CERTAIN ASPECTS OF MACHINE VISION IN THE ARTS – extended version

Marc Böhlen
Department of Media Study
231 Center for the Arts
Buffalo, New York

TEL: 716 645 6902
FAX: 716 645 6979

Toronto – Buffalo April 2003

CERTAIN ASPECTS OF MACHINE VISION IN THE ARTS

Abstract

This essay attempts to consider the consequences of machine vision technologies for the role of the image in the visual arts. After a short introduction to the topic, the text gives a practical overview of image processing techniques that are relevant in surveillance, installation and information art practice. Example work by practitioners in the field contextualizes these more technical descriptions and shows how computational approaches to digital imagery can radically expand the use of the image in the arts. A final note on possible future areas of investigation is included.

Keywords

Machine vision, computer vision, visual arts, machine vision art, computer vision and installation arts, new media, surveillance, surveillance art forms, information arts

Introduction

The domain of the arts is wider today than is has ever been. The last decade has seen the appearance of activities in the arts previously reserved for science and engineering practices such as robotics, machine vision, data mining and bioengineering. This essay will discuss some conceptual, philosophical and practical issues in machine vision as they apply to an extended art practice that blurs preconceived distinctions between engineering methods and artistic practice. This text is neither a treatise on machine vision[1] nor a history of automation of vision[2], nor concerned with issues of machine vision particular to human computer interaction[3]. This text is an attempt to place machine vision into a critical cultural context, and to show how machine vision allows for new forms of inquiries into imagery far beyond those of surveillance, and how machine vision used to extract information from image data changes the role of the image in the arts.

Overview

Our visual perception apparatus is so highly developed that we often equate what we see to a depiction of reality itself. Three color sensors, one each sensitive to the red, blue and green spectrum, transduce together with intensity sensitive cells in the retina a curiously short interval of the electromagnetic spectrum into a data stream that our cerebral cortex discerns as objects in space. We have not always “seen” as we see today. Our visual apparatus evolved from less complex precursor sensory systems. Evidence that the human retina carries more than one copy of the middle wavelength sensitive cone ospin,[4] suggests that this development is not over. Tomorrow we may see things differently.

Early vision research began with investigations into amphibian visual perception[5]. Later, research on the macaque monkey brain has shown that primate visual activity encompasses separate systems for visual perception and visual control[6]. In primates, vision is a dual system where one stream controls object recognition and another object-directed action[7]. While much of our understanding of primate vision comes from research on the level of neurobiology, much of what we know about human vision comes from physiological experimentation and case studies of pathologies. It is not clear how much of the low-level neurological findings in the macaque monkey and other primates actually apply to the human visual system. Nonetheless, there is evidence that this perception-action distinction is inherent in human vision as well[8].

Vision is more than sight. Much of the activity of the visual system has nothing to do with sight per se. The papillary light reflex, and the visual control of posture[9] are examples of what is called reactive vision. Experience from everyday visual perception also confirms the assumption of two vision modalities in humans. The juggler can follow the trajectory of moving objects but cannot, and need not, remember particularities of these objects' appearance, while a few moments of viewing breathtaking scenery can be material for memories. Everyday experience further confirms a complex interweaving of these vision modalities. In the context of machine vision, gazing is a particularly interesting kind of seeing. It differs from watching, that is intentional and anticipates, goal oriented, an eventful action. Gazing is looking with no particular intention, a kind of removed surveying of events. The many forms of seeing are intimately linked to thinking; vision is meshed into the cognition process.

Machine Vision

A cultural paradigm shift occurred when automation reached into the realm of perception. The automation of core involuntary perception processes we perform by biologic necessity turns the high regard attributed to sight since antiquity inside out.

Machine vision has many forms. X ray, gamma ray, radar, sonar and tomography are all methods of encoding spatial relationships by transducing atomic, electromagnetic or sound energies. Radar and sonar, for example, map distance to an object by time of flight, the time it takes for a signal to be reflected from an object back to the point from which it was emitted. X ray, gamma ray and radar are active where our sensory system is blind and dumb. They are not limited to the bandwidth in which the human eye and ear operate.

Machine vision includes the act of interfacing to the world. Sensing and transducing light energy is part of machine vision. Limitations and particularities of the sensing system modulate all following processes. Computer vision generally deals with operations that follow image creation. Here, however, the terms will be used interchangeably. In this text I focus on machine vision defined as the automation of vision processes that mimic the human visual system and operate in the same spectra as the human eye.

Parallel to the long term and possibly utopian end goal of strong artificial intelligence, the aim in machine vision is the full synthesis of human vision, the replacement of human seeing. Not surprisingly, the domains of machine vision and machine learning overlap and compliment each other. Machine vision systems today can robustly recognize faces and track automobiles. This success is a result of a marriage between findings in neurophysiology and computer science. It was David Marr's work[10] that laid the foundation for a computational approach to understanding vision. By grounding vision in the domain of mathematical information processing, Marr made available to vision research the general procedures of digital information applicable to any type of quanantized data. By abstracting the signal processing procedures from the locations in which they occur, the retina and brain, Marr laid the foundation for building artificial systems capable of synthesizing human vision-like perception. Marr's formulation takes into consideration the results from neurophysiology and maps vision primitives measured in amphibia[11] to primary operations. By Marr, such primitives combine to form the basis for higher and more evolved forms of seeing. Marr's approach begins with a 'primal sketch', the most significant intensity changes in an image, followed by a '$2_{1/2}$ dimensional sketch' that delivers depth information, and a '3D representation' that includes object properties. The grand machine vision narrative Marr conceived has not fulfilled itself to date. Many high level visual processes and the link to thought and selective memory remain inaccessible to synthesis. Despite this, incremental but substantial progress has been made in solving particular problems in the automation of vision.

Interestingly, machine vision excels as some perception tasks while it fails at others humans perform effortlessly. Color constancy is an example in case. Color constancy describes the ability of the human visual apparatus to accommodate for changes in ambient lighting conditions and color environment when seeing color. To a human, a red apple always appears red, while machine vision systems cannot readily correct for color under variable ambient light. Because machine vision is very selective, it can miss much of the richness of visual information human beings enjoy through sight. In this regard, machine vision is a poor form of seeing, one that conveniently allows for certain types of precision but negates the existence of what it can not perceive.

Categories of Machine Vision

There is no reason why artificial visual perception need be constructed along the example of human vision. However, the belief that our visual perception apparatus is a near optimal solution to the problem of light capture and processing has lead vision research, from the first camera to current computer perception, to attempt to reconstruct key elements of human vision. Cameras are built to mimic the human eye. In the digital camera, the retina is typically replaced by a charge-coupled device that transduces the light energy impinging on the sensor into electrical signals. Quantization is necessary to map the continuous signal to discrete values. The result is a translation, a discrete map of the visual stimulus. One needs a useful data container to hold this information. One such suitable descriptor is the

matrix, an ordered set of vectors. For the purpose of image manipulation, one can think of an image as a simple two-dimensional matrix, a table with rows and columns. Such an image matrix can be denoted as Im[i] [j], where Im is the image matrix, i the position of the row of a particular pixel and j, the column location of the same pixel. Each cell contains discrete information about the total image. Each of these i x j cells or pixels is encoded by n bits for color, intensity and other properties. The density of the pixel arrangement defines the resolution of the image on the screen, the ratio of pixels per unit area. A 24bit per pixel color scheme, for example, reserves eight bits for each of the three colors red, blue and green for all elements of the image matrix. By this scheme a simple color image is a multidimensional mathematical representation in which three dimensions or layers define color and two dimensions define location.

Matrices can be performed on by operations of linear algebra, such as subtraction, addition, multiplication and division. Since images can be mapped as matrices, all operations defined on matrices can be applied to images. Of the many operations that linear algebra can perform on matrices, a subset is useful for image data.

The following section divides machine vision procedures into four main categories: fundamental operations on individual images, complex operations on individual images, fundamental operations on a sequence of images and complex operations on a sequence of images. In each category some of the main ideas are discussed. The list is incomplete. It is only a subset of all the known procedures, but can form a guideline for the uninitiated and a point of departure for the interested reader to continue from.

Fundamental operations on an individual image

Fundamental operations on an individual image include geometric operations, color space operations and filtering.

Geometric operations include spatial transformations such as rotation, scaling and translation of the image matrix. In all cases the approach is to multiply the original image matrix with the appropriate transformation matrix[12]. Rotation is achieved by multiplying the image matrix by a rotation matrix. The entries of the rotation matrix depend on the axis about which one rotates the object and the dimensionality of the space in which one rotates. For example, the rotation matrix that rotates an ellipse clockwise around its origin in the plane by the angle Θ, is the 3x3 matrix

$R = [\cos(\Theta)\ \sin(\Theta)\ 0;\ -\sin(\Theta)\ \cos(\Theta)\ 0;\ 0\ 0\ 1]$.

Translation is similarly achieved by multiplying the image matrix by a translation matrix. In this example the matrix is:

$T = [1\ 0\ 0;\ 0\ 1\ 0;\ tx\ ty\ 1]$.

This results in a shift by tx in the x direction and ty in the y direction.

Color space operations alter the color components of an image; adding a constant to the R band of an RGB image results in a red shift. Some color operations require an image to be transferred into an alternate representation. An RGB image implicitly contains information about luminescence; as a property it is not directly available to linear operations. To change the luminescence of an RGB image, for example, it is necessary to map the image into the HSL (hue, saturation, luminescence) representation. This can be done through a multistep conversion algorithm[13]. Thereafter the brightness of an image can be altered directly by linear operations.

Filtering an image is achieved by convolution of the image matrix with a convolution kernel of appropriate size and weight. In the discrete case of a two dimensional convolution, the kernel itself is a matrix or mask, usually a few pixels wide. The convolution proper involves successive multiplication and summing of the convolution kernel with sections of the image, starting from the top left to the bottom right of the image. One can imagine the kernel as a small window sliding over the original image, operating on the respective overlapping area, only to be shifted in the next step. Smoothing or blurring is a typical convolution based filter. In order to smooth an image one convolves it with a particular mask that results in evening out the details. In such a mask, all entries are usually of the same value. Blurring is but a form of local averaging. The following 3x3 matrix could be used to smooth an image:

B = [1/9 1/9 1/9; 1/9 1/9 1/9; 1/9 1/9 1/9]

Convolution with this kernel can also be interpreted as a lowpass filter. It has the same effect as the removal of high frequency information and blurs the image. Different kernels have different effects when convolved with a given image. Image sharpening is the inverse of blurring. Here one chooses a mask in which the center values are much larger than those at the edge. A kernel that can be used to sharpen an image would be

S = [-1 -1 -1; -1 9 -1; -1 -1 -1]

with a central value significantly larger (here 9 times) than the others and opposite in sign.

Complex operations on an individual image

The second category of operations comprises complex operations on an individual image. It contains, amongst others, methods of segmentation and pattern matching.

Segmentation is the process of distinguishing objects from the background in which they are set. Thresholding and edge detection are examples of segmentation operations.

Thresholding is the process of binning a large set of numbers into two or more categories based on a control or threshold value. Eight bit grayscale images with values between 0 and 255 can be converted to binary black/white by thresholding all gray values into either white or black, depending on whether they are above or below a given control value.

Edge detection is a problem of fundamental importance in image analysis. The human visual system is very sensitive to abrupt intensity changes and edges. Much effort has been placed into the synthesis of similar procedures in machine vision. Often, edges characterize object boundaries. Synthetic edge detection makes use of this to delineate objects in the image matrix. This is a two-step process that comprises filtering and detection. Filtering is applied first to remove noise from the image. The detection component finds edges of objects in an image. Mathematically, this means finding local changes of intensity values in an image. In order to find such changes, one usually uses the first derivative or gradient of the image information as it assumes a local maximum at an edge. For an image Im(x, y), where x and y are the row and column coordinates respectively, one typically considers the two directional derivatives or gradients. Local maxima of the gradient magnitude identify edges in Im(x, y). These gradient matrices are formed by convolving the original image with appropriate kernels. One such kernel is the Sobel operator. The gradient in the x direction is created by convolving Im(x, y) with

Sx = [-1 0 1; -2 0 2; -1 0 1].

The gradient in the y direction is created by convolving Im(x, y) with

Sy = [-1 -2 –1; 0 0 0; 1 2 1].

When the first derivative achieves a maximum, the second derivative is zero. For this reason, an alternative edge-detection strategy locates zeros of the second derivatives of Im(x,y). One well-known differential operator used in these so-called zero-crossing edge detectors is the Laplacian operator[14]. The results of a well-designed, robust edge detection algorithm yield visually pleasing results, at times with a similar sensitive delineation as a drawing by an accomplished draftsman.

Pattern matching requires objects and properties to be found first in a single image and then matched with similar findings in other images. Matchable properties include texture, color, geometric shape and frequency components. The matching algorithm between images depends on the properties to be compared. There is no universal, unique and precise notion of similarity, particularly since similarity has both a quantitative and a qualitative aspect. Typically one contents oneself with a numeric representation of quantitative similarity. This is how one can imagine the process: Assume that representational features have been found in an image and that there are j features and i classes. The similarity S of the object with the i th class is then given by

$S\,i = \sum(w\,j\,s\,j)$

where w j is the weight for the j th feature. The weight is selected based on the relative importance of the feature. The similarity value of the j th feature is s j. The similarity S i is then the sum of all products of the weighted features times the corresponding similarity value.
The best similarity value is indicative of the feature of the image. Fitness or 'closeness' is usually translated into a distance measure. Least squares, correlation and k nearest neighbor methods are commonly used to evaluate distance measures.

Fundamental operations on multiple images

All of the above methods are described with regards to a single image. Applying such operations to a sequence of images allows one to find features that are not apparent in a single image. Some features become salient only when viewed in the context of other images.

The most important operation on sequential images is image differencing. It is a very simple procedure, computationally cheap and fast. This powerful method calculates the numeric difference between the matrix values of two images. For single band or binary images this gives a quick indication of change due to motion between two similar image frames[15]. Image differencing can be performed rapidly because subtraction as well as addition use few computation cycles, as opposed to multiplication and division based operations. Usually, sequential images are much more alike than different. In real time (30 frames/second) video streams, sequential images differ only slightly for many applications. The resultant difference information is much smaller, essential, and can be operated on much more efficiently. Because of these features, image differencing is the basis of motion segmentation and is often used as the basis for robust tracking algorithms. Tracking, in simple and complex forms, allows one to tally information over time, to log and keep track of events. The diachronic nature of tracking gives a window onto properties that single image analysis cannot reveal. With this, time becomes an image manipulation parameter.

Complex operations on multiple images

In this category machine vision unfolds its full potential. For this text, the most important applications include object tracking and navigation. They are built to a large part upon the imaging primitives described earlier in this text.

Object tracking is the repeated finding of an object's location in sequential images. In tracking, an object must first be segmented from the background. Many methods can be applied to this end, but filtering and select feature detection are often part of it. From the result set the tracking algorithm must select a subset that fit a certain goal criteria such as geometric, texture, color properties or feature factors. From this pruned result set, a tracking algorithm must then find the position of the sought objects in the image plane. This is often achieved by calculating the center of gravity of the result set. These procedures are then repeated for each image such that the goal object is continuously

located in a steam of images. Applying such a tracking algorithm to a moving object results in a changing stream of x and y coordinates that indicate the path of the object across the image plane. To track an object over a field of view that is wider than the angle of the camera lens, the camera itself can be moved in sync with the object's projected path. Pan and tilt operations adjust for the motion of the target and are reactively based on feedback from the data stream of the moving object's position.

Navigation is a special kind of tracking. Usually one selects a few key features to extract and track from a video stream and uses this information in a feedback loop to control the heading of a vehicle. Speed is of essence is such procedures. Reaction time to change course must be accounted for in any navigation algorithm. Machine vision is generally susceptible to performance degradation under variable lighting conditions. In navigation, this can have dangerous consequences. For this reason, much care is given to adaptive processes that can dynamically adjust to change in ambient conditions, for example by varying color thresholds by neural networks[16].

Many more such derived operations on sequential images could be listed here. Many, such as optical flow from moving cameras and scene reconstruction in active vision are too involved to address in such a short survey[17]. Shape by motion detection, however, is interesting enough to mention. Shape by motion detection is an involved set of operations that reclaim from sequential images information on the shape of an object tracked and the relative motion of the camera. It requires prior knowledge in the form of estimations of shape and of motion. A two-stage approach is employed when applying shape from motion to video. In the first stage, two dimensional point features are extracted and tracked though the image sequence. In the second stage, the resulting feature tracks are used to recover the camera's motion and the three-dimensional locations of the tracked points by minimizing the error between the tracked point locations and the image locations predicted by the shape and motion estimates[18].

All of these procedures operate on images that are captured from a single camera. It is also possible to apply them to image data from multiple cameras. In some cases, this delivers different information than is available from a single camera. Stereo vision, for example, employs two cameras in a fixed and calibrated configuration, just as our eyes are fixed in our eye sockets. With this, stereo can deliver distance information by measuring the disparity in simultaneous images from both cameras. Adding additional cameras adds further relevant information. Multiple view interpolation creates continuous scenes from single overlapping images.

Beyond Surveillance

Earlier in this text I singled out gazing from the many modes of human seeing. Actually, one can interpret machine vision as a kind of gazing. The machine vision camera must capture each image frame in totality before it can operate on it. There is no pre-selection of features and areas prior to image capture. The machine cannot see selectively. It is forced to see everything in its field of vision. Each entry into the image matrix must be filled

before further operations can occur. Only after this will the numeric processing of the image data begin. This is in contrast to human vision that is more distributed and selective. The retina preprocesses image information and delivers to the cortex a subset and representation of the complete visual stimulus. As humans, we can voluntarily see selectively by focus of attention. That is the act of watching. But when humans gaze, they de-emphasize the early filtering, look with no immediate intention, and act very much like a machine must. Machine gazing and human gazing are similar in this regard. A gazing machine, however, is usually understood as a threatening machine since its resultant data is often recorded. The angst of surveillance is a construction of gazing and recording. If one removes the recording event from this process, surveillance becomes again a neutral gaze.

The next section describes work from practitioners who make use of machine vision in their work in a variety of ways. The focus is set not on conformance with standard procedures, but rather on understanding how artists interpret the results of machine vision processes as an expressive medium. In order to properly situate the use of machine vision in each of these pieces, short descriptions of the work beyond the topics of this text are included. As the reader progresses through the examples it will become clear how many of the above described procedures, and some not discussed here, become new methods of creating meaning and intention far beyond those of surveillance.

David Rokeby: Shock Absorber, 2001

Shock Absorber takes live feed from broadcast television and separates it into two parts in real time. One part contains all the movements, edits and high frequency visual stimulation. Changes and movements are visible, but anything constant in the image is not seen. The other part contains everything that is left over after the movements and changes are removed. In this case, a cut becomes a slow cross-fade. Newscasters' bodies are solid, but their eyes and lips are blurred. Figure 1 shows an example of these two separated versions of a video feed.

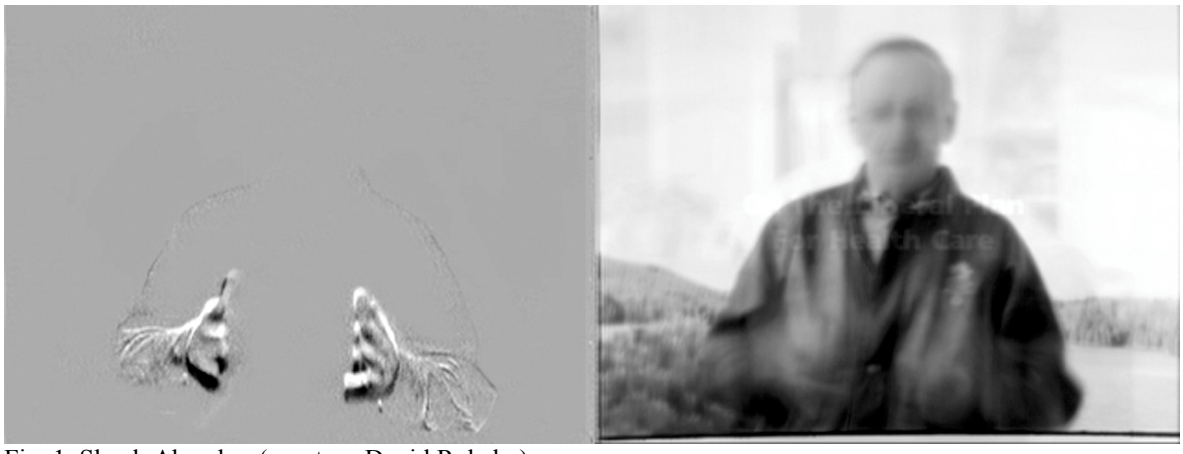

Fig. 1 Shock Absorber (courtesy David Rokeby)

This piece works from the insight that our brains build elaborate fictions as it constructs the images we see. Originally justified in the rational need for fast reaction and survival, seeing and social seeing have additional dimensions. For the artist, the television industry has developed a visual language that triggers the perceptual system in new ways. Changes and refinements are applied based on the feedback loop of the ratings mechanism. This work reacts directly to this new language and deconstructs its apparent cohesion it into two distinct readings, one nervous, agitated, and one barely moving and blurred into indistinction. This and related work leaves the content open. The artist describes this and related experiments as perceptual protheses. By this he means a work, which the viewer looks through. Meaning is emergent with the realization of the altered perception process.

Paul Vanouse: The Relative Velocity Inscription Engine, 2002

The Relative Velocity Inscription Device (RVID) is a live scientific experiment using the DNA of a multi-racial family of Jamaican descent. The experiment takes the form of an interactive, multi-media installation. The installation consists of a computer-regulated separation gel through which four family members' DNA samples slowly travel. Viewer interactions with an early eugenic publication within the installation allows access to historical precursors of this "race," while a touch-screen display details the results of this particular experiment.

The RVID is an assemblage of 3 different processes that have not previously been combined into a single apparatus in laboratory practice: Gel Electrophoresis, UV florescence imaging, and machine vision. Gel Electrophoresis is a common laboratory procedure for both separating and sequencing DNA, which has been re-purposed in the RVID for racing DNA. Typically, a gel is "imaged" in a special, opaque cabinet that contains UV-light. The scientist then views the DNA bands through a camera (since the DNA glows orange when stained and bathed in UV-light, the camera blocks the harmful invisible UV-light from the eyes of the scientist.) The RVID is built from a combination of UV-emitting clear acrylic and UV-opaque clear acrylic to allow the UV light to make the DNA glow as the experiment runs, while protecting the viewer from the harmful UV-radiation.

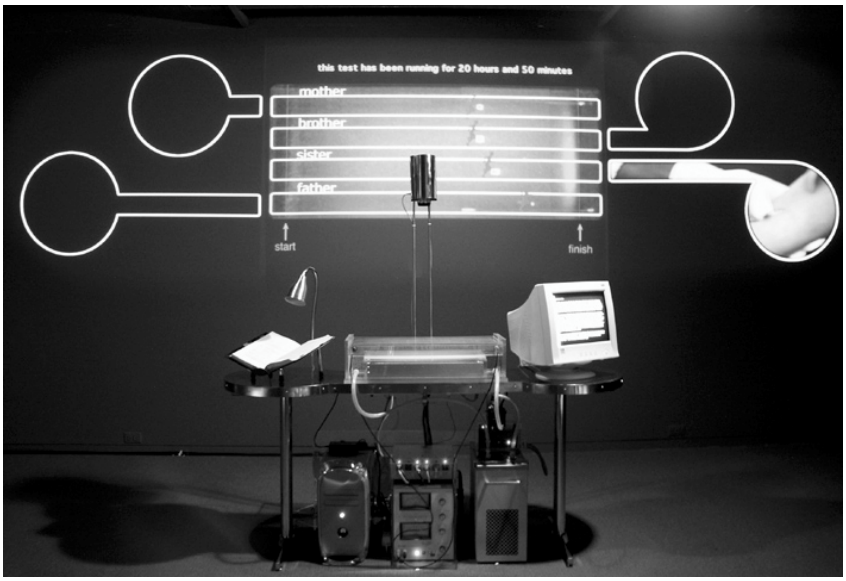

Fig. 2 Relative Velocity Inscription Device (courtesy Paul Vanouse)

The computer-controlled camera periodically grabs images of the glowing DNA, and machine-vision algorithms find each glowing sample. This last step is slightly difficult in the gallery context as background light levels change, the DNA florescence diminishes over time and the coherence of a DNA band is reduced over long periods (2 days) in the gel. The machine vision algorithm runs as follows: First, search the camera image for pixels containing the highest intensity orange values. Then, sort these pixels into groups of adjacent pixels. Then, evaluate which of these groups are brightest and have expected size and shape characteristics to determine the position of each DNA sample. It is through these steps that the software is able to determine the positions of samples at all points in the race and determine the winning sample at the end of each race. Here, machine vision is a neutral observer of a race of races.

Steve Mann: DECONference, 2002

This exhibit attempts to demonstrate how we have become interdependent on technological extensions of the mind and body, and hence, to some degree, have taken a first step to becoming cyborgs. The exhibit also attempts to show how authorities might view the cyborg being as a threat, therefore requiring that it be stripped of these extensions. The cyborg body, whether by way of a wearable computer, or by pens, pencils, portable data organizers, shoes, clothing, eyeglasses, and other personal effects, is potentially contaminated, and thus requires cleansing.

DECONference was conceived as a probe into society, and to understand the culture and technology of mass decontamination. In the exhibit, decontamination was deconstructed by literally building a futuristic mass decontamination facility, as might form the entrance to a space station or airport of the future, or as an entranceway into a high security government building or industrial facility such as a factory. In such a future world, an alleged need for

cleanliness might also be used as justification for a mass decontrabanding (including a search of personal belongings).

Initially all persons entering a "clean facility" are assumed to be potentially contaminated. Since it is not known which if any of these persons are guaranteed to be free of contaminants, everyone must undergo decontamination. One of the many procedures contaminees are exposed to includes automated prefab uniform measurement. This is achieved with a body scanner, comprised of an infrared sensor array and a DEC Alpha supercomputer. It calculates all body dimensions, and computes the optimal design for a uniform. This information can then be used to either custom-tailor a uniform, or to select from a fixed stock of pre-made uniforms. In this exhibit, the uniforms are all pre-made, of white Tyvek.

Fig. 3 Image of the infrared sensor array (courtesy Steve Mann)

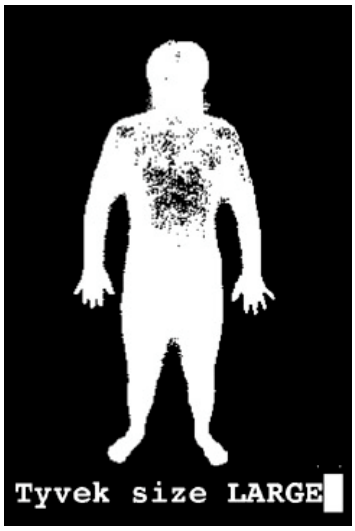

Fig. 4 Screen capture from the body scanner (courtesy Steve Mann)

The extracted body metrics of height, girth, and shoulder width reduce contaminees to a standardized metric and expose them to efficient control. This work employs computer vision, performance and street theatre tactics to create a forced scenario of authoritarian population control. It is a warning sign. Security and control demand a price: freedom of mind and body.

Marc Böhlen: The Open Biometrics Project and its Keeper, 2002-2003

The Open Biometrics Project and its Keeper is an access granting machine and data management system that utilizes finger scanning and pattern matching techniques to access a person's presumed right to enter a restricted site[19]. Of all biometric validation techniques, finger print classification is the most established and entrenched in law enforcement throughout the world. New imaging technologies replace the fingerprint with the digital

finger scan, and use computational similarity measures to match one scan to another. The Keeper makes use of this technical knowledge and differs in its interpretation of it. The Keeper has a defined policy of data acquisition and retention, a particular conception of biometric based uniqueness and works with the laws of data classification within system inherent limitations.

There have been many attempts to find technical solutions to classifying human beings based on body metrics. At least since Lavater, body metrics is a contested field of character validation. Indeed, critique of body, or more generally biometrics, can occur on a number of levels. While there is much circumstantial evidence, for example, that every human being has a distinct set of fingerprints this has never been statistically proven over all populations and peoples. The computerized automation of biometric validation is an even trickier issue. It is one thing to solve an isolated problem in biometric data interpretation, and a very different thing to devise and enforce a large-scale system on all members of a population.

With advances in signal processing and computation, it is becoming very convenient to automate any numerically tractable problem, however questionable the underlying assumptions may be. Numerous government and private agencies are working towards large-scale biometric identification systems. In the near future, no official government document will be issued without a fingerprint or an eye-scan. Of all biometric validation techniques, fingerprint-based validation is the most established and entrenched in law enforcement through out the world. This is where the Keeper comes into play. The Open Biometric Project and its Keeper crack open the clean fabrication of automated biometric identification at its root. Put your finger on this machine and it will show you what kind of information biometric readers extract from humans.

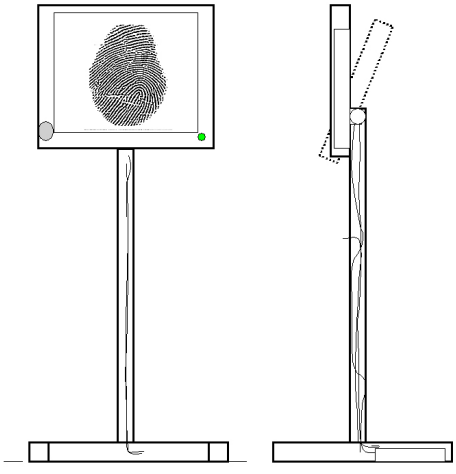

Fig. 5 CAD of the Keeper
(courtesy Marc Böhlen)

Fig. 6 Probable minutiae of finger scan, certified by the Keeper
(courtesy Marc Böhlen)

A fingerprint is made of a series of ridges and furrows on the surface of the finger. The uniqueness of a fingerprint can be determined by the pattern of ridges and furrows as well as the singular or minutiae points, local ridge characteristics that occur at either a ridge bifurcation or a ridge ending. The extraction of the minutiae points from a scan delivers the structural basis of identification. Fingerprint matching techniques that use minutiae-based methods first find minutiae point positions and angles and then compare their relative placements to a reference fingerprint[3]. The constellation and number of minutiae points build the basis for matching a one fingerprint to another. This is a rather delicate matter. Formerly a domain reserved for human forensics experts, minutiae extraction can now be translated into executable computer code. In the machine, both minutiae map and minutiae matching are found within degrees of error and translated into probabilities; an elegant form of quantifying likelihood. However, the results of these mathematical operations generate information that is valid within certain limits and under certain assumptions. The rules of probability theory ensure that the assumptions are computationally tractable. Error is translated into a fraction of unity. There is nothing wrong with this process. It is elegant, intelligent and conceptually sound. But the results are not absolutes; rather a kind of suggestion. While the human in the loop might ponder the uncertainties of an assigned task, the machine is programmed to minimize ambiguities for efficiency and authority. The imperative of erring on the side of caution in times of *Angst* only enforces the tendency to simplify such complex operations. Here, machine vision techniques challenge hard and fast one-way classification that biometric based validation can be misused for.

Fig. 7. The Keeper in use.

Machine Vision as a Method of Perceptual Intervention

A common denominator runs through the works described above. They all challenge the viewer to rethink what they see and ponder the chain of seeing, thinking, believing and behaving. It is easier to understand this development from a more removed vantage point.

In the arts, the history of automated vision is long and complex, and begins possibly in a dark cave. Visual representation is a core creative act, exemplified in ancient Greek history by the legendary Xeuxis whose painted grapes are said to have appeared so realistic that birds flew by to eat them. The desire to encode reality created the need for tools and methods to facilitate the task. The craft of making brushes and preparing pigments, and the formulation of the perspectively encoded third dimension that began almost 1000 years ago with Alhazen and continued in the early Renaissance in Italy with Giotto, Alberti and Masaccio, have all contributed to the automation of vision by the machine.

The last 150 years saw radical additions to this endeavor. Photography, film, cybernetics, and military research redefined the distinction between mimesis and visual perception. Machine vision can be seen as yet another step in the desire to organize visual data; automation principles of the early 20th century carried into visual perception. In this context the image becomes data and information. Image creation becomes data processing and all data processing tools become potential image manipulation operations. Appreciation of automated vision in the arts is often limited to the compositing of visual information. By this I mean the composition of image elements and the filtering and manipulation of global image properties. The well-know image enhancement application 'Photoshop' is an example of image processing software that operates on this level. The Renaissance goal of pictorial representation of reality terminates in Photoshop-like image processing since any object can be seamlessly added or removed from an image. Pictorial veracity is no longer a given. But machine vision allows yet more forms of intervention into visual data. Images can be mined for meaning. Now, images can be queried for content in color, text and object classification, and hence recognition and extraction of meaning. For the first time in the history of pictorial representation and perception, semantics can be extracted from an image by a machine. The data mining of streaming images changes the role of the image in visual culture in a fundamental way. The depth of meaning extractable from images is still shallow compared to human visual intelligence capabilities. The redefinition of mimesis in pictorial representation due to automation and insights from neurobiology and computer science described above extend the discussion of visual representation into the domain of practical philosophy. The use of imaging techniques that see differently is a technical catalyst to thinking differently about seeing. Machine vision can be used as a tool to question the epistemology upon which habitual vision is based and functions thus as a method of practical philosophy or intervention into the unconscious processes of perception. This is a new type of inquiry in the arts. However, it is important to understand that machine vision is very limited. Precision is no substitute for interpretation and accuracy is not related to truth. The convenience of machine vision techniques is seductive.

What lies ahead

The urge to see, interpret and know everything will drive machine vision development in the future. It is not unreasonable to assume that one will soon be able to register and access everything a human being might see in his/her life through synthetic vision. Furthermore, we can expect very high (>500 frames/second) capture and processing speeds from large numbers of networked high definition cameras to redefine the notion of 'real time' image processing. Nothing will be left unseen. One can expect artists to react differently than scientists to these new possibilities.

Parallel to the increase in bandwidth we can expect significant developments in recognition and classification techniques. Researchers at the University of Sussex[20], for example, are investigating the possibility of predicting crowd behavior based on observations from simple surveillance cameras. Together with phenomenal data mining[21], that seeks to find phenomena that give rise to relationships in datasets beyond the apparent relationships themselves, automated prediction may become a contested area of investigation for machine vision artists.

Short Bibliography

[Chen, 1998]
S. Chen, "Learning-based vision and its application to autonomous indoor navigation", PhD Thesis *Department of Computer Science and Engineering, Michigan State University*, 1998.

[Davies 1990]
E. Davies, "Machine Vision: Theory, Algorithms and Practicalities", *Academic Press*, 1990

[Faugeras et al 1998]
O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin, and I. Zoghlami. "3-d reconstruction of urban scenes from image sequences", *Computer Vision and Image Understanding*, 69(3):292-309, March 1998.

[Foley, Dam 1982]
J. Foley, A. van Dam, "Fundamentals of interactive computer graphics", *Addison Wesley*, 1982

[Goodale Humphrey 1998]
M. Goodale and K. Humphrey, "The objects of action and perception", In: *Cognition* 67, 1998, p. 181 - 207.

[Jain, Kasturi, Schunck, 1995]
R. Jain, R. Kasturi, B. Schunck, "Machine Vision", *McGraw-Hill Inc.*,1995.

[Lettvin, Maturana, McCulloch, Pitts 1959]
J.Y. Lettvin, H.R. Maturana, W.S. McCulloch, W.H. Pitts, "What the frog's eye tells the frog's brain", *Proc. Inst. Radio. Eng*., 1959, vol 47, p. 1940 – 1951.

[Mann 2001]
S. Mann, "Intelligent Image Processing", *John Wiley and Sons*, 2001.

[Manovich 2001]
L. Manovich, "Modern Surveillance Machines: Perspective, Radar, 3-D Computer Graphics and Computer Vision", in CTRL [SPACE] - Rhetorics of Surveillance from Bentham to Big Brother, edited by Thomas Y. Levin, Karlsruhe*: ZKM / Zentrum für Kunst und Medientechnologie and Cambridge, Mass.: The MIT Press*, 2001

[Marr 1982]
D. Marr, "Vision", *W.H. Freeman and Company*, 1982

[McCarthy 2000]
J. McCarthy, "Phenomenal Data Mining: From Data to Phenomena", *Computer Science Department, Stanford University,* 2000.

[Pentland and Cipolla 1998]
A. Pentland, R. Cipolla, "Computer Vision for Human-Machine Interaction", *Cambridge University Press*, 1998

[Rowe 1997]
M. Rowe, "The Evolution of Color Vision", The Talk Origins Archive.

[Strang 1986]
G. Strang, "Introduction to Applied Mathematics", *Wellesley-Cambridge Press*, 1986

[Shapiro Stockman 2000]
L. Shapiro and G. Stockman, "Computer Vision", *Prentice-Hal*, 2000.

[Strelow et al 2001]
D. Strelow, J. Mishler, S. Singh, and H. Herman, "Extending Shape-from-Motion to Noncentral Omnidirectional Cameras", *The Robotics Institute, Carnegie Mellon*, 2001

[Troscianko et al 2001]
T. Troscianko, A. Holmes, J. Stillman, M. Mirmehdi, and D. Wright. "Will they have a fight?", In *European Conference on Visual Perception 2001, Perception Vol. 30 Supplement*, pages 72--72. Pion Ltd, August 2001

Citation of works

David Rokeby, Shock Absorber, 2001
Justina M. Barnicke Gallery Toronto 2001, and the Art Gallery of Hamilton, Canada, 2002

Paul Vanouse, The Relative Velocity Inscription Engine, 2002.
Henry Gallery Seattle, and International Symposium of Electronic Arts in Nagoya Japan, 2002.

Steve Mann, DECONference 2002
Performance, September 2002 at the Decon Gallery, Toronto

Marc Böhlen, Keeper of Keys, 2002-2003.
International Symposium of Electronic Arts in Nagoya Japan, 2002, and the Symposium on Language and Encoding at the University of Buffalo, 2002.

Artist biographies

David Rokeby has won acclaim in both artistic and technical fields for his new media artworks. A pioneer in interactive art and an acknowledged innovator in interactive technologies, Rokeby has achieved international recognition as an artist and seen the technologies which he develops for his work given unique applications by a broad range of arts practitioners and medical scientists. Rokeby's best known work, Very Nervous System (1986-90) premiered at the Venice Biennale in 1996, won the first Petro-Canada Award for Media Arts (1988) and is permanently installed in several museums around the world. The technology Rokeby developed for this work is widely used by composers, choreographers, musicians, and artists. It is also used in music therapy applications and is currently being tested as an activity enabler for victims of Parkinson's Disease.
Rokeby has twice been honored with Austria's Prix Ars Electronica Award of Distinction (1991 and 1997). He has been an invited speaker at events around the world, and has published two papers that are required reading in the new media arts faculties of many universities. He recently received a Governor General's Award in Visual and Media Arts.

Paul Vanouse has been working in interactive electronic media since 1990, critically exploring the intersections of big science and popular culture. A fundamental strategy of this work has been to create playful interactive situations for public participation that induce a skeptical ambivalence toward entrenched cultural contructs. He is an Assistant Professor of Art at SUNY Buffalo and a Research Fellow at the Studio for Creative Inquiry at Carnegie Mellon University. Vanouse's electronic cinema, installation and performances have been exhibited in Austria, Brazil, France, Scotland, Belgium, Chile, Spain, the Netherlands, Denmark, Canada, Germany, Australia, New Zealand and widely across the US. While Vanouse often designs his work for public spaces, it has also been exhibited in major museums including: The Carnegie Museum of Art and the Andy Warhol Museum in Pittsburgh, The Walker Art Center in Minneapolis, The TePapa Museum in New Zealand and the Louvre Museum in Paris.

Steve Mann has written 139 research publications (39 journal articles, 37 conference articles, 2 books, 10 book chapters, and 51 patents), and has been the keynote speaker at 24 scientific and industry symposia and conferences and has also been an invited speaker at 52 university Distinguished Lecture Series and colloquia.
Dr. Mann has been working on his WearComp invention for more than 20 years, dating back to his high school days in the 1970s. He brought his inventions and ideas to the Massachusetts Institute of Technology in 1991, and is considered to have brought the seed that later become the MIT Wearable Computing Project. He also built the world's first covert fully functional WearComp with display and camera concealed in ordinary eyeglasses in 1995, for the creation of his award winning documentary ShootingBack. He received his PhD degree from MIT in 1997 for work including the introduction of Humanistic Intelligence. He is also inventor of the Chirplet Transform, a new mathematical framework for signal processing, and of Comparametric Equations, a new mathematical framework for computer mediated reality.
He has also given numerous Keynote Addresses on the subject, including the Keynote at the first International Conference on Wearable Computing, the Keynote at the Virtual

Reality conference, and the Keynote at the McLuhan Conference on Culture and Technology, on the subject of Privacy issues and Wearable Computers.

Marc Böhlen makes machines that reconfigure expectations towards automation processes. After graduating from the Robotics Institute at Carnegie Mellon he was on faculty at the University of California and the Center for Research and Computing in the Arts, both in San Diego, and is currently faculty member at the University of Buffalo in the Department of Media Study. He has presented papers and exhibited artwork nationally and internationally including at the New York Digital Salon, the Andy Warhol Museum, the American Association for Artificial Intelligence (AAAI), the Association for Computing Machinery (ACM), and the Institute for Electrical and Electronics Engineers (IEEE). Recent work has been featured at the iMAGES International Film Festival Toronto 2002, ISEA2002 in Nagoya, the APEX Gallery in New York, Version3.0 at the Art Institute of Chicago 2003, and the International Garden Festival of Grand-Métis, Reford Gardens Québec, 2003.

Endnotes

[1] There is an extensive bibliography on machine vision in general. The interested reader can start at the Computer Vision Homepage of Carnegie Mellon University:
http://www-2.cs.cmu.edu/~cil/v-pubs.html
or the Compendium of Computer Vision at the School of Informatics of the University of Edinburgh:
http://www.dai.ed.ac.uk/CVonline/

[2] Issues particular to computer graphics and automation of perspective are discussed in Manovich's text (see bibliography).

[3] [Pentland and Cipolla 1998]

[4] [Rowe 97]

[5] [Lettvin et al 1959]

[6] [Ungerleider and Mishkin, 1982]

[7] [Goodale and Humphrey 1998]

[8] [Goodale and Humphrey 1998]

[9] [Goodale and Humphrey 1998]

[10] [Marr, 1982]

[11] [Lettvin et al 1959]

[12] Detailed descriptions of all the matrix operations mentioned here can be found in Strang's text (see bibliography).

[13] [Foley and Dam 1982]

[14] A discussion of the particularities of the Laplacian and other operators are beyond the scope of this essay. For details see [Jain, Kasturi, Schunck, 1995]

[15] Negative numbers are clamped to zero.

[16] [Chen 1998]

[17] [Faugeras et al 1998]

[18] [Strelow et al 2001]

[19] KK is created with the help of Richard Pradenas, JT Rinker and Nicolas Canaple

[18] [Troscianko et al 2001]

[21] [McCarthy 2000]