Bridging the Gap Between Vision and Commercial Applications

Richard A. Young
Dept. of Analytic Process/AP50, GM R&D Center
30500 Mound Rd., Warren, MI 48090-9055

Abstract

There is a large gap between academic vision research and industrial machine vision applications. A cross-fertilization between academic and industrial researchers is of mutual benefit: it leads to practical applications of benefit to industry, as well as offering real-world challenges that deepen research understanding. A cross-disciplinary atmosphere between academia and industry should be encouraged by appropriate funding sources in government, academia, and industry. Three real-life projects in the author's experience at General Motors demonstrate the potential pay-offs of such an approach. An understanding of human hyperacuity mechanisms (i.e., vernier acuity, or the ability to judge distances between lines at higher accuracy than the resolution limit of the human photoreceptors) led to a subpixel method for finding circuit board locations to extremely high accuracy using a low-cost machine vision system (Young, SPIE Proc., v. 728, 1986). Research 011 the neural basis of primate color vision led to a proposed new type of color display, based on opponent mechanisms, that should be cheaper than conventional RGB displays (Young, SPIE, v. 1250, 1990). Also, a study of primate motion vision led to a proposed new low-cost motion sensor (Young & Lesperance, SPIE Proc., v. 1913, 1993). The vision machines on the plant floor, and the sensor and display devices needed for vehicles of the 21st century, will greatly benefit from closer interaction between academic and industrial researchers.

Keywords: vision, subpixel, hyperacuity, color, motion, neural networks

1. Introduction

Both academically-oriented and industrially-oriented colleagues must share their ideas and experiences for the U.S. to achieve globally competitive advantages in its core manufacturing industries. I propose a future seminar with attendees encouraged to provide examples from their own experience of bridging the gap between academic research and industrial applications. Ways to improve the transition from academic research to industrial applications should also be discussed. The pursuit of the application of knowledge is something which gives benefit to all companies, to academic research which must ultimately test its ideas for scientific advancement, and to society in general. A free-wheeling discussion of how to better bridge the gap should be encouraged as part of the program of this proposed seminar, with the results of the discussion to be summarized in an appropriate medium for wide-spread distribution.

I will describe several real-life cases where the academic study of human vision either led to or could lead to practical industrial applications in machine vision and electronic imaging technology. Using models and knowledge from the academic study of human vision, my colleagues and I at General Motors developed a number of industrial applications with advantages over previous engineering approaches to electronic imaging and computer vision technology. I will first briefly review the basics of early biological vision. Then I will describe three projects:

- Supersight: An advance in fast, highly accurate computer vision. The study of human hyperacuity mechanisms led to a subpixel method for finding printed circuit board locations to high accuracy.
- Opponent-color display: Basic research on primate color vision led to a proposed new low-cost opponent-color liquid crystal display for automobile dashboards.
- A speed sensor: A basic research study of primate motion vision led to a proposed new low-cost speed and direction sensor.

2. Oh Say Can You See? The Physiology of Vision

One way of developing advanced machine vision systems—suitable for applications ranging from assembly robot control to "smart cars" in intelligent vehicle highway systems—is to imitate the computational mechanisms of human vision.

A simple, concise model of the fundamental mechanisms of vision as observed in the primate eye and brain is described in this paper. Biological vision provides a basis for efficient, robust machine vision systems for form, motion, color, and potentially for stereo and all other types of machine vision as well.

How can we see? One answer lies in the receptive fields of visual cells in our eyes and brain—the regions where light "turns on" a nerve signal.

Millions of such fields analyze and filter the patterns of light that fall on the retina. The output of these fields provides the basis upon which conscious visual perception is eventually constructed by higher brain processes. That is, perception itself is derived from the information as filtered and analyzed by such fields (Figure 1).

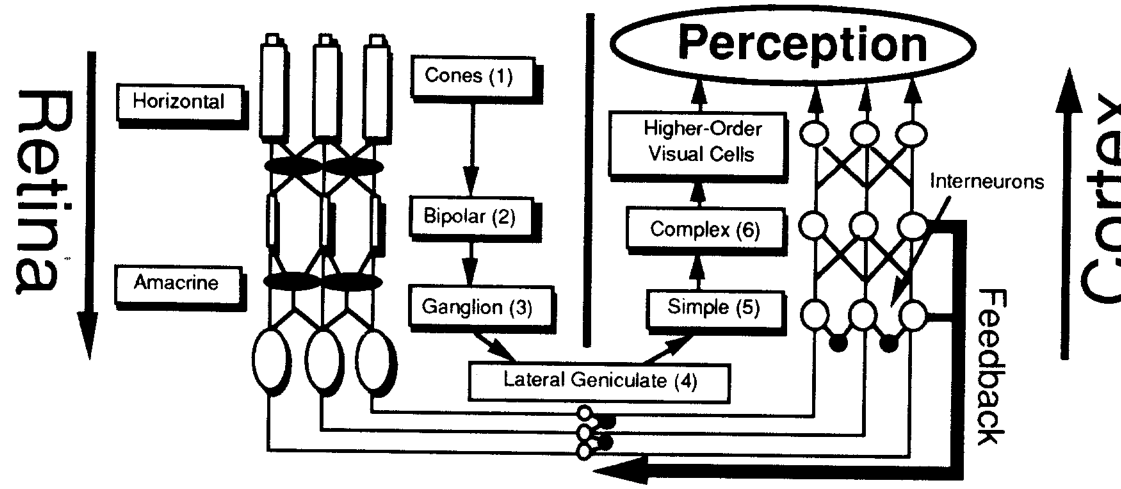

Figure 1. The primary visual pathway consists of retinal (left) and cortical (right) stages of processing. Light falls on cones (1), and nerve impulses are transmitted through bipolar and ganglion cells to the Lateral Geniculate Nucleus or LGN (4). The simple and complex cells (5 and 6) filter the information for "higher order" processing. At the final level of processing, perception arises by an unknown process that may be based upon all the neural activity in the brain acting together as a whole. The Gaussian derivative model so far has the same functional properties as "early vision" from stages (1) through (5).

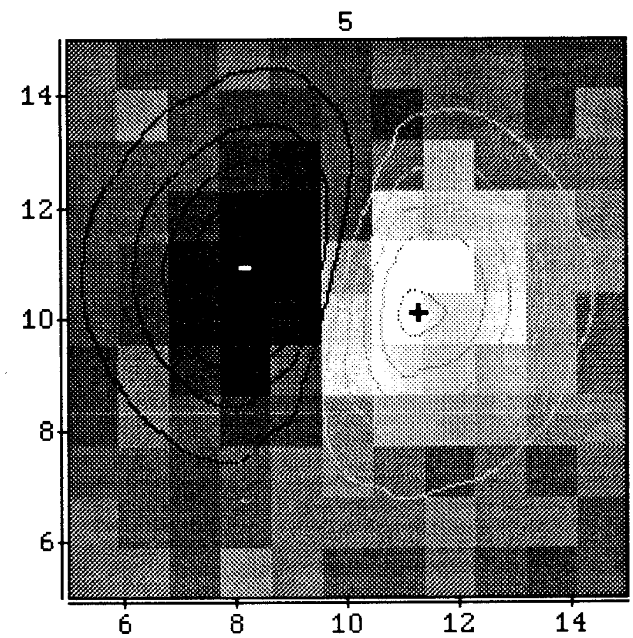

Figure 2. Response of visual simple cell (Stage 5 in schematic on Figure 1) to a small spot of light differs depending on the precise location on the retina where the fight falls. This graph is a receive field map for a particular cell. (The horizontal axis is the left-right direction on a video screen in front of the animal at which a small spot of light about 1/2 sec of visual angle is turned on. The vertical axis is the up-down direction on the screen.) The intensity of the cell's response to the small spot of light is measured at different locations. Dark squares represent areas that actually decrease the cell's response below the resting level of activity; lightest squares represent the most intense positive response. From this map, we know that this cell will respond strongly to a light-dark border or "edge" placed at about an 80° orientation. The contour plot represents the best fit of the Gaussian first derivative function created by the model described in this paper. This filter provides a near-optimum first-stage edge analysis for a vision system.

Young[1] illustrates the major anatomical structures and physiological processes underlying visual receptive fields. Theoretical models of those field shapes were also tested and investigated.

The Gaussian derivative model provided the simplest and most concise description of the receptive fields of the models tested (for example, see Figure 7). Gaussian derivative machine vision spatiotemporal filters, based upon the biological data, produced robust estimates of the spatial and temporal derivatives of the image. Such filters are suitable for many types of vision analysis, using only linear, separable Gaussian derivative filters or their linear combinations.

So a partial answer to "How can we see?" may be that receptive fields in the early visual system serve as robust derivative analyzers in space and time. Using such analyzers in an artificial vision system may help robots (or cars) to "see."

3. Supersight: An Advance In Fast, Accurate Computer Vision

3.1 Biological Vision and the Nyquist Resolution Limit

The human vision system has limited resolution capability since there are individual receptors in the eye rather than a continuous "sheet" of photoreceptive material. The "cones" are the receptor cells in the retina which we use for the initial stage of daylight vision. They receive the light entering the eye and translate it into electrical signals for transmission to later stages of vision in the eye and brain. In Figure 3 you can see a picture of cones in the central region of the eye. The spacing between the cones is seen to be less than 10 microns, the length of the marker at the top of the figure. This inter-cone distance translates in the human eye into an inter-cone spacing of about 30 seconds of arc. There are thus about 120 cones per degree of visual angle in the central region of the human eye.

Here is an over-simplified explanation of the visual resolution limit. If we shine a light source so as to put alternate white and black bars across each column of cones,

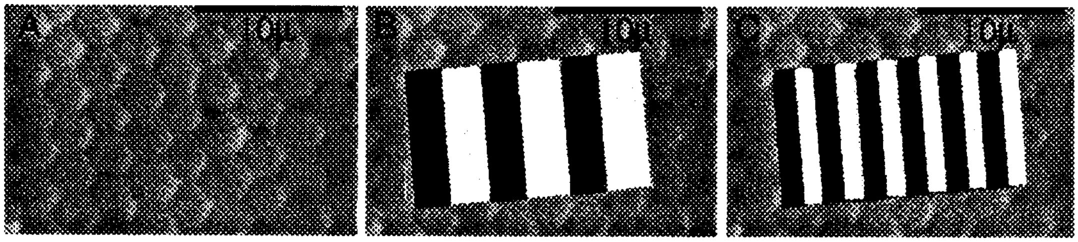

Figure 3. A. End-on view of human cones in central fovea.[2] B. A grating at the resolution limit of the human eye. C. A grating beyond the resolution limit of the eye.

you would expect to distinguish the bars as separate if one column had white bars, the next column had black bars, the next column had white bars, and so forth (Figure 3B). The visual system need only compare whether one column of cones next to another column of cones had a different activation level to say whether the bar grating was present or not.

Any spacing of the bars closer than these alternating columns would lead to an inability to discriminate the white and black bars. To understand this better, consider what would happen at the extreme where both the white bars and black bars were entirely within one column of cones (Figure 3C). The bars would not be discriminable from each other, since the white and black bars fell entirely within one cone and their effects would cancel. No cone would be activated more than its neighbor. Only by simulating separate cones differently can something be discriminated as distinct from something else. Therefore, the minimum spacing at the resolution limit would approximate alternating columns of cones as in Figure 3B.

Each cone is about 30 seconds of visual angle. One white column plus one black column would therefore total 30 seconds plus 30 seconds of angle, or a total of 1 minute of visual angle. That means that we could put about 60 black/white pairs of lines in 1 degree of visual angle, and still cover alternating columns of cones. This simple analysis gives a reasonable approximation to the resolution limit of the human eye, which is a grating of about 60 cycles per degree, when tested psychophysically.

That means that if we viewed a "picket fence" at a sufficient distance so that the spacing between the pickets was less than 1 minute of arc, we would not be able to distinguish the individual bars. We would see the fence as a uniform gray area. In general, this resolution limit is known as the Nyquist resolution limit.

The same Nyquist resolution limit applies to a machine-vision system, since video cameras likewise have individual "receptors"—the individual solid-state photodetectors used to receive the light.

Although we cannot determine *resolutions* greater than the Nyquist limit, can we determine the location of objects to an accuracy greater than the Nyquist limit? This issue is that of *subpixel accuracy*. The basic question is: is it possible?

It is necessary to understand that there is a fundamental distinction between "resolution" and "accuracy." *Resolution* refers to the discrimination of multiple stimuli in the field of view. *Accuracy* refers to the ability to determine the absolute or relative locations of the stimuli.

At first consideration, accuracies smaller than the size of the detectors does seem impossible. Consider an extremely small spot whose image is so well-focused that it falls entirely within the boundaries of a single detector element. The detector element would then give the same response regardless of the spot location, making any such subpixel accuracy technique impossible.

However, if the spot of light extends across more than one detector element (which is typically the case because of lens diffraction, which spreads the light out), then information across multiple detector elements can be combined, and higher accuracies than the Nyquist resolution limit should theoretically be obtainable, even for small spots. In theory, there is no upper bound on the accuracy that can be achieved, with the only limiting factor being noise.

So subpixel *resolution* is impossible for a stationary camera or eye; subpixel *accuracy* may be possible. The question is: for a single spot or line, or for a pair of spots or lines, can the relative accuracy be determined at a value finer than the Nyquist limit?

In a machine vision system, for lines or edges or any other extended object, multiple pixels will always be activated, and hence subpixel accuracies should theoretically be achievable by the techniques to be described. These techniques require some *a priori* knowledge of the sought object, but this is always the case for industrial applications where the desired goal is to locate an object or a feature of an object.

The biological visual system likely uses such techniques to achieve its capability of about one-tenth of a receptor element for tasks requiring the relative alignment of two dots or two lines. This human capability is familiar in the task of aligning two lines in vernier instrument dials. Such "vernier acuity", or "hyperacuity" as it is often called in the biological vision literature, has been studied for many years.[3,4] The question is: "Can we make use of this biological knowledge to build a machine vision system with subpixel capability?"

3.2 Machine Vision Subpixel Capability

Ask any manufacturing specialist the attributes a computer vision system needs in order to win friends and influence engineers, and he'd probably say: "It better be fast, accurate, cheap, and easy to use." Engineers familiar with computer vision systems know that is asking for a lot. Usually, high accuracy equates with unacceptably slow image processing times and relatively costly hardware.

At GM, however, we have developed a machine perception system that only needs a low-resolution video camera to locate industrial parts with high accuracy. In fact, this new vision method (called SUPERSIGHT) can locate important image features with accuracies higher than that implied by the pixel resolution of the camera. Image-processing algorithms substitute for conventional high-resolution cameras and extensive electronics.

The image processing is fast, too. Employing a camera with a low resolution of only 100 by 100 picture elements (pixels), SUPERSIGHT can, for example, locate circles in a test pattern to a one-tenth-of-a-pixel or better accuracy. With conventional techniques, achieving this accuracy requires a 1000 by 1000 pixel camera, requiring the processing of a million pixels. Since the SUPERSIGHT method processes 100 times fewer pixels, it's virtually 100 times faster, and the camera is much lower cost than higher-resolution cameras.

At the time of its development in the early 1980's, the sub-pixel accuracy we obtained with low-resolution cameras represented a new level of speed and performance capability for the computer vision industry. Today, subpixel methods are routine in many commercially available machine vision systems.

3.3 General Approach

The key elements of the SUPERSIGHT approach are new digital image filters for noise rejection, and statistical methods that combine information across pixels to locate predefined image features.

The digital filters—called directional Gaussian derivative filters—are applied in two steps. The initial Gaussian filter acts as a low-pass filter screening out high-frequency noise generated internally by the camera. This filter blurs or smoothes the image slightly. Next, the computer applies a derivative filter that sharpens up the edges and, with a coarse accuracy, locates the edges in the previously blurred image. The result: a simpler, less noisy, but still accurate representation of the information in the image.

In the next step, clustering algorithms are engaged that separate the edge pixels that lie on the object sought from those that do not. The way these algorithms work varies depending on whether the object sought is a straight line or a circle. Once the pixels are clustered, special regression routines (developed by Wes Meyer of the GM R & D Mathematics Department) fit the clustered circle or line pixels to subpixel accuracy levels—this is the "true" edge.

SUPERSIGHT resembles typical machine vision programs in that it analyzes images in terms of a stored model. However, rather than an image model, SUPERSIGHT stores a mathematical model which is the key that enables the method to, in effect, transcend the implied pixel resolution of the camera and achieve subpixel accuracy. The current version of SUPERSIGHT is set up to locate lines or circles with the aid of the clustering algorithms that segment edge pixels and the regression routine that fits the segmented pixels to a model. Working in the same GM R & D Center laboratory, Ronald M. Lesperance has generalized these clustering methods so they can recognize any arbitrary shape. The associated regression routines which would work on any arbitrary shape have yet to be developed.

The SUPERSIGHT algorithm is illustrated in Fig. 4 (next page).

3.4 Applications

This type of computer vision—very fast and highly accurate at a relatively low cost—is ideal for dimensional measurement of parts as well as part location tasks.

To this end, GM has packaged the software for use in the "DelcoVision II" machine vision system developed by our subsidiary Delco Electronics in Kokomo, Indiana. Delco Electronics has put SUPERSIGHT software on-line to locate circuit boards to a 0.001 inch accuracy in an automated manufacturing operation. To date, several million printed circuit boards for car radios have been made with the help of SUPERSIGHT. We have also used the SUPERSIGHT methods for a flexible manufacturing approach to assembling alternator housings.

4. Opponent-Color Display: A Low-Cost Color Display

Reconfigurable color displays are an attractive option for instrument panels, radio, and heat/ventilation/air-conditioning displays, and for new displays such as navigation and night vision. Designers have thus far been stymied in attempts to bring these color displays into production vehicles, largely because of the high cost of traditional three-channel "RGB" (red-green-blue) technology. An opponent-color display will reduce these costs by using only two channels of color information, yet still preserving a wide range of perceived colors.

"How much color detail do we really need in a display?" I asked. The human brain processes 90% of the color information in the neurons of the brain in two channels only: a broad-band channel, interpreted by the brain as black-and-white, and an opponent channel, interpreted by the brain as blue-and-red[5] (see Figure 5). The weighted combination of these two channels gives rise to a wide variety of perceived colors, containing 90% of the color information that we can see in the natural world.

I worked with GM R&D physicists and electrical engineers to develop a two-channel reconfigurable display that uses liquid crystals and color dichroic polarizing film technology to control levels of black-white and red-blue at each pixel. A simulation of the display produces a quite believable color image.[5,7]

4.1 Two-Channel Color—Methods and Results

Land-type. We first tried red and white color channels expanding upon the well-known two-channel photographic approach proposed many years ago by Edwin Land.[8] We simulated a Land-type matrix addressable display with 256 by 256 pixels on a standard color cathode ray tube. The resultant image produces a wide range of perceived colors—not just red and white, but also greens, browns, etc.

Opponent-type. But why white and red channels? One brightness and one *color opponent* channel would better describe the major two chromatic components in the visual system.[6] (An *opponent* channel is one which switches between complementary colors such as red and cyan). Would a liquid crystal display using an opponent method cause even more colors to be seen than with the Land method? Yes—consider Figure 6.

The *brightness* signal is displayed in the even diagonal pixels (blank squares in Figure 6) as white/black. It looks

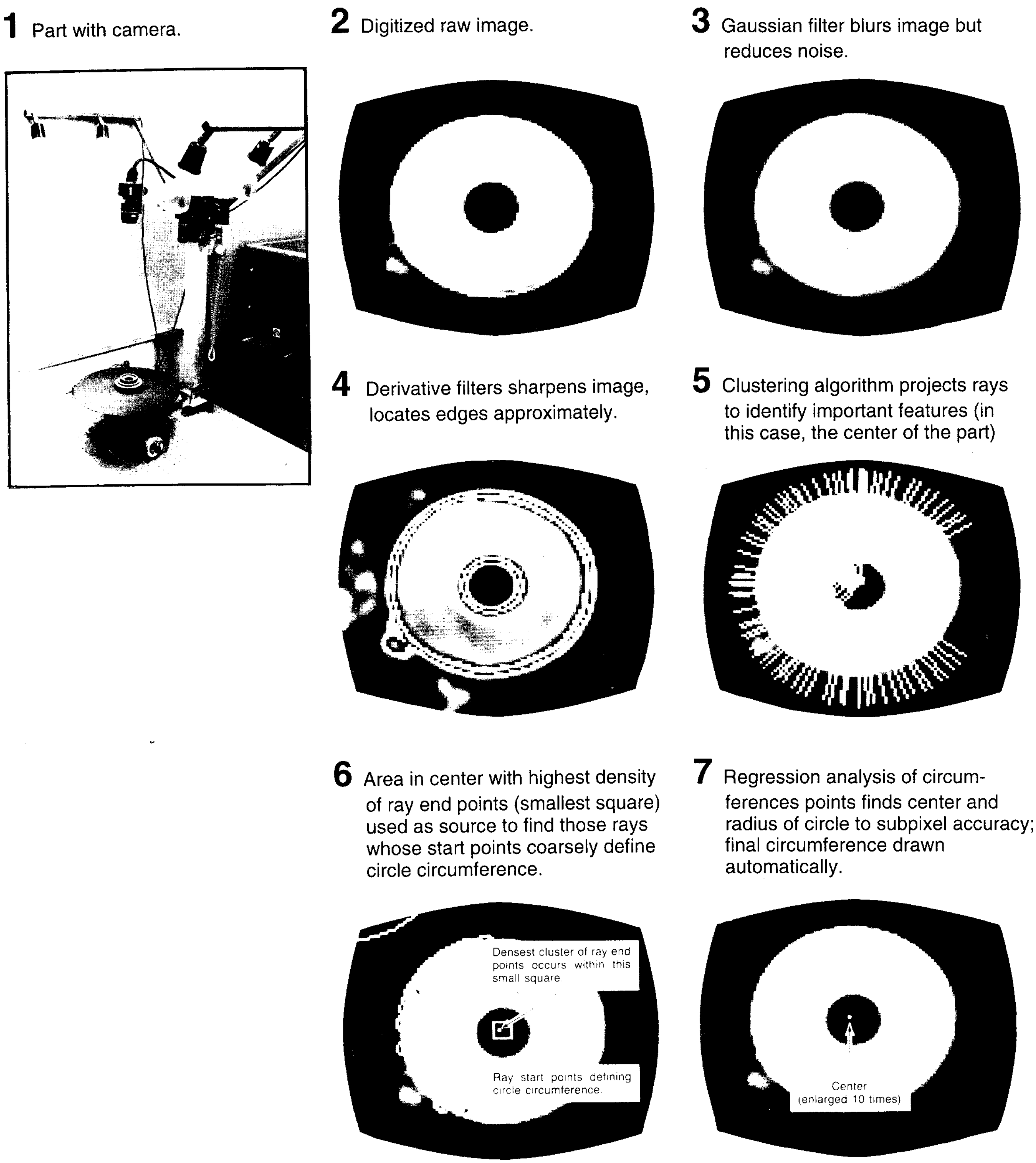

Figure 4. Using SUPERSIGHT to locate the center of a part to sub-pixel accuracy.

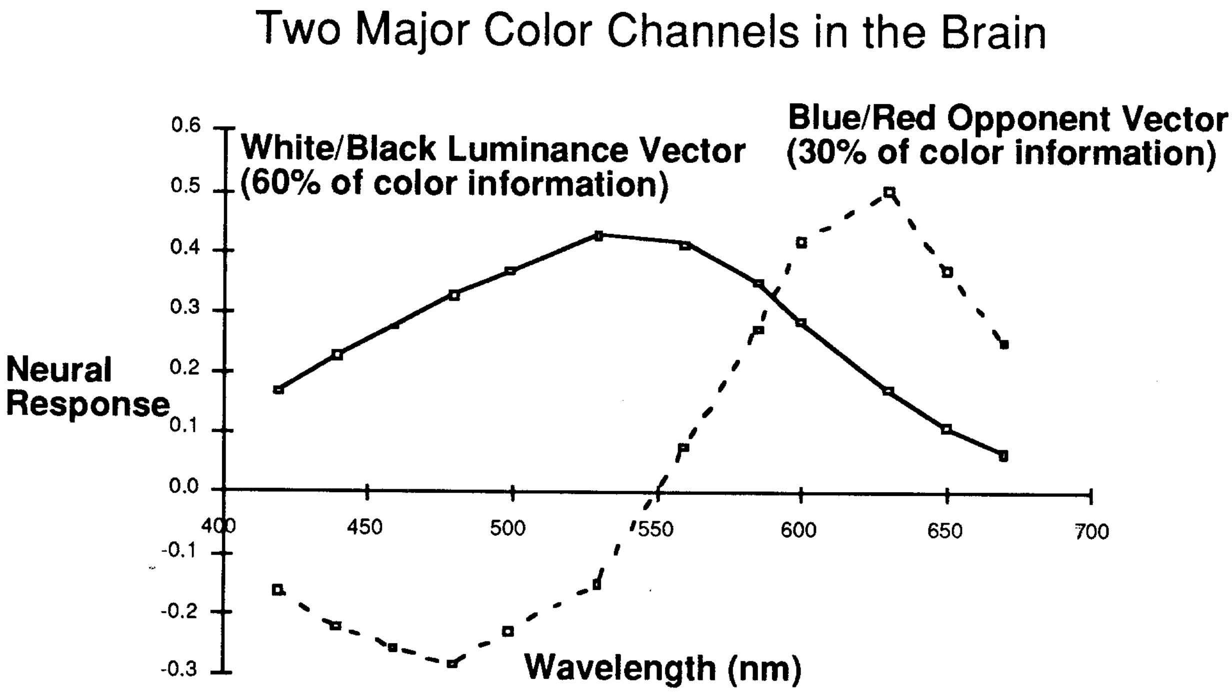

Figure 5. The two major color channels in the primate brain.[6]

like a black/white photo if seen by itself. The color *opponent* signal is displayed in the odd diagonal pixels (labeled "R/B" for "Red/Blue" in Figure 6). This channel is the red and blue opponent signal at each location. If red is more intense than blue in the image, the pixel becomes more red. If blue is more intense than red, the pixel becomes more blue. This color opposition can be produced by a dichroic polarizer placed on the liquid crystal cell. A test of the two methods using color photographs of the displays shows that about 75% of observers see a wider range of colors in the opponent-type than in the Land-type display.

R/B		R/B		R/B	
	R/B		R/B		R/B
R/B		R/B		R/B	
	R/B		R/B		R/B
R/B		R/B		R/B	
	R/B		R/B		R/B

Figure 6. Opponent-color filter. The first channel is displayed in the even diagonal pixels (blank squares). It is formed by taking the white/black (brightness) portion of the original image, and displaying it as white/black. The second channel is displayed in the odd diagonal pixels (squares labeled "R/B" for "Red/Blue"). This channel is formed by subtracting the red and blue portions of the original image (see text). The squares are small enough so the observer's eye visually merges the white/black and red/blue channels. When presented in combination, a wide variety or colors are seen.

4.2 Potential Opponent-Color Display Advantages

The new opponent-color imaging technology if used for automobile direct-view displays (instrument panels) might offer several advantages over RGB displays:

- Costs less and brighter than RGB liquid-crystal displays;
- Good color representation (90% of RGB gamut expected);
- Easier to package than CRTs (thin, flat, low weight);
- Safe (low voltage, no vacuum compared to CRTs);
- Faster driver reaction than to vacuum fluorescent or other monochrome-type displays;
- Better spatial resolution than vacuum fluorescent displays.

4.3 Two-Channel Opponent Color—Conclusions and Recommendations

Two-channel opponent-color displays offer cost or resolution advantages over conventional three-channel color displays, yet may still retain a wide range of perceived color. A prototype liquid-crystal two-channel display should be developed to test this prediction.

Car instrument panels that are reconfigurable offer a competitive advantage over conventional fixed segment panels. Advances in technology will make it more cost effective to produce a common hardware display and then customize it with software, than to manufacture specific instrument panels for every car model. Liquid crystal displays are preferable to cathode ray tubes in cars because of their smaller space. Color enhances quickness of information discrimination by the driver. But ordinary color liquid crystal displays require three channels and are expensive. A proposed new opponent-color technique should create a wide range of colors in liquid crystal displays with only two color channels.

5. A Proposed New Motion Sensor

5.1 Physiological Basis of Motion

The ability to see movement is common among all visual animals. Next to the simple detection of light or dark, the detection of motion may be the oldest and most basic of visual capabilities. In humans, motion is involved in a number of critical functions: detecting moving objects, segmenting scenes, extracting visual depth from moving objects, navigating in an environment, avoiding collisions, determining self-motion, and tracking moving objects via eye and head movements.

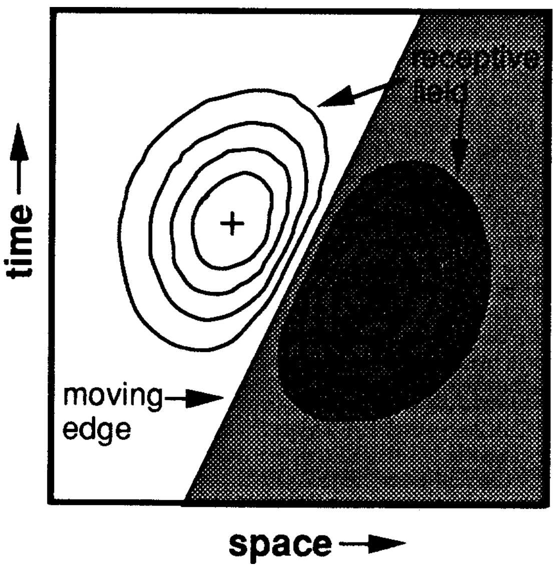

Figure 7. Space-time plot (space is the horizontal axis, time is the vertical axis) of a moving edge and a Gaussian first derivative-like receptive field, optimized to analyze the edge's motion. Note that we have the same shape of field as in Figure 2, only now the field is oriented in space-time rather than just oriented in space, as previously.

How do we see motion? To answer this question, we must first understand the physiological properties of simple cell visual receptive fields in the brain, the first place at which sensitivity to the direction of motion arises in the visual pathway of the primate. We therefore studied the spatiotemporal shape of "receptive fields" of simple cells in the monkey visual cortex. Receptive fields are maps of the regions in space and time that affect a cell's electrical responses. Fields with no change in shape over time responded to all directions of motion; fields with changing shape over time (see Figure 7) responded to only some directions of motion.

We tested a new model[9] that describes not only the static spatial properties of these cells, but their motion properties as well. This model is an extension of the GD model for static spatial vision.[2] We compared the new GD motion model with other models of receptive fields in visual cortex. We found that a GD model fit these fields well, in a transformed variable space that aligned the centers and principal axes of the field and model in space-time (Figure 7). The model accounts for fields that vary in orientation, location, spatial scale, motion properties, and number of lobes. The model requires only ten parameters (the minimum possible) to describe fields in two dimensions of space and one of time. Figure 8 shows a canonical view of such fields. A difference-of-offset-Gaussians (DOOG) provides a plausible physiological means to form GD model fields.[1]

5.2 Machine Implementation

Can a machine analyze motion the same way we do? Because of its simplicity, the GD model improves the efficiency of machine vision systems for analyzing motion. We implemented a simplified version of the GD motion model in a working machine vision system. The system produced robust local estimates of the direction and speed of moving objects in real scenes. By so doing, we not only discovered a more efficient computational algorithm for motion, but also furthered our understanding of the possible functional significance of biological motion fields.

5.3 Sensor Implementation

With this research based on the biology, we now have a highly efficient computational method to measure the speed and direction of moving edges.

It would be useful to extend this method to determine the speed and direction of moving textures, and then implement the method in a simple, inexpensive visual sensor if possible. Such a sensor might be used, for example, to

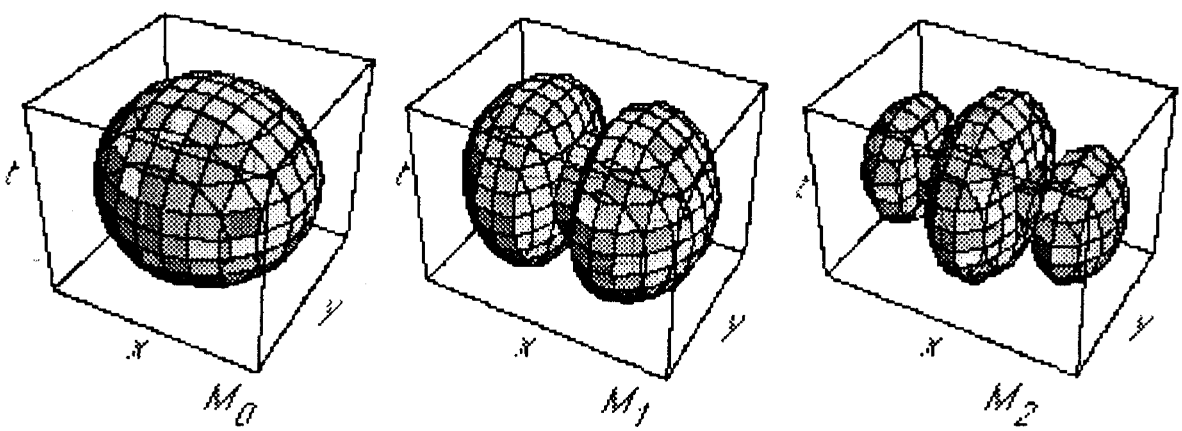

Figure 8. 3-Dimensional view of canonical Gaussian Derivative fields in space and time. This one set of canonical fields, when rotated in space and time, provides a good fit to all simple cell receptive field types, both directionally-selective and non-directionally-selective, of blob, edge, and bar types, that have so far been observed in primate visual cortex.[9]

measure the motion of the road under vehicles. If the true ground speed and direction of the vehicle could then be determined, it would permit improved braking and steering control, and hence vehicle safety.

Some general development objectives for a generic "ideal" speed sensor for use in the transportation industry are that it must be capable of:

- Operating over a wide variety of surfaces, including ice, snow, and gravel;
- Maintaining an accuracy within required margins over a wide range of speeds;
- Operating under the range of intensities of light arising from surfaces, including night use, given that an illuminating light can be used; and
- Providing the signal output at a bandwidth and delay time suitable for real-time control of the vehicle.

A "true ground speed sensor" that is reliable, accurate, and low-cost has been a "holy grail" for vehicle engineers for a long time. Technologies to date are too expensive (e.g. radar or inertial guidance units), too prone to error, or to bulky to be useful for commercial vehicles. There is therefore a need for further research and development in this area.

Since motion is so basic, there are many other important national priorities that would benefit from such an improved motion sensor. These include applications in:

- Intelligent vehicles—collision warning, traffic speed monitoring, and crankshaft speed sensing for misfire detection;
- Manufacturing—monitoring of material handling and robot motion control;
- Security—monitoring direction and speed of possible intruders;
- Fluid dynamics—measuring particle motion to verify fluid flow;
- Medical diagnostics—measuring speed of blood flow.

6. Summary and Conclusions

Industrial vision engineers generally seek out practical solutions to their immediate plant needs, with little or no knowledge of the academic vision research, either of a biological or engineering variety. Likewise, academic researchers spend lifetimes in pursuit of furthering an understanding of the biological mechanisms of vision, or of exploring toy vision problems in university laboratories, or even just in simulations in computers, without a clue as to the potential practical benefits of their research, or without the benefit of the increased research understanding that would come from testing their theories and models in real-world situations.

A cross-fertilization between academic and industrial researchers would therefore be of great benefit: it could lead to practical applications of benefit to industry, as well as offering real-world challenges that stimulate and encourage a broader and deeper research understanding. The vision machines on the plant floor, and the sensor and display devices needed for vehicles of the 21st century, will greatly benefit from closer interaction between academic and industrial researchers. Appropriate funding from industry, universities, and government should be made available to researchers to encourage such interaction.

7. References

1. R. A. Young, "Oh say can you see? The physiology of vision," *Proceedings of SPIE* **1453**, Feb. 57 1991.
2. C. R. Curcio, K. R. Sloan, R. E. Kalina and A. E. Hendrickson, "Human Photoreceptor Topography," *The Journal of Comparative Neurology* **292** pp. 497-523, 1990.
3. G. Westheimer, "Diffraction theory and visual hyperacuity," *Am. J. Optom.*, **53**, p. 362, 1976.
4. R. A. Young, "Locating industrial parts with subpixel accuracies," *Proceedings of SPIE* **728** Oct. 30-31 1986.
5. R. A. Young, "Getting more for less: A new opponent color technique for two-channel color displays," *Proceedings of SPIE* **1250**, Feb. 15-16 1990.
6. R. A. Young, "Principal-component analysis of macaque lateral geniculate nucleus chromatic data," *J. Opt. Soc. Amer.* **3**, pp. 1735-1742, 1986.
7. C. Reis, "Toward real two-channel color: New approach to an old problem," *Advanced Imaging*, Vol. **6**, No. 2, pp. 52-69, Feb. 1991.
8. E. H. Land, "Color vision and the natural image, Parts I and II," *Proceedings of the National Academy of Sciences*, **45**, pp. 115-129 and pp. 639-644, 1959.
9. R. A. Young and R. M. Lesperance, "A physiological model of motion analysis for machine vision," *Proceedings of SPIE* **1913**, pp. 48-123, Feb. 1-4 1993.

published previously in SPIE, Vol. 2411, page 2