Lasse Scherffig

It's in Your Eyes

Gaze Based Image Retrieval in Context

Zentrum für Kunst und Medientechnologie Karlsruhe

Lasse Scherffig

It's in Your Eyes

Gaze Based Image Retrieval in Context

ZKM | Institute for Basic Research
In cooperation with Media Art Net
Edited by Hans H. Diebner
2005

Center for Art
and Media
Zentrum für Kunst und
Medientechnologie
Karlsruhe

Lasse Scherffig

It’s in Your Eyes

Gaze Based Image Retrieval in Context

This book has originally been submitted as a requirement for a M.Sc. in Digital Media at the University of Bremen in March 2005

It has been supervised and reviewed by:
Prof. Dr. Frieder Nake, University of Bremen
Dr. Hans H. Diebner, ZKM | Institute for Basic Research, Karlsruhe

Lasse Scherffig has been supported by a scholarship of the ZKM

Layout by Anneke Kleine-Brüggeney and Lasse Scherffig

ISBN 3-928201-34-4

Contents

It's in your eyes
I can tell what you're thinking
My heart is sinking, too
It's no surprise
I've been watching you lately
I want to make it with you

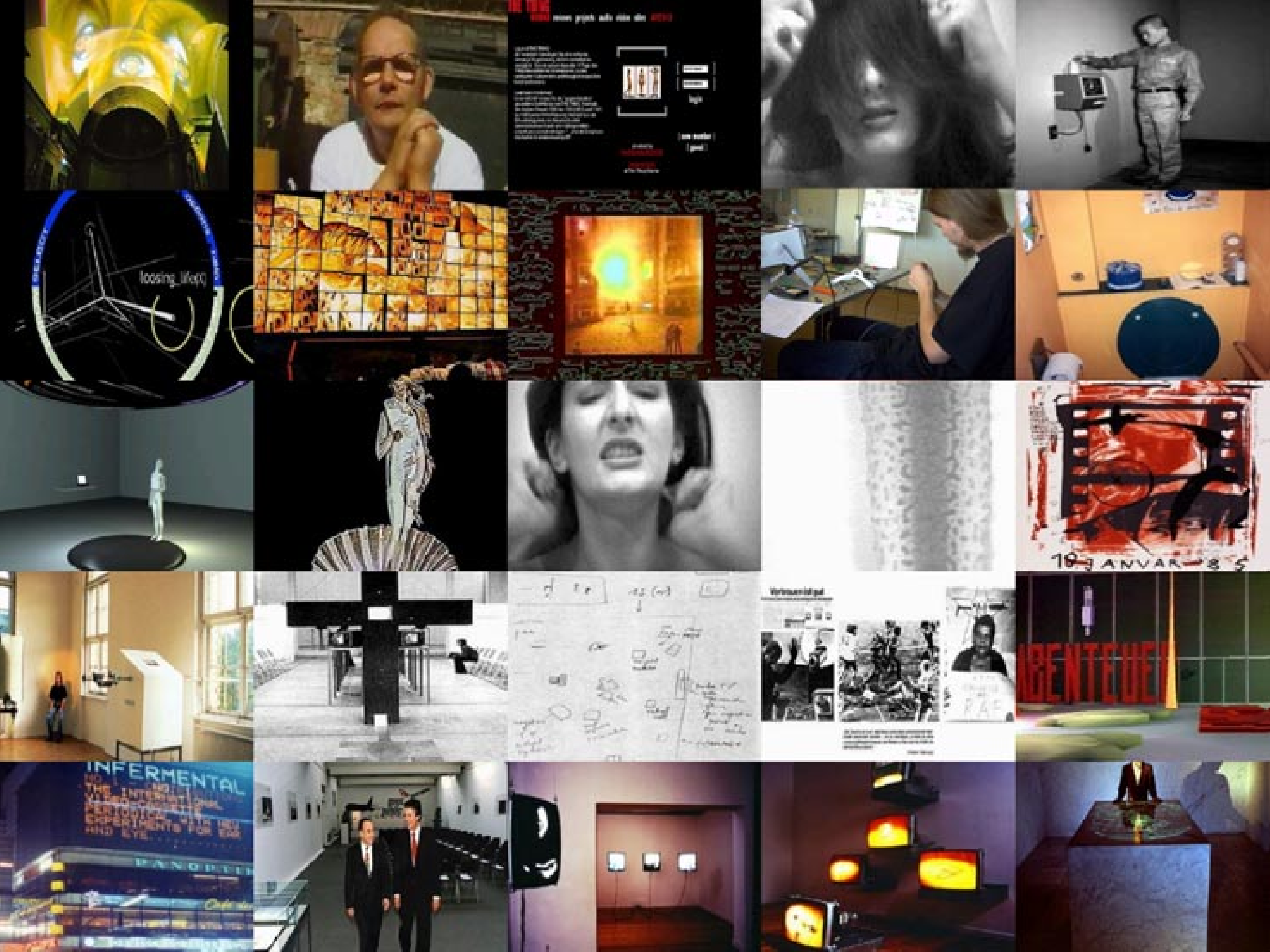
[Minogue, 2002]

Ich würde fast vermuten, dass bei diesen Abtastungen der berühmten Augenbewegungen [...] eben nicht mehr der Mensch der Beobachter ist, sondern zum ersten Mal der Beobachtete wird. [...] Augen und Mund, die ja weithin als letzte Refugien der sogenannten Intimität gelten, werden also eingespeist in eine Rückkopplungslogik, die nicht unsere, sondern die der Maschine ist.

In case of those scans of the famous eye movements [...] I almost would assume that man is not the observer anymore but for the first time becomes the observed. [...] Eyes and mouth, which widely are considered the last hideaways of so called intimacy, are hence fed into a logic of feedback which is not ours but that of the machine.[1]

[Kittler, 1992, pp. 75–76]

1 Rough translation from the German original by the author. Chew on the idea that we will be eaten (eingespeist or *fed*).


loosing_life(x)
INFERMENTAL
EXPERIMENTS FOR EAR
AND EYE

Rudi Hinterwaldner – a bricoleur in action

0 Introduction

0.1 A Method Against Method

The master thesis, the examination regulation says, has to show that the student is able to develop a new solution to a problem from the field of digital media [Prüfungsordnung, 2003, §14]. The expectation behind such a sentence is an expectation of our time: Working is solving problems. Publications around computer science and other disciplines that do some sort of engineering often stick to this pattern: A problem is described and then the work carried out is presented as unique (optimal, usable, efficient, fast) solution for that problem.

This method ignores the twofold nature of descriptions: Any description can work as a definition and vice versa. Or, in other words, descriptions are productive[2]. This lesson has been important for a great deal of contemporary philosophy. It is especially to be found in deconstructive thought. Judith Butler, for example, builds much of her critique of feminism on the idea that the descriptive categories of gender are constituting the very idea of gender identity – they are constituting the identity whose result they are said to be [Butler, 1990, pp. 37–49].

Following that idea, descriptions of a problem – be it in computer science or elsewhere – most likely already entail the solution. They also entail the way work is carried out. Often they are formulated having the solution and a specific method in mind. Having realized this problem for the idea of problems, Terry Winograd and Fernando Flores even decided to replace the term 'problem' by the term 'breakdown'. The former being loaded with the idea that problems objectively exist, the latter acknowledging the subjectivity induced by context dependency (or being-in-the-world as Winograd and Flores would say referring to Heidegger) [Winograd and Flores, 1986, p. 77].

2 The descriptive sentence stating that all ravens are black, for instance, not only entails what is and is not a raven. It also entails the way *how* to look for ravens.

One of the prototypes of
Eye-Vision-Bot

It was Jacques Derrida who mentioned that the engineer is a myth. Real work as it is carried out, he says, is 'bricolage', a do-it-yourself style of work that entails using the means currently at hand – no matter if they were made for this purpose or not [Derrida, 1972, p. 125]. Taking over the idea of bricolage as method from Claude Lévi-Strauss, Derrida radicalizes it by arguing that any engineer is merely a theoretical creation of a bricoleur.

Interestingly computer science seems to be *the* field where the top-down, engineering approach (dubbed software engineering) contradicts the actual bottom-up practice in which work is carried out. Only in the last few years, this bricolage nature of software development has been taken serious by approaches such as extreme programming.

That is what the following most likely is: bricolage. At the heart of this there are two things: two years of studying digital media at different places and *Eye-Vision-Bot*. The latter being some hardware, some software and some important concepts. Partially developed by me, mostly by others.

This text follows from the reaction of both. In any case it assembles ideas around that topic – subjective as any text is. In the best case it shows a perspective of digital media as and in context.

While context, as understood here, is fundamentally cultural this is not about uniting the famous two cultures: the natural and cultural sciences. Instead I will try the nondisciplinary approach.

The context – or rather the multiple contexts – of this thesis can be seen in its threefold structure which is marked by art, science and the interface. Of course, this structure does not work. Defined by the imaginary engineer it only serves as the point where the bricoleur starts working.

This master thesis, I therefore say, naturally follows from confronting my digital media studies with *Eye-Vision-Bot*[3].

Others

Eye-Vision-Bot was conceptualized by the physicist Dr. Hans H. Diebner and the psychologist Sebastian Fischer. Its software was designed and programmed by Sebastian Fischer and Lasse Scherffig. The museum setup was planned by all people involved, mocked up by Sebastian Fischer and Lasse Scherffig, prototyped by the designer Ruth Weber, prototyped again and finally build by the sculptor Manfred Schmieder.

3 Note that here I present my own approach as a reaction (solution) to the fact that problem descriptions are productive (problem).

0.2 Eye-Vision-Bot

Form... Since September 2004 the media museum of the ZKM shows an installation that is based on eye tracking[4]. It is called *Eye-Vision-Bot*. It consists of a seat, a table and a projection. The table hosts a device which combines a mask and a camera. Both are set up in a way ensuring that the camera always faces the right part of the mask – the place where the right eye of anybody looking through it would be found. The height of both can be adjusted – relatively to the table – by turning a small crank. On the table there also is a small green button.

If unused, the projection shows some pictures and some text. Both together are intended to explain which kind of behavior the installation requires in order to function properly. A visitor obeying these instructions will sit on the chair and adjust the height of mask and camera to a level allowing her to look through the mask without feeling terribly uncomfortable. She then will press the green button.

Now the projection will change. A short text will give further instructions. Then little white dots will appear, one after another. The visitor will focus each dot until the next appears. Nine dots later the projection will change again.

A grid of five times five pictures will fade in. These include photographs, sketches, texts and screenshots. The visitor will soon realize that each picture she lays her eye on is growing towards a size somewhat bigger than the rest and is becoming brighter. Once left by the eye it will shrink again, another one will grow.

The game will continue. After some time the grid will fade out and be replaced by another one. Same size, different pictures. These will be – some obviously, others not – related to those the visitor found most interesting

4 For the moment, eye tracking is some method of finding out where someone looks at by looking at her eyes. We will learn more about it later.

ZKM

The ZKM is the *Zentrum für Kunst und Medientechnologie* or *Center for Art and Media*. Located in Karlsruhe, it comprises a media museum and several institutes. Purpose of the latter is production and development in the field between media and the arts. The *Institute for Basic Research*, where *Eye-Vision-Bot* has been created, is one of them.
See `http://www.zkm.de` and `http://basic-research.zkm.de`

before. She eventually will find out that she is searching. She, in fact, already started finding.

Eye-Vision-Bot in the museum

...and Function Hidden inside a wall of the museum there are three computers: one responsible for the projection and coordination of the others, one holding all images and databases that are involved and one that is tracking eyes via the camera.

While nothing happens, the computer doing the projection is displaying an explanatory screen. As soon as the green button has been pressed, the event management system[5] on this machine initiates two processes: one for calibration of the eye tracker the other for image retrieval. Both yield some network communication:

The computer connected to the camera is signaled to start the calibration process. It responds sending coordinates for calibration points. These are displayed, one after another, on the projection. The language, or protocol, both speak is UDP. The latter process starts a new thread. This communicates with the computer holding the images and databases. Their communication at this point is carried out in MRML (see p. 26) – based on TCP/IP. Some random images from the *Media Art Net* (or MAN, see p. 12) database are requested.

While the calibration is running, requesting images is finished and the identifiers of a number of images arrive at the computer responsible for the projection. These identifiers are fed into a new thread which downloads them from the computer holding the images. HTTP is the language for that.

Calibration successfully finishes. The computer doing the projection is notified. It then turns the images that just were downloaded into *openGL* textures and places them on the projection to fade them in. At the same time, the eye tracker constantly sends coordinates of the gaze currently measured. When fading in is done, the image that currently is under these coordinates is zooming and getting brighter. In addition, the viewing time

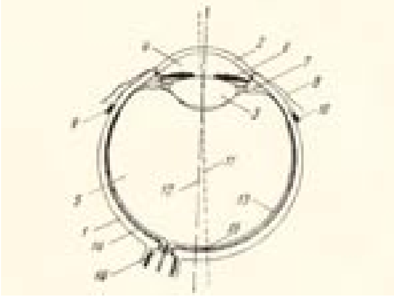
5 Event and window management, multi threading and some more routines stem from the wxWidgets framework, an open source project offering extensive functionality for the C++ programmer (see `http://www.wxwidgets.org`).

– the total time an image has been watched so far – for each image is recorded.
After some time – while still images are shown[6] – new images are requested based on the viewing times. Some via a query by example – a query for images that are structurally similar to those most looked at. Others in relation to the categories the pictures that were watched have in the MAN database. Thus, some of the new images stem from content based image retrieval, others from category (or keyword) based search. For each image, content based image retrieval also returns a relevance value – a scalar value denoting how much the image returned is similar to the images the queries were based on. Again, extra threads are used for requesting and downloading them. Again MRML and HTTP is spoken.
Downloading finishes and the images are faded out. Once fading is finished all old textures are deleted and the new images (mixed together and sorted by relevance) are turned into textures and faded in.
As soon as the green button is pressed again, a new calibration starts. As soon as for some time no eye has been detected by the eye tracker, the system returns to the initial loop.
Everything – which image is watched how long while which others are shown, which algorithm yielded what result, etc. – is stored in the central database[7].

In form and function, *Eye-Vision-Bot* appears as an image search system (or image retrieval system, as some would say) that is controlled by gaze. Being built in a modular way, many variations of what is described above are possible. Being discussed in context, it will be treated as work of art, scientific system and interface. Its central idea always is the computer reading ones wishes – the images one seeks – from the eyes.

6 The main reason for having a multithreaded architecture is being able to search for images while still images are shown and zoomed.

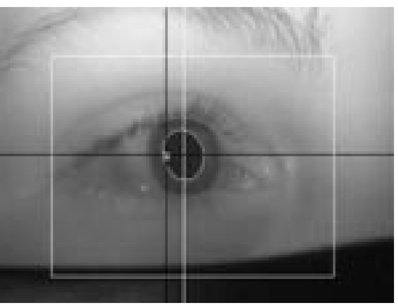
7 If you are interested in technical details, a good start is downloading the open source program *openbaar* at `http://openbaar.sf.net`. Most of the technology of *Eye-Vision-Bot* has been directly transferred into it, making it a good start for understanding how *Eye-Vision-Bot* has been implemented.



The eye: Number 2 marks the cornea, the retina is 13 and the fovea 15.

0.3 The Eye

The human eye has been of interest for the sciences as well as for the arts for a long time. Part of this interest has always been directed at what it is looking at. To understand that, its outermost layer as well as some of its internal features are important.
Between the eye and the world there is the cornea, a transparent layer that, as the lenses below, is responsible for a part of the refraction that is necessary for focusing. It, besides, protects the rest of the eye from the world. Seen from outside it forms a part of a sphere. If light is thrown onto the eye and hence onto the cornea it partly is reflected. Due to its spherical form these reflections always are at the same place no matter where the eye is directed at. Corneal reflections are invariant against eye movements. These characteristic reflections are called Purkinje images and are very important for eye tracking [Duchowski, 2003, p. 60].
Transduction from light to neural activity happens at the other end of



An eye as seen by the camera of *Eye-Vision-Bot*: The black cross marks a Purkinje reflection

the eye: at the retina. For the neurobiologist the retina is a structure made of different kinds of cells. Most of these are nerve cells – anatomically the retina even is seen as part of the brain. Among these there are two kinds of cells that react to light. These are called rods and cones.

To cut it short, rods and cones react to light differently. Moreover, they are not equally distributed: Cones are mostly to be found at the fovea – a small central part of the retina –, rods are found in the surrounding areas. Because of their differences in the way light is transduced as well as in their distribution, the retina has some important features: At the fovea spatial resolution of what is perceived is quite high but temporal resolution is low. In addition, color is mainly perceived there. In contrast, the periphery shows no color vision, a high temporal and a low spatial resolution.

To cut it even shorter, there is a small part of the retina that can perceive detailed and colored images while the rest can mainly perceive movements without color or details[8].

This yields a central problem: The part of the visual field that is perceived in color and at high resolution is very small. The angle in which foveal vision takes place lies between one and five degrees [Duchowski, 2003, p. 17]. For *Eye-Vision-Bot*, for instance, from these figures follows that no more than seven percent[9] of the projection can be perceived at a time. Because of that, the eyes are in constant motion scanning the scene they perceive. Because the temporal resolution of foveal vision is quite low, however, this motion has to stop frequently. Eye movements are hence characterized by short and very fast phases of movement followed by longer pauses. The former are called *saccades*, the latter are named *fixations*.

Besides saccadic movements there are other forms of eye movements serving other purposes[10]: These aim at keeping a moving target in focus or keeping the current image on the retina constant while the head moves. As opposed to saccadic movements they are smooth and comparably slow.

There seems to be no doubt that conscious interest and attention are closely related to saccadic eye movements. We look at what is interesting. How much of the process controlling this is happening consciously or attentional remains questionable. A mixture of preattentive and attentive processes is thought to control saccadic eye movements [Kandel et al., 2000, p. 502]. Their relation and “the neural mechanisms of attention and conscious awareness are one of the great unresolved problems in perception and indeed all of neurobiology” [Kandel et al., 2000, p. 504].

8 See [Kandel et al., 2000, pp. 507–522] for details of retinal processing and [Kandel et al., 2000, pp. 548–571] on perception of form and motion.

9 Given a distance of 4 m to a projection of 2.7 times 2.1 m (as in the museum setup) and an angle of 5 degrees. For 1 degree this number drops to less than 1 percent.

10 See [Kandel et al., 2000, pp. 782–800] for details on eye movements and their control.

1 Art

1.1 Eye Tracking as Art

As a piece shown in a collection of media art *Eye-Vision-Bot* is a work of art. This is not the first time eye tracking has been used in the arts.

At the latest with the arrival of media art as concept the observer was moved into the center of attention (see p. 15). Since the fine arts over decades centered around the image and media art – intentionally or not – is part of this tradition, the role of the observer's eyes for media art cannot be overestimated.

Tracking the eyes of the observer makes a promise: By looking at the organ reception of art is thought to be carried out by, reception of art and aesthetics can be understood. Artists hence will eventually be able to look through the eyes of those watching their work. A textbook entitled *Cognition in the Visual Arts* from 1996 promises the same: "The eye" it says, is "a window to the mind" [Solso, 1996, p. 132]. This idea will be of importance for the rest of this thesis.

MAN

MAN is the *Media Art Net*. It is a bricolage mainly consisting in a website (`http://www.mediaartnet.org`) and some books. Its objective is providing various forms of access to media art and its theoretical context. Its image database is what people see when they use *Eye-Vision-Bot* in the museum. The *Media Art Net* was created by various institutions, the ZKM being only one of them.

Also of importance will be a dichotomy: Eye tracking, for Andrew Duchowski, either is *interactive* or *diagnostic* [Duchowski, 2003, p. 131]. As any dichotomy I will use this one in spite the fact that it can be questioned. In this particular case Robert Jacob and Keith Karn, for example, state that both ways of usage "turn out to be analogous" [Jacob and Karn, 2003].

Diagnostic eye tracking has been applied to reception of art since the 1930s. It took some more time to transfer this approach to creation of art.

Some of Jochem Hendricks' *Augenzeichnungen*

An early work of art doing so are the *Augenzeichnungen* (roughly: Eye-Drawings) presented by Jochem Hendricks in 1992 [MAN, Augenzeichnungen]. These are a series of paintings that were created by recording eye movements and then plotting the recorded trajectories. While one can say that these turn the eye – "the organ of perception" – into an "organ of expression" [MAN, Augenzeichnungen] this transformation does not happen in realtime and therefore resembles the diagnostic approach.

The machine behind the *Augenzeichnungen*

A pseudo child-drawing created by a grown-up scientist using his eyes

Data is recorded and turned into something. In science this something often is statistics; in the arts it is a picture.

The connection between the pictures of the series and scientific eye tracking becomes most obvious by comparing the pictures of Hendricks with the figures provided by famous eye tracking studies such as those of Yarbus (see p. 21). Hendricks' act does hence not consist in developing a form of expression – a visual language. The act consists in realizing the artistic power of scientific images from that field. It consists in appropriating and recontextualizing their aesthetics.

For that reason the *Augenzeichnungen* function on two levels: They function as images as well as they function by appealing to the idea of looking through the window to the mind. The way we perceive is given form. One can pretend that "the invisible is made visible by means of a trace" [MAN, Augenzeichnungen].

Others in the context of fine arts have worked along the same line. John Tchalenko has analyzed eye movements of painters while drawing images [Miall and Tchalenko, 2001] – changing the object of observation by trying to understand creation instead of reception. He has performed research in painting with the eyes, too [Tchalenko, 2001]. The images produced in these experiments seem to quote the *Augenzeichnungen*.

Stripping off the ideas of understanding either reception or creation of art there also is research on utilizing the eye tracker as a device to enable disabled people to paint [Hornof et al., 2004].

But back into the past: With probably the first interactive eye tracking installation, the group *ART+COM* takes another path. In the installation *Zerseher* (roughly: dis-watcher) from 1991 eye movements are recorded while a painting which is shown on a screen is watched. The recorded data then is used – in realtime – to distort the picture where it is looked at. The picture that becomes dis-watched this way is *Boy with a child-drawing in his hand* by Francesco Carotto; being "the first documented child-drawing in art history – an adequate metaphor for the state of computer-art at the early 1990s" as *ART+COM* say [MAN, Zerseher].

The approach in this piece, besides focusing on interaction, is very playful. In the words of *ART+COM*: "In the past an old master might leave an

John Tchalenko's studies on drawing and writing with the eye impression in the mind of the passive onlooker – now the onlooker can leave an impression on the old master" [MAN, Zerseher]. Nevertheless, a central topic of interactive media art – observing the roles of work and observer – can be seen here.

Remarkably, the *Zerseher* is made at about the same time as eye tracking was introduced for military command and control. While the *Augenzeichnungen* can be seen as a retrospective appropriation of eye tracking as research, *ART+COM* are at the cutting edge of development. And while those studying gaze for command and control define the *Midas Touch Problem* (see p. 32) *ART+COM* make art of this problem.

Dis-watching
Boy with a child-drawing in his hand

1.2 Closed Circuits

The *Media Art Net* classifies *Eye-Vision-Bot* as a closed circuit installation [MAN, Eye-Vision-Bot]. Closed circuit art borrowed its name from closed circuit television. When systems of video cameras and screens that are directly connected by wires needed to be named the latter term was coined. In such systems generally record, transmission and reproduction of what is filmed happen in real-time [Helfert, 2004, p. 190]. This, in contrast, does not necessarily hold for mass media. While the latter was called broadcast television or open circuit television [Kacunko, 2004, p. 51], the term closed circuit television (CCTV) was introduced do denote such smaller and directly interconnected systems.

A special case of closed circuit television are the ubiquitous camera surveillance systems of our times. Since they are the dominant form of CCTV the term CCTV is often used to denote them[11]. As a closed circuit installation – we may note for the first time – *Eye-Vision-Bot* is a surveillance system.

Already in the 1970s closed circuit television was discovered as material for the fine arts. Bruce Nauman is one of the first to use this technology

11 Responding to the growing amount of CCTV surveillance, there also are art projects directly dealing with it: The *Surveillance Camera Players* – staging actions in front of CCTV systems – and *iSee* – a system for automatic calculation of surveillance free routes through Manhattan – are amongst the most famous. See `http://www.notbored.org/the-scp.html` and `http://www.appliedautonomy.org/iSee` (both accessed 11-01-05)

Live Taped Video Corridor
by Bruce Nauman

in installations. For his work *Live Taped Video Corridor* [MAN, Video Corridor] he placed two monitors at the end of a long and narrow corridor. One monitor always shows a recording of the empty corridor the other shows the picture of a camera placed at the corridor's begin. Approaching the monitor now implies creating a shrinking image of oneself on it – as it means going away from the camera. Nauman with this piece paradigmatically stands for a confrontation of the observer with "technologically mediated self-perception", as Heike Helfert notes [Helfert, 2004, p. 190]. A paradigm radicalized, for instance, by Peter Weibel in *Observation of the Observation: Uncertainty* [MAN, Observation]. Here three cameras and monitors are ordered in a circle in a way that an observer inside this circle will only be able to see herself from behind – no matter how much she twists and turns herself. Setup and title of the installation not only touch the question of the mediated self. They also directly link to the question of how – from the perspective of system theory – systems can observe and understand themselves.

Someone observing her observation

For the art historian Dieter Daniels such installations mark a transition. They clearly are pieces of interactive art but with closed circuit and related forms of art the role of interactivity fundamentally changes. Between the 1960s and the 1990s, he says, a paradigm shift takes place [Daniels, 2000, p. 172].

Starting in the 1960s, Daniels argues, Intermedia and Fluxus art focused on the social. Interaction hence was understood as something happening between people and therefore as something inherently political. Technology in this context only appears as a means towards a social and cultural utopia.

Until the 1990s this role was reversed – it has "been turned on its head" [Daniels, 2000, p. 174]. Now technology became an *a priori*. Since then, the social as well as culture are to follow from it. Simultaneously the observer in closed circuit installations as described above is not an active and equal participant anymore. She rather is a proband in an experiment. The topic of these works is not participation but reflection. Reflection of the relation of medium and observer. Reflection also of the self-referential system of the fine arts.

It is no accident that both the *Augenzeichnungen* and the *Zerseher* stem from the early 1990s. One recurring on scientific experiments and both being pieces that reflect on reception of art.

The attitude behind this kind of interaction is, Daniels says, "an attitude producing the very opposite of creative participation, namely the radical conditioning of a viewer through a work". [Daniels, 2000, p. 176]. As part of this tradition *Eye-Vision-Bot* asks people to condition themselves – it produces behavior.

The character of scientific experiments also defines those closed circuit installations that are not interactive at all. These consist of media that instead mediating are turned towards themselves. Most people working with a video camera will at least once have turned it towards the screen showing its picture. They then created a closed circuit (in a sense also a mirror hanging across from a mirror is one[12]). In such experiments, one can say, the subject is not the observer but technology itself.

For his installation *Two Cameras* [MAN, Two Cameras] Dieter Kiessling placed two video cameras facing each other. The area each one films is mostly the lens of the other. The auto focus function of each camera constantly tries to focus the other lens which unfortunately is in motion since it tries to do the same. The result is a complex dynamics of movements and pauses.

This illustrates an important fact that holds for many closed circuit installations: They create complex behavior from feedback. We may note that the same holds for *Eye-Vision-Bot*.

But in the context of such installations – where the circuit is evident in the technological setup – the closed circuit nature of *Eye-Vision-Bot* becomes questionable. If this computer system forms a closed circuit then any web browser does so. Just replace the eye tracker by a mouse or a keyboard and you get a regular computer setup. In a museum context *Eye-Vision-Bot* shows some structural similarities to the closed circuit concept. From the computer science perspective, however, it is just an interface.

This is true. And this allows us to reconnect the contemporary science of interaction with old-fashioned cybernetics. The feedback loop at the heart of the latter also is at the heart of any closed circuit installation, it plays a constituting role for closed circuit art, as Slavko Kacunko notes [Kacunko, 2004, p. 42]. This loop also is founding *Human Computer Interaction* (or HCI, as the academic field dealing with it is called). The very notion of the computer interface cannot be thought without the closed circuit.

When the science fiction author William Gibson invented the idea of cyberspace he did so without ever having used a computer. He just observed the circuit formed by an arcade game player, a game console and a television set (as reported by [Nagula, 1989, p. 358]). This closed circuit (formed, as Gibson says, by photons, eyes, nerves, the hand and cables and electrons) gave birth to cyberspace and still is the foundation of the I in HCI.

Two Cameras by Dieter Kiessling

Cybernetics

Cybernetics originated as a theory of systems that are coupled with an environment by observing its state and acting accordingly. Its central idea is that of the feedback loop as the base of regulatory behavior. Being formulated in a mathematical setting it soon was transferred to other realms, such as the social sciences. Contemporary constructivism and system theory are two of its most prominent offspring.

12 On the role the mirror plays for closed circuit art see [Kacunko, 2004, p. 43].

1.3 Aesthetics and Information

Being an interface for image search, *Eye-Vision-Bot* is a magic machine: It can computationally handle aesthetics. It assumes that when someone looks at images a *category* of these can be identified and other images from that category can be found. The category of an image, however, may be seen as a function of its aesthetics.

Although the question what is beautiful and why is very old, aesthetics as a philosophic field is quite new. It was introduced by Gottlieb Baumgarten in the eighteenth century [Scheer, 1997, p. 54]. He defined aesthetics as a theory of perception, placing aesthetics near what today psychology and neuroscience are dealing with. Philosophic aesthetics since then always has been in a place in between the social and natural sciences – the disciplines this thesis was meant to ignore.

As Brigitte Scheer notes, theories of aesthetics always are to be seen in context of certain epistemologies which they may complement or reject [Scheer, 1997, p. 4].

Katherine Hayles has described a central "perception" of the twentieth century: virtuality. Virtuality for her is "the cultural perception that material objects are interpenetrated by information patterns" [Hayles, 2001, p. 69]. While not being an explicitly formulated epistemology this perception dominates various discourses, be it popular or philosophic. Its central feature is a dichotomy of information and matter.

This dichotomy has not been introduced complete. It rather has been "constructed" at various "sites" [Hayles, 2001, p. 69]: Molecular biology with its idea of informational DNA defining material bodies, for instance, is one of them. The whole process of constructing it was accompanied and reinforced by technology [Hayles, 2001, p. 69]: Military technology, namely in World War II, "made the value of information real" [Hayles, 2001, p. 71]. Until today, the idea of information as distinct from and defining form and function of matter has become ubiquitous. It is accompanied and reinforced by computing machinery but also by the way various sciences think their objects.

Of course, virtuality is not totally new. It rather continues an old dichotomy of western thought: that of spirit and matter [Hayles, 2001, p. 72]. While it was constructed, however, also an aesthetics emerged that complements it: information aesthetics.

In 1933, the mathematician George David Birkhoff introduced what possibly is the first attempt towards an objective method for measuring aesthetics [Weibel and Diebner, 2001, p. 179]. This "aesthetic measure" simply defines 'beauty' B by the ratio of 'order' O and 'complexity' C .

$$B = \frac{O}{C}$$

The definition of O and C raises a number of problems that will not be discussed here. In any case, applying the statistic approach to aesthetics may be seen as a blueprint for various proximate developments such as the countless transfers of Shannon's model of technical communication to semantic contexts. During the 1960s the work of Birkhoff unfolded much influence in Stuttgart where Max Bense was teaching at that time. Here, information aesthetics was broadly elaborated.

For Bense, information aesthetics aims at developing an objective aesthetics, free of irrationalism and against speculative art babble[13]. The numerical approach of Birkhoff played a minor role in it: It was given a state-of-the-art foundation by coupling it to the notion of information science [Bense, 1969, p. 272] and it was refined by incorporating a possibly infinite number of aspects whose complexity and order may be defined individually [Bense, 1969, p. 308]. The major work, however, consisted in developing a detailed semiotic theory of artistic processes [Bense, 1969, p. 260] – hence leaving the grounds of numerical and context independent analysis.

Discussing the work of Birkhoff, Peter Weibel and Hans Diebner suggest two refinements of the aesthetic measure: First, complexity could be defined as algorithmic complexity – the only complexity measure that is defined objectively [Weibel and Diebner, 2001, p. 182]. The complexity of the least complex algorithm sufficient for producing a certain image would then be the complexity of that image. Next, the observer could be included into the aesthetic measure by not using the complexity of the image but that of the algorithm sufficient to produce the eye movements while watching the image [Weibel and Diebner, 2001, p. 184]. Both approaches fulfill what Bense demanded: an objective measure for aesthetics.

The main problem such objective measures carry with them is that they eliminate context – all forms of Hayles' virtuality, by the way, are marked by elimination of context. This simply follows from the strive for objectivity. Bense as well as Weibel and Diebner see the limits of the aesthetic measure and try to transcend them: the former by formulating a broad semiotic theory of aesthetics, the latter by recurring on the hermeneutic tradition [Weibel and Diebner, 2001, p. 187].

Herbert Franke is among those who simply ignore these limits and insist on objective beauty. He radically understands aesthetics as a theory of perception. But he equals human perception and computational data processing [Franke, 2003, p. 91]. It has been proven, he argues, that consciousness is able to take in 16 bit per second and to hold no more than 160 bit at a time [Franke, 2003, p. 91]. Beauty, he says, is the exact matching of these numbers: Everything that can be perceived at exactly 16 bit per second is beautiful [Franke, 2003, p. 92] and interesting [Franke, 2003, p. 93].

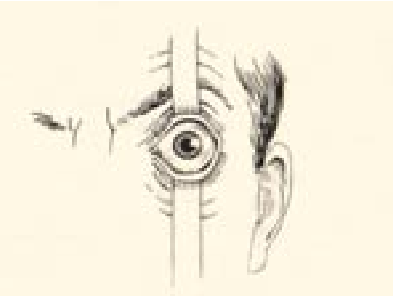
To some extend *Eye-Vision-Bot* employs an idea similar to measuring image complexity by algorithmic means: structure based search (see p. 26). It also makes use of expert knowledge – which especially in contemporary art is inherently context dependant – by using the image categorizations of *Media Art Net*. Combining both approaches definitely is not the last word in accessing the aesthetics (or semantics[14]) of an image. Integrating the user into the search process by using simple feedback loops is another technique used in *Eye-Vision-Bot* (see p. 33). To me, especially the latter seems to be of great potential.

13 "Das allgemeine spekulative Kunstgeschwätz" [Bense, 1969, p. 258]

14 The relation of aesthetics and semantics is complicated – here both words are used in a somewhat confused manner.

2 Science

2.1 Eye Tracking as Science



Eye tracking as the program of looking through the window to the mind constituted itself quite fast. Its history normally is told by scientists[15]. Instead retelling it, I will only point out some of its parts.

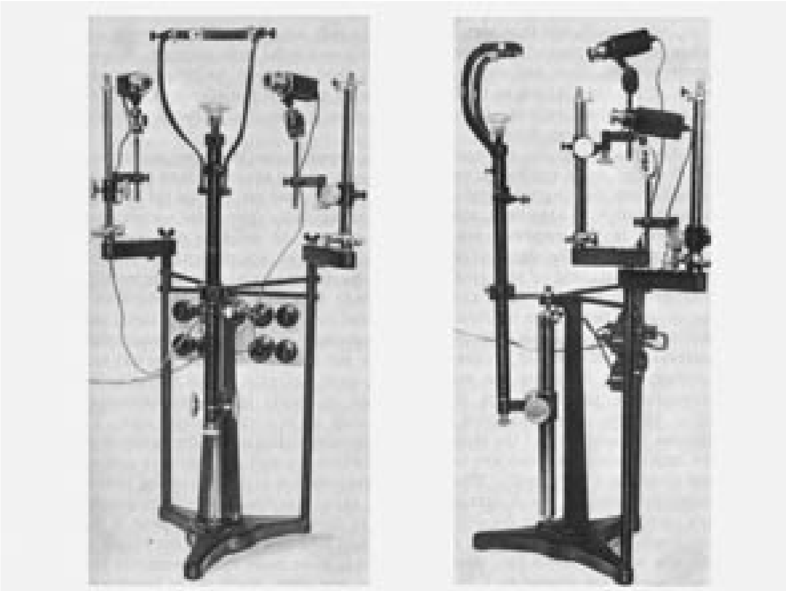
A French professor named Emile Javal realized that during reading the eye frequently stops moving. While he probably was not the first one to notice he most likely was the first one to publish that he noticed. That was in 1879. Later, he and his colleagues performed several observations of eye movements. Since their field of interest was movements of reading eyes this already defined the field in which most of eye tracking research was done later on: psycholinguistics. This also gave birth to a basic paradigm of eye tracking research: The differentiation between fixations and saccades was introduced.

The method employed by Javal was very simple: observation. Eyes observed eyes. A bit later ears observed eyes – when his colleague Lamare attached a microphone to the eye he observed in order to count the noises induced by fixations [Huey, 1908, p. 19]. Unfortunately, all these observations implied having trained observers. Untrained observation or even introspection is, as the scientist Raymond Dodge says, "utterly unreliable" [Dodge, 1900, p. 454]. Dodge, after being a trained observer for more than five years went for the technical solution. He demanded and came up with machinery: an exposure apparatus. Now the history of eye tracking became a history of technology.

With his exposure apparatus Dodge was able to show that information intake requires the eye to pause. He thus introduced one more paradigm of all further eye tracking research: Fixations are since seen as crucial for inquiries into the process of visual perception.

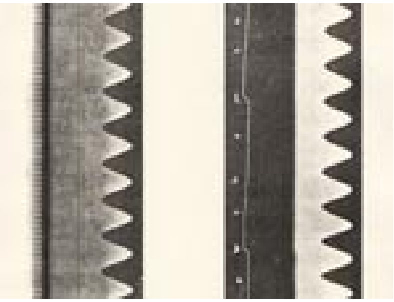
Apparently, these discoveries (fixations, saccades and the role of fixations

Early eye tracking machinery



15 For an overview of the state of the art see [Duchowski, 2003], for a summary of early work see [Yarbus, 1967] and [Huey, 1908].

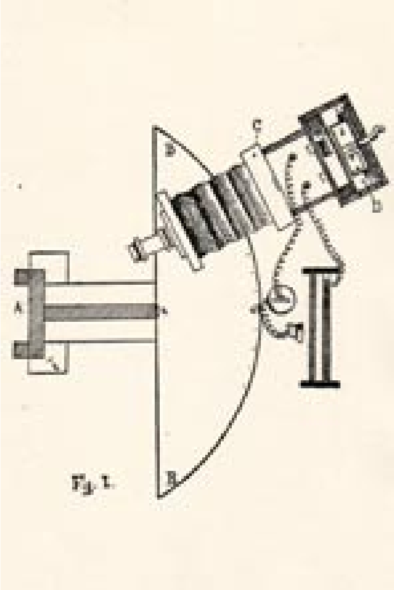
for perception) were made before the neurobiological base for them has been formulated. As already noted, the roles of fixations and saccades today are explained within a theory of the physiology of the visual system (see p. 10).



The first photographic record of an eye movement ever, made by Dodge and Cline

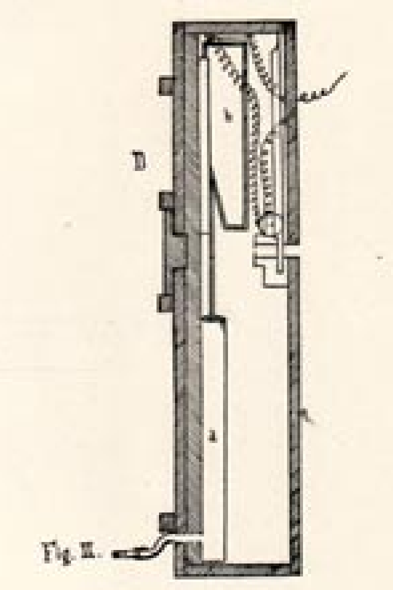
Technically the apparatus of Dodge still was simple and still involved trained observation and even introspection. As an exposure apparatus it did not observe the eyes but control what is presented during their movements. One or more colored pieces of cardboard were shown in a way that they could not be seen from each of two points subjects were asked to fixate after another. On the way from the first point to the other their gaze would cross the colored cardboard. The subjective experience of the subjects (whether or how the colors were noticed) in relation to whether their eye stopped moving between both points was measured. The latter condition was controlled by the setup of the apparatus and by observation.

Empirically his results were not as clear as it is often claimed: Vision, according to Dodge, also happens during eye movements and not only during the pauses. Only that moving vision always is “fused” and is – as he speculates – “systematically ignored” [Dodge, 1900, p. 464].



The apparatus used by Dodge and Cline

It was also Dodge to replace the trained observer by photography. In 1901, he and Thomas Cline introduced a machine that mapped the light reflected from the cornea onto a photosensitive film [Dodge and Cline, 1901]. Doing so they invented the method of measuring eye movements by using light reflected from the eye. The film they used was continuously moved behind a small slit. The reflection of a bright line from the cornea could be filmed this way. Since the film moved one dimension of the picture on it reflected the horizontal position of the eye while the other reflected its change over time. As only one spatial dimension of movement was recorded, only the speed of horizontal eye movements was measured by this apparatus. However, Dodge and Cline give no account of how they dealt with the spherical surface of the cornea. It hence remains unclear if their recordings really show what Dodge and Cline interpret – in spite the fact that corneal reflections are invariant against eye movements (see p. 10).

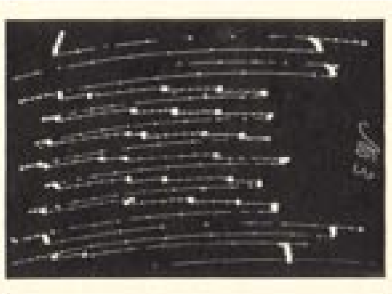


The plate holder used by Dodge and Cline

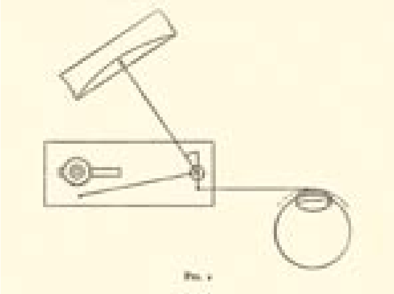
Replacing the observer by machinery was carried further when Edmund Huey constructed a mechanical recording apparatus [Huey, 1908, p. 26].

A cup on the cornea was connected to an aluminium pointer that – via a mechanical deflection – touched smoked paper on a turning cylinder. This way each movement of the eye was recorded mechanically. The setup most likely was inspired by seismographic devices of the 19th century. Huey's direct inspiration, however, were experiments of Ahrens who attached mechanical pointers to the cornea in 1891. "He was unsuccessful but he had given a valuable suggestion", as Huey states [Huey, 1908, p. 20].

Fortunately in Huey's experiments, "no inconvenience was felt, as the corneal surface was made insensitive by the use of a little holocain, or sometimes cocaine" [Huey, 1908, p. 25]. A time stamp was added to the recorded data by adding to the pointer an electrical current that was interrupted by a tuning fork. The current appeared as dots on the smoked-paper.



One of Huey's records

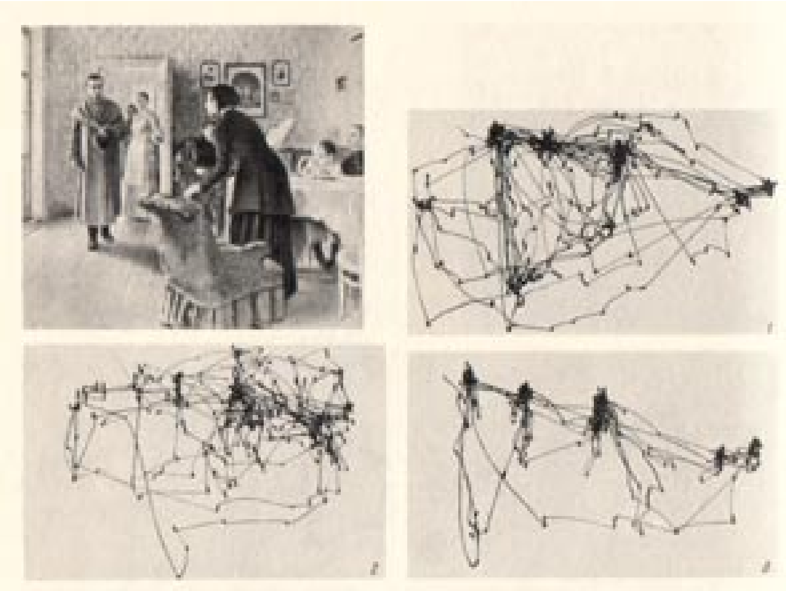


The mechanical recording apparatus of Huey

But photography was not forgotten. Moreover, soon film was ready to record eye movements. Here, too, a beam of light reflected from the cornea was filmed – as introduced by Dodge and Cline. Those working with film, however, took into account the spherical form of the cornea and attached reflecting material to it that would move when the eye moves (as, for instance, [Yarbus, 1967]).

While the psycholinguistic approach dominated for a long time in 1935 the first study on how people look at pictures was published under the title *How People Look at Pictures* (as reported by [Duchowski, 2003]). Another paradigm of eye tracking experiments stems from these studies: The trace – the spatial sequence of fixations over time – now became important. In studying reading eyes the spatial arrangement of the trace was mainly determined by the way we read. On a picture this is different: From now on the trace on a picture again and again is thought to say something about its meaning or its aesthetics.

One of Yarbus' famous figures:
One picture, one subject, different trajectories, depending on the task



Today perception of images – scene perception – is most often related to the work of Alfred Yarbus. The studies of Yarbus were very influential – his book *Eye Movements and Vision* [Yarbus, 1967] is one of the most important books on eye tracking. Large parts of it are devoted to methodology. Even larger parts deal with basic research on the nature of fixations and other forms of eye movements. But the part that probably had the deepest impact is the rather small last chapter in which Yarbus examines perception of images – complex objects, as he says [Yarbus, 1967, p. 171]. While many of his experiments are evaluated by means of statistics this part lacks scientific results in the sense of statistical significant data. However, it shows interesting images.

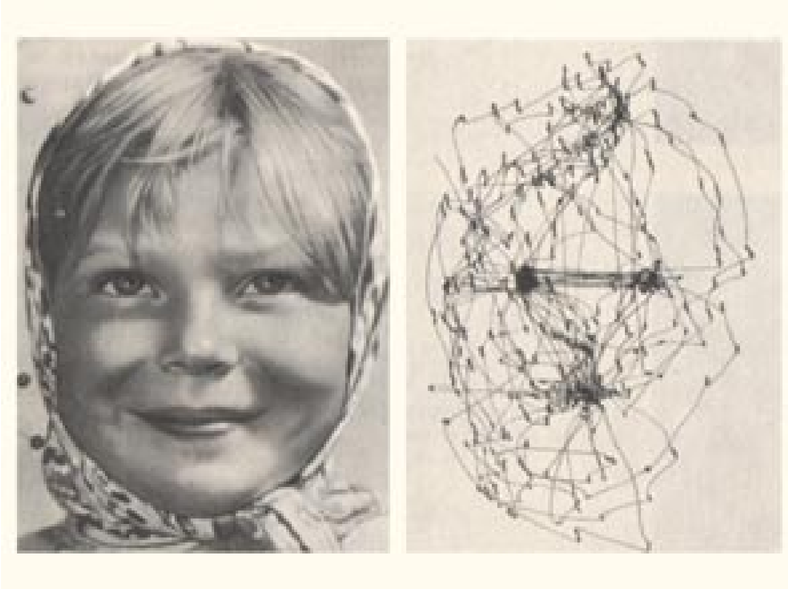
These images suggest what since then often is taken for granted: The trace on a picture depends on the person looking at it but is similar for different persons. The trace also depends on the task the person is to perform. The trace, it often is argued, not only reflects the structural characteristics of the image looked at, it reflects the attention of the observer – "the invisible is made visible by means of a trace" [MAN, Augenzeichnungen]. For Yarbus the images are enough to conclude that the eyes "fixate on those elements of an object which carry or may carry essential and useful information" [Yarbus, 1967, p. 211]. These suggestions prevail over more recent experiments that speak against that intuition: John Henderson and Andrew Hollingworth, for instance, in a review article summarize many experiments that failed to show that gaze reflects semantic informativeness. Nevertheless they state that "qualitative analysis of the figures presented by Buswell (1935) and Yarbus (1967) suggests that it [semantic informativeness] does influence overall fixation density in a scene region" [Henderson and Hollingworth, 1999, p. 250]. The problem all research in this direction suffers from, the problem of separating visual complexity and semantic informativeness – or structure and context –, hence often is subordinated under the power of these early images.

While neurobiology also claims a close relation between foveal vision and attention (see p. 11) it does not view both as being identical. In contrast, a dissociation between visual attention and gaze has been described [Duchowski, 2003, p. 9]. However, most authors believe that attention largely determines eye movements making them constantly play their role for the "great unresolved problems in perception and indeed all of neurobiology" [Kandel et al., 2000, pp. 504–505].

Besides the quest for understanding perception, since 1947 eye tracking has been used in ergonomic evaluation of interfaces [Jacob and Karn, 2003]. The idea behind such work until today is that observing the eyes of someone using an interface – as, for instance, the cockpit of an airplane – may help to improve this interface. Today the question if and how long a banner on a Nascar racing car has been fixated probably is of comparable interest [Duchowski, 2003, p. 200].

In addition to filming the eye, until the 1970s also electro-oculography – measuring the activity of the muscles responsible for eye movements – was widely used. There also were experiments with systems in which the position of a wire coil attached to the cornea was observed by measuring how the coil disturbed an electromagnetic field surrounding the head of the subject (find more in [Duchowski, 2003, p. 65]).

The era of all of these early eye tracking methods ended when computing machinery became fast enough for digital image processing in realtime. With the help of computers one can observe two independent features



Girl from the Volga and three minutes of watching, recorded by Yarbus of the eye. By comparing both, gaze direction – "the point of regard" – can be computed quite accurately without attaching anything to the eye [Duchowski, 2003, p. 58]. In the most common case, Purkinje images are measured in relation to the center of the pupil. In contrast to the former, the latter changes for different gaze directions. Computation of the point of regard is then possible if the eye tracking system has been calibrated. To do so, Purkinje images and pupil are recorded for fixed points shown on the surface the subject will look at during eye tracking. Later the point of regard on that surface can be computed in relation to the calibration data.

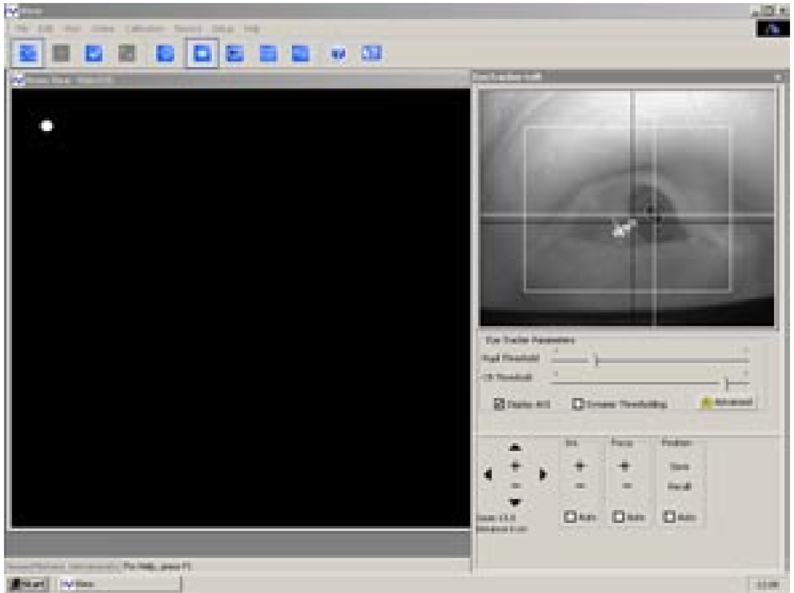
Today, recording often is done in the infrared part of the spectrum. This allows to use an invisible beam of light to cause the reflection – further reducing the invasive impact of eye tracking. The SMI eye tracker used for *Eye-Vision-Bot* employs this technique.

The idea behind all this research still is that of looking through the eye and into the mind. Be it on the quest for understanding art and aesthetics or of attention.

Eye-Vision-Bot recurs on this tradition. But it does not apply the paradigms shown here: A clear separation of fixations and saccades is not done. The trace on a single picture is not taken into consideration as well. Instead, it rather tries to directly make use of viewing times in the search process by having images compete for focus (see p. 29).

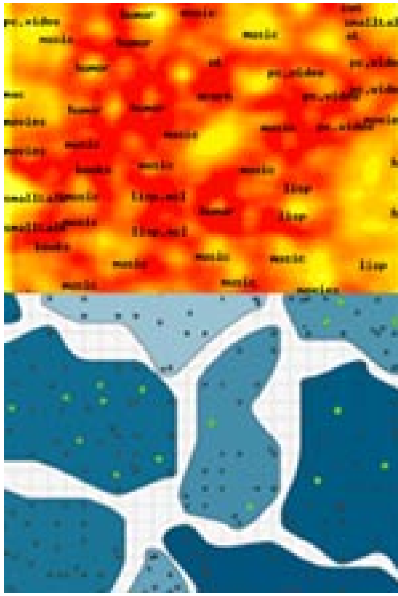
SMI

SMI is *Sensomotoric Instruments* – a company that mainly sells eye tracking devices. The device used for *Eye-Vision-Bot* is their *iViewX* system. It consists in an infrared camera and a computer connected to it. The software provided for the latter allows to control some of the eye tracking activity via UDP communication. This is how the system is integrated into *Eye-Vision-Bot* (see `http://www.smi.de`).



Inside the *iViewX* software: Someone is looking at a calibration point

2.2 Searching Images



Kohonen maps

The most obvious role *Eye-Vision-Bot* plays is that of an image retrieval system – a search engine. As such it allows interaction with a collection of images from a database. Interaction with the database mainly is based on the question which images are shown and how they are shown. Most of the work in this field of computer science – computational search – draws on the rich tradition of text search. Image search is no exception of that. Since many methods employed here are the same for any kind of data I will in the following speak of a data point denoting what for *Eye-Vision-Bot* is an image but for other search systems may be a text, a video or whatever.

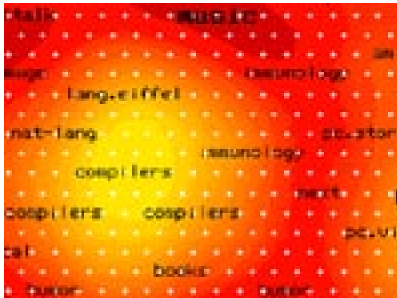
Presentation At some point of any search process the things that are searched have to be shown – they have to be presented. In presenting data points there are two extremes: Either the source the data stems from can be shown as a whole or particular points of data can be shown. Of course one also can think of any mixture of both.

Rendering a large data source as a whole in most cases means mapping it onto space. Often this is done in two, sometimes in three dimensions. During the last years these techniques became a fashion in scientific research as well as in the discourse of data mining. The best known example possibly is the work of Teuvo Kohonen, the inventor of the Self Organizing Map (SOM). However, there are other approaches such as multidimensional scaling and various forms of clustering data.

Self Organizing Maps (sometimes dubbed Kohonen maps) also may be used to cluster data. They are formed by a set of numerical units (often called neurons) that can produce numerical output for certain input vectors. In contrast to the neural networks normally used in computer science the neurons in a SOM have a distance metric that allows to ascribe positions in space to them, making it possible to draw them as maps as well as using them for drawing maps of large amounts of data.

Kohonen used Self Organizing Maps in order to create two dimensional representations of large document collections that contain millions of entries [Kohonen et al., 2000]. These then were presented as maps to browse through. His approach has been applied to the massive amount of data in news groups and also to the media art archive of *netzspannung.org*[16].

Although these methods have been quite popular other systems are mainly in use. At the moment the most successful search engine probably is famous *Google*[17] which can be used to search for images or text. This system



16 For a map of news groups see `http://websom.hut.fi/websom` (accessed 02-15-2005), for the media art archive see `http://netzspannung.org` (accessed 02-15-2005).

17 Found at `http://www.google.com`

Finding media art images with *Google*

employs the other extreme: A subset of the results of a search query is shown as ordered list. *Eye-Vision-Bot* is similar. At each point in time some particular results of the current search process are shown (see p. 43).

Interestingly the metaphor used by *Google*, *netzspannung.org* and *Eye-Vision-Bot* is the same: The spatial relation of data – their place on the map or their presence and order in the result set – is based on the metaphor of meaning. Data points appear together because they are related in sense – at least ideally.

In contrast, the contemporary metaphor behind a file system and the desktop, Stephen Johnson notes, is space [Johnson, 1997, p. 170]: Things appear together (say, in the same folder) because they have been placed at the same place. In 1997 he predicted a shift towards meaning as metaphor – towards the "semantic interface" [Johnson, 1997, p. 170]. It is important that this shift presupposes that computing machinery can handle semantics – or even aesthetics if you insist that *Eye-Vision-Bot* is working with aesthetic categories. Enabling machines to deal with semantics is one important idea behind the invention of metadata.

Comparison No matter if the search process is to happen on the basis of particular data points, on a map of all data or on anything in between, it requires one fundamental ability: The search engine has to be able to compare data points. Comparison of data can again be qualified by two broad categories: Either it is based on the structures within the data itself (in structure based search) or on metadata.

Metadata is data about data or, as Nicholas Negroponte, founder of the famous MIT *MediaLab* and preacher of the computer as friend[18] called it, "bits about bits" [Negroponte, 1996, p. 18]. It is used to add features to data that are not naturally part of that data. The simplest forms of such describe file formats or modes of compression. More confident approaches describe the content of data in terms of semantics. Metadata may be used for automatic categorization and for comparison. This, however, requires a predefined language for metadata – an ontology.

The words of this language are its categories. Sometimes even the values the categories can take are predefined and often at least the range of these is. Given that, comparison becomes relatively easy. Instead comparing data categories can be compared. Relations between categories can be incorporated into the comparison. Often metadata even include categories that already link to other data points, freeing the search system from searching for appropriate data by itself.

The artist and theoretician Lev Manovich entitled one of his essays *Metadata, Mon Amour* [Manovich, 2002]. In this essay he argues that today for the first time in history storage devices are freed from the need

18 See the book *Being Digital* [Negroponte, 1996].

to sample reality and compress culture. Instead reducing a historic tale to its key actors and events one now may just record everything. There finally is enough disk space. This freedom in turn produces the need for metadata: "Storage media", according to Manovich, need to become "active". And activity – automated processing and selection – presupposes "metadata describing the semantics" [Manovich, 2002]. He hence demands metadata to allow what Steven Johnson is predicting: meaning as metaphor for data organization.

Following this argument we soon see the central problems metadata create: They have to be created and they presuppose an ontology. As Manovich also notes, metadata allow to build "maps as big or larger than the territory" [Manovich, 2002]. The cost of building them hence may also become bigger than the cost of creating data. Automatic creation of metadata therefore is one of the hot topics for artificial intelligence. If metadata are intended to describe semantics this, however, brings back the problem that metadata were meant to solve: Automatic creation of metadata needs machines that can handle semantics.

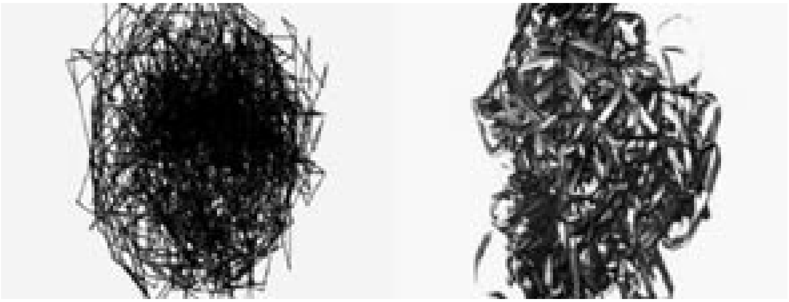
If searching is done on the basis of data itself, structures within are detected algorithmically. In text search statistics of word frequencies can be used for that, in image search computer vision algorithms are the means of choice. Images then are compared according to their visual characteristics: The distribution of color values, textures and shapes can be measured algorithmically. Searching images on that base is named content based image retrieval [McG. Squire et al., 1999].

Eye-Vision-Bot uses both metadata and searching inside data: Some of the images are requested on the basis of categorizations of the MAN database, others via content based image retrieval. The MAN database – the categories and the categorization of data according to these – were created by hand, using expert knowledge. Content based image retrieval is done using the GIFT.

The GIFT actually redirects structure based search to searching metadata. Since a realtime comparison of those features that can be compared by computer vision is computationally too expensive some features of all images are extracted once and then stored in a data structure.

The GIFT

The GIFT is the *Gnu Image Finding Tool*. It is an open source application developed at the computer vision laboratory of the University of Geneve. GIFT constitutes a server through which image collections can be accessed. Communication with a GIFT server works on the basis of the *Multimedia Retrieval Markup Language* (MRML, [Müller et al., 2000]). This language – a protocol – is designed to allow communicating with media sources that allow query by example (QBE). See `http://www.gnu.org/software/gift` for the software, `http://viper.unige.ch` for the computer vision laboratory, `http://www.mrml.net` for information on MRML and `http://openbaar.sf.net` for the MRML client that is an offspring of *Eye-Vision-Bot*.



According to the GIFT, the similarity of these images is 0.25

This structure, called inverted file, is designed to allow efficient search for similar data [McG. Squire et al., 1999]. It is one example for automatically created metadata.
This example shows that the distinction between data and metadata is problematic. As metadata technically are data the distinction is a distinction rather in usage than in nature. Looking at famous *Google* the distinction is blurred even further.
As a text search system, *Google* searches large amounts of text employing structure based search. As an image search system, the text accompanying images is used as metadata for searching images. The most important form of metadata used by *Google* is the link: The *pagerank*™ algorithm, as their algorithm responsible for computing relevance is called, essentially adds to structure based search the idea of treating "a link from page A to page B as a vote, by page A, for page B"[19].

The search process of *Eye-Vision-Bot* comprises as follows: A random set of images is selected and presented. While the images are watched, viewing times are gathered and stored. Searching new images happens on the basis of these. Both types of search based on metadata and based on structure are performed. In the former, images belonging to the same work of art and the same category as those that were most watched are searched. In the latter, the GIFT is asked to return images that are structurally similar to those. Here, two different algorithms defining the way the GIFT weights the images the query is based on are used in two queries. The results of both search methods then are mixed and presented again (as to be seen on p. 43–46). Integrating these concurrent approaches towards the similarity of images, *Eye-Vision-Bot* makes no definite claim on where to find the meaning of an image. Seen as semantic interface, semantics here are approached by combining different approaches: internal structure, metadata and, as we will see, relations and interaction.

19 As explained on `http://www.google.com/technology` (accessed 14-02-2005).

2.3 Adaptive Agents and Data Mining

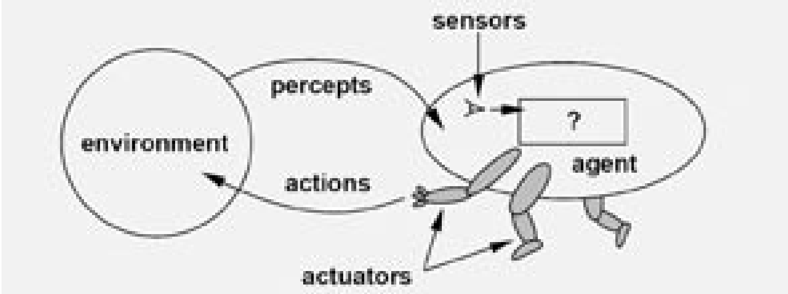
Agents The term agent advanced into the center of artificial intelligence during the nineties of the last century [Russel and Norvig, 2003, p. 28]. An agent is an input output system. Its input is defined by sensors, its output by actuators. The work of an agent can be rated by a performance measure. The world it acts in is named its environment. These terms – Performance, Environment, Actuators, Sensors or PEAS – form a complete description of an agent. That, at least, is the way a standard textbook on artificial intelligence puts it [Russel and Norvig, 2003, p. 40].
Though normally not mentioned in the artificial intelligence discourse, the ancestor of the agent is the cybernetic feedback loop. It is, in fact, identical to the agent. Interestingly, it seems that artificial intelligence forgot that when it took its detour over the expert system and neural networks[20].
Being the central concept of contemporary artificial intelligence, agents do not necessarily work with real world data stemming from analog sensors. Data miners also are agents. Only that their sensors process and their actuators produce digital information. Their environment is, one could say, software. Such agents are named software robots, softbots or

20 For a history of artificial intelligence as seen from its own perspective see [Russel and Norvig, 2003, pp. 17–28].

– even more briefly – bots [Russel and Norvig, 2003, p. 42]. That is where *Eye-Vision-Bot* got his last name from.

If an agent is intended to work intelligent or rational, Stuart Russel and Peter Norvig say, its goal is maximizing performance. For them this “principle of maximum expected utility [...] could be seen as defining all of AI” [Russel and Norvig, 2003, p. 590]. In turn this means that the performance measure serves as a definition or description of an agent’s task. Here, too, descriptions are productive.

In order to achieve this goal an agent will, given a certain input, decide to create an optimal output. In a simple reflex machine this mapping has to be build in. Other agents will have an internal performance measure that triggers the decision.



Agent and environment: a closed circuit

Adaptivity An adaptive agent is an agent whose behavior changes over time. It learns. In addition to the decision function it will have a learning function. These two are named “performance element” and “learning element” [Russel and Norvig, 2003, p. 54]. The role of the learning element is approximating an optimal performance element while modifying the performance element currently used to the better.

All agents that use the performance element before it is optimal work with estimations. They decide on the basis of their current knowledge while constantly updating it. They, as the computer scientist says, bootstrap.

Designing an adaptive agent the crucial question becomes what to adapt on. Machine learning generally is automatic approximation of a target function [Mitchell, 1997, p. 8]. Often the domain of this target function is semantics. That is the case when the agent is doing content based image retrieval which is meant to retrieve images whose content is to be similar semantically. That also is the case when an agent is to create metadata that describe semantics.

Taking up the question of what is determining aesthetics we can ask what is determining such functions: Can semantic content be accessed via context free analysis of structure, as propagated by information aesthetics, or do we need context?

Vilém Flusser pointed out that there is a common confusion of the words fact and data [Flusser, 1991, p. 203]. The Latin root of the former is *facere* the root of the latter is *dare*. Facere is to make, dare is to give. But what we nowadays call data is not given. It is no *a priori* but it is made. Data, therefore, are facts. Successful data mining may remind that and mine from what people do. And one way of accessing what people do is looking at the relations they create.

Actually this is already done. Some of the successful search systems of today are going into this direction: They pay little or no attention to content as structure. Instead, they focus on relations – on context. As already mentioned, *Google*’s *pagerank*™ algorithm uses links as relational metadata. Relational content based retrieval makes use of the interlinking between data points instead of their content [Jeh and Widom, 2002].

Amazon – the famous web shop – also does not compare the content of books or music. But *Amazon* comments ones buying behavior with hints that sometimes really are useful. The rule behind is simple: “People who liked this book also liked...”[21].

Of course, the little (or big?) brother of military machinery has understood that, too. As machinery was introduced to eye tracking in order to replace

21 Just visit `http://www.amazon.com` and look for any book they offer to create such a suggestion.

the trained observer the same reasons are pushing computing machinery into surveillance. There just are not enough trained observers to watch all the video data created in closed circuit television surveillance.

Hence surveillance is being automated on a large scale. Algorithmic CCTV – closed circuit television surveillance by algorithmic means – has been used in London Underground stations and elsewhere [Graham and Wood, 2003, p. 244]. And an important feature of algorithmic surveillance is the possibility to work with relations of data by linking the knowledge of distinct databases [Graham and Wood, 2003, p. 239]. Adopting a term by Gilles Deleuze, Stephen Graham and David Wood argue that database based surveillance of individuals creates 'dividuals' – consisting in relational data – on which surveillance systems act "without the knowledge of the original" [Graham and Wood, 2003, pp. 238–239]. Despite the problematic idea of the original in-dividual, the idea of relation over content may be found here.

Video stills from algorithmic CCTV

The search queries *Eye-Vision-Bot* creates are based on viewing times. We can claim that the viewing times may carry information about a user's interest because images are put in relational context: Several images are shown together. The paradigm behind searching images with *Eye-Vision-Bot* hence is not understanding what people search for by analyzing where they look. It rather is analyzing how images perform when they compete for attention (see also p. 33).

Besides, *Eye-Vision-Bot* actually is less adaptive than agent. But it already contains much of what is needed to become adaptive: It is, as we know, a surveillance system. Everything it does also is logged. One reason for that is that from the very beginning of the project there was the idea that the data sources for *Eye-Vision-Bot* could become "active". Why not adapting the categorization of the *Media Art Net* based on how people look at it? This, actually, might turn out as an interesting answer to the question how media art could be categorized: simply by mining from what people do with it.

3 Interface

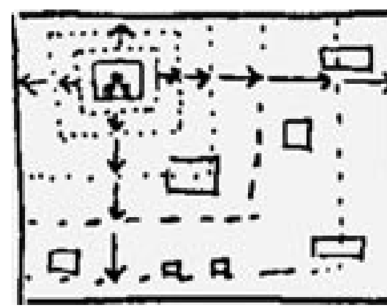
3.1 Eye Tracking as Interface

WoW

As we saw, eye tracking for a long time has been a (more or less scientific) program for looking through the window to the mind. It then was discovered by those searching for devices for disabled people. When, however, computing machinery became a means for analyzing eye tracking data and realtime analysis of that data became possible eye tracking became a topic for computer science. This was not because of the role the computer played for diagnostic applications but because of the emergent possibility of controlling computing machinery with the eye. It was because of the possibility for interactive applications.

In the context of the *Association for Computing Machinery*[22] in 1982 one of the first articles setting up this field was published: In *Eyes at the Interface* Richard Bolt asks the computer science community to "think of it as 'eyes as output'" [Bolt, 1982, p. 360]. Until 1995 this term has even made its way into Negroponte's book *Being Digital* [Negroponte, 1996, p. 136]. Eyes as human output are eyes as computer input which leads us to Kittler's (naive) idea of feeding persons into the machine (see p. 4). The belief behind both is similar: The computer that is able to look through the window to the mind will finally be able to become a friend, as Negroponte thinks, or an enemy eating us, as Kittler believes.

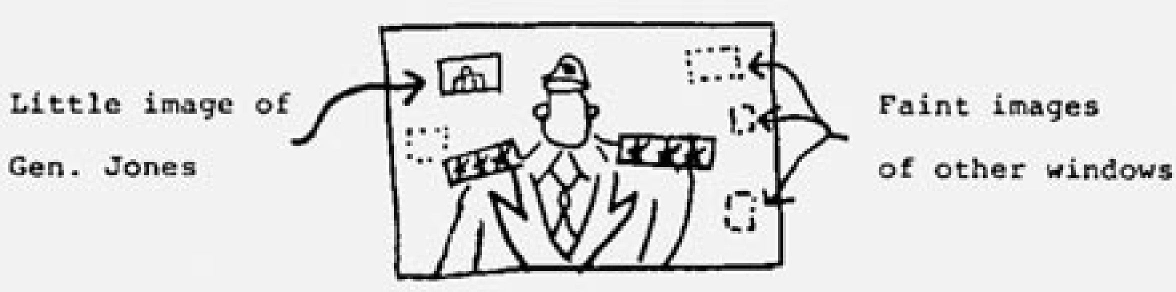
Bolt's article was written on the background of an experimental computer system he called *World of Windows* or WoW [Bolt, 1981, p. 111]. On a closer look, WoW turns out to be a direct precursor of *Eye-Vision-Bot*. Designed as a prototype it was meant to demonstrate how gaze could be used to control a number of visual information sources. Images, moving images in this case, were shown on a projection. The image in focus was halted and then zoomed to full quality. Sound accompanying the image currently focused was played. In 1981 all this was made possible by a remarkable bricolage of computers and three laser disc players. The content of the windows were short video loops of up to five minutes length. Zooming (or rather switching to full size) and halting clips was done by switching between the three players. Strikingly Bolt proposes *Eye-Vision-Bot* as an application of WoW: For him "there is implicit in the fact of such a display some kind of information-gathering network 'behind the scenes', deciding what is to be gathered, gathering it, placing it on view in certain spots of the large display [...]" [Bolt, 1981, p. 112]. What technically was impossible in 1981 is added by *Eye-Vision-Bot*: "a greater supportive database with which the observer can interact" [Bolt, 1981, p. 112].



WoW will zoom General Jones

While this early research aimed at using gaze as additional input channel – Bolt's inspiration was human communication with its mixture of gesture, gaze and speech – later the aim was changed. The point and click paradigm of mouse based interaction had become too strong to permit

22 ACM: see `http://www.acm.org`



WoW has zoomed General Jones

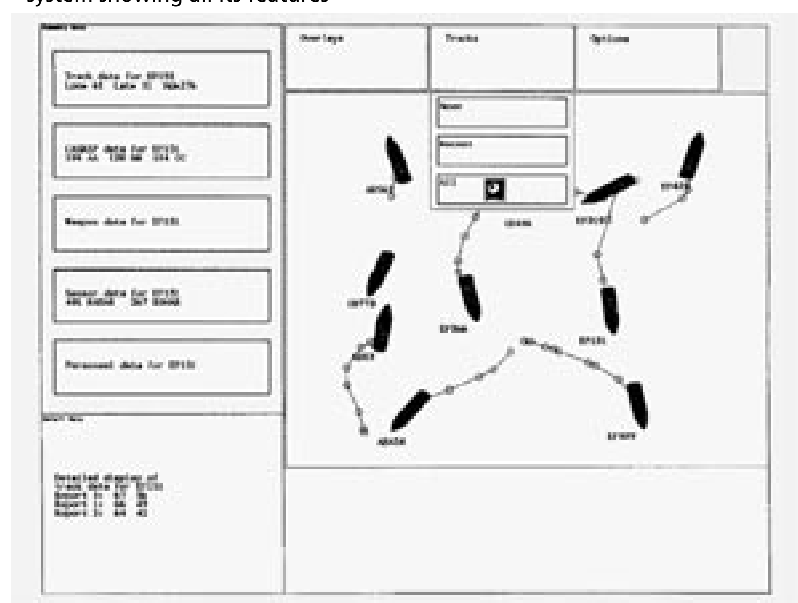
other paths. In 1987 the eye tracker was intended to become a better – that is faster – mouse.

For disabled people this idea seems obvious – still “the archetypical gaze-based pointing application is eye typing” as Duchowski writes [Duchowski, 2003, p. 207]. It is less natural to replace the mouse by the eye when moving the mouse poses no problem to the user. Nevertheless, the eye tracker soon is seen as “natural ‘pick’ device” [Ware and Mikaelian, 1986, p. 183]. Therefore, in the 1980s eye tracking was tested in detail in comparison to mouse and joystick. These early tests suggested that eye tracking *may* be faster – if there is not too much demand for accuracy [Ware and Mikaelian, 1986, p. 187]. In 1986 Colin Ware and Harutune Mikaelian conclude: “To summarize all this in a sentence; where speed is of essence, cost is no object, sizes are moderate and it is important that the hands are reserved for other activities, the eye tracker may be the input device of choice” [Ware and Mikaelian, 1986, p. 188].

Something to learn from the history of computer science is that computing machinery from its very beginning was military machinery[23]. This is where the evaluation of eye tracking as input device directly leads: In times of war, speed is of essence and cost is no object. After Bolt, whose work was funded by the DARPA[24], a naval research laboratory comes with the next milestone study about eye movements as interaction technique: The article *The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get* [Jacob, 1991], until now is *the* reference for research on eye based interaction.

The project behind this article centered around a naval command and control system. Selecting objects (ships on a map in this case)

Jacob’s naval command and control system showing all its features



23 Friedich Kittler has given several nice examples. See, for instance, *Krieg im Schaltkreis* [Kittler, 2000].

24 The *Defense Advanced Research Projects Agency*, see `http://www.darpa.mil`

A gaze contingent display

– a traditional mouse task – was done by gaze. Other tasks, such as scrolling text, were tested, too. The results were not too promising but Robert Jacob, the author, states that “it is amazing that it can be done at all” [Jacob, 1991, p. 168]. Besides for the ingenious title the paper is well known for giving a name to a central problem of gaze based pointing: It introduces the *Midas Touch Problem* [Jacob, 1991, p. 156].

In Greek mythology King Midas turned everything he touched into gold. Unfortunately, he thus became unable to touch anything at all. When the eye substitutes the mouse it touches everything it sees – just because perceiving a scene presupposes scanning it with the eye (see p. 11). Eyes as mouse – “to overload a perceptual organ [...] with a motor task” [Duchowski, 2002] – hence makes the mouse inusable. After identifying this problem much research dealt with working around it – Jacob did so by using a button or dwell time as indicator to discriminate visual scanning from pointing. Instead, the *Zerseher* of *ART+COM* makes use of this problem. So does *Eye-Vision-Bot* and so do gaze contingent displays.

The gaze contingent display possibly is the most recent gaze based interaction technique. Generally, any display that changes according to gaze position is gaze contingent. *Eye-Vision-Bot* hence also could be called one. The term, however, has been used to describe displays that all follow the same idea: Information in the part of the screen (or projection) that is currently perceived in detail is rendered in more detail than peripheral information. This can be used, for example, to save computational effort when the level of detail of 3D animations is triggered that way [Duchowski, 2003, p. 211]. This has the important effect that the gaze contingency is not perceived by the observer: The area that is in focus is rendered in high quality, the area out of focus is not perceived in high quality anyway (see p. 11).

While making use of realtime information *Eye-Vision-Bot* steps away from trying to be a better mouse. It hence in some respects is closer to a gaze contingent display than to a mouse based pointing application. Unlike both it tries to make use of the unique features of eye based interaction. Instead overloading a perceptual organ with an explicit motor task it allows the eyes to do what they always do: to look around. With *Eye-Vision-Bot* we thus are closer to the ideas of Bolt than to those of most contemporaries.

The terminology of eye tracking applications in this case is not definite. After dividing all eye tracking applications into being diagnostic or interactive Duchowski divides the latter into selective and gaze contingent applications [Duchowski, 2003, p. 132]. *Eye-Vision-Bot* does not fit into this scheme. This is because pure gaze contingent applications would not permit feedback during interaction and the idea behind a real selective application is the mouse – and hence conscious interaction.

Instead, reviewing the current state of interactive eye tracking applications, Jacob and Karn divide those according to the nature of user’s eye movements and the nature of the machine’s response [Jacob and Karn, 2003]. Both, they argue, can be natural (whatever that might mean) and unnatural yielding a taxonomy of four categories.

Eye-Vision-Bot, in this terminology could be rated as a device using natural eye movements – movements that do not have to be controlled consciously – and an unnatural response. A natural response in Jacob’s terms would be the way “people respond to other people’s gaze” [Jacob and Karn, 2003]. Of course, zooming an image on the short and searching images on the long run is not a natural response in that way. Jacob, however, also places his own work – the naval command and control system – in the category of natural movements and unnatural response. One might argue about that.

In spite of the problems scientific research on the relation of attention, semantic informativeness and gaze has raised, the idea that *what you look at is what you want* prevails until today. Attention and gaze often are just set equal. If this is justified then almost always the famous last chapter of Yarbus' book is consulted (see p. 22).

Eye-Vision-Bot basically ignores the question. It is usable in any case – be the eye tracker a mind reader or an input device controlled by intentions – because it fully integrates the user into the interface and replaces the user model (see p. 34) by interaction. To do so, two forms of feedback are used: Direct feedback consisting in zooming the image in focus makes the process of decision between images transparent. Indirect feedback consisting in showing and replacing sets of images visualizes the search process. Both together make transparent how the system works in an – hopefully – intuitive way. Success, however, still owes much to how user and interface adapt on each other.

3.2 Interface?

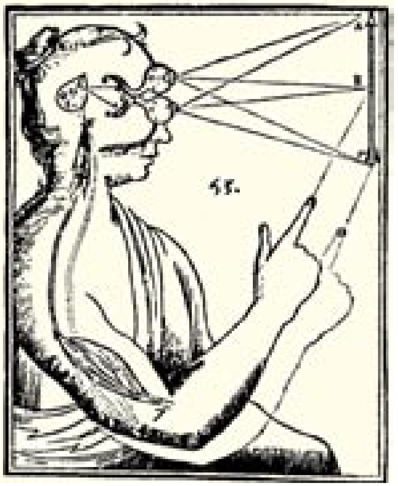
For the programmer an interface is a set of more or less related functions whose arguments and return values are clearly defined. Such interfaces are essential for the modularity of object oriented programming languages. They form the logical connection between parts of a program that – to some extend – can be thought and programmed independently. They define the functionality of objects. Someone using an object that implements – as the computer scientist says – the definition of the corresponding interface may assume that it will work the way it was defined without bothering with how it does so. Objects meet at the interface.

The interface we talk about starts from there – rhetorically not historically. In its most classical view human and computer meet at the interface. The computer in this view is more or less the Turing machine. The human in this view is the Cartesian subject. Between both there is a gap [Norman, 1986, p. 31].

On one side there is mind on the other there is the physical world. In between there is a translator. For Descartes, mind body interaction took place at the pineal gland [Beckermann, 1999, p. 50]. Today human computer interaction takes place at the interface. And as the base of interaction between man and machine is the cybernetic feedback loop (see p. 16) interface is the point where we can cut[25] the loop into its two constituents. Still the academic discipline concerned with this kind of psychophysical interaction is reflecting this thought in its name: *Human Computer Interaction* or HCI.

When design hit the screens the interface gained dimensionality. For the designer the interface often is a surface. It is "a contact surface", Brenda Laurel said (as quoted in [Fuller, 2003, p. 99]) before she dropped this notion of interface somewhat later. As we saw, the system or agent on

25 The German term for interface is 'Schnittstelle', which literally translates to 'location of the cut'.



Mind body interaction at the pineal gland – input from the eyes is transmitted to the mind and output for the hand is received there

one side of the feedback loop can be seen as an input-output system (see p. 17). The interface then is the surface that holds the two dimensional representations of input and output. Often this surface coincides with the shiny surface of the product. Of course, this surface view is induced by the screen and ignores that it always comprised three dimensional tangible elements – as mouse and keyboard.
Nowadays the field of HCI is transforming into something called *Interaction Design*. And for some *Interaction Design* is "designing interactive products" [Preece et al., 2002, p. 6]. But with this transformation the interface has not solely become product, it also was opened for a larger perspective that includes classical ideas from architecture and design. This way the interface gained another dimension: *Interaction Design* for Terry Winograd is "designing a space for people" (as quoted in [Preece et al., 2002, p. 70]). It hence is more than a surface – it is space.

Finally we may follow Brenda Laurel who states that the interface is even more than space. Interface, for her, is all there is.
This movement from the point towards everything, though being mostly motivated rhetorically, may be elaborated at some points.

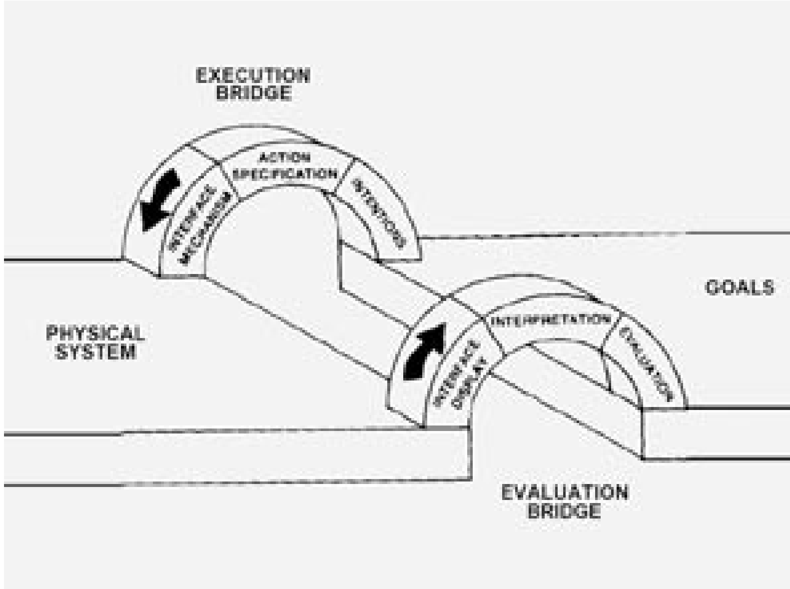
Point Stepping back a few dimensions, interface is seen as the point of translation between man and machine. This view is not only shared from the perspective of technology: While trying to talk about technology as culture [Johnson, 1997, p. 1] Steven Johnson says the same. "The interface serves as a kind of translator", he says [Johnson, 1997, p. 14], it renders the digital world for the analog observer.
Seen as such a translator between mind and the physical world the interface has to "bridge the gap" [Norman, 1986, p. 43][26]. To do so Donald Norman sees two possibilities: We can "move the system closer to the user" or "move the user closer to the system" [Norman, 1986, p. 43]. Of course, transforming people into machines does not feel right, if you are not into transhumanism. Hence for Norman follows the need for the other: moving the system to the user. As the base of that he demands a cognitive theory of the user's actions [Norman, 1986, p. 37]. He therewith arrives at the user model.
The adaptive agent is closely connected to the user model simply because what an adaptive user interface does often is nothing but user modeling [Müller, 2002, p. 13]. A model of the user is learned and used to estimate his wishes. This implies a fundamental problem: Computational learning (as well as explicit programming) of such a user model presupposes that it is possible to formulate what is important about the user in terms of computation. Interestingly, this presupposition directly contradicts the problem that created the demand for the user model: the gap between the Cartesian subject and the physical machine[27].

26 It is interesting that the whole argumentation of Norman reflects the debate around the mind body problem. Or, to be more precise, it reflects the anglo-saxon (that again is Cartesian) tradition of that debate.

27 Remarkably, metadata also arrive at demanding what they have been designed to replace (see p. 26).

User modeling, thought to its end, implies transforming people into machines.
In this light the words of Kittler may be reconsidered: A user modeled by a computer may be read as being "fed into a logic of feedback which is not ours but that of the machine" (see p. 4). This is not about the computer becoming and deleting us, but about human acting formulated on computational grounds.
As mentioned, *Eye-Vision-Bot* does not explicitly model users but tries to incorporate user action into the search process (see p. 33). However, the possible inter-actions still are limited by the way it was modeled.

Space While the academic field of *Interaction Design* is quite new the idea of "designing a space for people" is not. Terry Winograd and Fernando Flores early left the traditional path of computer science by incorporating the hermeneutic tradition into their work with and about computers.
In this course, computers are rethought as "designed in language" and "equipment for language" at the same time [Winograd and Flores, 1986, p. 79]. Computing machinery hence is placed in the space of communication – promoting "a shift from an individual to a social perspective" [Winograd and Flores, 1986, p. 78]. Design in this sense defines "the space of what can be said" [Winograd and Flores, 1986, p. 78]. These words are important for the politics and practice of designing the interface.
Insisting on the social perspective Winograd and Flores approach what for Daniels came before interactivity – namely Intermedia and Fluxus (see p. 15). This focus on the computer as being part of langauge bypasses the observer-as-proband situation, which for the computer interface could be dubbed a user-as-proband situation. In that light, it is no surprise that in *Human Computer Interaction* testing interfaces against users (or vice versa) is a major doctrine.
Politically, this doctrine may be read as typically capitalist – after all it is about justifying design decisions by mapping them onto the numerical scales of user performance. Bypassing it may open the process of designing interfaces towards the social – towards a Fluxus tradition, as one could say following Daniels.
Practically, the path Winograd and Flores take anticipates a trend in the contemporary discourse in *Human Computer Interaction*. A trend marked by the shift of the name for that field towards *Interaction Design*. People involved in promoting this shift are opening up the academic field towards those disciplines that traditionally deal with designing space and communication: design and architecture. This trend might eventually dissolve this discipline as it currently is.



Norman's Cartesian dualism of the interface: a physical system and non-physical goals

Everything But back into our movement: If interface is all there is it also is nothing. Nevertheless, in her book *Computers as Theatre* Brenda Laurel arrives at this point [Laurel, 1993]. The basic idea behind the book is applying Aristotelian poetics to *Human Computer Interaction*. Doing so Laurel tries to rethink the idea of *Human Computer Interaction* by throwing away the interface concept. That is throwing away her own idea of a "contact surface" – and starting from scratch. As the Aristotelian poetics are a theory of theatre Laurel views the interface as a representational setting – a stage – in which humans and computers perform. For her, in theatre "the stage is all there is" [Laurel, 1993, p. 16]. She therefore concludes that as soon as computers are involved "the representation is all there is" [Laurel, 1993, p. 17]. She thus frees the theory of HCI from separating between computer and interface.

In contrast, computer science and its derivatives[28] constantly favor such a separation. "Separate the design of the interface from the design of the system", Norman formulates this "principle of modularization" that is central to computer science [Norman, 1986, p. 60]. This separation is reflected in the way *Eye-Vision-Bot* was introduced here (see p. 8). This separation also is behind the *Model-View-Controller* design pattern that was used in programming it.

Laurel does not deny the separation itself. "In computerese", she says, "two kinds of representations are acknowledged: internal and external" [Laurel, 1993, p. 46]. But she goes on saying that "an internal representation has no value by itself" [Laurel, 1993, p. 46]. That is because it is not perceived – just as the script behind a play in the theatre.

Saying so she neglects the fact that the script behind a play indeed is of importance[29]. Also in dealing with computing machinery the way something is represented internally – the way it is done – affects the way we perceive it. It is true, however, that we always only perceive the interface of a program. If there is no feedback loop between human and computer, the computer does not exist. We can hence take up Laurel's idea that the representation is all there is and proceed from that: The internal representation – the script – always to some extend is part of the interface and vice versa. Their separation is an impossible separation.

Just surf the web. The look and feel of websites to a large extend is formed by the technology behind – by internal representation. Through websites we can see HTML and Macromedia Flash or databases and content management systems. It is no accident that computer scientists often publish which programming languages and tools they used in creating something. They implicitly acknowledge that the internal representation can be seen from the outside – that there only is internal and external representation at the same time.

Among the first trying to establish a critique of every-day software and

MVC

Design patterns are formalized abstract solutions for common problems. They were first used in architecture but gained much popularity in object oriented programming. One of the most famous design pattern is called *Model-View-Controller* or MVC [Cooper, 1998, p. 10]. The common problem of having a computational system (a model) that needs to be represented visually (by a view) is solved by it: In a MVC architecture, model and view are separated. They only communicate through the controller. All three are defined by interfaces stating how the controller communicates with the others. The advantage of such an architecture – besides allowing independent work on each part – is that views or models may be exchanged while not changing the other. Applied in real work, the strict separation of model and view often is not that strict. The data passed in between may to a large extend reveal the internal structure of both.

28 HCI and related fields still to a large extend are derivatives of computer science – in style and methods.

29 Her whole book seems to ignore recent forms of theatre. Forms that are not equaling theatre and representation anymore and that are at least as recent as the 20th century.

its interfaces that goes beyond usability engineering – a "software criticism" [Fuller, 2003, p. 11] – is Matthew Fuller. In his essay *It looks like you're writing a letter: Microsoft Word* [Fuller, 2003, p. 137] he pays great attention to how the object oriented architecture of *Microsoft Word* – the internal representation – influences the way the user is conceptualized – or modeled. "The user", Fuller writes, "becomes an object" [Fuller, 2003, p. 142]. Critiques aiming at how structure and interface of *Microsoft PowerPoint* model the way knowledge is conveyed even have reached the popular discourse [Tufte, 2003].

In *The Impossibility of Interface* [Fuller, 2003, p. 99], Fuller works with three definitions of interface. The architectural setting of an American prison together with its social structure, for him, represents a kind of interface that is "distributed throughout and indivisible from the system it is part of" [Fuller, 2003, p. 99]. A computer interface used in a bakery he sees as a device for "monitoring and control of a [...] map of separate elements that can be changed but not altered" [Fuller, 2003, p. 99] – an interface that divides the user from the processes she controls. That is the second type. The third type of interface he distinguishes is an "independent associational structure" that allows to control processes that are separated from it [Fuller, 2003, p. 99]. This type shares the division of interface and process with the second type but features an involvement of the user in the processes she controls that is lacking in the second type.

While the first type is close to some thoughts of this and the upcoming section and the third is close to the way computer scientists want to see the interface, it is important how Fuller works with the three types: They all, as he says, "operate one inside the other at different moments" [Fuller, 2003, p. 112]. The movement from the single point towards space and finally everything has to be seen in that light. At no point there is a complete description (that is definition) of interface.

This is why it makes sense to apply the MVC design pattern to *Eye-Vision-Bot* – thinking it as a translator rendering the zeros and ones of an image search system. This is why it is also necessary to think it as closed circuit, as surveillance system and as whatever else there is.

The medium The interface has an interesting correspondent in cultural theory: the medium. The computer itself has been described as a medium [Laurel, 1993, p. 126] and first of all both concepts share that they are hard to grasp and somehow in between. Of course, rendering the world of zeros and ones also may be read as mediating. After all, interfaces "mediate between the psychological and physical representations" as Norman writes [Norman, 1986, p. 45]. And was not the Pineal gland a mediator, too?

A certain contemporary discourse about media shares a view that already appeared here – when closed circuit art was discussed. The shift towards media art for which technology is an *a priori* (see p. 15) resembles a popular line of thought in media theory. A line of thought "that grasps social and cultural transformation as an effect of the media" [Daniels, 2000, p. 174] – that understands media technology as preceding culture. Dieter Daniels associates Friedrich Kittler but also Marshall McLuhan and Italian Futurism with this kind of thought [Daniels, 2000, 174]. And for Kittler and others these media nowadays are computational media or computers. Interface as medium in this light is what computing machinery does to people.

What this discourse tends to ignore is that media always mediate between people. They imprint their characteristics on what is mediated and on how this is done. But they still have to be written and read. If the similarities of media and interface are of importance here then because interface as medium reminds us that interface is what people do to people.

3.3 Interface!

When Jacob identified the *Midas Touch Problem* in 1991 he was able to state its cause. It exists, he argued, "because people are not accustomed to operating devices just by moving their eyes" [Jacob, 1991, p. 156]. Several years of research were not able to change his opinion: "The problem with a simple implementation of an eye tracker interface is", he writes in 2003, "that people are not accustomed to operating devices simply by moving their eyes" [Jacob and Karn, 2003].

Eye-Vision-Bot faces the same problem. We can claim that once calibration has successfully finished the interface in most cases is used properly: The interaction technique of just looking at images without a specific pointing task (as opposed to Jacob's command and control system) seems to be quite self explaining. However, people have to accustom to calibrate, to not move their head and to set up the device in correct height *before* pressing the button. Of course, being mainly developed for scientific research and for medical applications eye trackers are not made for a museum. The scientist as a behavioral corrective is missing here. When designing the museum setup of *Eye-Vision-Bot* we therefore joked that it would be nice to have a robotic device grabbing people's heads and not releasing them as long as they use the system. The mask is what is left of that.

The perfect interface probably has been described by Nicholas Negroponte: It is a drilled soldier. An admiral would just "bark orders" to him. Orders he then would type into a computer system. A naval command and control system, by the way. This was in the mid 1970s [Negroponte, 1996, pp. 97–98]. Negroponte uses this story in order to promote the multimodal, intelligent interface. The interface that understands. "My dream for the interface is that computers will be more like people", he writes [Negroponte, 1996, p. 101]. What he actually describes, however, can be read as what interfaces demand: drill. Besides, it is questionable that the admiral was able to use the soldier input system without adapting to it.

Also in the world of the dualist interface designer Donald Norman the "interface at the human side" can be changed by "training and experience" [Norman, 1986, p. 45]. He too, while arguing in favor of intelligent interfaces that "do as I am intending, not as I am doing" [Norman, 1986, p. 51], with these words acknowledges what is essential behind efficient interfaces: changing the user.

This user had to be created. As the artists Margarete Jahrmann and Max Moswitzer put it: "The videogame arcade is a place at which we were trained as computer users"[30]. This training to a large extend is training in *algorithmic* behavior: "The similarity between the actions expected from the player and computer algorithms is too uncanny to be dismissed", Lev Manovich in a text on the future of narrative states [Manovich, 1998].

30 See their homepage at `http://linx3d.konsum.net/content.html` (accessed 11-01-05)

UML

UML is the *Unified Modeling Language*. It is a graphical language that is aimed at modeling systems. As such it consists of different sets of diagrams for different purposes. These diagrams essentially model objects and their relations. There are diagrams for temporal courses of action, for structural relations, for interactions of users and systems and many more (see [Si Alhir, 1998]). So far, UML has become an industry standard for planning and documenting software projects.

Computer games, he continues, "demand that a player executes an algorithm in order to win".

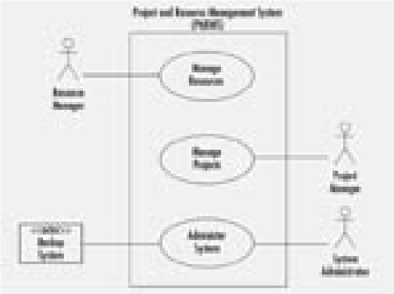
As I argued, user modeling ultimately means formulating the user in terms of computation (see p. 34). Often this process is behind both design and evaluation of interfaces. There always is "a model of the human being imposed", as Fuller states. This, for him, constitutes a "fatal flaw" [Fuller, 2003, p. 13].

In contrast, Norman is demanding this model for designing the interface. The cognitive theory of action he is seeking may have the advantage (if you think it is one) that the user's mind is taken into account. And such theories of action – "action science" [Carroll and Rosson, 1992, p. 181] – already are an industry standard in designing the interface. Scenarios and use cases are the base for these.

Scenarios are narrations. They informally describe a situation in which an artifact, as John Carroll and Mary Rosson say, is used. In designing a system a number of such scenarios may be written. All of these together "constitute a narrative theory of the artifact in use" [Carroll and Rosson, 1992, p. 190]. With the help of UML one can transform such scenarios into more formal representations[31]. The UML *Use Case Diagram* is intended to write these formal representations down [Si Alhir, 1998, p. 71].

In her poetics of the interface, Brenda Laurel uses the notion of a script as analogous to a program [Laurel, 1993, p. 44]. These scripts are not only, as the normal notion of program would suggest, telling the computer what to do. They also are "orchestrating human response" [Laurel, 1993, p. 93].

Scenario, use case and Laurel's script make evident that designing an interface in any case means defining the way people behave. Designing interfaces from its very beginning comprises modeling both: interface and user. It would be naive to assume that the feedback loop at the heart of the interface would only change the computer side. Both, in fact, cannot be thought separately. Following what Daniels wrote about interactive art, interfaces imply radical conditioning of the user through the system.

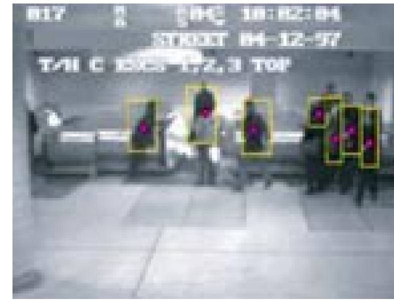


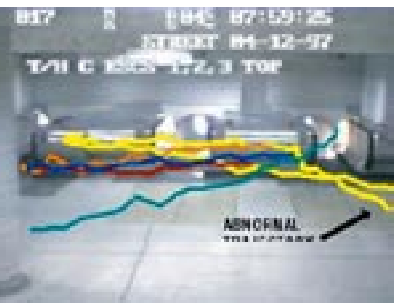
A *Use Case Diagram*: There are different symbols for human and computational actors, although both are "equivalent" [Si Alhir, 1998, p. 72]

31 "Scenarios are well-defined representational constructs of concrete or conceptual entities that are conduits for a sequence of message exchanges among other entities" [Si Alhir, 1998, p. 62].

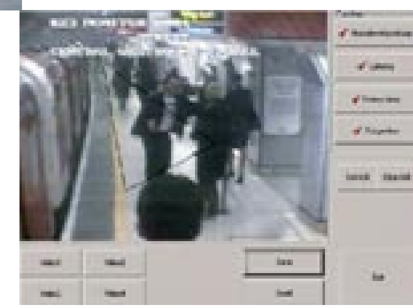
Discipline Few people would disagree that interface and structure of *Microsoft Word* influence the way people write. It is also not questionable that the state of involvement of a user – the way a user is accustomed to it – influences her work with Word. People such as Fuller also suggest that *what* is written with *Word* depends on both.

The way the term interface was used here, however, aims at more than applications of *Microsoft Word* type. We do not have to follow Fuller – reading the architectural and social structure of a prison as interface – and still arrive in broader contexts. We can stick to interfaces as those feedback loops that involve computing machinery. The London underground stations are such interfaces. As mentioned, algorithmic CCTV is employed there (see p. 29). If we now imagine these systems as interfaces between the closed circuits of surveillance and the guards reading them we would identify them as diagnostic applications. But we already saw that the dichotomy of diagnostic and interactive applications cannot be hold up easily (see p. 12). And the guards definitely will form a loop back to those that were filmed by the system. As a result of that loop, those under surveillance will modify their behavior: “In British towns, young black men have been shown to develop elaborate practices to exploit CCTV system ‘blindspots’”, Graham and Wood note. Of course if behavior is detected, exploiting blindspots means changing ones behavior in addition to just avoiding the space that is watched. Algorithmic surveillance thus is an interactive application.





Algorithmic CCTV working with trace: there is an “abnormal trajectory” to the right



Diagnostic or interactive?

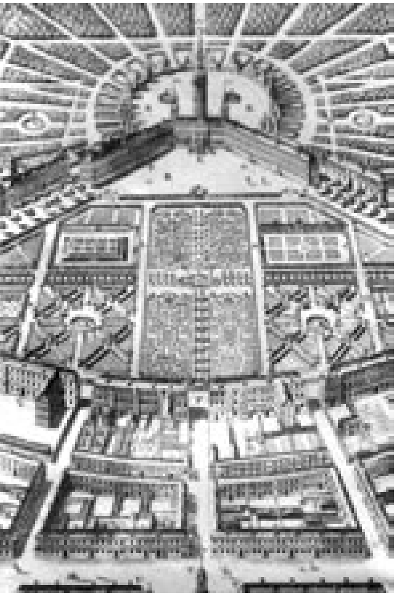
Graham and Wood identify two forms of algorithmic surveillance: systems that work with biometric features and systems that work with *trace* [Graham and Wood, 2003, p. 243]. The latter can be used – as its promoters hope – to detect behavior. Such systems are said to help preventing suicides since the suicidal supposedly show a behavior easy to detect [Graham and Wood, 2003, p. 244]. Such systems, as the company *Ipsotek* says, may also be used to detect loitering beggars [Ipsotek].

Any space under algorithmic surveillance in this light may be read as a perfect case of interface. As interfaces they change their users – they enforce a “digital rule” [Graham and Wood, 2003, p. 236]. As digital systems they were designed having a use case – a narration of correct usage of that space – in mind. There are “social and political assumptions that software producers embedded (unconsciously or consciously) into their algorithms" [Graham and Wood, 2003, p. 250]. Public spaces as interface in this sense may be understood as descriptions of what certain people think is the correct computational approach towards what those people think is the correct definition of behavior there. And descriptions are productive.

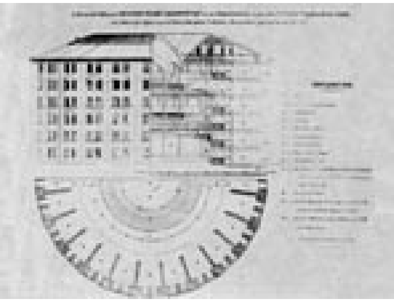
Surveiller et Punir The original title of a famous book by Michel Foucault is *Surveiller et Punir* [Foucault, 1994]. This approximately translates to the English phrase "surveillance and punishment"[32]. In terms of input and output (and in spite the fact that these terms may not seem appropriate here) surveillance is input while punishment is output. Both together make up a feedback loop. Interfaces hence can be read as systems of surveillance and punishment, they are disciplinary systems in a Foucaultian sense.

The book, whose topic is the birth of prison, traces the development from medieval to modern law enforcement. A central movement described by Foucault is the shift from medieval martyrdom towards punishment and finally to discipline. Discipline is introduced when – in the eighteenth century – the human body is discovered as being teachable [Foucault, 1994, p. 173]. In the following, drill and conditioning becomes *the* form of inscribing power into social structures.

At the base of discipline Foucault identifies a number of strategies. Among those are precise control of space and of activity. Controlling activity comprises control of scheduling, interconnection of body and gestures, interconnection of body and objects and exhaustive exploitation [Foucault, 1994, pp. 181–197]. As methods of drill he identifies hierarchical surveillance, normalizing sanctions and test [Foucault, 1994, pp. 221–238]. These finally yield panoptism [Foucault, 1994, p. 251].


Karlsruhe in 1736 – a panoptic city

Panoptism can roughly be described as a situation in which for a central observer total visibility of all subjects is given while the observer for the subjects remains invisible. The blueprint for this concept is a prison designed by Jeremy Bentham in 1789. This building placed all prisoners in a circle of single cells that could be watched from a central tower. These situations yield an automatic functioning of power since the awareness of total visibility creates subjects that observe themselves – that put themselves under self surveillance [Foucault, 1994, p. 258]. Remarkably, with this development the function of discipline becomes inversed [Foucault, 1994, p. 269]: From a sanction for those that violated the rules it becomes a *general* form of defining behavior. Discipline becomes *normal*.

Bentham's Panopticon, 1789


Such systems or disciplinary procedures, according to Foucault, tend to spread automatically [Foucault, 1994, p. 271]. What may have started at certain institutions makes its way into society. Christian school, for instance, in the eighteenth century spreads from observing pupils into observing their homes and parents. In becoming normal, the mechanisms of discipline are turning into norms. Once introduced artificially they become naturalized [Foucault, 1994, p. 391].

The references one can draw now are obvious: User modeling as defining the user, the use case as a definition of correct behavior, the feedback loop as surveillance system and means of punishment – all of these point into *Surveiller et Punir*. A detailed study of any aspect of the interface under a Foucaultian light would be possible though too extensive to be done here. That interfaces indeed tend to become norms of behavior is evident – be it in the case of the point and click paradigm, the QWERTY keyboard layout or just that people in the museum who use *Eye-Vision-Bot*

32 The English version of that book, however, is entitled *Discipline and Punish*.

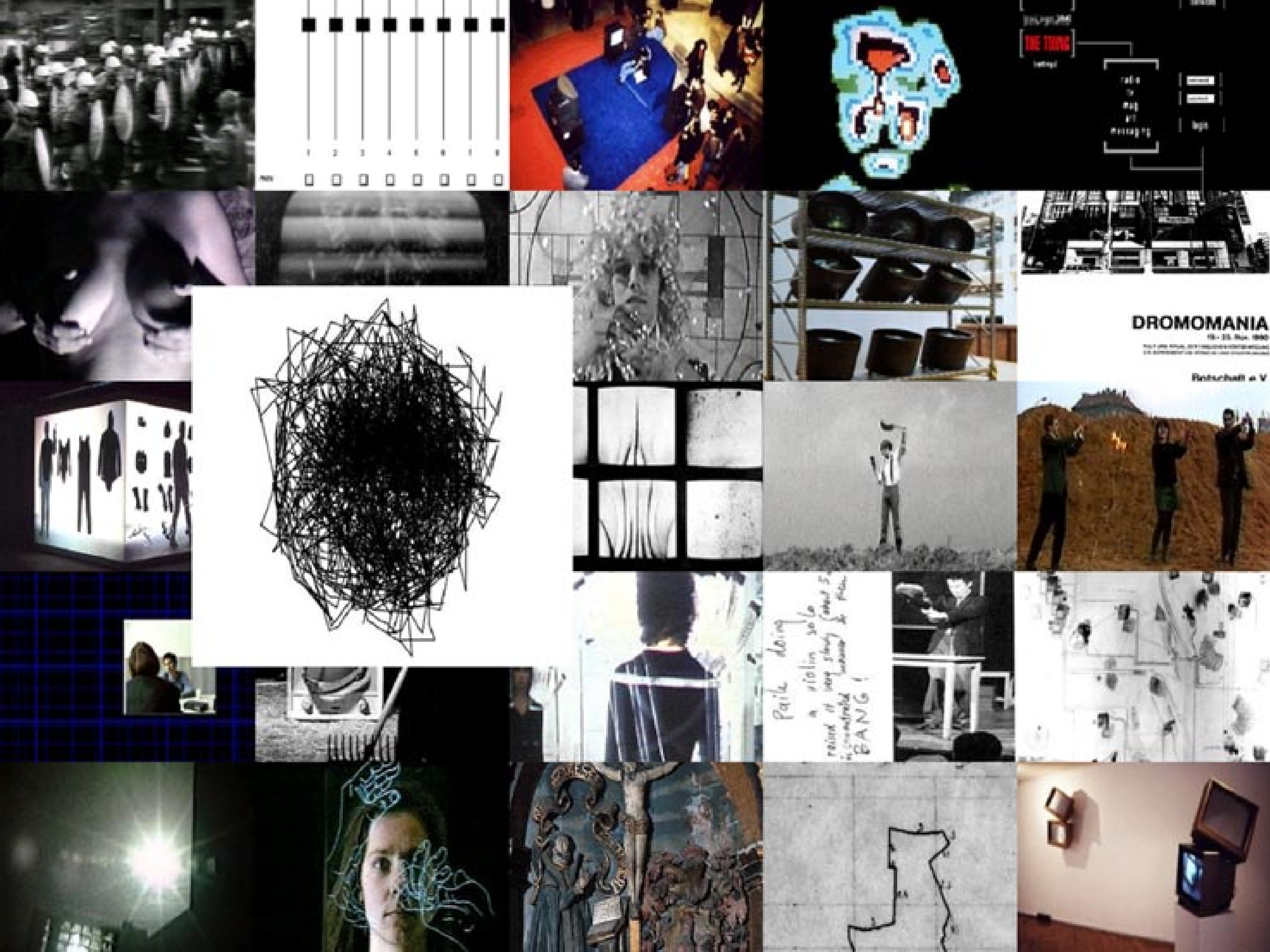
often declare their wrong behavior or even their wrong eyes as being the cause for the misbehavior of *Eye-Vision-Bot*. Maybe interaction with any device can be said to be natural merely in the context of the disciplinary system that naturalized it.
No matter if it is called the disciplinary society or, as Fuller says following Deleuze, the "society of control" [Fuller, 2003, p. 104]: Its norms and behaviors are, among other places, defined at the interface. Recurring on his example of the prison as interface Fuller concludes that there are forms of interface "which can only be met with its destruction" [Fuller, 2003, p. 116].

Of course, this is only half of the story. Fuller, too, demands the destruction of certain interfaces only while demanding the creation of others.
Eye-Vision-Bot constrains behavior, demands becoming accustomed and narrows down the number of possible paths through an image database. But reducing the number of possible decisions is not necessarily a bad thing. It rather is what we expect from a search engine. Generally this feature of interfaces is one thing we need them for: reducing complexity[33]. It is, however, evident that reducing complexity by the interface always happens in favor of certain behaviors and decisions that were inscribed by technology and design. It simply is impossible to escape the problems raised here. Designing the interface instead demands dealing with them.
Possible ways to start doing so are theorizing context – as done here – and making the interface transparent. The latter is work currently in progress: Another interface will visualize the decisions *Eye-Vision-Bot* takes as image search system. Ironically, this only is possible because *Eye-Vision-Bot* is a surveillance system and everything it and its users do can be reconstructed. Developing interfaces that to a larger extend are transparent by *themselves* is a task for the future. Unfinished business.

33 Interestingly, at this point we meet the medium again: In Niklas Luhmann's famous system theory of the social the role of media ultimately is reducing complexity in order to reduce the improbability of communication [Luhmann, 1984, p. 220].


4. Parteikonferenz der SED für den
STUBNITZ
PABLO PICASSO
Amazone
Platz des
unsichtbaren -Mahnmals
STATT
REAGAN


Structural similarity
GIFT relevance: 0.105
Structural similarity
GIFT relevance: 0.108
Structural similarity
GIFT relevance: 0.145
Structural similarity
GIFT relevance: 0.097
Structural similarity
GIFT relevance: 0.104
Structural similarity
GIFT relevance: 0.108
Structural similarity
GIFT relevance: 0.435
Structural similarity
GIFT relevance: 0.405
Structural similarity
GIFT relevance: 0.437
Same work
M+M
Call an Artist
Structural similarity
GIFT relevance: 0.114
Structural similarity
GIFT relevance: 0.442
Winning image
Same work
Structural similarity
GIFT relevance: 0.232
Structural similarity
GIFT relevance: 0.120
Structural similarity
GIFT relevance: 0.442
Same work
Same category
NEUESTER MUSIK
Structural similarity
GIFT relevance: 0.202
Structural similarity
GIFT relevance: 0.126
Structural similarity
GIFT relevance: 0.129
Structural similarity
GIFT relevance: 0.152
Structural similarity
GIFT relevance: 0.140
Structural similarity
GIFT relevance: 0.200


THE THING
DROMOMANIA


Structural similarity
GIFT relevance: 0.172
Structural similarity
GIFT relevance: 0.170
Structural similarity
GIFT relevance: 0.170
Structural similarity
GIFT relevance: 0.165
Structural similarity
GIFT relevance: 0.166
Structural similarity
GIFT relevance: 0.173
Same work
Structural similarity
GIFT relevance: 0.731
Same work
Structural similarity
GIFT relevance: 0.250
Structural similarity
GIFT relevance: 0.175
Structural similarity
GIFT relevance: 0.736
Winning image
Structural similarity
GIFT relevance: 0.746
Structural similarity
GIFT relevance: 0.189
Structural similarity
GIFT relevance: 0.177
Structural similarity
GIFT relevance: 0.736
Structural similarity
GIFT relevance: 0.739
Structural similarity
GIFT relevance: 0.740
Structural similarity
GIFT relevance: 0.204
Structural similarity
GIFT relevance: 0.178
Structural similarity
GIFT relevance: 0.180
Structural similarity
GIFT relevance: 0.182
Structural similarity
GIFT relevance: 0.179
Structural similarity
GIFT relevance: 0.187

References

[Beckermann, 1999] Ansgar Beckermann, *Analytische Einführung in die Philosophie des Geistes*, de Gruyter, 1999

[Bense, 1969] Max Bense, *Einführung in die informationstheoretische Ästhetik*, in *Ausgewählte Schriften Band 3*, pages 259–336, Verlag J. B. Metzler, 1998 (originally published in 1969)

[Bolt, 1981] Richard A. Bolt, *Gaze-Orchestrated Dynamic Windows*, Proceedings of the 8th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH, 1981

[Bolt, 1982] Richard A. Bolt, *Eyes at the Interface*, Proceedings of the 1982 Conference on Human Factors in Computing Systems, 1982

[Butler, 1990] Judith Butler, *Das Unbehagen der Geschlechter*, translated by Kathrina Menke, Suhrkamp, 2003 (originally published in 1990)

[Carroll and Rosson, 1992] John M. Carroll, Mary B. Rosson, *Getting Around the Task-Artifact Cycle: How to Make Claims and Design by Scenario*, ACM Transactions on Information Systems, Volume 10, Issue 2, pages 181–212, 1992

[Cooper, 1998] James W. Cooper, *The Design Patterns Java Companion*, `http://www.patterndepot.com/put/8/JavaPatterns.htm`, 1998 (accessed 02-21-2005)

[Daniels, 2000] Dieter Daniels, *Strategies of Interactivity*, in *Media Art Interaction – The 1980s and 1990s in Germany*, Rudolf Frieling, Dieter Daniels (editors), Springer, 2000

[Derrida, 1972] Jacques Derrida, *Die Struktur, das Zeichen und das Spiel im Diskurs der Wissenschaften vom Menschen*, translated by Rudolphe Gasché, in *Postmoderne und Dekonstruktion*, pages 114–139, Peter Engelmann (editor), Reclam, 1990 (originally published in 1972)

[Dodge and Cline, 1901] Raymond Dodge, Thomas S. Cline, *The Angle Velocity of Eye Movements*, Psychological Review, 8, pages 145–157, 1901

[Dodge, 1900] Raymond Dodge, *Visual Perceptions During Eye Movement*. Psychological Review, 7, pages 454–465, 1900

[Duchowski, 2002] Andrew T. Duchowski, *A Breadth-First Survey of Eye Tracking Applications*, Behavior Research Methods, Instruments and Computers, 34(4), 2002

[Duchowski, 2003] Andrew T. Duchowski, *Eye Tracking Methodology – Theory and Practice*, Springer, 2003

[Flusser, 1991] Vilém Flusser, *Digitaler Schein*, in *Medienkultur*, pages 202-215, Stephan Bollmann (editor), Fischer, 1997 (originally published in 1991)

[Foucault, 1994] Michel Foucault, *Überwachen und Strafen – Die Geburt des Gefägnisses*, translated by Walter Seitter, Suhrkamp, 1994

[Franke, 2003] Herbert W. Franke, *Kunst als Datenverarbeitung*, in *Algorithmik – Kunst – Semiotik*, pages 88–97, Karl-Heinz Rödiger (editor), Synchron, 2003

[Fuller, 2003] Matthew Fuller, *Behind the Blip – Essays on the Culture of Software*, Autonomedia, 2003

[Graham and Wood, 2003] Stephen Graham, David Wood, *Digitizing Surveillance: Categorization, Space, Inequality*, Critical Social Policy, Volume 23, Number 2, pages 227–248, 2003

[Hayles, 2001] N. Katherine Hayles, *The Condition of Virtuality*, in *The Digital Dialectic: New Essays on New Media*, pages 69–94, Peter Lunenfeld (editor), MIT Press, 2001

[Helfert, 2004] Heike Helfert, *Technological Constructions of Space-Time*, in *Media Art Net 1*, pages 183–199, Rudolf Frieling, Dieter Daniels (editors), Springer, 2004

[Henderson and Hollingworth, 1999] John M. Henderson and Andrew Hollingworth, *High Level Scene Perception*, Annual Review of Psychology, pages 243–271, 1999

[Hornof et al., 2004] Anthony Hornof, Anna Cavender, Rob Hoselton, *EyeDraw: A System for Drawing Pictures with Eye Movements*, Proceedings of the ACM SIGACCESS Conference on Computers and Accessibility, 2004

[Huey, 1908] Edmund Burke Huey, *The Psychology and Pedagogy of Reading (Reprint)*, MIT Press 1968, (originally published in 1908)

[Ipsotek] Ipsotek, *Intelligent CCTV IPS16 – Data Sheet*, `http://www.ipsotek.com/downloads/ipsotek/_cctv/_ips16.pdf` (accessed 03-02-2005)

[Jacob and Karn, 2003] Robert J.K. Jacob, Keith S. Karn, *Eye tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises*, in *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*, J. Hyona, R. Radach, H. Deubel (editors), Elsevier Science, 2003.

[Jacob, 1991] Robert J.K. Jacob, *The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get*, ACM Transactions on Information Systems, Volume 9, Issue 2, 1991

[Jeh and Widom, 2002] Glen Jeh, Jennifer Widom, *SimRank: A Measure of Structural-Context Similarity*, Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002

[Johnson, 1997] Steven Johnson, *Interface Culture*, Harper Collins, 1997

[Kacunko, 2004] Slavko Kacunko, *Closed Circuit Videoinstallationen*, Logos, 2004

[Kandel et al., 2000] Eric R. Kandel, James H. Schwartz, Thomas M. Jessel, *Principles of Neural Science (Fourth Edition)*, McGraw Hill, 2000

[Kittler, 1992] Friedrich Kittler, *Gespräch mit Peter Weibel*, in *Friedrich Kittler – Short Cuts*, pages 68–88, Peter Gente, Martin Weinmann (editors), Zweitausendeins, 2002 (originally published in 1992)

[Kittler, 2000] Friedrich Kittler, *Krieg im Schaltkreis*, in *Friedrich Kittler – Short Cuts*, pages 249–268, Peter Gente, Martin Weinmann (editors), Zweitausendeins, 2002 (originally published in 2000)

[Kohonen et al., 2000] Teuvo Kohonen, Samuel Kaski, Krista Lagus, Jarkko Salojärvi, Jukka Honkela, Vesa Paatero, Antti Saarela, *Self Organization of a Massive Document Collection*, IEEE Transactions on Neural Networks, Volume 11, Number 3, 2000

[Laurel, 1993] Brenda Laurel, *Computers as Theatre*, Addison Wesley, 1993

[Luhmann, 1984] Niklas Luhmann, *Soziale Systeme*, Suhrkamp, 1987 (originally published in 1984)

[Manovich, 1998] Lev Manovich, *Database as a Symbolic Form*, `http://www.manovich.net/DOCS/database.rtf`, 1998 (accessed 02-21-2005)

[Manovich, 2002] Lev Manovich, *Metadata Mon Amour*, `http://www.manovich.net/DOCS/metadata.doc`, 2002 (accessed 02-21-2005)

[McG. Squire et al., 1999] David McG. Squire, Wolfgang Müller, Henning Müller, *Content-Based Query of Image Databases – Inspirations from Text Retrieval: Inverted Files, Frequency-Based Weights and Relevance Feedback*, Proceedings of the 11th Scandinavian Conference on Image Analysis, 1999

[Miall and Tchalenko, 2001] R. Chris Miall, John Tchalenko, *A Painter's Eye Movements: A Study of Eye and Hand Movement during Portrait Drawing*, Leonardo, Volume 34, Number 1, pages 35–40, 2001

[Minogue, 2002] Kylie Minogue, *In Your Eyes*, Maxi CD, Parlophone (EMI), 2002

[Mitchell, 1997] Tom M. Mitchell, *Machine Learning*, McGraw Hill, 1997

[MAN] Dieter Daniels, Rudolf Frieling, *Media Art Net*, `http://www.mediaartnet.org`, (accessed 03-02-2005)

[MAN, Augenzeichnungen] Media Art Net, *Jochem Hendricks: Augenzeichnungen*, `http://www.mediaartnet.org/works/augenzeichnungen` (accessed 03-02-2005)

[MAN, Eye-Vision-Bot] Media Art Net, *ZKM Institut für Grundlagenforschung: Eye-Vision-Bot*, `http://www.mediaartnet.org/works/eye-vision-bot` (accessed 03-02-2005)

[MAN, Observation] Media Art Net, *Peter Weibel: Observation of the Observation: Uncertainty*, `http://www.mediaartnet.org/works/beobachtung` (accessed 03-02-2005)

[MAN, Two Cameras] Media Art Net, *Dieter Kiessling: Two Cameras*, `http://www.mediaartnet.org/works/two-cameras` (accessed 03-02-2005)

[MAN, Video Corridor] Media Art Net, *Bruce Nauman: Live-Taped Video Corridor*, `http://www.mediaartnet.org/works/live-taped-video-corridor` (accessed 03-02-2005)

[MAN, Zerseher] Media Art Net, *ART+COM: Zerseher*, `http://www.mediaartnet.org/works/der-zerseher` (accessed 03-02-2005)

[Müller et al., 2000] Wolfgang Müller, Henning Müller, Stéphane Marchand-Maillet, Thierry Pun, David McG. Squire, *MRML: A Communication Protocol for Content-Based Image Retrieval*, Proceedings of the International Conference on Visual Information Systems (Visual 2000), 2000

[Müller, 2002] Martin E. Müller, *Inducing Conceptual User Models*, Phd Thesis, Institute of Cognitive Science, University of Osnabrück, 2002

[Nagula, 1989] Michael Nagula, *Die Nullstelle der Wahrnehmung - William Gibson und der Cyberpunk*, in *Mona Lisa Overdrive*, pages 337–361, William Gibson, translated by Reinhard Heinz, Heyne, 1989

[Negroponte, 1996] Nicholas Negroponte, *Being Digital*, Vintage Books, 1996

[Norman, 1986] Donald A. Norman, *Cognitive Engineering*, in *User Centered System Design*, pages 31–61, Donald A. Norman, Stephen W. Draper (editors), Lawrence Erlbaum, 1986

[Prüfungsordnung, 2003] *Prüfungsordnung für den hochschulübergreifenden internationalen Master-Studiengang Digitale Medien (Medieninformatik/Mediengestaltung) der Universität Bremen, der Hochschule Bremen, der Hochschule für Künste Bremen und der Hochschule Bremerhaven*, 2003

[Preece et al., 2002] Jennifer Preece, Yvonne Rogers, Helen Sharp, *Interaction Design – Beyond Human-Computer Interaction*, Wiley, 2002

[Russel and Norvig, 2003] Stuart Russel, Peter Norvig, *Artificial Intelligence - A Modern Approach (Second Edition)*, Prentice Hall, 2003

[Scheer, 1997] Brigitte Scheer, *Einführung in die philosophische Ästhetik*, Primus, 1997

[Si Alhir, 1998] Sinan Si Alhir, *UML in a Nutshell*, O'Reilly, 1998

[Solso, 1996] Robert L. Solso, *Cognition and the Visual Arts*, MIT Press, 1996

[Tchalenko, 2001] John Tchalenko, *Free Eye Drawing*, Point, Issue 11, pages 36–41, 2001

[Tufte, 2003] Edward Tufte, *PowerPoint is Evil*, Wired Magazine, http://www.wired.com/wired/archive/11.09/ppt2.html (accessed 02-21-2005), 2003

[Ware and Mikaelian, 1986] Colin Ware, Harutune H. Mikaelian, *An Evaluation of an Eye Tracker as Device for Computer Input*, Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface, 1986

[Weibel and Diebner, 2001] Peter Weibel, Hans H. Diebner, *Realität, Aktualität, Ästhetik und Interpretation*, in *Studium generale zur Komplexität*, Hans H. Diebner (editor), pages 179–191, Genista, 2001

[Winograd and Flores, 1986] Terry Winograd, Fernando Flores, *Understanding Computers and Cognition – A New Foundation for Design*, Addison Wesley, 1986

[Yarbus, 1967] Alfred L. Yarbus, *Eye Movements and Vision*, translated by Basil Haigh, Plenum Press, 1967

Images

Pagewise, depth-first

Title A collage by Anneke Kleine-Brüggeney using edited screenshots of *Eye-Vision-Bot* (using images from [MAN])

5 *Eye-Vision-Bot* (using images from [MAN]), screenshot by Lasse Scherffig

6 A bricoleur, photo by Lasse Scherffig

7 A prototype, photo by Anneke Kleine-Brüggeney

8 The ZKM, photo by Anneke Kleine-Brüggeney

9 *Eye-Vision-Bot* in the museum, photo by Anneke Kleine-Brüggeney
Eye-Vision-Bot in the museum, photo by Anneke Kleine-Brüggeney
Eye-Vision-Bot in the museum, photo by Ulrike Havemann

10 The eye, detail from [Yarbus, 1967, 6]

11 *Eye-Vision-Bot* sees an eye, screenshot by Lasse Scherffig

12 One of the *Augenzeichnungen*, from [MAN, Augenzeichnungen]
One of the *Augenzeichnungen*, from [MAN, Augenzeichnungen]
One of the *Augenzeichnungen*, from [MAN, Augenzeichnungen]
One of the *Augenzeichnungen*, from [MAN, Augenzeichnungen]

13 *Augenzeichnungen*: machinery, from [MAN, Augenzeichnungen]
Augenzeichnungen: machinery, from [MAN, Augenzeichnungen]
EyeDraw, from [Hornof et al., 2004]
Free eye drawing, from [Tchalenko, 2001]

Besides those who directly were involved creating *Eye-Vision-Bot*, I want to thank Florian Grond, Rudi Hinterwaldner and Andres Burbano for help and discussions. I also want to thank Alex Gerdemann, Cordula Patzig and especially Anneke Kleine-Brüggeney.

Contact Information

Editor Hans H. Diebner, hd@zkm.de
ZKM | Institute for Basic Research
Lorenzstraße 19
76135 Karlsruhe, Germany

Author Lasse Scherffig, scherffig@scheimlack.de
Designer Anneke Kleine-Brüggeney, anneke.kb@gmx.de

ISBN 3-928201-34-4