Proceedings of the 2001 IEEE
International Conference on Robotics & Automation
Seoul, Korea • May 21-26, 2001

PRACTICAL ISSUES IN PIXEL-BASED AUTOFOCUSING FOR MACHINE VISION

Ng Kuang Chern,Nathaniel Poo Aun Neow* Marcelo H. Ang Jr.*

Center for Intelligent Products and Manufacturing Systems
Department of Mechanical Engineering
National University of Singapore
*{mpepooan,mpeangh}@nus.edu.sg

Abstract

Different autofocusing methods exist for many cameras today. While not ignoring commercially available methods requiring specialized hardware, this paper focuses mainly on pixel based autofocusing algorithms as applied to CCD camera systems. Different measures of image sharpness are compared. For each of these, different algorithms for searching the best lens setting are assessed in terms of performance as well as their applicability to various situations. In addition, several other factors potentially affecting camera focusing are also discussed. Based on the information obtained, this research attempts to formulate a robust autofocusing algorithm.

Keywords

Autofocusing, focusing, machine vision, image processing.

1. Introduction

Focusing is an important aspect in many applications involving machine vision. The degree of focus in an image is a factor in determining image quality. For example, a focused image might contain some details not present in an unfocused image of the same scene.

Not every application requires automatic focusing. Fixed focus settings may be used when the depth of field is large or the camera to object distance is known. Often, a camera's built-in autofocusing system is adequate for the task involved. However, certain applications may require a greater degree of control for focusing to obtain a sharp image. It may be necessary to have a focusing window that may be selected or changed dynamically. A good focusing system is thus required to ensure the reliability of the images obtained.

Most autofocusing systems involve hardware built into the camera, which may often be difficult to customize. Nevertheless, more flexible techniques exist, allowing a scene to be focused based solely on the information obtained from a camera's CCD array. Only the (x, y) position and the red, green and blue intensities of each pixel is known. In this paper, this type of focusing shall be referred to as "pixel-based focusing". Despite the disadvantages of this technique to traditional focusing methods, such as speed and cost, occasions arise where the added flexibility and degree of control of pixel-based focusing make it the preferred focusing method.

"Pixel-based" focusing involves several parameters. The first consists of the focusing window, or the region of the scene that is to be focused. Next, a quantity indicative of the image sharpness, or a "sharpness function", is required. Following this, a searching algorithm to find the global maximum of the "sharpness function" must be chosen. In addition, several other factors may need to be considered—fluctuations in scene illumination, the depth of field, as well as lens aberrations. In the case of color CCD cameras, the option arises of choosing the red, green, blue channel of an image for focusing, as well as using either grayscale or all three channels.

This research covers the practical issues in implementing a pixel-based autofocusing algorithm. Each of the parameters used for focusing, such as the sharpness function, the searching algorithm, or the focusing window, may be customized for a particular application. Making appropriate decisions regarding each of these parameters requires some background and experience. In this paper, results are presented to enable one to make a more informed choice when designing a pixel-based autofocusing algorithm.

This paper is organized as follows. In Section 2, we present different sharpness functions reported in literature as well as new ones we introduced. Section 3 provides a discussion on performance and practical issues in implementing the sharpness functions. Various search algorithms including new ones are presented in Section 4. In Section 5, we summarize our work and provide recommendations.

2. Sharpness Functions

The sharpness function computes the sharpness or degree of focus on an image or a region (area) of an image. At different lens positions, the sharpness of an image changes. Autofocusing means automatically moving the lens position such that the sharpness is maximized, i.e, image is in best focus. In literature, there are eight different sharpness functions. The first two, the "amplitude" method and the "variance" method, are quite similar. They are:

0-7803-6475-9/01/$10.00© 2001 IEEE

grey level "amplitude", $\frac{1}{n}\sum\left|I(x,y)-\bar{I}\right|$, and

grey level "variance", $\frac{1}{n}\sum\left(I(x,y)-\bar{I}\right)^2$

The next two, the "Tenengrad" and the "Laplacian" are based on standard edge detection masks. Each of them are used with a threshold setting of zero, as suggested by Krotkov [7]. In the "Tenengrad" method [9, 10], the Sobel horizontal and vertical operators

$$i_x=\frac{1}{4}\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix}, \text{ and } \quad i_y=\frac{1}{4}\begin{bmatrix}1 & 2 & 1\\ 0 & 0 & 0\\ -1 & -2 & -1\end{bmatrix},$$

are used to find the strength of the horizontal and vertical gradients. The image sharpness is then defined as

$$\frac{1}{n}\sum\left(S, \text{for } S>T\right),$$

where $S=i_x{}^2+i_y{}^2$ and T is the threshold. The "Laplacian" [7] is almost identical, except that the operator

$$L=\frac{1}{6}\begin{bmatrix}1 & 4 & 1\\ 4 & -20 & 4\\ 1 & 4 & 1\end{bmatrix}$$

is used and the image sharpness is

$$\frac{1}{n}\sum\left(L, \text{for } |L|>T\right).$$

The "Fast Fourier Transform" (FFT) method [3], as its name implies, evaluates image sharpness from the Fast Fourier Transform of the image. In this implementation, the image's grey levels are placed row by row into a 1D array [2], and its FFT evaluated. Since the FFT is evaluated using the successive doubling algorithm [8], the array is first zero-padded before performing the FFT. From the real and imaginary parts of each element of the new array, the quantity is evaluated as the sharpness function.

$$\tfrac{1}{n}\sum\left|\left(real^2+imaginary^2\right)\times\tan^{-1}\left(imaginary/real\right)\right|$$

The "Sum-Modulus-Difference" method [4], or SMD, simply sums up the differences between adjacent pixels and is defined as $\sum\left|I(x,y)-I(x,y-1)\right| + \sum\left|I(x,y)-I(x+1,y)\right|$.

Two histogram methods, the "histogram entropy" [4, 9] and the "histogram of local variations" [7], work by evaluating a quantity from the image's intensity histogram. If the intensity histogram is, $h(i)$, where $h(i)$ is the frequency of pixels of intensity i, then the histogram entropy is defined as

$E = -\Sigma\ \{h(i)\ \ln\ (h(i)), \text{ for } h(i) \neq 0\}$.

In the "histogram of local variations", the intensity histogram is evaluated with pixel intensities compressed logarithmically and the gradient of the line of best fit through the points, m, is evaluated. The quantity, m, is at a minimum for the sharpest image. Since there are 256 gray levels, $i = 0$ to 255, $\Sigma\{\ln(i+1)\}$ and $\Sigma\ \{\ln(i+1)\}^2$ is known and m may be evaluated as

$$\frac{256\times\sum\log(i+1)\times h(i)-1167.26\times\sum h(i)}{60354.1}.$$

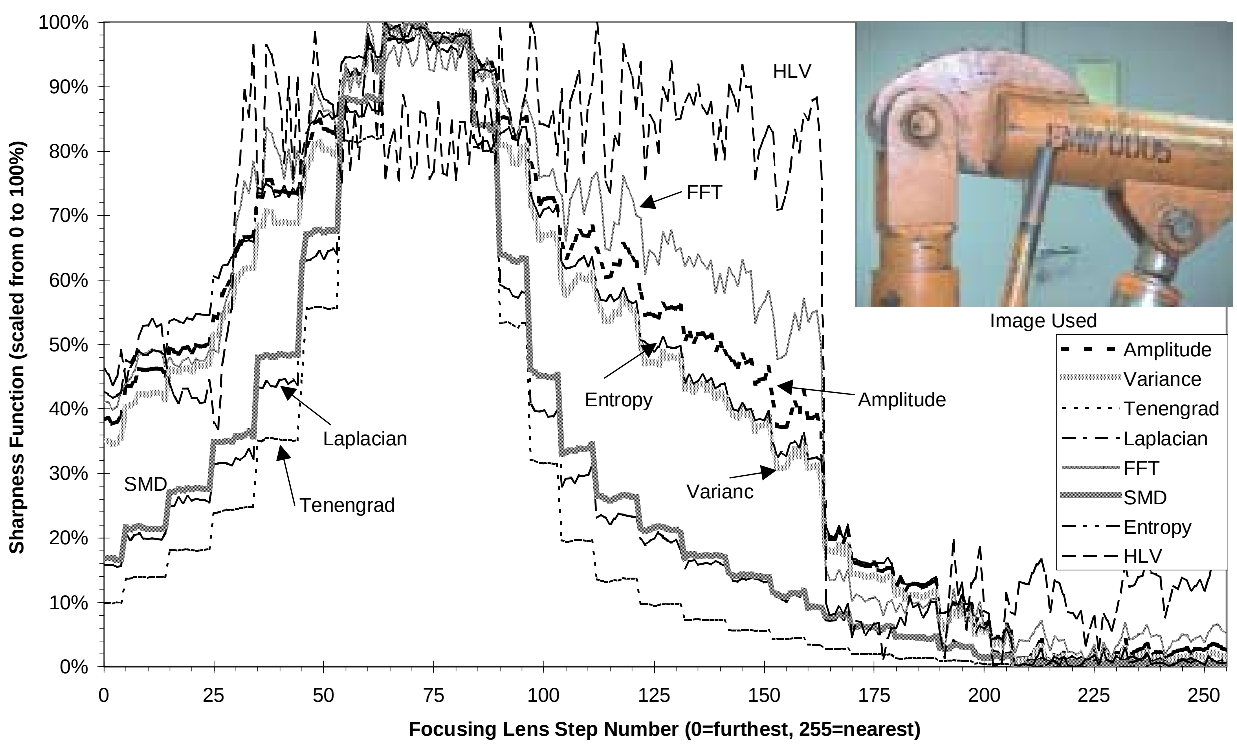

Figure 1: Plot of Image Sharpness (as determined by different sharpness functions) vs. Lens Position

Varying the position of the focusing lens will change the image, and hence the image sharpness. Thus, a graph of the magnitude of the image sharpness against the camera's lens position may be obtained. The JAI camera was used, where 256 different lens positions could be set. For the purpose of repeatability, the image at each of the 256 lens positions was saved to the hard disk before processing. To fit the eight different sharpness functions to the same range on the vertical axis as well as for comparison purposes with results by Krotkov [7], each was first scaled to the range 0 to 100% before plotting them.

The graphs are then analyzed to determine whether the global maximum of each sharpness function corresponds to the focused image, as well as whether the global maximum for each sharpness function agrees with the rest. Observation is made as to whether or not these graphs change, if a different color channel is used with the sharpness function. Figure 1 shows the image sharpness plots for different sharpness functions using the red channel of the frame grabber.

The shape of the sharpness function graph is an important indicator to determine the ease of which the global maximum may be found as well as to check the accuracy of the position of best focus. The shapes of the Tenengrad and Laplacian and SMD are particularly good for this image. The variance method also works acceptably well, its main problem being its low signal to noise ratio. The image of best focus was determined as frame #72 (lens position 72) by the amplitude and variance method. The Tenengrad determined frame #73 as the sharpest image while for the Laplacian and sum-modulus-difference functions, the frame was #69. The histogram entropy method obtained frame #64 while the histogram of local variations incorrectly obtained frame #145. The depth of field as determined by the size of the flat region of each of the sharpness functions was from frame #64 to frame #83. Visual inspection reveals virtually no difference between the frames in this range.

All the sharpness functions, except the HLV were able to find an image near the point of best focus without too much difficulty. The results differ from the results obtained by Krotkov, whereby the Laplacian method failed on a high contrast cross as well as on text. In addition, the entropy method could not accurately find a focus on text. From these results, as well as those by Krotkov, the best functions appear to be the variance and the Tenengrad, followed by the SMD.

For this reason, only the first two functions, the variance and the Tenengrad were chosen for obtaining "depth maps" later.

The same procedure was repeated to obtain plots for the green and blue channels. No significant difference in pattern was observed between the different plots for the three sharpness functions tested, the "variance", "Tenengrad" and "histogram entropy" methods. Figure 2 shows the plot for the Tenengrad function using the 3 different color channels. In the tests, the global maximum did not differ much between each colour channel.

3.0 Performance Issues

The speed of focusing depends partly on the speed of the camera's focusing motor and partly on the speed of the focusing algorithm. In terms of speed or performance, greyscale focusing may be done at higher frame rates than colour focusing, including colour focusing based on only one channel, because the hardware grabbing for a colour image takes longer than a greyscale grab.

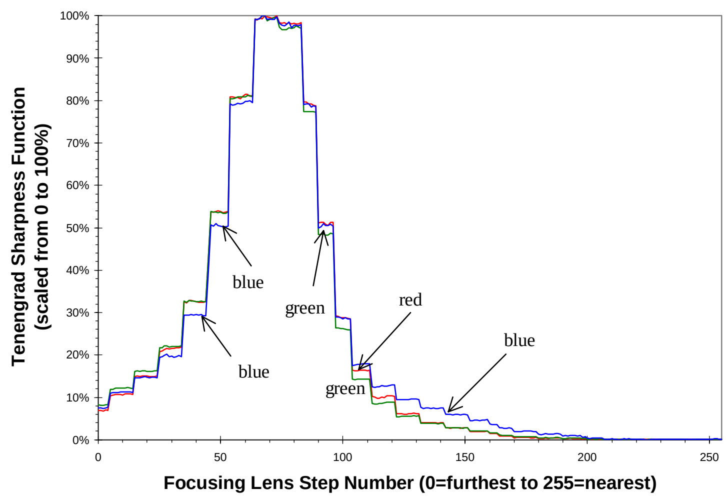

Figure 2: Difference in Using Different Color Channels for Focusing for Tenengrad Sharpness Function

If the greyscale channel is used for focusing, the sharpness function may not be able to detect an edge between, say a red patch and a green patch of the same intensity. Similarly, if only the green channel is used, as suggested above, an image consisting predominantly of red, blue or magenta (a mixture of red and blue), but little green may have insufficient contrast for focusing.

Ideally, all three channels of red, green and blue should be used and combined in some way (for example, using a weighted average). However, the tripled increase in processing time is a considerable price to pay for this extra accuracy. Nevertheless, there may sometimes be applications whereby accuracy is much more important than processing time and this method may find its use there. In most other situations, greyscale focusing should be adequate.

Of the different sharpness functions, the fastest methods are the variance methods and the histogram methods, averaging about 20ms to process a 768 x 576 focusing window. On the slower end are the Fast Fourier Transform, the SMD and the Tenengrad, clocking 2500, 65 and 68 milliseconds respectively. The timings are for a Pentium II-300 machine.

Although the Fast Fourier Transform method provides a good way of obtaining the degree of defocus, it is generally too slow for use in focusing, even with the successive doubling algorithm. A 1-D transform that only includes a row or column of pixels may accelerate this, but this will only allow focusing on a 1D region of the scene, rather than a two-dimensional region. In addition, it is useful to have a larger focusing window as it is easier to ensure that the object of interest falls entirely within the window. Large window sizes allow focusing under small movement of the object in question as well as camera vibration. The main problem with the Fast Fourier Transform technique is its slow speed of about 2.5 seconds or 0.4 Hz using a 768 x 576 focusing window.

The image grabbed in any single frame will differ slightly from a subsequent grab due to noise and small changes in scene illumination. To test the degree to which this affects each sharpness function, another 256 images of the same scene were grabbed consecutively and saved to the hard disk. Next, the sharpness of each image was determined using each sharpness function. The results show that for the same scene taken at different times, there is a significant variation in the sharpness function. Figure 3 shows the variation of the Tenengrad sharpness function for the same scene taken at different iterations. It has not been tested whether the amount of noise in each function is independent of the scene captured or the scene illumination.

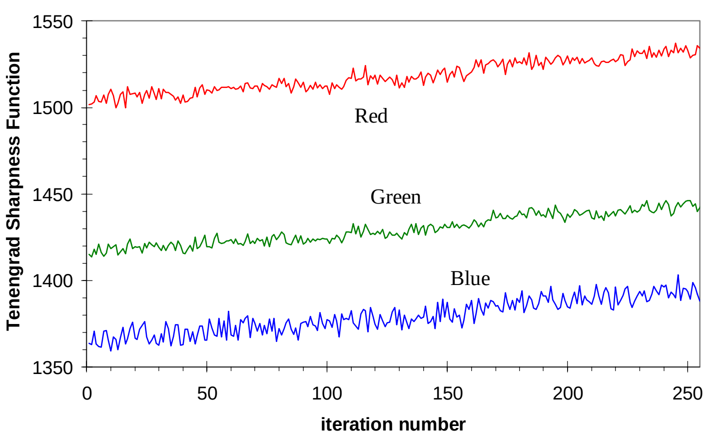

Figure 3: Variation of Tenengrad Function with Time.

The values of the sharpness functions are not constant with time, even without moving the camera's lens. This is because the pixels in the image fluctuate with time. Observation of a single pixel near the center of the image revealed a variation in intensity of about ± 7 to ± 8 for each color channel. To test whether this effect was entirely due to scene illumination, a lens cap was used to cover the camera lens. However, this did not eliminate the noise totally—the variation in intensity was reduced to about ± 3 to ± 4. The light source is in this case was the room's fluorescent lighting, powered by a 50Hz AC supply. Roughly half the noise may thus be attributed to scene illumination and the other half due to noise in the analog PAL signal.

4. Searching Algorithms

The searching algorithms suggested by Krotkov and the Matrox Imaging Library were evaluated. This section describes the global search method and the Fibonacci search method and then goes on to introduce two more search methods—the search by percentage drop and one still image technique. In addition, two search refinement methods are introduced, the search by centre of area and the search by pulsing. The problems encountered with the development of these searching methods led to some additional safety checks to make the algorithm more robust.

In the global search method, each lens position is scanned and the sharpness of its image calculated. After all lens positions have been scanned, the searching algorithm attempts to move the lens back to the position where the best image was obtained. Krotkov recommends the Fibonacci search technique [1] as the optimal search strategy [6, 7]. This strategy is based on continuously narrowing the search region by subdividing it according to the Fibonacci sequence. The required number of iterations for this search is the least integer N, such that FN ³ the initial search interval. For 256 different lens positions, the first Fibonacci number just exceeding 256 is F13 = 377, so this search will require at most 13 steps to find the focused image. However, this search technique can only be implemented on a camera for which the focusing lens may be controlled by specifying its position. Moreover, in the case of the camera system used, the lens motor moves too slowly for the Fibonacci search to work well. The Matrox Imaging Library in version 6.0 provides a smart search technique that repeatedly halves the search region into smaller and smaller portions. In addition, a global search technique is also supported. The main advantage of the global search method is that it generally guarantees that a local maximum will not be mistaken for a global maximum. The problem with this function-based method is that the sharpness function is affected by noise in the image. Thus, even if the lens were to return to the exact position corresponding to maximum image sharpness, the sharpness function would not return the same value. Thus, some allowable degree of error must be allowed in the sharpness function, meaning that the lens will be "close to," but not "at" the focus position. .

We introduce the "Searching by Percentage Drop" as a modification of the global search, where not all lens positions are scanned. Rather, the lens positions are scanned until a percentage drop by a predetermined amount is detected. To calculate the percentage drop in the sharpness function, the formula used was not (fmax-f)/fmax, (where f is the value of the sharpness function), since some sharpness functions have ranges which do not start at zero, (e.g. 7.5 to 8.5 or 400 to 700). The formula was modified to (f-fmin)/(fmax – fmin) in order to account for this. However, it should be noted that edge-detection based functions, namely the Tenengrad and the Laplacian, do not suffer from this problem since fmax >> fmin. In the search by percentage drop, several parameters must be passed to the function—the noise amplitude in the image, the initial direction of search, the magnitude of the percentage drop to look for, the minimum number of steps before starting to search for the drop, as well as the criteria for determining whether the lens has reached its minimum or maximum position. This technique aims to improve upon the global search technique by reducing the distance for which the lens motor must move. The speed of this technique is determined in part by the allowed drop in the sharpness function. Setting too low an allowed drop might cause the searching algorithm to be caught in a local maximum. Too high a drop would result in a large overshoot. In addition, if the sharpness function does not decrease by a large enough amount after hitting the maximum, the technique will fail. In addition, there is some difficulty in determining when the lens has overshot its limit. This algorithm is also subject to the accuracy of the several parameters it is supplied with—the noise amplitude, the magnitude of the percentage drop, and the condition to determine when the lens has reached its maximum position.

We introduce the "Searching by Centre of Area" method as a way to further refine a search. When the point of best focus is near to being found, a baseline is specified and centre of area method uses the region of the sharpness function above this baseline for its calculation. The location of the centre of area of this region is in general, not the maximum point of the sharpness function. However, due to the shape of most sharpness functions, the curve is almost flat at the top possibly due to the depth of field. For this reason, the centre of area is a good estimate for a location that will fall within the depth of field of the camera. Another possible method is to find the average of the two limits of the depth of field. However, this estimate is not as good as the centre of area method in a typical sharpness function as shown in Figure 4.

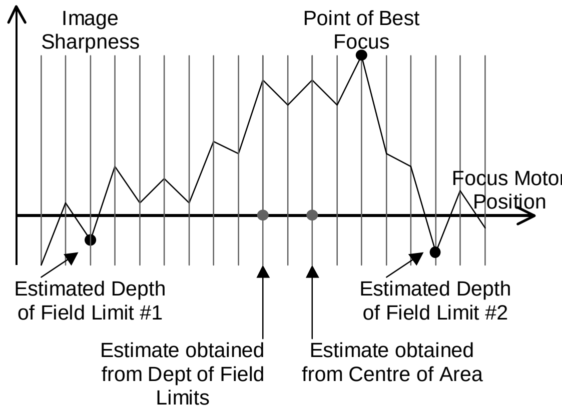

Figure 4: Centre of Area Estimate for Point of Best Focus

Some information that must be passed to this function include the baseline value, the noise amplitude of the sharpness function and the initial direction of search. It should be noted also that if the noise is reduced by smoothing the curve in Figure 5-1 above, the point of best focus as measured by the smoothed curve would almost coincide with the estimate obtained by the centre of area. In general, the accuracy of this method will depend on how close the value of the base is to the maximum of the function. This method works best after another search method has already found the approximate location of the maximum point. This method works as a good alternative to finding the minimum and maximum positions of the depth of field and taking the average. This is because finding the minimum and maximum positions is quite different as there may be several points where the sharpness function crosses the baseline. Noise may further decrease the accuracy of determining the minimum and maximum positions. One problem with this method is that if the sharpness function never drops below the baseline specified, the search method will fail. This problem is only likely to occur if the focus position is not near enough. In addition, the assumption that the centre of area is near to the function maximum is only true when the focusing position is near. For these two reasons, the search by the centre of area is recommended only as a search refinement method.

5. Conclusions

Based on the good performance of the variance and Tenengrad methods in determining image sharpness, not only in these tests, but also in the results by Krotkov, these two methods have been determined to work well in the crane scene analysed. The implementation of these two methods in the different search algorithms was successful, while in the crane scene, these two methods failed where the image was uniform.

Of the search algorithms tested, the global search and the search for the percentage drop were found to be the fastest initial search methods. The other two methods, the centre of area method and the pulsing method were found to be more suited for refining a search when the focusing point is near to being found.

As the graphs comparing the effect of colour on image sharpness show, the wavelength of light apparently does not play a significant role in affecting the position of best focus for the crane scene tested. Focusing based on only one colour channel, red, green, blue, or grey is likely to be sufficient for most situations.

The best performance gain comes from using greyscale values for focusing as well as smaller focusing windows. Despite the differences in speed of the various sharpness functions, this effect is small compared to image grab time for all the sharpness functions except the Fast Fourier Transform.

In implementing a camera system with pixel-based autofocusing, the following may be taken into consideration. Firstly, one must ask whether focusing is required. If the depth of field is large, a fixed focus setting may be sufficient. In addition, the built-in autofocus of the camera is often adequate.

In terms of sharpness functions, the variance method and the Tenengrad have been found to be adequately suited for the task, based on the results of this project and the results obtained by Krotkov. The combination of the three searching algorithms the “search for percentage drop” method as the initial search algorithm, with the “search for the centre of area”, followed by the “pulse search method” for further refinement, appears to be sufficient to handle most situations. However, some tweaking of the criteria to determine when the camera lens has reached its limits is necessary to improve the robustness of the “search for percentage drop” method so that it can handle the situation where the initial lens position is unknown.

For the purposes of focusing, the camera with lens position feedback is highly recommended, due to its combination of speed, repeatability, as well as its ability to give position feedback. In order to further reduce the noise in the sharpness function, a camera following the IEEE1394 digital standard may be used as more become available.

Continuous focusing is difficult to achieve using the currently available hardware. The sharpness function is only an indication of the degree of defocus, but it does not provide information as to where to move the camera lens. The hardware focusing method by Pentax shows that at least two images are required to know the direction in which the lens should be moved.

Acknowledgments

The support of Singapore’s National Science and Technology Board, Jurong Shipyard and Sembawang Shipyard is gratefully acknowledged. Autofocusing is used in our pipe measurement project for these two shipyards.

References

1. Beveridge, G. S., Schechter, R. S. Optimization: Theory and Practice, McGraw-Hill, New York. 1970.

2. Davies, E.R. Machine Vision: Theory, Algorithms, Practicalities. Academic Press Limited. 1990.

3. Horn, B. K. P., "Focusing," *Technical Report AIM-160*, Massachusetts Institute of Technology, May, 1968.

4. Jarvis, R. A., "Focus Optimisation Criteria for Computer Image Processing," Microscope 24(2), 1976, pp. 163-180.

5. Johnson, S.M. “Optimal Search for a Maximum is Fibonaccian,” Technical Report P-856, RAND, Santa Monica, California, 1956.

6. Kiefer, J., “Sequential Minimax Search for a Maximum,” Proc. Am. Math. Soc.(4), pp. 502-506. 1953.

7. Krotkov, Eric Paul. Active Computer Vision by Cooperative Focus and Stereo. New York: Springer-Verlag. 1989.

8. Press, W. H., Teukolsky, S. A., Vetterling, W. T., Flannery, B. P. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press. 2nd Edition. 1992. pp. 504-508.

9. Schlag, J. F., A. C. Sanderson, C. P. Neumann, and F. C. Wimberly, “Implementation of Automatic Focusing Algorithms for a Computer Vision System with Camera Control,” Technical Report CMU-RI-TR-83-14, Carnegie Mellon University, August, 1983.

10. Tenenbaum, J. M., “Accomodation in Computer Vision,” Ph.D. Dissertation, Stanford University. November, 1970.