HUMAN STEREOSCOPIC VISION IN ARTIFICIAL RETINAS

HAUPTSEMINAR NEUROENGINEERING

TECHNISCHE UNIVERSITÄT MÜNCHEN

SOMMERSEMESTRE 2014

SUBMITTED BY:
HELENA HALASZ

SUPERVISOR:
MARCELLO MULAS

ABSTRACT

Human beings have the ability to perceive the environment in three dimensions, enabling them to perform tasks that require precise hand-eye coordination and to accomplish motions that other animals that may not be able to see depth cannot do. Due to the placement of the human eyes in the same planar field, the two retinas project the same image from two slightly different angles, known as binocular disparity. This is incredibly difficult to reproduce computationally, for a number of reasons. First, science has not yet determined the neural mechanism of combining two images into a singular one, and how those two images can encode depth. Furthermore, computational energy required to emulate the millions of neurons firing in the visual system is not yet available, and as such, neuromorphic systems cannot yet perform with the same efficacy and speed as the human brain. Additionally, solving the biologically-inspired correspondence problem, that is, the ability of the brain to naturally find matching points in two images, is difficult to emulate.

A number of stereovision systems have been made in recent years, each with a varying degree of visual capabilities: some can navigate through doors while others are able merely to perceive the contours of objects in grey-pixelated form. The following paper discussed the possible solution to the correspondence problem, a stereo-matching algorithm to combine two images, and the development of faster and cheaper stereo-vision systems. With the advent of newer technologies as well as studies using macaque monkeys and cats, scientists are able to draw more knowledge from biology in order to emulate the biological marvel that is the human visual system.

Contents

INTRODUCTION

The human visual system is an incredibly complex neurophysiological process that is still shrouded in mystery. In recent years, researchers have attempted to uncover the inner workings of the visual cortex, in an effort to help the estimated 285 million people worldwide that suffer from blindness and visual impairment regain some of their vision. Of the total, approximately 78% of the impaired have lost eyesight as a result of illness, and thousands of others in accidents or war.[1] Furthermore, advancements in neuromorphic vision systems would enable scientists to replicate the biological processes in humanoid robots, and, furthermore, to potentially lead to enhancements of the human body.

This paper will focus on the perception of depth using stereopsis in artificial retinas, and will look at the visual system anatomy, the geometry of binocular vision, and a number of existing technologies in order to better understand stereoscopic vision.

BACKGROUND:

1.1 Visual System Anatomy

The human retina is an integral part of the central nervous system, as it not only contains the photoreceptors necessary for vision, but also converts the light entering the eye into electrical signals that can be sent to the brain for processing. The light is focused by the cornea and the lens, sending the beam directly to the photoreceptors (rods and cones) that line the retina. The ganglion cells located in front of the photoreceptors are shifted to the sides, as shown in Figure 1 below, in order to allow the light to hit the photoreceptors directly. This location is known as the fovea, and it is the reason that the human eye moves, to shift so that the object of interest is in direct contact with the fovea. To reduce optical distortion until the light reaches the photoreceptors at the back of the eye, the eye cavity is filled with a vitreous humor and the axons of the neurons of are unmyelinated, and thus both relatively transparent.

The absorption of light and its transduction into electrical signals is carried out by the photoreceptors. The visual information is transferred to the ganglion cells via the bipolar cells from the receiving photoreceptors, and the information is transduced via a three-stage biochemical cascade. The ganglion cells are capable of not only detecting weak contrasts and rapid changes in light intensity, but are also attributed to the processing of visual aspects such as movement, fine spatial detail, or color, which, as will be discussed later, are some of the visual cues for depth perception.

[1] Visual impairment and blindness. (n.d.). *WHO*.

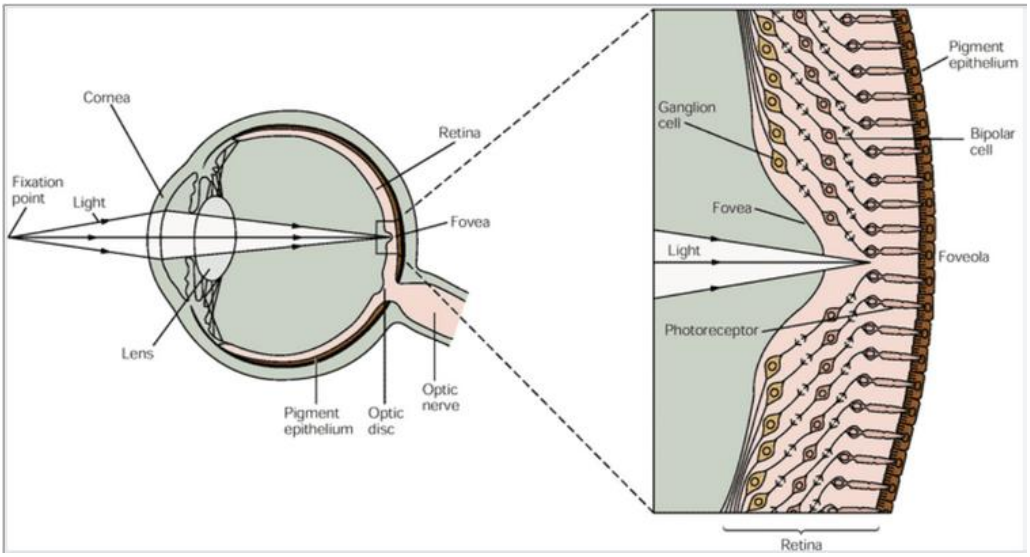


Figure 1 The human eye, showing the light entering through the cornea, going through the vitreous humor and hitting the back of the retina, at the fovea.

At this point, the electrical signals travel to the lateral geniculate nucleus (LGN) of the thalamus in the brain, and from there, to the primary visual cortex (which is also referred to as visual area 1, V1, striate cortex, or Brodmann area 17 in various literature) at the rear of the brain. Visual information is crossed--images from the right side are seen by the left retina and transmitted to the right thalamus, and vice versa.[2]

1.2 Binocular vision:

Vision can be separated into two types: monocular and binocular. The former refers to the usage of the two eyes separately, such as in the case of prey, and gives a wider field of view. The latter is the type of vision that humans, and a number of predators, have, and has various advantages. Binocular vision leads to single vision (the compilation of two distinct images into a singular sight), enlargement of the field of vision, compensation for certain blind spots, and, the focus of this paper, stereopsis.

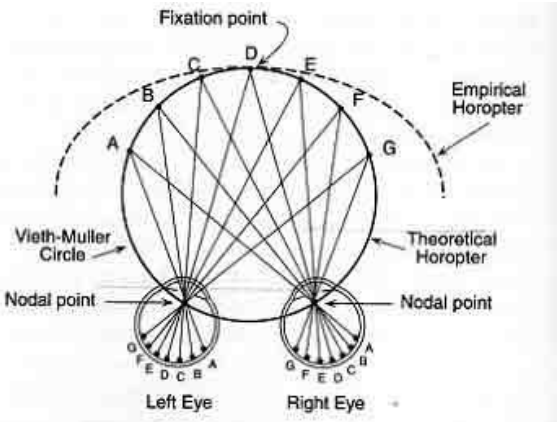
The geometry of binocular vision is very important (Figure 2). The physical horizontal separation of the two eyes, known as parallax, is what leads to 3-dimensional vision. Humans generally have between 50-70 mm between the two eyes, thus affording two views of the world that differ slightly in vantage point. [3] From this disparity, the brain can extract depth information, specifically position of the point in depth due to horizontal disparities. In absolute disparity, the difference in angles, defined for a single point, sends information about that point's position

[2] Prasad, S. Chapter 1 – Anatomy and physiology of the afferent visual system.
[3] Qian, N. Binocular Disparity Review and Perception of Depth

relative to a point of fixation. [Aside: Humans (and monkeys) display much finer depth judgment using relative, not absolute, disparity.][4]

The horopter is the locus of points whose images appear in corresponding locations in the retinas when the eyes are focused. As pictured below, the horopter, which was first described by Vieth and Muller in the 19th century as a circle that crosses the fixation and nodal points of both eyes, actually has a shape that changes with fixation distance. That is, for focus distances below 1 meter, the horopter is concave and is between the frontoparallel plane (not shown) and the Vieth-Muller circle. For focus distances greater than 1 meter, the shape of the horopoter gradually becomes convex. Charles Wheatstone, inventor of the stereoscope, was the first to establish that an object that is between the observer and the horopter creates images that have negative horizontal disparities, while an object past the horopter produces positive disparities (in 1828). Later, it was determined that a horopter has both horizontal as well as vertical components, forming a "space horopter," as discussed by Solomons in a 1975 paper.[5]

Figure 2 The Vieth-Muller circle, also known as the theoretical horopter.



Retinal correspondence is when the fovea share a common direction, and are therefore examining the same image. The topic of correspondence will be discussed at length, as it becomes a challenge in the computations of neurmorophic systems. Finally, sensory fusion occurs; two images that are similar in size, sharpness, brightness and location in the retinal areas are combined into a singular one in the visual cortex. [discuss epipolar geometry?]

1.2.1 Overview of cues for depth perception:

The main difference between depth perception and stereopsis, which will explored in detail, is that the latter depends solely on interocular retinal image differences, while the former can infer depth from both monocular and binocular cues.[6] The human visual system relies on a number of cues, both psychological and physiological, to determine both relative and absolute distances between objects. Certain physiological cues such as binocular parallax and convergence, which will be discussed in detail, require binocular vision, while certain psychological cues are determined with monocular vision.

[4] Orban, G. Extracting 3D structure from disparity.
[5] Gonzalez, F. Neural mechanisms underlying stereoscopic vision.
[6] Ibid.

Cues for depth perception that are strictly monocular cues are the following: [7]

i. relative size of objects, referring to depth perception that occurs when one of two objects that are known to be of similar size appears bigger than the other, such as in the case of trees in a forest appearing closer because they are bigger.
ii. interposition, or partial blocking of an object by another, gives clues to the relative proximities of the objects. An object that is behind another is thus deemed to be further away than the object in the foreground.
iii. aerial perspective refers to depth cues that come from the light that bounces off molecules in the air. Over large distances, the amount of small distortions grows and results in a fainter atmosphere in the distance. Thus, objects that are further away appear blurry and the brain can determine that the sharpest images are closer than blurred ones.
iv. shading, a more subtle cue, allows humans to determine the 3 dimensional position and shape of objects in space. It occurs because the human brain assumes that light comes from above, and computes images thus.
v. geometrical perspective refers to the apparent convergence of parallel lines to a point on the horizon
vi. texture gradient, which means that fine details of objects in the distance cannot be discerned as easily as the texture of objects close at hand.
vii. relative velocity, or motion parallax, a depth cue that cannot be replicated in a painting or image, and refers to the impression that objects that are closer to the subject are moving faster. A prime example is looking out of a moving car, only to see that the road directly beside the subject is speeding by, while mountains in the distance may appear to be stationary.[8]

Studies have shown that binocular cues (ones that require the use of both eyes) for depth perception are more powerful than the above-listed monocular cues. Binocular parallax, binocular convergence and accommodation are three such cues. The first, also known as stereopsis, refers to the horizontal disparity between the two eyes, while convergence refers to when the two eyes face inward in an attempt to view objects at a distance less than 10 meters. The brain calculates the angle of the eyes to determine whether the image being viewed is closer than another, reference, image. Accommodation, the third binocular cue and one that is only valuable at distances less than 2 meters, is an oculomotor cue that results from the ciliary muscles of the eye stretching the lens in order to change the focal length. Information about the contractions and relaxations of the muscles is sent to the visual cortex, which is able to determine distance and depth information.

[7] Gonzalez, F. Neural mechanisms underlying stereoscopic vision.
[8] Teittinen, M. (n.d.). Depth Cues in the Human Visual System.

1.2.2 Stereoscopic vision

The word 'stereoscopic' comes from the Greek word, 'stereos' meaning 'solid.' It refers to the ability of the brain to extract three dimensional information from horizontal disparity. Random-dot stereograms (RDS) are tests developed by Julesz that evaluate stereopsis in human beings. The test consists of two dotted images that are identical, save for a single row that is horizontally displaced, which can only be seen when the two images are viewed with binocular fusion (otherwise, if viewed separately, the displacement is invisible).[9]

1.2.2.1 Neural Basis for Stereoscopic Vision

Although the neural workings of the human stereoscopic vision remains a largely mystery, there have been advancements in the research that give scientists an idea of how stereoscopic vision is processed in the human brain. However, most studies have been conducted on macaque monkeys and cats, rather than humans, and thus the following information regarding the neural basis for stereoscopic vision will be drawn from those sources. Macaques are used due to the similarity of their brain to humans', while research in cat-vision is so abundant that it is convenient to draw conclusions from the existing literature.

Visual information and pathways from the two eyes remain largely independent until they reach the cerebral cortex.[10] The first steps of visual processing occur in the aforementioned V1. This was originally discovered by Barlow *et al.*, when it was realized that only very specific neurons in a cat's striate cortex responded to stimuli that was far away from the point of fixation, and certain *different* neurons responded to stimuli that was closer.[11] Recently, electrophysiological analysis of neural activity in macaques has shown that, contrary to previous beliefs, the neurons in the primary visual cortex (V1) cannot themselves determine depth, rather, they send the information about absolute retinal disparities within small patches of the visual field to where it can be processed.[12] Another study showed that neurons could discern the horizontal displacement of the RDS test, a rather complex image processing operation, already in the second stage of visual cortical processing, which was earlier than expected.[13] Detection of binocular disparities in the visual system must take place in a place where the two ocular inputs converge--therefore, this can happen no earlier than at the striate cortex, which was shown in studies with macaques, as well.

However, researchers do not understand the neural circuitry of sensitivity to the disparities that were reported to the cerebral cortex. There are three possibilities, none of which have been proved yet:

1) intrinsic horizontal pathways in the visual cortex send information coming from the two retinas: This hypothesis is based on the understanding that dendrites of the stellate cells that receive the majority of the input from the LGN could link to cortical cells via horizontal

---

[9] Ohzawa, l. Mechanism of stereoscopic vision: the disparity energy model.
[10] Nikara, T. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex.
[11] Gonzalez, F. Neural mechanisms underlying stereoscopic vision.
[12] Backus, B. Human Cortical Activity Correlates With Stereoscopic Depth Perception.
[13] Ohzawa, l. Mechanism of stereoscopic vision: the disparity energy model.

arborizations (branching). These cortical cells, which are already responsible for orientation tuning and the enhancement of surrounding interactions, could also link together disparate portions of the visual image, and thus be connected to the detection of binocular disparity.

2) inputs from cells in the LGN converge on the same cell in the visual cortex: Due to the nature of geniculate axons (that is, their restriction of terminal arbors in both parvo- and magnocellular layers), it cannot be determined if geniculate cells with differing receptive fields actually connect to the same cortical cell. However, there is a possibility that information regarding disparity from the two retinas is in fact integrated at an unseen stage by individual cortical cells.

3) feedback from higher systems carries information on retinal disparities: Given that the V1 area of the brain receives feedback from multiple other parts of the brain, scientists tested to see which loops contained information regarding disparity and found that the V2, V3 and MT areas all had some cells that were sensitive to the difference in angles of the two retinae.[14]

It is important also to look at the types of disparity sensitive cells and how they behave. In 1977, scientists Poggio and Fischer determined, by moving dark and bright bars at different depths in front of awake monkeys, that the cells could be grouped into four categories: tuned excitatory, tuned inhibitor, near and far cells. Of the 142 cells that were determined to be in the V1 and V2, 84% responded differentially to the 3-dimensionally moving stimuli, and 55% were found to be tuned excitatory, 12% tuned inhibitory, 28% near/far and a remaining 5% that could not be identified to belong to any of the four groups. Two years later, the same team identified that 74% of the cells in those same two areas of the brain were sensitive to horizontal disparity.[15]

A bit about the logic of neural connections in the brain. Most mature sensory maps have a so-called topographic organization, which means that pre-synaptic (afferent) neurons synapse on post-synaptic (target) neurons that are in the physical proximity. As such, spatially nearby target neurons only receive information from physically nearby afferent neurons. Competition arises when more than one afferent neuron is mapped onto a singular target neuron--for example, in the primary visual cortex, when various laminae of the LGN compete for control of visual cortical cells that were originally equally controlled by both eyes, but eventually, became dominated by one of the eyes. Thus, nearby cortical cells are typically controlled by spatially nearby ganglion cells, but the control can change quickly from one eye to the other if need arises.[16]

[14] Gonzalez, F. Neural mechanisms underlying stereoscopic vision.
[15] Ibid.
[16] Elliott, T. Developing a robot visual system using a biologically inspired model of neuronal development.

NEUROMORPHIC SYSTEMS

2.1 Neuromorphic systems specifically the silicon retina

Scientists are having trouble mimicking the human brain using supercomputers for various reasons. Even current state-of-the-art computer vision systems sample the world at a constant rate, resulting in series of static snapshots that contain highly redundant information. The low dynamic content makes transferring visual data, as well as storing and processing it, a fruitless and time-intensive process. Analyzing motion, which seems to come easily to humans, would require the use of several central processing units, and the computation time would be so slow it would not allow interaction in real time.[17] Developing an artificial retina as a neuromorphic system (that is, a computational system that attempts to mimic the biological processes present in the human body) would allow scientists to imitate the massively parallel, data-drive and asynchronous behavior of the human eye.[18]

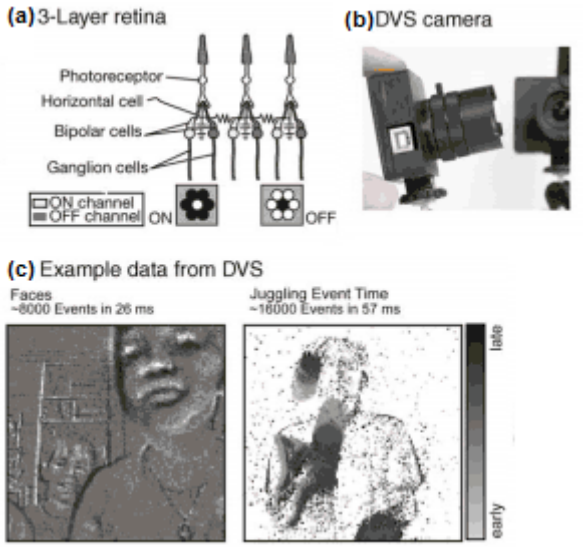


Figure 3 The basics of the silicon retina system. (a) A simplified 3-layer retina model, showing the positive/negative ON/OFF channels. (b) The DVS camera, which is connected to the computer interface via USB. (c) Data collected from DVS system shows that contrast changes for features as fine as faces [left] can be visible (using relative event time), and dotted spatiotemporal results (using space-time technique) can be achieved [right].

The retina does not see in snapshots, as standard imagers do, and, contrary to conventional cameras, the human brain is composed of innumerable slow asynchronous neural components that combine both analog and digital signal representations.[19] In standard imagers an external process regularly polls the sensor array at a rather high rate in order to capture all frequencies of interest, thus creating very redundant data. While traditional sensors regularly output frame-sampled intensity values, silicon retinas are fabricated to mimic the local processing, local gain control and asynchronous spike transmission properties that make the human eye so unique and successful.[20] Furthermore, silicon retinas have a higher dynamic range in illumination than conventional cameras (more than 10^5 compared to about 300, respectively) and higher sampling rates (>1000kHz versus less than 60 Hz).[21] For an overview of a simplified silicon retina system, please see Figure 3.

---

[17] Bartolozzi, C. eMorph: Towards Neuromorphic Robotic Vision.
[18] Wilson, H. R. Neural models of stereoscopic vision
[19] Delbruck, T., & Liu, S. Neuromorphic sensory systems.
[20] Ibid.
[21] Ibid.

2.1.1 Silicon as the material of choice:

Silicon is currently used to fabricate analog and digital computing chips, as well as neuromorphic electronic devices, such as the artificial retina. The transistor through which current flows, representing the circuitry of nervous systems, is the primary silicon primitive, because the material shares a lot of the same physics as neurons. [22] Furthermore, the human body does not reject the material, which is crucial. In studies done with silicon retinal implant microchip, it was determined that patients did not complain of any rejection, infection, erosion/chafing, inflammation, neovascularization, or retinal detachment or migration, demonstrating the efficacy of silicon as a biocompatible material.[23]

2.2 Address-event representation (AER):

Representing the human neuronal network with interconnected neurons on a silicon chip can pose quite a challenge. For one, the transistors are placed on a 2-dimensional substrate which allows only for a few layers, while in reality this limitation is not applicable, since neurons are in a 3-dimensional environment. Furthermore, making point-to-point neuronal connections is time-consuming and expensive, and the circuitry remains fixed after the initial chip fabrication, thus limiting the possible architectures that could be expressed.

On the other hand, the speed of current flow in a transistor is 10^7 times faster than the speed of ions in the human body. [Although, it is important to note that unpredictable current variability in transistors proves to be a hindrance because it can severely reduce precision--this aspect requires further improvements]. Even so, the tremendous difference in mobility is very advantageous, and is used to make up for the fixed circuitry limitation in neuromorphic engineering (mentioned above) by changing the mode of signal transmission. Instead of having axons connecting to dendrites via synapses, as occurs in neurons, the team at the University of Zurich assigned each of the "neurons" on the chips an address, which is transmitted off-chip containing *what* (which neuron) and *when* (when it fired) information. Because each 16-bit address can transmit spikes from up to 2^{16}, a huge wealth of information can be simultaneously transmitted before another neuron spikes. Additionally, access to such detailed information paves the way for further processing [expand].[24]

Unlike in conventional sensors, in an AER system the retina pixels act autonomously in deciding whether collected information is worthy of transmission. This mimicry of the biological system ensures that only non-redundant events are transmitted, which decreases computational time and increases efficiency in power dissipation, and because pixels autonomously initiate communication, outputs are able to be transmitted with short latencies [latency being defined as the delay between stimulation and response]. Furthermore, cost of processing data outputted

---

[22] Delbruck, T., & Liu, S. Neuromorphic sensory systems.
[23] Chow, A. Y. The Artificial Silicon Retina Microchip For The Treatment Of Vision Loss From Retinitis Pigmentosa.
[24] Delbruck, T., & Liu, S. Neuromorphic sensory systems.

from silicon retinas can be reduced by a factor of 100, because the event-driven computing results in less redundant data to be evaluated.[25]

2.2.1 Addressing Challenges

A paper by Tobi Delbruck of the Institute of Neuroinformatics, UNI-ETH Zurich, raised some questions that explore some of the continuing limitations in AER technology. For example, why is it that, to this date, there are no reliable color vision sensors? Past attempts have failed because color separation based on wavelengths-absorption has proven very weak, and traditional color filter technology has only recently become available for use, and is, unfortunately, still very expensive. Furthermore, is it possible to simultaneously reduce pixel redundancy in spatial, temporal and spectral stimulus environments?[26]

2.3 Event-based dynamic vision sensor (eDVS)

The eDVS is a specific event-driven AER sensor that focuses on detecting contrast changes and outputs them as a sequence of digital pulses. As in an AER system, each pixel autonomously responds to varying contrasts, "achieving a wide intra-scene dynamical range" that enables it to respond to larger ranges of contrast. The neural-inspired method is a novel way of encoding light and the temporal variations by focusing only on transmitting changes in the scene at the time they occur.[27] By taking advantage of the dynamic characteristics of the visual field, an eDVS system is much more efficient and uses less memory, because it is able to digitize, transmit and store only the pixels that have changed. Furthermore, the sensitivity to relative light intensity changes enables it to function in uncontrolled lighting, such as outside, giving mobile robots more freedom.[28]

2.4 Neural Inspired Algorithm:

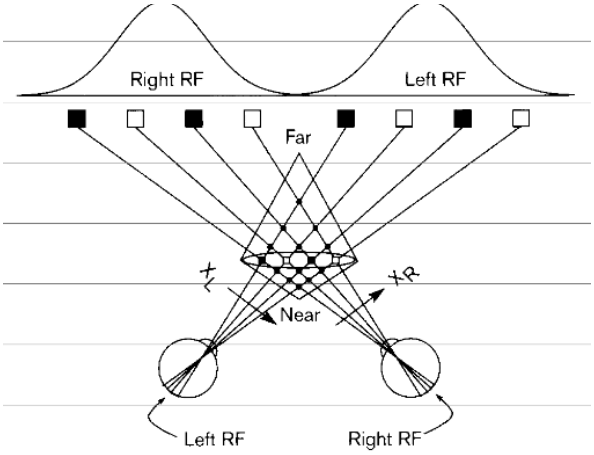
2.4.1 The Correspondence Problem

As discussed above, the human eye naturally matches the left and right scenes into a singular image, but even biologically, the process is a difficult one. In neuromorphic computing, for each pixel in the left image, the corresponding pixel must be found in the right image, and matched to create a stereoscopic image. This can be quite difficult to achieve, especially because there are no visible monocular features that could guide the process.[29]

---

[25] Ibid.
[26] Delbrück, T. Activity-Driven, Event-Based Vision Sensors.
[27] Benosman, R. Neuromorphic computer vision: overcoming 3D limitations.
[28] Piatkowska, E. Asynchronous Stereo Vision for Event-Driven Dynamic Stereo Sensor Using an Adaptive Cooperative Approach.
[29] Ohzawa, l. Mechanism of stereoscopic vision: the disparity energy model.

Figure 4 Diagram of the receptive fields (RF) of binocular cells, showing a row of targets (contained within the ellipse) and all the potential matches/false matches. This depiction was originally developed by Julesz and his team



The field of view, or receptive field (RF) of a singular visual cell, contains numerous image features, which, when crossed with the RF of another visual cell, could have several intersecting points (see Figure 4 to the left). For a row of intended targets, such as those in the ellipse, there are innumerable matching points, depicted by the intersecting points within the larger diamond shape. All of the matches lying outside the ellipse can thus be considered false matches, and a neuromorphic stereovision system should be able to consolidate the intersecting points, determine the false matches and find a consistent solution. For computation purposes, a filtering command could focus only on the points within the ellipse for analysis, thus resulting in responses only to the actual target matches.

Registering the contours of images is the mere beginning of achieving neuromorphic stereovision. Next, the human eye determines shape in the third dimension, which is a process that does not take place in the V1. Computationally, to determine borders based on differences of depth, an antagonistic (subtractive) convergence of V1 complex cells' outputs may be employed but the 1998 Ohzawa paper did not delve into that.

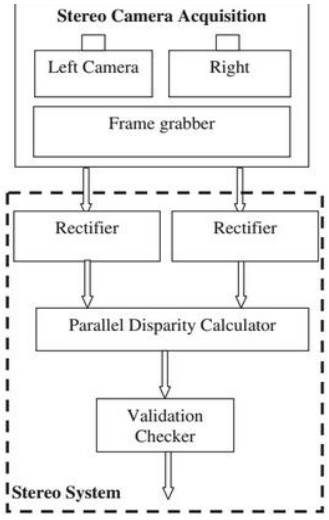
The above gives an overview of why the correspondence problem is an issue in neuromorpic computing. A 2012 journal article by Paolo Zicari team developed a low-cost stereo vision system, and implemented a process called rectification, in which the correspondence problem can be simplified into a one-dimensional computation. That is, given that *P* is a generic point in the scene and PRl(xl,y) and PRl(xr,y) are the coordinates of the 2D projection of the left and right images, respectively, the pixels are horizontally aligned. Using simple triangulation, the disparity can be accurately calculated. Errors in the disparity map can occur due to occlusions (pixels that are visible in only the right or left images, not both) and false matches (errors due to noise, varying exposures and lack of texture), but even so, the rectification was able to help simplify the problem.[30]

The stereo vision system works in three main steps: pre-processing, stereo matching and post-processing (Figure 5). The first phase aims to rectify the tangential and radial distortion that results from the very nature of camera lenses, and uses a pre-determined MATLAB Calibration process to apply distortion correction to the raw images taken by the stereo camera. In the second phase, an SAD-based stereo matching algorithm is used.

[30] Zicari, P. Low-cost FPGA stereo vision system for real time disparity maps calculation.

The CC method, which requires the computation of two disparity maps, is the most common approach, but due to this, it is also incredibly computationally intensive. Because the disparity map *disp* is first attained by running the matching algorithm from the left image (the reference) to the right (the candidate), and then run again from right to left, the algorithm is actually executed twice. The authors discuss a few different approaches found in literature, such as the uniqueness check method (UC), employed by authors L. Di Stefano and M. Marchionni, has a reportedly lower computational cost, since the algorithm is only ran once. But they themselves used an injective consistency check method (IC) in their proposed stereo system.[31]

Figure 5 Diagram of the Zicari stereo camera.



First and foremost, vectors and matrices that depend on the intrinsic (principal point and the perspective projection matrix) and extrinsic parameters (rotation and translation vectors) are preliminarily obtained. This helps define the orientation and physical position relationships between the images taken by the two cameras. The team used the following 6-step algorithm to rectify generic pixels:

$$\textbf{(1)}\quad \begin{bmatrix} x_{raw} \\ y_{raw} \\ 1 \end{bmatrix} = KK \times \begin{bmatrix} xd \\ yd \\ 1 \end{bmatrix}$$

$$\textbf{(2)}\quad \begin{bmatrix} xd \\ yd \end{bmatrix} = (1 + k(1)\cdot r2 + k(2)\cdot r4 + k(5)\cdot r6)\cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} k(3)\cdot a1 + k(4)\cdot a2 \\ k(3)\cdot a3 + k(4)\cdot a1 \end{bmatrix}$$

$$\textbf{(3)}\quad \begin{aligned} &a1 = 2\cdot x\cdot y; \quad a2 = r2 + 2\cdot x^2; \quad a3 = r2 + 2\cdot y^2 \\ &r2 = x^2 + y^2; \quad r4 = r2^2; \quad r6 = r2^3 \end{aligned}$$

$$\textbf{(4)}\quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} xx/zz \\ yy/zz \end{bmatrix}$$

$$\textbf{(5)}\quad \begin{bmatrix} xx \\ yy \\ zz \end{bmatrix} = M \times \begin{bmatrix} x_{rect} \\ y_{rect} \\ 1 \end{bmatrix}$$

The algorithmic computation is performed individually for each pixel in both the left and right images. Then, SAD is used to calculate the parallel disparity between the reference and candidate images. This method is efficient because it does not require heavy arithmetic

[31] Zicari, P. Low-cost FPGA stereo vision system for real time disparity maps calculation.

calculations, and enables the easy design of parallel structures. A *W x W* window of pixels is used, where *d* refers to the varying distances between the values of Min*d* and Max*d*:

$$SAD(xr,y,d) = \sum_{i=-\frac{W-1}{2}}^{\frac{W-1}{2}} \sum_{j=-\frac{W-1}{2}}^{\frac{W-1}{2}} |PRr(xr+i,y+j) - PRl(xr+d+i,y+j)|$$

Finally, the team employed a cost-effective disparity validation circuit, which used a purpose-designed asymmetric consistency check instead of the aforementioned CC, such that only one disparity map needs to be analyzed.

$$\exists k | (xr - Nc \leqslant k < xr) \text{ and } xr + disp(xr,y) = k + disp(k,y)$$

$$\exists k | (xr - Nc \leqslant k < xr) \text{ and } xr - k = disp(k,y) - disp(xr,y)$$

The two equations ensure that the disparity values computed for the reference pixels in the right image that are the same as points in the left image are overlooked, thus minimizing computational time.[32]

Robotic Applications

A number of teams have implemented the above concepts in revolutionary bio-inspired technologies, as well as non-neuromorphic systems, in both standard cameras and artificial vision systems, and in this section we will explore a few case studies of robotic applications within the past twenty-five years.

3.1 Trinocular Stereo Algorithm in *José*

While a lot of focus has been placed on replicating the neural connections of the human eye, some teams have attempted to achieve stereoscopic vision with non-neuromorphic systems. Don Murray, et. al, of the Computer Science Department of the University of Britich Columbia developed a working stereovision based mobile robot, dubbed *José*, that has the ability to autonomously navigate unknown surroundings while mapping an occupancy grid. While historically mobile robots have used sonar detection or laser sensors for landmark sensing, *José* uses a multibaseline stereo technique originally developed by Okutomi and Kanade in 1993 that compares three instead of two images. A triangular head with three wide-angle cameras takes three images, and each pixel in the pre-determined reference image (for instance, the right image) is compared with pixels along the epipolar lines in the left and top images. A SAD algorithm, comparing the left/right and top/right image pairs, results in a combined score, and because the technique uses two comparisons, unlike the binocular method, the probability of error is reduced, as it is less likely to have the same mismatches in the same exact pixels. Furthermore, validation algorithms are run to improve the results. A "sufficient texture" test verifies that there is adequate variation in the pixels being compared, while a "quality of match"

[32] Zicari, P. Low-cost FPGA stereo vision system for real time disparity maps calculation.

test normalizes the sum of all scores for a specific pixel, and if the value is not below a threshold (that can be changed via parameters), the pixel match is deemed to inadequate.

While the experiment was deemed successful, there is a key improvement that is needed. The debilitating challenge of looking past, under or over objects remains in occupancy grid mapping done autonomous navigation of mobile robots. Damaging collisions can occur when certain objects, i.e. tables, are partially occluded because the left/right image pair cannot successfully match the features of the obstacle because the pixels are aligned with the edge, and the top/right image pair views the table too differently due to occlusion and angular differences. [33]

3.2 Advanced Driver Assistance Systems

A silicon retina-based stereo vision system, developed as a pre-crash warning for side automotive impacts, uses an analog chip that outputs intensity changes. The system, which can assist with high beam maintenance, can detect cars approaching to prevent collisions and can help drivers merge lanes or park, uses a bio-inspired 3-dimensional vision system. For each pixel that exceeds the pre-determined intensity change threshold, the retina sends information containing the coordinates of the pixel (x,y), a timestamp (because the AE can occur at any time), and the polarity signaling rising intensity ("on" event) or falling intensity ("off" event). For the setting of the threshold, 12 bias voltages are available in the retina, and every pixel is connected to the others via an analog circuit, allowing for the measuring of the intensity. [insert examples from study if needed].[34]

The advantages of using silicon retina sensor as a pre-crash sensor is that this kind of sensor overcomes limitations of traditional vision systems by being able to sustain a 1ms resolution and functioning in a variety of lighting conditions, with a dynamic range of approximately 120dB. This allows it to react quickly to real-time events, and because it is able to do some minor pre-processing in the sensor itself, it reduces computing power and memory requirements.[35] Such advantages for a pre-crash sensor can naturally be tied back to improving stereoscopic vision in the artificial retina, and so successful results, such as those published in the Austrian Institute of Technology by Jurgen Kogler, is promising.

3.3 Neuromorphic computer vision system from the Insititute of Neuromorphic Engineering

A team lead by Ryad Benosman at the INE developed a complete sterevision event-based framework with which to compute depth based on two asynchronously functioning silicon retinas. Using two eDVS models with an AER of 128x128 pixels, the system was programmed to detect changes in contrast that crossed the threshold of a 15% intensity difference. Unlike

[33] Murray, D. Using real-time stereo vision for mobile robot navigation.
[34] Kogler, J. Bio-inspired stereo vision system with silicon retina imagers.
[35] Ibid.

traditional cameras, an eDVS system can time events with a faster resolution of 1 microsecond and a frame rate of several kilohertz. Like in general AER systems, changes in the scene are represented by a +1 or -1 polarity (negative and positive contrast changes, respectively) and a lack of response denotes redundant visual information that is thus not transmitted.

Before data is collected, the system is calibrated to estimate the distance between the two cameras, and while the position of the cameras do not change, the system requires only a single calibration. Then, upon receiving visual information from the two cameras, the scenes are matched based on identifying scene point projections, relying on the similarities of neighboring pixels' gray levels.

3.4 Low-cost stereo vision system

Zicari, et. al, compared current technologies in stereo vision to a proposed system, evaluating efficiency of software and hardware in real time implementation and bringing the focus to cost-reduction. The authors of the paper considered achieving high-speed systems as important as finding low-cost solutions, and presented a stereo vision system employing a single XilinxVirtex-4 XC4VLX60 field-programmable gate array (FPGA) chip, as opposed to the XC4VLX200 chip used in a paper by S. Jin (link to FPGA design and implementation of a real time stereo vision system), which had a commercial cost that was nine times that of the proposed system.

Experimentally, it was shown that the IC method (discussed in detail above) allowed for lower percentages of error than the UC method, and that the performance of the proposed system was not altogether dissimilar from the far more expensive CC method. Furthermore, the IC method seemed to be less sensitive to the size of the windows than the UC one. Overall, the team was able to demonstrate that, not only was the proposed system of better quality, but it was also less expensive.

3.5 Humanoid vision system on iCub robot

The robot developed by the Italian Institute of Technology uses AER and employs a data processer as well as two asynchronous eDVS technologies in its visual sensory system. The team went further, to design an asynchronous sensor that can evaluate both temporal and spatial contrast.[36] The image resolution of the vision system was kept at a minimum, of 64x32 pixels, and the robot was able to compute the relative distance of objects, as well as its own position and velocity, resulting in slight propioperception.[37] But, because the aim of the experiment was to create a humanoid robot with maximum degrees of freedom and humanlike-abilities (such as reaching for objects), the researchers were more focused on replicating human behavior than just developing a visual system.

---

[36] Bartolozzi, C. eMorph: Towards Neuromorphic Robotic Vision.
[37] Metta. BabyBot -- Robot Modelled On Two Year-old Child -- Takes First Steps.

3.5.1 Robot Navigation Algorithm

In an aside: once the robot can combine binocular images, the next step is using that information for navigation. A team lead by Sylvain Lecorne, France, developed an algorithm to analyze environmental information and allow the robot to find obstacles, such as doors. In the simulation (for now, it is simply a simulation)it first calculates the displacement of relevant pixels, finds the distances between the points and applies a filter to discover where exactly the emptiness (a.ka. door) is located. [Another algorithm, for purely navigational purposes, allows the robot to drive through the doorway.]

The contrast based algorithm finds the contrast of a specific point but using a Sum of Difference calculation in red, green and blue between the target point and the surrounding pixels. The output is given by the intensity of a grey point, meaning that points with high contrast appear light and those with low contrast appear dark. The formulas employed by the team are the following:

$$Contrast(x, y) = \sum_{i,j} Diference(Pixel(x, y), Pixel(x + i, y + j))$$

$$Diference\ (p1, p2) = (abs(p1.red - p2.red) + abs(p1.blue - p2.blue) + abs(p1.green - p2.green))/3$$ [38]

During the span of the experiment, the robot was able to survey its surroundings in a complete 360 degree turn, locate the door in the room and successfully navigate through the door into an adjoining room, where it completed a similar procedure to navigate into the next room. With such technological advancements, the future of robotic navigation has the potential to become increasingly more autonomous.

DISCUSSION AND CONCLUSIONS

While the development of stereoscopic vision systems has improved drastically over the years, there yet remains a lot of work to be done.

4.1 Potential improvements

The human brain is estimated to be made up of 86 billions of neurons, an unknown percentage of which contribute to the processing of visual information.[39] Replicating such vast amounts of connections will require far stronger computers, but, with the advent of new technologies every day, immense sensory system advancements will severely impact the development of artificial retinas capable of stereoscopic vision. Even now, multi-neuron chip

[38] Lecorné, S. l., & Weitzenfeld, A. Robot navigation using stereo-vision.
[39] Herculano-Houzel, S. Equal Numbers Of Neuronal And Nonneuronal Cells Make The Human Brain An Isometrically Scaled-up Primate Brain.

systems are growing in size as the technology is developed further, which is evidenced by an as-of-yet unpublished working sensory system that has 65,000 neurons on a singular chip. Additionally, it is believed that multi-chip sensory systems can exhibit cortical visual properties akin to those of the human eye, such as stereopsis, motion sensing by tracking and orientation selectivity.[40]

A 16-electrode retina implant that is currently running human trials showed that, although the learning process was incredibly slow, all six patients with the implant benefitted from a restoration of light perception. Another implant, with 60 electrodes, is in Phase II of clinical trials, and results have not been determined yet. In the meantime, devices with 250+ electrodes are being developed, and theoretical modeling has calculated that an implant with 1000 electrodes could have the potential to restore relatively good functional vision, to the point where the patient that had previously little-to-no eyesight, could regain the ability to read and recognize faces.[41]

4.2 Challenges

One of the major challenges that researchers are faced with is the lack of knowledge about the neural processing of visual information. A vast majority of the literature today relies on animal studies to give a background on how the human brain handles vision, and thus, various assumptions between humans and non-human animals, that may or may not hold true, must be made. Furthermore, primate visual processing systems may differ from cats', thus negating the conclusions that were drawn under the assumption that the two are very similar.[42] Until more effective imaging technology is developed or human trials can be done, it may be impossible to know the inner workings of the brain, enough to accurately replicate it.

Additionally, the inherent property of adaptability found in the human body is sorely lacking in neuromorphic systems, and as such, these systems are unable to alter operating parameters when faced with a capricious environment.[43] However, advancements in the topic are being pursued. A team in the Department of Electronics and Computer Science at the University of Southhampton attempted to implement a biologically inspired synaptic plasticity model on a robot with a binocular vision system. The full algorithmic model for neural plasticity, while not entirely relevant to the topic at hand, showed that robots can independently develop bio-inspired topographic maps, and can be found here.[44]

---

[40] Delbruck, T., & Liu, S. Neuromorphic sensory systems.
[41] Chader, G. J. Artificial vision: needs, functioning, and testing of a retinal electronic prosthesis.
[42] Backus, B. Human Cortical Activity Correlates With Stereoscopic Depth Perception.
[43] Delbruck, T., & Liu, S. Neuromorphic sensory systems.
[44] Elliott, T. Developing a robot visual system using a biologically inspired model of neuronal development.

BIBLIOGRAPHY

Axenie, C., Conradt, J. (2014) Cortically inspired sensor fusion network for mobile robot egomotion estimation. Robotics and Autonomous Systems, Special Issue on “Emerging Spatial Competences: From Machine Perception to Sensorimotor Intelligence”, preprint.

Backus, B. Human Cortical Activity Correlates With Stereoscopic Depth Perception. *Department of Psychology, University of Pennsylvania*.

Bartolozzi, C. eMorph: Towards Neuromorphic Robotic Vision. *Procedia Computer Science, 7*, 163-165.

Benosman, R. Neuromorphic computer vision: overcoming 3D limitations. *Vision Institute Pierre and Marie Curie University*.

Chader, G. J. Artificial vision: needs, functioning, and testing of a retinal electronic prosthesis. *Neurotherapy: Progress in Restorative Neuroscience and Neurology, 175*, 317–332.

Chow, A. Y. The Artificial Silicon Retina Microchip For The Treatment Of Vision Loss From Retinitis Pigmentosa. *Archives of Ophthalmology*, *122*, 460-469.

Conradt, J., Simon, P., Pescatore, M., and Verschure, PFMJ. (2002). Saliency Maps Operating on Stereo Images Detect Landmarks and their Distance, Int. Conference on Artificial Neural Networks (ICANN2002), p. 795-800, Madrid, Spain.

Delbrück, T. Activity-Driven, Event-Based Vision Sensors. *Inst. of Neuroinformatics, UNI - ETH* .

Delbruck, T., & Liu, S. Neuromorphic sensory systems. *Current Opinion in Neurobiology, 20*, 288-295.

Elliott, T. Developing a robot visual system using a biologically inspired model of neuronal development. *Robotics and Autonomous Systems, 45*, 111-130.

Gonzalez, F. Neural mechanisms underlying stereoscopic vision. *Progress in Neurobiology, 55*, 191-224.

Herculano-Houzel, S. Equal Numbers Of Neuronal And Nonneuronal Cells Make The Human Brain An Isometrically Scaled-up Primate Brain. *The Journal of Comparative Neurology, 513*, 532-541.

IST Results. (2006, May 2). BabyBot -- Robot Modelled On Two Year-old Child -- Takes First Steps. ScienceDaily. Retrieved June 27, 2014 from www.sciencedaily.com/releases/2006/05/060502173527.htm

Kogler, J. Bio-inspired stereo vision system with silicon retina imagers. *AIT Austrian Institute of Technology GmbH*.

Lecorné, S. l., & Weitzenfeld, A. Robot navigation using stereo-vision. *ENSEIRB, France, ITAM, Mexico*.

Murray, D. Using real-time stereo vision for mobile robot navigation. *Computer Science Dept. , University of British Columbia Vancouver*.

Nikara, T. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Experimental Brain Research, 6*, 353-372.

Ohzawa, l. Mechanism of stereoscopic vision: the disparity energy model. *Current Opinion in Neurobiology , 8*, 509-515.

Orban, G. Extracting 3D structure from disparity. *Trends in Neurosciences*, 466-473.

Piatkowska, E. Asynchronous Stereo Vision for Event-Driven Dynamic Stereo Sensor Using an Adaptive Cooperative Approach. *IEEE, AIT Austrian Institute of Technology: Safety and Security Department*.

Prasad, S. Chapter 1 – Anatomy and physiology of the afferent visual system. *Handbook of Clinical Neurology, 102*, 3-19.

Qian, N. Binocular Disparity Review and the Perception of Depth. *Neuron, 18*, 359–368.

Teittinen, M. (n.d.). Depth Cues in the Human Visual System. *Depth Cues in the Human Visual System*. Retrieved April 29, 2014, from http://www.hitl.washington.edu/scivw/EVE/III.A.1.c.DepthCues.html

Visual impairment and blindness. (n.d.). *WHO*. Retrieved June 4, 2014, from http://www.who.int/mediacentre/factsheets/fs282/en /

Wilson, H. R. Neural models of stereoscopic vision. *Trends in Neurosciences, 14*, 445-452.

Zicari, P. Low-cost FPGA stereo vision system for real time disparity maps calculation. *Microprocessors and Microsystems*, 281-288.