N94-30455

S15-52

2451/

/ 7

DUAL USE OF IMAGE BASED TRACKING TECHNIQUES: LASER EYE SURGERY AND LOW VISION PROSTHESIS

Richard D. Juday, Ph.D.
Manager, Hybrid Vision
Tracking and Communications Division
NASA Johnson Space Center
Houston TX 77058

ABSTRACT

With a concentration on Fourier optics pattern recognition, we have developed several methods of tracking objects in dynamic imagery to automate certain space applications such as orbital rendezvous and spacecraft capture, or planetary landing. We are developing two of these techniques for Earth applications in real-time medical image processing. The first is warping of a video image, developed to evoke shift invariance to scale and rotation in correlation pattern recognition. The technology is being applied to compensation for certain field defects in low vision humans. The second is using the optical joint Fourier transform to track the translation of unmodeled scenes. Developed as an image fixation tool to assist in calculating shape from motion, it is being applied to tracking motions of the eyeball quickly enough to keep a laser photocoagulation spot fixed on the retina, thus avoiding collateral damage.

INTRODUCTION

This is a difficult paper to write, with respect to striking an appropriate balance between being complete on the one hand, and too thinly spreading out information of use to the target audience on the other. I will try to leave much of the technical development to the reference material, presuming a technically competent reader who is not an expert in these disciplines.

NASA's Johnson Space Center (JSC) has been actively developing its image based tracking technology for two medical applications: noncontact retinal position stabilization for laser photocoagulation, and image warping as a prosthesis for certain field defect forms of human low vision. In the retina surgical application, JSC is working with Pinnacle Imaging, a small company dedicated to the development of surgical and robotic methods and equipment for the surgical market. In the image warping application, we are working with the University of Houston's College of Optometry (UHCO) and the University of Pennsylvania. As an Agency, NASA has formal connections with an industrial and medical group that has built a heads-up display system that can directly incorporate our results into prosthetic hardware.

Ophthalmic laser procedures

In the eye surgery application, our central problem is maintaining the aim point of a laser during retinal photocoagulation. Similar image processing hardware may allow tracking assemblyline parts in jumbled orientation, as on a conveyer belt, with no conditions on their placement or orientation. NASA is developing an applicable technology: image correlation based tracking. The eyeball (and the retina with it) undergo continual small-angle jitter, along with the larger jerky saccades that are typified by eye motion while reading. Both are detrimental to the photocoagulation procedure. Incorrect locations on the retina may be struck by the laser; in addition to causing collateral damage to the retina, this limits how close eye surgeons are comfortable working to the fovea (the highest resolution part of the retina -- the part we use when looking directly "at" an object).

JSC has patented[1] a technique that provides fixation; that is, regardless of exactly *what* is in the scene, the input sensor is given the information to allow it to stare at a single location in the scene even as the spacecraft moves or jerks. This is important to simplify analysis of the dynamic scene for information about the three dimensional nature of the unknown surface. "Optic flow" is the technical term for the intra-image motion that results from a shifting perspective of a three-dimensional surface, and computation of optic flow is far eased if fixation is maintained on some point in the image. Our method is based on what is technically described as the optically derived joint Fourier transform correlation. Full details are presented in the patent and in other literature[2,3,4,5].

Field Defect Low Vision Prosthesis

NASA has an Agency presence in image processing technology for low vision sufferers. JSC's Tracking and Communications Division and an associated organization, the University of Houston's College of Optometry (UHCO), originally proposed image warping for field defect amelioration. We have developed hardware that either exists[6,7] or is in advanced design[8,9] that implements our specially designed mathematical image transformations[10,11,12,13], and we have also begun clinical studies whose initial results[14,15] are favorable to the notion that video rate image warping may yield increased visual function for such disorders as central field blindness (maculopathy) or peripheral field loss (retinitis pigmentosa). If known portions of the retina are dysfunctional, an image may be warped so as to minimize the part of the world view that lands there, while retaining a balance between distorting the world view too much and losing part of that world view. We intend that the end result of this program will be visual prostheses (high tech "glasses"[16]) that are more capable, smaller, lighter weight, and less expensive compared with other methods. We intend that the current program will divine mathematical image warpings that maximally increase visual function as tested by reading speed, facial recognition, and so on, but implementing the warpings in inexpensive, lightweight hardware remains an unsolved problem. Among other practical items, we need to determine the range of visual function loss that can benefit from our technique, and then find a minimum set of transforms that span that range.

Fourier Optics

Fourier optics is an important discipline in which the diffractive properties of light become important in the way information is carried and processed. Journals such as Applied Optics, Journal of the Optical Society of America-A, and Optical Engineering carry current articles concentrating heavily in this discipline. Societies such as **SPIE - The International Optical Engineering** Society publish many Proceedings per year in the discipline. I will limit the discussion here to saying that we bring information to a Fourier optics correlator, usually in video image form, and that an element known as a spatial light modulator (SLM) has its light transmitting properties altered by that video image. Then as coherent light passes through the SLM, the information within the video signal is encoded onto the light beam. Diffraction can then be arranged so as to take the Fourier transform of the encoded image, and subsequent transmission (or reflection) of the light through other elements causes the information to be processed and then presented in simpler form (i.e. bright spots).

Optical Correlation Based Pattern Recognition

(Technically incorrectly, this section will mix some concepts from conjugating Fourier optics [the so-called VanderLugt, or 4-*f*, correlator] and joint transform correlation [JTC]. I am going to go ahead and do it, even if the expert would squawk; this is more of a tutorial heuristic device than a technical compendium. Just don't jump me for what looks like an error. There is not room to be rigorous. I am also using one-dimensional notation for a two-dimensional signal, pointing out that to do so has become common in Fourier filtersmithing.)

Suppose we have a function $f(x)$ and we wish to filter information from it with a linear system whose impulse response is $h(x)$. The convolution is $y(x) = f(x)*h(x)$:

$$y_{conv}(x) = \int_{-\infty}^{+\infty} f(x-\tau)h^*(\tau)\,d\tau \qquad (1)$$

but we will not technically distinguish between a convolution, $f*h$, and a correlation, $f \star h$

$$y_{corr}(x) = f \star h = \int_{-\infty}^{+\infty} f(x+\tau) h^*(\tau)\, d\tau \tag{2}$$

since a time reversal of $h(\bullet)$ changes between the two, and $h(\bullet)$ has been arbitrary. As is well known in linear system theory, convolution can be expressed as the inverse Fourier transform of the product of transforms of the functions. That is, if $f(\bullet)$ and $F(\bullet)$ are a transform pair, as are y and Y, and h and H, then

$$\begin{aligned} y(x) &= \mathcal{F}^{-1}\{\mathcal{F}\{f(x)\} \times \mathcal{F}\{h(x)\}\} \\ &= \mathcal{F}^{-1}\{F \times H\} \end{aligned} \tag{3}$$

Also as is well known, there is a shift in $y(\bullet)$ corresponding to a shift in the input signal $f(\bullet)$, a property known as shift invariance. Correlation uses this property to locate a signal that has been identified by the presence of values of $y(\bullet)$ rising above an identification threshold. Optics offers the ability to do the transforms easily by the diffraction of light (once the signal is impressed on it) and to do the multiplication easily (by interacting the light with an SLM on which the function H has been created). There have been three difficult parts. One has been to manufacture SLMs on which desired functions for encoding the signal and the filter. The second is how to make appropriate use of the SLMs we do have; signal theory prescribes ideal values for the filter values, H, that are ordinarily not realizable. Recently[17] we have achieved a breakthrough in optical filter theory that make some of the job easier, and we continue developing modulators and their characterization[18,19] for other aspects of the problem. The third is how to accommodate the rotation of the retinal image that occurs as the eyeball rotates within the socket, as in response to signals from the vestibular function. Pattern recognition by image correlation is usually quite sensitive to rotation, so we continue our effort to achieve tailored amounts of rotation invariance, and the results will be fed into our activities under Memorandum of Understanding with our medical partner, Pinnacle Imaging. This effort is ongoing under university grant (see Formal Colleagues, below) following concepts laid down several years ago[20].

Somewhat simplistically, a Fourier optics image processor converts the video image of the object it has been trained to find into a bright spot of light, and then infers information from that spot. The difficult parts of the information extraction are done off-line and at leisure; the easy part is what happens on-line and in real time. As distinct from most digital image processing, this system determines the correlation between highly synthesized reference objects and the viewed object, and then extracts the information from those correlations. Synthesizing a reference object is a congruent process to obtaining the inverse transform of the optimal filter that can be realized within the limitations of the filtering spatial light modulator. The correlations run at a very high frame rate (tens to thousands per second), allowing comparisons of the input video object with all members of a library of views of that object. Careful crafting of the reference image set has the effect that the library element producing the highest correlation corresponds to the actual position and attitude of the viewed object.

In further distinction from most digital processing, optical pattern recognition operates on the whole input image, rather than individual features, and thus is robust against partial obscuration. The difficult technical challenges are principally in the physical devices -- spatial light modulators -- that encode the incoming video image and the correlation image.

VIDEO RATE IMAGE WARPING

Shift Invariance in Optical Pattern Recognition

As is well known in signal theory, the convolution (equivalently the correlation) of two signals is a linear and shift invariant operation. Also well known is that scale and rotation changes in a Cartesian geometry do not produce shift invariance. If we know $[x_0, y_0]$, the center about which scale and rotation changes occur, though, we can convert scale and rotation into translational shifts by performing a coordinate transformation. The log polar transformation is what we seek.

$$\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix} \tag{4}$$

shows the conversion between the Cartesian $[x,y]$ and the polar $[r,\theta]$ as coordinate systems. Go one step further and define u as $\log r$. Then as the original image coordinates undergo a change expressed as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = k \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{5}$$

we will find that the coordinate pair $[u,\theta]$ undergoes the change

$$\begin{bmatrix} u' \\ \theta' \end{bmatrix} = \begin{bmatrix} u + \log k \\ \theta + \alpha \end{bmatrix}, \tag{6}$$

so that the rotation through α and scaling by k become translations by α and $\log k$. For video rate pattern recognition that is to be shift invariant to scale change and rotation about $[x_0,y_0]$, and the difficult part is to create the log polar transformation at full video rate.

The Programmable Remapper

With the Programmable Remapper we have solved that problem, at least on a laboratory scale. Its generality allows us to create virtually unlimited geometric transformations at video rate. Conversion to practicality as a prosthesis remains an unsolved, but significant and potentially productive, interest. Read the References for technical details, but suffice it for the present purposes to say that the Remapper can wreak an arbitrary static geometric transformation (at video rate and incurring only a three frame delay) on an incoming video image. We conceived the Remapper as a tool for creating a log polar video image, but we rapidly understood its potential for manifesting other warpings. Human low vision field defects was the first such extracurricular application.

Field Defect Application for Low Vision Humans

As reported in a number of the papers in the References, we have investigated the applicability of image warping for various field defects. For examples of images, see in particular Ref[12]. One of the two principal ones is maculopathy, in which a lesion causes loss of function at the high resolution central portion of the normal field of view. The other principal application is RP (retinitis pigmentosa), also known as tunnel vision, where peripheral vision is progressively lost. Our intent is to move the active portions of a video image so that it falls onto portions of the retina that are still functional, thus having a more nearly complete mapping of the world into the visual cortex (albeit at the expense of local distortion in the normal world-to-cortex representation).

The fovea is the portion of the eye where the resolution is the highest; it is the part of the eye one naturally places into conjunction with the part of the world scene where visual detail is sought. It is the part of your field of view you use when looking "at" something. Loss of foveal function is particularly expensive to detailed functions like reading and facial recognition. Our method is to "rubber-sheet" a video image. Imagine that a television screen is made of rubber, that you poke a hole in the middle, and then stretch that hole out until it is just larger than the blind spot in the center of the field of view. There are many loose parameters in this description, but the core idea is there. In our initial tests, we used normally-sighted volunteers, and we simulated the central blind spot by forcing them to look at an obscuration on a television screen with text scrolling by. An eye tracker followed the gaze of the eye, controlling mirrors that directed the gaze onto the obscuration. The words scrolled by at various font sizes, angular rates, etc. There was significant (though highly variable from subject to subject) increase in reading rate with image warping. Our next studies will use actual low vision volunteers, it will not force foveation, and we have some advanced mathematical transformations to try. These studies are beginning in Fall 1993.

To deal with the converse problem, we wish to squeeze the normal world field of view into a smaller solid angle. This is easily accomplished, of course, with a minifying telescope, and such is one weapon in the ordinary armamentarium of the RP patient. The problem is that the angular subtense of all elements in the field of view is reduced. Even though you see more of the things one needs to interact with in moving about (door frames, walls, stairs, etc.), the reduced size of objects may give difficulty in reading signs or otherwise resolving the world. Our approach is to give spatially varying magnification, so that localization of gross objects is possible, but full central magnification is maintained. Perhaps in Spring 1994 we will begin field tests with the Remapper in locations like grocery stores, to see if visual functionality is improved.

JOINT TRANSFORM TRACKING

When NASA was recently contemplating autonomous lunar and planetary landings, this project proposed two techniques for lander vision, and we set about demonstrating certain of the technical elements. The first technique was navigation as guided by image correlation, in which images of landmark features would provide reference to absolute geometric coordinates. The second was image fixation by joint Fourier transform correlation, in which unmodeled image structure correlates with itself and shows sequential error in pointing. It is the JTC, as applied to tracking retinal motions, where we have been vigorously spinning off to medical practice.

Dynamics of Eyeball Tracking

Tracking the human eyeball is very challenging. Consider the problem from the other side for a moment; think how quickly you can jerk your eyes over a really large angular distance and settle your viewpoint in on a small object that has caught your attention. Now consider keeping a laser beam fixated on an individual structure within the retina during that motion, to a precision at the same resolution of the smallest object you can resolve. The harmonic content of the eyeball motion extends to as high as 170 Hz, with speeds up to several hundred degrees per second. Those are really astounding figures. Nyquist sampling theory indicates immediately that to track the motion, about 400 independent measurements per second are necessary. If the eyeball motion is able to break lock (e.g. if the eye jerks during a blink), then it is important to know that track lock has broken and in which direction to move to reacquire, but the precision requirements on knowing the translation are reduced since the surgical laser will be inhibited immediately upon loss of track. The optical engineers in the project (the author cheerfully admits to this persuasion) think that the JTC is the preferable method because of the ease with which two qualities are achieved: spatially variable resolution, and a wide field of regard. If tracking is done by digital methods by time-domain correlation, the size of the field of regard is directly impacted by the serial nature of the digital correlation. In contrast, optical processing is inherently parallel, and the size of the field of regard is limited only by the number of pixels in the spatial light modulators. The project currently is headed for a showdown between digital and optical implementation, as the digital method (if practical) would be slightly less expensive to implement.

THE JOHNSON SPACE CENTER PROGRAM

The Hybrid Vision project at JSC has several civil servants (Richard Juday, Tim Fisher, Shane Barton, and Jennifer Yi), NRC post-doctoral fellow Colin Soutar, and in-house contractor personnel employed by Lockheed Engineering and Sciences Corporation (Stanley Monroe, Carlton Faller). Additionally we have either active or recently concluded contracts, fellowships, and grants at a number of universities and corporations (University of Missouri, Carnegie Mellon, Tennessee Technological University, University of Colorado at Boulder, University of Houston, Physical Optics Corporation, Boulder Nonlinear Systems, Physics Innovations) and working relationships with other Government laboratories (Army Missile Command, Air Force Rome Laboratory, ARPA, Air Force Wright Laboratory, Sandia), and others. The project bibliography for the past seven or eight years has over fifty entries including four issued patents (others in work). We are active in image warping, optimal filter theory, optical correlator architectures and applications, spatial light modulator development, with specific thrusts in robotic vision and human low vision. The author will provide copies of the bibliography upon request.

ACKNOWLEDGEMENTS

Financial, moral, and technical support from our military colleagues, mentioned just previously, is particularly appreciated.

REFERENCES

1. U. S. Patent 5,029,220; Richard D. Juday; **Optical joint correlator for real-time image tracking and retinal surgery**; Jul. 2, 1991.

2. Jerome Knopp and Richard D. Juday, "Optical joint transformation correlation on the DMD", Proc. SPIE **1053**, 208-215 (1989).

3. Eddy C. Tam, Francis T. S. Yu, Don A. Gregory, and Richard D. Juday, "Autonomous real-time objects tracking with an adaptive joint transform correlator", Optical Engineering **29**, 314-320 (April 1990).

4. Eddy C. Tam, Francis T. S. Yu, Aris Tanone, Don A. Gregory, and Richard D. Juday, "Data association multiple target tracking using a phase-mostly liquid crystal television", Optical Engineering **29**, 1114-1121 (September 1990).

5. K. L. Schehrer, M. G. Roe, and R. A. Dobson, "Rapid tracking of a human retina using a nonlinear joint transform correlator", SPIE Proceedings vol. **1959**, April 1993 (Orlando).

6. U.S. Patent 5,067,019; Richard D. Juday and Jeffrey B. Sampsell; **Programmable remapper for image processing**; Nov. 19, 1991.

7. Timothy E. Fisher and Richard D. Juday, "A programmable video image remapper", Proc. SPIE **938**, 122-128 (1988).

8. U.S. Patent 5,208,872; Timothy E. Fisher; **Programmable Remapper with single flow architecture**; May 4, 1993.

9. Timothy E. Fisher and Richard D. Juday, "An improved architecture for video rate image transformations", Proc. SPIE **1098**, 224-231 (1989).

10. Richard D. Juday and David S. Loshin, "Some examples of image warping for low vision prosthesis", Proc. SPIE **938** 163-168 (1988).

11. Richard D. Juday and David S. Loshin, "Quasiconformal remapping for compensation of human visual field defects: advances in image remapping for human field defects", Proc. SPIE **1053**, 124-130 (1989).

12. David S. Loshin and Richard D. Juday, "The programmable remapper: Clinical applications for patients with field defects", Optometry and Vision Science **66**, 389-395 (1989).

13. Richard D. Juday, Alan T. Smith, and David S. Loshin, "Human low vision image warping: channel matching considerations", Proc. SPIE **1705** (1992).

14. David S. Loshin, Janice Wensveen, Richard D. Juday, and R. Shane Barton, "Design of reading tests for low vision image warping", Proc. SPIE **1961**, Orlando, April 1993.

15. Janice Marie Wensveen, "Reading Rate with Simulated Central Scotoma", Master's thesis, University of Houston College of Optometry, August 1993.

16. Anonymous, "High-Tech Help for Low Vision", NASA Tech Briefs vol. **17** no. 2, pp. 20-22 (April 1993).

17. Richard D. Juday, "Optimal realizable filters and the minimum Euclidean distance principle", Applied Optics **32**, 5100-5111 (10 September 1993).

18. Colin Soutar, Stanley E. Monroe, Jr., and Jerome Knopp, "Complex characterization of the Epson liquid crystal television", Proc. SPIE **1959**, Orlando, April 1993.

19. Colin Soutar, Stanley E. Monroe, Jr., and Jerome Knopp, "Measurement of the complex transmittance of the Epson liquid crystal television", Optical Engineering (accepted).

20. Richard D. Juday and Brian Bourgeois, "Convolution-controlled rotation and scale invariance in optical correlation", Proc. SPIE **938**, 198-205 (1988).