Reprinted from SPIE Vol. 360—Robotics and Industrial Inspection

Image-based focusing

Ted Selker
Computer Science Department, Stanford University
Stanford, California 94305

Abstract

Lens focusing using a hardware model of a retina (Reticon RL256 light sensitive array) with a low cost processor (8085 with 512 bytes of ROM and 512 bytes of RAM) was built.

This system was developed and tested on a variety of visual stimuli to demonstrate that: a)an algorithm which moves a lens to maximize the sum of the difference of light level on adjacent light sensors will converge to best focus in all but contrived situations. This is a simpler algorithm than any previously suggested; b) it is feasible to use unmodified video sensor arrays with inexpensive processors to aid video camera use. In the future, software could be developed to extend the processor's usefulness, possibly to track an actor by panning and zooming to give a camera operator increased ease of framing; c) lateral inhibition is an adequate basis for determining best focus. This supports a simple anatomically motivated model of how our brain focuses our eyes.

Introduction

Given an object field, a lens and an image field (figure 1.), ideal focus exists if the light level distribution on the image plane (sensor) varies spatially only by coefficents of enlargement. In this condition with a magnification of 1, object plane points map uniquely onto image plane points. Defocus is characterized by an object plane point mapping onto distribution area in the image plane. The area to which object plane points map on the defocused image is called blur-circle. The larger the blur-circle, the worse the focus. This blur attenuates high spatial frequency components.

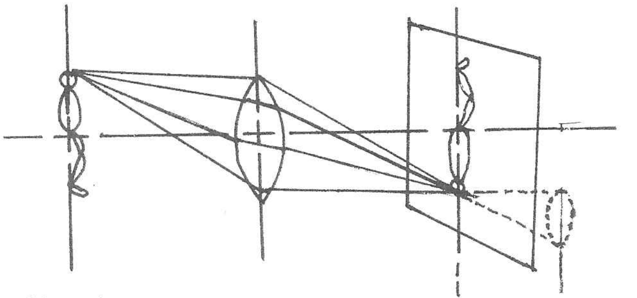


figure 1

To understand defocus better let us look at the limiting image focus problems.

Case 0, no fluctations in light level: If an object plane is all one luminescence, it is not possible to determine ideal focus. The redisturibution of object plane points onto image plane blur-circles will only change the source of image light on the image plane, not the image light level. Light from adjacent object plane points will overlap in the image plane. The defocus will not introduce spatial fluctuations in light level.

Case 1, one fluctuation in light level: The next simplest object plane is that of one light level change. This can be most easily considered in a one-dimensional image plane. Consider this, the dark to light step function (figure 2a.). Such an image has the attribute of having only one light level change in ideal focus. The value of the step change is equal to the difference in light level accross the step. If the the system is defocused, blur-circles of all points from the light side within blur-circle radius of the step will map their light to the dark side of the step. The step will change into a gradient of light level change. Light level will only be altered within the blur circle radius of the step light level change. In addition, if the blur-circles from the light part of the step maps off the dark end of the image plane (or the dark part, off of the light end of the image plane) then the total light level change along the line will be decreased.

Case 2, two fluctiations in light level: With two adjacent, opposing light level changes along the image plane (figure 2b.) we get the bar; the simplest example of interacting edges. Here, if the blur-circle caused by defocus is larger than 1/2 the width of the light bar, then the maximum light level which the bar had in the in focus condition will not be attained. In this case all points on the light section spill light onto the dark sections.

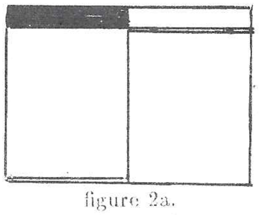
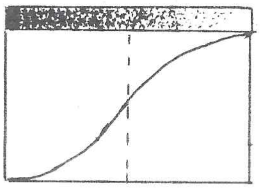


figure 2a.



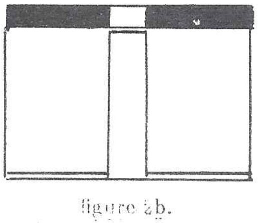
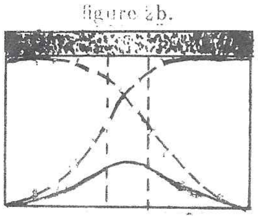


figure 2b.



This attenuation of image intensity difference across the image occurs whenever one of two things happens:

1. The attenuation occurs when two opposing light level changes in the object plane map to points on the image plane which are closer together than the blur-circle radius of the focus state of the system.

2. The attenuation will also occur when the gradient caused by the blur of an edge maps partially off of the image plane. This happens when light level changes and edges of the image plane are closer together than the radius of the blur-circle radius of the focus state of the system.

Ocular focus for humans and animals is, in most cases, an unconscious effort. The physiology of accommodation (figure 3.) allows us, in most cases, to automatically focus our eyes. As well, we have alternative circuitry which allows us to override this auto focus. Although our visual system has corroborating information for focus such as object relative size, vergence of our eyes and chromatic aberation of our lens, our eyes, for the most part our visual system uses the clarity of the image that the retina produces as the basis for focusing. Such an image based focusing system is termed passive, systems which focus based on other evidence such as sonar are called active focusing systems.

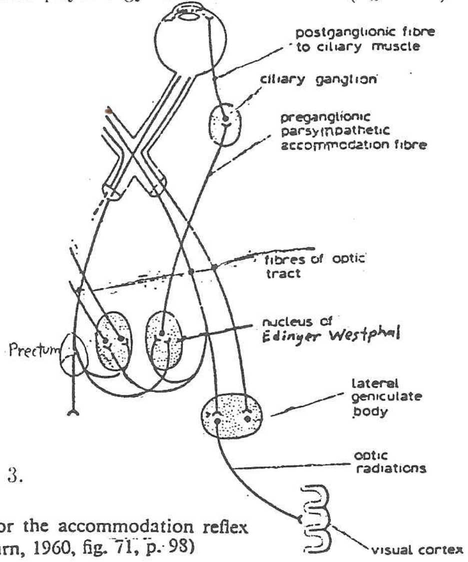


figure 3.

Pathways for the accommodation reflex
(Modified after Wyburn, 1960, fig. 71, p. 98)

Commercially Available Products

Although passive focus systems have been shown to be possible for some time[2 4 5] the first auto focus systems on commercially available cameras were the active Polaroid sonar distance measuring focuser and the trianglating Honeywell Visitronic range finder 1978. Both of these systems work and have sold well. However, neither of these systems is self calibrating; neither relies on the actual image projected to the image plane of the cameras they focus.

Leitz first showed a passive focus system in 1976; Correfot [1]. This system uses an optical grating mounted on a vibrating tuning fork to sample focus with two sets of sensors. These sensors are focused on the exit pupil of the lens. At this point in the optical system, the light intensity will be diffused when the image plane is in focus. When a comparison of the two sets of sensors trained towards different parts of the exit pupil has minimum local light level difference, the system is in focus. The Honeywell TCL system first disclosed in December 1978 [1] uses the same type of information, replacing the tuning fork image sampler with a spatially distributed sensor array. These systems use an interesting property of the spatial transformations which occur in a lens system. They do not, however, use the image plane data itself. For this reason systems like these will never be implementable using the camera's image plane sensors without modification.

The most modern passive focus entry incorporated into the Pentax ME-F camera this year is the patented Pentax EFC system [1]. This system uses two linear light sensing arrays; one placed slightly in front of the image plane, one slightly behind. If these arrays are at a calibrated distance when the camera is in focus a correllation will show them to have equal spatial light level distributions. This system uses the light level distribution seen on the image plane as focus data. This system does, however, share the drawbacks pointed out above. It requires a separate sensor system, specialized optics for the focus sensor, and special calibration for the focus sensor.

Focus Research

Two treatments of image based focusing [2 5] have shown that the separate sensor problems with which commercially available systems now contend are not essential. In 1968, Horn described a system which he had implimented[2]. He used a computationally expensive, straight forward focus algorithm. This algorithm moves a lens to maximize high frequency terms of a Fourier transform of the image. This will maximize the high frequency information availible in the image. In 1970, Tenenbaum described and implemented a much simpler algorithm[5]. He maximizes high frequency content in an image by maximizing the sum of thresholded light level differences across an image. Such an algorithm and similar methods such as the sum of the squares of light level difference are exciting for their computational simplicity and low memory overhead.

The System

Our focuser[4] demonstrates how an even simpler algorithm than Tenenbaum's is adequate for focusing. The thresholding which was believed crucial we show is not.

We will call the sum of the absolute value of intensity differences on an image the Luminescence Difference Sum (LDS). This LDS is attenuated by defocus by the two methods described in the introduction. When the period of local light or dark intensity extrema is smaller than the blur circle radius of the focus system, these intensity extrema will be attenuated. The sums of the differences in light level across the image will reflect this.

The LDS will decrease if any high frequency extremum in an image is lost. There is one class of images which will not show decreased LDS in defocus. If the image intensity is constantly decreasing across the image, high frequency fluctuations will only smooth out the gradient in defocus; they will not eliminate any extrema. Thus the LDS will not decrease. It is, however, hard to imagine such a dark to light gradient object plane which one would want to record.

For all object planes for which at least part is not a monotonically darkening or lightening gradient, the LDS will decrease if high frequency information is lost. Moving the lens to see if the LDS is decreasing or increasing, then moving it in the direction indicated to maximize the LDS will find the focus position with the high frequency components spatially resolvable. This inclusion of all high frequency light level changes in an image is the metric of best attainable focus.

Experiments

An 8085 computer was interfaced to a Reticon RL256G linear array and a focusing lens Using this system on a range of data (see samples in figure 4) focus algroithms were tested. An algorithm which took the difference of the square of light level divided by 255 was tested as well as the LDS. This function was chosen to test the sum of the difference of squares of light level algorithm in an 8 bit machine. Such an algorithm should cue on high frequency content, and this removes the monotonicity problem described above.

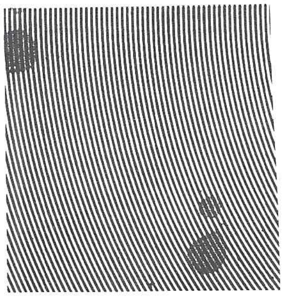


figure 4a

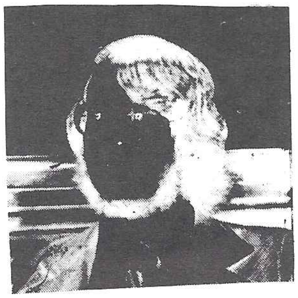


figure 4b

Results

The LDS focusing algorithm was capable of focusing on the wide range of data tested. The more complicated the object was, the further the system could be out of range and still pick up edges on which to focus. The Difference of pseudo squares did not compete with the simpler LDS algorithm favorably. This is due to the fact that the "squares" lookup table loses significant bits for low light situations. The LDS focusing algorithm was able to track focus when the lens was moved away from the object plane at greater than 1/2 the maximum motor speed on any of the data.

The LDS algorithm's step response varied with the object it analysed. Figure 4a for example will become a flat field in the image plane when the width of the lens's blur circles is 1/2 the spatial frequency of the grating. Such an image is easy to track but has bad step response. The best step response was attained for object planes such as figure 5b with a mix of high and low spatial frequencies. Such an object will have a more even gradient in its focus function over focus change. With figure 5b using a 124 mm lens the system will focus in the correct direction even if the lens is focused at 30 cm and the object distance is 60 cm[4].

Discussion

The results of this autofocus project are extremely encouraging. The entire cost of the hardware including the computer development system was well under $500, in a camera with a digital sensor, the cost would essentially be the cost of the processor.

With a slow (2 mhz) 8 bit processor, 512 bytes of Ram and 512 bytes of ROM software was be written which could run a focusing algorithm in real time. This could easily be the basis for a control system for video cameras.

The system was able to focus on a wide variety of real and contrived images. Optical stability problems and motor speed were major impediments to performance. When the lens used changed directions it moved the 256 pixel image 25 pixels laterally by rotating the lens perpendicularly to its optical axis. This added a large noise function which limited the focus speed. Hysteresis in changing motor directions required software timing so that the system could know the direction the system was focusing in.

As a final speculative note we would like to admit that the ideas for this autofocus project were motivated by the anatomy of higher animal's optical systems. We would argue that a metric similar to the LDS is in play in the visual system. Three layers of neurons reside inside the eye itself, these cells impart a spatial function to the Retinal Ganglion cells. Most of these cells respond to bright spots in what is called an on center off surround manner. This means that the cell will respond less to a spot which it is wired to sense if there is light shining on the area surrounding this spot on the retina. A convolution of such cells with competing fields gives a function which is proportional to some difference of light level across the area the cells represent. This would be our LDS were the character of the lateral inhibition linear. Of course since the LDS is the simplest such algorithm, and since it works, a higher order "difference" would only pick out high frequency information better.

Most synapses (neuron connections) are like a convolution; somewhere between 100 and 10000 input neurons connect to one neuron which, when enough of the input neurons fire will turn on. The Retinal Ganglion cells synapse once in the Prectum on course to the Nucleus of Edinger Westphal. There a synapse is made to the focus driving neuron; the Ciliary ganglion. It is then proposed that it is possible that the accommodation of our visual system to focus could be driven for the most part by a very straightforward connectionistic algorithm similar to the LDS described in this paper.

References

1. Goldberg, N., Inside Autofocus: How the Magic Works, *Popular Photography*. Feb 1982.
2. Horn, B., *Focusing*, MIT Artificial Intelligence Memo. No. 160, 1968.
3. Polaroid Focuses With Sound Waves, *Electronics*, pp. 41-42, May 1978.
4. Selker, T., *An Image Based Focusing System: An Alternative for Cameras, a Model of Eye accommodation*, COINS Techincal Report, pp. 81-19.
5. Tenenbaum, J. M., *Accommodation in Computer Vision*, Stanford Artificial Intelligence Report No. Cs 182, 1970.