Depth Perception using Focus with Hand_Eye Robotic System

Jorge Dias, A. de Almeida, Helder Araújo

Dept. Eng. Electrotécnica-Univ. Coimbra
Largo Marques de Pombal, 3000Coimbra,Portugal

Abstract: *Vision systems are a possible choice to obtain sensorial data about the world in robotic systems. To obtain three-dimensional information using vision we can use different computer vision techniques such as stereo, motion, or focus. In particular, this work explores focus to obtain depth or structure perception of the world. In practice focusing can be obtained by displacing the sensor plate with respect to the image plane, by moving the lens or by moving the object with respect to the optical system. Moving the lens or sensor plate with respect to each other, causes changes of the magnification and corresponding changes on the object coordinates. In order to overcome these problems, we propose in this work, to vary the degree of focusing by moving the camera with respect to the object position. In our case the camera is attached to the tool of a manipulator in a hand-eye configuration with the position of the camera always known. This approach ensures that the focused areas of the image are always subjected to the same magnification. To measure the focus quality we use operators to evaluate the quantity of high-frequency components on the image. Different types of these operators was tested and the results compared.*

1. Introduction

Vision systems are a possible choice to obtain sensorial data about the world in robot systems. These systems help to understand the three dimensional world where these systems develop and to control their movements. One important task in these vision systems is to extract depth information from two-dimensional images which normally is followed by the three-dimensional reconstruction of the scene. These three-dimensional data is essential for later stages like object recognition, scene interpretation or path planning. To obtain good results in these later stages is essential to have good data as input and this implies to have accurate depth measurements or measurements with uncertainty known. In vision, depth information can be obtained by use different techniques like, stereo, motion, or focus. Stereo one of the most studied techniques for depth evaluation, but the algorithms proposed until now are very dependent of the correspondence technique used. Motion is another technique that can be used to obtain depth information but this very dependent of the accuracy of the results of the optical flow algorithm used. Focus is another possible technique to depth recovery and the algorithms proposed until now normally explore the focus technique by controlling the positions of lens or the sensor plate. This work presents a different approach to the problem controlling the focus by changing the position of a vision system attached to a manipulator's tool.

Previous work

The focus analysis has been used to automatically focus imaging systems or to obtain coarse depth information from the observed scene. One of the first publication concerning with this type of problems was made by [4] where he proposed focusing imaging systems by using the Fourier transform and analyzing the frequency content in the image. Pentland in [8] proposed two methods for finding the depth-map of a scene. The first method was based on measuring the "blur" of the edges of a defocused image. Grossman in [3] reports the results of some experiments based on this same principle. This method requires the knowledge of the location and magnitude of the edges of the image. The Pentland's second method is based on comparing two images formed with different aperture diameter. Subbarao and Gurumoorthy also recovered the depth using the blurred edges by a different technique [11,12]. All these methods recover the scene depth directly from defocused images. Recently Hwang [15] proposed a two-phase algorithm where the defocus process is modeled as a two-dimensional Gaussian point spread function.

On the other hand, there are several algorithms based on depth from focus, rather than depth from defocus. Some of them use active range finding with infrared or sonar sensors to focusing the images. Another algorithms use criterion functions to measure the sharpness of focus which would be equivalent to getting the sharpest focus and after recover the depth of the object. In [6] is evaluated and compared the performance of different of these focus criterion functions and also a method to estimate the depth of an image area by actively changing the positions of the lens. In the same work Krotkov compares the algorithm proposed by Tenenbaum[13] and based on the gradient magnitude maximization and the algorithms based on the minimization of the histogram entropy proposed by Jarvis [5] and Shlag [10].

2. Active focus

Focused and defocused images

We can describe defocused images as processed versions of focused images. Normally real imaging systems are composed of several lenses but conceptually we can combine all the imaging elements into a single thick lens with a focal length f as shown in the figure 2.1. To analyze the image formation process we can use two different approaches. using the classical geometric optics or diffraction theory. The geometric optics uses ray-tracing process to explain the image formation and is identical to a first-order approximation model since the geometrical distortion of the image is neglected. The classical physical

CH2976-9/91/0000-1312 $1.00 © 1991 IEEE

optics and the diffraction theory are used to obtain exact results but since the cameras used have sufficient spatial resolution to make diffraction effects significant we can consider only geometric effects to analyze the image formation process.

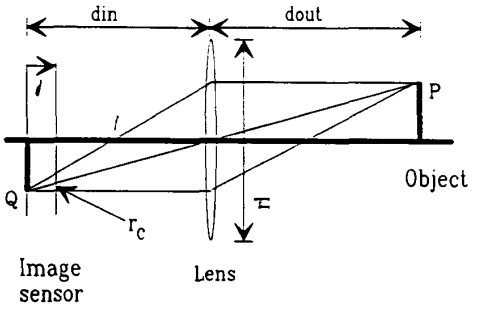

Figure 2.1. - Formation of focused and defocused images using thick lens. The defocused image is achieved changing the place of the image sensor by δ.

Assuming that the lens are thick and the front and back focal planes of them, are planes normal to the optic axis situated at equal distances *f*, the *Gauss lens law* holds,

$$\frac{1}{d_{out}} + \frac{1}{d_{in}} = \frac{1}{f} \qquad (2.1)$$

where d_{in} and d_{out} represent the distance to the image plate and the distance to the object imaged.

To explain the effects of defocusing an image we can trace the rays that light take trough the optical system. Using this principle, for a point at infinity d_{out}= inf we obtain by equation (2.1) $d_{in}= f$. In this case the focused image plane is placed at *f* and the image of the three-dimensional point **P** is also a point **Q**. If the image plane is displaced by an amount $\delta = |d_{in} - f|$ then the image formed on it will be a circle with diameter $2r_c$ since if the aperture of the lens is also a circle. Using simple geometry and by similar triangles we find that

$$\frac{r_l}{d_{in}} = \frac{r_c}{\delta} \quad \Rightarrow \quad r_c = \frac{\delta\, r_l}{d_{in}} \qquad (2.2)$$

where r_l and r_c are the radius of the lens and the circle respectively. The physical optics explains the circle formation by the distribution of the energy received by the optical system. This area have equal shape as the aperture of the system. For points not far from the lens are not in the optic axis, the arguments and the equation (2.2) are the same.

Focus measure

To automatically measure the focus of a region in the image we need a criterion to measure the sharpness in that region. The notion of defocus have the inherent sensing of loss of information compared to focused images, and it is equivalent to the loss of image quality of the

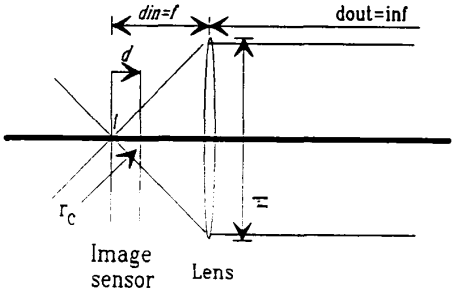

Figure 2.2. - Displacement of detector plane causes defocusing for points in the infinity or near of the lens.

image. Analyzing this phenomena using physical optics, the defocusing appears due to the loss of high-frequency components of light energy arriving to the optical system. If we are using an optical system with aperture circular and **P** is a point on a visible surface in the scene then **Q** will be its focused image as we can see in figure 2.1. If the image plate is changed to another point in the optical axis the image of **P** will be not in focus and a circular image appears on the image plate detector. Analyzing this phenomena by physical optics we can conclude that the intensity within the circular patch is approximately constant and the blurring image can be modeled by a convolution between the ideal image ***I(x,y)*** and a circular function ***circ()***. Assuming that the camera is a linear shift invariant system, this is expressed by

$$I(x,y) * circ(x,y) \qquad (2.3)$$

where the function ***circ()*** is defined as

$$circ(x,y) = \begin{cases} \frac{1}{\pi r^2} & if\ x^2 + y^2 <= r^2 \\ 0 & otherwise \end{cases} \qquad (2.4).$$

This function has the form of a "pill-box" as shown in figure 2.3.

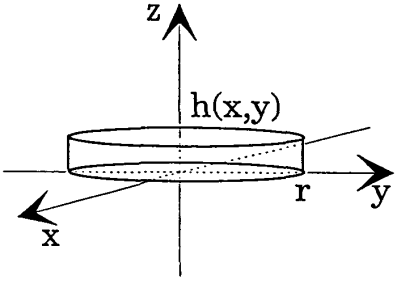

Figure 2.3 - Form of the function ***circ()***.

Tacking the Fourier transform of this circle function we obtain

$$F\{circ(x,y)\} = CIRC(f_x, f_y) = 2\pi r^2 \left[\frac{J_1\left(r\sqrt{f_x^2 + f_y^2}\right)}{r\sqrt{f_x^2 + f_y^2}} \right] \qquad (2.5)$$

where J_1 is a Bessel function of the first kind and order one (see [9] or [7] for details). This transform is circularly

symmetric and consists of a central spike and a series of concentric rings of diminishing amplitude. The form of this point spread function is based purely on geometric considerations but, for a practical camera system we must consider various other effects. If we ignore lens aberrations, a source of distortion of the image is due to diffraction caused by the wave nature of the light for coherent or for incoherent monochromatic illumination. If we try to obtain the transfer functions for these intensity patterns we will see that they are different of the expression (2.5) (see [11] for some examples). Using white light for illumination, the intensity pattern will have a cumulative effect of different intensity patterns produced by the lights of many different wave lengths. The net effect of all these phenomena can be described by a two-dimensional Gaussian point spread function as

$$h(x,y)=\frac{1}{2\pi\sigma^2}e^{-\left[\frac{x^2+y^2}{2\sigma^2}\right]} \qquad (2.6)$$

where σ is the spread parameter which is proportional to the radius of the blur circle [Subbarao 87]. Using this transfer function, the blurred or defocused image $I_d(x,y)$ formed on the sensor plate can also be determined by convolution the focused image $I(x,y)$ with the blurring function $h(x,y)$

$$I_d(x,y) = I(x,y) * h(x,y) \qquad (2.7).$$

Analyzing this defocusing process in the frequency domain we can express the equation (2.7) as

$$I_d(f_x, f_y) = I(f_x, f_y) \cdot H(f_x, f_y) \qquad (2.8)$$

and the transfer function as

$$H(f_x, f_y)= e^{-\left[\frac{f_x^2+f_y^2}{2}\right]\sigma^2} \qquad (2.9).$$

We can see by the transform $H(f_x, f_y)$=, that low frequencies of the light passes and the high frequencies are attenuated. Another important factor is that as the image plate is displaced and the defocusing radius increases the spread parameter σ also increases. Since the defocusing process works like a low-pass filtering process, the bandwith of the light energy decreases with the defocusing increase.

Concluding that the defocusing of optical system attenuates the high-frequency components arriving to the image plate, to measure the focus quality we can quantify the high-frequency content of the image. Several focus measure operators as base for different criterion of "sharpness" have been proposed by vision researchers. Often of them had been analyzed by Krotkov [6] in his thesis where he compares them, using only information in the image.

The objective to automatically control the focus is to find an operator that behaves in a stable and robust manner over a variety of images such as images from outdoor or indoor scenes, text or images with different illuminations.

More or less directly, the operators that have been proposed measure the high spatial-frequency content of the image. The Fourier transform is an first candidate for a criteria function, but it due to its high computational complexity and it needs of special-purpose hardware to improve the calculation speed, it is not normally considered. Since the defocusing affects the edge characteristics it is natural to use an edge detector for the criteria function to measure the quality of focus. Based in this methodology we estimate the gradient $\Delta I(x, y)$ at each image point and simply sum all the magnitudes greater than a threshold value. The focus criterion function only searches a maximum of this sum for different defocused images. Another technique to measure the focus can be based on the filtering of the images by an high-pass filter like the Laplacian

$$\Delta^2 I = \frac{\delta^2 I}{\delta x^2} + \frac{\delta^2 I}{\delta y^2} \qquad (2.10)$$

where $I(x,y)$ is the image intensity at the point (x,y). The focus criterion function can be then computed by summing the absolute value of the Laplacian at each image point with magnitudes greater than a threshold value. The criterion function, and like above, only searches a maximum of that sum between different defocused images from the same scene.

If we take an image from a scene with half part black and the another half white and the image is taken with an optical system defocused, we can see that image histogram normalized tends to have an uniform distribution. If the image is enhanced and a sharply focused edge is obtained, the same histogram tends look like two spikes one corresponding to each side of the edge. Defining the histogram entropy as

$$Ent = -\sum_n P(n)\, ln(P(n)) \qquad P(n) \neq 0 \qquad (2.11)$$

where $P(n)$ denotes the frequency of the occurrence of grey-level n then Ent takes its maximum value when all $P(n)$ are equal. By this definition, the blurred edge image will have greater entropy than the sharp edge image and the criterion will be to minimize this entropy.

Another fact is that high grey-level variance of the image histogram is associated with sharp quality of the images while low variance is associated with blurring. The blurring reduces the amount of grey-level fluctuation and the variance can be a criterion to measure the defocus. Adopting the standard definition of variance as

$$\sigma^2 = \frac{1}{N^2}\sum_{x=1}^{N}\sum_{y=1}^{N}(I(x, y) - \mu) \qquad (2.12)$$

where μ is the mean of the grey-level distribution. Using this principle the criterion is to maximize σ^2, which indirectly corresponds to maximizing the integral of the power spectrum of the intensity distribution.

The figure 2.5 shows the results for the application some of the operators explained above for a sequence of images similar to images given on figure 2.4.

$$d_{in} = \frac{d'_{in}}{1 + \frac{r_c}{r_l}} \qquad (2.14).$$

Combining the equation (2.13) and (2.14) we obtain

$$d_{out} = \frac{d'_{in}\, r_l f}{d'_{in} - f - r_c \left(\frac{f}{r_l}\right)} \qquad (2.15).$$

where $\left(\frac{f}{r_l}\right)$ is known as *f-number* of the lens.

Using the equation (2.15) we can recover the depth of the object from an image, focused or not, if we know radius r_c. But the real problem with this method for depth recovery is to evaluate the r_c from the image. Another two techniques using the same principle is presented in [11] and [15] where they made a formal analyses of the process and proposed algorithms to evaluate the depth using defocused images. In practice all these processes need a previous calibration of the camera and the results obtained are very coarse.

3. Active focus as approach to depth recovery

As explained in the point **2.3**, defining d'_{out} as the distance to the object for an image perfectly focused we can recover the depth of the object by moving the camera until obtain a focused image of the object if we know the position of the camera relatively to the object. For this approach we need to do the calibration of this distance d'_{out} since we do not move the lens or the image plate.

To solve this problem we need to know with a reasonable accuracy, the relative transformation between the 2D images and the referential of the three-dimensional world. In our case the camera is placed on the last link of the manipulator which have a referential associated known as ***{TOOL}***. One referential called ***{CAM}*** is placed in the camera and two more, ***{BASE}*** and ***{W}***, represents the referential of the manipulator and the referential of the world respectively.

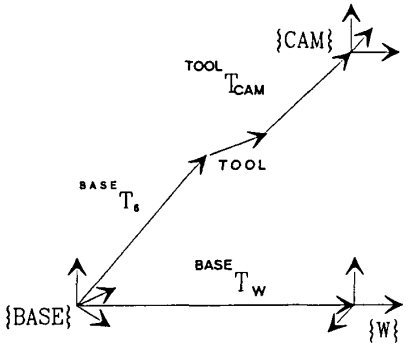

Figure 3.1 - Geometrical relations between the different referentials used in the system

One object is normally referenced to the referential ***{W}*** and the position of the referential ***{CAM}*** is possible to know by commands to the manipulator [2] and by the calibration of the extrinsic and intrinsic parameters of the camera [14] [1].

Calibrating the position d_{out} where the image is focused and after searching the position where a particular object is focused, the depth to the object is easily recovered by the geometric relations between the manipulator the camera and the referential ***{W}***. To focusing the object we do movements in the direction of the optical axis until obtain an image of high-quality with the object positioned within a window. Automatic window selection based on these criterion requires some form of selectivity or can be used to automatically do the correspondence. It is also necessary to guarantee that the feature stays within the window. The evaluation window must either be large enough so that in the course of the movement the focusing distance the point does not travel outside the window.

To obtain the transformation corresponding to this movement we begin by defining a line ***N*** that passes by the point that we are trying to focus. Defining the vector **n** as the direction of this line, the plane XOY defined by the $\hat{k}$ must to achieve the vector ***n***. This equivalent to rotate by an angle of *arccos* $(\hat{k}.n)$ with $\hat{k}.n >= 0$ - see figure 3.2. with the axis of movement given by $\hat{k} \Lambda n$.

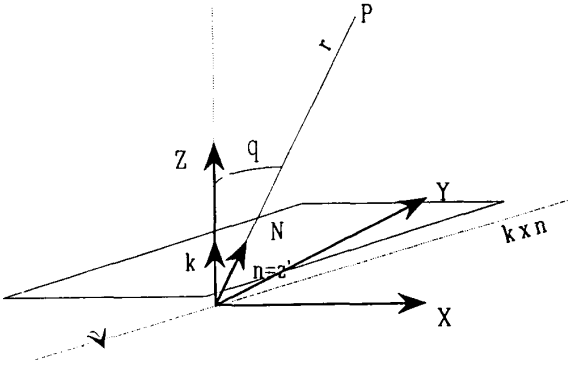

Figure 3.2 - Movement of the referential {CAM} to place the optical axis coincident with the vector ***n***.

The rotation matrix resulting of this movement is given by:

$$R = \begin{bmatrix} \left(1 - \frac{n_x^2}{1+n_z}\right) & -\left(\frac{n_x n_y}{1+n_z}\right) & n_x \\ -\left(\frac{n_x n_y}{1+n_z}\right) & \left(1 - \frac{n_y^2}{1+n_z}\right) & n_y \\ -n_x & -n_y & n_z \end{bmatrix} \qquad (3.1)$$

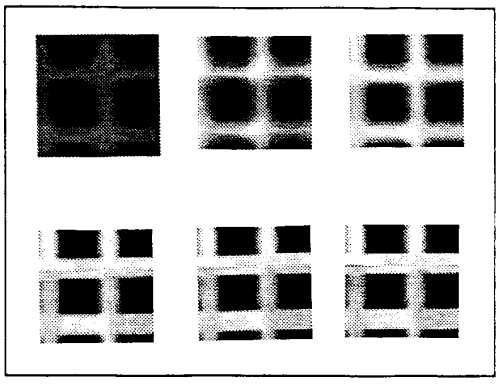

Figure 2.4 -Some of the frames from the sequence used to test the operators for different focus quality criteria.

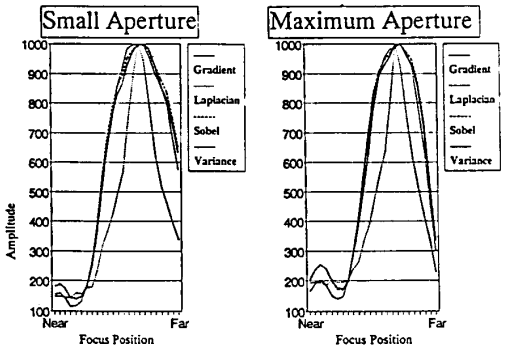

Figure 2.5 -Diagrams describing the evolution of the operators outputs explained above and for a sequence of 8 images.

Since in the normal scenes are impossible to have all object focused, it is necessary to select evaluation windows containing features or a region of interest to evaluate the focus. Conversely, if the window contains the projection of two or more object points lying at different distances, then the criterion function will in general have more than one peak as show on figure 2.6.

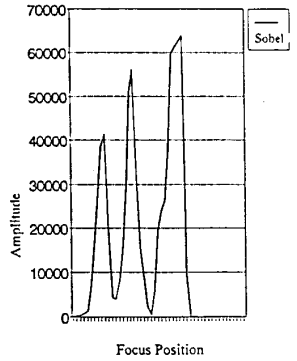

Figure 2.6- Using a stack of images of a scene with objects at different distances, the criterion function presents more than one peak.

Using this fact we can establish an algorithm to obtain a depth map for a sequence of images with different focus and for a scene with objects at different distances. First we divide the images in small windows until obtain windows with only one peak of focus criterion function (without local maximum). These windows on the image have a peak corresponding to a different distances on 3D space. This is equivalent to establish a depth map by regions. This approach is currently used in another work using motorized lens.

In practice focusing can be obtained by displacing the sensor plate with respect to the image plane, by moving the lens or by moving the object with respect to the optical system. Moving the lens or sensor plate with respect to each other, causes changes of the magnification and correspondent changes on the object point's coordinates. In order to overcome these problems, we propose to vary the degree of focusing by moving the camera with respect to the object position. In our case the camera is attached in a hand-eye configuration and with the position of the camera always known. This approach ensures that the focused areas of the image are always subjected to the same magnification. Another effect of moving the camera is a slight reduction in brightness as the lens moves away from the sensor plate, distributing the same energy over a larger area (this can be compensated by normalizing the intensity values).

Depth evaluation

Using the equation (2.1) for a situation of an image perfectly focused we have

$$d_{out} = \frac{f\, d_{in}}{d_{in} - f} \tag{2.13}$$

and for a particular lens, the focal length f is constant. If we fix the distance d_{in} between the lens and the image plane we can focus an image by changing the distance from the object to the lens d_{out}. Defining d'_{out} as the distance for that condition and obtained by calibration we can recover the depth of the object by moving the camera until obtain a focused image of the object if we know the position of the camera relatively to the object. We propose this approach as another process for depth perception on active vision with the advantage that this approach ensures that the focused areas of the image are always subjected to the same magnification. In our experimental setup the camera is attached in the last link of one manipulator and in a hand-eye configuration. The position of the camera is always known and the movement is made to maintain the object into a window centered relative to the center of the image. This is achieved by a previous movement that puts the object into the window and all the other movements are made in the direction of the optical axis. In practice we have to check if object is into the window and rectify the trajectory when necessary.

Another method developed by [8] uses defocused images to directly recover the depth of an object. From the equation (2.2) and displacing the image plane by $\delta = |d'_{in} - d_{in}|$ with d'_{int} the distance to the image plate we obtain the equation

4. References

[1] A.T. Almeida, U.C. Nunes, J. Dias,; H. Araújo, J. Batista,: A Distributed System for Robotic Multi-Sensor Integration. Int.Journal of Industrial Metrology ,1990, pag. 217-229

[2] E. Capote, J. Machado, F. Bastos, J. Dias, "Utilização do DDCMP". Internal Technical Report, DEE-University of Coimbra, 1989 (in portuguese)

[3] P. Grossamann, "Depth from focus", Pattern recognition Letters, Vol 5, No 1, 1987

[4] B.K. Horn,"Focusing",MIT Artificial Intelligence Laboratory, Memo N°160,May 1968

[5] R.A. Jarvis, "Focus Optimization Criteria for Computer Image Processing", Microscope, Vol24, No2, 1976

[6] E.P.Krotkov, "Active Computer Vision by Cooperative Focus and Stereo", Springer-Verlag Series in Perception Engineering, 1989

[7] Jae S. Lim, "Two-dimensional Signal and Image Processing", Prentice-Hall, 1990, pg. 29-30

[8] A.P. Pentland, "A New Sense for Depth of Field", IEEE Trans. on PAMII, VOl PAMI-9, No4, July 1987

[9] W.F.Schreiber, "Fundamentals of Electronic Imaging Systems", Springer-Verlag Series in Information Sciences, 1986, pg36

[10] J.F. Schlag, A.C. Sanderson, C.P. Neumann, and F.C. Wimberly, " Implementation of Automatic Focusing Algorithms for a Computer Vision System with Camera Control", Tech. Report CMU-RI-TR-83-14, Carnegie Mellon University, Aug. 1983

[11] M. Subbarao, "Direct Recovery of Depth-map I: Differential Methods", Proc. of the IEEE Computer Society- Workshop on Computer Vision, 1987

[12] M. Subbarao and N. Gurumoorthy, "Depth Recovery from Blurred Edges", Proc. of the IEEE Computer Vision and Pat. Rec., 1988

[13] J.M. Tenenbaum, "Accommodation in Computer Vision", PhD Thesis, Stanford University, Nov, 1970

[14] R.Y. Tsai: A Versatile Camera Calibration Technique for High Accuracy 3D Machine Vision Metrology Using off-the-shelf TV Cameras and Lenses. IBM RC 11413, October 1985

[15] T.Hwang, J. Clark, and A. Yuille, "A Depth Recovery Algorithm Using Defocus Information, TR No 89-2, Harvard Robotics Laboratory, USA, 1989