hands of the consulting surgeon.



The consulting surgeon sees exactly what the operator sees, and he cannot only give acoustic advice, but also point directly to parts of the operating area. His hands are filmed by his head position cameras and the picture is transmitted to the operator.
So this is truly a two-way communication both with speech and images.

Integrating external infrared or radar pictures
(virtual headup display)



An example how to use a vision simulator to save expensive hardware. A radar sensor has been mounted to the car's front (a far infrared camera could also be used and mounted closer to the driver's viewpoint; both technologies have one big advantage, to be able to see through fog). Instead of a headup display, the vision simulator does the job. The picture is transmitted by an IR or radio emitter in the dashboard.
Perspective correction information for the virtual overhead display is provided by the position sensor cameras

The car's instrument panel could also largely be replaced by a virtual version of course, like with the washing machine example. A routing guide could also be transmitted to the vision simulator and seamlessly be integrated into the field of view, probably much better than with an LCD device mounted somewhere on the panel, and also better than with a classical headup display, because any information could seamlessly be integrated into the real landscape.

Intelligent sunglasses

Currently available solutions for the mask display will not be suited for night driving because of their high intrinsic absorption. Nevertheless, if we find a better technology, the mask display could also serve as an anti blinding mechanism, selectively dimming the headlights of approaching cars.

The position sensor cameras would provide information about extraordinarily blinding lights, which we would then use to dim appropriate areas in the mask displays.

Correcting for the slightly different perspectives of the cameras and the user is very easy, as we only need the distance information that comes from correlating the camera images.

This way, blinding lights could be dimmed without affecting the entire field of view. With such an equipment, one could perhaps even directly view into the sun (nevertheless, don't try this !).

Especially when driving, anybody would probably highly appreciate such a feature.

For day driving, it would of course work with a simple LCD mask display already.

Avionics applications

Avionics exhibit some similarities to the automotive application presented, yet here the vision simulator could also replace the headset, even with advanced features like noise cancellation.

Nowadays, advanced color screen displays are replacing classical instruments even in general aviation. These units are a major advancement, but they are pretty expensive.

As headup displays are already appearing in cars however, these will also be seen even in low budget general aviation airplanes.

If everybody would use a vision simulator as discussed here, displaying would almost entirely be done with it. The only hardware add-on for an airplane to keep up to date with the most modern features, would then be a data link that transmits all information from the 'classical' instruments, and also receives some input from the vision simulator.

The vision simulator could generate a very versatile display, head-up style for example, and a lot more.
It still needs software to process instrument data and operator inputs, to generate an ergonomic display (instruments, maps etc.).
Some orientation features, like virtual maps, GPS, etc. would just be part of the vision simulator anyway, but a necessary addition would be electronic air maps, approach charts, ATIS and other avionics information, and all software for processing and display.

Radio hardware should of course remain part of the airplane, it requires a lot of RF power and special hardware features that won't make sense in a vision simulator. The same applies to most of the 'classical' instrumentation. We would not even have to drop any of these to make room for new displays. So the entire concept would be one of great redundancy. Which doesn't mean that it would come for free. Instruments and sensors need data links to be integrated into an advanced system.

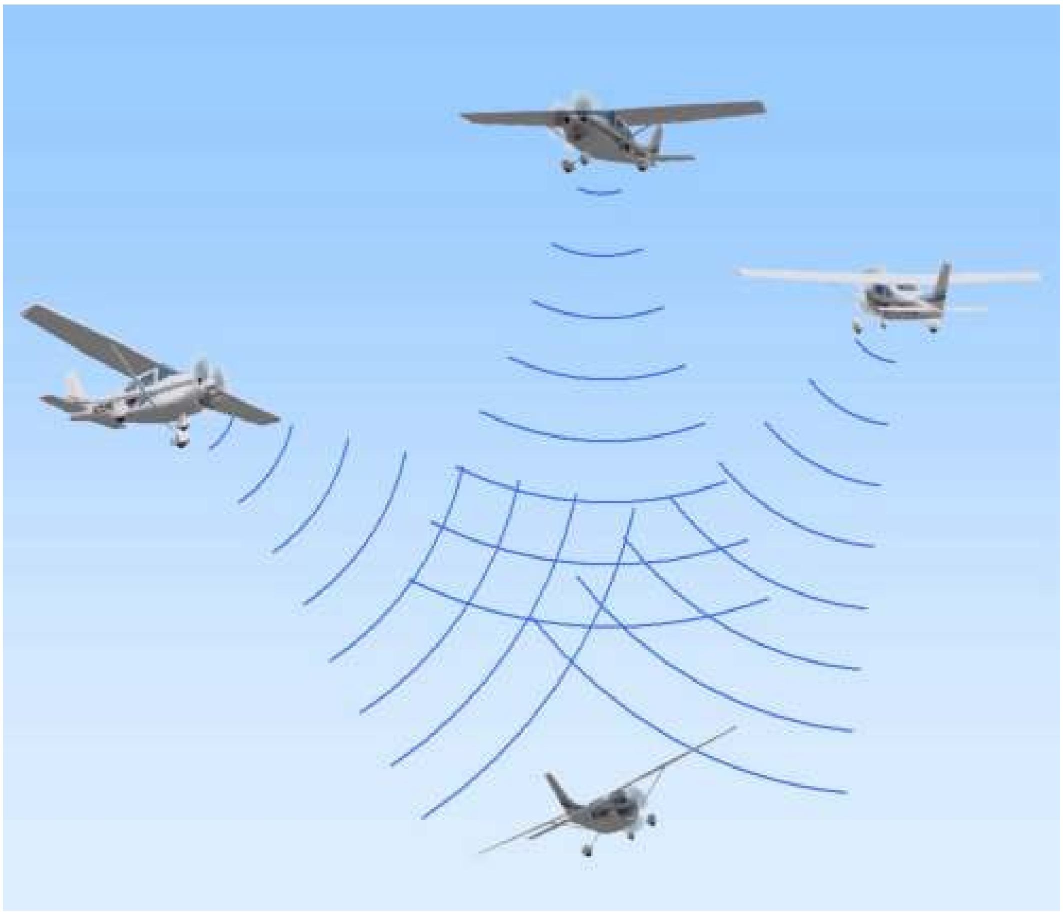
What's most important: the virtual display would not only save the costs for the extra screens, it would greatly surpass them.
Blocking direct sight, like huge conventional instrument boards tend to do, would be a thing of the past. We could use images from cameras outside the airplane and generate a virtual view through the walls, introducing an entirely new level of visual overview and situation awareness.

Maps, trajectories, also a predicted flight path of our own airplane and of others, could just be projected into the landscape and the skies, and be visible through walls and wings.
First steps in this direction have been tried, in [38] for example, but anything like this will need a good affordable vision simulator to get into common use.

Virtual Radar

Every airplane has a so-called transponder, a radio transmitter that continuously transmits the aircraft's altitude and ID code. This usually helps radar devices to classify echoes and to generate an informative display for traffic control.

There are plans to have transponders transmit GPS data as well, because then it would be simple to get a traffic overview without a radar, so a relatively simple device could provide any pilot with a virtual radar image of any traffic within radio range.



A vision simulator could make this up as a really 3-dimensional picture, letting the pilot see other aircraft right through the cabin wall or the wings, removing any of the usual dead angles.

The aircraft itself would not even need to have a built-in GPS, as the vision simulator would already have one, yet I think it would be better for safety (redundancy).
Speaking of redundancy: Multi function color screens always come in pairs, to have a backup if one of them fails. With Vision simulators generally available and probably cheap enough (this would not be specialized GA stuff but consumer electronics), having some spare ones wouldn't be a problem at all. It would therefore come out a lot cheaper than today's high end instrumentation, without any loss of safety.

Virtual video conferencing

Vision simulators could easily be used as a videophone. If the user places himself in front of a mirror, the environment cameras could pick up his image and transmit it. In turn, his own simulator could display the image of the communication counterpart(s).

The impression however would not be ideal: in order to display the communication partner(s), a central part of the mask display would have to be darkened, otherwise the user would see himself and his partner as a mixed image. This would cause his vision simulator to look like dark sunglasses.

A simple solution is to use the eye tracker picture. We would anyway need a very good eye tracker for a good vision simulator, one that should cover the entire eye. So this picture could easily be inserted into a face picture, resulting in a true and real reconstruction of what is actually there.

The only problem here could be that we probably use infrared lighting for the eye tracker, and other light sources could be dark. Hence, color wouldn't necessarily be accurate. We may have to reconstruct color. Not a big problem, if we only once before provide the system with a true color picture of the user's eyes.

In a room with several installed cameras, an entire array perhaps, more options would be available that I will cover in the virtual media part of this book.

Simple speech connections will undoubtfully remain a frequent application as well, especially when walking, driving etc.

Our hardware would make this easy, we only have to consider how to dial. If not by voice, we would probably use some virtual eye operated keys in the field of view.

The odd way ("meet my avatar")

Another Method of virtual meeting would be to use avatars:

The device could be taught to 'know' how its bearer would look. It could also pick up eye movements just by the eye trackers. If we would add some micro cameras to record the face (just near field down from the lower end of the glasses) and to eliminate the dead angle in gesture recognition, a sufficiently intelligent software could generate a quite vivid and exact simulation of the user's actual look, even without the obstruction by the simulator.

It would even be possible to synthesize position and environment. The user could, for example, lie on his back at a beach, while the software would depict him sitting at a desk in an appropriate business outfit. It could even compensate physiognomy changes that result from non vertical positions. It could also shave its user, and so on...

Quite a peculiar concept, but anyway, physiognomy, look and gesture could become entirely realistic, so the main objective of video conferencing would be achieved.

Needless to say that such a concept could be expanded to several partners simultaneously, that they could be arranged in a virtual conference room independent of their actual being sitting, standing, on their outfit or environment. We could even generate perfect eye contact between all participants, something extremely difficult with any conventional teleconferencing system.

I have very little doubt that this can be achieved perfectly enough to be totally acceptable as a professional tool. There already are some experimental teleconferencing systems using image processing and perspective correction to restore eye contact [10], so the solution discussed is not so surprising at all.

The downside: one could also abuse such a system, not only to look better, but to deceive people in many ways. Anyway, this is not too bad if people know about these possibilities and are prepared to be suspicious about what they see. A manufactured communication avatar is not too different from a faked document, is it?

Another nice side effect of avatar imaging: Bandwidth requirements are extremely small. Indeed, the capacity of a simple classical telephone line may suffice to transport both good sound and a crisp and vivid image. With today's rapid increase of available bandwidth in most of the communication channels available, this may not seem so important. Nevertheless, bandwidth will always remain a cost factor, and the difference between a really good TV transmission and what has just been proposed here, is tremendous. So in case of wireless communications, or a satellite link, this advantage may still count alot.

Corrective glasses

Quite obviously, a vision simulator could in some cases replace corrective glasses, by overlaying sharp pictures from the position cameras onto the real scene. The different perspective could be mended by image processing, and an advanced vision simulator should be able to position the overlay precisely, but it isn't foreseeable how such an overlay would look and feel. It could be better than progressive glasses just for reading, but using this for short sightedness would involve using overlays all the time. The better way in my opinion would be to add some individual corrective lenses to the sim for far adaptation, which is not a problem, and use the vision simulator's features for reading only.

Today there are also advanced surgical methods for eye correction. Correcting just short sightedness by laser ablation has literally become a fashion. Yet this doesn't help with the reading problems due to lack of near adaptation, experienced by most people over the age of 50. It appears that replacing the entire lens with a soft synthetically one might entirely restore adaptation. There aren't enough long term experiences with this so far, but maybe in the not-too distant future it will be possible to correct any sight problem the surgical way. Glasses could then become a thing of the past maybe. So you may ask, why introduce them again by wearing a vision simulator? Anyway, there's no other way so far to make a minimized man machine interface; and after all, why are so many people walking around with sunglasses even at night ?? There are even at least two products integrating an MP3 player with sunglasses right now. All a matter of fashion.

Seeing with the ears

One class of partial implementations that we shouldn't forget concerns sound: generating spatial sound impressions complying with the pictures of virtual objects is something that's necessary for a vivid impression with both virtual devices and media, and definitely needs to be a component of the vision simulator's software. This leads to another thought: we could also generate sounds just from the images, letting things start to hum, whistle or crackle.

Entirely weird? Not at all. There are people who can't see but have learned to orientate themselves just by making noises and listen to the echoes. It can work extremely well. Even a special generator for these echo blimps is already available.

Yet couldn't it allow for an even better sound image if we just use the camera images and generate spatial sound as if the objects were actively emitting it? Spatial discrimination improves a lot if we can turn our head and listen for the changing sounds. Many things already have a very characteristic sound image (knock at them and you know). We could also encode object color, size and speed with different sounds. So vision simulator technology could even provide for a very efficient and affordable implementation of solutions for visually impaired people. Imagine the cameras recognizing signs and inscriptions from ordinary objects and make them sound different. Eye or finger pointing could then tell the device to read the text. The cameras could recognize persons from a distance, recognize many other things and just tell about them, and so on.

This is all not as far fetched as it may first appear, and indeed there is already at least one project with similar objectives [59].

Let's consider this further with the vision simulator in mind: many visually impaired people are still able to use eye pointing to a certain degree of accuracy. Imagine different applications windows arranged in 3D space clearly separated, sounding according to status or content, also clearly separable, and reacting to pointing by reading their contents or accepting entries, perhaps with a mouse pen. Imagine a person wearing the specialized vision simulator device taking a book, opening it, and the position cameras would right away start 'reading' it loud to the earphones. With eye pointing, the device would also know when to start reading and if at the left or the right page. In addition, finger pointing could be used to select certain paragraphs or sentences.

All absolutely intuitive, ergonomical and natural.

Using the position cameras for general orientation needs some additional thoughts. Simply recognizing locations, what we typically require of a vision simulator, does not require to identify objects in a way that would determine their nature or meaning. It does not even require separating different natural objects from each other. Meaningless scene details are entirely enough.

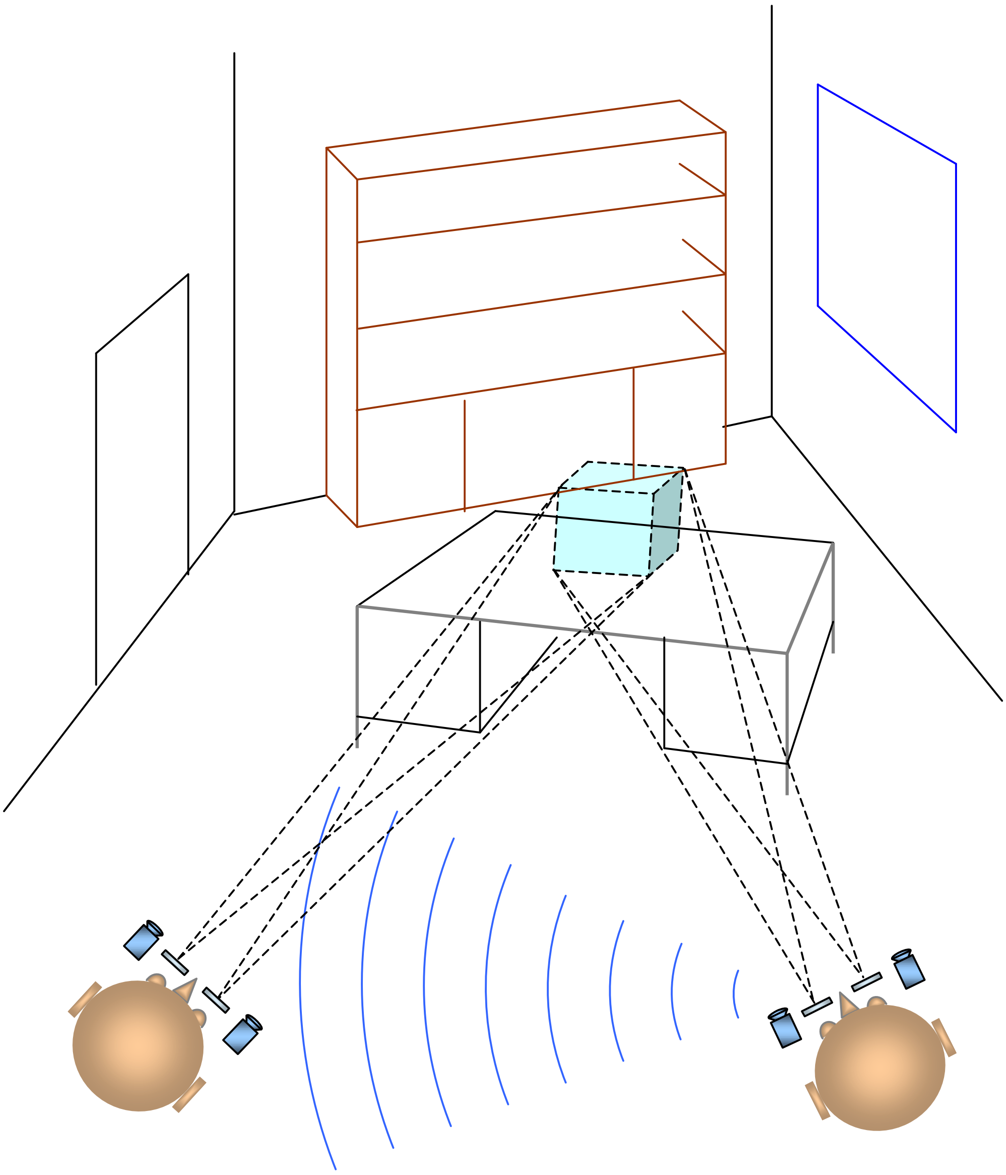
Image recognition in the classical sense, as we would require for an intelligent orientation support system, is different. We would need a lot more sophisticated software and probably some support in case of ambiguities in the image interpretation (a glass door, for example is a difficulty), or in case of insufficient light. In these situations, an additional depth sensor - maybe a simple ultrasound device - would help quite a lot. Many things to develop, yet certainly way easier than creating the complete perfect vision simulator, especially because we won't need those high end displays here.

There is also an amazing project with an entirely different approach, as it translates pixel patterns directly into spectral patterns, sort of a sweep signal [58]. Surprisingly, people can learn to interpret these patterns very well, and actually 'see' the image that a camera delivers. It's astonishing that this works at all, and once more it reveals how flexible a human brain can adapt to the environment, although the learning curve seems to be pretty steep.

Currently this works with a single camera and is about to be extended to stereo view and more. The most fascinating development is a version that works on a camera mobile phone and is nothing but a free software for the phone. Minimum cost and maximum effect.

This may as well be a basis to add some of the features discussed above, like automatic object recognition or classification, automatic reading, spatial sound impressions and so on. Just imagine what could be achieved if mobile phones were already built like vision simulators. This could lead to fascinating products very fast, even in numbers allowing for a commercial amortization, and being based on a widely used, common vision simulator technology, thiese would be very affordable as well.

Sharing a virtual object

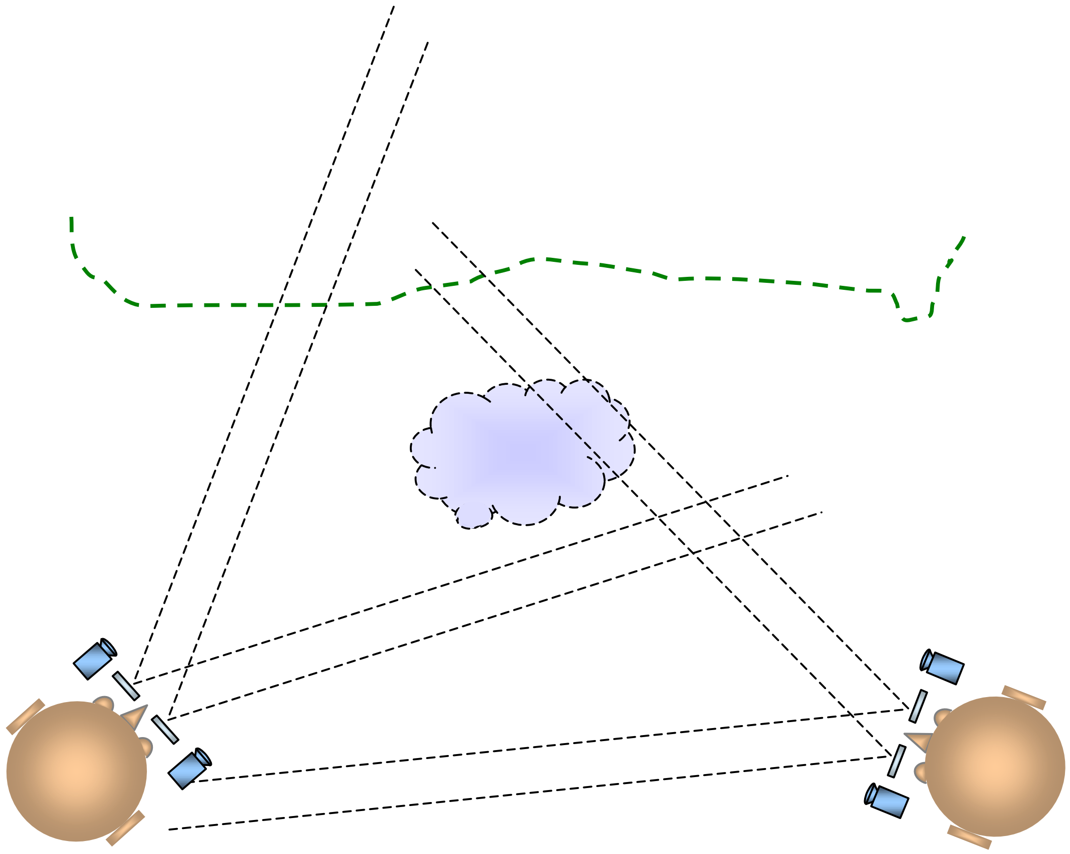


This is the most powerful and versatile feature of really complete vision simulators as I see them:

The right user has defined himself an object (maybe a virtual monitor) sitting on top of a real desktop in an office. This object is real to him, stays fixed to the real environment and is obviously no more accessible if he leaves the room (like any real object also wouldn't).

He can make this object visible and accessible to others by selecting this feature (how that is done, is defined by the 3D operating system he uses), then his vision device transmits the object's data by infrared, radio or similar, so that others can receive it. The receiver's vision simulator then produces a picture of the virtual device in the perspective correct for him. It depends on the sharing allowances given, if the other user can also operate or even move the device. In game applications, he could even steal it.

Virtual holography ('holodeck')



It is self evident that a vision simulator with appropriate software could generate the illusion of totally immersed 'real' objects, as with a 'holodeck' that everybody knows from the Star Trek series. (of course, without tactile feedback or the danger of getting hurt; the simulation of real intelligent beings from fairy tales will also take some more time to develop).

Virtual showcase, Virtual exhibition, public objects

In a radio networked environment, anybody could set up a node that transmits object descriptions and possibly communicates with vision simulators passing by.
Imagine virtual window dressings, or an entirely virtual art exhibition, especially useful for sculptures.

The nodes delivering the object descriptions also have to transmit some orientational data, descriptions of real objects in the same environment for example, and virtual object position relative to them. They could even serve differential GPS data to further facilitate the task (information about the actual deviation of GPS at location).

Objects Types

Objects may be static, moving, interactive, restricted to certain users, for free, or payware. They may be shared by groups of users or just be generated for one.
Detailed type descriptions should allow to determine if a certain object should be accessed automatically, by user command or not at all.

Object allowances

Sometimes it needs to be secured that public objects do not cause hazards. Street traffic for example has to be protected from such interference.
Objects should also not be allowed to molest anybody.
Therefore is will be necessary to establish rules and licensing structures to inhibit the abuse of public objects.
I will address these issues more detailed in the Design chapter.

Some more applications:

- Virtual advertising towers, placards, posts etc.
- Virtual attendants, guides, policemen, road runners... (should only be possible if authorized).
- Location dependent information: signs, road signs, labels and popup windows of all kinds.
- Virtual environments, guidance etc. in shops.
- Virtual road and town guides.
- Virtual Internet everywhere, with all services imaginable.
- Virtual media: this is a wide area also comprising novel recording technologies. It will enable truly holographic films and many other things. See media chapter.
- Virtual cinemas (would it really make sense? sound, effects and snacks could still be real; the virtual display makes perfect 3D and surround possible; more on this in the media chapter).
- Virtual training and job guidance, showing how to do things, hands-on, with explanations, virtual hands and tools showing techniques, even intelligent observation and commenting of user actions.
- And still many more

Additional applications will appear in the fiction and the design chapter of this book.

Some others - that don't need the fully implemented vision simulator - are outlined in [3].

Markets

For highly complex products, there are typically three major market segments that can provide for a stepwise amortization of development costs for new technology: Military (partly), Office and Home.

Other professional applications (medical, industrial, architectural etc.) will only create niche markets in comparison.

In military applications, costs are typically not most relevant. With defense budgets sometimes under pressure, it is also important to know that simple pilot applications could easily be implemented in medical environments. This could well be helpful to lay the foundations for a new technology trend.

Not only a trend though. Expect a revolution.

Yet basic technology developments for such markets only, are generally not recommendable. Although medical applications are an ideal area of experimentation, also with lower technical requirements, and even though there are some military products that have to be rugged but can be kept relatively simple, the billions to be invested into the new technology may only be amortized with millions of pieces sold. Being 100 or 1000 times smaller than the office or home markets, specialized segments might well accept higher product costs, but scarcely 100 or 1000 times as high.
I won't try to guess any numbers here, but from the applications it is obvious that the market impact of this technology will be extreme.
Many major manufacturers will have to rethink about their abstinence in certain consumer or business relevant markets, especially the camcorder or the mobile phone market, where a lot of manufacturing and component technologies for future vision simulators are located. The same applies to the ignorance about certain components, e.g. high resolution micro displays.

Ecological and economical side effects

The ecological and economical advantages of virtual objects technology will be immense.
Today, fast changing technology obsoletes equipment every 2-5 years. This will not be different with VR equipment, but with our virtual devices concept, only very small amounts of physical equipment will be discarded, while virtual devices are simply reprogrammed. With such an interface, virtual paper would finally have the potential to really replace its physical counterpart, because ergonomic constraints are far better met than with any present concepts.

In conjunction with a communications network supporting virtual offices and work at home, not only office buildings will become obsolete in great parts, but working environments will as well be reduced to a minimum. Virtual devices will replace traditional hardware and filing concepts, and working conditions will become independent of location.
Given these effects, and the even broader impact on everyday applications, this technology will be for the 21st century what the car was for the 20th.

Anything we're considering here, is a logical part of the advent of information technology, or the 2nd industrial revolution. It's already a fact that even though we should have learned something, some mistakes of the first industrial revolution are obviously being repeated.

As any new technology, this one may require major conversions. Well, it's trivial that people do invent technology to save work or resources. It should increase wealth. Everybody's, hopefully (cars don't buy cars).

The economy is not to blame if changes don't unfold in a smooth way. Businesses are to earn money, nothing else. Even if they know better, their lobbies are only there to multiply individualist interests which, if turned into politics, could as well turn out damaging or suicidal to their very originators. Preventing new technologies or seeking cheaper labor instead of investing may be short term strategies for one company, but turn out bad for all in the long run.

What is due in many more aspects as well, is to adjust the legal system to entirely new facts and circumstances.

The diligence to create appropriate structures is due to politicians, none else. It's quite obvious that they usually don't meet the challenge very well. In fact, not so many people at all seem to have a real clue about what's going on here. Politicians have the mandate, the power and (in most cases, I hope) also the will to act adequately, but are too often lost in the labyrinth of opinions and propositions flooding on them. Lobbying runs wild these days.

This is an utterly important matter. Information may become more valuable than hardware, the plains are open and the wagons run. Some are trying hard to get an unfair share, you'd guess.

Bottom line: it's every single citizen's own responsibility to take care about this. Not only deciders, anybody has to *really* learn about technology. Voters have to ask for whom 'their' candidates are really acting, or who pays their campaign for example (takes just a few clicks!).

Big Brother

With everybody wearing a vision simulator, having orientation cameras that are also able to record anything seen, being connected by wireless networks, it may be possible that evil governments want to have access to all that data.
The better part of the possibilities may be that pictures of terrorists could be distributed and an 'intelligent' software in the vision

simulators could ring alarm if it sees them. In the worst case however, everybody could be tracked by secret services just anytime anywhere.

With enough storage capacity, the vision simulator could also do less controversial things, like finding the lost keys, because it recorded where they have been seen last time.
It could amplify memory by photographing texts and numbers we see. With some intelligent algorithms, the device could also select those informations that are relevant and dismiss others.

Otherwise, if somebody succeeded to pirate a vision simulator, he could spy out the user's secret numbers and passwords. Just as with any computer. Yet with the new technology, even murder could be committed, by sending someone faked road scenes when driving, for example.
This however would require a severe hacking of the vision simulator's software, which is only possible if all safety measures are neglected.
It's not an impossible scenario, anyway, if operating systems are further delivered in a default state that is just utterly unsafe; and despite of all the warnings, this is still the rule.

'Secure computing', understood well, could inhibit unauthorized changes to operating software, if not greedy copyright advocates would steadily be pushing laws and technologies that deprive the user of his capabilities, force unwanted communications and information transfers, and inevitably open barn doors for possible attackers.

The really important copyright to address, is not the copyright to any pseudo cultural kindergarten stuff, but anybody's copyright to his own thoughts.

Privacy will really be a paramount issue. Computers, even today, have practically become extensions of our mind. Information stored must be protected from governmental and other intrusion, or we will more and more get an equivalent to thought control, not to mention the security threat to businesses, institutions, the entire society from criminal or terrorist attackers.

With a device that is used by anyone, anywhere and all the time, this applies even more. As vision simulators are just the last step before implanting chips, all the issues that would then finally arise have to be addressed in the course of this evolution. What's evident already: computer privacy is a human right [68], [69].

One step further

A propos implanting chips. This is something that would not end with just interfacing.
Just for fun, let's speculate a bit: It is easily foreseeable that information processing structures will one day surpass the density of comparable brain structures*. They will also be able to reproduce typical brain structures [12]. Logical elements like gates will almost get to the size of single atoms.

---

* In 1965, Gordon E.Moore stated that the number of transistors on a single chip doubled every 2 years. This has been proven true ever since (p.73). Today we have chips running at 3 GHz (billion cycles/second), having 100 million transistors with 60 nm structures, consuming up to 100 watts. It would take 30 years until the absolute limit of structure size (atoms are approx. 0.3 nm) is reached, or only 15 years for a structure that could still work like current chips (2nm). This would yet be chips 1000 times as complex. As smaller components also get faster, we could see a 30 fold increase in clock rate or alternatively a 30 fold decrease in power consumption after the same time. A chip of 1cm^2 could have more than 1000 billion transistors and either a clock speed of 100 GHz (even many times that, with some new semiconductor materials) or a power consumption of only 3W (at 3GHz or a lot more). In less than 20 years, a tiny very low power chip as capable as a today's PC could be an integral part of super light vision glasses or, if the connection problem is solved, even be implanted for permanent personal use. In 30 years or a little more, a single layer 'brain chip' of 100 cm^2 could have 1000 transistors for each of our 100 billion brain cells, take about 1 watt and *run a million times faster* than any biological structure.

Quantum error correction could first be used to make them reliable, yet full fledged quantum processing will also be possible, not only resembling many of the pattern processing capabilities of the brain but most probably surpassing them. The structures required may get orders of magnitude smaller than comparable neuron structures, even without the quantum.

Concurrently, processing and signal speed will be (already is) million times faster than with any biological structure.

A problem one could bring up is wiring. Yet a single glass fiber can theoretically carry about 10^{15} bits/second (read out an entire super brain chip in *less than one second*). A formidable tool for interconnection even in an ultra large and complex structure.

Mere density of course won't suffice at all. It could only work if we know how to model the brain's structure. The actual time scale for these developments may therefore depend on the progress in neuroscience rather than in chip technology, at least if Moore's law really persists for some more decades. Decoding the brain in 30-40 years is not really that likely.

Nevertheless, it will some day be possible not only to implant chips, but also to use them to think, provided that a suitable interface technology can be developed. It may be that our natural brain will then play a role just as it is currently taken by the cerebellum. Due to its higher computing power, practically all of the personality will then finally reside in the chip.

You might argue that intelligent machines could also emerge, that some day would compete with us for predominance. Well, I don't buy this. Computers are depressive. Making them independent beings (such with a real will and ego) would require to implement unrestricted self preservation (the instinct that makes us tick), and any sane human beings would violently prohibit any appearance of it. Terminator won't happen. What is realistically conceivable in my opinion, is a continuous mind migration from brain to 'chip', first expanding and then accommodating an existing human being's mind. There might come a time when bearers of such chips would refer to those still using vision simulators as the 'four-eyed' (now you know why the fiction part of this book is called "Adventures of a Four-Eyed").

As it is quite likely that a 'brain chip'* will - contrary to the biological original - have a backup feature, it may become possible to copy an entire personality. Then we'd get some copyright problems, you bet !

There would also be profound dangers of privacy breach and slavery, to say the least. The advantage of all the hassle could be some sort of immortality. Therefore if such a technology would work, it would certainly be developed. The intelligence achievable is overwhelming. Such a being could 'talk' to a billion people simultaneously, if sufficient communication channels were established. The world would become a village.

The time frame for all this could be less than a thousand years. In cosmological terms, the current stage of human development would then literally only last for a second. Hence, if we are looking for extraterrestrial intelligence, we should keep in mind that we may be looking for super intelligence in (quite exactly) 99.9999% of all cases. These would probably not be interested in talking to us at all. As a science fiction writer once put it, we might currently be just an ant hill beneath the highway.

As we won't need super intelligence to develop a perfect vision simulator, just for the time being let's stick to this far less exotic and far more acceptable technology we are dealing with now.

---

* You may object that there is no conceivable method of interfacing. Yet consider that our two brain halves are interconnected just by a relatively narrow trunk of 'only' about 200 million nerves, the corpus callosum. In the past, surgeons have experimented to cut this apart for the treatment of very severe epilepsy. From the results, one can conclude that people survive this but that many cognition processes must normally be communicated through this connection. If we imagine a chip reconnecting the brain halves and thereby tapping the wires, so to say, we could conceive that a 'third brain half' could be implanted between the two natural ones. Where it could get its energy from would of course be another problem. This is all extremely spun and quite irritating, but it shows that speculations about such technology are not entirely clueless. Simple neural adapters like cochlea (inner ear) implants have long been in use now. For some current research on neural interfaces, see [66] and [67]. *RESISTANCE IS FUTILE.*

Conclusion

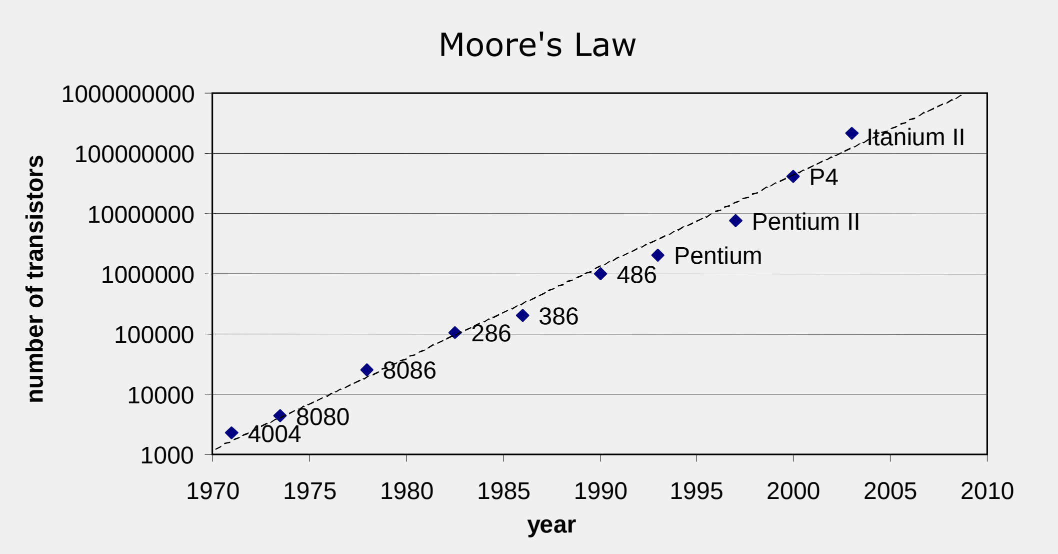
The evolution of information technology tends to minimize hardware and connect the user interface as closely to the human senses as possible. Replacing Hardware by virtual devices and reducing the remaining parts to the absolute minimum, would be a logical step in this development.

This Technology will save material and costs while offering an overwhelming area of new possibilities. Its openness to the environment versus the concealedness of classical virtual reality, along with its communicative capabilities, will lead to new applications not yet even thought of.

As many traditional products will become obsolete, it will be crucial for any information technology business to be present in the new markets evolving from the developments considered.

Virtual devices could definitely save a lot of natural resources and generate additional wealth. Dumping a 20g vision simulator is not as bad as dumping a 20 kg monitor, for just one example.

So this development should in the end be to everybody's advantage.



Evolution of processor complexity on the example of Intel processors
(dashed line: doubling every 2 years)

The materials presented here have originally been found in the ruins of one of these ancient office buildings that became obsolete during the 21st century. It may be quite amusing to see the difficulty of life before brain chips.

part 2: Fiction-Adventures of a Four-Eyed

Home office

This place here is not very inspiring. I guess I should change the wallpaper. No problem, although the edges of the furniture look a bit soft with all the pieces of virtual wallpaper inserted between them.

Maybe I should just tare down the walls. Replace them with some panoramic view. There are these super webcams, 360 by 360 degrees high definition. These guys are making a living by selling the ultimate view. Considered that people are paying millions extra for an apartment, just for the view, it's no wonder how good these webcams are selling now.

Swiftly I find one on a Caribbean island, on a hill, overlooking a gorgeous bay. It's night, like here. Now the walls and ceiling are gone, it feels as if I was really there. Well, let's dim the light, it creeps in at the edges of my glasses. Spot on the desk, it's the only real thing remaining, and the working windows are of course still there.

Hey, a little too real. No, not simulated mosquitoes. But that disco over the bay is really too loud, and then there are those little frogs chirping in the tree beneath me, those super loud little ear killers. This really hurts.

That the neon light of the disco are displaying some ads is OK (this is a free webcam), but what have these polar lights to do in the Caribbean, also forming ad texts all the time?

This is annoying. Fortunately, here's a better webcam. The only thing still missing is some breeze. I tell my computer to start the ceiling fan, blowing upwards, so it makes some whirl. Quite OK now.

With daytime scenes, it's a bit harder. I can simulate sunlight by adding some light on the real objects simply with the goggle display. Sun heats. I guess I'll install an infrared beamer at the ceiling. This may complete the illusion (and make me sweat).

The room itself won't get any better this way. It already looks like a bunker with careless installations all over. Only the virtual room dressings, ambience lights and so on make it livable. No good. I should do something about it.

The link

Then that wall breakthrough. We have a branch office in Springfield, and sometimes my wife or I are looking for our business there. So we had the idea to connect the houses.

First we thought about display wallpaper. It comes in nice rolls of 20" by 80" that are glued on like ordinary wallpaper, edge to edge, and then connected by some clamps screwed to the wall.

We would still have to set some cameras right into the wall, and even though these tiny devices would come with optics only 1mm wide, any of them would require a hole in the wall as well as in the display, and they would have to be wired.

One company offers an integrated system with this, but then these very tiny cams have low resolution or suffer from snow. The perspective also isn't always right with such a real screen. They can calculate it for one person, but if there are two, each has to wear glasses so 2 different pictures can be multiplexed. The glasses are also necessary if just one person would like to have stereo view.

The next choice was a 12-projector multi perspective cell system with a holographic mirror wall. It concentrates light from the different projectors to different areas of the room, dynamically generates perspectives if someone is there, does not stray ambient light and is transparent in directions not programmed to reflect. So the cameras can also sit behind it and need no holes. A very nice system but most varieties do not allow for different pictures for each eye (no real stereo) and even then they're too expensive.

So as we wear our little vision simulator glasses almost all the time, why the hassle ?

We decided just to install a camera array in each house and leave the display to our toys. I found one vendor who proposed a grid like assembly, vertical tubes with several cams and mikes each, that are connected by horizontal tubes at the floor and at the ceiling. Looks like behind bars, I told him, and he wasn't so amused.

Another system was more convincing: Tiles of about 50x50cm. including a camera, a microphone, and also a flat speaker box.

Together they could provide an almost holographic sound field, as the salesman in the Floppy Hints shop told me. More spatial fidelity than just the simulation in the headphones (which works quite well though). This was easily installed, only plugging the tiles together and a nail here and there. They even come in 10 different wallpaper coatings already. It works amazingly well. I often think she's really there. Three times I ran right into that wall. Now I know why these tiles have that soft surface, as in a padded cell. Good job, guys.

Excuse me, I think my doorbell is ringing. A tunnel to the front door opens. It's Rudi, our dog. I point to the control field of the virtual tunnel view and the direct camera picture appears in front of me, together with the door controls. It also establishes the speech connection. I say "come in" and open the door. OK, let's continue with...

just a minute, there's a virtual owl bringing an e-mail. It's pecking at the window, wanting me to really open it, instead of flying right through!! At least this obstinate bird can't bite. The envelope opens and releases a giant dragon's head, giving me a fireblasting "you forgot to charge the car!". OK, I plead guilty. I guess she had to hook up to one of these 'in-drive loading' services that load your fuel cells while driving. They're also called 'highloaders' (guess why). At least the extra expense saves you the wait at the gas station.

Now really back to that wall: you may wonder if the broadband connection (it's high definition of course) wouldn't suck up my wallet. It turned out quite cheap though. The networking people told me that one single optical fiber, with multi spectral modulation, could carry 10 millions of those connections at a time. Amazing. I guess they're even still too expensive. After this installation, my wife noticed that the real look of our rooms also needed a facelift. No doubt. Especially in the morning, before taking the goggles, it's really enough to look into the mirror, the house around it should at least look a little better.

Some work

After all the 'home improvement', I really have to get back to some work.

Finishing this book was almost out of control. I had so many documents and notes, and in order to forget nothing of it and to check the entire material for consistency, I had to keep many files open at the same time. If I had nothing but one of these medieval display screens, it would have been a nuisance, cluttered windows all stacked upon each other, not really good to use.

Maybe electronic folders are way easier to use than real ones, but if you work with many files at once, there is no match to the method of pinning everything to the wall. In a software project we once had all the walls of a big room covered with listings. It really helped. Except that we had to jump around all the time, just as the CPU would jump from subroutine to subroutine.

This here is even better. All the docs are arranged around me in all directions, like controls in a spaceship. I guess these are a hundred or so and still all visible and in reach.

My wife just comes in to ask something. Could have used the link. Anyway, the windows in her direction are automatically dimmed as the position cameras register relevant motion.

She needs a document I have. It's just one of the windows in front of me. I grab it and throw it to her. A paper pigeon flies off. She sees a display appearing, asking if she is willing to accept. She looks at 'yes' and the sheet appears in front of her.

This feature is called SEEHIS ("see it how I see it") in contrary to SEEWIS ("see it where I see it"). With the first method, an object is placed to someone else in the same relative perspective. This is good for copying documents. The other method is for cooperative work. It does not copy but shows all participants the same object in the same place and allows any to manipulate it. Public objects also use this method but cannot always be manipulated by others.

Ok that's managed, let's go on with the work. To activate a window I point at its controls with my eyes. I even move the cursor this way. For writing and other inputs when working with documents, I like to use a mouse pen.

I could write directly into the air with the pen, but that's far too jittery and tiring. I remember when those parcel guys always presented me with an electronic notepad to sign as a recipient. I wonder if anyone has ever tried to verify who did those chicken scratch figures the device had stored. Writing without a support is something for a professional painter. I prefer to write on a desk. This pen that I always carry with me works like a PC mouse, on any surface. I do not need a keyboard because all applications now are able to recognize handwriting.

Now the work is getting schematic. Proofreading, and I'm sure I won't get all errors even after a dozen passes. Could also use some stimulation while doing the boring stuff.

As these windows clutter up the virtual wall TV that I normally use, I open another little TV window that I can sort into the crowd. Looks pretty conservative compared to some fun skins they once made for such applications, but this is not unnormal, designs always revert to the simple if that proves to be the best.

The satellite viewer application however sticks out, it comes as a virtual 3D globe that can be flipped around by just touching and turning. It only transforms into an ever flatter shape when closing in on some place.

TVs and icons

I really guess I need a break. The TV application follows me step by step as I walk around in the house. Any of those 20-something virtual TVs that I've set up in the home just show me the program I was tuned to. Normally this is what's intended, but now it sucks. I touch the switch on one of those, and all turn off at once. Actually they are reduced to little icons that I can point at to reactivate. I could have used the option to revert them into virtual paintings, but the little artifacts of the mask displays stop me from defining these virtual room decorations all too often. It only works well if the TV was defined onto an already black part of wall. The icons use no mask, it is not bad if they are transparent and the advantage is that there can't be any artifacts.

Reactivating them if they are not in reach, is no problem. I have configured most of them to 'staring' mode, i.e. just looking at them for more than a second will do it. If I rise my eyebrows it even works instantly. My daughter just twitches her nose for the same purpose. Her sim has an ingenious electrical field sensor to record physiognomy. Mine still relies on a 'flies eye' stripe of optical sensors along the lower rims of the glasses.

I'm just figuring out what it would have cost to buy this many TVs, not to mention the space and the energy they would take and the wiring as well. Meanwhile these TV apps for the vision simulator cost nothing, they are always included or can be loaded as freeware. Given the fact hat I could even set up a giant movie theatre wall anywhere I want, or even a full fledged surround theatre, that it can even display fully holographic images, there is just no comparison to ancient screen technology.

Only in the sleeping room do I still have a big display (a laser projector pointed to the ceiling) and speakers, as it is not always convenient to carry even these very light glasses that I have now.

The downside of the cheap supply is that eventually all rooms get cluttered up with virtual items like chess boards, pets, stickers, controls and icons of all kinds, maybe even a leftover monster from a video game. Fortunately I can hide or dump them without needing a real trash container.

Outdoors

A bicycle ride through the woods will be perfect right now. Let's do that. The dictating function of my sim is really useful here. Being able to activate it just by eye pointing and just to talk into the machine what comes to me, is way better than with any ancient recording device. Works hands free, allows to assign files to where they belong from the very beginning, stores everything with timestamp and dynamically backs up anything over the radio network. I like this especially as many of my best inspirational ideas come to me outdoors. This way I don't forget anything. It's also pretty useful in a car or any other busy situa-

tion of course. Orientation is another advantage. Not the usual stuff. Bicycling in the wild, I really appreciate the satellite pictures that show me what the environment actually looks like.
A very versatile way of seeing them is the 'transparent landscape' mode, that lets me see behind way bends, buildings, hills and woods. It lets me see certain things even miles away. Simply merging the satellite picture into my field of view would obscure it and be dangerous. The subtle methods of blending in the see-through information you can get now are much better.

A friend of mine even takes this for car rallies and racing. Seeing bends ahead with this virtual tripmaster allows him to speed like mad, and seeing obstacles like stranded cars improves safety. This also applies to normal street traffic, of course.
I remember when in the old days I once traveled on a highway, the traffic radio warned of a horse running around on one highway nearby, an armchair sitting on the left lane of another, and a car going the wrong way on a third one. Fortunately not on my route. Stranded cars are easily announced by their emergency transmitter, but for things as those mentioned above they've fortunately invented the voluntary 'anticipation network', that either delivers pictures directly from sim to sim, or over the street safety network. So you can literally see through the eyes of the guy ahead, intelligently filtered for importance and timely and spatial fidelity, of course.

Bicycling makes me hungry. There's a recipe service that promises to guide you while cooking. I once logged in and it worked quite well, picking up my kitchen inventory through my sim's cameras, telling me what I could make from it, then guiding me through any steps necessary, timing the cooking, and so on.
The downside was that afterwards my cupboard and kitchen walls carried some lists of things missing, but also several ads with 'one-look order' functionality, and my stove received a sticker "look here to buy a new one" and another for a service contract. It took me an hour to get rid of all this. So much for free recipes.

Concerto

Doing sports myself is much better than watching, but nevertheless I appreciate nature films or live events that use the full surround technology.

The triathlon last week was one of those in several ways. Most fun was the helicopter ride above the swimmers, bringing this immersive experience to the max.

The athletes were also using vision simulators, not all in the water (although the cameras can provide crisp underwater view with laser based units), but certainly afterwards when bicycling, because it allows for total orientation and control and, of course, they are perfect sunglasses.

Some people also like the virtual sitting in the football stadium, but that's not for me.

I better appreciate nature films, that really get exciting this way, as many visitors of surround movies had experienced even in the olden days. Back then, one had to go to a special theater. Now I can have this at home all day.

This evening they transmitted one of these holodeck like events: A violin quartet with artists, recorded by a complete surround camera array, that could be displayed right in the middle of my living room.

Not that I always prefer this kind of music. Yet being able to go around the orchestra and hear any instrument clearly separated, with this perfect spatial simulation, is something really attractive.

Having the news speaker sitting almost bodily in front of you is just nothing compared with this one.

I could even choose to have the walls and ceilings of the real concert room displayed, but that's dangerous if you walk around, as you will definitely run into your own furniture, or the safety function will rip the image. It's just an option for sitting in one place.

Needless to say that the full spatial sound requires to use the headphones, because speakers would just involve the acoustics of your own room. I only use them to enhance the bass, as those headphones could never make your stomach feel it.

Fairy tales

Recently the fairy tale channel featured the 'history of media', a 3D movie with total surround, virtual characters running through the room, and so on. Yet do they really think even a child would now buy it that 'once upon a time', they tried to prohibit recordings of TV transmissions? In this story it must even have been illegal to play old vinyl disks with a digital amplifier (as this could mean digitizing) or in presence of friends. As this was no interactive game movie, I could not test these suspicions myself.

It went on with storage crystals getting so powerful that the entire 'cultural' heritage of mankind could easily fit onto something the size of a sugar cube*.

Inevitably, just this kind of gadgets emerged. Those 'Final Chips' were heavily prohibited, but it couldn't be stopped because nothing that people want can be stopped and they had been so totally deprived of all their consumer rights that almost nobody had any sense for legality with these things any more.

It also proved what experienced technologists had known all the time, that it's an illusion to stop such things by technical means, or the damage gets way bigger than the benefits. During this time - they called it the 'media prohibition' - crime was booming again, as for outlaws it became another license for money printing.

Some guys with trick or treat hats then asked to install surveillance cameras in any home to see if anyone played illegal stuff. As this wouldn't work with vision simulators, they finally asked that any private computer should be entirely accessible to them at any time. It became finally evident that ill designed laws had led to a point where there could either be no copyright or no privacy any more. The backdoors to computers and 'magic mirrors', opened by some copy protection schemes, had then already contributed to billions in damages by attacks of villains and evildoers that eventually even paralyzed the entire kingdom and destroyed colossal amounts of data.

---

*Even a simple, flat ½"plastic chip with engravings several atoms wide, can secretly store way over 10000 movies or 100mio. books or pictures. Replicating it by injection molding (like DVDs) is dirt cheap and done in no time

A more harmless occurrence was a TV virus that caused certain politicians always to appear with puppet strings attached and their suits stuffed with dollar bills all over. Well deserved.

Scientists presented studies that the costs for implementing and maintaining this 'technology' and for the collateral damages it caused amounted far higher than its benefits. As if it hadn't been enough already, huge problems that cybersquatting[1], patent sniping[2] and patent landmining[3], accompanied by clueless law-making apparently had accumulated over a long time, added to all the disaster. Rightout bizarre, bits and bytes had been assigned more rights than humans, and digital slavery[4] was spreading.

The 'intellectual' property mania turned against the entire economy. Any engagement with hardware or programming became as heavily restricted as money printing, small IT businesses couldn't emerge nor grow any more because they could neither afford enough patents nor meet the security requirements for dealing with classified chips and algorithms. Education in IT and electronics became confined to a small caste of shamans, and progress had been frozen forever.

Everywhere, signs like this were harassing harmless pedestrians:

> This text is copyplighted. Reading it and
> storing it in your brain is a violation of the
> devious malicious copyplight act of 3001 and
> constitutes a filial crime.
> Please move to the next frolice station
> immediately and report your crime. Do not
> surf over GO and do not look away $200.

---

[1] Hijacking rights on data paths, information resources, web addresses, trade marks on common words, names, colors, genes, natural substances etc

[2] Hoarding rights on trivial ideas, common knowledge etc.

[3] Landmining by issuing numerous trivial patents.

[4] Depriving persons of their informational rights and possessions.

At this point, just before the guys in the white jackets came, everybody realized that this couldn't lead anywhere, and politicians - who had taken two generations of evolution to acquire some knowledge about information technology - established new rules, a legislation granting both fairness and freedom (hey, don't they always claim doing just that ? and what's the result ?).
The wide acceptance of the new regulations made copy protection almost unnecessary after all.
They also declared it unlawful to track anyone's physical movements, personal business, private life or communications without a court decision. So the intrusion by tireless data collectors also finally came to an end.
Most important, a law was passed, that granted the privacy of personal computers ('mind extensions'). This was even turned into an amendment to the Constitution.
Nevertheless, could anybody even in a fairy tale[1] be naive enough not to recognize that no property is safe without freedom, as only freedom can grant justice and only justice can protect property? Privacy in turn is essential to protect freedom, isn't it. So protecting property by taking away privacy is a contradiction in itself and anybody trying this will shoot himself in the foot.
In this story, Orwell's prophecies [11] had almost all come true. The reversal came shortly after some lobbying groups like 'Take Them to the Cleaners' or 'Privacy is Crime' had in a last effort, tried a little too hard to oppress their opinions onto the world's legal structure in a fictitious '2nd Cyberwar'[2].

Sometimes, things are turning for the better and all live happily ever after.

---

[1] Probably it hasn't escaped your attention that in reality, even more bizarre and sinister developments had been pushed ahead once upon a time.

[2] It had long been recognized that all knowledge and legal structures also belonged to what had initially been called 'cyberspace'. So this 'war' didn't take place in the web, and no viruses were used to fight it. The main weapon in this 'war' instead were law proposals tirelessly written by hosts of lobbyists.

This was made for kids, of course. Issues like 'private' banking and the virtual goods black market weren't mentioned, these in reality induced even more tendencies to install Big Brother in any computer, because it impaired tax revenues. Unsurprisingly, in these cases no business lobbies were ever asking for more control.

Some official's greed for information got so bizarre, that they entirely forgot that privacy is a human right [68], [69] (yet did any of these law poets ever care about constitutions and such?).

I'd really like to know what your Senator would say if just after the postman, a snoop would regularly appear at his mailbox, open all his letters and read them thoroughly. Of course the guy would put everything back in afterwards, nice, ain't it.

As secret communications generally can't be stopped, not even proven (if data are hidden in sound or picture files by steganography), it finally turned out that total control will ever remain an illusion and one had to turn back to more conventional methods of investigation.

Since the laptop days it had also dawned to just anybody, that the possibility of the perfect digital copy never made any difference at all. Most people aren't even able to tell the quality difference and if so, they usually don't care anyway.

Inhibiting all varieties of copying requires to place a policeman in every living room. 1:1 hardware copies cannot be inhibited unless the technology is entirely sealed and locked, and playing them cannot be blocked unless there are only playback devices that only work when online connected to a dedicated 'rights server'.

It turned out that such stupid makeshift stuff would never work in any acceptable manner at all, unless an entire host of also makeshift and likely unconstitutional laws would be pushed forward, trying to enforce the functionality of an inherently flawed technology. As digital goods are totally different from hardware in that they can be reproduced and distributed at very little cost, implementing the equivalent of a daily house search for everybody just for their control is meanwhile considered brainless anyway, and just a fundamental failure to adapt to the market. Nowadays, the fight of the plastic carvers against the web is often mentioned alongside with the fight of the gas companies against electricity. So it finally became common consent, that the primitive crowbar technology once named 'Digital rights management' (DRM) was clueless and even counterproductive, as nobody was actually willing to buy anything without getting to own it, even less so from the very same people who would just use the money to promote even more 'legal' atrocities against them [75].

The 'six postulates' became standard in any engineering textbook :

I. It's impossible to prevent any physically identical copying unless the entire information technology gets classified and restricted.
II. It's impossible to prevent any content copying unless the entire information technology as well as analog technology is classified and restricted.
III. It's impossible to manage any detailed digital rights without an entirely classified and restricted digital technology and enforced communications to and surveillance by centralized servers.
IV. It's absolutely dangerous to bind any information content to any specific hardware, as such content would be destroyed if this hardware becomes unavailable.
V. It's absolutely dangerous to bind the usability of any information content to any remote servers as such content would be destroyed if these servers become unavailable.
VI. 'Digital rights management' can only work in a perfect police state. The only thing that digital technology contributes here, is that the perfect police state can be implemented much easier.

Fortunately, the archaic habits of distributing media on round pieces of plastic are now long forgotten. So an entirely different system came into being. If I download something now, I might get it customized with my name, the originator's signature[1] and maybe an asymmetric watermark[2], but then I can do with it what

---

[1] A hash (sort of checksum) calculated for a data set, using a secret key. Anyone can check for integrity of the entire set using a related public key.

[2] Making at a large number of subtle (invisible/inaudible) pseudo random changes in a file, determined by a secret key. A related public key allows to derive enough of the changes for a very safe validation of the watermark (even from only a part of the file). Yet trying to *remove* all of them - especially if we know only the public key - would not work, or add so much noise as to render the file useless.

I like, as long as I don't redistribute it illegally. I can sell items second hand by just adding the buyer's ID and my signature. It's for digital media the equivalent of a receipt, and the watermark equals a serial number. Perfectly simple and all that it needs.

Frankly, there has never been anything wrong about pursuing people causing damages to or making money from the work of others, as long as the methods are complying with human and constitutional rights.
Nonetheless it's rightout perverted to ever even think of making '1984' come true just for the sake of greed.

Flight scenes

Yesterday I went to Tampa, via New York JFK airport. Needless to say that I had booked my ticket by the web and the certificate was loaded into my pocket computer. So I had nothing to do at check in but to take the glasses off shortly for a face control (their system wasn't really up to date). Booking was all done wireless. I only had to OK that transaction once for the sake of safety. Again I was happy that I didn't have to use a mobile phone or other ancient stuff that would occupy my hands and make me run against lanterns while trying to read the display.
As usual, the onboard movie was as bad as the music from the armrest. I also wonder why they still have these screens built into the back of the seats. They already have transmitters as well, sending a vision simulator the movies, the music, and having it display and operate the virtual controls for it.
The only time I enjoyed an onboard movie was long ago, when they played Apollo13, when it was all new. Good that I have my own harddisk in that pocket thingy, that can simulate a giant movie theater with the high resolution glasses and carries my entire movie collection. Yet once you have it, it starts to bore.
At least they don't tie these nickelodeons to the hardware any more, as some did once upon a time. I've never had a computer that actually *itself* liked watching movies or listening to music.

Just being able to download personalized files without insane restrictions is much less inconvenient. Some retail stores may even still burn disks for you, if you live in the digital outback. Good that my shelves are no longer filled with plastic garbage that's only been invented to make the stuff look more 'valuable'.

I decided to ignore that inevitable psychotic B-movie. Tuned to the outside cams of the plane instead. Hey, that's the best thing ever invented. Especially as I didn't manage to get a window seat. I gave it a try and let the entire plane interior vanish in favor of the outside picture. These cams merge together quite seamlessly, so the impression is absolutely perfect. It felt like flying without any plane at all, just alone in mid air. Quite disturbing the first time.

Then I tried to have the seat blended in, so that it wasn't looking as if I was sitting on plain air (as a default, these goggles never hide your own body, so you won't get disoriented), but then it looked even weirder. I wondered if they shouldn't replace those seats with broomsticks, that would be looking quite cool. The first class could offer dragon's necks instead.

The flight attendant suddenly brought me back to the inside. The motion detector had decided that she needed attention and blended her into the picture, the same time blending in the plane interior by 50%. This surely destroyed the illusion.

After she had handed me the drink, I stayed with the outside view most of the time, except for some phone calls and a little web search. The clouds were spectacular this day.

I have been told that in the past, some airlines prohibited the use of camcorders and even cassette players in their aircraft, because they feared interference with onboard electronics or radio communications. Really mad. I understand that they didn't like the use of mobile phones, but didn't they have any engineers to tell them that all these other devices emit a thousand times less power?

Actually I think that just some yellow press campaigns had impressed people more than facts. Maybe I should wrap this book in yellow ?

General aviation

I remember renting a general aviation aircraft last year, doing one of those useless dinner trips just for the fun of it.

You guess that even the remote rustic restaurant we headed for had these virtual menus that let all dishes appear right before you on the table as soon as you're seated.

This airplane had more gadgets. I could see all the other aircraft around, virtually, right through the cockpit walls.

It also had a really good set of external cameras, so I could see anything out there as well as in the big airliner. Quite nice, really.

I never did such a soft landing before in a rented plane: I could actually see the landing gear and the runway right through the bottom. There were even some thermal cameras that would go through fog and would make night flying a cinch.

Fighter planes have been operating these for a long time of course, and the ability to aim and shoot those disk rockets to all directions at any time made dog fights a little obsolete.

The real-world mapped navigation system is really so much better than just a conventional GPS. You'll know when you have learned how difficult it sometimes is to find small airports from the air, in all that haze, even if you know that they must be right ahead of you. I have seen many experienced pilots having the same difficulties.

The reason is simple: you only see things if you know what to look for. From the air, landscape changes a lot with weather, time and season. I remember once canoeing with a fiend and just for fun I didn't change directions when we were right approaching an alligator. He didn't see it before it just ditched into the water only 10 ft. away. I only saw it because I knew what to look for.

So in unknown environments, a direct insertion of directions into the real scene is a huge advantage.

These virtual sight applications are quite common now. With my sailboat, the sonar provides 3D pictures of the ground, that can be seen through the bottom of the boat, and outside cameras let me look right through the sail. Also no toys, these ones, could really avoid collisions or falling dry.

Games and sims

I'm not a game fanatic. I just run some of them from time to time, to check out what's going on with technology, but generally the flight simulator is the only one I really keep (that's not actually a game though).

Escapism has been a problem with computer games from the beginning. People fled into virtual worlds, and some even got insane with it. High end vision simulators also offer much less dangerous and more useful things, like dressing the rooms for free or having that million dollar view just out of the backyard window. The borderline between the sane and the insane will always be hard to draw.

With these shoot-and-run games it has become really dangerous. People have been reported running into cars or falling down stairs, just acting on virtual scenes. Many games were completely transported to the real world, and sometimes you see someone running down the street shooting at imaginary opponents with a real looking gun. Of course, those guys are usually caught quickly, if they are lucky enough not to be fired at by a nervous policeman. It went even further, actually: some gamers really thought they were immortal and jumped from buildings, some started to fight with real swords. Now you need a sanity check to join certain online games.

My Flight simulator is much safer. I still use a yoke and some rudders (how would you simulate these), but the cockpit is a 100% reality like 3D reconstruction that can actually be operated by any switch and control that it shows, though without tactile feedback. The good thing with this application, that it allows for a fully realistic VFR (visual) flight. The only downside, that flight figures are almost as likely to cause vertigo as in a real plane. Only gravity effects are missing.

They built simulators like this decades ago. Back then I could try out one of these ultra expensive installations and wondered when it would be possible to get something this good at home. Took some time.

What I also can't do with this simulation, open the doors and trample at the left one or the right one to steer, as in that real Cessna 182.

From the past

Back to the JFK/Tampa journey. After landing, I really enjoyed those green arrows on the floor, directing me right to the connecting plane. Even that virtual officer approaching me "Sir, you may miss your connection if you take this way". These guys always know your schedule. I think you can also ask them for the next restroom or so.

I remember a flight along the same route long ago, when right before landing they began to announce all the connections possible and their gate numbers. I don't know why, but somehow I mixed up gate numbers. After running through the Airport half way towards what I thought was the right gate, I looked at some screen wall and saw that we were wrong.

I was with my wife and a friend of ours, but they also hadn't noticed the error. Having made it back to the right gate (that was just next to where we had arrived), the airline guys had already checked in some people from the waiting list (our flight had been late anyway), and only one seat was left.

Would a mobile phone or notepad with a guiding software have helped here? Certainly not: no hands free, no time, no way. Only a vision simulator could have done that.

They could offer us two more seats in a plane from La Guardia, going one hour later. We decided to give our friend that last seat and to head for La Guardia ourselves. Well, that was like an excursion into the NY underground: First a drafty roadside under a crossroads where we waited for the bus a long time, then a long and sometimes slow night ride through the wet and rainy industrial outskirts of town.

Finally arriving at La Guardia, we were so late that the front desk didn't even issue us new boarding passes but told us to hurry right through. Then a clerk at the baggage check didn't want to let us pass because our tickets didn't fit a flight from this airport. Remembering that our company would be waiting in Tampa probably with no idea where to go, I gathered the rest of my politeness remaining and shouted something like "Sir, I will definitely not miss this flight", together with a few other words unmistakably explaining what we had just been through. He must have felt that I was about to run right over him, so he gave way.

We got more to do that day. Not that I even remember meeting our companion and getting the car, but the hotel we headed for was a bit out of town and we had to find a tiny road going off the highway and then some more corners to take until we arrived about 2:30 AM.

I'm just telling you all this because I wish I had my vision simulator back then. There have been some approaches towards using mobile phones for similar services, but it's simply so that I need my hands when travelling, and the displays were tiny. A display that would insert a data screen at a fixed place in my field of view would also not be the real thing, feeling just like dirt on my glasses, and using a notebook even more so. There is no other good interface but the vision simulator.

A rental car

Entering that rental car right now, I remember how it was in the old days. They had stuffed more and more electronics into these cars, making them more expensive and less reliable all the time.

A 'technology' magazine even once featured an article saying that soon some car company would probably have a CEO who was an IT specialist, because information technology would make up 90% of the value of a car. Rightout bizarre.

When I was a student, I turned my car into sort of a gadget carrier, and I learned for sure what does *not* make sense. All my friends back then made the same experiences.

Fortunately, that's all over now. My car is a car, has 4 wheels and is made for driving. All those electronic gadgets, those 26 stereo speakers, TVs in any seat, fax machine, even the radio, have been replaced with my pocket computer and vision glasses. The car provides the motor, some outside infrared and radar cameras, anti lock brakes, and so on, but most of the superfluous IT and media stuff went where it belongs: right to the user. I only added a subwoofer that gets its signal from my sim.

I really appreciate that they made a law last year, that any car must have direct mechanical or hydraulic brakes and steering, with only a certain amount of simple *servo* support allowed.

None allowed that would depend on electrical or other power. It will save lives, I guess. That 'X-by-wire' stuff was a really bad idea, at least for the safety relevant features.

At the rental desk, the clerk had shown me several models by transmitting full fledged 3D pictures. I could go around each car, see it as real and in full size, open the doors and look inside, but I couldn't take seat of course (a guy at the next desk tried, you guess what happened).

Good that I had insisted to get a signed confirmation for my reservation price. With this particular rental company, I knew that they would always have 'forgotten' the reservation price and come up with a much higher one at the desk.

This time, before I showed my reservation ticket (a digitally signed virtual one of course), the clerk offered me an upgrade to a luxury car for $5 a day. After considering shortly I said yes, and then he came up with a rental contract almost twice as expensive as my reservation. I handed my confirmation to his sim, and now he surely was in a fix, because he had to give me that car for my confirmed price plus $5, which he had never intended at all.

Walking to the parking lots, I was wondering why nearly all staff at the airport looked completely identical. I had never heard about any cloning program going on.

We saw our car from 300 ft. away, through walls and floors, by a flashing light on top (virtual, of course), as if it hadn't been enough to see those guiding lights on the floor and walls all the time. The car unlocked when I looked at its door (the clerk had handed me the virtual keys just like real ones, only that it was a vision sim object), and when I got in, a short introduction to its controls, also the virtual ones, started to play, with illuminations and little hands pointing on anything being explained, switches virtually moving, and so on.

I stopped it because I already knew this model. It had hardly any controls on its instrument board, that, except for the styling, was almost as Spartan as that of a 1957 Beetle. At least if looked at with the naked eye. The car could be conducted this way, but for convenience and safety, a vision simulator was recommended.

Well, a flashy firework of controls, goodies and gadgets appeared as soon as I started, and first I swapped the dashboard to show the radio controls (an unusual feature for a car today), so I didn't have to tune when driving. I could of course have used the radio of my vision simulator, but it couldn't really provide the bass this car could let loose (button headphones only provide good bass when they fit tight, and I hate that).

It was not so unusual for this rental car to have a stereo. Rental companies usually choose cars that can also be operated by technophobes, so this model did not only have the usual backup touch screen to access some functions in case one's vision simulator went defective, but it had some other 'real' gadgets as well.

I myself feel quite comfortable with my vision sim, it's only 20 grams, and this new model now has those 10nm chips that reduce the total power consumption so far as to integrate batteries for an entire week or even to run it from the solar cells. With no cable to my pocket unit anymore (which is also not heavier than a bar of chocolate), this doesn't bug at all.

I can hardly remember how it was to drive without active sunglasses, the mask display of my sim* I mean, that takes out blinding lights selectively. Be it headlamps of cars approaching, or the sun reflected in the mirror or rear window of a car ahead, those smart glasses will provide a tiny speck of dark before it and let you see at ease.

My glasses had automatically gone into 'car safety' mode. It's a largely hardware based (and therefore reliable) mechanism that prevents displaying any objects that are too large, too bright or entirely opaque, ensures that even dimmed down objects can still be seen, and forces any of my own software control windows to arrange with the control panels of the car.

My own radio controls, for example, would appear in the car's steering wheel, the secondary (virtual) switchboard over the front window, or the middle console, never on the road.

---

* Slang for 'vision simulator'

The middle console is entirely virtual of course, a big progress from these old fashioned cars with their overloaded consoles and tunnels that served nothing but to wedge the driver, as in a racing car. They should have delivered those ridiculous design monsters with sort of a shoe horn, to help people squeeze in.

As soon as I told the car where to go, those glittering guiding snakes readily appeared and lined out any single turn, even the right way out of the parking area. The headup display that the exterior infrared cameras sent to my sim, together with the anti blinding function, almost turned night into day.

I could see for miles now. The road ahead continued even beyond the visible horizon, and there, small and distant was my destination, although 35 miles away, together with the estimated arrival time. Even further at the horizon, some tinted clouds with little symbols on them predicted rain approaching from the west, but that didn't affect us.

What I really liked with this car were its rear cameras. Not only for parking. I had forgotten to bring with me one of these convex rear mirrors that I'm used to, and the side mirrors also weren't my taste. Fortunately, I had a little software that could simulate me some rear mirrors from the outside pictures.

Most traffic signs were hybrid already, showing symbols as well as transmitting virtual appearances. These virtual enhancements can easily be changed and also be confined to be seen by drivers concerned only. Their use is limited however, for safety reasons.

Suddenly one of the conventional traffic signs flashed. The sim camera had seen it and thought I might be too fast, so it made a little overlay in the display. That sign however was meant for another road that ran beneath ours. Good that I steer myself.

There are still too many irritating traffic situations to leave it all to computers. After many years of trying, engineers had finally been able to sort out what works and what doesn't. Their machines, at least until now, never managed to match a driver who had 30 years without accidents, and there are many.

Maybe the 'superbrain' chips they make now could learn to act like a human? Rudi already has one and I fear he will be smarter than me some day. He is already good at rapping.

Yet these brain chips have to be trained. The problem is that too much general knowledge plays into some situations in real road traffic, and this can only be acquired by a complete human being. Hence, these chips indeed are just good as brain amplifiers, connected to real brains, rather than as autonomous units that would try to learn such a hard thing as safe driving, without a clue what's really going on at all.

Hardly on the highway, I soon got hungry. I increased the alien display allowance so I could see all these commercial road signs telling me where to find the next KrustyBurger's or such. That's a real improvement: They could tare down many of those huge, ugly advertisement towers and replace them by transmitters. Only if I want to see what's offered I turn this function on and see all the ad posts littering the landscape. Inevitably some people hacked the permits, just to show people signs they didn't want. Some also abused the capabilities of the feature: A business once displayed a very real looking, perfectly animated burning vehicle on the parking lane, with a flashing sign "we also roast beef". Bad taste, indeed (the sign, as to the beef I don't know).

Thinking of this, it suddenly came to me why all those guys at the airport had been looking alike. They must have been using avatar transmitters, and I hadn't set my sim to reject these things.

Merry Christmas

Some people have now started to use that road sign feature to dress up their home at Christmas. It's obviously cheaper and easier just to program virtual lights rather than climbing roofs and trees to install everything. The city supports it, because it saves lots of energy as well.

You have to get a permit for your setup, only then everybody can see it, but apart from this it's really more easy than the traditional way. The permit avoids traffic jams and accidents caused by setups that are too distracting. The only ones suffering from it are the city officials, they have a hard time every year, drawing the border line between allowed and disallowed.

In the mall

This evening I had to go to the supermarket, buying some necessities. A new gadget drew my attention when strolling along the shelves. They offered a virtual sewing machine. Quite peculiar. I peered at the price tag, and a short ad movie showed up.

Of course, should have known it, this was a toy, a simulation for kids. Guess I was a little flooded because I had programmed my sim to automatically extract and display the expiration date of any food products. It's a little freeware gadget from the consumer association, not really liked by the stores.

What they even like less is the second gadget I use, that looks at tags or reads out the info from these solar powered radio tags and automatically not only compares prices, but tells me if there are specials at another place in the store, also compares to other stores just by surfing the web in the background while I shop. Everything is of course displayed at the item addressed, never just stuck before my eyes.

Ten years ago, some guys even tried to lobby a law that would forbid using their active price tags as basis for a web search. Most peculiar for those times, they didn't succeed with it.

Some had even ceased to use radio tags because these allowed vision simulators or other special computers to sort out prices even more easily, for example comparing anything there is on that shelf within a second. Some encrypted RFID and barcodes, but that put off too many customers.

So now they try the old fashioned way, to name some products in a way just to aggravate comparison. As the only thing these price tags have is the printing, especially the bar code, that is read by the vision sims cameras, one anyway has to get the rest of the data from the store's computer.

I went right out with the shopping cart, and when passing the gate anything in it was accounted for, because it had already been registered by the cart through its RFID tags and also been assigned to me trough automatic interaction with my sim. Some items with classical barcodes had been registered through my sim's cameras and a certified piece of software. These carts could even register the weight of the entire load for a final check. When leaving the shopping area, the list of all items appeared directly on top of the cart, even with small pictures for each item, so I could control in real time that everything was accounted correctly. The bill was stored to my pocket computer and backed up via the next web hotspot. Everything under control.

I decided I had a little time left to walk about in the mall.

This year they replaced their Santa with a virtual one. Bad idea, really. How could that cyberfake lift the kids, like the real one ? Couldn't even take a normal photo with Santa*. These virtual guys behave exactly like ghosts: cast no shadow and can't be seen in a mirror. How about wearing his head under his arm ? I guess they will realize, sooner or later.

These window dressings however galore. Some items look alive, I point at them and they speak (sort of "buy me", what else). Others start a little movie. And there are some that are entirely virtual. Some windows sometimes even display different goods for different people. Some are adults only, for example.

The fashion shop just installed a new system. Entering the store I'm being asked if it should do a microwave body scan to get my size. Well, that's not dangerous with data protection being quite rigorous now. So why not.

Now the mainly radio chip based inventory system highlights anything in the shelves and racks that would fit. I enter some styles and items I'd possibly be interested in, to reduce it a bit. Then just by looking at something, I'm trying on this and that. The system sends me a simulated projection of me with that garment, walking and turning in the aisle. Quite silly, but much easier than trying anything on for real. I would do this as well if I'd really liked something, but compared to the tiresome shopping ventures I remember from the past, this is a lot less exhausting.

Some teens nearby also seem to use the system extensively, although they have to rely to the sleeve displays of their 'intelligent' designer sweaters (especially girls still don't like glasses and accept any hardship to avoid them).

---

* My vision sim as well as some cameras can do it, by recording what I see.

At least they always have some equipment with them. Some people ('technophobes') still live without any computer, and just in case I have to show them a document or something, I took care that my new glasses have one of these pin head sized laser projectors. Although small, this unit draws tremendous power however. Not just something for all occasions.

Once the store has recorded my size, I can also visit their web shop try anything on. So I may as well go home now. Anyway, why buying garment if anybody wears a vision sim - could as well just simulate it.

Reading a book

Sometimes the old paper stuff is still useful. Before we had these ergonomically designed virtual interfaces, reading something on a computer was much less convenient than just having it in paper. Worst of all, the whole thing was confined to bulky machines (I also consider a notebook too bulky for the purpose).

There had been some attempts to build special e-book readers, some with LCD displays (not a good idea), some with those polymer e-paper displays (usually poor in contrast), and all with some terrible disadvantages, be it resolution, handling, and so on. Anyway, it doesn't pay to build a computer just for one purpose.

Now it's better, my vision sim can emulate a book where I can flip pages with my fingers and even make dents and other naughty things. I can even use it in bed, as easy as a real book or a dedicated e-book reader, but try that with a notebook.

What is not possible with all conventional solutions, just to lean back and read, the pages large and above in the air. I even like to write this way, for small corrections and such, even if I have to hack on an enlarged virtual keyboard, 'search and destroy' style.

Nevertheless there are so many real books that I still use. I have set up an autostart to the scanner application that stores pages I have read, so I can access them anywhere later on. It does this simply through the position sensor cameras. It also 'binds' me a book of those pages and sorts them just by the page numbers it recognizes. I guess a lot of people are using this feature. You never know. Companies have taken numerous measures to keep their employees from scanning every document they see, as well as from filming the entire site. But that's another story. Let me only say so much that it has been established as a personal right to record one's own life anywhere anytime, in picture and sound, except for using the data to trace others or to distribute things without allowance.

With implanted computers expected to become a mass product in a few years, this was mandatory. Scientists even expect that technology will soon allow anybody to acquire *biological* memory capabilities like certain 'savants'[1], without the disadvantages.

Here I perfectly agree. I remember when I was at highschool, I once tried to learn 20 pages from the history book for a critical test. After repeating the first half of it for many times, I suddenly noticed that I stored the remaining pages just while reading. I couldn't repeat this experience later on, maybe I was too lazy to try hard enough, but I'm absolutely convinced that everybody could achieve it. So in any case, forbidding anybody to record what he sees or hears is an attack on basic personality rights.

My sim now really ensures that I almost never forget anything any more. At least if I can still figure out how to find it in those terabytes of data. If that app wouldn't also store an OCR version of texts, (that still cannot always replace the facsimile) and a signature of the pictures, I'd sometimes really be lost trying to find something again.

The most peculiar thing with vision sims and printed media are books and newspapers that have code stamps on their pages, or hidden[2] in the pictures, causing the reader's sim to display animated pictures (from the web) right in the pages, as in the ancient Harry Potter movies.

---

[1] Persons who remember anything unfiltered, for example thousands of books word by word. Usually also a disease as other capabilities are seriously affected, and as there often is no real control over the flood of data.

[2] Watermarks embedded in the pictures, too faint to be seen by the human eye but detectable by electronic cameras. Can contain quite a lot of encoded information (keyword: steganography).

The downside has been that before privacy laws were tightened, you never knew if your reading habits weren't traced and by whom. Some books even charged you for reading.

Nowadays it's granted that you've already paid for the extra web content with the book. It's also a permanent value, as web add-ons of vintage books are routinely transferred to the world online library.

Some of my books don't even need this service, as they have built-in storage and radio chips and generators or solar cells, and can deliver their own enhanced media as soon as the book is opened.

Stealing a vision simulator

What I hadn't expected in this neighborhood: somebody managed to steal my glasses. Don't know what he wanted to do with it. Those devices are so heavily coded that probably nobody could crack them. The only use would be as replacement parts, but who would care to steal a $5 laser unit (those items have become dirt cheap since they are built by the millions).

The guy also couldn't hope to get into my computer with this device, as it would only work for someone with just exactly my iris pattern and anyway it would instantly lock up when the connection was ripped.

I just bought a new pair of glasses, the most expensive part of them being the corrective lenses for my eyes. The vision simulator part of the new device was far more advanced now, an ultra light low power unit.

When my pocket computer got lost 2 years ago, it was more money but less hassle, as I didn't have to wait 2 days for new correction glasses like this time. I had good web connection right before, so all data were backed up in a most current state. Only the last hours of the vision cams were lost, as I don't upload them to the public network in raw format. This kind of bandwidth is still not for free.

Lost keys

Usually I'm running the camcorder function of my position cameras all the time. They record anything I see. Sometimes this is really useful. Two days ago I lost my keys.

Fortunately I have an advanced recognition software with a special object tracking feature. It saw and categorized the keys any time I used them, and allowed me to find the decisive second in my personal recordings in an instant. These cameras also catalog every item I store in the cupboards and allow a random search for it.

The downside of the recording function is that you never know if the guy next to you hasn't rented his vision simulator cameras to the police or someone else, although this is heavily restricted now. When laws were more negligent and total surveillance quite common, everyone not wearing a vision simulator was rightout suspicious of being a drug dealer or something.

An unexpected bus ride

Never thought that I would ever have to face this: my car in the garage for repair and my wife out with our second one. And then suddenly I need to go downtown for a meeting. Need a cab or a bus I guess. The weather is a bit gray but it doesn't rain and I'll use my mood lifter software to make the sky blue and add little patches of sunlight to the scenery.

Guess I'll get a hint about transport from the web. Well that's pretty amazing: the bus company has a new front-end that can lead me from my home just where I have to go, traces my location, displays guiding arrows and continuously informs me where any and each bus is that I could use, when I would arrive and what it costs. The app also lists alternate routes, with the subway for example (it's obviously a front-end made by the city), or can even point me to a bicycle rental. Buses aren't always ideally scheduled or routed at all, but with this kind of guidance I guess they can be considered an option for occasional users the first time in history, even more so as they automatically schedule and call a cab for route parts they can't serve.

Nice, that one. I only wonder why I never encountered something like it for a classical mobile phone. I've heard of several plans doing this or that. Probably it all didn't really succeed because of ergonomical reasons, price tags, whatsoever.

Keyword 'ergonomical': once long ago I encountered a railway ticket machine that could quickly and easily display any line and schedule in the nation just by walking the menus with two big 'mouse wheels'. I've never understood why they didn't equip mobile phones with mouse wheels. Really no idea.

The meeting of the ghosts

This afternoon we had a virtual conference with one of our overseas partners. They just upgraded their system. Now they also have one of those surround camera arrays, and their transmission allows for a perfect 3D reconstruction of each participant. No comparison to their old screen based system any more.

Sometimes I have to pinch myself to realize that these guys are not really sitting here. This is so perfect that everyone can even walk around in the room, and around one anothe. Needs a host of cameras, and a lot of computing power and bandwidth (even though it always only transmits those details that any of the participants needs to see). But who cares, this saves a fortune in travel costs, and lots of nerves as well.

With the pseudo holographic acquisition cards originally designed for media production, such setups can now be flexibly installed anywhere, and they have also become very affordable.

We have had this system for many years in every office. Most being home offices of course. The camera arrays allow to simulate virtual wall breakthroughs and doors in any direction, connecting these offices as if they were adjacent in the same building, or even in one big room. I can't just walk over to the next office of course, or I'd hit my - very real - wall (rubber walls are popular these days). In this case I have to tell my vision simulator to virtually transport me there, so the rooms are merged. They also provide a silly 'beaming' effect if I want. After this, it just feels as if I was really there.

These camera arrays may be a larger installation, but they are definitely better than the cheap 'avatar' solution that tries to synthesize my own picture from just what my vision simulator can get. The only real synthesizing effort with the camera array, except for perspective, is that eyes have to be simulated in a person's picture because one's vision sim's mask display will normally cover them in order to insert the picture of the opponent. The basis for the eye pictures is provided by the vision sim's eye trackers, so the merged picture is entirely real and accurate.

For just calling someone, I usually just position myself in front of a mirror. This only works for face to face connections, but it's very simple and requires very little synthesizing.

My washing machine

Even the most die-hard computer junkie sometimes needs fresh garment. Guess I have to start my washing machine. This one has a virtual control panel as almost any one you could get these days. I load the machine, then I think it's pretty stupid to bow in front of it to set up the program. I detach the virtual panel and return to my couch, to finish the job right there.

Don't even think I'd use the standard panel. There are numerous guys writing their own fancy controller applications for such a frequent washing machine, so I downloaded one of these that would best fit what I was thinking a washing machine should look like. It has some washing programs that are quite different form those the manufacturer originally built in. Most important, it only requires a few clicks to start what I just need it to do. Bingo, click that panel away, can restart it later on to see how far that washing has gone. This machine even has a drum camera. To see if the cat has gotten in, I'd guess.

Plants, lamps and old stuff

Some people have virtual pets, many have virtual plants. Those v-pets I really deem a bit crazy. I'm also not quite so comfortable with virtual plants. They don't have any good influence on air and climate, so I prefer to take the extra work of growing the real ones.

Of course I use some gadgets that make it less cumbersome. My watering system is quite handy: little units with a tank and a moisture sensor. They can serve up to 5 flowerpots each, are solar powered from the normal room lighting, and have wireless network.

I've configured individual virtual controls at each flowerpot, that are each activated only when I hit a little button (it´s the only thing remaining visible when not in use, so this stuff doesn't clutter up the view to the flowers).

So I can give each plant an individual program right in location.

That 'green thumb' software keeps track of the plant's appearances by 3D recording them and lets me literally see them grow by recalling last month's pictures.

I can also access each control remotely everywhere I want, or have them all in a neat list. The same possibilities I have with my window blinds, the heating, the pool, anything. All of the controller units consist of just a chip, no displays, no keys, smallest dimensions, small power consumption. Those far from AC current use little solar units or long lasting batteries.

There's also no lamp in the house that hasn't a virtual control interface, even though it simply duplicates nothing but the power switch. These controller units are made by the millions and are dirt cheap now. Many are nothing but simple infrared or radio receivers. Some carry only a bar or block code that tells my sim how to address them, especially the simple switches. Some I have to program into my sim. Before sims became common, switches were in fashion that had cameras of their own and could see if you looked at them. I still use some of these.

The older hardware stuff standing around here required some setup as well, but in most cases my sim just looked it up in a web database just by its picture, and immediately came up with the right programming codes and even a picture of the original remote that I could operate wherever desired.

One or two needed special treatment, but my sim also works as a learning remote, no problem.

What I like most with it, that no remotes are lying around any more, and that I can just operate anything here with eye pointing alone. Sort of magic, that is.

Garage job

I think I have to do some car maintenance today. One of the bus operated relays seems to be defective. The controls show that the right high beam doesn't work, but that it's only the relay that doesn't respond, not the lamp.

This isn't something you need a central computer for, nowadays. If these cars have any general purpose computers, it's just a recording unit or a tripmaster, noting that interferes with anything that could be safety relevant, and all units independent of each other, the only concept that can avoid complex and potentially dangerous malfunctions to develop.

When they finally realized that super processors drawing lots of power and being run with complicated operating systems were just not appropriate for relatively simple tasks, they managed to standardize a few bus based universal switches and relays. These are now produced by the hundreds of millions and can be purchased at any gas station. As with all technology, once mature it finally becomes reliable.

Cars were mature before electronics (except for the cylinder head gasket), then they became progressively immature when more and more experiments were built in. These were mainly used to secure revenue by hosts of patents and part prices rightout insane. Finally, when car electronics also became of age, things became reasonable once more.

Something I always wondered about was how long they sticked to driving valves with a camshaft. My last fuel engine had electromagnetic ones and could 'drink' almost anything liquid.

Meanwhile they've also standardized things like the motor control unit (actually several modular units for better service) - all simple dedicated chips that are identical for any car and just get teached in with a little program - and several other standard items (no, this car here has none of these combustion engines any more, all electrical).

None of these modular controllers can just quit with a software error any more, and they are all autonomous. Together with two power buses (one for backup) and the also redundantly wired fiber bus, these ACME* [28] standard components make cars as reliable and simply maintainable as my son's tricycle.

So the switch told me the relay was defective. I've just opened the hood, and now my sim shows me right away where it is.

The maintenance recording unit - also a totally independent device - serves me with detailed 3D drawings and wire plans, overlaid to the motor cabinet, and lets me see through everything there is.

Maintenance applications like this one had been implemented with simpler display units long ago. Meanwhile it's common and cheap, and car manufacturers have found out that it is a good idea to integrate this knowledge right into the car. Many other hardware products have meanwhile also been equipped with this feature.

This defective relay even flashes, of course the maintenance unit has also noticed it, by listening to the bus communications. So changing it is done in a second, The socket contains a code that tells the relay what to do from now on, and that's it. I didn't even have to go to the next gas station for this. These items are so current that I have some in stock all the time.

---

* The Original Illustrated Catalog of ACME Products, by GP Markham. ACME has built a reputation by supplying all kinds of equipment to many well-known comic characters. The product range comprises anything from birdseed to atom rearrangers to cars or jet propelled pogo sticks.

Living history

Last year I was in Italy and decided to visit some historic sites for the first time. Twice I had been in Venice before, but never in Rome or Pompeii.

Meanwhile they have installed virtuality transmitters all over. In Venice for example, you can now visit the Arsenal and watch ancient shipwrights producing battleships in series, in a perfect assembly line technique.

Walking through the ancient ruins of Pompeii, you can chose to see the buildings as they are, or with virtual reconstructions added, so the entire city may look exactly as it was before the volcano (or at least as the historians think it had looked).

This is very impressive. One could build such a site entirely virtual, but it's by no means comparable to having the real thing alongside with the enhanced views.

Every two hours the volcano breaks out, pyroclastic flows are rushing down the slopes of mount Vesuvius, ashes are raining (virtually, I really appreciate if this doesn't get too real), and horrified citizens are running through the streets and sometimes right trough you. A really ground shaking sound system supports the volcano eruptions, but they're using it very cautiously because it could damage the ruins.

All pictures and local sound events are provided by a cellular network of image transmitters, each to be received from a short distance only. Any vision simulator can decide which of the virtual objects offered are near enough to reproduce, and build a complete virtual scene combined from these structures.

For a large area, an ancient town for example, this has proven the most efficient method. Transmitting views of thousands of buildings and other things from just one central facility would be too slow in most cases.

You could also see the entire site from a helicopter. It looks a bit as with an ancient flight simulator, with buildings appearing one by one as you approach.

In Rome, in the Coliseum, it's been accomplished with one central cinema transmitter. All of a sudden you find yourself in a blood thirsty crowd, squealing and yelling at the atrocities taking place in the arena. Your fellow visitors also suddenly change, their shorts and t-shirts replaced by togas and the omnipresent Klompen sweat shoes replaced by sandals.

If you're standing in the arena, it's even more thrilling, although I got the impression that the lions always go for the more nutritious guys.

Many historic sites now have their virtuality transmitters. Landmarks also do, some only working similar to virtual signs, but many are like virtual cinemas. On an ancient battlefield you may suddenly be in for a shock, if you didn't watch for the acceptance levels of your sim.

The most impressive makeups may be sea battles. You can charter a boat and watch the battles of Trafalgar, Midway, or Lepanto if you like, at their original sites or at similar locations where special theme parks have been installed. Original sites are the most exclusive, of course.

This is lightyears ahead of the monster cinemas that once dominated theme park assemblies. You can actually get right in the middle of the battle, in case of the sea variety with a real ship, and if you avoid running right trough the virtual ones it will definitely feel quite real. Too real maybe. Will it show people the insanity of these events, or will it just blunt their feelings? Hard to say in any case.

With all virtual landmark sites or theme parks it's advisable to shield the open edges of your sim, so you don't see the nothing beneath the display area, that would disturb the illusion. The sims they offer for short time rent at the sites are all built this way.

Did I mention the dinosaur park ? Real beasts would be more impressive, but I would not like to find myself so close by them.

Lost in space

An exceptional type of theme park application has emerged with the combination of vision simulators with underwater setups long used in astronaut training. You get stuffed into a wetsuit and a helmet with advanced display and oxygen supply. Then they throw you in a water tank where your position is fixed by jet blowers. It feels weightless, although your stomach still knows it isn't (which is an advantage, as vomiting into the helmet would be horrible). The really important part however is the vision simulation: you can look around and there is no doubt you're floating in free space, 100 miles above the earth, and the image is not only a perfect illusion but also crisp enough to see tiniest detail on the earth's surface. At one side, myriads of stars are twinkling (you'll tell me they aren't, as there is no atmosphere, but it sounds better), on the other side the sun is terribly blinding. Couldn't be more real. Even an entire space suit is simulated when you look at yourself. The ultimate 'kick' comes when you're all of a sudden accelerated towards the moon, fly around it, and then blast off towards mars and Jupiter like as with a warp drive.

Feeling no acceleration force here is really SciFi, but who knows- recently the there was an article about magnetic nano particles connected with proteins. An astronaut stuffed up with this should weigh 20 pounds more, but they claimed that in the field of some superconducting magnets, he could stand 20 or 30 g of acceleration. Absolutely weird.

The planets program takes at least an hour, as there is so much to see and you don't want to accelerate sunrises and sunsets on alien planets, nor other spectacular occurrences.

Any picture material available from space missions has been incorporated with the simulation, in ultimate detail, and a sophisticated software merges it into a perfect world model.

Optionally, you can have your own flight controls or even a spacecraft simulated, to conduct your own expedition program. Even researchers like to use these high end simulators, to get a real look and feel of their object of studies.

Something I personally will never do, is to book a vacation especially for these occasions.

I'd prefer some weeks of pure nature, without a vision simulator.

BOONDOCKS OR BUST U

The following pages have also been found with the historic 'files from the rubble' (early 21st century). Many of the envisioned technologies were in common use a few decades after. It may be of interest to compare the suggested solutions to the partly far more elegant technology of later vision simulators.

part 3: Technical Design

General considerations

In the following I will present and analyze possible technical solutions for a virtuality interface that seamlessly integrates into everyday environment.

Let's shortly repeat the fundamentals:

Main objective is to enable the creation of virtual devices, virtual 'hardware' objects that are fully integrated into reality, with the default option of being fixed to real objects or locations, just as most real things. This implies that the visualization equipment must be able to orientate itself in a natural environment, in order to 'know' where it is and how to accommodate visual perspectives.
A special form of object and scenery recognition is required, three dimensional image analysis (stereo viewing) and other technologies are necessary. I will show a realistically feasible approach to this.
Also very important is manual interaction. I will show an approach to manual or visual steering with virtual keys that needs extremely little computing power.
The optics necessary will go far beyond currently available solutions. Anything published so far does not fulfil the objectives according to size, weight, and wearability.
One reason may be that large industrial companies, who would have the means, have regarded this technology as a niche application not really worth much effort. The other reason is that creating 'perfect' optics is not the smartest approach.
It is quite evident that unconventional ways of construction have to be explored.
Dynamic geometry and focus correction, 'exotic' optical elements like dichroic filters or holographic lenses and mirrors, even non planar displays, have to be considered.

We will be exploring some possible approaches to very light and convenient constructions.

We will further discuss additional components for 'mixed virtuality', wireless interfacing/networking, virtual object handling and security, and more.

Let's recall why we are doing all this:

Advantages over conventional VR solutions

Natural objects stick to the environment, not to the user's field of view. This should also be the default option for virtual objects. It keeps sight free and enables the use of many more virtual objects at once. Our entire capability to deal with complex 3D environments can be unleashed this way.

Information interfaces should be reduced to the very basic, they should just deliver pictures and sound as directly to the user's senses as possible and become so light and small that they go almost unnoticed.

Advantages over conventional 3D display solutions

All varieties of electronic 3D displays known so far, force the eye to focus to a different distance than the stereo perspective implies. Most do not care for perspective changes by head movements. Most cannot adjust for different viewing distances. In one or the other way, none so far reproduce all aspects of stereo viewing accurately. This results in a more or less unrealistic impression and considerable eye wear.

With the approach taken here, it should be possible to resolve these problems entirely.

Last but not least, the virtual display area and resolution of a vision simulator can be orders of magnitude larger than that of any static screen, as only a part of the entire angular range has to be displayed at any given time.

Why it can be done

A human interface has to be flexible. It has to endure or adapt to motion and all sorts of dynamic influences. Dealing with this will make display glasses, a vision simulator as I'm calling it here, much more complex but also, in case of optics, more simple.

In principle, there would be no other optical part necessary in a vision simulator than a single concave mirror. This way we get a large field of view and considerable enlargement with very little weight. Inevitably we will see image distortion, but this can obviously best be dealt with by software based precompensation.

We may also try to develop displays with uneven pixel distribution or with a convex surface. Organic LED displays for example could be made this way.

Lenses or prisms, except for small ones, are not as well suited for the purpose, because of weight. A curved mirror is our first choice as the main optical element, because it's not such a good idea just to use a planar glass to mirror images into the eye, as this doesn't only look bad, but also doesn't allow large viewing angles. A mirror will probably be a good choice for a second optical element as well, if necessary.

We may need aspherical optics and other more unusual things. The complexity of the task will require extensive optical calculations and a lot of unconventional methods. Mirrors and other elements could also be holographic ones, if we deal with monochrome light.

Focus and aperture may be critical, but it is absolutely necessary anyway, that we provide a dynamic compensation for this, based on the input of an eye tracker. This also allows us to do away with special fixtures of the VR device to the head. An acceptable device should always be freely movable like ordinary glasses, without becoming dysfunctional. With eye tracking and a servo system, we can achieve this.

An eye tracker and dynamic adjustments of focus and geometry will provide entirely new options for display design.

Given the light weight optics and the possibility to use small displays, the entire vision simulator (without computer) can be made almost as light as any modern pair of ordinary plastic glasses.

We also need a perfect position and motion sensor for our device. Key ingredients of position sensing are stereo cameras attached to the glasses. Usually, sensing positions with such devices is deemed anything but trivial. Yet for our approach, we do not require the cameras to *recognize* anything.
Instead, I am proposing an approach that relies on simplest image details as lines, angles, etc., normalizes these image 'atoms' to a standard perspective, size, orientation and illumination, then stores them together with geometrical pointers to other 'atoms' and to virtual objects. The atom and pointer data structure enables a description of more complex scenes.
We do not need any separation or recognition of real objects. It is of absolutely no concern if image atoms or structures describe one object or are distributed between many objects. It is especially unimportant to assign any meaning to any object. We do not try image recognition but image remembering, and only of the very simplest and most prevalent detail.
This somehow resembles the way humans are believed to recognize images. As soon as devices become available that simulate the human visual and neocortex structures [12], these may perhaps greatly accelerate the process.

The data structure would enable us to search for a resemblance of newly seen image structures to stored ones very quickly. As we never need to store any image data from places where we did not place virtual objects, the database remains relatively small. Given the power of modern processors, the search for similarities can therefore be carried out in a very short time.
A GPS or similar device will further reduce the difficulty, by excluding irrelevant places and quickly resolving ambiguities.
Once relevant detail is recognized, even standard video cameras can provide for a millimeter accurate position sensing.

Acceleration sensors will allow to compensate for effects resulting from the limited speed (frame rate) of cameras and displays. Head movements can be predicted from speed and acceleration at a given time, quite precisely, due to the relatively big inertia of the head.

Gesture recognition has always been a challenge. It's also not even really necessary. I don't appreciate any solutions that require gloves or other special equipment. Recognizing human hands or free gestures I don't consider either, as it has always been difficult and it's even a bit peculiar: imagine the computer mouse would have been introduced with no keys, and anything had to be entered with it by gestures. Queer, isn't it?

Instead, we should let the display project virtual keys or sensitive areas in objects, and scan the stereo camera's outputs for any correlated signal change indicating a movement exactly and only in these areas.

This is very easy and needs very little computing power. Adding some simple heuristic checks can secure the algorithm against false triggering. Once a finger is identified in a key area, it can as well be traced to some extent very easily, in order to enable more complicated actions (grabbing and positioning an object, or throwing an object, for example).

Last but not least, we would usually *look* a the key we are pressing. Hence, we could use the eye tracker cameras to implement eye pointing and this could improve security and also enable remote key pressing.

It would not be a problem to activate the surface of entire objects to be 'touched' and switched or dragged by hand or eyes. Indeed, gestures only matter if we want to drag objects. In this case we could implement a 'lock' function and then follow the hand. It's also possible to exploit accelerated movements in order to virtually 'throw' objects, in a heuristic context with environment structure, because this would be one way, for example, to define a video screen on a wall. This is still way below the difficulty of mainstream gesture recognition.

Dynamic action like this has already been used with a 3D 'mouse' device by a small Israeli company: they have built a device that traces an ultrasound source with 3 microphones attached to a display. A Game application they delivered allowed to throw virtual rings over a stick, by just mimicking the throwing action with the ultrasound emitter. The rings 'thrown into the display' would then appear in there and fly as if they were real. It worked perfectly well, so I have no doubt that this kind of action will be an easy way to handle remote virtual objects.

Eye pointing however will be the most elegant way to deal with remote objects anyway, as it doesn't only deliver an unambiguous direction information, but also a distance information because of of eye convergence ('squinting'). Nobody usually peers at anything for more than a second without a reason, and blinking could also be exploited for some actions, so this could all work as a complete way of interfacing.

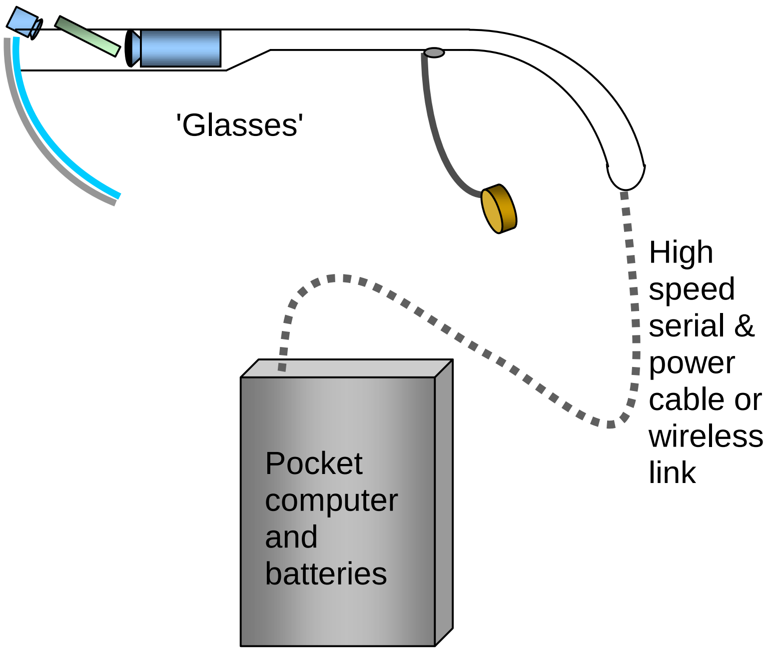
Sharing virtual objects with other people would be most natural. Everyone involved should be able to manipulate them. This is another field for research and I will also address it in this chapter.

I don't think it would be desirable to replace all real input devices with virtual ones. A keyboard for example is difficult to implement in virtual. Combining virtual machines with a foldable or rollable *real* keyboard, for example, could be the better way, without compromising portability. Using an active pen or pointer device could also be of advantage.

Given all the different tasks to address, the vision simulator will still be a very complex device and will need considerable research in many different fields to get it accomplished in a marketable way.

In this chapter, we will explore possible solutions to many of the tasks mentioned. This will inevitably be far from comprehensive. It is mainly intended as a proof of concept. If vision simulators will have the huge impact on technology I am expecting, first implementations will initiate an extraordinarily dynamic development, leading to many surprising new twists anyway.

Hardware assembly



Quite obviously, we want to reduce the weight of the glasses assembly as far as possible.

A powerful CPU won't fit into this part in a foreseeable future.

So the aim is clear: the glasses should only contain the most necessary hardware. Images, sensor and actuator data should be transported in both directions by radio or a thin cable.

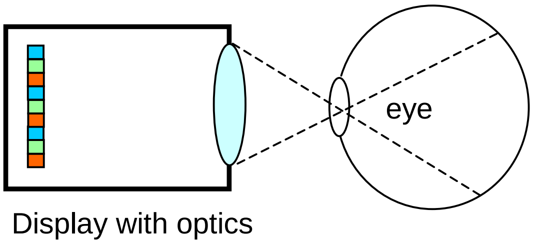
A modern serial bus like USB2 can transport over 400 Mbit/s over a single wire, that could also be used to carry power. So we need only one very thin cable. Any computing whatsoever is performed in the pocket unit that also takes care of all other interfacing, like wireless networking.

In the glasses assembly, we only need data compression, serial bus, display driver and camera chips, some actuators, to mention the most obvious. These must be optimized for size and power consumption (the pocket computer as well, of course).

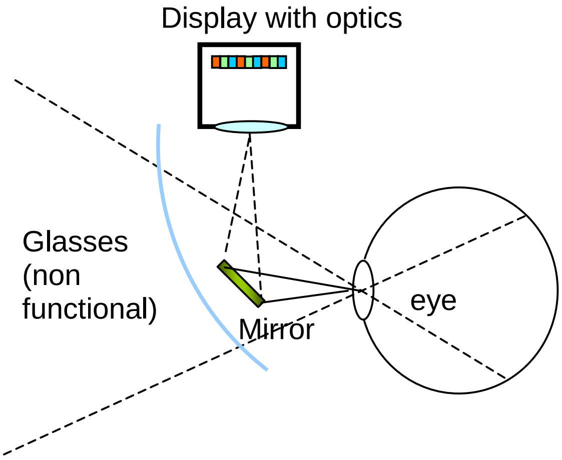
It is by no means physically impossible to reduce the power consumption of the glasses unit far enough as to use separate batteries and wireless data connection. Then we could even drop the pocket computer if we could link to a home or office network. Even real soon now. Read on for details.

Some basic display designs

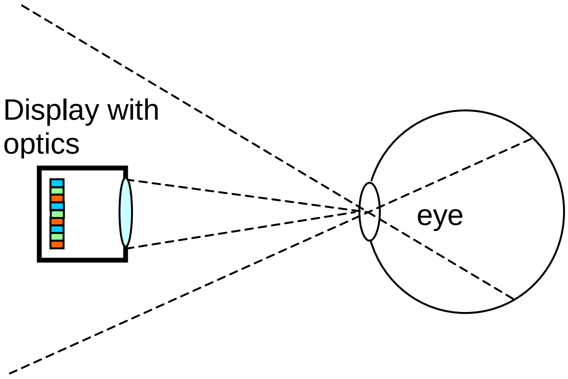
Let's first have a look at some conventional VR display types:



Immersive displays

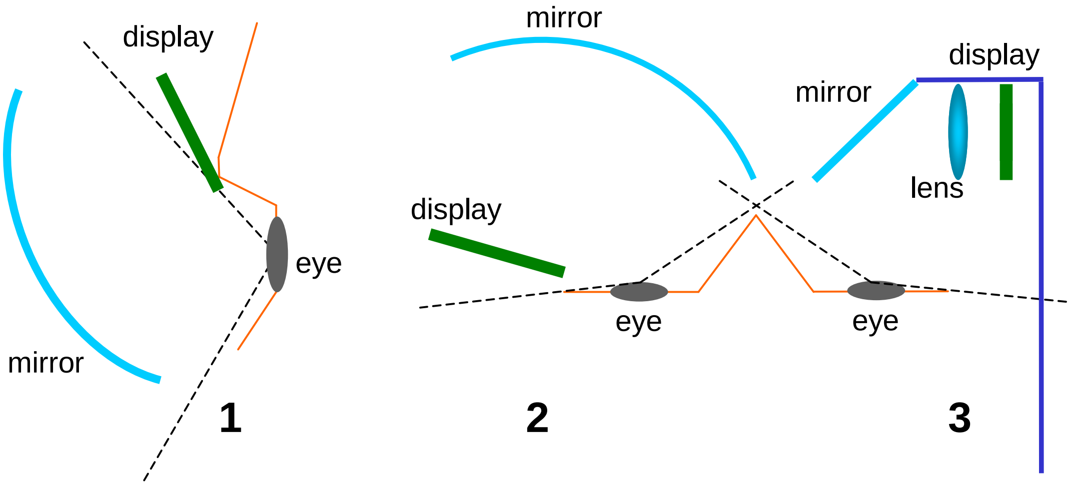


Typical personal TV display (one-eyed)



Personal data display (one-eyed). Allows to look around the display by turning the head

Some varieties of see-through displays



I won't consider any assemblies that don't allow direct sight to the real world, as the eyetap [25] or other approaches that require the user to see everything through cameras. These won't be acceptable for everyday use in a foreseeable future, if at all.

What we need here are see-through displays mirroring additional items into the field of view. The question is what the mirror should look like and where we place the display.

A natural dead angle is at the eyebrow and above. A display positioned there covers almost no direct sight and allows for a relatively wide projection image (1).

The same optical assembly, with a conventional display at the *side* (2), could cause some obstruction and have a smaller projection angle sideways, where humans have a large viewing range, over 180 degrees. Yet with small display assemblies, laser or holographic for example, this could also work quite well.

A typical commercial data display assembly (3) causes even more obstruction, although very light and convenient ones are available.

I will therefore start the design considerations with the display positioned above the eye. All these assemblies could as well be arranged as in example (2).

Some examples of current displays

MicroOptical CV-3 video viewer
monocular, color, 640x480 pixels, 20^0x16^0 35g(viewer) [14]
(photo courtesy of MicroOptical Corporation)

Liteye LE-500 [54]; monocular, 800x600 pixel OLED display, see-through optional, 28^0 field of view, 120 grams
(photo courtesy of Liteye Systems)

i-glasses theater: 112g, 230000 pixels
(photo courtesy of i-O Display Systems, LLC)

i-glasses 3D Pro, 600x800x3 color pixels
(photo courtesy of i-O Display Systems, LLC)

Kaiser Proview XL40 ST: 1024x768 green, 45^0 diagonal field of view, see-through, 900g (photo courtesy of Rockwell Collins)

Cybermind Visette 45 XSGA 1280x1024 pixels, 750g,
see-through optional (photo courtesy of Cybermind BV)

Saab Addvisor 150: 1280x1024 color, 46^0, see-through
(photo courtesy of Saab Avitronics)

NVIS NVISOR SX [55], binocular 1280x1024, 1kg, 60^0.
Eye tracker (Arrington Research, [56]) optional
(photo courtesy of NVIS)

Arrington Research Viewpoint PC-60 eye tracker [56],
mounted on glasses and on a HMD

eMagin Z800 3DVisor [60], binocular 600x800 color, OLED on silicon displays, 40^0 diagonal field of view, equipped with an acceleration based tracking system. (photo courtesy of eMagin Corporation)

Mirage LightVu [53] OEM optical module for display glasses. The glasses unit weighs 40 grams. The 'nanoprism' assembly (non transparent, see-around) creates a virtual 42" screen from a single display (photo courtesy of Mirage Innovations)

An overview of current products can also for example be found on the EST website [29], or at Stereo3D.com [50].

In conclusion, there are some products with see-through and high resolution, that would allow to implement some of the new applications we have discussed, but mainly for scientific evaluation, development, and professional applications, as these current civil high end augmented reality displays are too bulky and heavy to wear in everyday life (not to mention the costs, but this problem would vanish with mass production). Other products, as the MicroOptical CV3 and the Mirage LightVu are examples how light display units could actually be built, although these models aren't just made for the applications considered here.

Extras like earphones or position sensors are present in certain products as well, but no one has even nearly all features we need. The 1280x1024 pixels resolution and ≈45^0 diagonal field of view offered by some current products are but a little less than what we would like to have (e.g. 1600x1200 and >60^0, ideally 3000x2000 at 90^0 or more; the FOHMD discussed in the first chapter already met or exceeded these specifications).

Typically, current display glasses have no dynamic compensation for position change (fit), no eye trackers (some, optional), no camera based exterior position sensors, no mask displays, and so on. For experimental setups, we could add some of those features. The entire software basis has to be developed as well, and when new markets will emerge, products will certainly be further developed and improved, concurrently with new applications.

We see that it's still quite some challenge to get to a really comfortable vision simulator, one that seamlessly integrates into everyday life.

One major task will be the display and the optics. I will therefore give high priority to considerations and suggestions on possible optical and display solutions, including eye tracking and related features.

The future: ACME [28] 'Hairstyle' display. Eye projector and cameras are concentrated in one tiny unit. The user wears a monocle with a totally passive holographic screen coating. The computer unit is contained inside the periwig. Less sudorific products are the hat based 'Stetson' and 'Sherlock' models.

Display technologies

This is a short list of currently available technologies with focus on near eye applications. It can neither be entirely complete or go into much detail. We will have a closer look on some technologies later on.

- **LCD** (Liquid Crystal Displays) use nematic liquid crystals between two glass plates with polarizer filters each. With electric fields applied, these liquid crystals turn the polarization vector of passing light, so transmission can be varied. Standard LCD designs are not so easily adaptable for a glasses display as we need it, because very small pixels are difficult to achieve (the minimum thickness of the LC fluid sets some limits). They may be a first solution for mask displays [13].

- **LCOS** (Liquid Crystal on Silicon) are LCD displays using a silicon wafer as a driving backplane. They work reflective.

- **FLC** (Ferroelectric LCD) are liquid crystal displays with 'chiral smectic C phases' that exhibit ferroelectric properties. They are more difficult to make than normal LCD but exhibit very short switching times (>10µs rather than >10ms for typical LCD).

- **F-LCOS** (Ferroelectric LCOS) displays combine LCOS and FLC technologies. They can be produced with very small pixel sizes (currently 8µm, but 4µm or less said to be possible) and are the technology of choice currently for high resolution VR goggles as well as synthetic holography [33], [51]. These displays are very fast, color is typically produced by sequential illumination, switching the light source rather than working with color filters in the display itself, advantageous especially for reflective displays.

- **DMD** displays consist of microscopic mirrors that are individually tilt by electrostatic forces (more later). Projectors using this technology are quite current already. These displays would actually perform better with smaller pixel sizes, in contrary to LCD. We could use them - with LED illumination - just like normal displays, but perhaps with laser light in an holographic mode as well.

- **Holographic** displays are an application currently with F-LCOS or DMD technology, that provides little computer generated holograms. As patterns to be displayed require resolutions in the micrometer range, these displays can only be made in small dimensions and must be used with enlargement optics [31], [36]. They may also become a good choice for vision simulators.

- **LED** (Light Emitting Diodes) could be ideal, but in order to produce light, semiconductors need to have threshold voltages (band gaps) corresponding to the energy of the photons we want. Red light photons have 1.5 electron volts, blue about 3, while silicon has a band gap of only 0,7 V. We need gallium arsenide or gallium nitrite for example, and combining these materials with a silicon driver chip for a display array is difficult. There is a product, but it's not yet available over 800x600 resolution.

- **OLED** (organic or plastic Light Emitting Diodes) could work for miniature displays, yet at the high resolutions we need, passive matrix driving would probably not suffice, so current products use a silicon driver chip with OLED coatings on top [60], even though organic transistors are in principle already available. A fully organic (plastic) design could also provide convex or concave displays, according to optical requirements. Making blue OLEDs and life expectancy are the main problems with this technology.

- **Polymer** displays are based on organic chemistry, like OLEDs. Formerly, only passive varieties ('electronic newspaper') were summarized here, but now there are also self luminating ones. Reportedly, high resolution versions are in test. Life expectancy is still not good (1 year for example). There is one variety that switches from transparent to color band absorption [7]; this may be an alternative for mask displays.

- **Transparent OLED/Polymer** displays are like a glass plate when inactive and pixels are caused to 'glow' when addressed. These displays may be useful for some novel design approaches. They are not directly useful for our approach to mixed reality however, because there is always the problem that the display and things behind it can't be focused to at the same time.

- **LASER** is very promising but also has difficulties, especially as the beams have to get thicker with high resolution (up to 2mm), and deflection has to become faster at the same time. We will discuss some aspects of laser displays later on. We'll see that with an advanced optical design, these may still become a good solution for our purpose.

- **GLV** (Grating Light Valve) is a very new variety that works with microscopic aluminum strips being switched to form an interference grid to reflect different laser colors in different angles. Current versions are only linear arrays that need a rotating mirror to scan an entire image. This is used for large projections and probably not as good for our purpose. If two dimensional arrays could be manufactured, this might be an option for holography, especially as we won't need accurately planar mirrors as with DMD, as soon as we get down to sub wavelength sizes.

- **Electrowetting** devices use microscopic oil drops that can be concentrated or spread over a pixel area by electrostatic force, to switch it from total reflection to transmission. It could be considered to make such a display with black drops, for masking purposes, yet this would likely have problems with stray light.

- **Dyed Guest Host** displays use microscopic dye disks diluted in liquid crystal fluid for example, that can be tilted by electric fields. Oriented parallel to the light path they hardly influence it, while perpendicular they form a 'solid' wall. It could be an ideal mask display, but contrast is currently only about 1:5 [33].

- **iMoD** displays [61] are working with thin transparent membranes approx. 1 um over a silicon chip; Light rays reflected from the chip and the membrane interfere, resulting in brightness change and colorizing (the same effect that creates colors on butterfly wings). Applying voltage draws the membrane to the silicon by electrostatic force, changing brightness. As the display elements are bistable (have to be, otherwise color would change), brightness is modulated by pulse width, exploiting the fast switching speed of ≈50 µs.

- **CRT** (Cathode Ray Tubes): I only mention them for completeness; miniature index color CRTs are available (UV index strips between colors provide for correct addressing of the phosphors). With our basic design (below), we could perhaps think about using a mini CRT backwards, peering through the glass cone from behind. This would also allow to construct a convex screen. Weight and size however would prohibit such an approach.

Eye physiology

Let us recapitulate some facts about human vision and their impact on display design (more about color, in the media chapter):

1) Our eyes can see sharp at the very center of view only. We don't normally notice it, but any time we want to see anything really crisp, we will direct our eyes to it. Larger scenes are perceived by quick eye movements between certain points of interest (saccades). Other detail may simply be overlooked this way.

The resolution in the center of view is 1 arcmin per pixel at most (60 pixels/degree). Interestingly this is just the maximum optical resolution possible at a pupil diameter of 2mm. A perfectly sharp display would need 3600 pixels for 60 degrees of viewing angle.

Outsides the very center of view, our eye resolution gets extremely unsharp in comparison (the exact crispness requirements are a complex function of contrast, focus of attention [76] and other factors).

The consequences for the design of a near-eye display are tremendous: Focus is necessary at the center of view only. Everything else can be unsharp. If we could adjust focus dynamically, fast of course, according to viewing direction, we could use a display with 'bad' optics that is really never sharp everywhere.

We need an eye tracker to do this, and we need to dynamically adjust not only focus but geometry as well, yet this effort gives us many degrees of freedom with our display design.

A display that is never overall crisp, nor linear, could actually be the right thing to have.

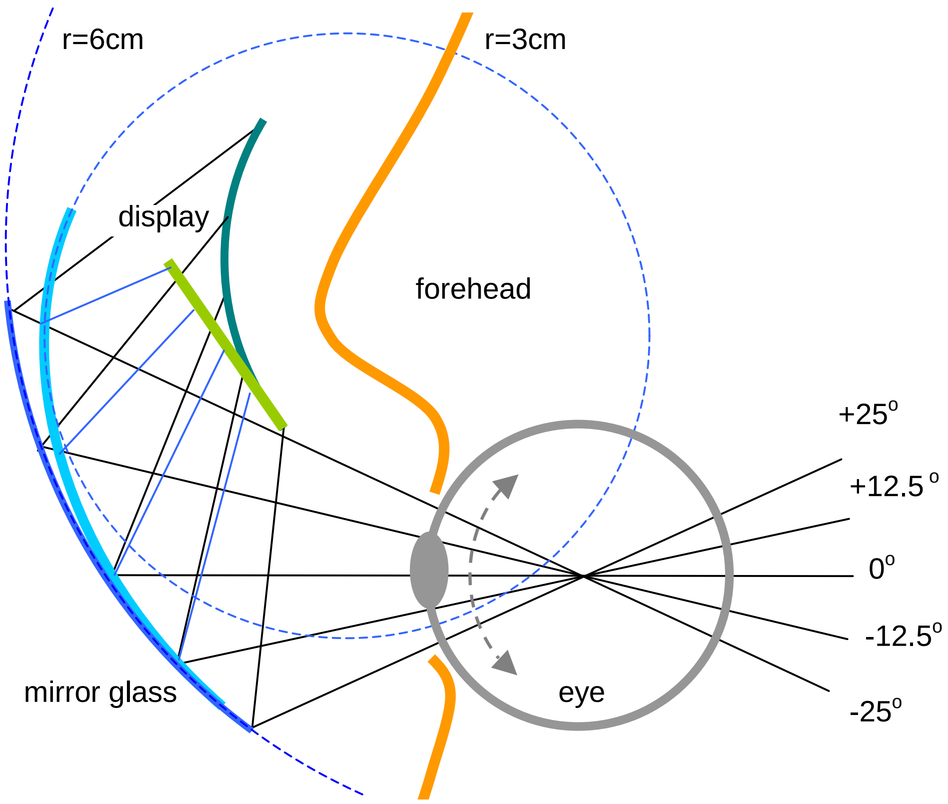
Eye movements can be very fast, so we need a fast eye tracker to follow them, faster than a usual camera at last. Eye tracker cameras don't need high resolution, but should always be 5..10 times faster than normal video cameras. Another application for an eye tracker could be to reduce bandwidth in communications, or computing effort with image synthesis as in games, with video, or just with ordinary virtual objects, if only those image parts are rendered or transmitted in full quality that are currently being looked at.

2) Another important feature of the human eye is a very large field of view, over 180^0 horizontally (though not for each eye separately, because of the nose). We do not necessarily have to achieve this much, but the general message is, the more the better. For a really good vision simulator, we should target viewing angles of 90^0 or more.
The computing power necessary could be reduced by concentrating most rendering efforts just to the center of view as described above, still providing a very high effective resolution.

3) We can see over a tremendous range of light intensity. It starts from about 1/1000 lux (starlight) and goes up to 10000 lux (Sunlight), sometimes even more.
Pupil diameters can vary between 1 mm in bright light and almost 9 mm (at night), but the maximum decreases with age, so for elder people the maximum usually is 4 or 5 mm.
A display integrating objects into reality, has to provide them with due brightness. Fortunately the display is small, so overall power consumption is not a big problem.
Current LED lights can provide huge intensities already. Part of the intensity also depends on the optics. Certain measures like micro lens arrays could further concentrate light towards the eye.
The best efficiency is achieved with laser displays, either scanner based or holographic: according to design, nearly 100% of the light produced can be sent into the eye. For this reason, laser displays currently are the technology of choice for combat applications, even though the laser diodes themselves are still pretty bad at converting electrical power into light, much less effective than current LEDs at least.
In any case, the bottom line is that we have to demand very high brightness from any vision simulator design, but it will not require much energy.

More details about eye physiology in general, you may e.g. find in [81].

An optical design study



Although the real thing may be laser or holographic displays together with holographic mirrors (see later), we may well be stuck with conventional designs for a while. So let's have a look at them.

This obvious design I devised 13 years ago but it's still current: starting from a very simple layout with a spherical mirror (dark blue, r=6cm, hence focal length=3cm), we can construct some beams (black) and see that the plane of sharpness (dark green) is itself about spherical.

To really show the focusing, we would have had to start from 2 parallel beams from the pupil for each angle, and show that they meet on the plane of sharpness. As we know the focal length already (r/2), only the reflection angles are shown and the reflected beams are cut at focus length.

We have to take into account that the user will direct his eye to the viewing direction in order to see sharp. The eye is principally a ball that turns around its center. Therefore the different central viewing beams have to meet at its center (I have not shown the pupil for all directions as this would have looked a bit too crowded). I don't consider off axis focus because I assume that this design requires dynamic eye tracker steered focus anyway and the eye can only see sharp at the center.

We see that we get a spherical focus plane, i.e. we need a spherical display or some more optics.

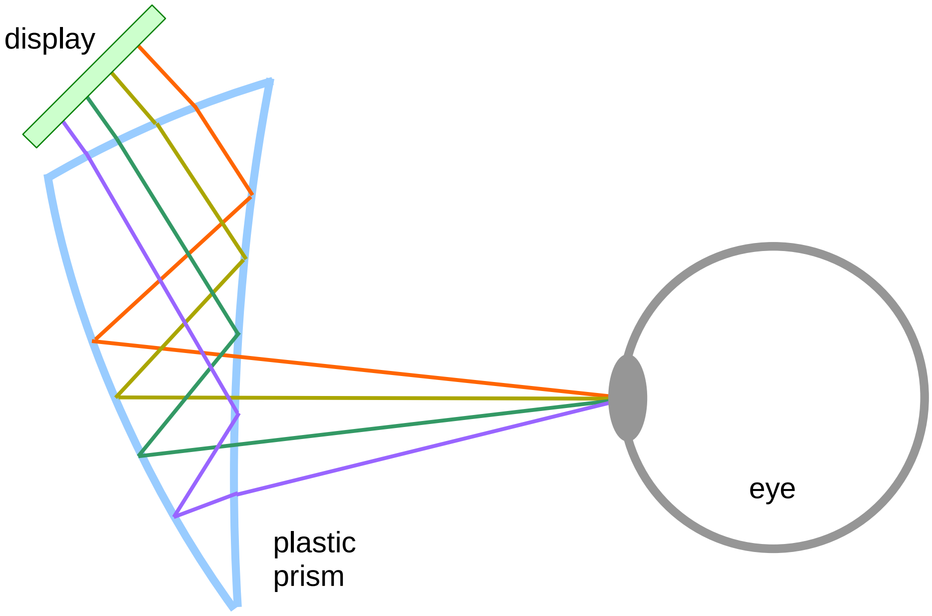
It's quite obvious that by positioning a second shaped mirror before or after the focus point of the first one, we'd get a lot more degrees of freedom for a construction, together with better magnification, which allows for a smaller display and thereby less weight. It is also quite conceivable however that such a construction has a smaller exit aperture, causing problems when the user moves his eyes or when the glasses sit incorrectly. It also requires a full fledged optical calculation to make this up, so I will not further delve into this here.

We could as well make a single main mirror glass aspherical at the top (here I have shown a gliding change to r=3cm), resulting in a flat focus plane and a smaller display area. Focus length at the top is now 1.5cm, which means 2x stronger magnification, i.e. we have to use a strong geometrical correction in the display image to get an undistorted view for the user. The design will also have to be extended into the 3rd dimension of course, making things a bit more difficult in the end. Nevertheless it's already become pretty clear, that we could provide a big field of view this way (50 degrees vertically here; moving the optic closer to the eye could even yield 60 degrees), and this with a relatively small display and a single mirror.

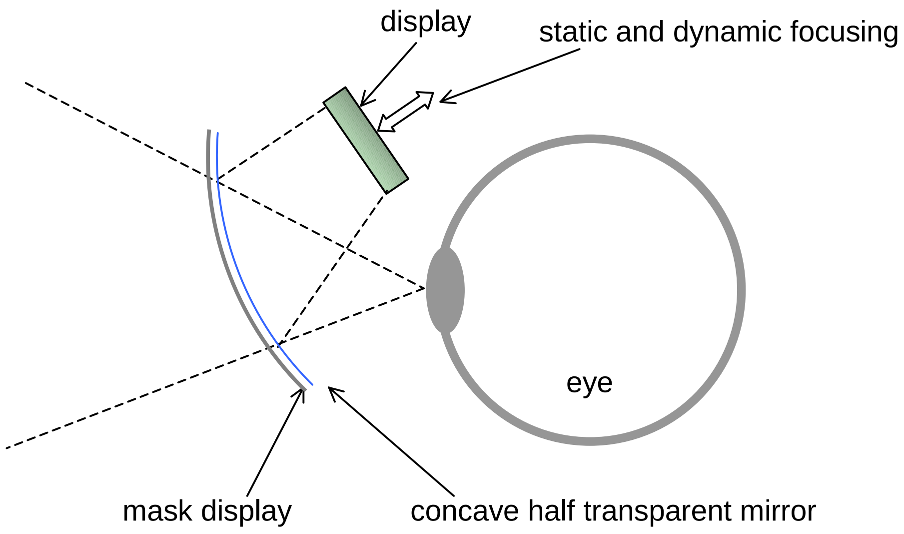
A mirror can be thin, plastic is an adequate material, so the entire assembly can be very light. If these glasses move away from the eye just by gliding a bit down the nose, as spectacles tend to do, then we'd get a somewhat smaller field of view, but the image generator - with help from the eye tracker - could keep perspective and focus in order, so we don't need a special fixture to the head. This is an essential advantage compared to many current designs. The concave mirror largely resembles typical sunglasses.

It combines light weight with strong enlargement and no color aberration. Its geometric image distortions are electronically corrected in the display resp. the display generating computer.

As stated, a second mirror could improve magnification and give more degrees of freedom for the optical construction.
Another approach with 2 optical surfaces (actually three, as there is a surface in front of the display that acts as a lens) has emerged very recently:



This patented prism based optical system of a commercial product (eMagin,[61]) has strong magnification and a decent viewing angle (not yet quite as much as we'd like to have). The prism is very light (about 11g). Obviously this doesn't allow see-through, as the prism would heavily distort any direct view, but if we'd replace the front and rear surfaces by thin glasses with mirror coatings, this could work just as well. With 2 surfaces in the line of direct sight, advanced mirrors technology will of course be required, to keep light absorption in an acceptable range.
There are more things we have to consider with display glasses than just the optical path. I'll show these with the simple one mirror approach, but everything basically applies to more complicated designs as well.



The display should be axially displaceable, which serves for focusing. A dynamic alteration of focus allows, in conjunction with an eye tracker (not shown here), to simulate the true distance of an imaginary object displayed (later on we'll see that with laser displays, the focusing problem is somewhat reduced). A lateral displacement could also be used, to compensate for incorrect fit of the vision simulator and for image shift due to eye movements, but it is more practical to use a larger display and do this by shifting the image electronically.

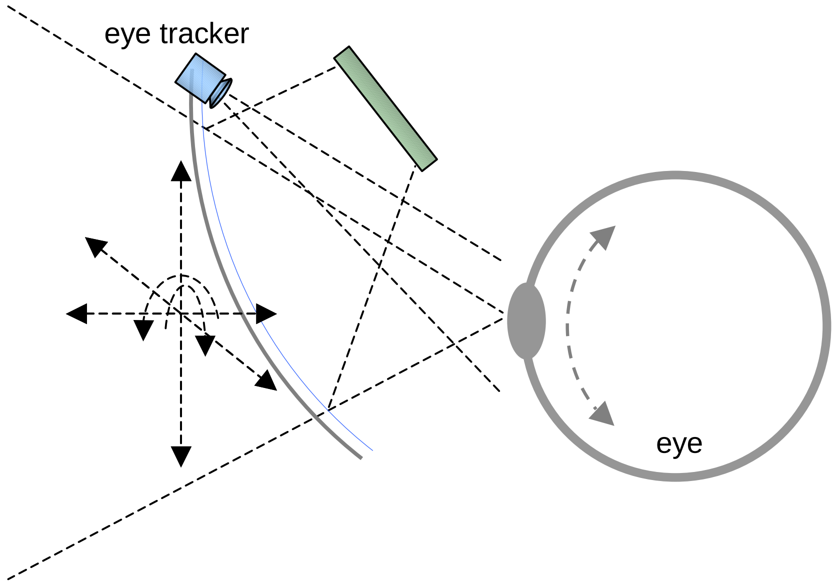
The mirror is semi transparent and does not shield outside view at all. If the display uses narrow band primary colors, a dichroic mirror coating for example could improve both transparency and reflection to almost 100%.

In order to cut out parts of the real scene for virtual image insertion, a mask display (light valve) is arranged outside the mirror. Preferably it is a curved assembly here, which may be the largest challenge with this approach.

This assembly looks and feels exactly like any pair of normal glasses except for the little add-ons at the upper edges of the lenses, containing display, eye tracker and perhaps also position sensor cameras. The front glasses can of course be laid out as sight correction lenses, i.e. also serve as traditional eyeglasses.

Eye tracking

It is absolutely vital for an acceptable visual interface, that it is not entirely fixed to the head. It should not be any more disturbing than regular (plastic) glasses. Hence, a certain freedom of movement of the device relative to the eye has to be accepted and dealt with. This can only be accomplished by sensing the device's position, for example with an eye tracker. The eye tracker is also necessary because pupil movements require complex image adaptation in order to avoid dynamically changing optical distortions.

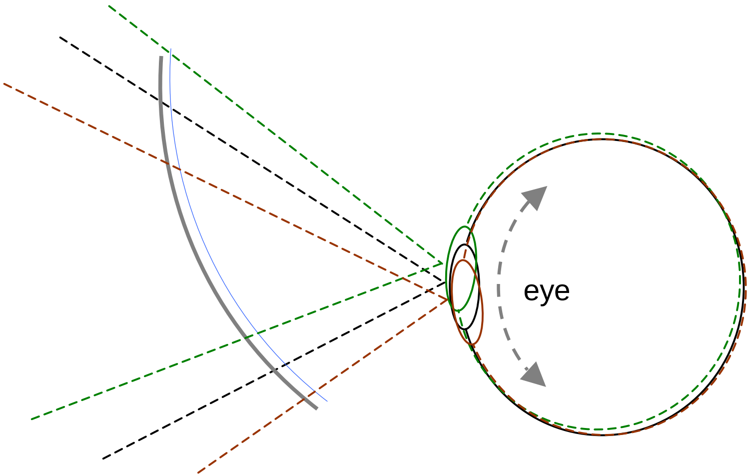


Example assembly with an eye tracker and usual degrees of freedom for conveniently wearable glasses

The picture illustrates possible movements of a vision simulator relative to the eye. The eye tracker, possibly some more sensors (measuring distance and tilt), and the image synthesizer software will have to compensate for all of this, dynamically.

The displayed image has to undergo resizing, shifting, tilting, trapezoid distortion etc, in order to constantly appear geometrically perfect to the user. A first work on this type of image compensation can be found in [5].

It is always important to remember that the eye is not steady like a camera but is in permanent motion, to center in on details.



Even by just looking around in the virtual scene, some very subtle perspective changes may occur, that could reveal the artificial nature of the picture and, more important, affect orientation in virtual space. Anything from an unreal impression up to headache or vertigo may result.

For virtual image parts at infinity this is not of major concern, as in this case lateral movements of eyes or display will not change object positions (this also applies to headup displays that we will treat later on). The problem remaining in these cases is that the pupil could leave the exit aperture of the optics. This aperture is smaller with small display chips (strong enlargement) and very small with laser scanners. It's the same effect we encounter when trying to find a comfortable looking position with binoculars.

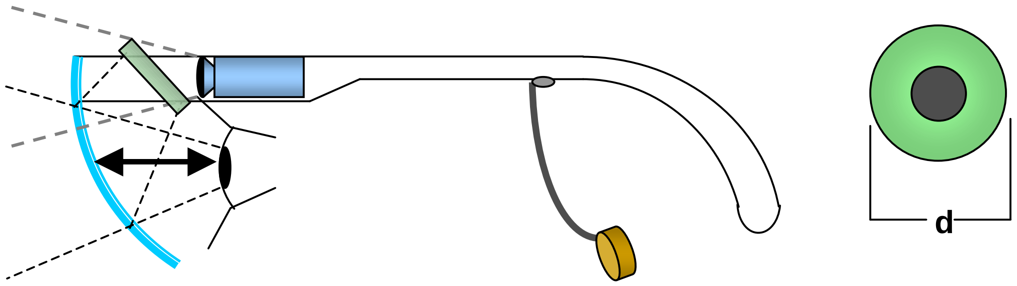
Eye trackers are not anything new [82],[86] but I will consider some design aspects in conjunction with the vision simulator.

An eye tracker basically consists of a micro camera and some image analysis software or hardware (locating the pupil and the iris in an image is just basic image processing (e.g. the Hough transform). The micro camera could use the main mirror as part of its lens, a camera integrated into the display, etc.

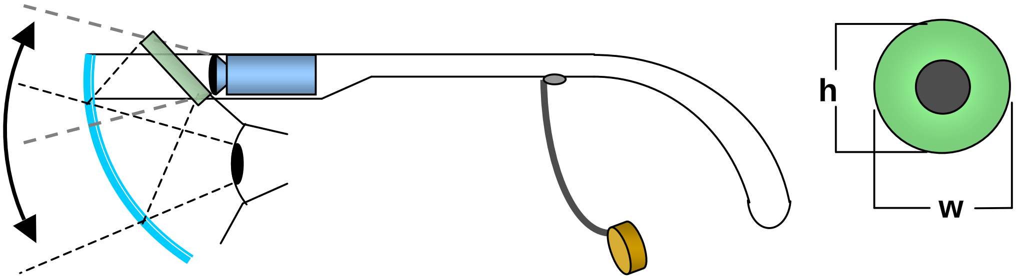
An eye tracker camera does not need to have high resolution, but a faster image frequency than the normal video rate of 2x25 or 2x30 half frames is necessary. 800Hz are already available [86].

One could also conceive a specialized light detector array with an integrated pupil recognition, that would be ultimately fast. It could even be made just slightly larger than the pupil image and be moved physically, actually *tracking* the pupil image this way.

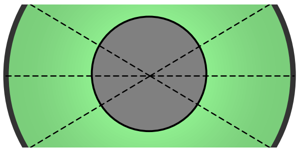
Sensing eye distance by Iris diameter



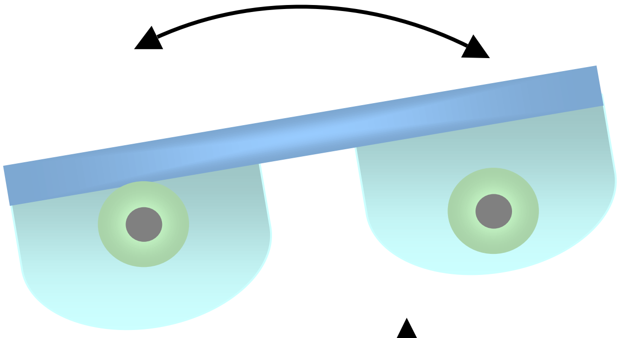
Detecting tilt by width/height comparison of the iris or pupil



Finding the center of partially occluded iris or pupil from visible edge arcs.

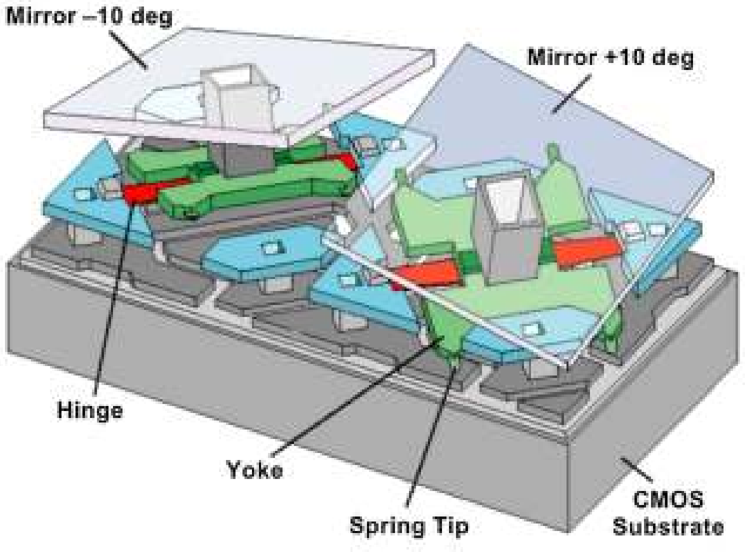


Lateral tilt detection can be achieved by comparing the vertical eye positions.

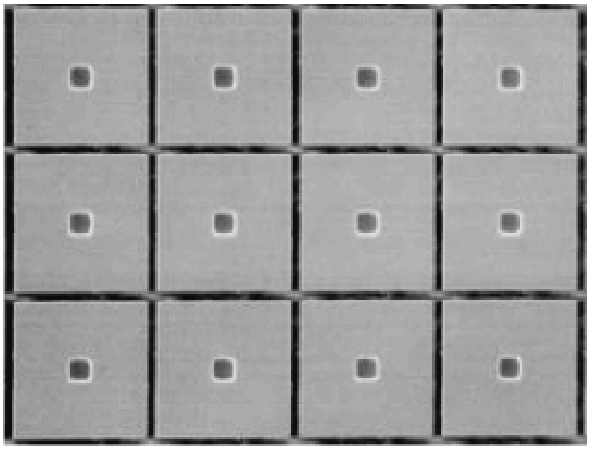


Here we exploit the fact that eyes usually don't move like this:

A DMD display design and eye tracker



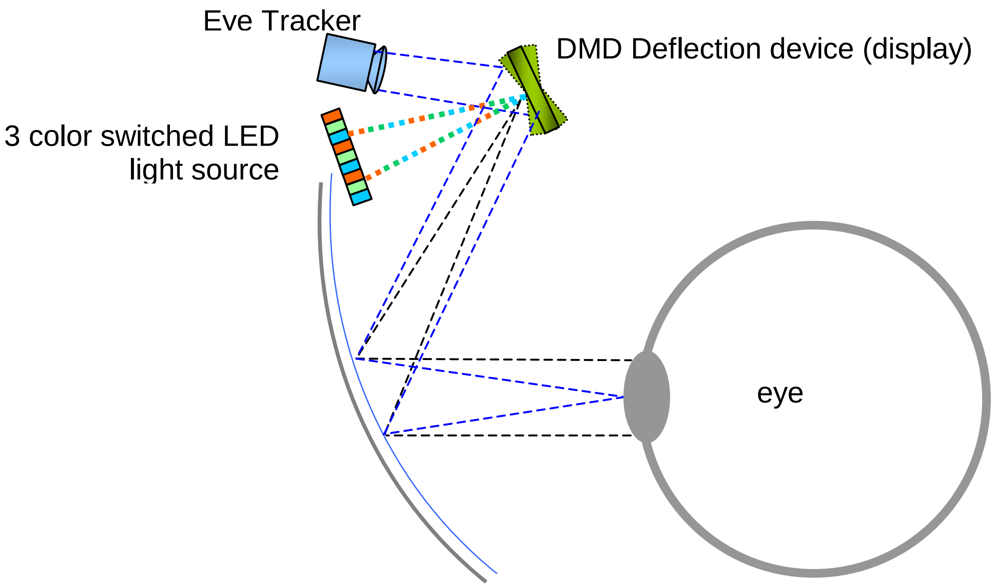
Two DMD mirrors



DMD mirror array

Digital Mirror Devices consist of tiny silicon structures, forming a CMOS driver array as well as elastically suspended Aluminum coated mirrors that are moved by electrostatic forces.

The entire structure is etched from a single, solid silicon chip. DMD mirrors can provide deflection frequencies up to 100 kHz and have shown to work reliably for at least 10^{12} cycles. This due to the hyperelasticity materials are showing at these tiny dimensions. One mirror is about 16x16 µm in size. The array is produced on silicon wafers with technologies similar to normal chip production. The mirrors are arranged in a very dense field. Almost 100% of infalling light can be reflected (images courtesy of Texas Instruments).



Principal assembly and focusing scheme for a DMD design

DMDs mirrors can tilt by ± 10^0, and extremely fast. Images are created by pulse width modulation of single mirrors. The light source is periodically reflected into or out of the projection optics. Color is produced by alternating light sources.

A possible advantage: as DMD elements also deflect light in 'off' position, we could place an eye tracker camera accordingly, so that it would get light from the eye just in this position. This would results in a good eye tracker image with little space required. Rays from the eye will also approach the eye tracker parallel, as required for a camera.

Such an optical assembly could allow to get a hardly distorted picture of the eye area, good for videophone applications.

The light source shown in this example consists of an interleaved array of LEDs in the 3 basic colors (red, green, blue). The array is not placed in the focus of the optics (the DMD is), so the colors

intermix. These LEDs can be fast switched and eliminate the need for color wheels as in conventional DMD projectors (there are LED-DMD projectors already in production, but they suffer from the low luminosity of the LEDs; this is not a concern here as we need but a very small image).
The light source arrangement has to take into account that DMDs are basically mirrors, i.e. the light is not strayed and lighting directions are important.

The drawing is far from being exact. We would have to take into account that current DMDs are tilting diagonally.

Hard to say if with our approach to optical design yet discussed, the viewing angle relative to the display would enable us to move the light source out of the optical path. It would certainly be better for this if the mirrors would indeed tilt up and down as sketched in the drawing.

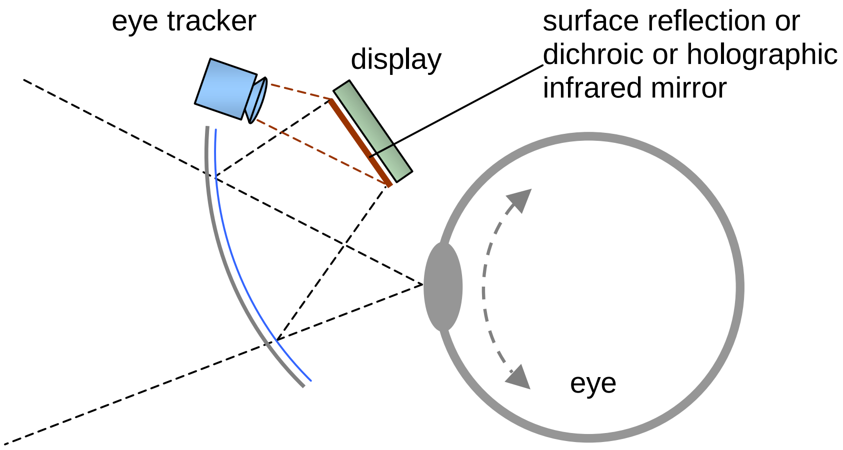
We would also have to find a curvature for the mirror glass that allows for a still more vertical assembly of the display than in the optical design study shown before.

If we could achieve this, the reflective behavior of the DMD would help us to concentrate most of the light directly to the eye, i.e. we would get very good energy efficiency and a very small leakage of display light to the outside.

Current DMDs have mirrors of 16x16 µm, 1000 pixels = 16mm. Given the display size in the optical assembly we have discussed in detail above – approx. 2.5 cm high – we would already arrive at 1600x2400 pixels for example, very good already.

Building smaller mirrors should also be easy, because DMD have been around for 10 years now and silicon process technologies have improved a lot. In principle, these smaller these mirrors are the better they get (speed, mechanical wear). So DMD is a technology that could actually give better results with smaller pixels.

Centered eye trackers for 'normal' displays



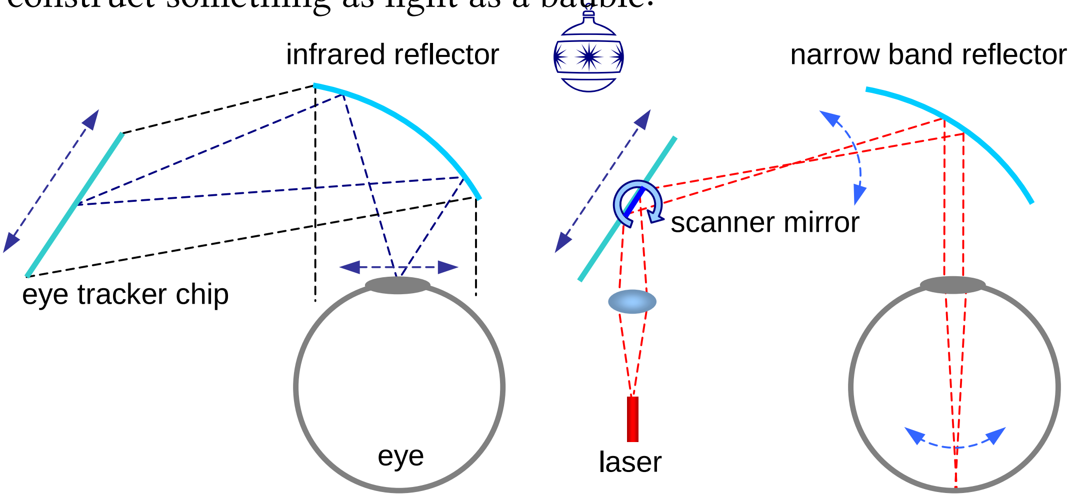
The DMD eye tracker approach may be adapted to other display types as well. As an eye tracker camera needs some light to see, we would probably use infrared light to avoid conflicts with other requirements. Using surface reflection or a dichroic mirror for infrared (one that would have to be transparent to visible light), just at the surface of a LED or similar display, we could have the eye tracker camera looking directly into the optical path.
Vice versa, we could also place the eye tracker first (with a different main mirror curvature) and use it as a reflector to pass the remaining light to additional optics and the display.

Through-the-display eye tracker

The best way for an eye tracker camera would always be to see by the main mirror anyway. We could try to etch or bore (or e.g. with OLEDs, just leave out) some tiny holes in between the display pixels and let a camera see through this mesh. The picture acquired this way would be a bit foggy (try to look through a piece of cloth), but it could work. Another option would be a 'fly's eye' construction, with little lenses and single camera pixels arranged in between the display pixels. As the camera pixels then have to be very small and we need a supporting matrix for readout, this would work best on a silicon chip, so one would want to combine it with a silicon based display type.

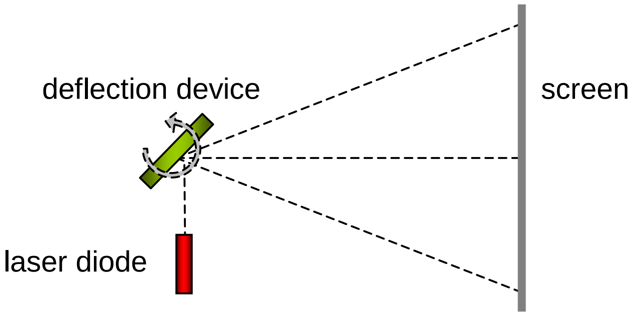
Optics for an eye operated cellphone

Cellphones may be the first mass market product implementing essential parts of augmented reality technology, especially eye tracking and eye steering, with simplified glasses, monocular maybe and without mask display. Imagine dialing and web link clicking just by gazing, hands-free. A large virtual display screen would be another great advantage, and it could be provided with very simple and light optics. The eye tracker system is meanwhile available as a single chip [86], cheap and tiny, so we can start to construct something as light as a bauble:



We combine an eye tracker (left, shown bigger than real) and a laser unit (right, more on lasers later on) in the same optical path. The tiny laser scanner mirror is integrated in the center of the eye tracker chip (hardly affecting its function). Just a single, small dichroic or holographic mirror (we will also address this later in the book), over 90% transparent, brings the picture into the eye. We will also need something to ensure that the laser beam will always enter the pupil in case of eye movements. For simple applications it could be sufficient or even better to see the display in only one direction, but we sometimes want to use eye pointing at large angles, this gets important. In this application, we may move the entire tracker/scanner chip, which would also allow to use a very small tracker camera chip, hardly bigger than the pupil. This would of course contribute to weight and power drain (in the gram or milliampere range, but anything counts here).

Laser displays



Basic laser projection assembly

An important type of displays for our purpose are laser scanners. Laser scanners use the narrow light beam of a laser diode, modulated for brightness changes and deflected in 'horizontal' and 'vertical' direction, writing an entire image onto a screen or directly into the eye.

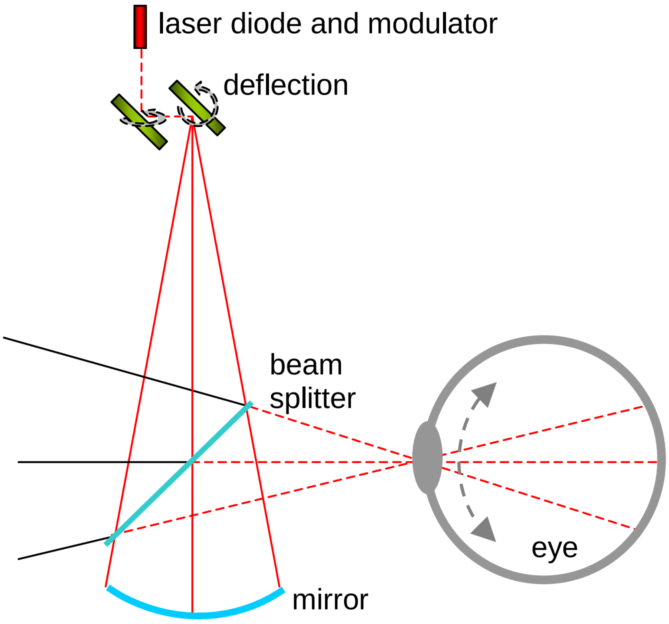
The advantages are simplicity of the light source, the ability to generate pictures with extreme brightness (almost 100% of the light produced can be concentrated into the eye), and some independence of focusing. Problems are how to modulate and how to deflect the beam, and also how to get a crisp image.

Usual laser applications use beam diameters of 0.5...1 mm. This appears quite large. Making it smaller will however cause problems with beam divergence. This is easily understood if we consider that light behaves like waves. With an exit surface smaller than the wavelength, we would quite obviously not even get a beam anymore but a spherical wave up to 180 degrees wide. Vice versa, we need to start from a certain beam diameter to be able to focus it to a very small spot. So we actually need an even larger beam if we want high resolution. Indeed for perfect resolution we have to deliver the pupil a 2mm wide beam, and this has to fit, as it is as large as the pupil itself. More about this later.

Making green and blue laser diodes has been difficult, but with the advent of HD-DVD and Blue Ray disk drives these items are even getting cheap. So the prerequisites for color laser scanners are now available.

A classical laser design

Let's have a look at an early laser design, that has been analyzed in [27]. It serves as a see-through data display. It is monochrome red at 640x480 resolution. Data are displayed in a constant position in the user's field of view. The brightness, typical for a laser display, allows its use in bright sunlight, e.g. in military applications.

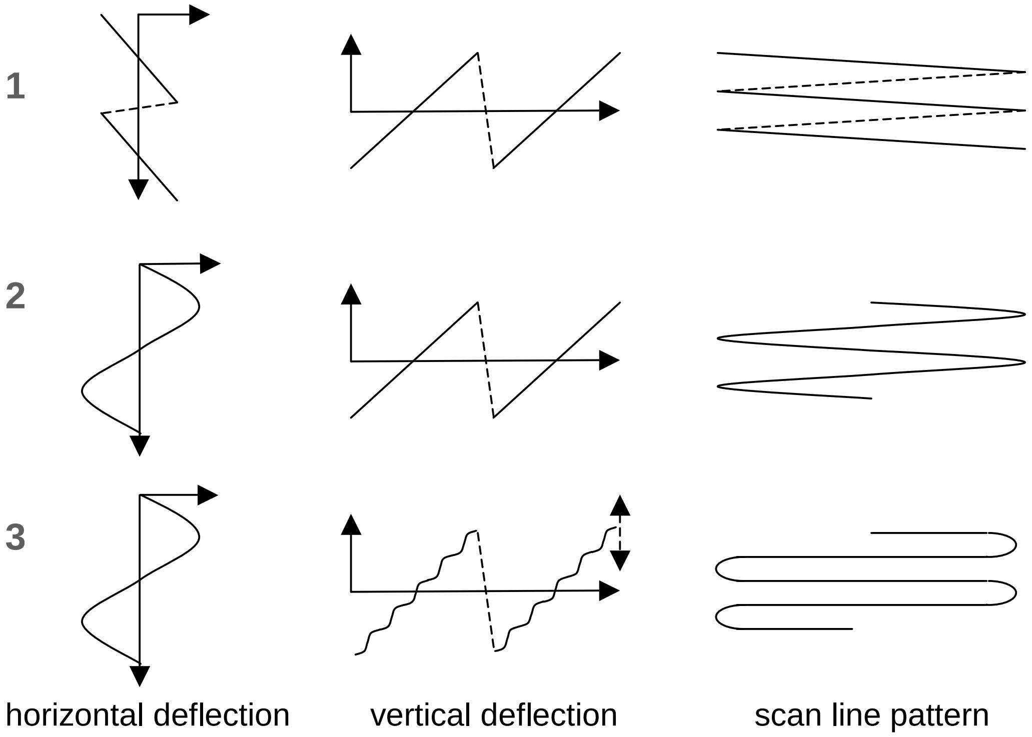


This design comprises a laser source, horizontal and vertical deflection mirrors in MEMS (micro electro-mechanical systems) technology, and a beam splitter/mirror assembly. This is a simple optical arrangement with hardly any geometry errors. Yet it's not the small convenient design that one would like for general use.

Later designs have the optics over the forehead and only an half transparent mirror in front of the eye. Information about current products and new developments can be found at [15].

This example illustrates a typical problem with laser displays: if the user moves his eyes, the laser beam may miss the pupil, especially when the latter is very small, adapted to bright sunlight for example. The design reviewed uses a special assembly that allows to split the laser beam resp. to displace it sideways in steps. If the pupil moves, at least one ray bundle is always there to enter the eye. For details of this tricky technique, see [27].

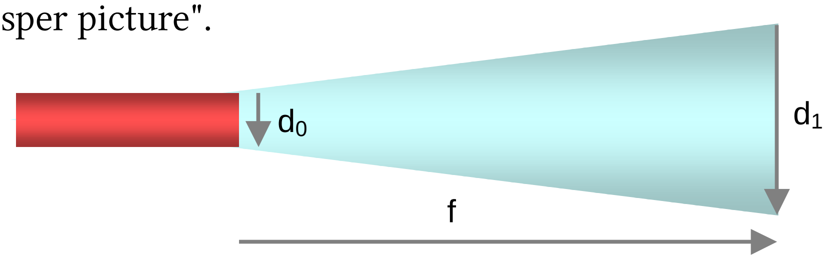
Beam deflection modes



Laser scanners draw pictures line by line, just as a classical TV picture tube. In a TV, a saw-tooth like voltage is used to deflect the electron beam from side to side with nearly constant velocity, then bringing it back even much faster, to draw the next line instantly (1). Moving a mirror this way would incur very violent accelerations even though such mirrors are very tiny devices. For the vertical (slow) deflection this could still work, but with the horizontal (fast) deflector we already operate at the edge of feasibility anyway. The smoothest way to move a mass forth and back is in a sine wave like motion. Yet this results in an uneven line and brightness distribution (2). Brightness could be compensated for, but we see line patterns at the image edges that would only disappear at very high resolutions. One method to address this problem would be to modulate the deflection mirror with a small high frequency sine wave also in the vertical, that would keep the lines about horizontal over most of the display area (3). The edges could then simply be dimmed out and we get a fairly even line pattern with smoothly moving mirrors.

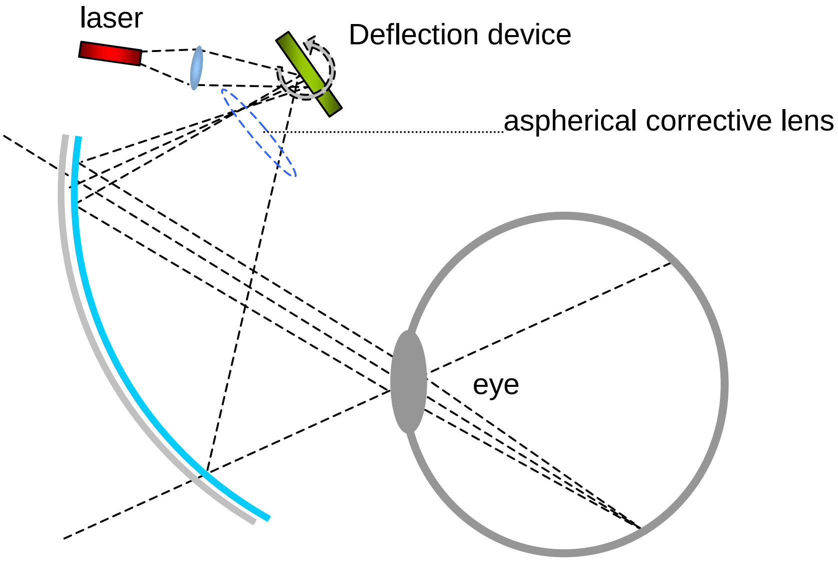
A laser design approach

Before we discuss a display design, let's first have a closer look at beam deviation. This is identical to the convergence needed to get to a certain focus diameter. The problem is that a smaller focus point just *requires* a larger convergence (=deviation). It's therefore not simply possible to say "make the beam narrower and we get a crisper picture".



Deviation of a laser beam with given source diameter d_0

We can apply textbook formulas for a so called Gaussian beam. For small angles ($x \approx \tan x$), these boil down to $d_1 \approx 1.27\, \lambda\, f / d_0$. ($\lambda$ is the wavelength). Later (p.261) I'll show a very simple way to get to this relation, as it is analog to the maximum crispness of a lens. Instead of 1.27 we see a factor of 1.22 there, but that's only resulting from the quite deliberately chosen amount of 'blur' allowed. So we may simply take $d_1 \approx \lambda\, f / d_0$ as a rule of thumb.
Example: $\lambda = 0.5\ \mu m$, $d_0 = 20\ \mu m$, $f = 20\ mm \longrightarrow d_1 \approx 0.5 mm$. Hence for a pixel size of 20 µm on the retina we have to deliver a laser beam 0.5mm wide and we would also need a deflection mirror of 0.5mm diameter, if we don't want to degrade beam quality. With a picture area on the retina of about 25x20 mm and this pixel size, we could deliver an image of 1280x1024 pixels, the same as a typical 19" computer screen.
Let's now have a look at an approach starting from the basic design for classical displays as considered before.
If the laser source would emit an ideal, very narrow and coherent beam, the eye's lens would influence geometry but not focus (the narrower the beam, the less), making the design much simpler. We could let the image originate at a single point. Distortion could be compensated by just modulating the deflection unit. Yet it isn't all that simple.



Basic adaptation of a laser unit

In the drawing, the pupil and the deflector are positioned at about 1.5…2x the focus length of the mirror. The pre-shaping of the laser beam is only hinted (we could also for instance conceive a concave deflector mirror). For a crisp spot at the retina with the eye focused to infinity, the beam would have to arrive parallel, so in turn it should converge at about focus length before approaching the deflector, where it would perhaps again have to be as wide as at the pupil. We could try to bring the deflector closer to the main mirror, in order to catch the laser beam at a smaller cross section, but this would also mean the deflector would have to deliver a larger angular range. Deflecting a laser beam at a narrow spot also raises the question if the resulting wavefront is still well defined.

A mirror as large as the beam diameter at the pupil would at least be 0.25mm even for VGA resolution (640x480), and almost 1mm for some good resolution (≥1600x1200), sufficient to generate virtual computer or HDTV screens with some reserve for surroundings and perspective. At this resolution, we need ≈100 kHz deflection frequency for an acceptable frame rate. Moving a mirror that fast is really difficult, to say the least.

Compensating focus errors of the main mirror could be easier than with the conventional displays: with a small corrective lens (picture), whose focal length changes over its surface.

The laser beam, although not entirely parallel, is usually less sensitive to focus changes than a classical assembly, as the effective pupil opening is no larger than the beam diameter. With low resolution displays, this could e.g. mean 0.4mm vs. 3mm, giving 8 times the focus tolerance of a classical display. At the highest useful resolution of 1/60 degree however, the beam diameter has to be 1.5-2mm, so in bright light the advantage then becomes zero.

Let's consider some more properties of laser projectors, or scanners as one could also call them.

Good news first: Laser scanners can be very small. Laser diodes are tiny, so the entire laser unit, including deflection, could almost be shrunk to the size of a sand grain, i.e. it would be the smallest and lightest display one could think of. There are some problems as well. Let's have a look at this.

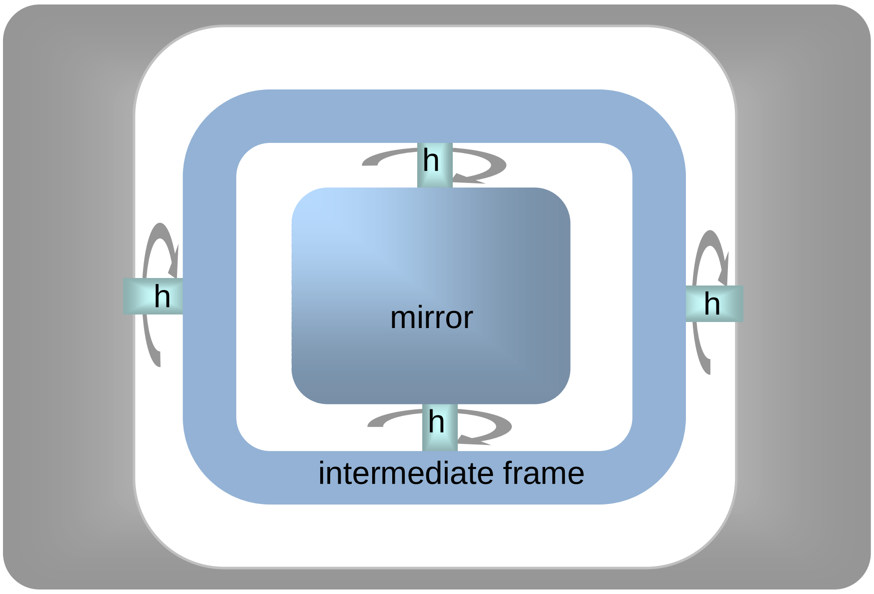
Laser diodes are very non linear and do only work above a certain power threshold. It's therefore not possible to simply modulate them by changes of supply current. They can however be operated with very short pulses, so a digital pulse width or frequency modulation may be possible.

The second challenge is with deflection. We first have to consider that measures have to be taken for the protection of the eye. The beam should never be allowed to rest, as it would burn holes into the retina if all its (minor but extremely concentrated) power remains on a single spot.

The deflection unit itself as stated is really difficult if we want 100 kHz or more for horizontal deflection. Achieving these frequencies is in principle possible with DMD mirrors. They have a resonance frequency of about 100 kHz and can endure this speed for years. Their size however is only 13 µm, way too small.

Another problem is that DMD mirrors are currently designed to operate bistable, not freely oscillating, and that they have only 20^0 total deflection range while we would like to have 40^0 or better. Anyway, we only need a single mirror and don't have to use the extremely complicated manufacturing process of DMD. We only need to know that hyperelasticity and small size allow us to create a super fast and reliable MEMS element from silicon or other materials by etching, laser cutting, or other methods.

In order to achieve large deflection angles, we have to fully decouple both axes with an intermediate frame and independent hinges (h) for horizontal and vertical guidance (picture).



Such a structure can be used with electrostatic or electromagnetic drive. Unwanted resonances and parametric effects have to be addressed by proper driver design and damping. Very high frequencies and very large deflection angles will be possible with this approach, but the mirror size required will be a problem.

Current projects can be found at [8] or [15], but anything published so far does not concurrently deliver the resolution and the deflection angles we'd require.

Yet something could be done with laser scanners that is much more difficult with other displays: producing a high resolution, eye movement steered *picture inlay* at the center of view.

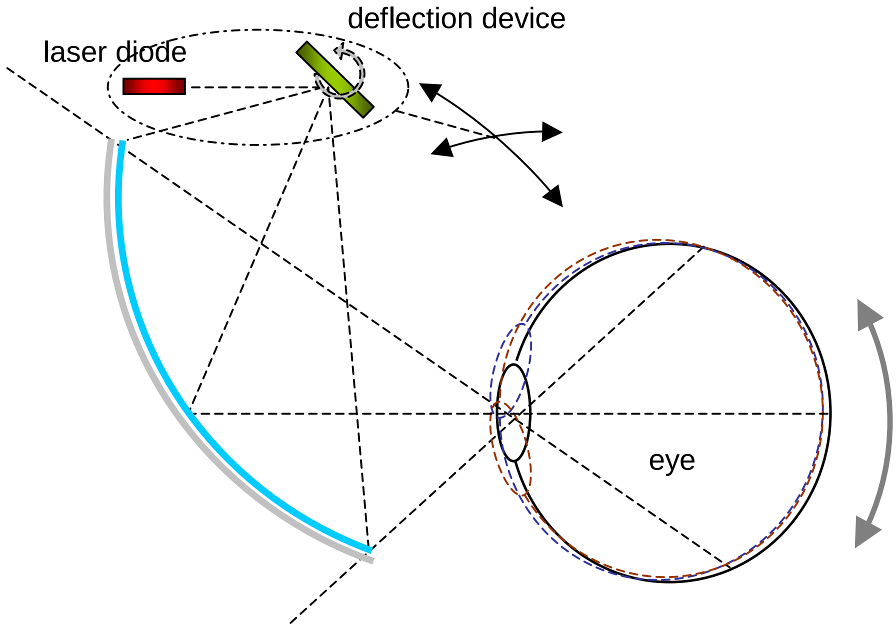
With small MEMS deflectors, we could perhaps place two of them side by side so near that all rays still enter the pupil. One deflector would act the usual way, the other would perform a quasi static offset deflection overlaid with a raster deflection of ¼ or even less the angle of the other. Aligning both images precisely is a bit difficult, but not so much if we dim the inlay area in the large picture. This allows to deliver a center resolution of one arc minute even if the total display area is 60^0 or greater, hence a perfectly crisp picture over almost the entire human field of view.

We could even use three deflectors and provide an inlay inside the inlay. With inlay sizes of $^1/_3$ and $^1/_9$, each picture would need to have no more than VGA size (640x480), to get the perceived image resolution up to a whopping 5760x4320.

The same could be achieved with just one MEMS element, by sequentially generating the main and the inlay pictures. It would need to be 3 times as fast, but for 3 VGA pictures this would still require less speed and agility than a single XGA picture. Alignment would be easier than with 3 elements, although we may have to care for errors due to the inertia of the mirror.

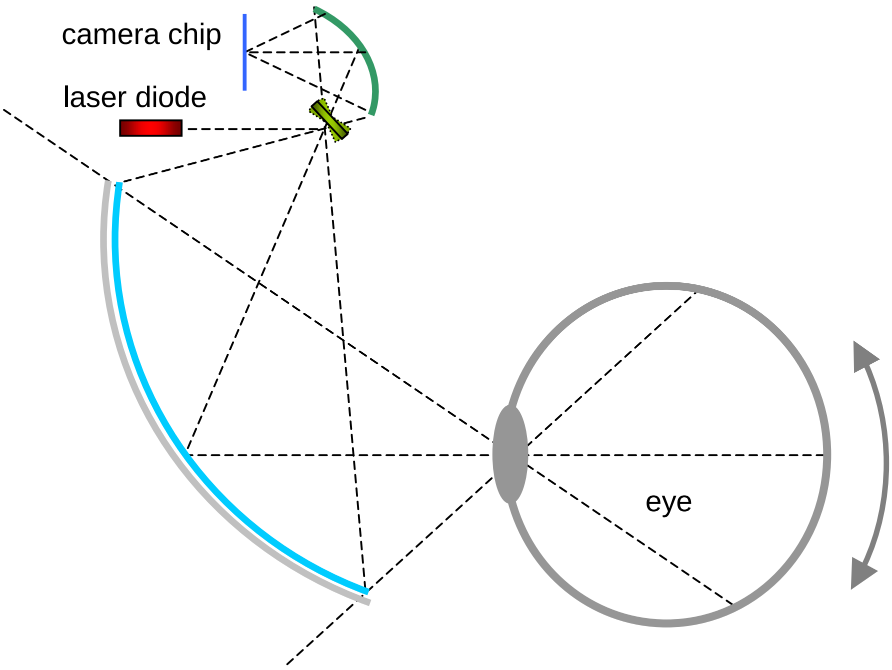
Alas, we will still have difficulties to handle a beam wide enough (up to 2mm) for the small focus spots we need with these resolutions. An holographic display acting as a mirror could perhaps solve the problem. I'll turn to this in the next section.

Back to our design attempt: as we will anyway need an eye tracker in order to show virtual objects correctly, we could compensate for eye movements by just moving the projection unit with the pupil (also suggested in [27]). A curved movement scheme would be of advantage, as it would help to keep deflection angles in the best range. As an alternative to the unabating movements we might consider an extremely widened laser beam, still parallel but wide enough to always hit the pupil.



Shifting the projector to compensate for eye movements

Eye tracker and laser display



The main mirror projects a crisp image area representing the entire eye just where the laser deflector is. In this virtual image area, the deflection mirror itself covers up nothing but the central part of the pupil, so it does not hide essential parts of the image.

Placing an eye tracker camera chip there has some difficulty, as the laser deflector has to move around, tracking the pupil. With a small eye tracker chip just covering the iris area and also moving around, this may be nicely integrated into a single unit though,

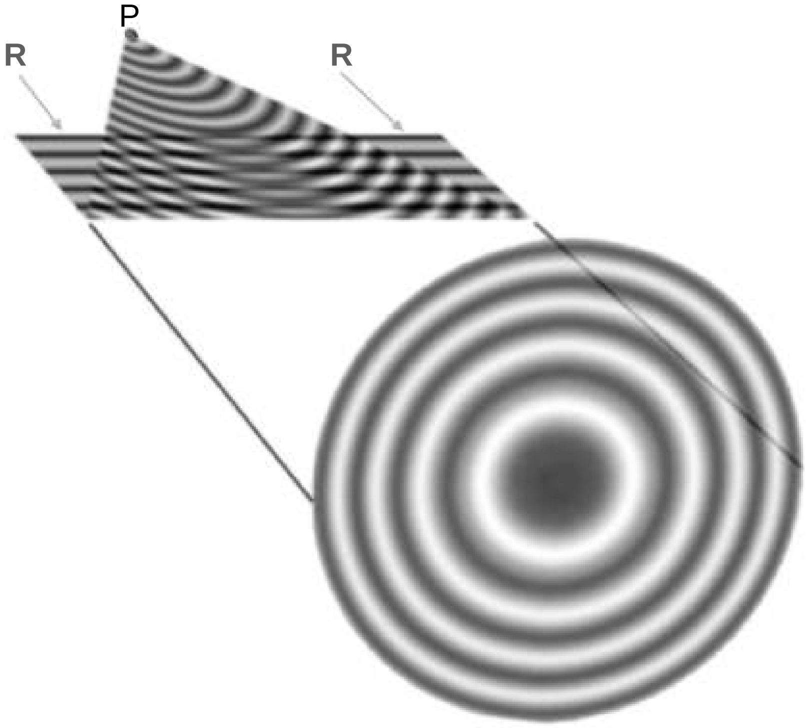
If we want to cover the entire eye area (as we have seen to be of advantage for some applications) we would need a pretty large eye tracker chip.

A second concave mirror behind the deflector, could produce a second focus plane, where we could then place the camera chip. We could also use a flat mirror and a lens, for the same purpose.

As this example here is drawn, the camera chip could be about 4 times smaller than the image area to be covered by the eye tracker. The second mirror could be laid out for ¼" or even smaller chips. With only the tiny beam deflector in the light path, at the location of a large intermediate image, the eye tracker picture would not be significantly occluded.

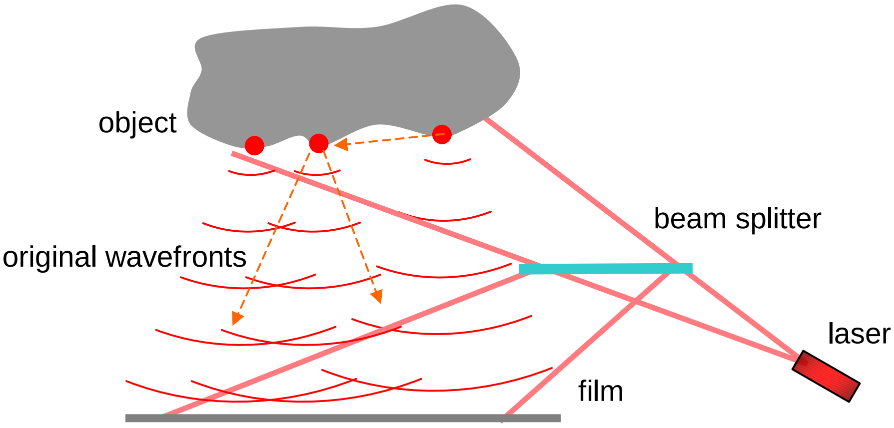
Holographic displays

Holography works with interference patterns. These occur when coherent waves (e.g. laser light) from different origins add or annihilate according to their phase and position. Recording a hologram is done by letting a direct laser ('reference') wavefront interfere with one reflected by objects, and 'freezing' the resulting pattern on a photographic film. If later on a single laser wavefront equal to the reference passes the photographed pattern, interference between the passed parts of the wavefront results in identical wavefronts as were previously emitted by the recorded objects. The principle works with reflective and engraved patterns as well.

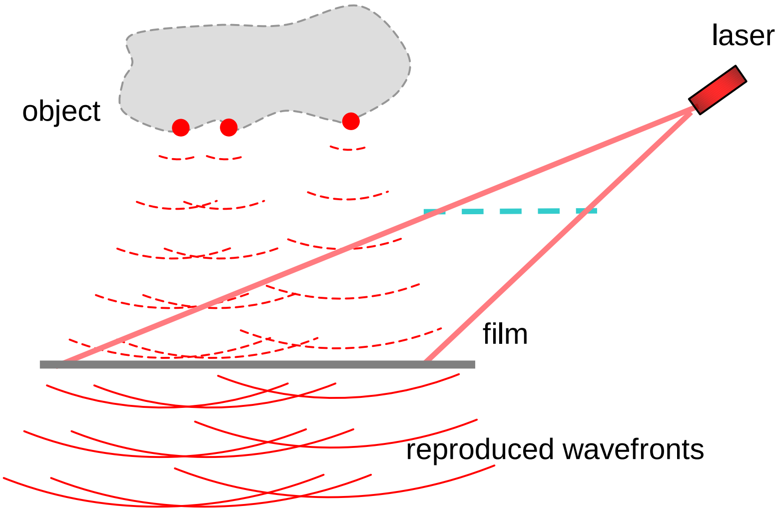


In order to illustrate this, we'll regard the very simplest of all holographic setups. Consider a point source of (coherent) light, e.g. one pixel P of a scene illuminated with a laser. In addition, we'll consider a totally planar wavefront R derived from the same laser source (i.e. entirely synchronous).
If we place a receptor (e.g. a photographic plate) in front of this, we get an interference pattern of concentric circles. (these are not equidistant; their distances are a function of the distance from the scene pixel to the receptor). If we now place the developed photo

plate in front of a planar wavefront as with the recording assembly, the concentric wavefront from the image pixel will also be synthesized, because the rings themselves act as wave origins and their addition just does the trick. Such a basic pattern of concentric rings is also called a zone plate and works like a lens.

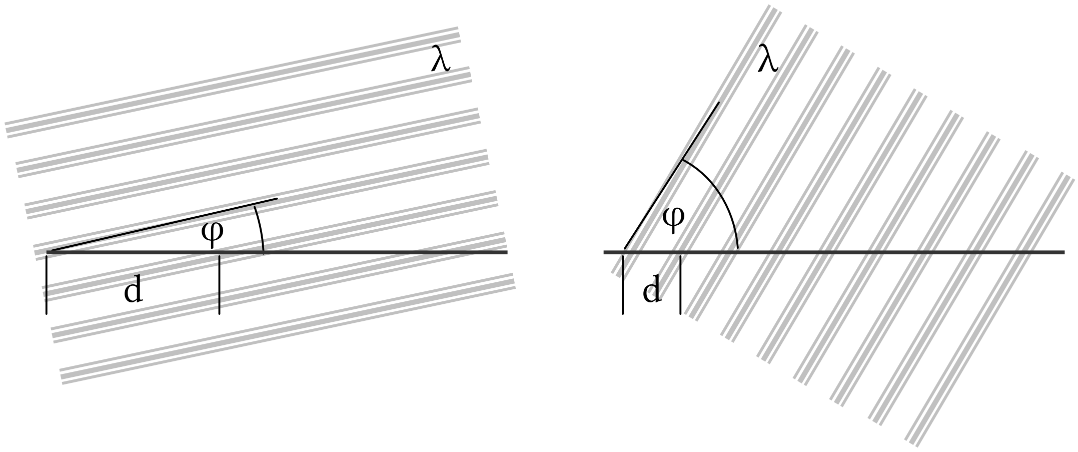


Simplified hologram photography assembly: a single laser delivers a reference beam to the film and illuminates an object. The wavefronts from three points on the object are shown as an example. Wavefronts from all points of the object interfere with the reference beam and form the hologram on the film. Note that self interference (dashed lines) causes noise with natural hologram recording.



If later on another laser of the same wavelength and position reproduces the reference beam and illuminates the holographic pattern on the developed film, the hologram pattern converts it into the same wavefronts as if the object points were still there. A front illuminated reflective hologram would do the same trick.

Superimposing many patterns, allows to reproduce many pixels simultaneously, together with their proper distance, hence the entirety of light waves from the original scene. We see that holography in essence is a very simple thing. It's quite natural however, that any tiny part of the pattern doesn't know if to deflect the beam left or right, so we get two exit wavefronts, one producing a 'conjugated' virtual image before the plane. Therefore, the actual recording assembly shown uses slanted beams. If so desired, a hologram appearing before the plane could still be generated and used as the main picture. There are several more tricks to apply with physical hologram recording, that can be found in specialized literature.



Interference patterns as on a photographic film can also be simulated with computers and reproduced with a display. We should note that with physical hologram recording, we can't exclude the above mentioned self interference. With synthetic holograms, this and other adverse effects can easily be avoided by proper pattern generation.

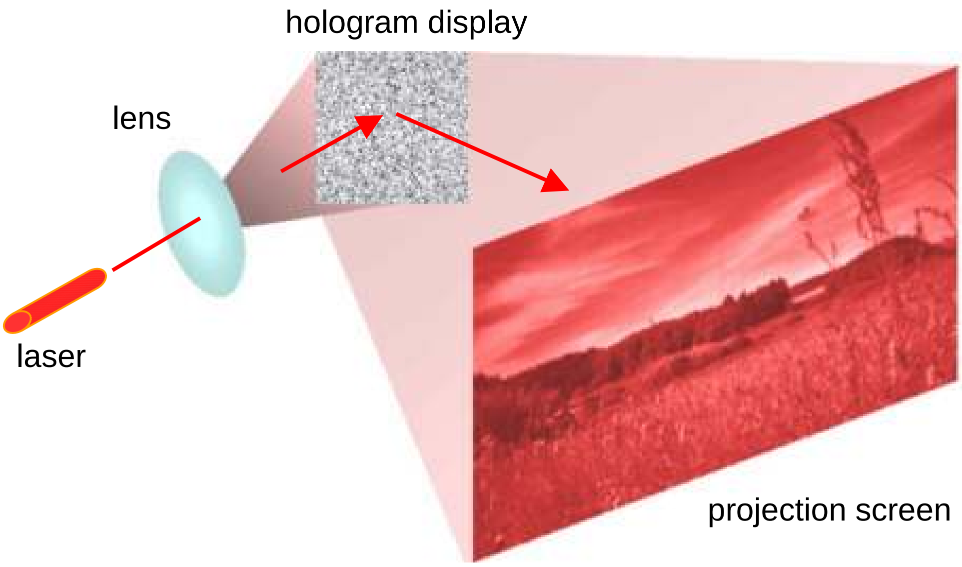
The resolution necessary depends on the angle of the infalling light: the larger the viewing angle we want to reproduce, the better the resolution required (picture). The gray values of the pattern constitute sine waves (wavelength=λ). According to the sampling theorem [34], these can already be reproduced with at least 2 pixels per wave, i.e. we need pixel sizes of about d/2, or a little less, because pixels are squares and resolution therefore is lower in diagonal direction. In any case, it is absolutely important to reproduce exact gray values.

If the display is illuminated almost perpendicular, φ is just the angle between illumination and resulting waves, so the pattern resolution just depends on the degree of deflection, or the viewing angle required.

For our reference beam, we are not confined to a planar wavefront of course. This was only chosen for an example. In practice, the reference 'beam' would be a concentric wavefront, originating from one point.

As wavelengths of visible light can be as low as 0.4 µm (blue light), we will need pixel sizes between about 0.2 and 20 µm, depending on the illumination and deflection angles desired.

For big screens, this would normally not require megapixels of resolution - as with classical displays - but *terapixels (trillions)*. This simply forfeits any attempt to implement it with current technology. With very small displays however, it's not quite so difficult. There even are first applications of such a technology.



Quite recently, a holographic projection display has been announced by Light Blue Optics (picture,[31]), based on a very high resolution reflective F-LCOS (Ferroelectric Liquid Crystal on Silicon) display and a special processor chip able to generate the interference patterns in real time.

This system is made to generate 2D projection pictures on a remote screen, which results in simpler hologram calculations than with full 3D. For this approach, an F-LCOS display with 13 µm pixel size (CRLO Displays Ltd, [33]), together with some optics, is already adequate.

We could expect this to work for a deflection angle of about 10 degrees total, already enough for a projector if we add some optics. F-LCOS pixels could be made even smaller, 4µm said to be possible. The structural size of the driver patterns itself is not the limit, as silicon structures can meanwhile be manufactured way below 0.1 µm.

There already is an LCOS display with 1920x1200 pixels and a pixel size of 8 µm (under the magnifier, Aurora Systems Co.,Ltd. [84], picture courtesy of Holoeye Photonics AG [51]).

If we get further down with pixel size (below 1µm), holographic displays could also be a solution for vision simulators. Alas, even with F-LCOS it's not so easy to accomplish. Liquid crystals need some thickness to provide a 90^0 polarization turn.

Other digital holography approaches are using DMD devices [36]. It's quite conceivable that DMD pixels could be made even smaller than F-LCOS. DMD are getting the faster the smaller they are. Sufficiently small structures don't even have to be exact mirrors, because a sub wavelength mirror will reflect anything as a spherical wavefront anyway. Generating a microscopic engravement pattern, as with GLV displays, would be enough. So this should actually not be too difficult.

For a hologram without artifacts though, we need to produce patterns with gray levels, and pulse width modulation won't work here (interference of photons emitted at different times can't realistically be expected).

What could be working are structures way below light wavelength, generating gray levels by sort of pixel dithering. With silicon chip structures of 60 nm ($^1/_{10}$ of red light wavelength) in mass manufacturing already, this is not unrealistic at all.

It would anyway make sense to integrate signal processing structures of significantly higher density right into the chip, in order to generate synthetic holograms in real time, a technology realistic within about 10..15 years (more in the media chapter).

Let's explain why holography could work for our purpose at all. With a 'normal' hologram, thousands of different viewing angles are contained in the wavefront. The larger the display area, the more there are. Hence, a large hologram contains a lot of partly redundant information. If we break a holographic plate or cut out a little piece, it still shows the entire picture, but from one direction only of course, like through a keyhole. This is like a 2D image, except that it still may have different focusing for different objects, hence a little remnant of 3D.

If we get even smaller, it works like a hole camera and also focus is the same over the entire image, i.e. then it's really 2D. Somewhere within this range, the mentioned projection display resides. As it only has to deliver 2D, it can be kept small and manageable. Clipping out still smaller pieces, the image would get more and more blurred as there won't remain enough lens rings to sufficiently define a pixel.

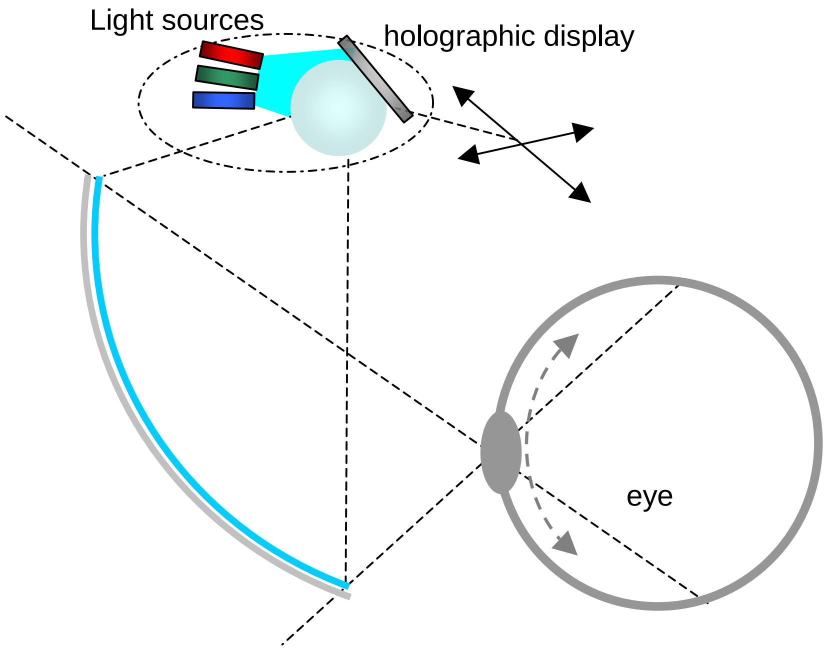
From the standpoint of information theory, the issue looks as follows: we deliver any information the little hologram has, just to the eye. No other perspectives etc. have to be generated. Hence, the number of pixels theoretically wouldn't need to be much larger than with a conventional display. As pattern line sizes are always dictated by wavelength, the simple trick to really achieve this is to use a very small display. Its image could then be enlarged by optics, e.g. we could look at it with magnifying glasses, project it to a screen or the like. We only have to accordingly generate the displayed objects, to appear in their real size after the magnification.

Let us now consider such a display in a vision simulator. Our main mirror, maybe together with a second one, could be designed for a lot of enlargement. We also have a requirement for 2D essentially, or only a little more: each picture is for one eye only, and the eye can only look at the display from more or less one perspective. The hologram has to be illuminated, as with the classical DMD version, except that we would use laser light.

With the vision simulator optics reviewed so far, we have seen that it is very difficult to get large magnification together with overall good focus and low distortion. Yet with a display chip showing a hologram, we have the option to add some 3D properties. In other words, we could assign any displayed pixel any virtual distance. It behaves a bit as if we were simulating lenses. Lenses recorded in a hologram look and behave like real lenses, and with a synthetic hologram, we could simulate not only lenses but fairly complex and dynamic optical behavior. So the image pixels just don't have to stick to the display plane. In a hologram they usually don't anyway. We could also do this according to where the user just looks at, i.e. to the virtual distance desired for each object.

The approach works with a lot of 3D information, hence requires more computing effort than the 2D projector mentioned. With a useful (large) deflection angle, less than 1µm pixel size would be desirable. With 10 times the pixel count of current displays, 3000x4000 pixels for example, the display would be 3x4mm. This is some way to go, but not impossible.

As we are looking at the display hologram through a magnifier, with inevitably extremely convergent light paths and a very distorted actual display perspective necessary in order to simulate a natural, wide field of view extending to infinity, it may actually be of advantage to have the image pixels appear before the display plane. The proper configuration has to be determined according to several factors. For example, it should ensure that every image pixel is represented by a pattern area on the display as large as possible, as this would result in better resolution.



Raw scheme with synthetic holography display (imaginary pixel positions visualized by sphere)

In the schematic drawing (picture), reflection angles are quite extreme, so some additional elements could be required for a proper design. The drawing can only be an illustration. One advantage we get, even dynamic focusing could be done just with the pattern computer. No moving parts for this any more.

Yet the display is very small. We therefore probably need a mechanical eye movement compensation, as with the laser display (if Moore's law still holds for some more years, we will later on be able to build the chips large enough to do without).

A holographic display could also correct for focus and geometry errors of our main mirror in a very comprehensive way, and even simulate complex and dynamically adapting optics. It could as well correct for eye defects. Many people have cylindrical eye distortions that can and will be corrected for direct sight, with corrective glasses that become part of the vision simulator.

For the virtual objects, it is normally difficult to include such a correction. Holographic displays could offer the flexibility to do this in software. Even with gliding correction lenses, it would likely be possible to perfectly adapt the virtual to the real sight.

Color, of course necessary, could in principle be done in one hologram. We would not need to join beams of separate color laser sources before the hologram, as all color parts of it can be computed for different origins. In the contrary, it would be better if the 3 color sources would be as far apart as possible.

The best way would be to switch between colors sequentially, even though this requires a 3 times higher pixel rate. With a possible micro mechanical display with pixel elements much smaller than a micrometer, switching them many million times per second wouldn't be a problem, and laser diodes offer switching speeds in the sub nanosecond range anyway. So color could be generated sequentially, and even pixel by pixel if we want it.

We see that synthetic holography displays could be a magnificent option, and the vision simulator could be a very rewarding application of this technology. The immense computing power necessary to generate synthetic holograms will be an obstacle for about the next 10 years (more in the media chapter). The first full fledged vision simulators will most likely use laser technology.

I won't just delve into more possible constructive details and variants like geometry, illumination etc., as this technology is still somewhat speculative and the issue would get very complex.

Let me only mention that holograms are good for a lot of other tricks. For example, we might even be able to overlay or switch holograms to let the display also serve as optics for an eye tracker camera (something very important that we have to keep in mind with all designs), similar to our DMD design example.

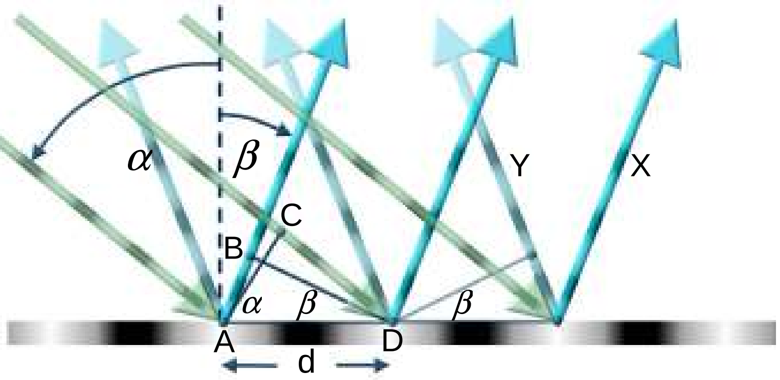
Generating hologram patterns for real 3D pictures is a lot more demanding than with the aforementioned projector application. In the media chapter I'll address this again and show that it should finally be achievable at least for the vision simulator display. We should also keep in mind that static holograms, translucent or reflective, could serve as versatile optical elements with a laser based display or projector (see 'holographic mirrors', below).

Holographic scanners

In a sense, what I have outlined here is really somewhat isomorphic to the laser scanner approach: imagine we would just use the hologram simply to *deflect beams*. It could then be much smaller and would have to change images much faster, but in the extreme, it could work exactly like the moving mirror reflector, yet with a larger surface and at almost any speed.

Feeding an entire hologram into the display, for any tiny change in deflection angle, is of course not possible. The according holograms should already be stored or calculated within the chip.

A plane mirror pattern would resemble the zone plate we already discussed, yet with only straight and equidistant lines. Just imagine a planar wavefront approaching a surface with very narrow reflecting lines. Any of these lines will become a source of a cylinder wave, and all these cylinder waves have different phases according to the angle of incoming light and pattern line distance. They will add up to another planar wave of a certain angle, very different from the reflection angle of a simple mirror surface. As this actually describes the behavior of any pattern detail of any hologram, it should be interesting enough to examine it in more detail (this is similar to the behavior of a diffraction 'grating'):



With a reflective hologram as depicted above, it can easily be seen that 'constructive interference' of waves from the different wave origins (the bright parts of the pattern) occurs when

$$\underline{CD} - \underline{AB} = n\,\lambda$$

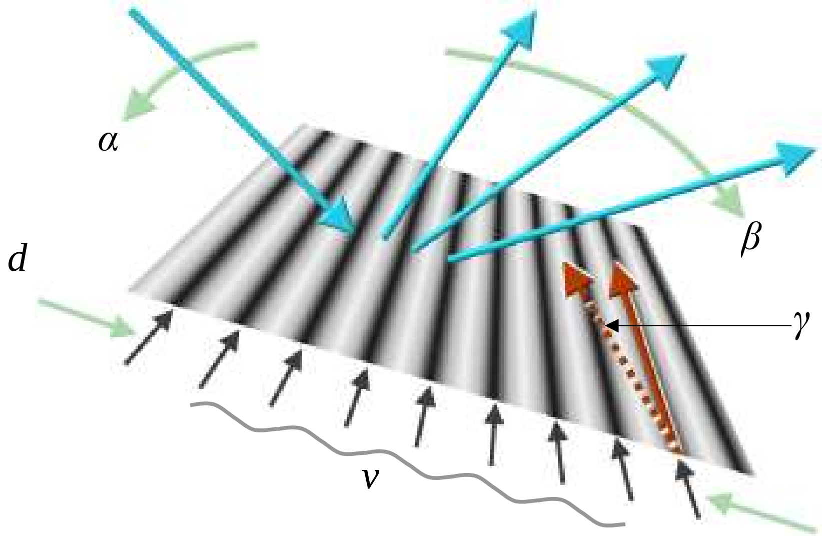
(a multiple n of the wavelength λ), but also when

$$\underline{CD} + \underline{AB} = n\,\lambda^*$$

(the underlinings denote distances). So actually two 'conjugated' departing wavefronts X and Y are generated.

---

* Hence, $n\,\lambda = \underline{CD} \pm \underline{AB} = d \sin\alpha \pm d \sin\beta$, (the 'grating equation')
$\beta = \pm \arcsin(n\,\lambda / d - \sin\alpha)$ where n=1,2,3,... (only n=1 is of interest)



So if we change the line distances *d*, we get different deflection angles *β*. If we actually calculate this, we see that an angle of 50^0 or 60^0 for the incoming beam for example would theoretically allow to alter the angle of the departing beam from about 10^0 to $>80^0$, without any unwanted modes ($n>1$) showing up. The conjugated beam can simply be absorbed by a black blind. Limitations are efficiency, the inferior definition for small angles due to large d values, and the smaller apparent mirror area for large angles (that can be partly compensated by a larger mirror area).

The straight line pattern could be generated by providing a sine pattern (for gray values) in just one dimension and feed the values *v* into the chip from the side, steering all pixel elements in a line at once. If we want to generate gray values using binary sub wavelength pixels, a dithering mechanism could perhaps be built into the chip. This is just one approach possible, others are likely to be found, yet it should show that it's really achievable.

By changing the line angle *γ* as well, we could deflect in two dimensions at once. Generating such slanted line patterns is pretty complicated, as the lines have to appear in very precise angles. A slanted propagation *γ* of the steering signals *v*, might be achieved by variably cross switchable steering lines, but this is a pretty coarse technique.

As the chip will essentially be 1...2mm small, we could also just tilt or rotate it for the slower deflection axis. There we won't need more than about 100 Hz, and this is not too difficult.

Several other operation modes are possible, for holographic elements. A small hologram could produce parts of the viewing area sequentially, serving each with its own partial image, staggered inlays for example with resolution progressing towards the center of view, as we have yet discussed with laser displays.

The deflection chip could even write a picture with different resolution parts in just one sweep, switching line frequencies even within lines, with no alignment errors. It could also address the edges more often, where the eye is more flicker sensitive.

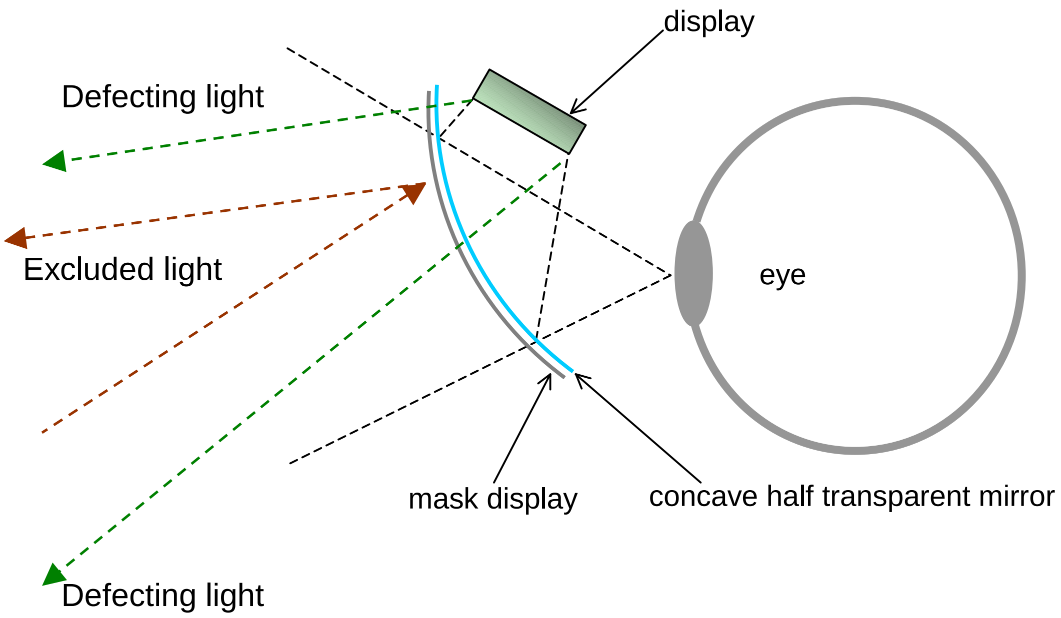
Holographic modulators and laser scanners essentially are but extreme varieties of a general wave function modulator for laser beams. No wonder that the development of high resolution displays with extremely small pixel sizes was not actually stimulated by display glasses but by laboratory applications like laser beam shaping, and that they're usually known as 'spatial light modulators'. What needs to be developed for the future, is a display chip technology for *extremely* small pixel sizes, maybe a micro mechanical one.

In conclusion, we have seen that conventional displays are already available in several varieties and at fairly acceptable resolutions. Yet building an optical assembly with a large magnification factor and viewing angle to allow for a very small, light display is quite difficult.

Due to a certain focus invariance, laser displays allow for very light constructions, yet they need a correction unit to keep the laser beam at the pupil with eye movements, and there is the problem that higher resolution (smaller pixels) needs a wider source beam and a faster deflection unit at the same time. Nevertheless, we may expect fairly good laser units 'real soon now', maybe tandem or multi mode units producing high resolution images with dynamic inlays, maybe with an holographic display for beam deflection.

Fully holographic displays could in theory combine all advantages, and it wouldn't be a big problem to build appropriate display chips, but holograms are so computing intensive that it may take 10..20 years until they could actually succeed.

'Half transparent' mirrors



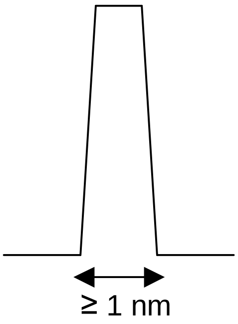
Let's now have a closer look on the 'half transparent' mirror that we need for the glasses, to combine the real with the virtual.
Just using thin silver coatings that allow half of the light to pass through while reflecting the remainder, is an approach still used but far from being optimal.
A significant amount of light will either uselessly pass the semi transparent mirror to the outside, or will be reflected back to the outside. The outside picture will be darkened, the display picture will uselessly be seen by external observers, glare and reflections will obscure sight, and the entire glasses assembly will look more like sunglasses, preventing direct eye contact.
One wouldn't really want to design something that always works like sunglasses, i.e. impossible to use at night, and produces unwanted reflections all the time.

All the adverse effects mentioned could be avoided by

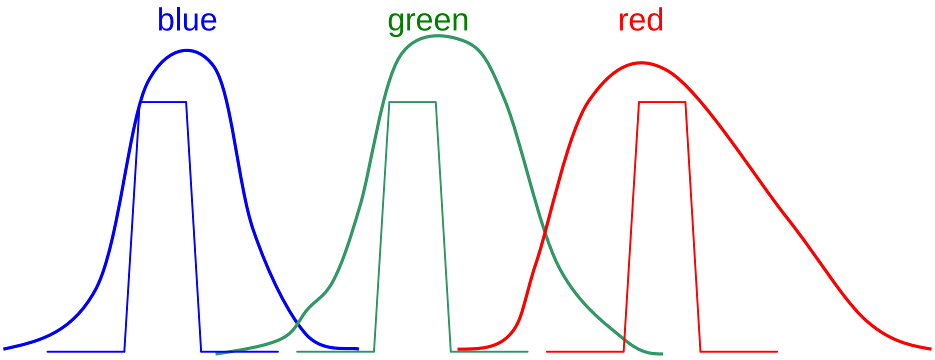
1. using directionally selective (e.g. holographic) mirrors
2. using spectrally selective mirrors (dichroic or holographic mirrors) together with spectrally selective displays.

Dichroic mirrors

Dichroic filters/mirrors are produced by deploying very thin coatings (usually 20...40 layers) on glass. The resulting filters can have a very narrow bandwidth, down to about 1 nm, compared to the entire about 300 nm of the visible spectrum. Contrary to dye filters, they do not absorb (transmissivity >90%), but selectively reflect parts of the spectrum. There are high pass, low pass, band pass and notch filters. Basically for our purpose we need notch filters, that reflect just one color, i.e. band pass mirrors.
There is no problem (except for costs) to use multiple filters stacked upon each other, in order to reflect several narrow color bands concurrently.



A typical dichroic filter curve



3 selective mirror bands vs. typical display emission curves

The principle will work the better, the narrower the original emission bands of the display colors are: In this case, more of the light produced will be reflected to the user's eye and shielded from the outside. We could therefore achieve an almost 100%

effectiveness if we use laser displays. With broader emission curves, a dichroic multi band mirror would at least improve color definition.

As the filter characteristics are defined by the thickness of coating layers, they also vary with the angle of infalling light.

With our typical vision simulator construction, light does not hit the mirror vertically. Therefore the filter frequencies would deviate. Yet it is no problem to adjust the coatings for this, even to produce coatings with a gliding change in thickness.
Some manufactures for example offer narrow band pass filters with a center frequency that varies from blue to red over the length of the filter, and with high precision. With a linear photo detector array, this makes a simple and good spectral analyzer.

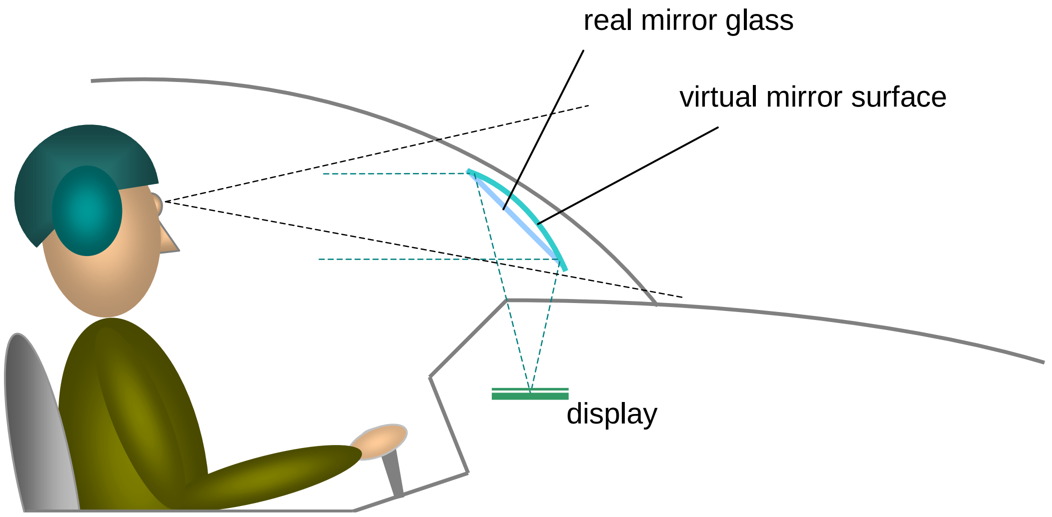
With a laser projector display, reflection angles vary only a little, and the light is perfectly monochrome, so we could use very narrow banded mirrors, resulting in an almost 100% transparency of the glasses. This would not be a 'half transparent' mirror any more, it would not visually absorb anything at all.
Likewise, hardly any of the laser light would get outside the mirror at all.

Dichroic mirrors may be relatively expensive to manufacture, but the advantages are convincing. In mass production, they could certainly be affordable.

It is meanwhile possible to deploy vacuum coatings and therefore produce dichroic mirrors on plastic glasses as well.

Holographic mirrors however, can be produced spectrally as well as directionally selective, and they may perhaps be cheaper to manufacture. Let's have a look at this.

Holographic Mirrors



Headup display (HUD) in a fighter jet plane (principle)

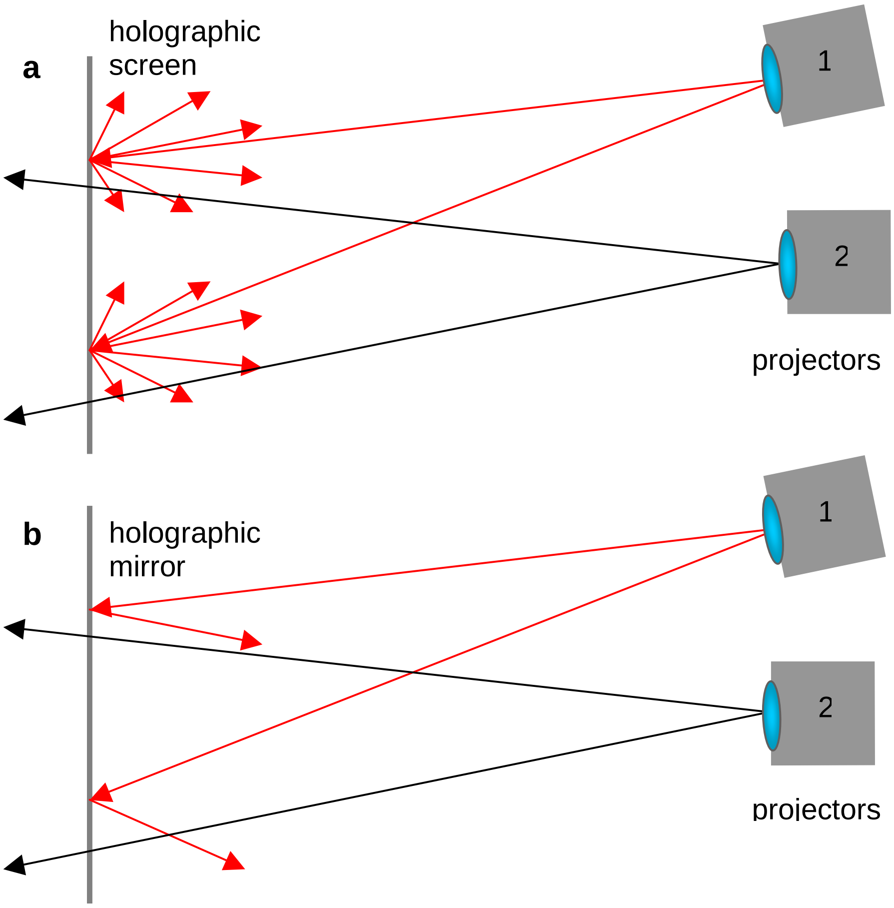
Holographic patterns cannot only represent images, but also optical elements like lenses and mirrors. So called volume holograms, recording a wavefront in a thick film layer (several wavelengths thick) exhibit even more interesting properties. Usually such holograms are recorded in plastic film, and the pattern after development is not represented in gray values but in zones of different diffraction index. They can be transparent to ordinary light and show their special behavior for light of a certain wavelength only, coming in at a certain angle only (constructive interference here has to comply with level and vertical patterns concurrently, which leaves only a few results).

A virtual mirror as in the picture above doesn't need to appear in the substrate plane, so a planar glass plate can exhibit the properties of a concave mirror, without any lens effects appearing.

This technology has been established decades ago, with headup and helmet displays in fighter jets, recently also with headup displays in civil airplanes. It's actually a bit overdue to advance into other areas.

The virtual image in a headup display appears at infinity, so eye adaptation changes are avoided and the image is position stable against the horizon even if the pilot moves his head.

Not only mirrors can be represented in a volume hologram, but objects like canvas screens, for example, as well. Such a mirror or a screen can be recorded by photographing a real mirror or screen with an holographic setup. After development, the reference beam can be replaced by the (laser) light of an image projector.

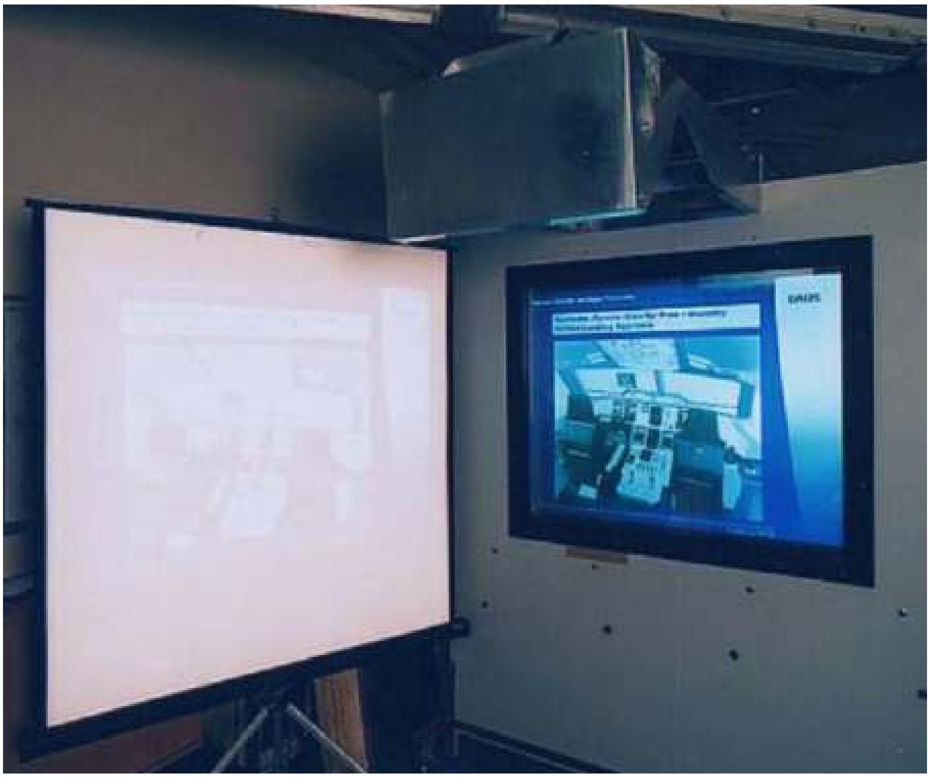


This way, the virtual screen in the hologram shows up in the projector light, with a decisive difference to the usually homogenous lighting: according to the image in the projector, any part of screen gets different intensities, forming the image on it.

With the proper production technology for the holographic screen, the object can only be lit from the same angle as with the original reference source (1) [42].

So any other ambient light cannot spoil the image. The light from the lower projectors passes the screen untouched (2). The reflection characteristics are those of the photographed material, so we have all options from canvas (a) to mirror (b).

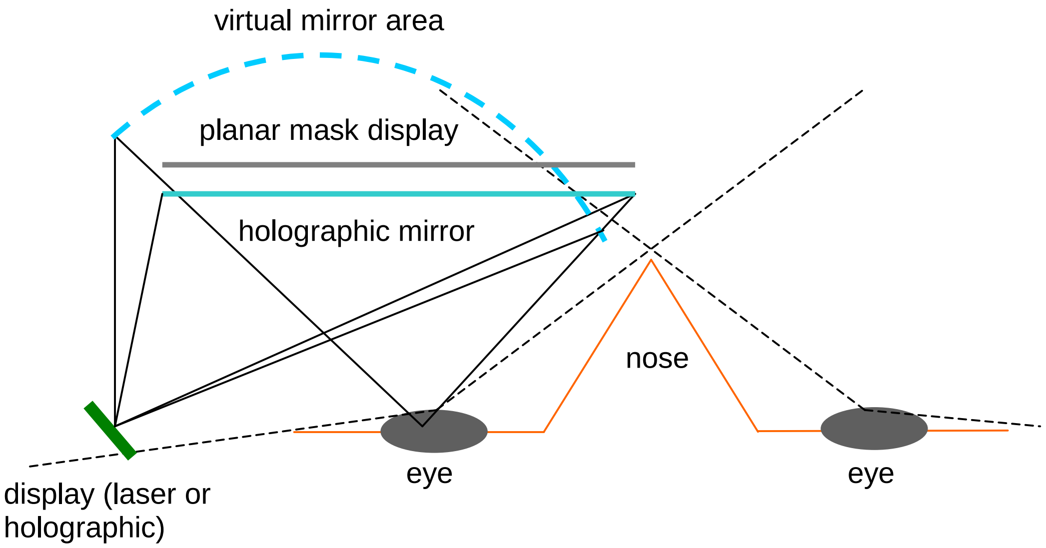
If an efficient mass manufacturing process could be developed, this technology could breed a new generation of projection TVs, that would really deliver the same contrast as any self luminating screen, and likely that these could be less expensive than current flat screen displays.



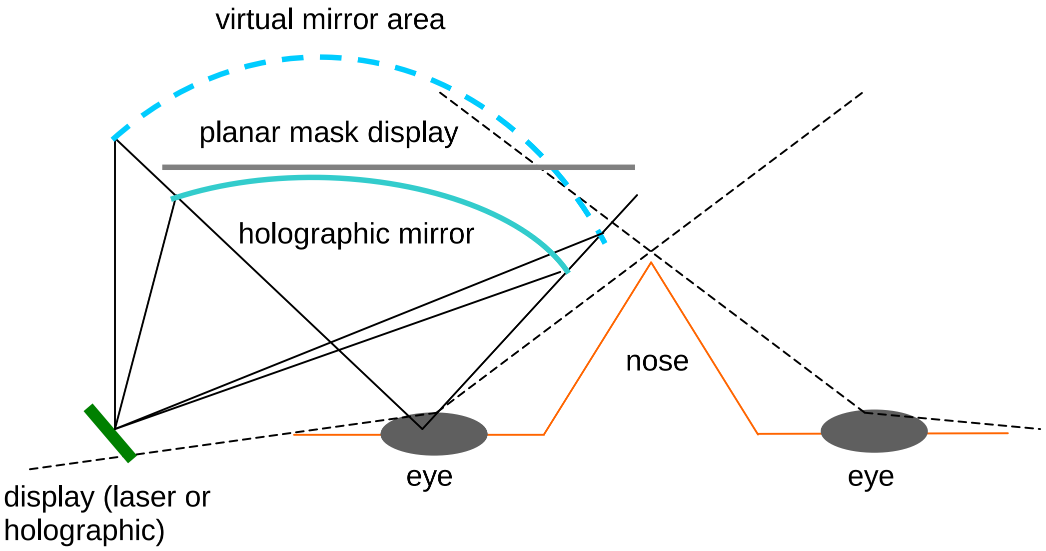
Comparison of a classical projection screen (left) with an holographic projection screen (right) [42]. While the image on the classical screen is almost entirely overlaid by ambient light, the contrast on the holographic screen is nearly 100% (image courtesy of EADS)

For color images, we need three screen layers for the primary colors (red, green, blue). Mixing holograms for 3 colors into one layer is possible, but then at least directional selectivity would suffer.

With display glasses, holographic mirrors could even take shapes that, if cast in real matter, would not be fitting into a physical design at all. Again, the holographic mirror can also work direction selective. Unlike even a dichroic mirror, it could hardly pick up and reflect any unwanted ambient light.



A hypothetical assembly with laser or holographic mirror and a projector sideways in the glasses' handle. Not needing the curved glasses, would make it possible to use a planar mask display, greatly simplifying this part.



In the first example it is quite clear that the reflection angle from the planar glass carrying the holographic mirror will become a bit large, especially at the inner edge. If we use a curved substrate, we could mend this and still use a planar mask display.

Although extreme angles would be possible, they would cause some disadvantages, and it would get quite difficult if we wanted to stack mount three films for the three basic colors (colors can also be combined into one hologram, but with even more restrictions). As holographic elements are usually thin polymer films either with imprinted or with photographically manufactured patterns, these could also be deployed onto a non planar surface.

Artist impression of an ultra light advanced vision simulator

The futuristic glasses illustrated here would have advanced mask displays - in this case perhaps planar ones - and holographic mirrors. The optical assemblies with position cameras, holographic / laser projectors and eye trackers are built into compact units in the handles. Also shown are the earpiece/microphone units.

It may take quite some time until we see something like this, but it's worth trying.

Let's discuss the advanced approach a little further: the glasses of such a device could as well each consist of two plastic glass domes, between them some yet unspecified mask display technology whereof we may only assume that a transparent driver circuit is deployed on the inner side of the outer glass dome for example, and a counter electrode on the opposite surface.

A holographic mirror would be attached to the inner side of the assembly, consisting of one or more thin plastic foils with appropriate microstructures. This holographic optical element could also be more complicated than just a mirror, e.g. it could have some properties of a virtual lens as well, making the entire assembly even more versatile.

The display of choice would most likely be a holographic light modulator and could consist of a single silicon chip with micro actuators forming a variable sub micrometer structure.

The hologram could be dynamically controlled to accommodate for focus changes and pupil movements, and it could also work as another optical element, correcting focus for the main mirror, for example. This way, the conundrum between enlargement, focus and geometry always haunting single element designs, could be resolved in a very elegant way. Some focus independence allowing for large magnifications is a property of laser scanners as well, but the hologram is the only approach that can deliver correct focus for different distances of virtual objects, and this even without adjusting for eye movements all the time.

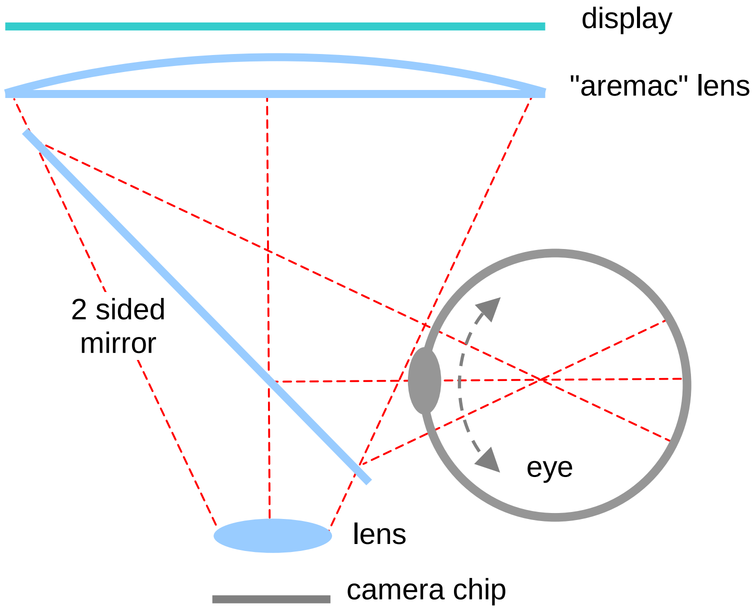
The main difficulty with this approach: synthetic holograms are difficult to calculate. Yet here we can facilitate the task a bit, as the holograms are viewed from one direction only (with a slight deviation just given by pupil size, that also establishes the focus effect), and we can restrict most of the effort to the image part currently viewed at, as human eyes have a very restricted crisp viewing area. The computation requirements nevertheless are still heavy. It may take some years until computer chips and algorithms catch up to the challenge.

As stated, a holographic display could also intermittently be switched into a mirror lens for an eye tracker camera chip. This isn't shown here, as well as the illumination.

Hence, the entire display hardware for one eye (the part included in the glasses, not the hologram computer that would go with a pocket unit), eye tracker included, would essentially consist of two silicon chips, a holographic foil, a thin mask display foil and a laser source the size of a sand grain, all together weighing less than 1g.

About eyetaps

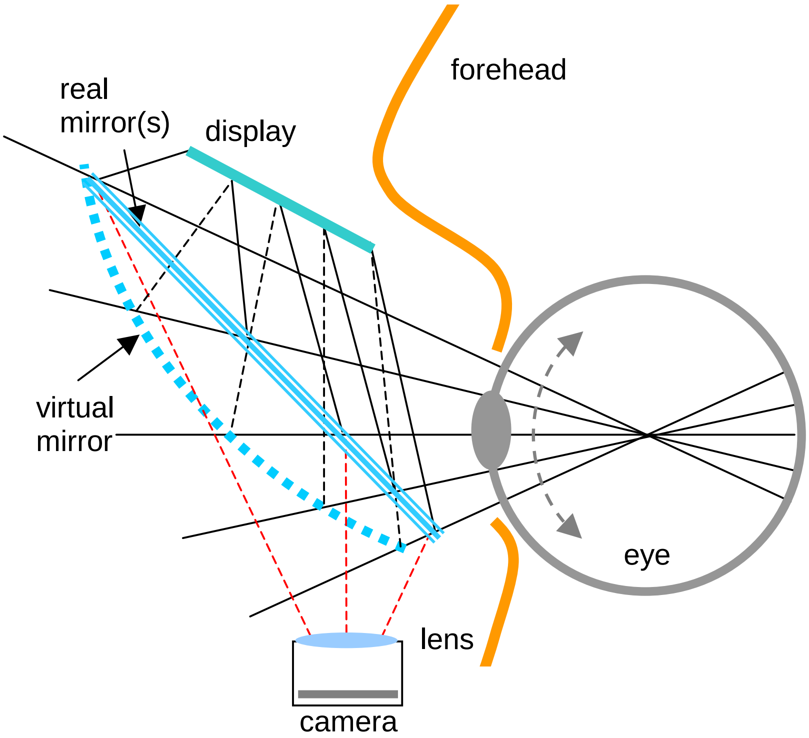
Just for completeness, let's have a look at a display variety that replaces the entire sight by artificial images from cameras. This is not what I think to be acceptable, but it's the currently most frequent approach to fully integrated virtual objects. This approach does not use (or need) a mask display, obviously. Its other disadvantages are however severe.
An eyetap is defined as a visual interface that intercepts the optical path to the eye, places a camera in the altered lightpath instead, and synthesizes the very same visual impression from a display, where also additional (virtual) image elements can be inserted.



The classical eyetap assembly uses a 'beam splitter' (actually rather a 2-sided mirror) that delivers to a camera the same perspective that the eye would get. Indeed this isn't really true already, as the eye will always move to center in on the point of interest while a camera stays static.
The other part of the assembly should reproduce exactly the same rays that the camera has stolen away, by a display that obviously needs a concentrating lens to produce this reverse light bundle. The assembly is called an 'aremac' (camera spelled backwards).

It's obvious that the example shown here collides with anatomy, as there is no sufficient space for the 'aremac'. Actually working designs of this type can only deliver a much smaller viewing angle.

A wide angle eyetap attempt



A design study for an advanced eyetap. The real mirror plate is 100% reflective on the outside and carries an holographic mirror on the eye side. The outside mirror reflects all incoming light, and the camera gets the same perspective the eye would have.

The holographic mirror is curved and focuses the display image towards the eye. As the virtual mirror lies outside the glass plane, there is a large difference between the virtual and the real angle the display is viewed under (solid and dashed lines).

Like any eyetap, this assembly has the apparent advantage that it needs no mask display. This here also has a larger viewing angle than other constructions. Nevertheless, it is not only as undesirable for everyday use as any indirect viewing device, it also looks quite awful because of the large flat mirrors and the cameras hanging from them.

Apart from the problems of specific optical designs and the capabilities of current cameras and displays, the eyetap idea has several other difficulties as well:

1) We need to move our eyes to see sharp in more than one direction. Therefore we can't just replace the eye with a camera and expect to get an identical optical behavior.
2) Replacing the entire view with an artificial one requires to reproduce a viewing angle of >180 degrees. The only alternative would be to have a display that leaves the outer skirts of our field of view open for direct vision. This requires 100% contrast and brightness identity of the display and the natural picture, a lot more difficult than with our approach.

Another approach could be, bringing some optics before the eye to create an intermediate crispness plane, where either a transparent- or a mirror- or a combined display could be placed. In case of a mirror display, a DMD could, for example, switch between the original picture and a projection of virtual objects.
If this was possible, it would have the advantage that (almost) no active elements are in the way of natural sight, but still the optics we would need are hardly conceivable at all, because they should deliver 180 degrees field of view (or at least seamlessly insert into the natural viewing range), shouldn't distort the picture and also shouldn't be heavy. Nothing like this is in sight.

Enough reasons to concentrate on the entirely different approach of just adding pictures to natural sight.

In order to insert objects without appearing transparent, this obviously requires a *mask display*; also not an entirely simple way, but one that's likely to work.

The mask display

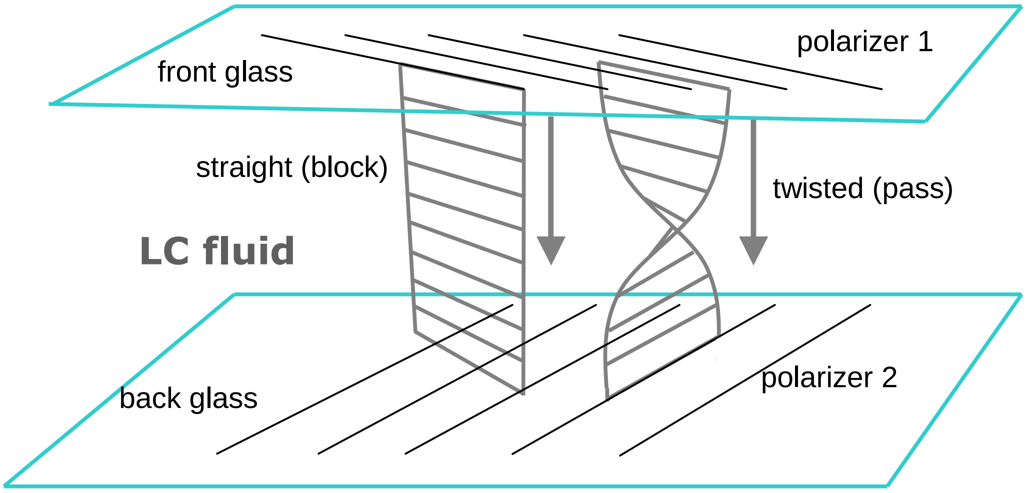
In the introduction we have already seen that a mask close to the eye will work, even though it can't be sharp. Maybe holographic technology could even lead to something better. The bad news is that there isn't yet an ideal candidate for the mask display. It is mandatory that such a display could switch from black to transparent (not opaque).

It is also quite desirable that the mask display could be built in a concave shape. Otherwise it could be suboptimal or perhaps look bad in front of the main mirror.

There is no need for high resolution, as this display will always be seen unsharp. Nevertheless, pixels should be completely adjacent to each other, being able to cover an area entirely, but not leaving a non transparent grid when not activated.

Last but not least, the mask display should be light and rugged, e.g. best built from plastic.

LCDs could do it, but 'transparent' here still means a significant light loss due to polarization filters:



Two polarizer filters (1),(2), normally block light from passing, but with a voltage applied between the two glass surfaces (that carry a transparent metallization for this purpose), the liquid crystal fluid twists the polarization vector of passing light, turning the device transparent again.

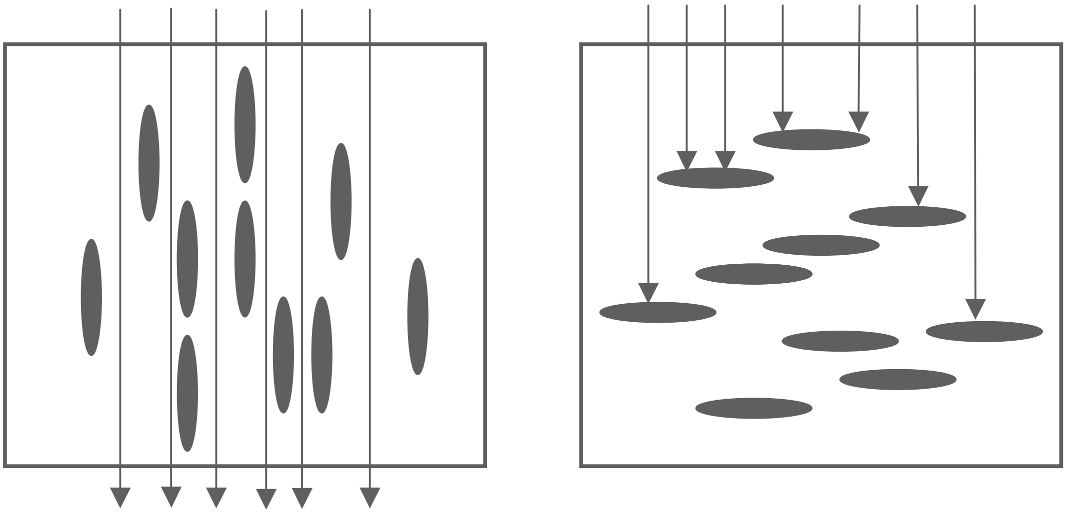
Obviously, an LCD panel loses at least half the light. Nevertheless, it would not look worse than some pretty weak sunglasses (you may check this with one of those fancy desktop watches that seem to consist of nothing but a free standing glass plate, and these aren't even optimal). It could even be much better if one could make a polarizer that turns everything to one direction instead of just absorbing one part.
LCDs are also not easily manufactured other than planar. By the way, there is a patent [13] that deals with all kinds of LCD shutters and may be interesting to read, but there is no word about using them as a mask to cover up or release selected image areas.

An LCD shutter could principally be built with plastic glass, if spacers between cells keep cell thickness constant or if we use domes, which are more rigid. Life expectance would be a problem with this, because of vapor diffusion through the plastic. We would also need a transparent structure to drive the LC fluid. Passive matrix drivers could be sufficient in some cases, for low resolutions. Conductive metal layers, thin enough to be transparent, can already be evaporated on plastic.

Active driver structures would provide more resolution and speed. Transparent circuits have just recently been demonstrated, with vacuum deployed layers based on tin- or zinc oxide. Organic driver structures like some developed for OLED, might also be developed in this direction. If the driver structures were flexible, the display could be manufactured in a planar process and blown to the spherical shape later on. It should also be possible to vacuum deploy driver structures onto curved substrates directly.

Displays that are genuinely made all of plastic, like polymer displays, are another option. Electrochromic polymers that switch between transparent and selective color band absorption, may be a candidate. There are red and green transparent colors currently available (each generated by absorbing the rest of the spectrum). Current research has led to a green variety that has 2 absorption bands [7] (3 layers could even allow subtractive full color mixing). For our purpose, it could still be necessary to use two absorber layers to cover the entire spectrum, and transmissivity is not yet ideal. Switching speed is good. A prevalent problem is durability.

Dyed Guest Host displays and similar technologies work with microscopic disks diluted in liquid crystal fluid for example, that can be tilted by electric fields.



Oriented parallel to the light path they hardly influence it, while perpendicular they form a 'solid' wall. It could be an ideal mask display, but contrast is only about 1:5 with current samples [33].

We could of course speculate about many other technologies:

Nano particles smaller than light wavelength, could perhaps be caused to agglomerate into larger particles to absorb light.

Electrowetting displays with dark drops could also be considered but aren't really likely to work here.
The solution could as well come from techniques similar to phototropic glasses. These use silver halides that fell out silver particles when lighted, just resembling a photographic film. We need something fast and electrically operated, but maybe this principle could lead us somewhere.

Energy consumption

As we are obviously dealing with portable units, power requirements should be as low as possible.

The display is not the only energy consumer in the vision simulator, but it's a good point to start at.

In case of the DMD version, it's already obvious that we only need to give the LED illuminator a certain beam focusing, in order to concentrate most of its light on the DMD and thus towards the eye. The optics always have to be focused in a way that parallel rays at the eye would converge at the display, hence parallel rays at the display (from a directed light source) would as well converge at the eye.
The same would apply if we use an active LED display with micro lenses concentrating the light of each LED individually.

So with DMD as well as LED displays, a great part of the light could be concentrated on the eye.
With laser displays, effectiveness can even be up to 100% starting from the light output of the laser diodes. The effectiveness of these diodes themselves is not ideal, but still acceptable.
Much of the energy of a laser display could be lost in a modulator. Laser diodes can't simply be dimmed down like LEDs, as they will soon go under the threshold power. Yet they can well be operated in pulsed mode.

Holographic displays could also concentrate light energy to the eye in a very flexible way. Even so flexible that some safety measures would be necessary.

Let's calculate the light power we need. Bright sunlight equals 10000 lux. If we face a white wall this bright, the eye will also be illuminated with approx. 10000 lux. Given a pupil diameter of less than 2mm and a retina area of approx. 20 mm in diameter, the retina would get 1/100, or below 100 lux.

Even then, nobody would deliberately endure this without taking some sunglasses, or dimming the mask display. Hence, 20...50 lux on the Retina are all that could be asked for. Given the fact that virtual objects cover only part of the viewing area, an average equivalent between 1 and 10 lux seems realistic.
10 lux equal about $1W/m^2$, and the retina area is about 1/2000 m^2. This means we need between 0.05mW and 1mW of actual light power entering the eye. In bright sunlight, that is.

Trying to do this with a large flat panel screen, even if we succeeded, we would need *kilowatts*.

The ratio between electrical power and light output can vary between 5 and 30%, and depending on display type, between 10 and >50% of the light would then actually enter the eye. Which gives a wide variation of possible results.

In every day use, full sunlight is the exception. Usually we have situations with 10 or 100 times lower brightness, which reduces power requirements a lot. With outdoor usage, we wouldn't normally use the display to provide a giant movie screen, and even if we did, we wouldn't endure maximum brightness.
We could therefore conclude that something around 1mW average electrical power would be sufficient for the display, even in relatively bright environments.

The computing power we need is much more critical. It's obvious that we have to externalize most of it. Even then we still need a communication link with the computing device that can carry several TV channels, 2 channels in high resolution.
We could use some compression in order to save bandwidth, but that needs more power, which could only be kept in an acceptable range maybe, if we use dedicated chips for compression and decompression (note that CRL Opto announced to develop F-LCOS displays with integrated MPEG decompression, see [33]).
Driving a synthetic holographic display would be difficult. The image processing chips would be power hungry, but if we exter-

nalize them we may have to transmit up to some GHz of holographic data. Which is still less of a problem than to turn the chips into real power savers.
Another power drain will result from communications, if we use wireless. As the distance to span would only be about 50 cm, a radio frequency output in the range of some milliwatts should be sufficient, even though we need so much bandwidth.

The sound system is really modest. Good button headphones can produce a tremendously loud sound just out of some mW (milliwatts) of electrical input.

In Summary, at some point in the future we may see a glasses assembly with an average power consumption of about 10 mW.

Some numbers: There are optical glasses that weigh 100g with their large normal glass lenses and that people still wear.
A 3.6 Volt/1 ampere-hour lithium accumulator for a mobile phone can be as light as 20g. This is 10 hours at 100mA, or 360 mW. If we would integrate 2 of these (perhaps in the handles for better balance), the entire glasses could weigh 70g for example, and could either operate 10 hours at 720 mW or 24 hours at 300 mW. This is all not too difficult to achieve. In a few years with some serious effort, an entire glasses unit could work in this range.

Indeed, we could just glue two of these very small cell phones already available onto some ordinary glasses and we would have most of the technology necessary for a first model of a vision simulator.

Hence, display glasses without power cable will emerge far earlier than anticipated.
When very low power versions finally arrive, some solar cells could also be used, together with miniature accumulators.
Under typical office conditions at 500 lux, 10 cm^2 of solar cells could deliver up to 10 mW, while the display illumination alone, for example, would need in the order of 0.1 mW in this case.

Dynamic image generation

It's obvious that we cannot realistically expect anybody to accept a device that is somehow 'fixed' to the head, like those gruesome makeshift VR displays that dominated the early days.

As an acceptable device cannot stick to the head more than some ordinary glasses, we have to detect the position of the glasses as well as the viewing direction and predistort pictures accordingly. This all adds to the basic requirement to generate dynamic 3D perspectives according to head position.

Head position can be detected by (without claim of completeness)

- Cameras registering real objects and their relative distance and angle to calculate their own position
- Tilt sensors using gravity to determine the vertical
- Acceleration sensors and rotation sensors, exploiting inertia
- Magnetic compass sensor
- GPS (Global Positioning System) and local varieties

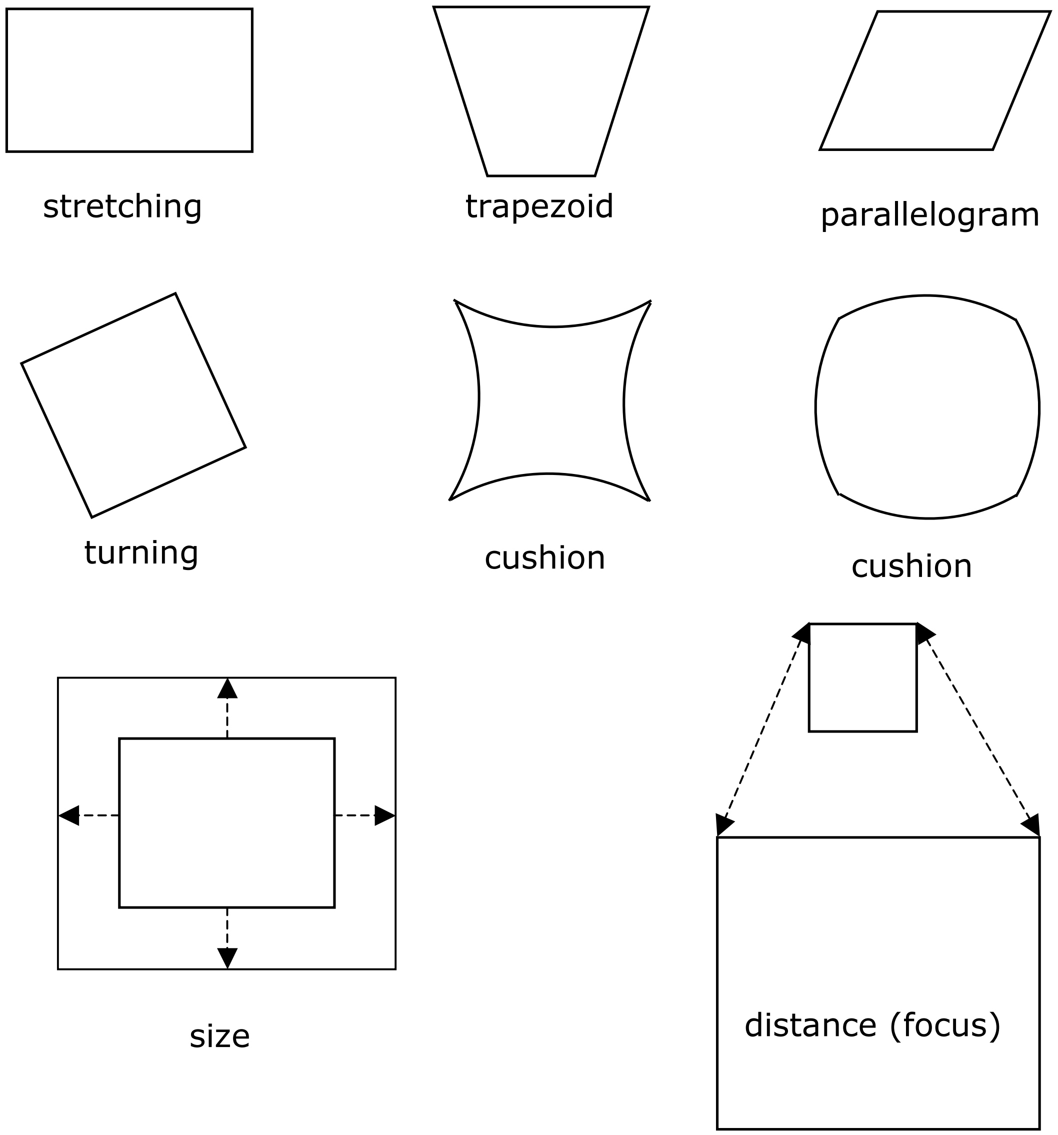
Tilt and acceleration sensors are fast and have high resolution. They are necessary to do any image compensations precisely and in real time. They are also necessary to do predictive position calculation, because image generation and usually the display as well, will have a delay of at least 1 frame (e.g. 20ms), which would result in a significant displacement with fast movements.

Displays building the image pixel by pixel, like CRT and (laser) scanners, offer the possibility at least to change their beam deflection in real time, hence they could react to acceleration sensor input directly, within microseconds. This would allow for a really rock stable image impression. Other displays like LCD build up an entire picture at a time and often have slow switching times. It's inevitable that with these displays, fast movements could result in sort of a stroboscope effects. A fast frame rate (75 Hz or more, like with computer monitors) can help to avoid this.

With a very small display and a huge enlargement factor of the optics, we could end up with a small aperture size, that would cause the image to vanish if the eye moves sideways. Compensating for this would require to shift the entire display. This could work like head positioners in harddisk drives, for example.

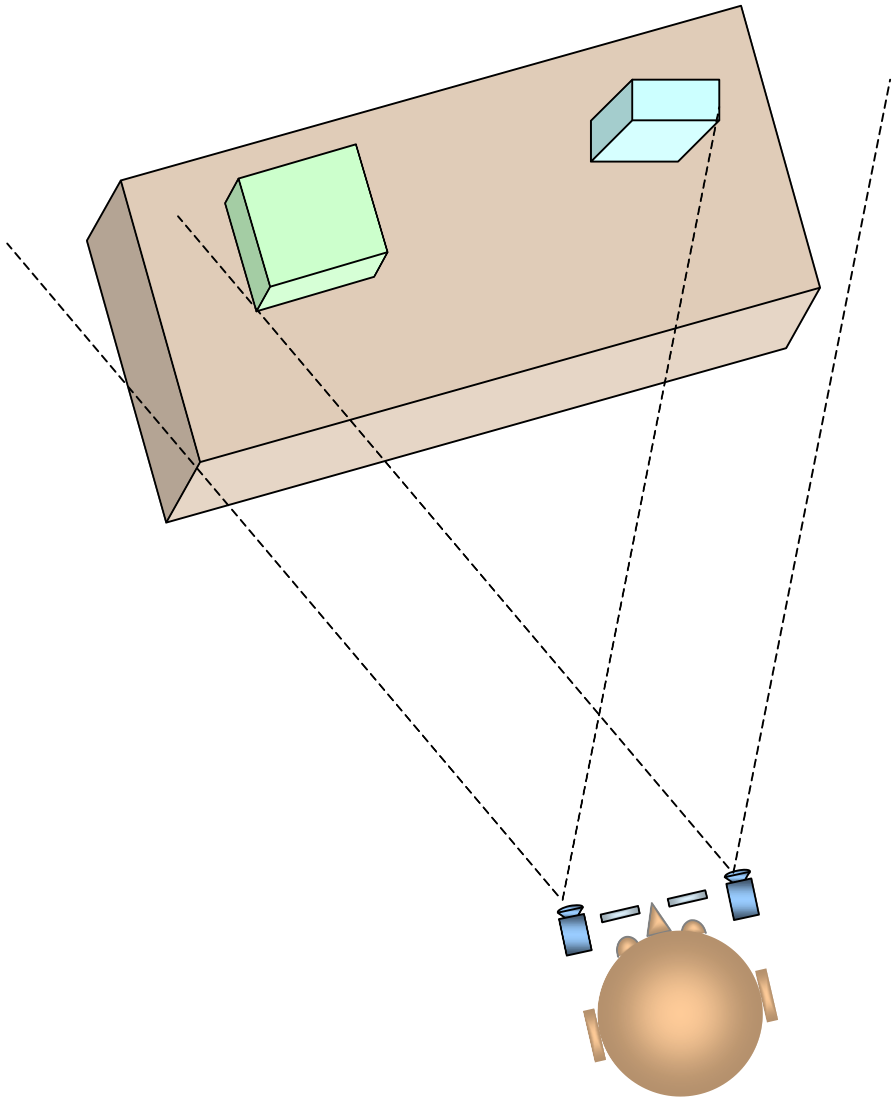
Moving the display to provide exact focus just for the image part currently being looked at, is another important ability as it provides for a realistic distance impression as well as it lowers the requirements for optical precision.

As it is absolutely mandatory to create a steady optical impression, any mechanical or electronic movements have to be accompanied by an electronic image predistortion.



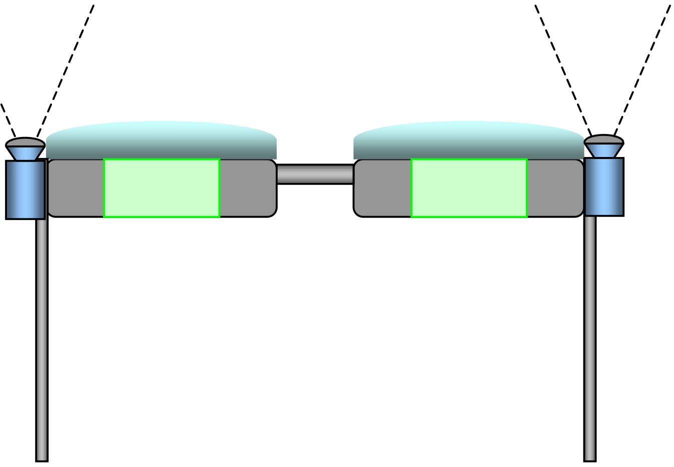
Types of distortion precompensation in the main display

Position and scene identification

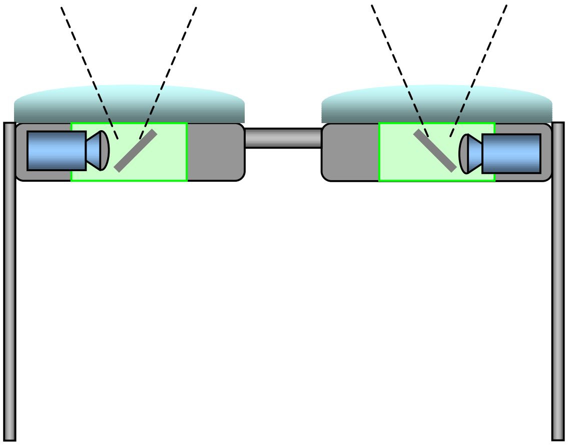


Apart from GPS and acceleration sensors that are also considered or even necessary, the 'right' way to built a position sensor for a universal augmented reality display – what I here call vision simulator – is to use cameras that recognize the environment and calculate their own position from the relative perspective. It is the best approach because it works like our own visual orientation.
This method has some difficulties that we will now have a look at. Note that in [22], advances in this area have been reported, and there is another thesis just finished [41].

Position camera assemblies



Side cameras have a larger stereo basis, maybe an advantage, but more deviation from the user's own perspective, and bad perspective matching if used for stereo camera recordings or teleworking.



Smaller cameras could be moved right above the eyes, or mirrors could be used to bring the camera lens locations as close to the eye centers as possible, achieving a closely similar perspective. Everything including displays, if small enough, could just be integrated in the upper section of the glasses. With laser projection displays, the cameras could even be positioned more freely.

We could also use holographic or dichroic mirrors or holographic prisms or lenses to steal away some light from the direct path to the eye and get a camera picture from the very own perspective of the user's eye. This would not work well in dark environments though, and the parallactic corrections we need with the simpler assemblies are not really a reason to accept these and other difficulties.

Ignorance required: Solving the orientation puzzle

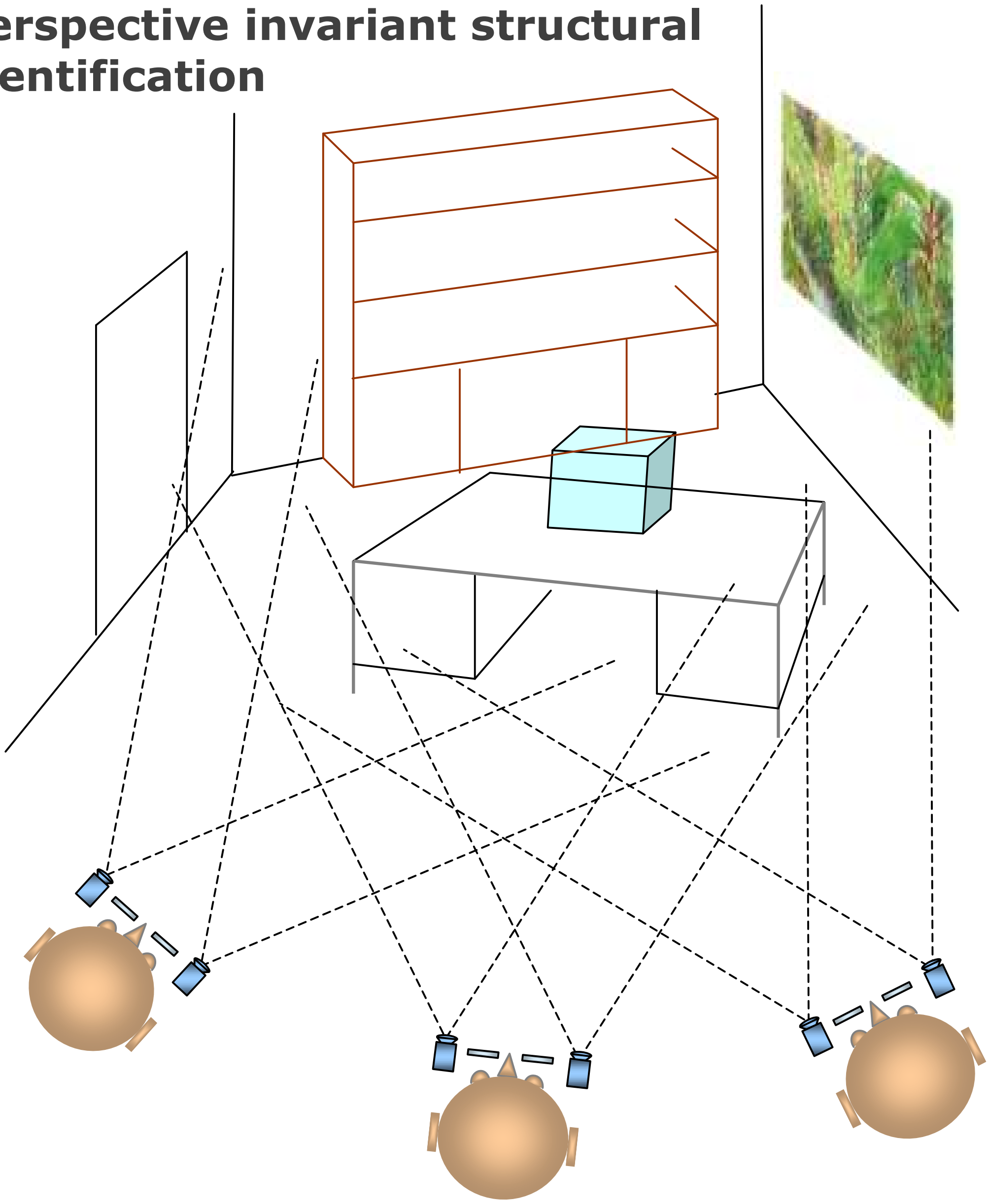
Let us first state that it is absolutely unnecessary for the system to know what it sees, not even to separate objects (it does not need to know that some edges define a desk, or a chair, or a sideboard). We do require nothing but to re-recognize edge structures later on.
This already is not so easy, as these structures look differently from each perspective. So the biggest challenge for our task is to develop a 3D structural search and recognition algorithm that is ignorant to perspective related variance of this structural data.

Could we achieve this, the comparison of the 3D variations (perspectives) would also yield very accurate position information about the viewer, enabling us to correctly restore any virtual objects that were previously defined in a certain place.
The system would know to have seen this room or scenery before, could also determine its exact position relative to it now, and could correctly display any virtual additions desired.

I will now outline some possible data structures and algorithms for visual position sensing in arbitrary scenes.
As said before, one difficulty can be avoided: we will never need to isolate real objects (like chairs, cars, etc..), not even to mention assigning to them any meaning, even not to find directly adjacent detail, but it is still not simple, because in practice, there are many influences obstructing and distorting image detail.

What we first have to achieve is simply to be able to compare a *detail* that we have seen with a *detail* that we see now, then their geometrical relations with other *details*, and this in the fastest way possible.
It's obvious and always supposed in the following contemplations that in practice we will deal with many subsequent pictures from varying perspectives that allow to assemble an omnidirectional view and even to link objects in different directions or rooms.

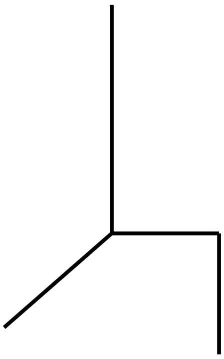
Perspective invariant structural identification



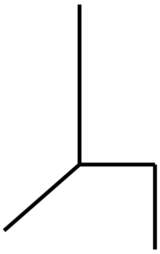
This picture shows several vision simulator users, viewing the same simple office scenario from different perspectives. (Also depicted is a virtual object on the desk, how this is dealt with we will discuss later). Real environments, like this office (concededly one of the simpler cases) exhibit some basic structural data that can be picked up by the stereo cameras of a vision simulator.

Easy to detect are edges of solid 3D objects, because these can be separated by stereo cross correlation, and the data gathered is already 3D. We also see an image at the wall; we will have to define special treatments for its 'surface patterns'.

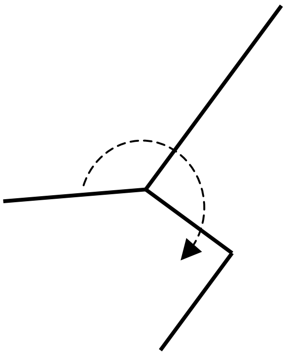
Basic examples of structural data variations



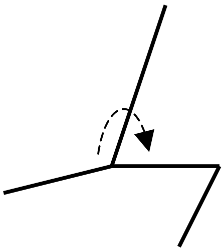
original edge graph



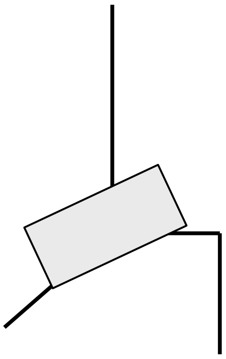
distance



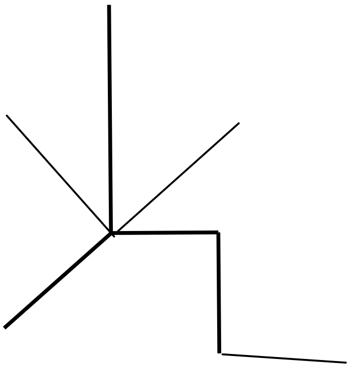
rotation



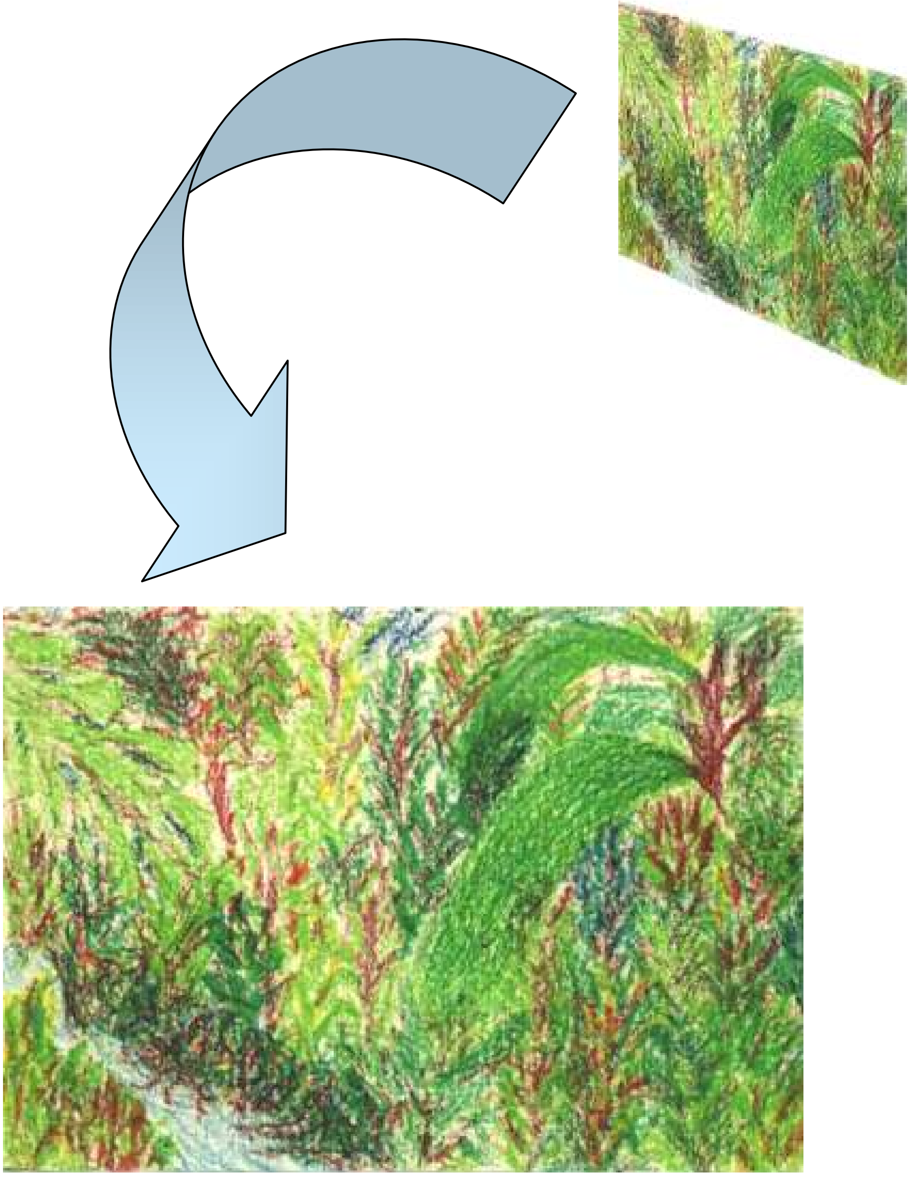
perspective



hidden parts



added detail



The approach shown here is that of normalization : If we could define some rules that certain structure shall always be stored in a certain orientation and in their original size, we would later on be able to search for them in our database with a good chance of success.

Obviously, a single angle, tripod or rectangle (to name some basic elements), is absolutely insufficient for a recognition. Yet we could store more complex structures starting from the detail, defining adjacent edges, angles, surfaces in a tree like data structure.

This could look as follows:

- Use simple cross correlation to match lines and edges between the images of the stereo cameras.
- Use this data together with gravity sensors to bring everything to the vertical and to its actual size.
- Use some heuristic algorithm to determine color and brightness in order to normalize these values for object surfaces seen.
- Store the geometrical relation between basic details, linked with the details themselves, so that they form a mesh that could define a certain scene just starting from some arbitrary detail without having to care for any other structures between them. This may be quite similar to our own way of orientation: We remember some details very accurate, but many others not at all.

The real size of detail objects ('atoms') can be obtained by simple calculations using stereo perspective. Orientation needs standards that have to be defined, for example a flat object will obviously be stored in a normal perspective just up front. If we store every detail object this way, and its orientational data relative to the real world and to virtual objects separately, we only need to apply the same normalization (that is, simple coordinate transformation) to any detail we want to compare to the database.

A picture at a wall (as a simple example) can be identified as a single and simple detail object and its real size and orientation can be obtained from the stereo vision by correlation and triangulation. The image can then be normalized to its real size and normal view. Storing images in order to compare them, could be done by storing coefficients of a discrete cosine transform, like with JPEG encoding, as one of the possible methods. Coefficients could be compared hierarchically from coarse to fine, enabling a search among thousands of pictures without much effort.

Surfaces of any rectangular objects are fully identical to images, which allows for a generalization of the approach. In case of irregular objects, e.g., a plant or a tree, one approach would be to store them just like pictures, from different sides (always keep in mind that we are really considering apparently adjoint structures, rather than really distinct real world objects). What we always need is a reproducible algorithm to define the directions. If we don't want to store their entire structure, we could store perpendicular views, for example. If there are straight sides, it pays to normalize relative to the most prevalent features.

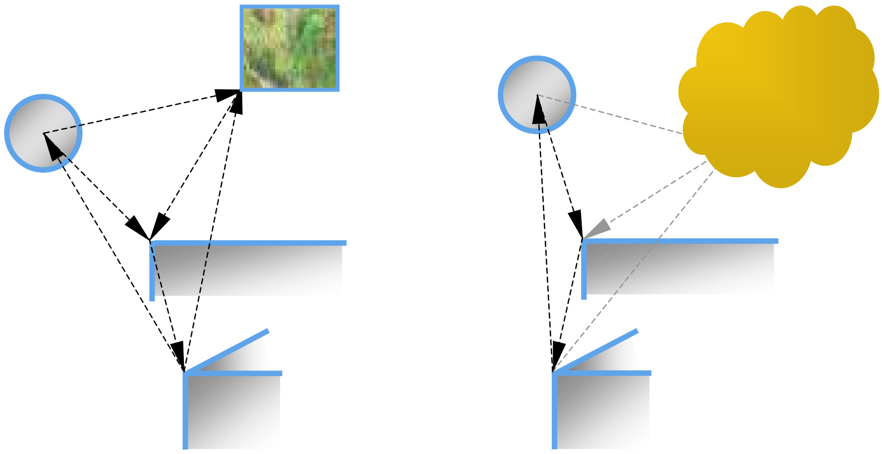
We will always have to store the angles for the real orientation towards a normal (usually North), in order to bring this data together with orientation coordinates. In a human made environment, we will usually be able to find some simple structures for orientation, and this may already suffice for many applications.

Nevertheless, even orientation in very irregular environments (outdoors, wilderness) has recently been approached in [22]. It would generally be very helpful – or even required – for the approaches discussed, to have some additional position sensors, at least for perpendicularity and also for North direction, in the vision simulator. Even in a simple and inexpensive version, a GPS would be a very natural extension. What could also be envisioned are local setups of time stamped transmitters, that could do for a building for example, what GPS is doing for the entire planet. We could also send signals from the vision glasses and account for their arrival times at different receivers in order to provide position information. These local varieties could require less power or smaller antennas that real GPS, would also be more accurate and have less reception problems, so they could really be an alternative for local applications.

Searching structural graph isomorphisms (remembering locations)

As any scene has many elementary details, we will have to store the same scene from many starting points and use interlinked pointers to keep the data structure from becoming too redundant.

Developing the best data structure is one of the main challenges we are facing here. This is very different from the classical image recognition approach, that tries to find separate real world objects. Although we also start from elementary structures, we do not care for real objects at all. What we care for is meaningless structure. A structure can consist of a square at one wall, a corner at the other and a circle that's really part of a lamp, for example, these linked by pointers noting their relation and distance values noting their relative position. It's absolutely irrelevant if these details belong physically or logically together in any way. Anything we care for is geometry, and maybe the pattern or color in a basic circle, triangle or square, or directly adjacent to a point, line or corner.



To the left, several primitives ('atoms') of a scene have been normalized and linked with pointers. To the right, although one of the primitives is occluded, the others together with their relating distances and orientation coefficients etc. (not shown) provide good confidence that we are indeed dealing with the same scene

If later on a first scan for a certain detail remembered delivers a million hits (not much for a current computer), we can subsequently exclude all false matches by looking at linked detail. This will rapidly close in on matches that are likely. Some more 'intelligent' tests could also be added, to get to a firm conclusion.

The original relations between our 'atoms' are vectors in 3D space. If we would try to actually compare two vector graphs we would end up with the same problem we already had with the 'atoms', but here it's not done with the normalization approach, as this would only work if we could implement it according to some precise global reference.

We don't however need to bother with this at all, we simply use the length of the vectors instead. Distances alone can define a structure, also in 3D, but not its orientation. So we have a simple way to match 3D structures by just matching distances, and this is inherently indifferent against rotation! Mirror symmetric structures aren't distinguished this way, but that's negligible.

The entire process again seems to resemble our own methods of orientation: it is obvious that we use heuristic selections of detail, rather than complete data sets, to recognize things and locations. Just look at a scene for some seconds and then try to remember any detail. Most probably, you won't even remember 10% of anything that could be relevant. Even in spite of this, you may still be able to recognize the picture many years after.

What we will have to deal with as well, image detail could be moved or covered by other objects. Multiple pointer paths in the data structure ensure that if any object is missing, the others are still interlinked around it. We need a heuristic approach to determine if the scene is still the same, despite of changes.

The approach in conclusion works like this:

- Identify a simple detail (corner, rectangle, circle etc...)
- Normalize the detail, including its surface pattern, to upright, facing, actual size, using position, direction sensors and stereo view.
  Normalize illumination.
- Store normalized data (patterns as coefficients of discrete cosine transforms e.g.) and its original position data (from comparisons, GPS data or such). If there are several possible normal views, store them all.
- Link detail objects by pointers and relative position data (vectors) to represent entire scenes. There is no need to have a complete set of all details of a scene, nor even any directly adjacent detail.
- We only have to store data from locations where we install virtual devices. This leads to an extreme reduction in storage capacity.

For later orientation, do this:

- Identify and normalize details as stated above.
- Search database for single details
- Follow links to see if related details are stored
- See if enough related details are in appropriate distance in actual scene. Continue search until a sufficiently large structure is found.
- Use the stored and the perceived distances and angles to determine own position.
- If available, simplify the process with GPS or other data.

A global orientation database ?

With an efficient object cataloging and search algorithm, a global object database may be possible, accessible by the web, offering the opportunity to re-recognize even objects and scenes that have never been stored locally.

We all know that we are able to instantly remember places where we had been even only once.

A vision simulator could even identify locations where it had never been before, just by matching its actual visual data with a global database. This would somehow make the GPS superfluous, or add to its functionality. At least, differential GPS would not be necessary any more, and we would be able to measure location to an inch or better.

The approach is especially interesting as GPS very often has reception problems, especially within buildings.

To achieve this, we would require a much faster database search capability than with our basic approach, where we only need to know places where virtual devices were installed.

Actual searches would always have to be carried out by the remote database server, obviously.

An approach towards utilizing image data as an aid for orientation has also been described in [3]. There, image data was just stored by a web connection, to be reviewed by human beings later on, in order to re-identify locations. Because of the perspective problem, this does not yield an appropriate algorithm for machine based orientation.

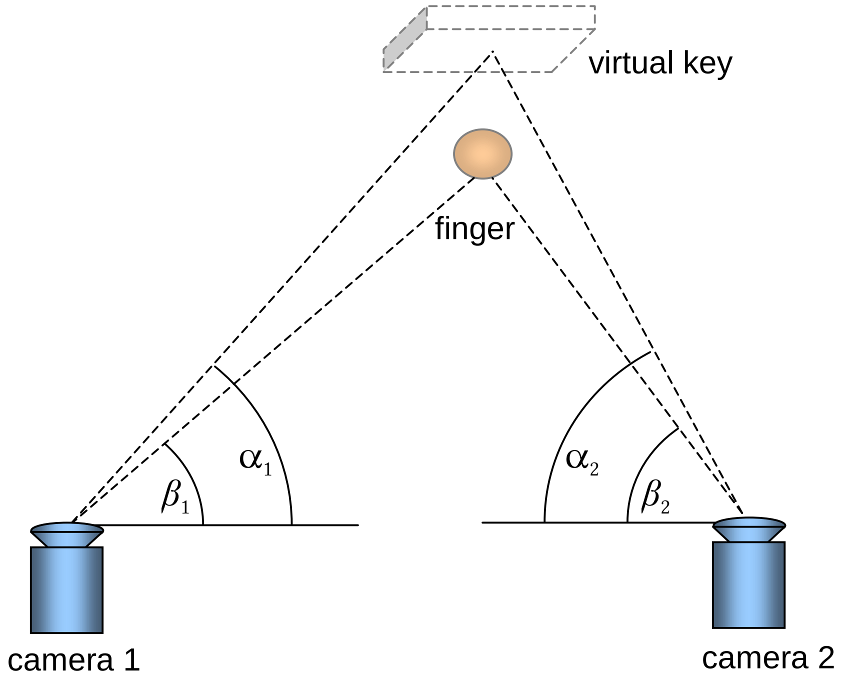
Hardware assisted image processing

Currently, substantial research is going on to create hardware that emulates the structures of the human neocortex [12]. One application would be image recognition. If these developments succeed, it would allow to save a lot of computing effort and power consumption, something very important to a portable system like a vision simulator.

Manual interaction – mastering virtual keys

Cameras attached to a display could not only give a direct means of virtual/real scene fitting. They might also be utilized to monitor the user's hand movements and gestures, allowing direct interaction with virtual objects, such as pressing imaginary keys, moving objects or defining new objects out of virtual menus projected into space. As people can hardly ever be convinced to always wear datagloves, this is a fairly important issue. Tactile feedback is not possible with this approach, but is also not necessary in most cases. A key pressing animation or a beep is usually sufficient.
Many research projects have already tried to implement gesture recognition. Common to all is the relatively high difficulty to recognize the hand and interpret the commands in complex environments. It has also been tried to recognize hand actions towards real or projected keyboards.

Much easier is the approach with Virtual Keys:



Consider we use the glasses assembly discussed above. A virtual key is displayed in the vision simulator. The position of the key is given by the relative angles α_1 and α_2. If the user tries to operate this key, a finger must appear at approx. some slightly smaller angles β_1 and β_2 in each of the stereo cameras. It must then stay there for a short moment and move towards the virtual button, increasing β_1 and β_2 to about α_1 and α_2.

With a simple correlation operation, a recognition of flesh tones and perhaps a simple shape recognition (finger shape), we can easily determine if something relevant happens.

When trying to press the key, the user must move his finger towards it. The x,y coordinates where it would appear in the stereo camera pictures (close before the key) are also known.

The principle advantage here lies in the fact that we only watch one certain point of interest, if a thing like a finger appears in the camera picture. Hence, the entire complexity of this approach is decisively lower than with the usual hand recognition schemes.

This setup could be further secured against accidental wrong inputs by some checks if this is really the finger of the user, or by requiring some special action like a 'double click' or others to trigger the virtual button. We could further reduce the effort by regarding only keys that the user looks at. This way we could also include keys that are out of direct reach.

With this approach it is *not* necessary to identify a hand or follow it, nor to recognize any gestures. In order to distinguish key pressings from accidental movements, light changes etc. one can identify color, shape and movement of any object, but only in the very small area of interest, hence with extremely low computing power. Once a key depression has been detected, the algorithm could then as well follow the finger, in order to make grabbing and dragging of objects possible. This tracking operation is still not at all as difficult as gesture recognition would be.

We have to keep in mind that we must program the mask display to shape an optical path for any real objects positioned in front of virtual objects. This also applies to the user's hand when he moves it towards a virtual object. This has nothing to do with hand recognition. Any near objects are sorted out by stereo distance measurement only.

A lot of basic things concerning virtual keys has also been covered in [26].

Cluttered spaces

Fingers can point to keys, surfaces. or discrete points in 3D structures. An entirely sensitive 3D structure makes no sense. So we always have a way of discriminating arbitrary movements from objects touching sensitive points or surfaces.
We have to consider that the position cameras could see fingers touch an object that for the user is hidden behind a more proximate object. This has to be excluded from interaction, so the user couldn't accidentally cause unwanted actions on objects he can't see.
This is even an issue with eye pointing, although it isn't quite likely that a user's eye parallaxes would correlate with a far object that's hidden behind a near one.
Program windows however could sometimes be hidden behind others, so we have to assign an eye pointing action to the right, the proximate one.

Eye Pointing

A great way of interacting would be to have the user just simply *look at* a key to be pressed. The eye tracker data always allow to determine viewing directions. People also don't usually stare at things without a reason. Normally, our eyes are moving restlessly, fixating different points in quick sequences named 'saccades'.

So if we stare at a virtual key for more than a second, it should certainly mean that we want to operate it. Our very center of view is extremely narrow. Hence it would only depend on the exactness of the eye tracker to discriminate even small distant buttons from each other.
An eye tracker could be self calibrating, by learning from the user apparently looking slightly beneath a virtual object and intelligently interpreting this as alignment errors.
It's therefore possible to start actions with eye pointing alone, or we could use it in conjunction with voice commands, or with

pressing a key at the simulator itself. So we could reliably act on keys or icons too far away to be touched.

We could use eye pointing to activate cells in a spreadsheet, to move a cursor, to drag windows, and so on. Together with a mouse or mouse pen, it would become incredibly versatile.

Another application could be with situations where we use our hands otherwise, riding a bicycle, a car or an airplane for example or operating complex machinery; it would be necessary of course to design this in a way that avoids any hazards resulting from occasionally looking away from more important things.

Eye pointing is not entirely new, as you probably know:

- A camcorder model once came with an eye tracker in the viewfinder, and it focused to the part of the picture being looked at.
- An experimental display screen once allowed the user to erase any parts of the image just by looking at them (ingenious nonsense, in a way).
- World famous physicist Stephen Hawking uses an eye pointing device to select words from a computer screen.
- Current commercial products can be found at [52], for example. Military applications have been reported [57], and the new Eurofighter actually has an eye operated head up display.

With vision glasses getting into mainstream use, everybody will be able to use eye pointing for many things that we might not even be aware of right now, and it will really be one of the most attractive features of this technology.

Additional components ('Mixed Virtuality')

Virtual writing with a real pen

A special pen with a modulated 'position light' could be used for an easy detection of the user's hand movements by the position sensor cameras. Virtual lines would be drawn when the pen is moved inside the virtual paper plane (this plane would normally better be chosen to fit to a real desktop's surface, or at least to a drawing board, to provide a tactile surface).

Adding a ball to the tip of the pen could allow to write very realistically on any hard surface. The ball could also provide signals about its movements. For example, it could have a surface structure that is detected by sensors, so it could work like an inverted optical mouse. We could also use a fixed tip with a high resolution optical sensor as in a normal optical or laser mouse. If we add a pressure sensor and transmit its output and the ball movements by wireless link, the pen could also work without being seen by the position cameras at all. It would then behave quite similar to a mouse, its pointer being displayed in the virtual window of an application or on the surface of a virtual object.

The virtual surface to operate on would first have to be 'activated', by a virtual key action, or just by looking at, for example.

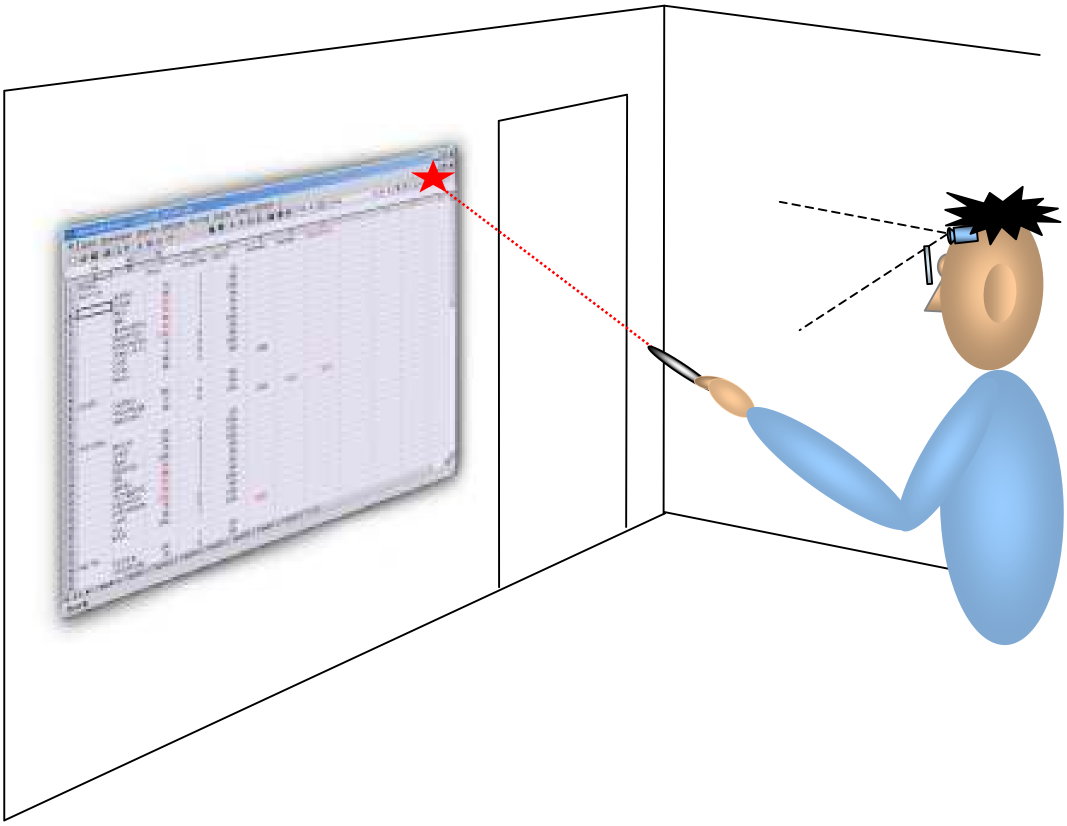
Another interesting add-on to the pen would be acceleration sensors. They could be used to carry out ballistic actions towards distant objects, without the difficulty of following hand or finger movements optically. They could as well be used to enhance the camera based detection in the basic 'into the air' writing example above.

There are many other possibilities for 'mixed virtuality', necessitating other dedicated hardware perhaps, that I will here leave to your imagination.

A laser pen

A pen could become an even more universal input device for a vision simulator. It could be extended with a laser pointer in order to provide a means of interaction with remote objects.

Arranging wall screens for example would be easy and accurate with it. Modulated light could be used to visually identify the pen and its beam with the position sensor cameras. Radio or infrared communications could be used to transmit pressing of the keys that the pen should certainly have.

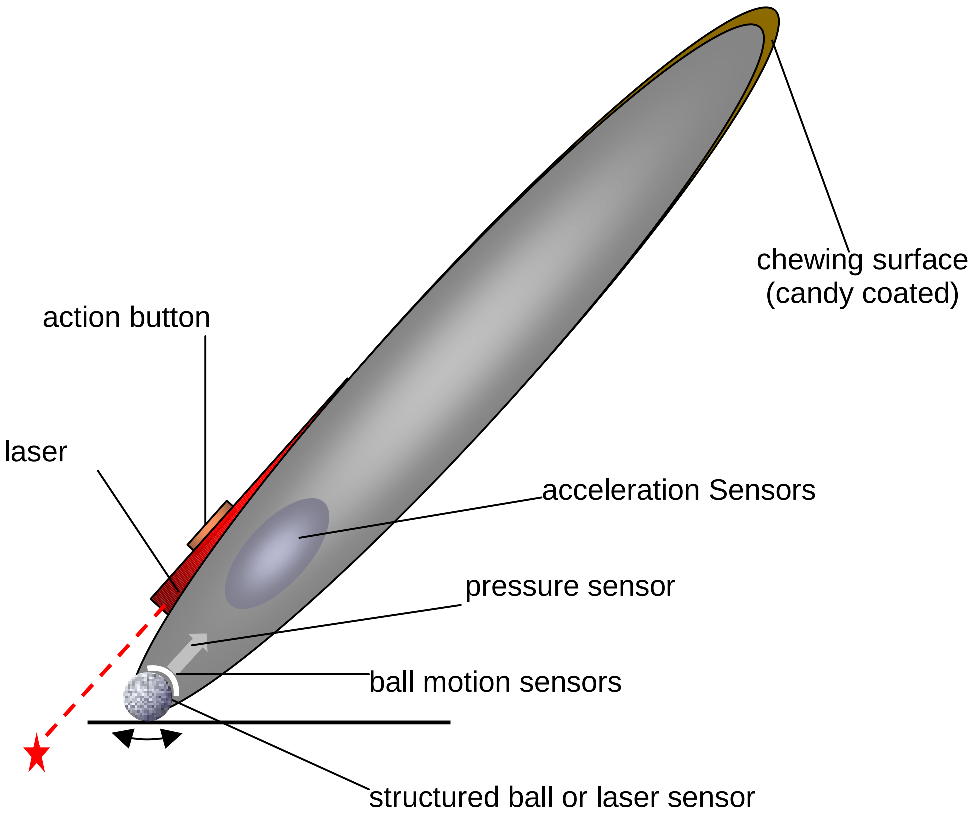


We could use a pen to click on the handle of a virtual window and enter drag mode. Then, by pointing to a wall and activating the laser, the window would automatically be mapped onto the wall. The laser pointer could now be used to manipulate the window from distance. Like a magic wand.

The laser pen works quite accurate and helps the vision simulator by directly showing a light spot at the desired place, revealing its accurate distance even on an unstructured surface.
We could also use finger pointing of course: this would only require first to recognize the finger, by pressing a virtual key for example (if we want to avoid the difficulties of full gesture recognition). Accuracy is a bit low with this approach.

The most exciting alternative for remote control remains Eye pointing of course, perhaps not quite as exact as a laser, exact but absolutely elegant.

The ultimate pen



Full featured pen mouse for use with a vision simulator

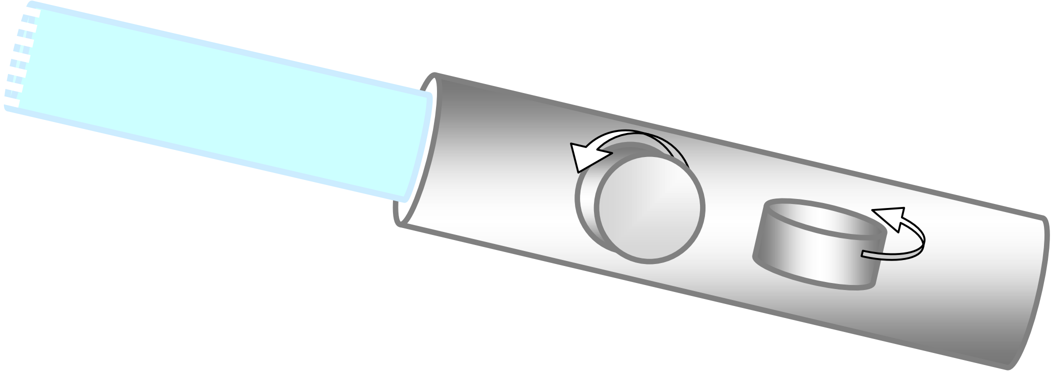
If we use a laser mouse sensor instead of the structured mouse ball, it's clear that it should not glide on its lens, as this would get scratched and blind. We should use a passive ball or round tip or ring with the sensor beneath or in the middle.

It is quite likely that such a pen could replace mice in almost all mobile applications even right now, as it needs not even a straight surface. You could 'write' with it on an armrest, on your leg, into your palm of your hand, just anything.

Maybe we should add an eraser...

Light Swords

I've mentioned light swords in the fiction part, so you may have asked yourself "how should I grip it?".
Obviously, there has to be some dedicated real handle for the virtual light sword. This isn't the actual object concerned in a game, but some universal hardware gadget, like a joystick.
It should have position and acceleration sensors and force feedback. The latter needs some thinking about. We could perhaps use some pulse magnets to kick around weights in the handle in order to simulate impacts.



What would work even better, if we use some rotating weights. These would work like gyros, creating a counteracting force to any rotation. This way, that the grip would feel heavy, just as if a real sword was attached to it, and these fast spinning weights could also generate a quite violent kickback if stopped instantly, simulating hits with the virtual sword (we need two counter rotating weights for each axis, only one is shown in the picture).
Some batteries would also be necessary of course, but such a handle shouldn't be too light anyway.

This handle could also be used for virtual guns, and even some decent applications like virtual gulf clubs, baseball bats, tools.
It could as well be attached to a real desktop with a little suction pad and used as a joystick, force feedback included.
Occasional gamers could of course use the 'ultimate pen' instead of a specialized device, for many kinds of games and more serious applications.

Wireless interfaces

Generally we have to consider that the entire computing power of the visual interface cannot be integrated into the glasses assembly in the near future.

We have to use some pocket computer, or in some cases external computer power of a company, campus or home network.

If we want to externalize computing power from the display interface, we have to transport 2 HDTV channels for the display, 4 normal TV channels for position sensor and eye tracker cameras, possibly 2 more TV channels for mask displays, audio in/output channels, and several sensor channels. It should be possible to transport all signals over some serial digital channel like Firewire or USB2. In this case the question is if any TV or HDTV channels should be compressed.

Encoding and decoding as well as compression/decompression and the interface protocol chips, will add some complexity and power consumption to the display interface, but less than if we add a processor.

Then we could also much easier use a wireless connection. With current varieties like Bluetooth however, bandwidth limitations are still severe. A single TV channel, if we use a low power consumption approach for compression, like MJPEG or DV, needs up to 35 Mbps, HDTV up to 200, while Bluetooth still usually is at 54 Mbps, with some extensions to come.

We could optimize compression a bit further, and wireless interfaces are evolving quickly, so this will surely become an option when any of the basic devices envisioned will be accomplished.

The bottom line currently is, that we may need a power connection from a pocket device for the very first implementations, so connecting to this device for data is straightforwardly done with a serial bus, sharing with power one single thin cable.

As soon as we can include low power compression chips, with glasses assemblies then probably being in the 50...100g range, the required batteries for wireless communications would easily fit in.

Wireless communications need an emission power proportional to the square of the distance spanned, and proportional to the bandwidth required.

So the high data rates we need will certainly contribute to power consumption, while the small distances to a pocket computer require much less power than with most other applications. Only the unsteady conditions of wave propagation near to the body may spoil this picture a bit.

External computing

In certain environments, we could run the vision simulator entirely without a pocket computer.

We could use a local computer network to run everything, including the real time adaptation algorithms and video processing of the vision simulator. This could be done in office or home networks, where the distance to the next wireless node is small (limiting the necessary radio power) and sufficient dedicated computer capacity is available.

We would need some more radio power for this variety, but it's still in a reasonable range.

It's obvious that we need secure encryption for any communications. With external computing, many other security questions arise.

Materials contributing to this kind of applications can be found in [3].

Security

A portable device carrying most of the user's private data in order to be useful in all situations, is very vulnerable to theft.

Notebooks are easily secured by fully encrypting their harddisks, although this is rarely used because most users are plainly unaware of the possibility. Data on such a machine is only accessible as long as the user is logged in and power is on. There is still a problem if a notebook is stolen while the owner is working on it, as the thief could get at everything currently accessible.

With the vision simulator, there are 2 items that could be stolen: The glasses and the pocket computing unit. The glasses do not carry their own data storage, so their theft is of monetary concern only, but we could use them to secure the pocket unit. We would totally encrypt the pocket unit anyway. Apart from quantum computers that won't be available so soon, there is no known method to crack a really good encryption. So this is really secure.

We would also encrypt the communications between the pocket unit and the glasses. It is no problem to make this really secure as well. The algorithm of choice would be working like the well known https* protocol. The only way to crack this is to fully break up the communications and insert a 3rd party device that processes every byte of it before transmitting it to the original destination. This way, encryption keys could be faked and the whole communications could be eavesdropped.

This is not so easy in our case. If we use wireless links between the pocket computer and the glasses, the distance is very small.

An eavesdropper would have to get very close, receive the signal of one part while jamming the other, and retransmit the packets modified afterwards, i.e. with a delay. It is quite obvious that this whole interference scheme is something that would work on a wire, but is utterly difficult to achieve with a wireless link. The unit would only need to watch for delays, radio interference and other anomalies.

---

* The secure internet protocol. Active if your browser shows the padlock icon. Uses public key encryption. This is a method working with key pairs, a public key that can be freely distributed for anyone to encrypt with it, and a private key that also allows to decrypt and that only the originator of the keys possesses. The mail encryption program PGP for example uses this method. With https, machine 1 crates a temporary key pair with a random number generator and sends the public part to machine 2. Machine 2 does likewise and sends its public key to machine 1. Now they can communicate safely in both directions.

We would also use encryption keys that are exchanged in a secure environment only. As this is not a communication with an alien partner as with arbitrary connections in the Internet, we would only use an automated key exchange if we establish the communication for the first time, for example if we replace a broken pair of glasses.

Afterwards, we need not touch the key at all. Once the keys are established, an attacker has no chance to get anything meaningful out of the data transfers [48], [70], [71].

Now if someone stole the pocket unit, he/she could not access it because he/she has no device that could communicate with it. It's like a notebook with no one logged on anytime.

If the thief stole both the pocket unit and the glasses, he could readily access the owner's data as long as the unit is powered. There is a lot of possible damage conceivable, that I won't discuss here in detail.

We could also imagine a thief just taking the glasses and using them while threatening the owner with a gun.

What we now want to have is something that prohibits the glasses to be used by unauthorized people, and also if they or the pocket unit are taken away.

There is one very elegant possibility to do this: Using the eye tracker for an iris analysis. This could work continuously and safe. As soon as the glasses are taken off, or taken by someone else, they stop working.

Another possibility to achieve this would be an implanted RFID like chip. This could be as small as 1 mm and would only work for distances of some mm. Not actually something anyone would like, I guess.

Anyway, the glasses could protect themselves from displacement.

Other types of attack may be more promising: If we know the inner workings of the display (quite probable if they are once produced by the millions), we could try to tune a radio receiver to some harmonics of the line frequency and exploit the modulations that are probably occuring.

The difficulty of such an attack can be increased by optimizing the wiring, lowering the power dissipation of the circuitry (useful anyway), shielding the device (a weight problem), randomly modulating the display line rate, or by actively producing interference noise to make the signals unusable.

With LED like displays, some form of eavesdropping could be just to look at somebody's glasses. Inevitably, some light would leak out, and the display would be quite 'large', some cm maybe.

So one could eventually catch a glimpse of what's going on in the display of their encounter, but that's not really a big issue.

With laser displays, one could eventually see a large projection of leaking light on the floor, even though very dim. In case we use color selective mirrors, that leakage would be so low that one couldn't likely exploit this.

The projector could theoretically be powerful enough to provide a good picture on a sheet of paper in a dark room, but normally the intensity would always be tuned to match the brightness of the environment, so this wouldn't work in practice.

The usual appearance of a laser display for other people would be a dim and very tiny dot of light right from the projection mirror.

With dirty glasses, one could probably see an image right in them. It's hard to figure out how large that problem could actually turn out to be, as it is also very dependent on the technology used.

With holographic projection at last, there's a good chance that this would be very safe. Eavesdropping several GHz of holographic data going to the display would be very difficult. Seeing something in the tiny displays from outside would as well be close to impossible, even more so if we also use holographic mirror glasses.

Secure private realm technology

For data security, we would use a pocket device for all processing but do dynamic backups on a remote server.

In this case it's advisable to encrypt anything before it goes out.

Incremental backups save bandwidth and can be done fully encrypted if all files are encrypted separately. The external service should not do anything but to provide storage. All control over it

should remain in the remote unit. The algorithm should enable restoration of all data by knowledge of password and possession of the appropriate software only.
This approach would remain unbreakable if the password is long and secure enough and always kept safely.

A secure remote data storage facility as envisioned is not fiction any more. First applications of this kind have become available quite recently, for use with portable computers [6].
I therefore do not consider this a technology specific to vision simulators. Some special problems may have to be addressed, but in general, current developments will certainly cover most of our special requirements.

We have discussed a possible attack scheme for communications between glasses and pocket unit. There a protection was given by the fact that it would be difficult to interfere with a very narrow and low power communication link. With the backup scheme one might think an interference could get easier, because of the larger distance between links, and because of the many network nodes possibly involved.
Yet this isn't the case, as the data never need to be decrypted. The backup server needs no key. The key entirely remains with the user, and anything entering the web is useless without it.
So the remote backup unit remains nothing but a storage, and there simply is no attack scheme for this protocol, other than hijacking the user's computer. Even if someone would alter the data stored, the result could be nothing but garbage, annoying but not a security threat.

If someone would set up a treacherous hotspot that appeared as if it was a trustworthy outlet but in fact would be the spider web of some cybercrooks, they could possibly interfere with a https session and do any kind of damage, but that's a quite common threat to everybody just walking around with a conventional notebook and a WLAN card and has nothing to do with the dynamic backup scheme discussed above.

Public objects

To provide publicly visible objects, it is necessary to announce their position and nature in a way that nearby computers can receive these announcements.

A viable solution would be to use a radio channel and transmit data packets with the basic informations continuously with a CSMA or CSMA-CD* protocol.

These transmissions need only provide position and nature of the objects. 'Nature' means type, basic information about size, importance, relation. If any nearby receiver decides that an object could be of interest, it should transmit a request, causing the originating machine to send further information (if the request is accepted), containing the image data of the object as well as shape and imaging information about nearby visible structures relative to which the object should be located. This precision locator information is necessary because GPS alone would not allow for a precise positioning.

Objects that should integrate and relate to real environment need to be positioned with an accuracy of millimeters that couldn't even be provided by differential GPS or other radio based methods.

The process somehow resembles web browsing, where the initial packet selection would equal the use of a search engine. The search in our case is also refined physically, just by the proximity of radio sources that can be received. In the next step, a http or similar connection to the web server, being the public node, is established, and the page, being the virtual object, is requested. The browser window in this case is the real environment.

---

* Carrier Sense Multiple Access with Collision Detection: A node 'listens' if the channel is free, then starts to send a message packet. Collision detection means that if 2 nodes started simultaneously, they can detect it from signal interference. In this case they stop and 'back up' (delay for a random time) before retrying. Many nodes can share such a channel without much administration. Ethernet (your computer's network card) works this way [39]. Wireless varieties often don't detect collisions [43]. They detect problems by missing receipt messages or the like, then randomly retry to send.

Content, as in the web example, can be static or animated, also interactive. The user could perhaps be allowed to manipulate it by hand, eyes, laser pointer, voice, virtual keyboards, or others.
With such a networking approach, where nodes randomly appear and disappear, only a probabilistic access algorithm would be appropriate of course. Contrary to a general misconception, these approaches are more reliable than deterministic ones even in real time environments, due to their inherent robustness [39],[43],[78],[79] (the delusion that only deterministic protocols could ensure reliable real time behavior, hitherto led to many peculiar developments, e.g. with industrial and car networks).

Shared objects

The same way public sources can create and distribute Objects, any vision simulator can provide for this as well.
Such objects could be abstract things like hard disk files, or they could be currently displayed objects or windows. We could as well envision that somebody lets his computer generate an animated replacement for his own person, a feature that could be used for gaming or virtual theatre play. Shared objects also comprise any kind of virtual devices belonging to real machinery, like the control panel of a washing machine, the instrument board of a car or a plane, input panels for home or office installations, virtual ambience objects like virtual window views, decorations or even virtual wallpapers belonging to certain rooms. Virtual Christmas decorations, or any kind of signs also belong to the category of privately published virtual objects.
Access rights to such objects have to be defined, as even objects in outdoor areas are not necessarily meant for everybody (a sign pointing to a phone switchbox for example could be confined to service people), and inner room decorations should not be accessible to people outside, even if theoretically they could configure their vision sims to see them right trough the wall. Object descriptions should therefore use encrypted protocols in many cases. What can be different is the way keys are handed out.

Interior objects could for example use an infrared link to distribute a key, hence anybody in the room would always have access, and these objects would behave just exactly like real ones, i.e. could be seen through windows but not through the walls.

Restricted objects

Public objects could easily clutter up public areas for anybody using a vision simulator. Some should necessarily be seen by anybody, like temporary traffic signs. Some could be considered harmless and therefore allowed for public distribution, like window dressings or posters. As virtual objects could have some disturbing properties, like being heavily animated, being too bright, sticking out from the wall (in case of 3D placards for example), even those usually harmless objects should comply to certain restrictions. Much worse would be the generation of virtual objects right on the road, the display of creatures attacking persons, and so on. Hard even to imagine all possible types of hazards that could arise from public virtual objects. They could not physically attack anybody, but nevertheless distract so much as to cause serious accidents. Any public object must therefore either comply to certain restrictive rules about its position and nature, or have a publicly controlled certificate, or the digital signature of a public authority, to be present in security relevant areas. Individual vision sims also have to be programmed in a way to reject objects if they would have the potential to seriously distract the user. A useful virtual object on a road for example could be a virtual road sign marking a recently occurred accident. It could even be allowed for private people to generate these, for example in conjunction with a car's hazard lights (where the object could just be automatically generated by the car itself).

Another handy possibility would be cars transmitting their position data like with the airplane virtual radar mentioned, so they could be 'seen' behind bends. These applications of course need a very secure structure of certificates to defy abuse.

Sound

It is quite obvious that a non immersive display has to be accompanied by 'open' headphones, that allow outside sounds to enter without much attenuation.

Button headphones would normally be the technology of choice.

Mounting such headphones can simply be done by attaching them to the spectacle's ear handles with flexible wires. This is surely better than just squeezing them in.

Reproducing surround sound with these earphones (as with any) requires some signal processing. This function can be found in today's sound cards as well as in some soft or maybe even desktop DVD players. It's state of the art. What is done is to emulate an artificial head recording, one that uses an original human anatomy to reproduce exactly the sound waves getting into the ear at the original site.

So far nothing special. What I would add is a dynamic sound field simulation that compensates for the user's head movements. So as with vision simulated scenes, it wouldn't be necessary any more to stay in place or seat to enjoy the full 3D impression. Even this will be something that could easily be addressed with today's technology. CPU's hardly take notice doing that real time.

We would also need microphones for this design, be it for speech steering or in order to use the vision simulator as a mobile phone or as a camcorder.

As stated already, a reasonable implementation would be to attach these microphones directly to the earphones, back to back, earphones pointing towards the ear channel and microphones outwards.

Using this device as a 'natural head microphone', of course has the disadvantage that it needs the head to be fixed for a professional result. Some sound processing however could greatly compensate for head movements, and we could also convert such a sound recording to multi channel if we desire.

In case of speech recording, a microphone at the ear might not be what's currently considered the best way. A microphone closer to the mouth (although still beneath it, to avoid breathe noises), is more usual. Yet we have 2 microphones, and using their output for some intelligent signal discrimination could provide for a very good separation of the owner's speech from environmental noise. So we could probably well do without any separate speech microphone.

Last but not least, let's mention again that we could add active noise cancellation, by feeding an inverted signal from the microphone to the headphone, to just compensate incoming noise with an actively generated signal of exactly opposite amplitude. This technique is quite common in aviation, and could be useful anywhere, if anyone had a sufficiently equipped vision simulator. We could also do the opposite. If button headphones stick too close (normally I wouldn't like that all the time, but it's sometimes useful to produce good bass), then we could avoid the user being closed out from environmental sound by just sending some of the microphone signals to the headphones.

Very recently, a new technology has emerged that can pick up the user's voice from the ear's auditory canal [63]. The sound comes from the oral cavity through the Eustachian tube. Even whispering can be picked up this way, in presence of environmental noise. As the earpiece has to fit entirely tight in this case and any hearing can only occur indirectly through microphones, this is something I wouldn't like for everyday use. It's very interesting for professional applications of course.

As stated in the introduction already, a very important software feature of a vision simulator will have to be spatial sound impressions complying with the pictures of virtual objects.

In order to support a realistic impression for any simulation or media reproduction, sound has to come from the visible sources.

As the user can move its head against these sources, classical surround sound has to be converted to the true angle and timing, and the characteristic influence of the user's head also has to be simulated (synthetic artificial head stereophony). The operations necessary can be found in any current soundcard already, maybe not exactly fit to our requirements, but it's state of the art, so let's leave this to other literature (e.g.,[73]).

We'll get back to sound once more in the media chapter.

Conclusion

I have tried to address the main technological aspects of vision simulators that should be able to generate really perfect virtual devices and objects.
Again, many of the ideas are still speculative, and what I have tried here is not only an overview of ideas, but also in some aspects a proof of concept (as far as possible in theory), i.e. to show that it should be achievable in a foreseeable time.

As to the status of technology, this can only be a snapshot, as new ideas and technologies keep arising every day.

I did not address sound too much, as this is all more or less state of the art and so much less difficult than the optical matters.

Although the difficulties of some approaches may seem to be overwhelming, and some are so complex that they also require large multi threaded research projects, we should not forget that this is a technology that will not only replace many others, but also have a huge economical perspective. Substantial efforts are therefore well justified.

In the following chapter, I will address some aspects of a very large application area having its own requirements, i.e. (virtual) media technology.
Here we have to think about appropriate 3D recording technology as well as about the relation between vision simulators and conventional screen displays (that will still play a role in some areas). So I will include some thoughts about 3D screen technology.
I will also use this theme to look into some basics of stereo pictures, hopefully useful for readers not too familiar with it, also better to understand some related topics in the preceding chapters.

Before imagination streaming, when media still needed hardware, possibilities were closely limited by available technology. This is a paper considering media technology at the edge of a vital intermediate stage called virtualization.

part 4: Virtual Media

Introduction

We have seen that a universal sensorial interface will allow for new media formats, greatly surpassing anything possible with classical screens.
Perfect 3D video is but one possibility, something hardly achievable with screens in a foreseeable future, in any acceptable or broadly affordable manner.
Producing truly three dimensional, or even screen independent media formats is anything but trivial already. We will see that we need object oriented or synthetic holography recording, because any user perspective has to be individually generated for each person, during playback, in real time.

Large conventional screens, unable to match the 3D capabilities of the vision simulator, will nevertheless stay around for a long time. Not because the current monster varieties of 'flat' screens are so good, but because they will be replaced by much better ones.

Organic (OLED) screens will quite likely be cheaply produced in large sizes and huge resolutions, probably flexible as well and mountable on any surface. Projection screens will be greatly improved as well, by the introduction of holographic screens that are direction selective and offer the same contrast as any self luminating versions, or even better.
Both varieties mentioned are thin plastic foils, promising a large potential for cost reduction with mass production, and they offer a much better energy efficiency than anything we have now.

So we will have to discuss which applications will probably run on a vision simulator and which will remain on real screens. Most likely, real screen applications will always have restricted 3D capabilities. The vision simulator is much better for this.

We will further regard the possibilities and problems of media production methods that could unleash all of these capabilities.

A large impulse for the development of new recording technologies comes from the increasing popularity of synthetic animation. Practically everyone could now use the 'levels editor' of a computer game to produce synthetic movies ('machinima'). Software for professional productions is many times more sophisticated of course.

What is inherent to all these productions, that anything is originally object oriented and fully three dimensional. Viewer perspectives have to be set by defining the 'camera positions', and only after this, a classical 2-dimensional movie emerges from the setup.

With these computer generated movies, producing for only one picture channel will, in almost any case, just be equivalent to throwing away information. Recording of multiple virtual camera channels would be the minimum requirement, to enable future reconstruction of 3D information.

Combining these virtual scenes with real ones is currently very difficult. At least it involves blue (or green) screen technique, where a single colored background is used to cut out the actors from it. First attempts to avoid all this and get the entire 3 dimensional scene just by stereo cameras are under way, but the difficulty is large.

Even the correct animation of virtual characters is still a problem. You may already have seen actors running around in black suits covered with white spots. These spots help the cameras to catch their movements in a way that can then be used to perfect the animations of a synthetic movie machine.

I'm nevertheless quite confident that an entirely three dimensional and object oriented production technology, perhaps with the help of additional 3D scanning equipment, laser scanners for example, will be accomplished in a way that could be routinely used even in less expensive projects.

What would basically be necessary is to record any scene with several cameras, in order to gather the raw material for any immediate or future 3D processing.

With audio, this has been a common practice for decades, not because surround or even simple stereo were already there, but because single instruments and voices were recorded separately and mixed down later on. Now the master tapes of these recording allow to produce surround sound with them, completely unforeseen at the time they were recorded. The original sources can be re-mixed to any format desired, allowing to exploit the same assets again and again.

So even though we might not be able to make full use of it right now, recording with multiple camera arrays may be a good idea anyway.
For people still using a 70 mm movie camera this may sound lunatic, but HDTV cameras are getting very affordable already and even larger formats are only a matter of pretty short time. Cameras are steadily getting smaller and also less expensive. It may be possible to upscale resolution by correlating pictures from several cameras in an array, so each of them would become even cheaper (more on this later). Recording will also no longer be done on tape but directly and routinely on the computer harddisk, where the data streams even of many parallel cameras are not a big deal any more.
Together with some adapted editing software, this won't be more difficult to use than any of the current systems.

What is necessary here is a standard format to record information about camera positions, focus length (for correct perspective), and also additional data from dedicated 3D scanners. We would also need a standard format for video compression (encoding) and a standard media format for permanent master storage.
Regrettably, no format has yet been standardized for the recording of camera position information, even though some proposals for this have been made.

Encoding is something special here, as pictures from different perspective also have large common parts that can be exploited to reduce data, just by encoding a 2D base image, then enhancing it with additional data for hidden surfaces, depth and so on. With such an approach, the total data volume may be reduced to a little more than that of a single camera. The advantage would be that we could also transmit this entire information to any viewer, where it could optionally be used by high end display equipment to produce any perspective according to the viewer's individual position.

Quite recently, experiments were carried out to encode multiple camera channels in the context of advancing the MPEG4 standard [9]. Neither time- nor spatial redundancy reduction however led to any significant reduction in bandwidth compared to separated channels.

These results are contrary to any reasonable expectations and basically they allow only one conclusion, that optimal encoding algorithms have not yet been developed. The experiments mentioned still are in a very early stage. I hope that this research will be continued. In any case, it would now be necessary to include the new compression standard AVC. This could also eventually lead to some improvements with stereo image encoding.

Already foreseen in the current standards, although not implemented except for research prototypes, is the description of three dimensional objects by a triangular mesh.

It may well be that the entire approach to extend current encoding technology by depth or surface information is not optimally suited for the purpose. It is just like doing collages with paper and scissors, while real objects can be transparent or foggy. I will later on describe an holographic approach. Holograms are just an optimal encoding of anything that light can show, therefore possibly much better suited for the task.

Standardization of certain frame formats or even certain camera array assemblies is not so important. At times of analog TV channels, even conversion of different image sizes has been a problem. Today we could almost do away with standards, at least these, because modern graphic chips do these conversions on the fly, all the time, without anybody even taking notice. Likewise, object oriented formats would leave the actual display arrangement entirely to the disposition of the viewer, restricted by his technical equipment only.

Storing scenes with object descriptions or other appropriate structural information would later on not only enable us to freely select the reproduction technology, but also to seamlessly integrate artificial objects and scenes. It would also allow for a universal manipulation of content, i.e. we will get additional degrees of freedom so far only available in pure computer animations.

At last, the objects displayed could be made interactive and reacting. This way, we would get a floating transgression from film to computer game to virtual device.

For this reason, concepts of augmented reality and of virtual media are essentially the same.

With hybrid projects, containing computer generated as well as natural scenes (hence, with any 'better' movie even nowadays), we see that such a technology would greatly simplify the integration of the different content. This should provide a strong motivation for further research in this field.

We could henceforward conclude, that the recording of three dimensional image information will become of fundamental significance in the foreseeable future and that it is absolutely urgent to develop standards for multi channel image recording as soon as possible.

In the following pages, I will try to systematically describe the methods outlined, and give some overview of the technologies involved.

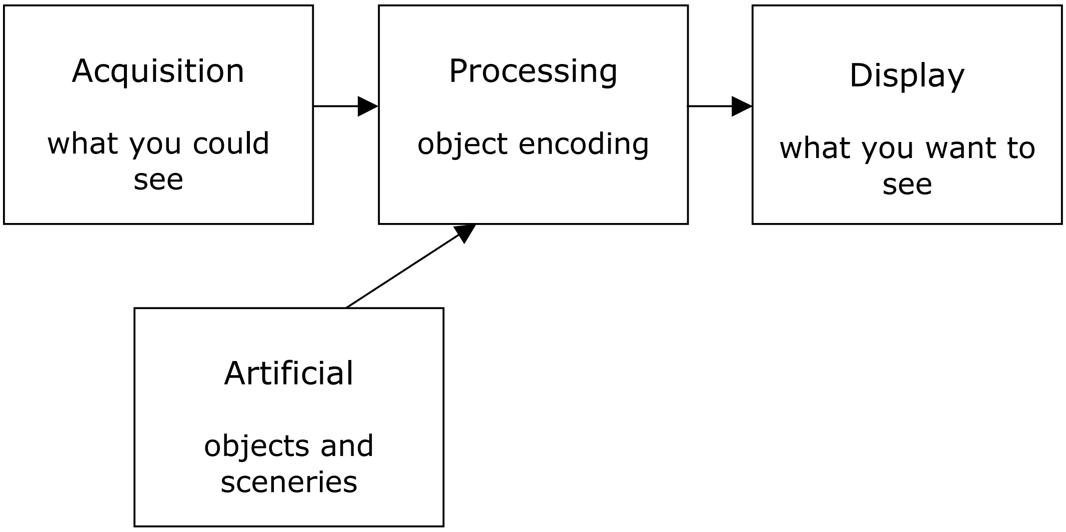
This may also be useful as background information for some things that I could only scratch in the preceding chapters.

In addition to an assessment of possible technologies, we also have to try some predictions about the viewing habits that might emerge with new technologies.

The 3 step paradigm – hardware independent media

The entire process of recording, processing, storage, and intelligent reproduction can basically be parted into 3 steps:

1) Acquisition: Transferring reality (and artificial resources) into data storage.
2) Processing: Anything from signal processing to cutting and directing.
3) Playback: Transporting the finished product into the human brain (via eyes, ears etc.).



Step 1) would in conventional terms mean shooting raw scenes or recording live music. It could also mean production of cartoons, for example.

In this paper, I will concentrate on the technical process of recording *reality*.

The central approach is to record exactly those signals that a human observer would acquire in the real scene. It is, in principle, based on conventional optimization of recording equipment

(avoiding any effort to record more than what meets the eye and ear but doing so perfectly and without redundancy), but I'll try to systematize this approach towards the extensive use of modern sensors and computing power to achieve a scene acquisition as complete as possible, including holography like pictures and sounds.
You say this may sound good but could never be achieved ? Just read on.

Step 2) would mean cutting etc. in conventional film or music production.

Here, the focus is on the technological process of conversion from crude recording signals into computer based reality models.

The idea is to use sensors as simple as possible, to extract the necessary data, to enhance the data where components may be flawed or missing due to limitations of the sensors, and to use reality and human sense models to achieve an optimal data storage and transport, including highest level and quality compression.

At this step, artificial (entirely computer generated) scenes would also be inserted.

Step 3) in this approach comprises the optimal preparation of data for a diversity of available playback devices (displays, speakers etc.).

Computer power is used to generate individually tailored signals, for example driving stereo glasses as well as pseudo holographic displays, any number of speakers according to the specific design of any playback assembly, synthesizing artificial head stereophony for in-ear-phones, or simply optimizing the picture signal for a given projector.

In this part, our focus here is on the evaluation and suggestion of reproduction devices best suited to the computer based media processing concept.

In general, I suppose that computers will extensively be used to make recording, transport/storage and playback as equipment independent as possible, the only central standard being the encoding, which should be an object oriented approach based on knowledge about the requirements and limitations of human perception.

In part, such a development has already occurred, for example in the fact that modern equipment gets more and more flexible regarding image resolution. In earlier days, resolution was hard-wired in the equipment, and all devices had to conform to standards. Now, low and high resolution pictures can be displayed on any kind of screen, because the necessary conversions have become so easy. HDTV enters the living room with only minor requirement for standards, as many of the devices involved can be scaled to different resolutions, image rates, etc.

Signal formats have also become flexible. With analog equipment, minor changes would have made everything incompatible. Today, Video playback devices master several compression and media formats, and more could be implemented by software, if desired.
We could anticipate that a multitude of formats could be used, that come with a description of their properties in a common language to program any device for their proper playback.

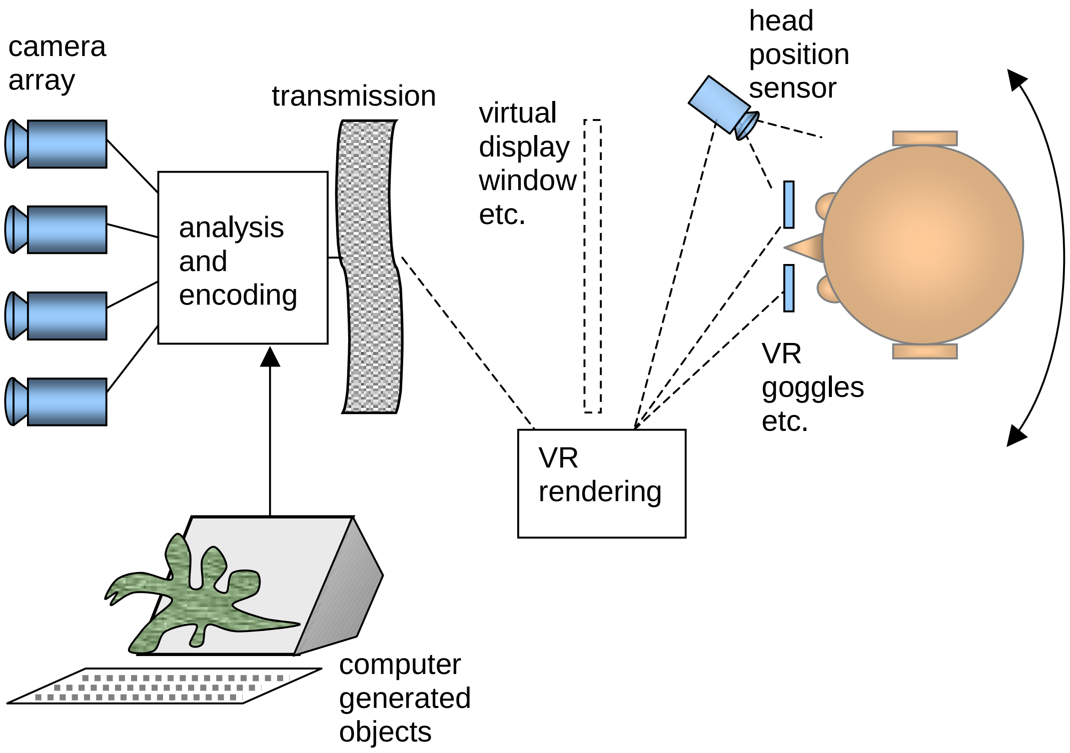
The MPEG4 and MPEG7 standards [80] are steps towards this goal, introducing image and media objects, not just image compression. The aim is to resolve recorded scenes into objects, which are first described by shape, then texture (involving conventional DCT* compression).

---

* Discrete cosine transform, used to encode 2-dimensional pixel arrays

For 3D, we would then add material properties, in order to be able to reproduce their appearances from different viewing angles, also such that are not identical to that of a real recording camera.
As stated already, it is quite questionable if the current mainstream approach is the ideal one. Holographic encoding may in the end prove to be better for many purposes.

The signal chain sketched at the beginning could for example look like this:

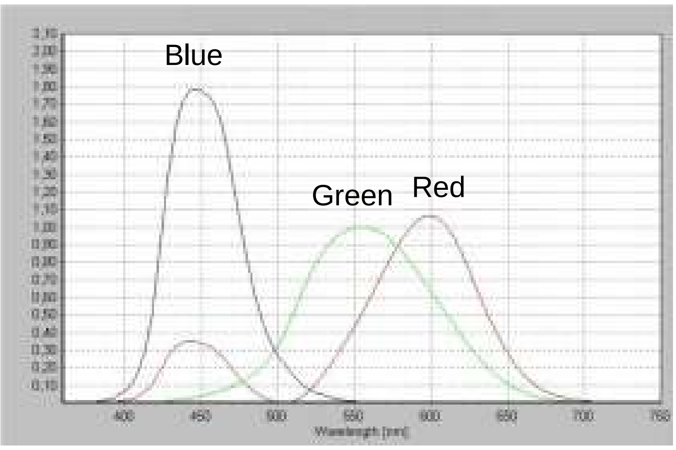


Here we have a camera array delivering - together with a signal processor - a 3D description of a scene, which is then encoded for recording and transmission.

At the receiver side, a virtual reality rendering processor together with a head position sensor delivers pictures to display glasses that emulate a virtual holographic display window, for example.

An example for step 1
(perfect acquisition): color

The human eye has 3 separate types of color sensors: One each for 'red', 'green' and 'blue'. Blue is at about a wavelength of <450nm (0.45 µm), green at ≈530nm and red at >600nm. The sensitivity curves of these receptors are not well separated. For many spectral colors, 2 or all 3 sensor types are reacting, although with different sensitivity. The Picture below shows how far the sensitivity curves are overlapping. It is obvious that the human brain does a pretty good job on signal processing here, because we are able to differentiate millions of colors, from these three only vaguely discriminated signals.



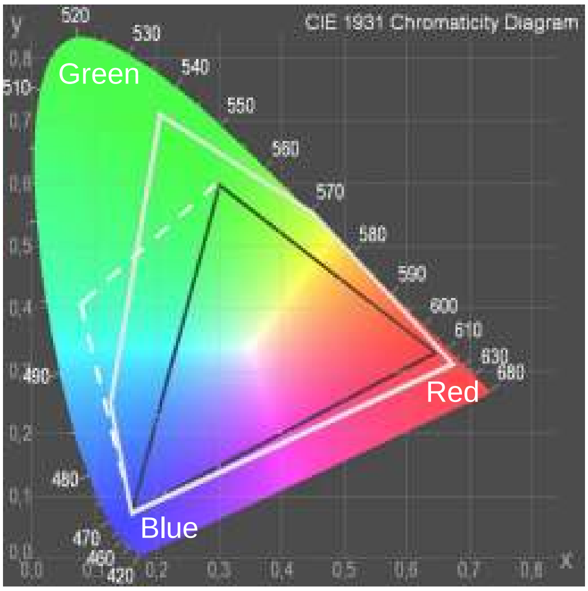
A modern professional TV camera has filter characteristics exactly corresponding to those of the human eye (picture). The 3 color outputs are then sent over a matrix circuit that recalculates them into three standardized R, G, and B color values to match the dyes of standard displays. This may create an additional overlapping and reduce the color space a bit, but with smart signal processing it has already been shown that this process is reversible.

As the color curves exactly match the human eye characteristics, they contain all information that could be visible. Color *recording* is already optimal with today's equipment !

But how does this hold against the fact that no current display can reproduce all possible colors?

The answer is simple: Most Displays have only 3 standardized primary colors, so we get a problem with fidelity, because any primary color from a display is again intermixed with the others by the viewing process, so we just can't address either of the 3 sensor groups *separately*.

Perfect color reproduction



This diagram shows the color space of the human eye (the outer elliptical curve) and the area of reproducible colors for TV (black triangle). The corners of this triangle are the color locations of the three standard dyes for displays. All reproducible colors lie in the area between these three locations. We see that these dyes are not all monochromatic (spectral): Otherwise they would be located on the curved border of the diagram. We also see that especially green here is a mixed color, and that there is a large deficit especially in the range of blue-green (cyan).

If we add some more colors, we can produce a larger area from the theoretical range, because we can simulate the original color much better.

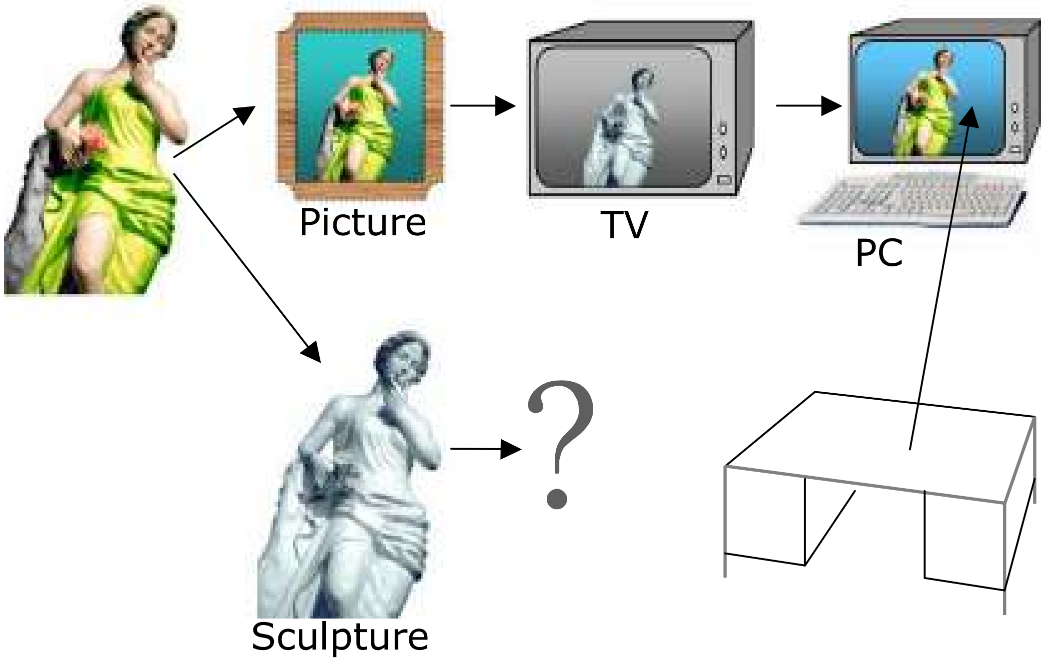
It is almost impossible to extend the range for extreme red or blue very far, because the eye's sensitivity there becomes very weak. In the green and cyan range, a greater improvement can be achieved, yielding a much more natural reproduction especially of landscapes (plants, sky, sea). Printing media already use 6 or more colors.

Recently, an Israeli company (Genoa) developed a 5 color system that shall be marketed in a high end projection TV together with Philips (gray line).

Also just recently, Eizo announced a computer display that uses an additional Cyan color (dotted white line) and is mainly targeted towards professional publishing and printing applications.

We see that any older video production could be reproduced with a color quality that only depends on the equipment. So color already conforms to the 3 step approach.

Evolution of 2D and 3D paradigms



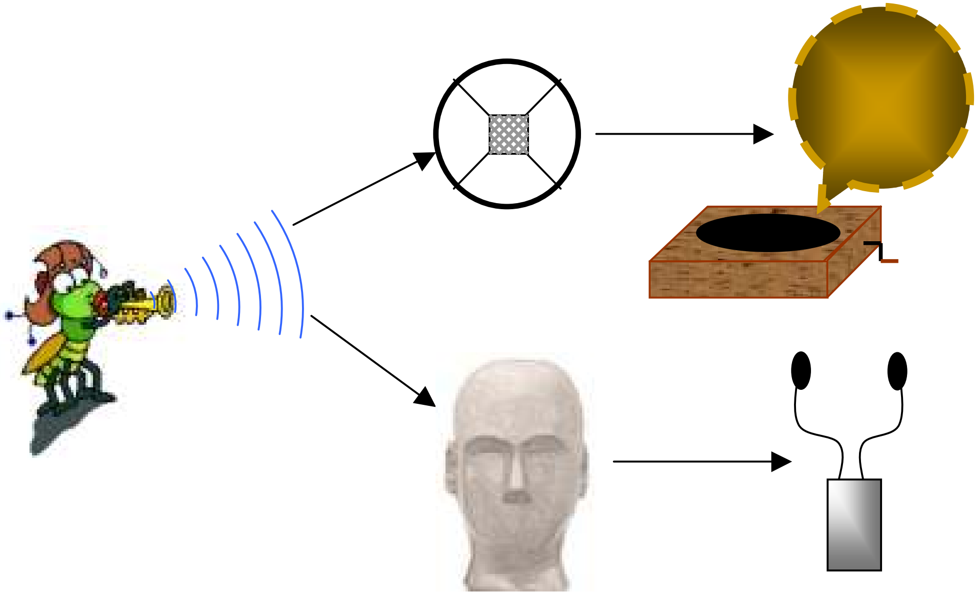
Let's add a little more systematics to see where the gaps are:

With visual media, traditional forms were painting and sculpture, representing 2D and 3D reproduction. In 2D, electronic technology led to the TV and then without any major change, to the usual PC screen. Natural environments like a desktop have been translated to a 2D screen representation, with all objects encased in that screen.

The classical 3D form (sculpture) has hardly any electronic counterpart until now, except for experimental augmented reality setups. Stereoscopic reproduction always lacks one or the other part of reality, be it dynamic perspective or focus. Also in all cases but VR , the object is always confined to a window.

Abandoning the window is useful for exhibitions (virtual sculptures), but also for certain kinds of entertainment media.
Characters in 3D real world games would also be of this variety.

Sound



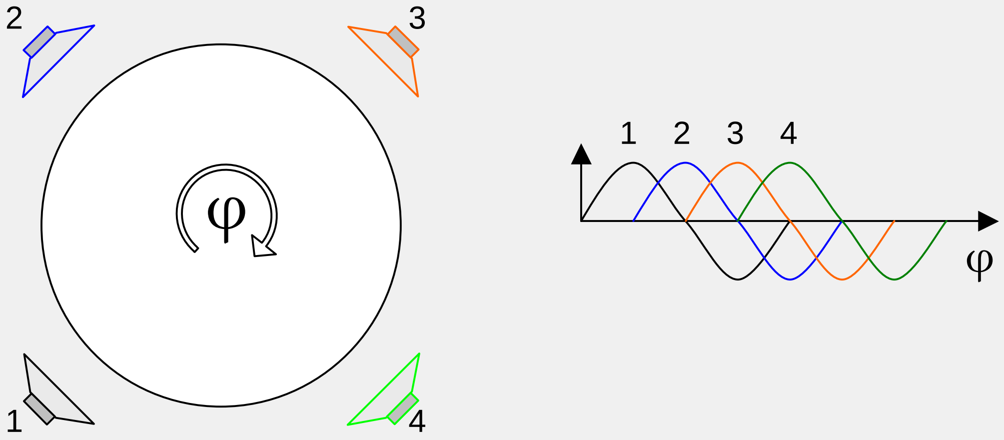
With sound, anything is much easier, not only because less processing power is required. The traditional monaural representation (microphone, gramophone or radio), could easily be expanded to a perfect 3D impression with an artificial head microphone and stereo headphones.

In the design part, some technologies suitable for the vision simulator were mentioned already. Talking about media, we'll just have to figure out what could become usual procedures with the assumed existence of the technology.
What could be a really new experience, with dynamic artificial head simulation, to go around the musicians in a virtual surround video recorded orchestra, being able to hear every instrument really separately.
Acoustic reflections in a room can be simulated quite easily. Current sound processors can even calculate multiple reflections in complicated environments in real time. This can actually be achieved with a PC soundcard and standard software.
Multi channel sound recording will remain standard in media production. Artificial head remained a niche technology for a reason, as it restricts proper playback to headphones in traditional technology and delivers much less directional separation

even with current post processing methods. It is also ill fitted to usual soundtrack production methodology.

Encoding multi channel sound has just recently seen a big progress with the introduction of mp3 surround [32]. Before, it had been usual to encode each channel (almost) separately, although multiple channels carry a great deal of redundancy. Now they need only little more bandwidth than a single one.

Historic example: Phase encoded surround



Imagine 4 speakers (1,2,3,4,) arranged in certain angles around one listener.

We encode the 4 signals by adding all of them into channel A, and adding them with different phase shifts in channel B (this example is taken as simple as possible, so it is not identical to real surround systems):

If we want to decode a certain speaker channel, we add A and B with a certain phase shift that equals the direction:

The signal of the opposite speaker becomes zero, the signals of the neighboring speakers get through with about –3 dB (70%).

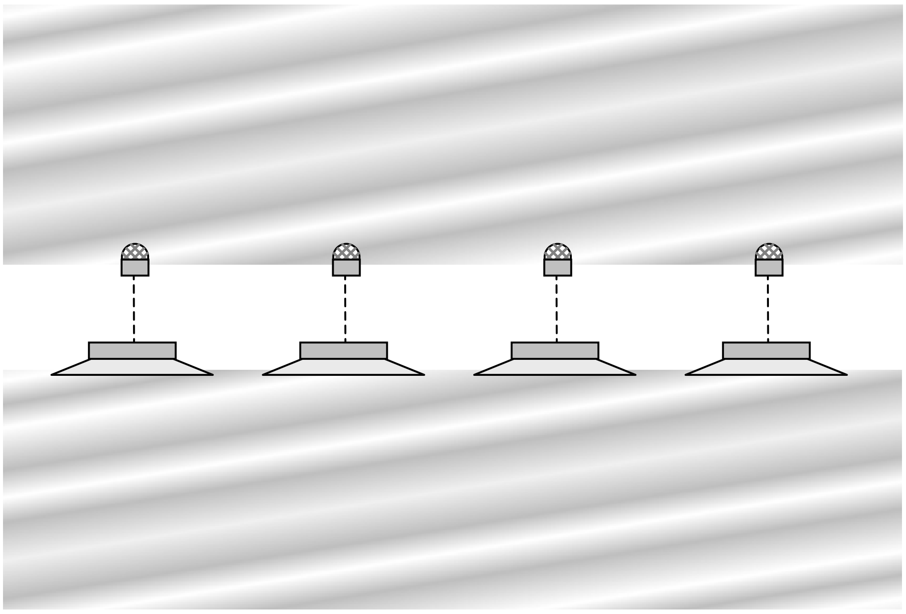
Hence, signal separation is not very good, but the system allows to decode signals for any speaker direction we could think of, so the number and arrangement of playback channels become independent of the recording situation.

All usual multi channel technology now relies on horizontal separation only. The more flexible usage habits that we may develop with vision simulators, could raise the interest in a really three dimensional sound recording.

Obviously, sound already carries more spatial information than picture. Simple screen formats come with surround and 3D sound

effects. In future, image technology will hopefully be able to catch up.
What should be developed, are flexible encoding technologies that allow to chose between several formats for different type of media. It may for example be useful to transmit virtual objects with their own sound channel.

A virtual orchestra could in the extreme case come with one channel per instrument and could then be walked around and viewed and listened to from all directions even with the acoustics of any listening room desired.
In this case we could as well consider some kind of holographic sound description that would cover the entire sound field around the orchestra.
This sounds more difficult than it is, as we only need a real or virtual mesh of microphones arranged on a sphere, and to encode their signals with a redundancy reducing algorithm.



A nearly holographic effect can be achieved even with a modest number of sound channels. For example, if we record with microphones at one wall on one room and reproduce with a speaker array over an entire wall of a second room, these rooms virtually melt into one.

With speakers, the acoustics of the listening room will overlay with those of the recording environment, so usually we only get something realistic from the most unrealistic of all recording situations, the multi channel studio recording and mixing of separate instruments (which is the most common recording method not at last for just this reason).

If we virtually link two rooms with a holographic sound wall as in the picture, this is not a problem because we would just expect the resulting virtually merged room to have both acoustical influences.

Speakers can also do things that earphones can't: shaking the ground, or the stomachs of the audience. This could be a reason to install real speakers.

The reproduction equipment we would prefer in most cases is much simpler: button headphones, nothing else. Due to the almost direct coupling to the ear drum, they can reproduce any frequency free of artifacts and phase or pulse errors, and these devices have also reached a very high degree of perfection over a long space of time. Their only disadvantage: if they do not fit tightly to the ear they can't deliver full bass because of acoustic shortcut. If they fit tight, they are inconvenient and attenuate natural sound.

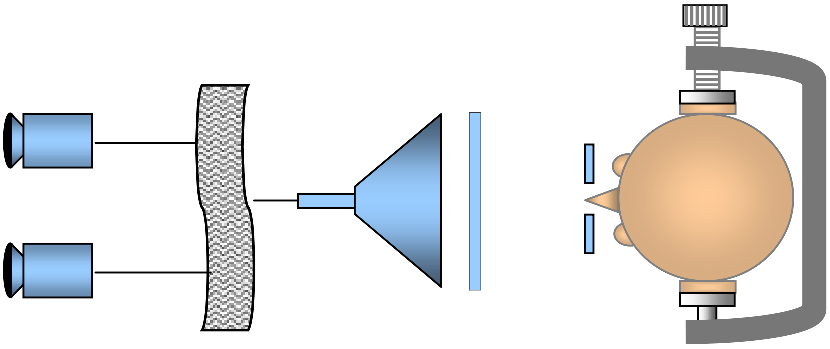
Conveniently wearing these things is normally impossible but not really the problem if we mount them flexibly with the glasses handles. Then they will sit loose however, and dealing with different levels of acoustic shortcut requires a little effort. Maybe an adaptive bass boosting could help here.

Otherwise if the phones fit too tight, natural sound could be passed through with help of the microphones that we would also install attached to them.

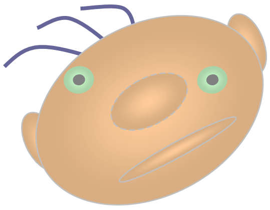
This is all quite simple. We could be glad if matters were so easy with *visual* media. There it's utterly difficult instead, but the sound example tells us what we should have: a display that is the optical equivalent of button headphones, i.e. that is as close to the eye as possible and just projects an image right into it, without any intermediate influences.

Perspective

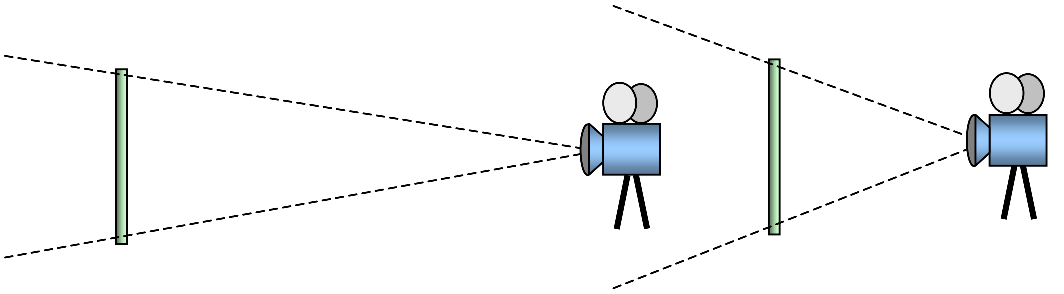
For completeness, I will shortly recapitulate some basics of stereo viewing and display. The experienced reader may skip this.



One problem with the conventional stereoscopic techniques described so far, is that in order to see the picture correctly, the viewer must literally have his head fixed in a screw mount.
The graphic shows a typical assembly with stereo cameras, a CRT display with a light shutter in front, and the viewer wearing shutter glasses. For a correct reproduction, display size and distance cannot be varied, zooming is almost impossible without very disturbing false perspectives, etc. Everything is defined by the recording equipment, once and for all. Any deviation from the right position and distance causes errors, and *sideways tilting of the head has catastrophic effects*, e.g. it's absolutely impossible to watch stereo TV lying on a couch (one would have to move one eye up and one down; try this). How could anybody ever have thought this could work anywhere anytime anyhow ? - And that's not all: an always present distraction is false focus, e.g. the viewer for example looks to infinity according to eye parallax, but has to accommodate to a screen only 6 ft away. This hurts. In a cinema where everybody is sitting upright and focus is near infinity it may work, but for TV it's not an option. The task is simply not possible to accomplish without very advanced recording, image processing and display technologies.
I'll now shortly list the major corrections necessary for a good reproduction.

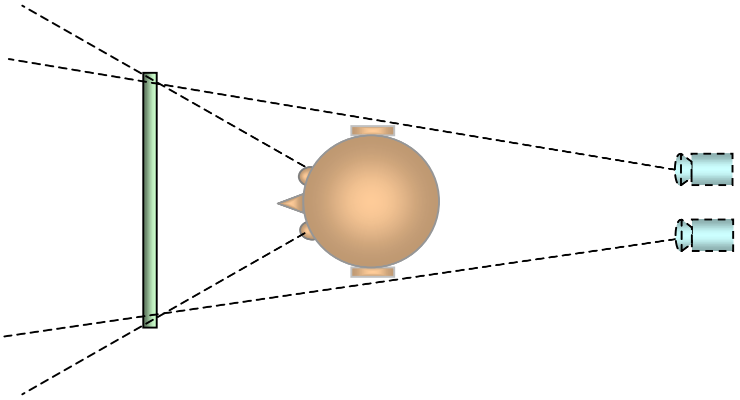


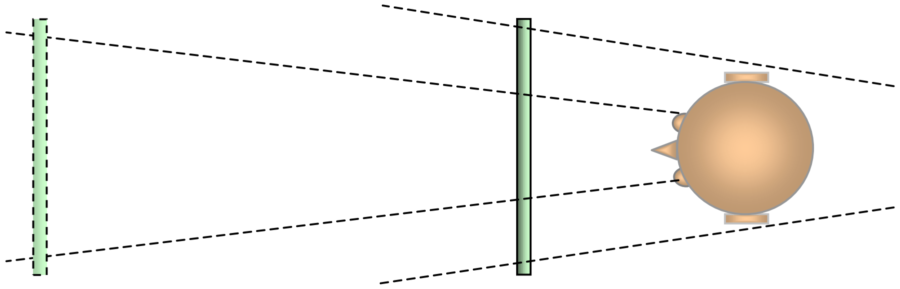
Perspective correction



Tele and wide angle shots have very different viewing angles. Due to this, different viewing environments with different screen sizes would be necessary, and zooming has always been a major problem for 3D filming.

In order to reproduce the original perspective, the viewer would have to be precisely at the position of the recording cameras.





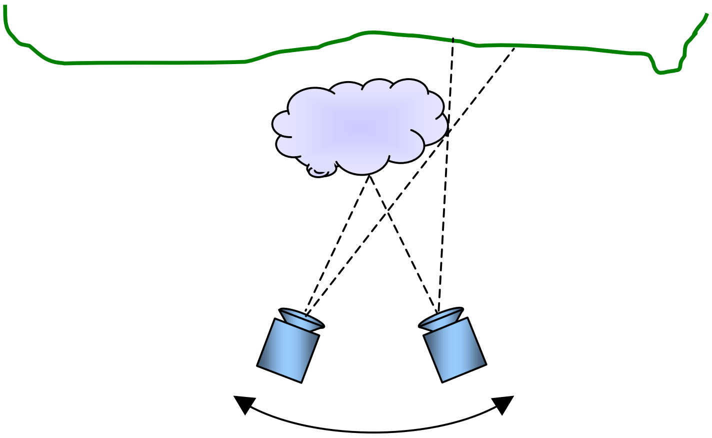
If viewer distance and original camera distance diverge too much, things get tricky. Images and objects appear very unreal, making the entire effort of 3D production quite useless.

In the example, in order to retain the original perspective, we would have to create a distant virtual screen window inside the

real display. This would render most of the real display area useless and require a very good resolution for the still active part.
This would not be a very reasonable approach.
If we had the processing power to generate a different perspective, we could retain the full display usage as well as constructing a physiological perspective.
The only useful strategy to shoot 3D zoom and tele shots currently is to enlarge the stereo basis together with increasing focal length. This allows for the same distance separation in a tele shot as in a normal shot and makes a natural looking reproduction much easier.

Panning effects

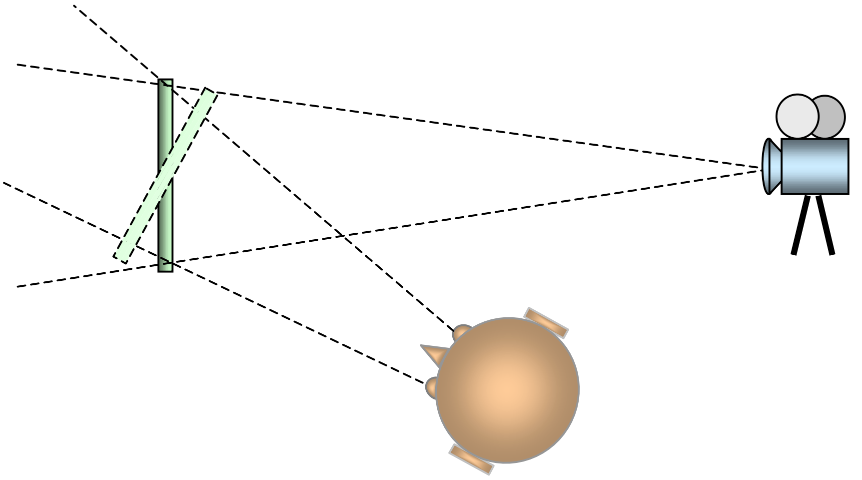
Imagine you are outdoors. There is a tree 50 yards from you and some others a mile away. Now if you move your head sideways by only 1 inch, you will already see the foreground tree move relative to the background. If it would not, you would know that something is wrong. That would not be a real scene.
Any 3D movie so far has this problem. In real world, we constantly use head movements to check perspective. So any 3D reproduction not dealing with this is significantly flawed.



This powerful trick for cameramen (that only a few actually use), can make any still shot look 3 dimensional:
Simply shift the camera sideways or upwards a bit while fixing one point. Everything suddenly has depth !

(There has also been a 3D shooting technology that wiggles the camera all the time, but this makes people seasick).
A certain type of 3D displays exploits this fact, by showing only one picture, but changing the virtual camera position according to head movements of the observer. This allows for 'looking behind' and gives a very 3 dimensional impression even though there is no real stereoscopy. These displays work for one viewer only.
Combining panning with a stereo display (polarized or shuttered). gives a very realistic impression. It is of course necessary to have an image generator that delivers all viewing angles, corrects perspectives etc. Focus errors remain. I'll discuss this later on.

Viewer position correction



If we had enough processing power, we could conceive to recalculate the perspective for the viewer's location. Though difficult, that's the only way to make things look real. What actually happens by itself with any conventional 3D screen, is that the virtual window seems to tilt towards the viewer and objects are distorted and start to look unreal.
As large correction angles would be difficult to calculate even if we could, another way would be to allow for some tilting towards the user's average position and to perform dynamic perspective corrections for small head movements only.

3D displays: a review

We already know that a vision simulator can do many things not achievable with screens of any type. As screens will also be around for various reasons, I will now give a short overview of stereo display (and connected recording) technologies.

The holography hype

Some decades ago, when holography was invented, many writers thought it would revolutionize film making. Pretty hasty indeed. True holograms need laser light to shoot, and film or displays with micrometer resolution to reproduce. They cannot be zoomed, adding color is difficult and may produce artifacts, and so on. Filming would actually require a studio with laser lighting and film formats several feet wide. Pretty weird. For shooting outdoors, no realistic ideas have arisen whatsoever. Effects like self interference are also severely limiting the quality of naturally recorded holograms. So the only still promising variety will be synthetic holograms generated by computers.

The VR disaster

In the beginning, many types of virtual reality glasses and goggles had been developed, most of them bulky, heavy, with primitive optics and low resolution displays that would make users seasick and send them to the chiropractic. Even now, no products have been sighted so far that were the opposite in *all* aspects. Technology strides forward, and without very high resolution, *eye trackers*, *mask displays* and semi transparent displays, nothing will really ever work perfect. So even the most developed VR glasses today are still not acceptable for our purpose. One reason may be that the bad reputation that VR acquired from early experiments has slowed down development.

The anaglyphic headache

A simple and ancient method to reproduce 3D pictures is to use red and green color to separate pictures. I only mention this for completeness. Wearing red/green glasses may be something one can adapt to for a little while, but it is quite inconvenient. Only black-and-white pictures are possible, and the method, although periodically revived on TV, is at best viable to reproduce some 3D pictures in educational books.

Polarizers and shutters

A screen or projection display can show 2 pictures sequentially or in different polarizations. Shutter or polarizing glasses must be used to deliver the pictures to the appropriate eye. Both approaches have the disadvantage to require glasses (then we could better use our vision simulator), and to allow no 'looking behind'.

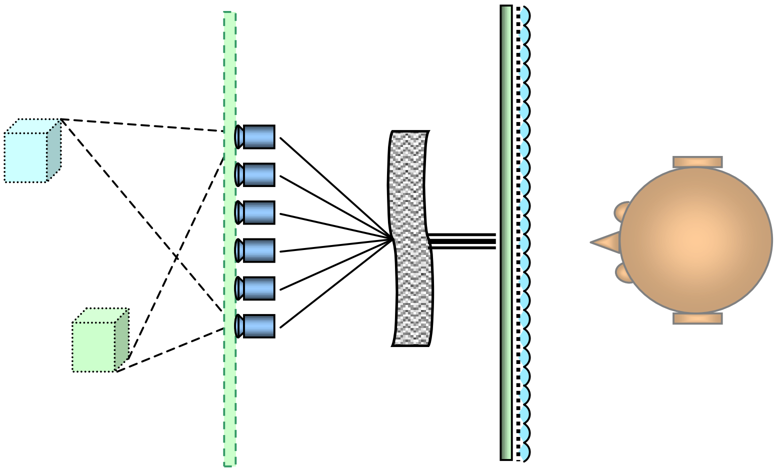
Given the fact that setting up a shuttered 3D display is very easy with any computer screen, and the shutter glasses aren't expensive, one may wonder why this technology is used so little.
It may just be that 3D is not considered necessary enough for most users, and it's certainly so that the problems with static 3D displays (false perspective, false size, false focus) are just annoying.

Perfectly displaying 3D is absolutely necessary for the concept of a vision simulator, including operation systems and applications, and I think that 3D video will probably just come as a side effect when this equipment will be in everyday use and no further effort has to be taken.

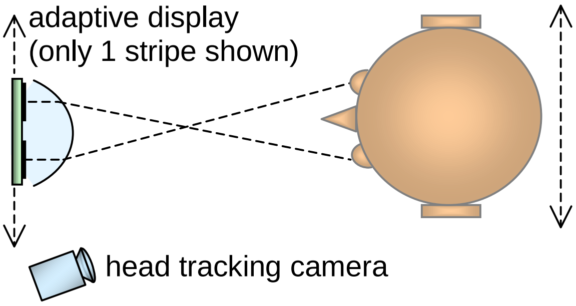
If the advanced recording and object oriented processing methods suggested here can be realized, this will also transgress into everyday use quite smoothly .

Auto stereoscopy

Auto stereoscopic displays typically use cylinder lens arrays to display different pictures for every viewing angle ('lenticular displays'). In Theory, this would work, but practically one can only display a very limited selection of angles and therefore has to cope with transition effects between them.

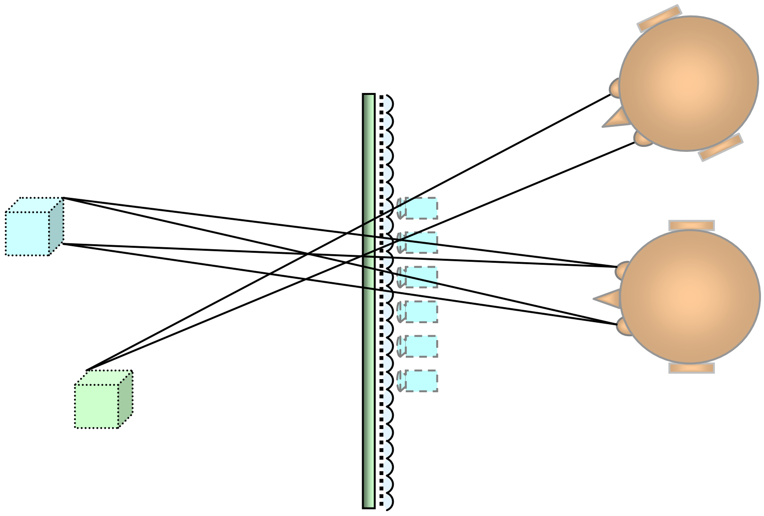


In an auto-stereoscopic assembly, two or more cameras pick up the scene. The cameras are arranged in an only horizontal row.
By arranging separate display stripes for each camera picture behind tiny lens stripes, each eye of a spectator should get only the image of a certain camera from a certain viewing angle. In practice, there are areas in between where the 2 images overlap.



If we move the display stripes sideways, we can adapt the optimal viewing points to different head positions, so that each eye always only gets its appropriate picture, and no overlapping or image switching occurs. This can only work for a single viewer.

We could further generate dynamic perspectives, allowing to view around objects.



Here I have visualized how the screen with its display and lens stripes virtually reproduces the camera array.
It is obvious that this assembly can also provide a certain view 'around' objects (regard the upper viewer position).
It is also obvious that with freely moving observers, many perspectives will consist of an overlay of 2 images (regard the lower viewer and the lower edge of the upper cube).

We see that auto stereoscopy, in spite of the advantage not to require glasses, has the problem that it produces artifacts because intermediate angles are simply achieved by overlaying 2 images.
What also becomes difficult is the realization of large displays, with accordingly large viewing distances from the display. In order to give every eye its own picture, very small angular deviations have to be resolved, which in turn reduces the 'good' viewing area, requires more cameras and more display stripes, and a pretty high optical precision.
Adaptive auto stereoscopy gives a 3D impression that works without glasses, but it is restricted to personal use.
So the main commercial application with this are notebook displays, that could as well be realized with shutter glasses (display

and glasses switching between 2 perspectives) at very little disadvantage and little effort.

3D effects in the typical applications involved (construction, chemistry etc.) may also be achieved by just tilting and turning virtual objects by software, so the practical advantage of any stereo display in this case is limited anyway.

With all non holographic displays, there is also no correlation of focus and distance. The user has to focus on the computer screen, while the object depicted may be at a different virtual distance. Real holograms and vision simulators could do better.

Finally remember the already stated fact that the viewer would have to keep his head in an upright position all the time. Nothing realistically acceptable for a home TV.

Pseudo Holography

Newest developments promise pseudo holographic displays that can deliver a much denser raster of viewing angles (0.8^0 for example) [30]. They seem to work quite similar to lenticular displays in principle, and image generation is still very demanding. As it is very new, a final conclusion cannot been drawn, but again this is 'head up only' 3D even in spite of the tremendous effort to provide dozens of perspectives.

Really holographic displays?

Holographic interference patterns usually have a resolution better than 1µm. This means we will not need millions, but trillions of pixels, in case of big screens. It simply busts up any attempt to do it with current technology. From the fact that single pixels lead to ring structures in a hologram, one might speculate about injecting pixel data in a flat nanostructure that would then generate holographic patterns by propagating waves.

The non linearity of the pixel ring patterns, the necessity to limit patterns at edges of foreground objects, to also generate patterns that originate outside the display window, make this a tremendously difficult idea.

The extreme increase in chip density still ahead of us in the next decade may deliver something that integrates this kind of processing and an ultra high density display on a single chip. These chips could also become bigger or be combined for larger screens, but that's really way ahead.

With very substantial efforts, moving holograms some inches wide have been demonstrated, with many drawbacks (horizontal perspective only), and that's about the top of the line now [65].

The conclusion is, that other than small versions for projectors or possible display glasses that we have discussed, there is currently no concept in sight for a truly holographic screen that could get to the market in less than several decades.

Vision simulators

The advantage of combining a vision simulator with computer based 3D recordings, is the possibility to calculate output pictures for different viewing positions and angles, according to the user's head movements. For example, a virtual holographic display window would then be projected with coordinate transformations to make it appear relative to any desired position in a real environment.

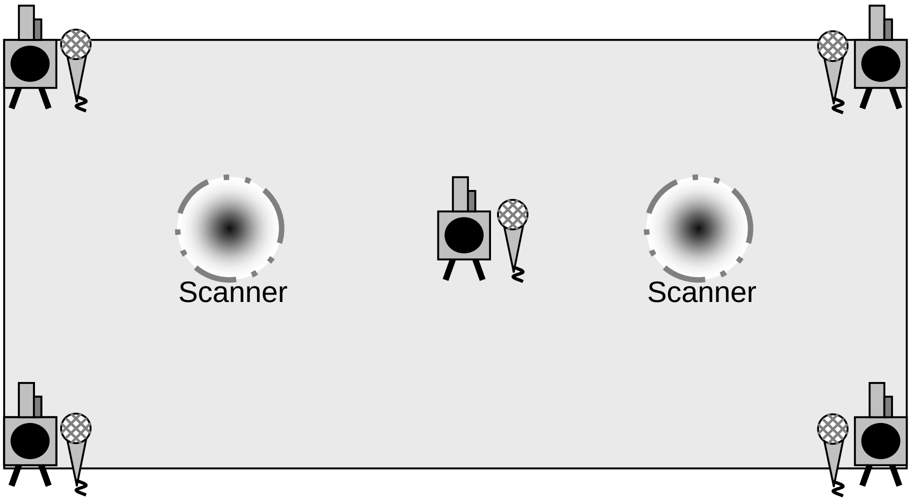
This technology is very demanding as it requires high resolution vision simulator displays for satisfactory results, but it would be more than rewarding.

Yet the tasks to perform for a perfect reproduction are less difficult compared to the *recording* of the pseudo holographic data we need as a basis, no matter if for screen or glasses.

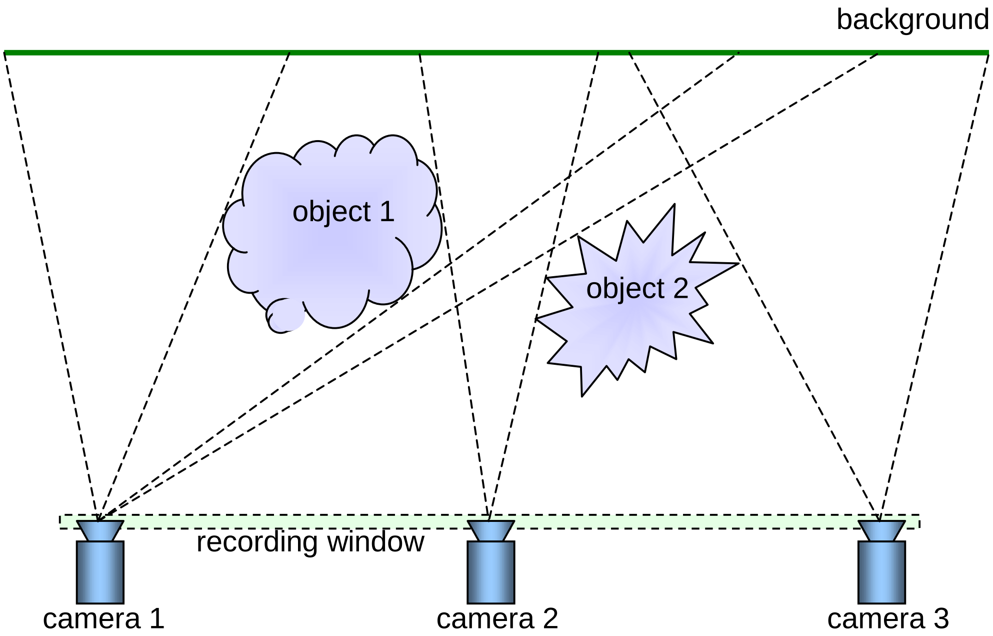
We should also recall that personal vision simulators will already be of great advantage even if we could only simulate classical two dimensional screens, as these could be set up anywhere in any size at no cost, and also for surround cinema, where we wouldn't necessarily need dynamic 3D as well.

3D image recording

Holography had a big media attention when it was first discovered some decades ago, but it never worked out. With today's technology, we could at least try to record as much data as there is in a hologram and can be perceived by a real observer.
What we need to do, is to place several cameras in a plane similar to an intended display screen, in order to record many viewing angles simultaneously.



The recording plane erects a virtual window, that could be reproduced in a similar window at playback and through which the scene can be viewed from any distance and angle, giving a realistic three dimensional impression, even with the possibility to see behind objects, as in a true hologram, if only we would be able to generate intermediate perspectives. Without this capability, the fixed window would be just as inflexible as conventional stereo video. As stated, it may nevertheless be wise to start with multi channel recordings at the present time, and hope that we could improve at a later date. We would need computers to differentiate the scene into objects of different distance and encode the entire picture in a descriptive, object oriented way for example. More about this thematic can also be found in [64]. In order to facilitate things, I propose additional distance scanners (laser, ultrasound, radar, or other). Object based *compression* was already laid out in the MPEG4 standard, but never implemented for common use.



This picture shows the (idealized) viewing angles of 3 cameras, with the shadowing and 'see behind' areas with some simple example objects.

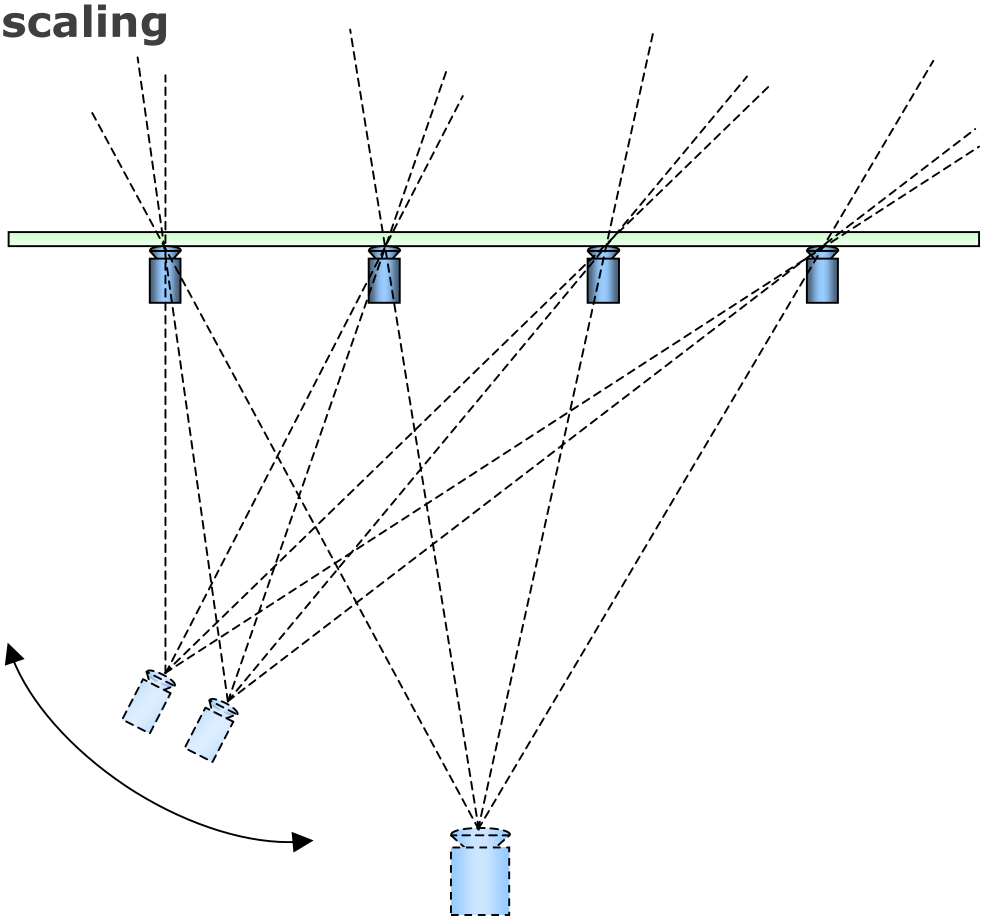
The recording is obviously not perfect, in that certain parts of a real scene will be missing as there is no real camera for all of the possible angles that a spectator could take when viewing the playback of this on a hypothetical, ideal 3D screen.

An array assembly would work better if more cameras and microphones were used. As cameras will rapidly get cheaper and smaller, we may expect to see arrays with many more cameras than depicted here, making the task much easier.

Such a recording technique that still confines the movie to a window view may be the usual type for a long time to come, but other possibilities, like surround, will also become more popular with vision simulators.

The following picture illustrates that there always are certain sections in the image of any of those cameras that resemble a section of the image a virtual camera behind the 'window' would have seen. By merging the appropriate image parts with some software help, the image of the virtual camera can therefore be constructed. This also applies to stereo viewing (two eye perspectives separately).

Upscaling



For a really perfect recording we will need many cameras. The more cameras, the fewer dead angles and the simpler the computing task. As it is practically impossible to align all cameras with single pixel precision (and also not reasonable, as they get slightly different pictures and perspectives anyway), we can assume their fine positioning to be random. So we would have to drop many pixels as they won't all fit into a common raster: a pixel in between just delivers no detail information

We could take advantage of this however. The cost of the array can be largely reduced if we use low resolution cameras and calculate a higher final picture resolution from overlaying and correlating their outputs. If the camera chips have pixels smaller than the raster size (some products have this anyway) as well as good optics, this is perfectly possible.

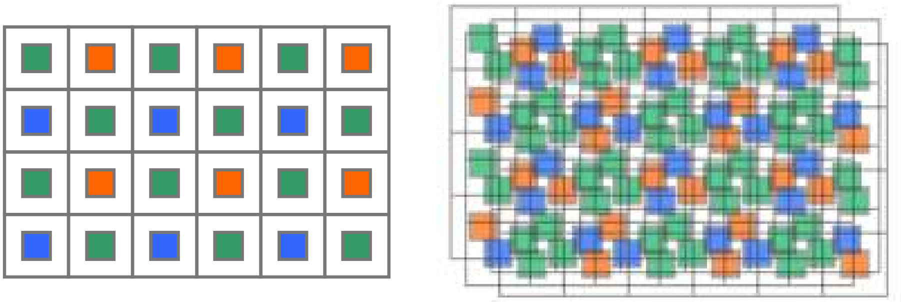
A resolution improvement of 2:1 or even 4:1 should be achievable, so an array of standard TV chips could as well record double HDTV as required for really large cinema applications.

Such an array may not be more expensive than just one dedicated double HDTV camera.

Let's explore this a bit further: smaller pixels are less sensitive or produce more noise, of course. Pixels with ¼ size have 2 times more image noise, for example. Adding up 4 signals just compensates for this (averaging n^2 signals reduces noise n times). For 4 times as many pixels, that is 4x resolution, the noise aspect would hence imply 16 cameras.

Under this aspect, an array of 16 times 9 cameras for example would be perfect to provide a resolution of 4x720 times 3x576, i.e. 2880x1728, which is really a cinema format (normal HDTV is 1280x720 in the US and 1920x1080 in Europe).

The computing equipment involved would first calculate raw correlations between different camera pictures to get the depth information, which is much easier here than with just 2 stereo cameras. Almost no dead angles will occur. With some additional depth sensors, this will be a cinch.

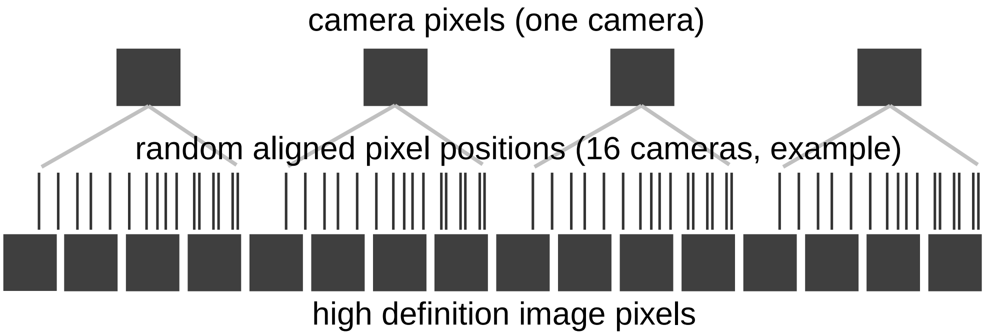


Section of a camera sensor with a sensitive pixel area of <100%, and random overlay of four pixel rasters (right)

In a second step, the computer would compare small image areas one by one, in order to find a correlation between pixel information from the single cameras and actual image detail. This way, camera pixels can be mapped to actual image pixels with high precision. This will only deliver actual fits if accidentally a camera pixel really fits into the higher resolution raster. Yet if we have 16x16 tries for just 4x4 raster points, the odds are good.

Again, nothing has to be precision aligned, not even the zoom factors, apertures and focus have to be 100% exactly fit, although they should be synchronized. Every final alignment and adjustment can and should be taken care of by software, that would by the way also deliver a classical image stabilizing function, without the usual raster misalignment or blur. The entire camera array will in the end be really affordable compared to the usual prices for cinema equipment.

The following example shows the mapping of a group of 4 camera sensor pixels to 16 high definition (4x resolution) image pixels in one dimension. We see that some image pixels get more properly aligned camera inputs while some may get almost none.



Calculating how well this would work is quite simple: that a certain high definition image pixel is aligned with one camera pixel by ±25% of the pixel width, has the probability $^1/_8$ in one direction (as we regard an 4:1 pixel relation and an alignment to 50% of a pixel width) or $^1/_8{}^2$, i.e. $^1/_{64}$ in two dimensions.

Inverse probabilities are easier, because they multiply[*]: that it's *not* the case, has a probability of $^{63}/_{64}$. We have 256 tries with 16x16 cameras and (assuming random alignment) this results in an overall probability of $^{63}/_{64}{}^{256}$, actually 0.018, or $< 2\%$. So only about one in 55 pixels will get no good input to derive the high definition information.

---

[*] Mathematically inclined readers may stumble over this as generally we would need the more complicated formula for the binomial distribution: with n trials at a probability p, the probability for k hits is

$$P(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } 0 \le k \le n \\ 0 & \text{otherwise} \end{cases} \qquad \text{where } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

As we only want to know the probability for 0 hits ($k = 0$), n above k and p^k both are 1 and we get $P(0) = (1 - p)^n$, as stated.

This results in a little high frequency noise, that can easily be smoothed out in the temporal axis, by averaging subsequent frames. Concurrently, this will result in a certain amount of 4x resolution information for any image pixel at any time, with a very high probability. Even with fewer cameras, we would get some decent coverage already. Real objects may of course be moving. The alignment will then be different for any frame. We will have to regard a certain displacement of image objects from frame to frame (motion vectors); hence, we won't average exactly the same pixels of subsequent frames, but those corresponding to the same (moving) scene pixels

Calculating all influences precisely and concurrently would result in some average frequency response for the entire recording system. Deriving this would be beyond the scope of this book.

Let's just have a look at the average pixel coverage for different numbers of cameras, using the same simple calculation scheme as above.

pixel accuracy: $^1/_2$x$^1/_2$	resolution multiplication		
number of cameras	**2**	**3**	**4**
4x4	0,356	0,637	0,777
5x5	0,199	0,494	0,675
6x6	**0,098**	0,363	0,567
7x7	0,042	0,251	0,462
8x8	**0,016**	0,165	0,365
9x9	0,005	**0,102**	0,279
10x10	0,002	0,060	0,207
11x11	0,000	0,033	0,149
12x12	0,000	**0,017**	**0,104**
13x13	0,000	0,009	0,070
14x14	0,000	0,004	0,046
15x15	0,000	0,002	0,029
16x16	0,000	0,001	**0,018**

Probability that a high resolution pixel is not aligned to at least one camera pixel by at least ±$^1/_4$ pixel width in x and y direction

We see that 12x12 cameras, for example would already be very good for a 4 times resolution upscaling, and 6x6 if we want 2 times the resolution (only about 10% imperfect pixels).

The 'imperfect' pixels can be filled at a high probability by the already mentioned temporal averaging. So even fewer cameras could suffice, but we need many of them to eliminate dead angles anyway. With some more cameras, we could as well get below 2%, so we wouldn't even need the averaging procedure.

Using an odd number of cameras, could perhaps help to distribute the uncovered pixels a little better.

All the perspective and upscaling processing would best be done in the cameras themselves, that could form a massively parallel processing unit if we equip them each with a little computing chip and wire these in an array, with high speed serial links for example. Still these camera units will be cheap, and even many of them would stay affordable in a professional cinema context.

The entire video data picked up by the camera array will finally boil down to a partially descriptive format not bigger than that from two conventional high definition cameras.

With HD camera chips getting cheaper in the future, the upscaling effort will become less important, but the electronic precision alignment of pixels will still be necessary, and also ¾ of all pixels gathered will still have to be dropped, as they won't fit into the same common raster: as mentioned already, it is simply not possible to ensure that all cameras get the same scene detail with the same alignment of their own pixel raster versus that detail. Natural scenes are not just flat test patterns but have a three dimensional structure that results in a different perspective for each camera.

It's also not realistic to believe that we could precision align several optics, including their focus and zoom factor, to an accuracy of an HD or even 2x or 4x HD pixel. Just for this reason we've just explored the upscaling approach for camera arrays, as we can really expect any raster alignments to be random anyway.

Hence, upscaling isn't just a weird and complicated idea, but a logical and consequential approach to turn some principal and unavoidable problems of camera arrays into a virtue.

Amateur video

One could hardly imagine the average video amateur walking around with a huge camera array.

This is something suited for big cinema or home movie, targeted at many spectators at different positions or even walking about. Simply taking the vacation or family video in a "what you saw is what you get" style is better done with the device we've already been talking about all the time: the vision simulator has stereo (HD) cameras, picks up anything anyway.

The recording conditions are good, as we usually hold our head quite steady. Focusing could not only be supported by looking at a target, the eye trackers could also deliver distance information right away by analyzing our squint, and we could even just show a frame border overlay instead of a complete viewfinder image.

Other than with the large cinema equipment described, the viewer position here is bound to the recording position.

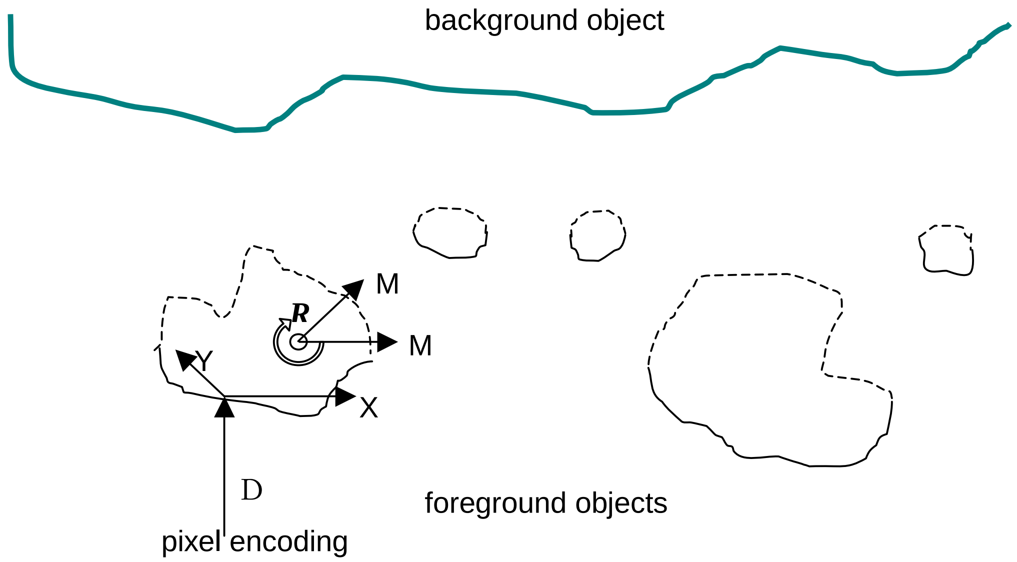
One major disadvantage of conventional stereo pictures however, the intolerance against sideways tilting of the head, could be dynamically compensated for with some not too complicated software.

It's obviously necessary to exercise some care about one's own head movements in order to get recordings that one would like to view later on. Much less inconvenient than with any current equipment anyway: no fumbling around with straps and cases and switches, no waiting, no peering into a viewfinder, no dragging and towing.

The cigarette box sized pocket unit that we expect to persist for the first generations of vision simulators might still be there, but then it wouldn't be larger than the smallest current camcorder, serve a thousand more purposes anyway, can stay in the pocket all the time, and will easily be able to store days after days of video.

So equipped with a vision simulator, a video camera with many bells and whistles will be available anytime, at no weight or cost, and it will work almost unattended and unnoticed. Could we imagine any better ?

Encoding

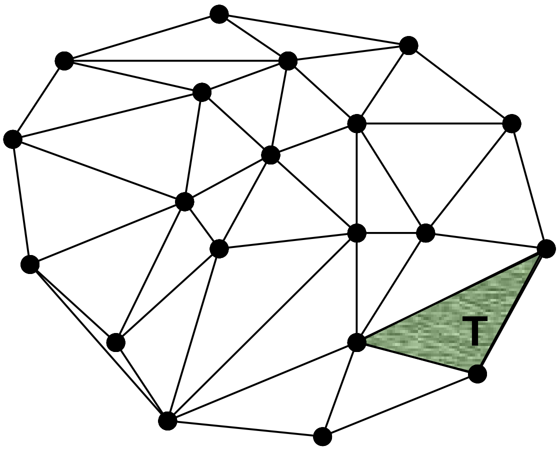


A computerized encoder should be able to separate all background and foreground objects, and to encode all the information gathered. Of course there are more difficulties than illustrated here; for example, transparent objects, fog, etc.

With glassy objects, the software must try to derive their properties from the pictures from different camera angles. A certain 'intelligence' will be necessary to derive a geometrical representation as well as a correct description of material properties, at least enough to allow for a correct calculation of intermediate viewing angles between the original camera positions.

Even though the software does not have to 'understand' what it sees, it becomes obvious at least with glassy, foggy and reflective objects, deriving a good object description becomes extremely difficult. Theoretically, current MPEG standards allow extensions for objects and surface textures as well as motion vectors, which would also allows for a high level of compression. It's however for a reason that so little thereof has actually been implemented.

While PC graphics cards will soon be able to synthesize arbitrary complex scenes so well that they look perfectly real, the opposite, i.e. deriving a description for such a synthesis from a real scene, is a lot more difficult.



In principle, 3D objects in graphics engines (e.g. game engines) are usually modeled as triangular surface meshes describing the spatial shape of the object (the more triangles, the better the fit), while textures (T) are used to fill the areas in the mesh. These textures can well be encoded and compressed like conventional images. With a sufficient number of texture elements and some smoothing applied, we can get a perfect reproduction of any real object. Game engines are based on predefined object descriptions as above, together with physical data describing the possible interactions of objects. A rendering engine then simulates a virtual camera to derive a perspective view of the scene as well as lighting and shadows according to virtual light sources that also have to be defined beforehand. The big advantage of the approach is that it profits from the huge advances of high end graphics cards driven by the game market.

Generating an object based representation of a real scene, requires to derive the object descriptions (mesh, textures) from real camera pictures. This has already been subject to intense research for a couple of years. It is not necessary to resolve the light sources in this case and to restore lighting later on (although this would facilitate the merging with virtual scenes). Lighting can be recorded as a property of the texture. Glassy, reflective and foggy objects, as said, cause difficulties with this approach, and time will tell how this will be resolved. Nevertheless, we may expect to see sophisticated encoding schemes of this type, and they'll seamlessly integrate with synthetic scene generation, enabling some very advanced hybrid production methods.

Holographic encoding

We may also have to think about other approaches. Here a little brainstorming (this seems to be getting into SciFi, but not really):

- Holograms can contain any of the nasty objects and reproduce them perfectly (they encode just anything light can show).
- The patterns of a hologram have a great deal of redundancy.
- Holograms can be computed (synthesized) from three dimensional picture data.
- The mathematics of the transformations are well understood.
- We could conceive a recording method where a hologram is recorded, then scanned and the vast amount of data it contains is reduced by computer algorithms, removing any unnecessary redundancy. Theoretically, this should be possible up to a point where the amount of data remaining is not much bigger than with the conventional methods regarded.
- Such a redundancy reduced hologram could be transmitted and a receiver could use it to reproduce not a hologram, but single custom tailored perspectives for single viewers instead.

This would greatly deviate from any current approach to image compression. One question is certainly left wide open, how to get at a hologram at recording, that could then be processed. Probably we would still have to derive this from camera arrays. It could be possible to develop a procedure to transform image data from a sufficiently dense array directly into holographic data.

The downside of such a method would be that we wouldn't get at an object based description and therefore the merging of real and computer generated data in movie production wouldn't get much easier this way. Objects however may just be too much of an abstraction, too much data. Procedures based on the mathematics of holography could perhaps enable us to merge 3D scenes as easy as a collage of paper clips, without any need for abstract description of surfaces, textures, materials and so on.

What is it that makes holographic data for large screens so vast? Is it the large number of perspectives? Not necessarily, as the hologram can't record much more than superficial information.

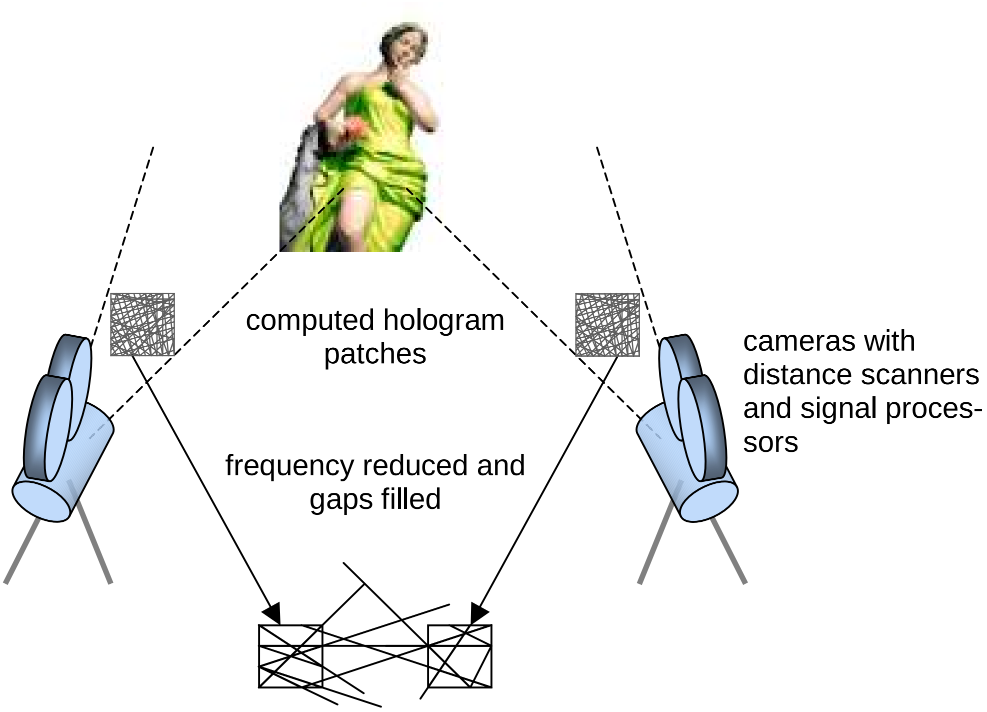
There is another fact that greatly contributes to the amount of data: a hologram is, theoretically, utterly crisp. Light wavelength is the limit (those little holograms that you can buy are not as crisp, but that is because they are viewed in white light and the lamps are not point like). A large screen image with 1/1000 mm resolution doesn't make any sense. And we can't just reduce this data by just blurring the picture. We could perhaps use a smaller window, as we have seen with holographic projectors already.

There is another way to do it: use a larger wavelength. We can't simply use microwaves for photography, of course. A millimeter wave hologram could theoretically be taken by a large area of micro antennas and wouldn't even need a reference beam because these elements could directly work with an electrical reference frequency, and an image could be read out from the array like with a giant camera chip. These waves however would go through things impenetrable for light and vice versa.

Ultrasound could also be an option, I've already considered this for medical applications in [1]. There a sensor array could be made from a single silicon wafer, but ultrasound of this wavelength doesn't work in plain air, beneath other problems.

The better way to do this kind of reduction would be to use a light hologram and then recalculate the patterns, synthetically that is, to look as if they were recorded with longer waves, ½ millimeter perhaps instead of ½ micrometer waves. Such a hologram would essentially carry all superficial and spatial information, yet with less detail resolution, in the order of ½ mm, only as much as we need for a large, apparently sharp natural image. The reduction factor possible is about one million. From the information theoretical view as well, we should indeed not need a much larger number of pixels than with a classical image.

It's nothing less than fully implementing step 1 of the aforementioned 3 step paradigm. Actually, anything 'you could see' is contained in the hologram, with the little constraint that we confine this to seeing through a certain window frame.



One way to achieve this could be with cameras in an array. Each camera could, together with depth sensors, just deliver a little keyhole hologram (just the size of the camera aperture). The most critical step then would be to use such little hologram patches from many cameras to calculate – or rather estimate – the holographic patterns for the areas in between. This should be accompanied by the conversion to a much larger wavelength.

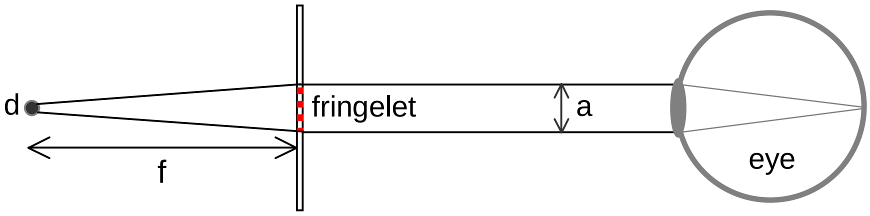
It is not yet foreseeable if for this stage, purely mathematical, object based or both methods would be appropriate. At least for the difficult parts, e.g. glass, fog, mirrors, a mathematical method based on the holographic patterns could have great advantages.

Maybe a real microwave hologram as discussed above could also be helpful, to fill the dead angles of the camera array.

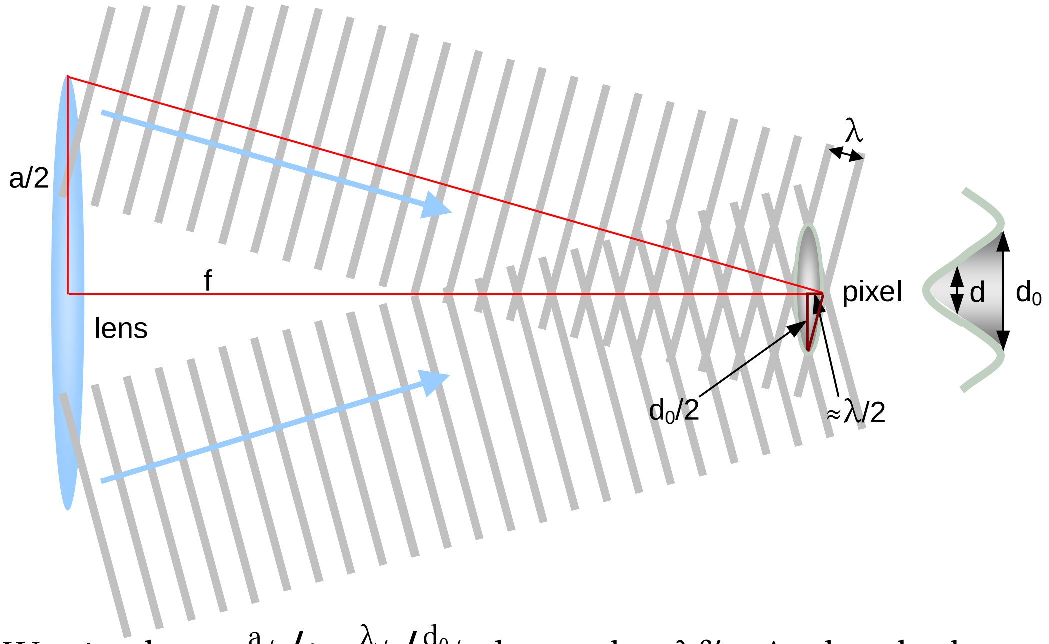
A hologram made this way already has a great deal of redundancy reduction, i.e. is relatively compact in a certain way: it encodes just anything a viewer could see through a certain virtual window. 'Seeing' in this case involves anything that could be achieved with light.

I proposed holographic encoding in the 1st edition of this book. Several months later, a first very short article about it appeared in scientific literature [83]. It really appears to be an entirely new idea. So I'll now explain it more thoroughly.

The hologram actually contains more than a certain viewer could see at any given time, as it has all possible perspectives included. Nevertheless, this is what we need, as we can't know the viewer's position beforehand at least if we don't deal with real time transmissions, and as visual media are usually produced with many simultaneous viewers in mind.



Let's compare the encoding approach to a light hologram. There, a certain image pixel (d) can only be represented by an area of fringes ('fringelet', actually a tiny zone plate) no larger than the viewer's eye aperture (pupil diameter) a. This also means that the theoretical resolution in the micrometer range is not achieved in practical viewing situations. The resolution a zone plate or a lens can deliver, follows very simply from the actual pixel sizes light wavefronts from different parts of the lens are able to form. With a slight simplification for larger focus lengths, the following picture allows us to see this easily:



We simply get ${}^{a}/_{2}/f \approx {}^{\lambda}/_{2}/{}^{d_0}/_{2}$, hence $d_0 \approx 2\lambda f/a$. As d_0 only denotes the rims of a sinusoidal intensity distribution, the actual pixel diameter is smaller, depending on the contrast values we require, e.g. $d = 1.22 \cdot \lambda \cdot f/a$ as found in textbooks. We also have to consider that wavefronts from inner parts of the lens (smaller a, larger d_0) will spoil the result towards larger d values. It would not be too difficult to derive an exact result for d by refining this model.

The formula resembles the Gaussian beam ansatz for laser beam focusing discussed earlier. If we imagine a laser beam with diameter a, originating from the eye and being focused by a lens positioned in the screen area, we see the analogy. Again, a wider beam can be focused to a smaller point than a narrow one. The factor there is 1.27 instead of 1.22, but this simply depends on the blur or spread we want to allow.

The formula also applies to both sides of a lens likewise, so it can tell us about the resolution of telescopes: A spy satellite with a large telescope (aperture a=1m) for example at λ = 500nm and an altitude of f = 100km could resolve d = 5cm (2") on the ground. We can also derive that it would take a lens or mirror almost a mile wide to see the remainders of lunar missions on the moon, or that an HDTV camera with optics could fit into an 1/8"(3mm) cube.

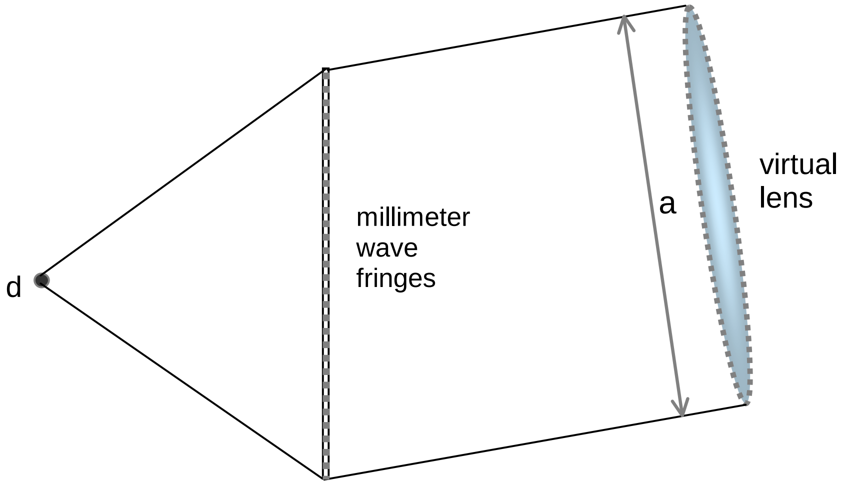
Applied to the eye (a ≈ 2...5mm, f ≈ 20mm), we get the strange result that theoretical crispness is lower at bright light, when the iris contracts and our lens aperture gets smaller. Actual d values for different lens apertures are between 3 and 6 µm, or ≈0.4 and 1 arcmin of angular resolution. Actually, 1 arcmin is what physiologists have measured and what TV standards are considering as 'crisp'. Hence, the actual resolution of our eyes comes quite near to the theoretical limits set by their size and optical aperture.

Now what do we get when viewing a real hologram ?

With a viewing distance f = 1 m, λ = 500 nm, a = 3 mm, we get a pixel diameter d = 0.2 mm. Due to our small eye aperture, viewing a real hologram we therefore don't see all of it's theoretical crispness (which can be about 300 times better).

Nevertheless, the pixel has to be encoded over the entire screen area, because viewers could come from any direction.

It becomes quite obvious here, why it is possible to make holograms that show entirely different pictures from different directions or unreel a film sequence as one moves by.



In case of our millimeter wave encoding scheme, the decoding algorithm would use the entire area in any case, somehow equaling a virtual lens as large as the screen. So we get about the same practical resolution as with the viewing situation for a real light hologram shown before (if $f = a$, then $d \approx \lambda$).

A problem would be remote objects that can only be seen through gaps between more proximate objects, as these could only be represented on a smaller screen area. It may be necessary to care for this, for example by partially inserting layers with smaller wavelengths (virtual holograms allow for such tricks).

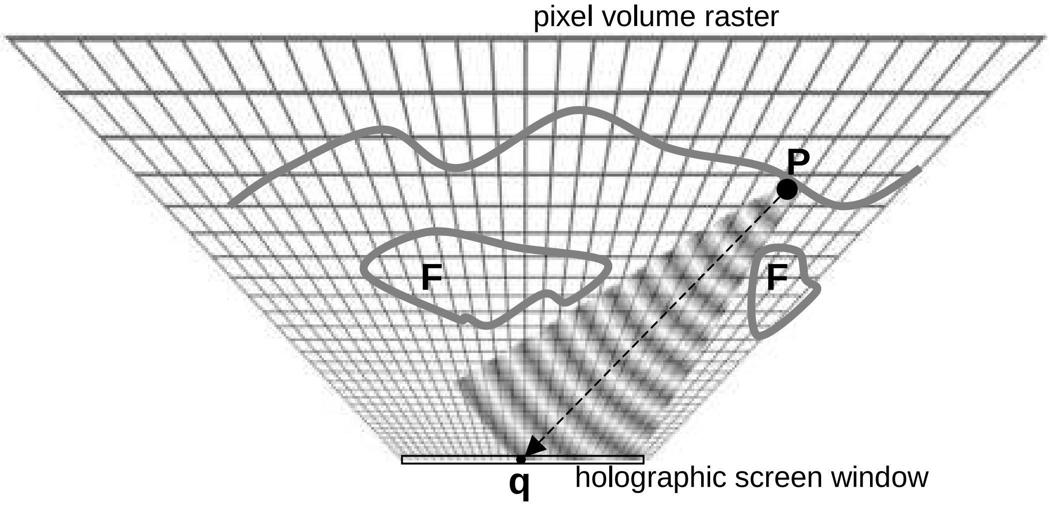
One problem remains with synthetic holograms: even in spite of the larger wavelength we use, the amount of computation we need will likely be tremendous. A lot of research still has to be done (e.g.[58]). The approach shown in the picture - interpolating between keyhole holograms - may or may not be effective.

Still the keyhole holograms have to be computed, and the interpolation may be complicated. Another method would start from a 3D object oriented representation of the scene, mainly based on surface descriptions and textures. This is only partly satisfying, as fog/lenses/mirrors are a problem with these scene descriptions.

Anyway, let's chose such a basic scenario and estimate the principle amount of computing, i.e. the feasibility of synthetic hologram generation. From some basic considerations concerning diffraction, we may guess that a millimeter wave hologram would need about twice the resolution of a classical picture, i.e. 4 times as many pixels.

With only basic algorithmic tricks applied, we could have to add up the interference patterns for all image pixels as they appear in the area of the holographic window. For 1 million pixels (very roughly 4x the count of a TV image), we'd have to add about 1mio. x 1mio. individual gray values. This sounds quite weird, but we'll likely find algorithms that reduce theses calculations by several orders. We may also greatly parallelize the process. Let's develop this a little further:

Assume we have a scene description consisting of surface pixels (a simple case). The number of pixel we have, will just be a little larger here than the number of screen pixels.



The further away scene objects are, the less resolution we need. Here we try to use a raster expanding with an angle corresponding to the viewing angle of the recording cameras (more can't be recorded).

The raster could expand up to a few hundred ft. of distance, anything more distant actually appears as infinite and can be allocated to the last pixel plane.

Actually this results in a pattern of single pixels, but from the distance they should appear as a solid surface.

Now we have to map all scene pixels (P) to every single screen pixel (q). The illustration shows the virtual wavefront coming from one scene pixel and getting to one screen pixel (wavelength shown larger than real).

Getting the wave amplitude that the scene pixel will cause at a given screen pixel, involves calculating the proper distance, then calculating the amplitude. The amplitude calculation at q can in principle be reduced to cutting off decimals before the point (if we just normalize wavelength to 1), then looking up a sine wave function in a simple table. Calculating the distance may be a bit more difficult but we shall find something easy as well. Then the amplitude (gray level) is added to the screen pixel, together with the (static) local gray level of a virtual reference beam. So this are still about 1mio. x 1mio. operations.

Yet if we use a large chip and dedicate *one complete calculation circuit* to any screen pixel, it's 'only' 1 million calculations/image in each of them. This is 60 millions per second, not a big deal at all with current chips easily reaching billions of clock cycles per second. Current high end CPUs or graphics processor chips have up to 100 mio. switching elements, this would result in 100 per pixel if we want 1 mio. parallel circuits. Actually, we may hardly need more than 1000 switching elements per circuit, i.e. just 10 times as many, and this should be in reach in about 5 years as to Moore's law. So this is not SciFi, it's quite close to accomplishment.

High definition pictures would of course need about an order of magnitude more processing power, but even this is not too far fetched, and the approach considered may even be far from optimal.

Yet actually it's a bit more difficult: Not all P/q combinations would contribute to the result, as certain pixels are covered by foreground objects F in certain areas of the hologram window (above picture).

This masking of distant pixels is more difficult than it first appears, as we have to tell any pixel element in our chip if to process this input or not. Doing this all the time would entirely spoil the gain expected from parallel computing. We could try to load this information only once and re-use it for neighboring pixels by just shifting it within the computing array and only reloading some pixels where necessary. This is not so simple as well, as we have to calculate externally what to do.

Another problem may arise from edge diffraction. Not a big deal with real wavelength holography, but the edges with our wavelength reduced image are not so easy to keep clean. We may have to adjust fringe patterns especially to deal with this.

We see that while there are some tricks already conceivable for true hologram synthesis, this requires a lot more.

And this is not all. In [65] it is shown that naturally recorded holograms have a lot of unnecessary patterns from object self interference, and those contribute to noise. This won't happen with our synthetic approach, but it's not the only reason not just to mimic the classical recording assembly: digitally produced hologram patterns can also be optimized to deliver less diffraction effects than natural ones.

In [65] a lot of effort also went into the reduction of computations. The results meanwhile led to research prototypes of real time synthetical holographic displays a few inches in size. This doesn't mean that the holographic TV is near: computing requirements and display complexity cause this technology to be tremendously complicated and expensive, and the images produced aren't yet perfect.

The data reduction techniques used aren't necessarily helpful for us as well: a millimeter wave hologram has a great deal of redundancy reduction already, and won't probably endure much more. The display hologram for a vision simulator is viewed through a strong magnifier and also doesn't endure any blurring, that's often accepted with life sized holography.

What we could well use are those optimized digital fringes, and we would obviously limit any patterns to the area size actually viewable due to eye aperture and optics. This would also concentrate light to the eyes only, and thereby help to save energy.

If somebody would find for the Laplace transform (the one describing the ring patterns) something as computer friendly as the Fast Fourier Transformation (FFT), this would really be a breakthrough. FFT itself is already used for flat holograms like with the projector shown in the design chapter, but not simply transferable to 3D problems.

Dealing with color will in principle increase the complexity of holographic encoding by less than ¼, because we can just add 2 color difference channels of half or even quarter resolution, as with classical video (here, half the wavelength).

Some of the reduction methods from existing holovideo research, especially blurring, could be applied here, so the reduced resolution could really be translated into fewer computations.

The processing chips necessary for synthetic high definition holography may be available in about 10 years according to Moore's law, with a little luck and some smart algorithms. A significant amount of processing power is already available with the processor chips of high end graphics cards.

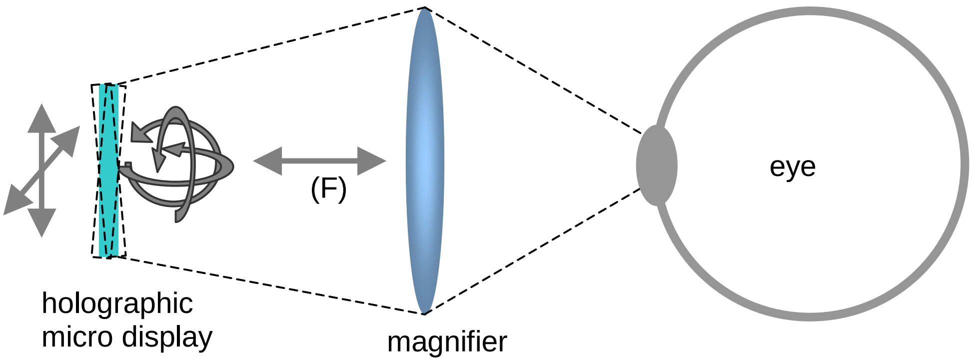
It's still a long way to go, even more so if we want to integrate the technology with low power portable devices, but we'll anyway have to wait a bit before camera arrays will be able to deliver the picture data at least.

Another demanding task is the decoding of a hologram into actual viewer perspectives for display glasses. It can be nearly as complex as to generate the hologram.

Last but not least, we also have to think about driving a holographic display in the glasses themselves as last stage of the transmission chain, and it may be rewarding but will get pretty complex to sort out possible simplifications with this, if the original picture source is already a hologram.

Just viewing a synthetic wavelength enlarged hologram, could easily be accomplished with a micro display exactly small enough to translate the mm wavelength back to light wavelength, and viewing this micro screen with a magnifier (that could as well be the mirror glass of a vision simulator).

For a realistic appearance, we would of course need to simulate head movements that would be possible in front of a real hologram screen. Otherwise seeing around objects - for example - would be impossible, and the entire reproduction would be no better than with a simple stereo movie. We could accomplish this by shifting and tilting the micro display opposite to actual head movements, measured with position sensors. This would include a lot of mechanics, but it would be simple.



Simulating movements to and from the screen (F) would be difficult however, as it would require to change the focus length of the magnifier or at least its distance to the eye, something almost impossible with the vision simulator optics we have discussed so far.

The ideal solution would of course try to do all these compensation movements electronically by changing the display image itself. Just shifting would be easy, but some other changes would require quite sophisticated mathematical operations on the hologram, that I'll leave for future exploration.

Another task with holographic displays in this context is a seamless integration of virtual objects with scene holograms, which will be necessary in several applications.

We see that this thematic is very complex, and a lot of research will be necessary to thoroughly explore it. Nevertheless, the feasibility appears to be given.

For recording or broadcasting, a further compression of holograms will be required, and it may well be possible. Holograms have a lot of regular patterns after all, especially if the original scene is not too complex. One could conceive compact coding methods for repetitious patterns, specializing on the representation of frequency, angle, curvature, frequency change. Like classical picture sequences, hologram films should also show a lot of redundancy over time - similarities between subsequent images - that could be exploited for further compression. If we manage to find something adequate (I'm quite confident about it), we may in the end be able to compress synthetic hologram sequences as perfectly as simple picture videos.

Giant screens

OLED displays are in a process of rapid development. These self luminating devices can be produced on flexible substrates. A simple production process is to print them, with inkjet printers. It's only natural to make them im stripes, roll them up and sell them as wallpaper. If this works, then the customer could just cover a wall with this at any desired size, screw some contact clamps on to interconnect everything, and a flat, cheap and efficient display would be ready.

If this doesn't work out, then there are holographic projection screens, that can work with little projectors and deliver the same contrast as a self luminating screen. This may even be getting the cheapest large screen ever.

Well, except for vision simulators. As some people might like the real screen more, if it works, we will look at the following from both points of view. This all applies to general TV and home cinema as well.

Virtual households and conference rooms

The impact of bandwidth

Glass fiber cables have a far higher potential than is widely recognized:

A single glass fiber could theoretically transport over 10^{15} bit/second (visible light has about 10^{15} Hz, and perfect modulation could yield more than 1 Bit/second per Hz of bandwidth). HDTV signals can be compressed to about 10 Mbits/second,

hence, 100 million HDTV channels would be the possible. With realistically achievable modulation technology (multispectral modulation), it would still be about 10 million. Given the fact that even a single overseas cable has a bundle of many fibers, up to about a billion simultaneous HDTV connections of any kind wouldn't be a problem.

The result of more and more bandwidth available is already becoming obvious: Far distance phone calls almost never go by satellite anymore, and they are already as cheap as local calls.
A recent study showed that only *2%* of the already available bandwidth of transatlantic fiber cables is currently in use !

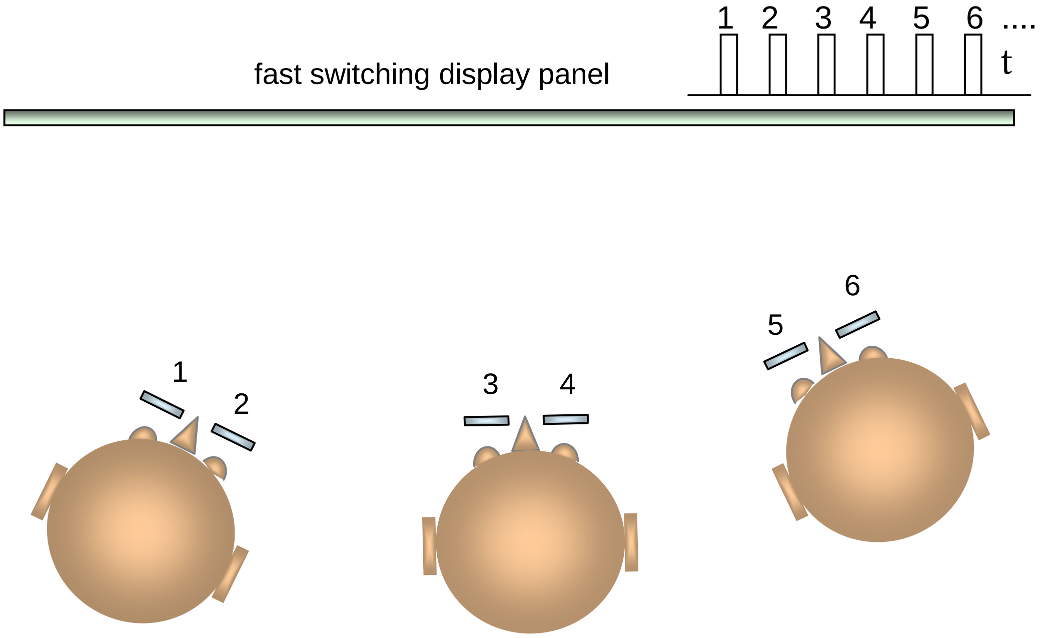
These millions of channels would make a permanent private HDTV connection affordable for anybody.

Who would need this, anyway ?

One possible application could be virtual households. Guess mom and dad live in Europe, grandma in the US and junior has a job in Japan. Or a company would like to have one or more permanent virtual connections between distant offices.
They could just cover a wall with display wallpaper, add some micro cameras with small lenses that hide in this large display area, and create a virtual wall breakthrough that would extend the room into another one at the other side of the world, 24 hours a day if necessary. If anyone used vision simulators, the display wallpaper could be omitted, of course.

In case of such virtual wall breakthroughs, or windows, most of the time the scene would be very static, and the display would have to generate pictures for existing viewers only (it would be possible to track them or use data from their shutter glasses or vision simulators to determine what they can see), so the originating source would have to send only that, and in high resolution it would be necessary only at the center of their field of vision.
Tracking is also absolutely necessary to generate the right perspective from the available camera perspective, otherwise the impression would not be entirely realistic.
With only necessary pictures transmitted, the average bandwidth requirement could in this case go well below 1 Mbit/second. Indeed, *billions* of high definition virtual wall breakthroughs could be held open permanently, at very moderate costs.

A pseudo holographic wall



With a display panel having a very short reaction time and being bright enough, one could switch between images for several viewers continuously.

Any viewer would only need simple shutter glasses.

Any eye of any viewer here gets its own image, so full synthetic holography is possible.

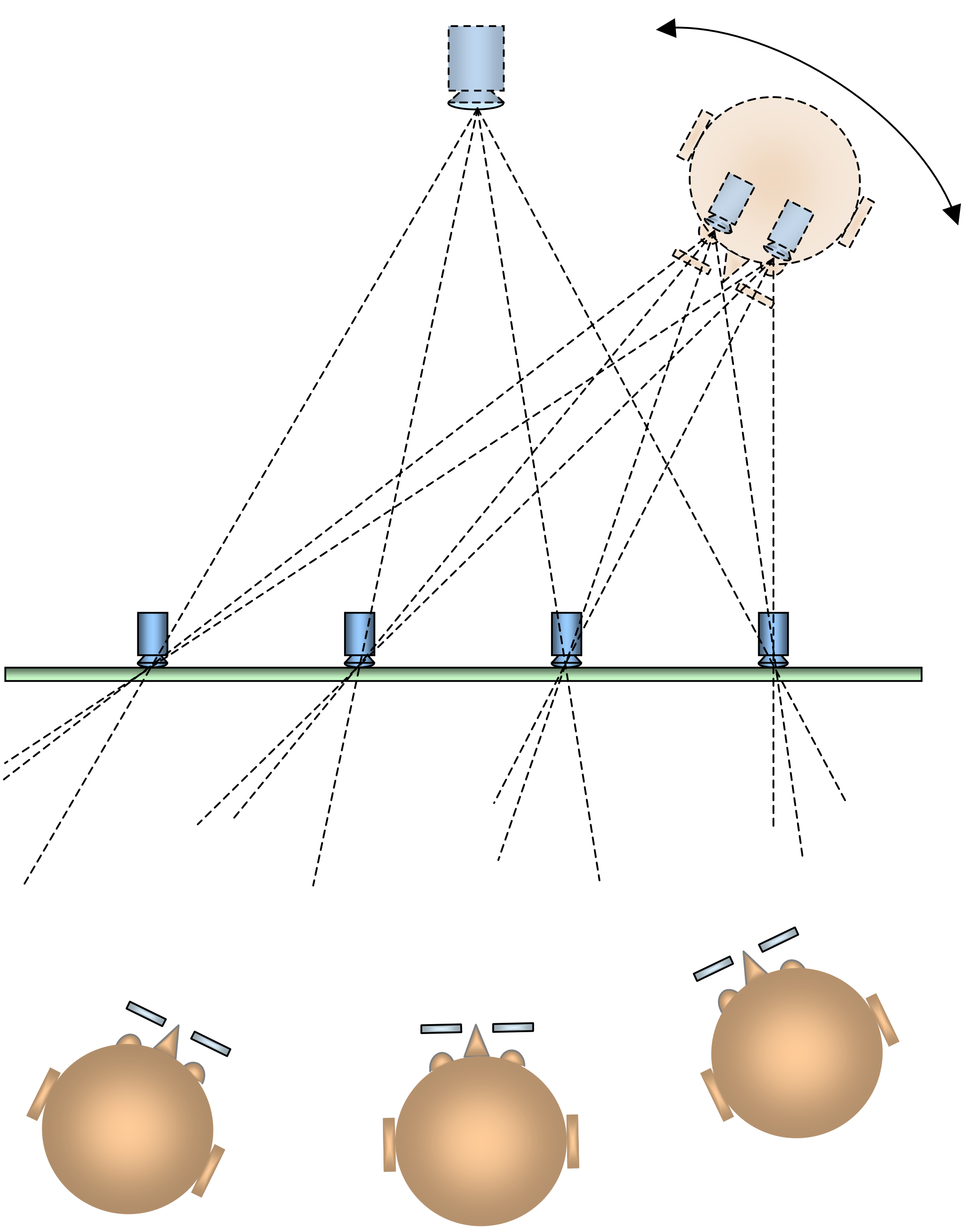
In this example, if the display simply switches from view to view, the shutter displays will only let pass less than $^1/_6$ of environmental light and therefore be like heavy sunglasses.

Better for this application would be a display that can provide short flashes, allowing the shutter glasses to just cut out those short flashes that are intended for the other viewers.

Giving every eye its own perspective would require head position sensors of course. This can be facilitated a lot by attaching some indicating devices to the shutter glasses.

A flashing IR lamp for example could easily be located by a simple camera, and the flashing sequence could encode the device number.

This assembly would so far be the only real alternative to full VR glasses. It still needs glasses and serves only a limited numbers of viewers, but those shutter glasses are extremely light and simple.



Perfecting the virtual wall breakthrough

The same wall display in a teleconferencing assembly. A camera that takes up the viewpoint of a participator at the other side is impossible to realize. Yet with a display this large, several pinhole cameras can be mounted in the panel without being noticed. There are always certain sections in the image of any of those cameras that resemble a section of the image a virtual camera behind the wall would have seen. By merging the appropriate image parts with some software help, the image of the virtual camera can therefore be constructed. This also applies to stereo viewing (two eye perspectives separately).

This way, we could also dynamically construct the right perspective for a viewer at any place at the other side of the wall.

If two rooms are equipped with this hardware, and we use a multi shutter display as described above, every participant gets a perfect stereo perspective all the time, e.g. the scene should look 100% real.

We could complete the assembly by a microphone array and a corresponding speaker array in the wall (flat speakers that fit behind wallpapers and even reproduce a lot of bass, are already on the market). Such a sound assembly would have pseudo holographic properties and make the illusion perfect.

Single person applications like this have been tried several times, one of the most recent being the Immersive Meeting Point by HHI [10], that implements some of the discussed techniques

Even with this, a virtual wall breakthrough with *real displays* will probably never be perfect:

- Auto stereoscopy is extremely difficult to achieve if many users have to be included.
- Switched displays still need glasses and are limited to a few users.
- Holographic displays are still more SciFi than achievable.

A realistic glasses-free approach could be a large holographic projection screen that takes light beams from different projectors and concentrates each of them on different viewer positions separately. Dynamically generating only the perspectives for certain viewers and switching them between projectors as viewers move, this could work for a limited number of persons. As with all current holographic auto stereoscopy, generating many images for the many different perspectives implies a lot of expense.

Screen based 3D: conclusion

We have taken a short but quite comprehensive look at 3D screen varieties, and all of them exhibit some problems

Many require to occupy a certain viewing position, or even to hold one's head straight. Even worse, many will cause headaches due to false perspective or focus.

A huge obstacle for general application is the inflexibility with display size and perspective. Modern media should work on a mobile phone display as well as on a wall sized screen.

The solution to the mentioned problems will definitely be computerized image processing. Computers have already overcome any format restrictions in conventional productions. Getting rid of formats, a generic description of recorded objects is called for, that later on allows to produce a device independent performance, and synthesize in real time whatever the desired display arrangement requires.

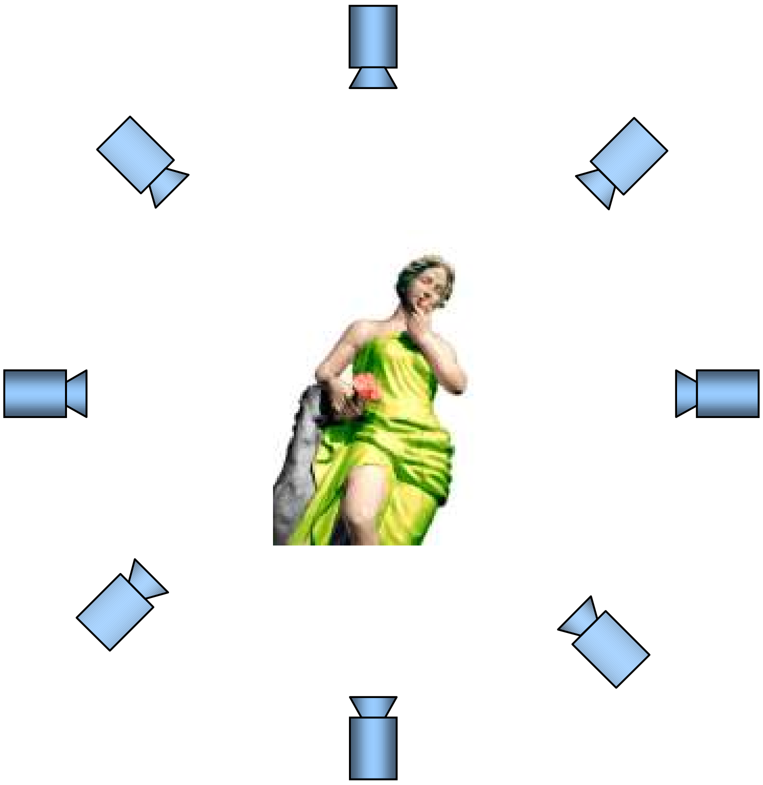
Large screens will only be widely acceptable for 3D usage, if any viewer gets his own custom generated perspective. Which in turn limits applications a lot.

I therefore anticipate, that the technology of choice will finally use VR glasses and integrate seamlessly into the wider and more general paradigm of a virtual environment including all sorts of virtual devices.

As soon as this technology is in widespread use, a virtual wall breakthrough, for example, will only need a camera array to be installed, all else will be just one more piece of software for already existing hardware.

One more bodily contribution to the implementation of virtual screens and windows may perhaps be a blackened wall in order to make the virtual objects insertion more easy. But that is not mandatory.

Optical surround
Beyond the screen



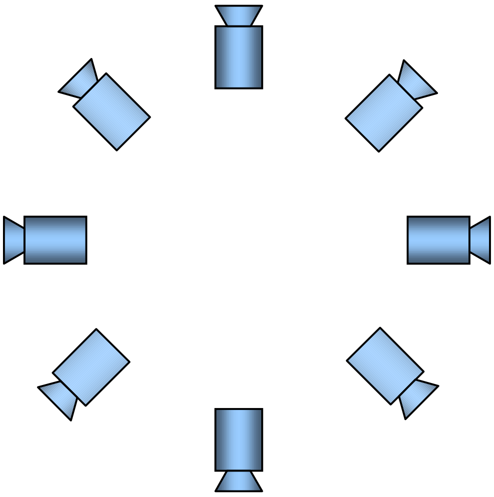
For special applications, it will be desirable to record objects, actors, scenes from everywhere, or from at least 180^0 up front.

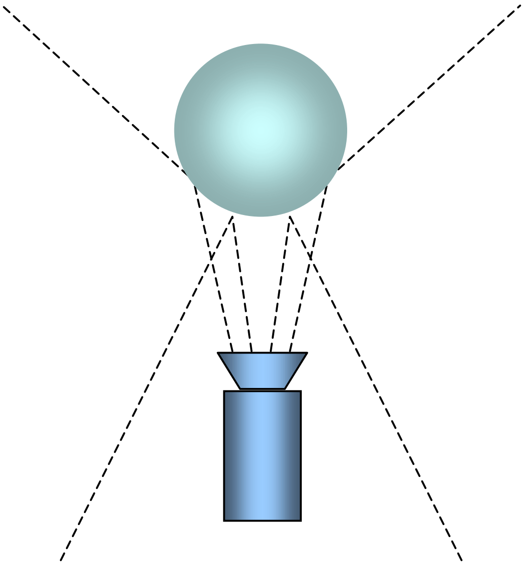
A chamber orchestra, a theater stage etc. would be good applications. This is only logical, if we use vision simulators, to get away from the screen paradigm and create a virtual stage that can be viewed from all around, for example.

It is necessary to do a lot of perspective calculations and object separation for this to work, and it will lead to an object oriented encoding of scenes, that would be similar to that of a scene synthesis like in 3D games. This is difficult but the logical course of further development anyway.

With such capabilities, the virtual orchestra could just sit in the middle of the living room, the virtual stage or the news speaker as well. Specific application patterns will evolve that we could hardly imagine today. It may seem as if this type of recording would not work with the holographic encoding scheme we have discussed. Where is the hologram 'screen' with surround technology?

We could conceive a virtual hologram screen as a *sphere* around the recorded objects. This would likely get very big, but then we could use a coarser resolution. Otherwise, as holograms can also show objects before them, we could also choose any smaller diameter desired, even *inside* the recorded objects.



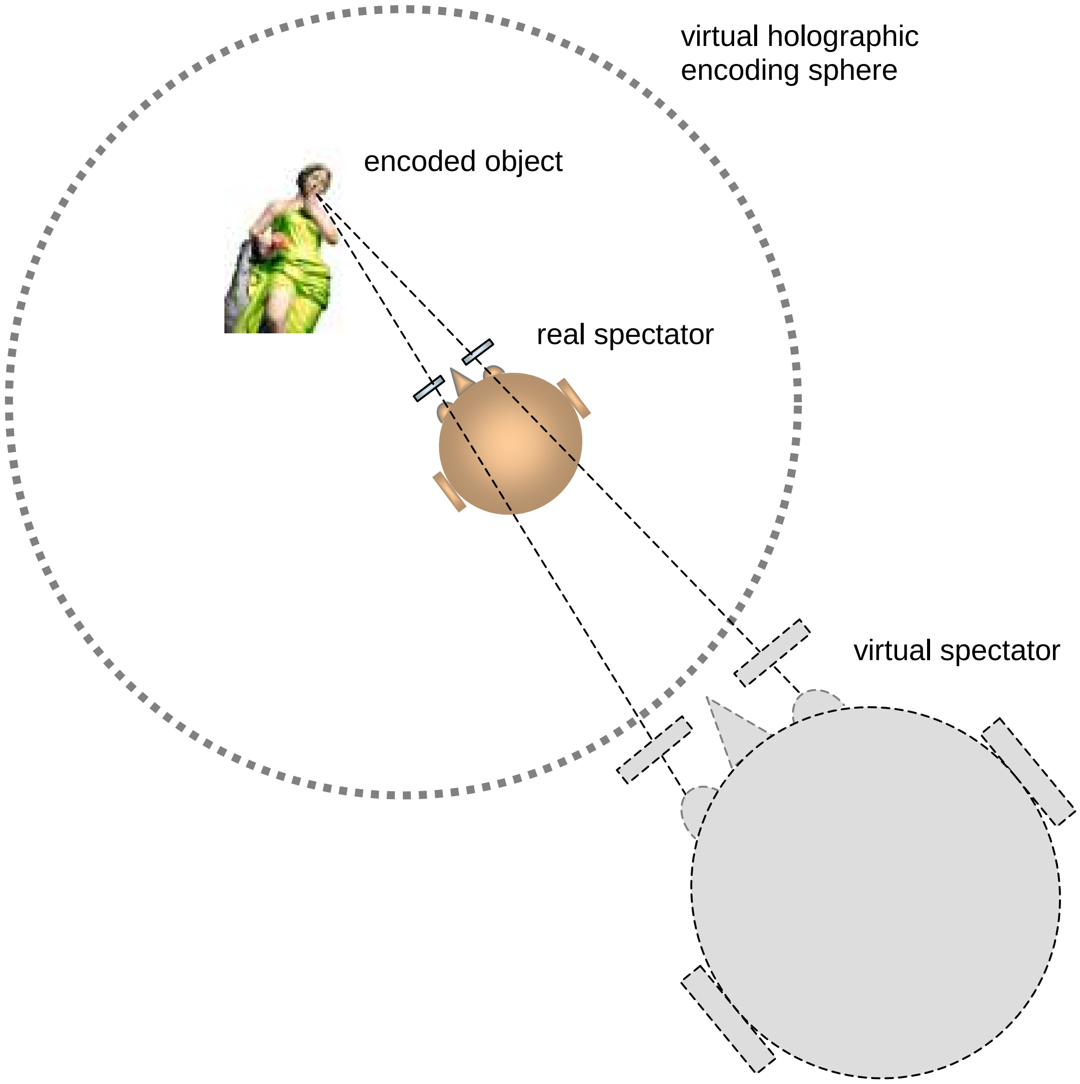


The other kind of surround video is available right now, and will become an everyday experience in the future. Either camera arrays or special optics (other than the simple mirror ball assembly shown above, there are a lot of sophisticated professional surround lenses available) can be used to record the required data. This type of recording should have many practical applications if vision simulators would become common equipment, as any place could then instantly be turned into a perfect surround theater and many program types like nature films, sports events etc. could really use it to produce an immersive experience.

We might even see both types of surround video in the same event, enabling us to walk in a virtual room and see the objects in it as well as the walls and ceilings on all sides in full 3D.

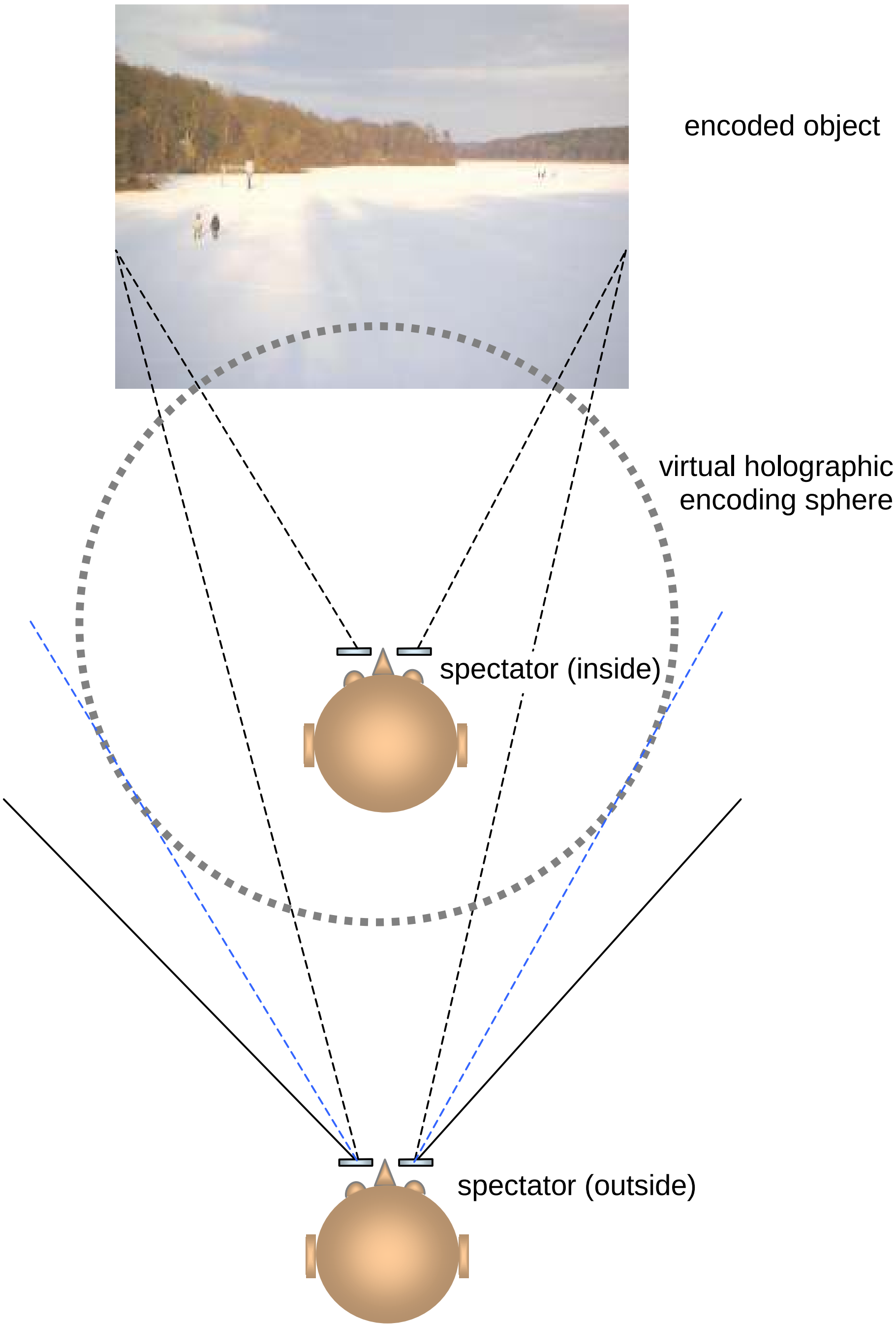
In this case we may chose the same large virtual hologram sphere to encode the outside and inside scenery. Outdoors scenes are also possible, of course.

The holographic circle



A virtual encoding sphere for the holographic method can be positioned far outside the depicted scene: The visual impression can later on be generated for a virtual spectator outside the sphere in a way that it entirely resembles the impression a real spectator inside the sphere would get, except for a different scale. The hologram patterns encoded could in this case be much coarser than with a small encoding sphere, according to the scaling factor involved, so the use of a large sphere does not increase the amount of data to be generated and recorded.

In and out of the sphere



The same virtual sphere could encode a scenery outsides of it. In fact, for a totally immersive cinema experience, it should be best to use a sphere as large as possible, because if a spectator gets outsides, we will still be able to generate images for him as long as the sphere is in the viewing angle, but not otherwise.

Virtual cinema

Today it's very dangerous to enter a US movie theater with a camcorder. Even only a suspected attempt of filming the movie could send you to jail. What if everybody had a personal vision simulator with all the cameras belonging to it? Would they collect all these items at the wardrobe ?

Or would movie theaters go out of business anyway, because everybody could perfectly simulate a big screen at home ?

The answers may be no.

If anyone had a vision simulator, movie theaters could just stop projecting films but instead transmit the picture by wireless communications. Viewers would get the picture by their own vision device. So everybody would always sit at the best place and the overwhelming expense for current Imax or surround projectors could be saved.

As the micro headphones of a vision simulator could produce quite good sound but not as well surround, and certainly not let the ground shake, movie theaters could invest in a good sound system (frankly, today's sound systems often are a torture, nothing but loud). Sound could even be diverted between the theater system and the user's headphones, to provide the same impression at every place.

The video transmission could be copy protected (recording inhibited) and encrypted. Admissions would not be sold in paper but as an electronic key sent to the user's device at payment.

Would anybody really visit such a cinema ?

Maybe yes. At least if the movie was new and couldn't be viewed anywhere else.

The technology would also offer new possibilities: Perfect 3D, just like holography, would be possible, and special effects like smells, moving seats, water splashing on the audience, like in some monster cinemas today, would be additional attractions.

So indeed cinema could survive the vision simulator.

Virtual media - conclusion

Cinema productions make heavy use of synthetic scenes already, and merge them with reality. These scenes are inherently 3D and object oriented, and enable us to render perspectives according to the viewer's own position rather than to a fixed camera position. These viewer perspectives could best be rendered and displayed in the viewer's own vision simulator glasses. This would, for the first time, deliver perfectly realistic 3D.

The bottleneck to this approach is the capability to record real scenes in a similar manner, i.e. to separate objects and gather their 3D surfaces, just like with the computer generated ones.

Therefore I did emphasize the necessity of multi channel multi position camera recording and 3D processing and encoding.

Virtual media is a very wide field, and so far we only scratched the surface a little, as far as it concerns the most important aspects in conjunction with our thematic.

Vision simulators could change our entire attitude and habits with media objects. For example, acting persons could come without image frame and background, our living room could become a theater stage or that of a concert hall. We could as well imagine total surround cinema, ideal for documentaries and outdoors themes, to be possible everywhere. If that isn't an attractive technology, then I don't know.

For the scope of this book, I think this may suffice. A little more is in the fiction chapter, and much more is left to your imagination.

The end of hardware – an outlook

In the course of this text it should have become obvious that the further development of computer communications and computer interfaces will inevitably include to move these as close to the brain as possible (i.e. towards the eyes and ears for the time being), which in turn will result in an ubiquitous and - hopefully - seamlessly integrated augmented reality.

I also hope to have proven that simplistic approaches may have been good for first experiments or specialized applications, but that the course now has to go towards more perfect implementations that - first and foremost - have to solve the problem of merging the virtual with the real.

This also implies lots of new applications, situations, usage habits, security problems, and so on, that couldn't realistically be addressed in a comprehensive way, even less in the course of a single book and surely not without a lot of experimentation and experience with the fully implemented technology. Nonetheless I hope my limited fantasy could shed some light on it and inspire the fantasy of others.
Even if there certainly are some contradictions remaining in the scenarios envisioned, especially as far as it concerns implications on society, laws etc., I think this may serve even the better to initiate, and contribute to, discussions on the thematic.

As with any technology, perfection will not be achieved in one step, so we have to expect many intermediate solutions and applications to come. An entire field that I did not really address here is about thinking out and implementing such intermediate products. It's of course about impossible to foresee such a certainly very complex development.
What I tried here instead, was to define the aim and the technologies necessary to approach the target, so we could easily determine the strategic relevance of this or that development for our final objective.

We have seen some applications that could do without one or the other feature, hence allowing to get to a product easier and faster.

The applications for vision impaired people discussed above, need about anything but the displays. So they are perfect test beds for camera based orientation systems, spatial sound synthesis, finger and maybe even eye pointing. The entire difficulty of high end main displays or mask displays is however avoided in this case.
Other varieties could need just these, but not sound: office applications, for example. So we could think about special office models.
For home entertainment, very simple models could be developed that would need the highest quality displays possible but could do without part or all of the spatial orientation functions. Other varieties of such devices could work with spatial orientation but without mask displays, instead requiring real black areas on the walls to insert virtual display screens.

For outdoor activities as well as driving, glasses without mask displays could be sufficient in many cases, with the advantage of lower light attenuation and some weight savings. These couldn't provide the intelligent sunglasses function of course.

A mobile phone integrated in glasses, as already described in the introduction, could be a great product and an ideal test bed for displays eye trackers, acoustic components and more.

The most difficult application area in the end would be night driving, if we want an anti blinding function, that of course wouldn't be possible without a mask display of very low light attenuation. It's also a tough application under security aspects, as we have to guarantee that nothing important is covered up unwantedly, or disturbed otherwise by the displays.

I think we have seen that there are many special applications that could be used to develop parts of the full featured vision simulator and address fairly big markets already. People may own

several different ones for different occasions, until it finally becomes easy to include all features into one device.
It will anyway remain hard to say what the further course of development will be, and how fast it will go.

Having thrived through so many aspects of the field, we could at least try to compile a short list of things to do:

1. A display design based on the presence of eye trackers, that allows for high resolution, viewing angle and brightness at minimum weight.
2. Technologies for a good mask display. This is even more basic research than (1). Very important.
3. Develop the best eye tracker possible, fully integrated with the display design.
4. Micro and nano technologies for laser, holographic and other displays as well as optical and mechanical elements.
5. Very low power image compression and decompression as well as serial communication chips.

There are also tasks that have already been performed somewhere but may have to be adapted, reinvented, revived or improved:

1. Anticipative motion sensing and rock stable rendering of objects.
2. Camera based position sensing and scene identification (not recognition).
3. Virtual keys. Add eye pointing !
4. Dynamic sound rendering.
5. High performance, low power, very light pocket computers.
6. All sorts of application software.

That's enough.

References

This is an individual selection from the many materials existing on the issue. It can't be comprehensive and isn't intended to be. It should be a collection of the most remarkable things I found, especially in the recent time. I also generally preferred to list materials that can be quickly accessed by the web. Some items are mainly included for completeness, because their title suggests that they would contribute to the theme but actually they might not, or use the same words for different things. You'll easily find out when checking abstracts or looking up book descriptions in online stores.

If there are keywords in the text that you'd like to know more about, you'll surely find something very quickly by a web search. If computer scientists were more up to date, you would also surely find lots of papers on the arxiv servers. Maybe next year.

Remark: I can't guarantee that the web links listed here will stay current. I'll try to list updates on http://www.theendofhardware.com.

[1] Rolf R. Hainich: Echtzeitdiagnostik und Sichtsimulation für die minimal invasive Chirurgie, 1993. Proposing Augmented Reality in medicine, as well as in mass market applications. http://www.theendofhardware.com/materials/Echtzeitdiagnostik.pdf

[2] Rolf R. Hainich: Integrative 3D Visualization. Conference opening paper, Workshop on Integrative 3D Visualization, Wiesbaden, 1994. A paper already defining and outlining all of augmented reality and technology solutions. http://www.theendofhardware.com/materials/i3p.pdf

[3] Steve Mann: Smart Clothing': Wearable Multimedia Computing and `Personal Imaging', 1996. http://wearcam.org/acm-mm96/

[4] Tim Suthau, Marcus Vetter, Peter Hassenpflug, Hans-Peter Meinzer, Olaf Hellwich: A concept work for Augmented Reality visualisation based on a medical application in liver surgery IAPRS, Corfu, 2002 http://www.cv.tu-berlin.de/forschung/AR/medizin_AR.phtml

[5] Henning Schaefer, Kalibrierungen für Augmented Reality. Diplomarbeit at Technical University of Berlin, 2003.

[6] IBM Continous Data Protection: ftp://ftp.software.ibm.com/software/tivoli/flyers/fl-tsm.pdf; http://www.redbooks.ibm.com/abstracts/SG246844.html?Open

[7] Gursel Sonmez et al., A RGB Polymeric Electrochromic Device, University of California, Los Angeles, 2003 http://www.wiley-vch.de/contents/jc_2001/2004/z52910_s.pdf

[8] Ulrich Hofmann, Beamer für die Westentasche, FhG-ISIT, 2004 http://www.fraunhofer.de/fhg/press/pi/2004/11/Mediendienst112004Thema6.jsp

[9] MPEG.org Report on 3DAV Exploration
http://www.chiariglione.org/mpeg/working_documents/explorations/3dav/report_on_3dav_explor.zip

[10] HHI Immersive Meeting Point: http://ip.hhi.de/imedia_G3/impoint2.htm

[11] George Orwell, 1984. Signet Book; Reissue edition (May, 1990), 336 pages, ISBN: 0451524934 (many other sources)

[12] Jeff Hawkins: On Intelligence, Owl Books, 2005, ISBN: 0805078533
http://www.onintelligence.org/

[13] US pat. 5,572,343, concerning LCD displays as light valves, http://patents.uspto.gov/

[14] MicroOptical Corporation
http://www.microopticalcorp.com/Products/HomePage.html

[15] Microvision, Inc. http://www.microvision.com

[16] James Fung: http://www.eyetap.org/~fungja/

[17] Wearable Computing Library: http://about.eyetap.org/fundamentals/

[18] Universal Display Corporation :http://www.universaldisplay.com/

[19] Oliver Bimber, Ramesh Raskar: Spatial Augmented Reality: Merging Real and Virtual Worlds, A K Peters, Ltd.,2005, ISBN: 1568812302
A very different approach, see also http://www.spatialar.com/

[20] Woodrow Barfield (Ed.), Thomas Caudell (Ed.): Fundamentals of Wearable Computers and Augmented Reality, Lawrence Erlbaum Associates, 2000, ISBN: 0805829016

[21] Tobias Hans Hollerer, User interfaces for mobile augmented reality systems, PhD thesis, 2004, Advisor: Steven K Feiner, Columbia University, Published Jun 2004, ISBN: 0-496-62842-6

[22] Bolan Jiang, Robust hybrid tracking for outdoor augmented reality, PhD thesis, 2004, Advisor: Ulrich Neumann, University of Southern California, Published Jan 2005 ISBN: 0-496-87626-2 (available as e-book)

[23] Grigore C. Burdea, Philippe Coiffet: Virtual Reality Technology, 2nd ed. with CD-ROM, Wiley-IEEE Press 2003, ISBN: 0471360899

[24] MIT wearable computing link collection:
http://www.media.mit.edu/wearables/lizzy/wearlinks.html

[25] Wearable Computing Library, eyetap:
http://www.eyetap.org/research/eyetap.html

[26] Virtual keys, US pt. 5,767,842 (1992/1998): http://patents.uspto.gov/

[27] Turner, Stuart L., Coupling Retinal Scanning Displays to the Human Vision System: Visual Response and Engineering Considerations. PhD thesis, University of Washington, 2000

[28] GP Markham: The ORIGINAL Illustrated Catalog Of ACME Products. *have fun !* http://home.nc.rr.com/tuco/looney/acme/acme.html

[29] EST-Engineering System Technologies. http://www.est-kl.com

[30] Tibor Balogh et al.: A Scalable Holographic Display for Interactive Graphics Applications. In Proc. IEEE VR 2005
http://www.crs4.it/vic/data/papers/ieeevr2005ws-holo.pdf
http://www.crs4.it/vic/cgi-bin/bib-page.cgi?id='Balogh:2005:SHD'

[31] Light Blue Optics / Cambridge University, Holographic projection display: http://www.lightblueoptics.com/
http://www.eng.cam.ac.uk/news/stories/pocket_projectors/

[32] Jürgen Herre et al.: An Introduction To MP3 Surround
http://www.iis.fraunhofer.de/amm/download/flyer/dl.html?f=introduction_to_mp3surround.pdf
http://www.iis.fraunhofer.de/amm/download/mp3surround/index.html

[33] CRLO Displays Ltd., microdisplays, http://www.crlopto.com/

[34] Claude Shannon, Communication in the Presence of Noise, 1948 Reprint: Proceedings of the IEEE, VOL. 86, NO. 2, Feb. 1998, also:
http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html
http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

[35] Oliver Bimber, L. Miguel Encarnação, and André Stork: Seamless integration of virtual reality in habitual workplaces - a website about some fundamental research projects in augmented reality.
http://www.uni-weimar.de/~bimber/research.php

[36] Michael L. Huebschman, Bala Munjuluri, and Harold R. Garner: Dynamic holographic 3-D image projection; see also:
http://innovation.swmed.edu/research/instrumentation/res_inst_dev3d.html

[37] C.E.Rash (ed.): Helmet mounted Displays in Aviation SPIE Press Monograph,Vol. PM93,ISBN0819439169, also: http://www.usaarl.army.mil/
http://www.usaarl.army.mil/hmdbook/cp_0002_contents.htm

[38] AR-NAV, FhG-FIT, http://www.fit.fraunhofer.de/projekte/arnav/index.xml

[39] Metcalfe, R.M., Boggs, D.R.: Ethernet: Distributed Packet Switching for Local Computer Networks. Coomm. ACM, 7/1976 (see also US Patent # 4.063.220, http://patents.uspto.gov/)

[40] Project ARVIKA: http://www.arvika.de/

[41] Tim Suthau: Augmented Reality – Positionsgenaue Einblendung räumlicher Information in ein See-Through Head Mounted Display für die Medizin am Beispiel der Leberchirurgie. Dissertation at TU Berlin, 2006

[42] Thorsteinn Halldorsson: Farbdisplays und holographische Bildschirme, EADS- Corporate Research Center, Ottobrunn. Darmstädter Kolloquium für Messtechnik, DAKOM 2005. http://www.eads.com

[43] Abramson, N.: The ALOHA System – Another Alternative for Computer Communications. AFIPS Conf. Proc. vol.37 (1970), p.281-285

[44] Lars Bönnen: FOHMD: The Fiber-Optic Helmet mounted Display and its Applications. CAE, Stolberg, Germany. In: Proceedings of the Workshop on integrative 3D visualization, Wiesbaden, 1994

[45] Jim Vallino's Augmented Reality Page (link collection): http://www.se.rit.edu/~jrv/research/ar/index.html

[46] Ronald Azuma's pages: http://www.cs.unc.edu/~azuma/azuma_publications.html

[47] Wikipedia article on Augmented Reality, contains many more links: http://en.wikipedia.org/wiki/Augmented_reality

[48] George Ou (http://blogs.zdnet.com/Ou) Is encryption really crackable? A short and thorough sweep-out of myths about 'cracked' encryption http://blogs.zdnet.com/Ou/?p=204&tag=nl.e550 [01.05.2006 22:29:28]

[49] Cyclon Systems, UK, manufacturer of wearable cameras for police, military and other services, http://www.cylonsystems.com/

[50] Stereo3d.com HMD comparison: http://www.stereo3d.com/hmd.htm

[51] Holoeye Photonics AG, spatial light modulators http://www.holoeye.com/spatial_light_modulators-technology.html

[52] LC Technologies, inc.: The Eygaze communication system, http://www.lctinc.com/PRODUCTS.htm

[53] Mirage Innovations, Manufacturer of the LughtVu glasses,http://www.mirageinnovations.com/

[54] Liteye Systems, producer of miniature displays, http://www.liteye.com/

[55] NVIS, manufacturer of head mounted displays, http://www.nvisinc.com

[56] Arrington Research, manufacturer of head mounted eye trackers. http://www.arringtonresearch.com

[57] Thomas Schnell, Applying Eye Tracking as an Alternative Approach for Activation of Controls and Functions in Aircraft http://cosmos.ssol.iastate.edu/isgc/RES_INF/VRR2000/Schnell_SEED.pdf

[58] Peter B.L.Meijer, vOICe, http://www.seeingwithsound.com/

[59] Larry Leifer and David Grossman: Blind Navigator: using object recognition to enhance blind mobility. http://mediax.stanford.edu/news/conference_nov03/dave_grossman.pdf

[60] eMagin head mounted displays, http://www.emagin.com

[61] Qualcomm, iMoD display technology, http://www.qualcomm.com/qmt/

[62] Jan Fischer, Interaktive Spezifikation von Domänen und Detektion partieller dynamischer Verdeckungen in Augmented-Reality Umgebungen; http://www.gris.uni-tuebingen.de/~fischer/janfischer.com/publications/fischer2002-thesis.pdf
J.Fischer and H. Regenbrecht and G. Baratoff, Detecting Dynamic Occlusion in front of Static Backgrounds for AR Scenes, EGVE, Zürich, 2003
http://www.gris.uni-tuebingen.de/~fischer/janfischer.com/publications/DynamicOcclusion.pdf

[63] Think-A-Move, Ltd., InnerVoice Pro
http://www.think-a-move.com/products.html

[64] Michael W. Halle, "Multiple Viewpoint Rendering for 3-Dimensional Displays", Ph.D. Thesis, Program in Media Arts and Sciences, MIT 1997, http://www.media.mit.edu/spi/spiPubs.htm
http://www.media.mit.edu/spi/SPIPapers/halazar/thesis-orig.pdf

[65] Mark Lucente, Diffraction-Specific Fringe Computation for Electro-Holography, PhD thesis, MIT1994 http://www.media.mit.edu/spi/spiPubs.htm
http://www.lucente.biz/pubs/PhDthesis/contents.html

[66] Cyberkinetics Neurotechnology Systems inc.
http://www.cyberkineticsinc.com/

[67] Donoghue Labs, http://donoghue.neuro.brown.edu/

[68] Privacy and Human Rights
http://www.gilc.org/privacy/survey/intro.html

[69] UN Universal Declaration of Human Rights
http://www.hrweb.org/legal/udhr.html

[70] Cryptool, free cryptography learning software by Deutsche Bank. Siegen University, TU Darmstadt, Secude, FZI: http://www.cryptool.com/

[71] Bernhard Esslinger (Ed), 2006 The Cryptool Script (cryptography):
http://www.cryptool.com/downloads/CrypToolScript_1_4_00_en.pdf

[72] Steven Feiner, links on user interfaces f. mobile & wearable computing, http://www1.cs.columbia.edu/graphics/courses/mobwear/reading.html

[73] Durand R. Begault, 3-D Sound for Virtual Reality and Multimedia
http://human-factors.arc.nasa.gov/publications/Begault_2000_3d_Sound_Multimedia.pdf

[74] Wayne Piekarski, Interactive 3D Modelling in Outdoor Augmented Reality Worlds, PhD thesis, http://www.tinmith.net/wayne/thesis/

[75] Electronic Frontier Foundation, working to protect your digital rights.
http://www.eff.org

[76] Thilo Womelsdorf et al.: Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nature, 2006
http://www.nature.com/neuro/journal/vaop/ncurrent/abs/nn1748.html

[77] Kiyoshi Kiyokawa et al. 2003, An Occlusion-Capable Optical See-through Head Mount Display forCollaboration of Co-located Multiple Users
http://www.hitlabnz.org/fileman_store/2003-ISMAR-occlusion_hmd3_kiyo_final.pdf

[78] Rolf R. Hainich, An Improved Ethernet for Real Time Applications, Real Time Data, Versailles 1982 http://www.theendofhardware.com/materials/ImprovedEthernet.pdf

[79] Rolf R. Hainich, Backoff Strategies for CSMA/CD with Real Time Applications, internat. conf. Kommunikation in Verteilten Systemen, GI/NTG 1983 http://www.theendofhardware.com/materials/CSMA-RT.pdf

[80] Home page of the Moving Picture Experts Group (MPEG) http://www.chiariglione.org/mpeg/index.htm

[81] Peter K. Kaiser, The joy of Visual Perception : a web book http://www.yorku.ca/eye/

[82] Eyes Tea, Berlin/Boston, http://www.eyes-tea.net/, with an interesting linklist at http://www.roetting.de/eyes-tea/who.html

[83] R. Ziegler, P. Kaufmann, M. Gross : A Framework for Holographic Scene Representation and Image Synthesis. To appear in ACM SIGGRAPH 2006 Sketch, Boston, USA, 30.July - 3.August, 2006 http://graphics.ethz.ch/Downloads/Publications/Papers/2006/Zie06a/Zie06a.pdf

[84] Aurora Systems Co.,Ltd., Manufacturer of high resolution LCOS micro displays, http://www.aurora-sys.com/home.htm

[85] Augmented Reality: Hyperlinking to the Real World, TechNewsWorld Nov.2006, http://www.technewsworld.com/story/54364.html

[86] Fraunhofer IDMT - Ultrafast 1-Chip-Eyetracker http://www.idmt.fraunhofer.de/de/projekte_themen/ultrafast_eyetracker.htm

[87] Zeiss - Head Mounted Displays project: http://www.zeiss.de/C12567A100537AB9/Contents-Frame/83F3D3C1FDFB5168C1256EA9002A8166

[88] SeeReal Holographic Display Technology http://www.seereal.com/en/holography/index.php

[89] Hinckley, K., Sinclair, M., Hanson, E., Szeliski, R., Conway, M., The VideoMouse: A Camera-Based Multi-Degree-of-Freedom Input Device, ACM UIST'99 Symposium on User Interface Software & Technology, pp. 103-112. http://research.microsoft.com/users/kenh/papers/VideoMouse.pdf

[90] R. Ziegler, S. Bucheli, L.Ahrenberg, M.Magnor, M.Gross: A Bidirectional Light Field - Hologram Transform. To appear in Computer Graphics Forum 26(3), Proceedings of Eurographics 2007. http://graphics.ethz.ch/Downloads/Publications/Papers/2007/Zie07c/Zie07c.pdf

[91] New Scale Technologies, Inc. - Manufacturaer of piezo electric micro motors. http://www.newscaletech.com/

[92] FhG-IPMS, Fraunhofer Institute for integrated photonic microsystems: micro scanners and micro mechanical light modulators. http://www.ipms.fraunhofer.de/en/products/microscanner.shtml

Index

Acknowledgements

There are many people who helped me with this effort. I can't honestly list them all, so I just wish to express my gratefulness to all of you, and for the reader I list a few of you here:

First of all, Sigrid, my beloved wife, who patiently endured me 'staring into the computer' for many months.

Walter Kroy, former head of research at MBB Aerospace, has supported the effort since the very beginning, chaired the first conference, and provided useful hints and links for this book. Thorsteinn Halldorsson of EADS has given me many helpful comments, especially on my holographic considerations. Lucilla Croce-Ferri of FhG-IGD/IPSI provided additional references on this issue and valuable hints about watermarks.

Dietrich Grönemeyer, who created interventional radiology, has supported my ideas already in their early stages. Georg Vallender and his colleagues at CAE have contributed a lot of motivation. Günter Hommel of Technical University Berlin, expert in robotics, didn't only give me that outstanding programming course very long ago, but also recently provided some interesting links.

Olaf Hellwich of TU Berlin enabled some important research in the field; Tim Suthau, who at his institute researched about visual adaptation and orientation, provided me with interesting materials about current near eye displays. Corell Nowak of EST helped with product pictures. A. Melhuish saved the English language.

David P. Keith, world's best flight instructor, did that gorgeous painting (p.93) on the cut out back of my T-shirt (after drying it).

Klaus Rebensburg of Hasso Plattner Institute and Uli Weinberg of Film Academy Babelsberg provided many inspirations with their 'n-space' invited lecture series.

The people at Booksurge LLC made this publication possible in the fastest and most innovative way. For a book about innovation, that's the way to do it, isn't it.

Finally, the honorable Balduin Egghead, who posed so patiently for many of my illustrations.

The Book

Draw these windows from your PC screen right into then air. Simply *look* to switch lights and open doors, to surf the web, to set up a virtual TV at your wall, a holodeck in your living room. See through things and see things about things. Never forget where you've put your keys anymore. See directions from your navigator, painted right onto the road. Replace that view from your window with the perfect one. Take your light sword and fight ghosts in the street, until they lock you up.

Virtual hardware**,** virtual objects will surround us, everywhere. It will work like magic. A single piece of real hardware will do it all, replace anything. Technology is ripe for it, now.

P*ersonal Augmented Reality* as we could call it, will develop a momentum as Personal Computing did, 30 years ago. Back then, white-coated switch panel gurus were pretending a PC would always be too expensive and John Doe couldn't use it anyway. Today, lack of fantasy and inspiration - especially in the hardware area - are producing the same baseless skepticism about Augmented Reality.

This is more than popular science - it is the blueprint for an entire technology, addressed to the interested public and to technology experts as well, explaining applications, technology, consequences, talking about vision technology and technology visions, showing advanced approaches like holographic optics and encoding, virtual media, and a great many more.

Even a science fiction part is included, giving the reader a hands-on experience right now.

The Author

Rolf R. Hainich is electronics engineer and computer scientist with many years of experience in complex hard- and software projects (real time processing and networking, media technology, communications, computer architecture, sensor technology, and more).

He also engaged in technology management, consulted to a large number of technology projects and companies, was chief consultant in public funding programs, started a private venture capital fund, joined the board of several high tech companies, authored many papers and studies.

Since the early 90's, he developed the idea of virtual devices, wrote papers and initiated conferences on the issue. This book goes far beyond.

http://www.theendofhardware.com

ROLF R. HAINICH
THE END
OF HARDWARE
A NOVEL APPROACH TO
AUGMENTED REALITY
2^{nd} Edition

BOOKSURGE LLC

ISBN 141965218-4
90000
9 781419 652189