Proceedings of the World Congress on Engineering and Computer Science 2007
WCECS 2007, October 24-26, 2007, San Francisco, USA

Depth Measurement and 3-D Reconstruction of Multilayered Surfaces by Binocular Stereo Vision with Parallel Axis Symmetry Using Fuzzy

Sharjeel Anwar, Dr. Shoaib, Taosif Iqbal, Mohammad Saqib Mansoor, Zubair Shafiq, Abdul Moeed,

Department of Electrical Engineering, College of E&ME (NUST), Pakistan

ABSTRACT— **Among various range finding methods stereo vision is worthy of notice since it needs no active media. However because of correspondence problem, the stereo vision has not been fully established as a computer vision system. In this paper we used fuzzy logic to bypass the correspondence problem in stereo vision for calculating the distances of objects.**

Index Terms— **Correspondence, disparity, Mamdani's fuzzy model, 3-d reconstruction.**

I. INTRODUCTION

There have been many matching algorithms for stereo vision proposed, which may be classified into feature based stereo vision and area based stereo vision methods. In feature-based method, Marr and Hildreth introduced the convolution operator V^2 G, where V^2 is the Laplacian operator and G stands for the two-dimensional Gaussian distribution, and adopted the zero-crossing (ZC) of the V^2G-filtered image as the features to be matched (Marr and Hildreth, 1980). A problem with matching the ZC is that ZC points may appear randomly in the region of little intensity change. If we try to match ZCs including such random ones, the probability of false matching becomes large.

In this paper we tried to use a different approach so that 3d reconstruction and distance calculation of objects can be done even for scenes with multilayered surfaces. We used fuzzy logic for this purpose. Our approach was to assign membership functions to two inputs, (1) sizes of the objects in the image, (2) disparities of the objects, to find the distances of all the objects in the scene, from the cameras. Disparity maps were constructed for this purpose after applying canny edge detector to both the image planes. This paper is organized in the following way, in section II, we describe the previous work, in section III, we describe the stereo vision system, we then discuss our system which is divided into two modules; section IV is the image processing module and section V is the fuzzy reasoning module. In image processing module we describe how disparity maps are constructed. In fuzzy reasoning module we describe the application of fuzzy in our system. In section VI 3-d reconstruction is discussed. In section VII results are shown and section VIII discusses conclusion and future work.

II. PREVIOUS WORK

There had been a lot of work done in the field of stereo vision, of which we would just like to mention some work related to our project. Previous work was done by Ding and Ronald [2] on correspondence free stereo vision, but their algorithm was only applicable to planar surfaces or scenes pictured from large distances. They had also not used fuzzy for this purpose. Yasamin and Monsour [5] used neural networks in robot navigation. Marr and Hildreth introduced the convolution operator V^2 G, where V^2 is the Laplacian operator and G stands for the two-dimensional Gaussian distribution, and adopted the zero-crossing (ZC) of the V^2G-filtered image as the features to be matched (Marr and Hildreth, 1980).

III. STEREO VISION SYSTEM

In what follows, we consider a "pinhole" camera model. This model connects a point of the 3D space to its perspective projection (i.e. its image) on the camera retina. This transformation is linear in the projective space and is described by the intrinsic and extrinsic parameters. The intrinsic camera parameters, such as the focal length *(f)*, the pixels size (k_u, k_v) and the image coordinates $(U_o, v_o)^T$ of the projection centre describe an affine transformation representing scaling, rotation and translation between the camera and the image references. The extrinsic parameters describe the rigid transformation from the world reference to the camera reference. This transformation is entirely defined by a 3x3 rotation matrix and 3x1 translation vector. The stereovision system we use is made of two cameras, in a configuration in which their optical axis were parallel. The relationship between the images of the two cameras is described by a 3x3 homographic *F* matrix named "fundamental matrix". This matrix has many important properties**,** among which the most important for our application is:

$$m^T Fm=O \qquad (1)$$

Equation (1) describes exactly the relationship between the homogenous coordinates of point *m'* belonging to the right image, and the homogenous coordinates of its counterpart *m* in the left image. In our particular case, the Fundamental matrix has the following form:

$$F= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \qquad (2)$$

This fundamental matrix also describes the epipolar geometry of our stereoscopic sensor. The analysis of this homographic form allows us to conclude that any point belonging to the left image will have a correspondent, if any, on the right image whose row's number is the same as its counterpart's one. Expression (2) for the fundamental matrix is obtained when the cameras' retinal planes can be deduced from each other. In parallel axis symmetry the translation is only along the x axis. That means that the cameras' retinas belong merely to the same plane. The step allowing us to compute the fundamental matrix exploits the Least Median of squares method**.** Comparing the actual fundamental matrix coefficients with their respective theoretical values from (2) gives us a relevant way to verify the precision of the binocular vision system configuration. Using homogeneous coordinates in a Projective space, we get:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_l & T_l \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad (3)$$

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} \alpha_u & 0 & u_0 & 0 \\ 0 & \alpha_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_r & T_r \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad (4)$$

Where:

$(x \ \ y \ \ s)^T$ and $(x' \ \ y' \ \ s')^T$ are the homogeneous coordinates of the projection of point P_i on the left and right image respectively.

α_u and α_v, are defined as follows: $\alpha_u = \frac{f}{k_u}$ and $\alpha_v = \frac{f}{k_v}$.

R_l and R_r are the rotation matrices, expressing rotations between the scene reference and the left and right cameras references, respectively.

$(X_i \ Y_i \ Z_i \ 1)$ are the homogeneous coordinates of point P_i in the scene reference. T_l *and* T_r are the translation vectors, expressing translation between the scene reference and the left and right cameras references, respectively.

A and A' are the perspective projection matrices. They express the transformation between the cameras coordinates system and the image coordinates system. In our case, we use a pair of identical cameras, thus we will consider $A \approx A'$.

The transformation between the scene reference and the cameras references can be expressed as a θ radian rotation around x-axis and a translation. Thus, rotation matrix and translation vectors can be written as follows:

$$R = R_l = Rr = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \qquad (5)$$

$$T_l = [\delta/2, h\cos\theta, h\sin\theta]^T \qquad (6)$$

$$T_r = [-\delta/2, h\cos\theta, h\sin\theta]^T \qquad (7)$$

OUR SYSTEM:

Our system consisted of two modules

iv. Image processing module

v. Fuzzy reasoning module

IV. IMAGE PROCESSING MODULE

CONSTRUCTION OF DISPARITY MAPS

Similarity measurement between pixels: Matching between two pixels can generally be expressed in terms of similarity estimation between two areas. We have chosen to compute this similarity with the normalized correlation coefficient:

$$C(x,y) = \frac{\sum_{y'=0}^{y'=h}\sum_{x'=0}^{x'=l}\tilde{M}(x',y')\tilde{I}(x+x',y+y')}{\sqrt{\sum_{y'=0}^{y'=h}\sum_{x'=0}^{x'=l}\tilde{M}(x',y')^2\sum_{y'=0}^{y'=h}\sum_{x'=0}^{x'=l}\tilde{I}(x+x',y+y')^2}} \qquad (8)$$

Such as:

$\tilde{M}(x',y') = M(x',y') - \overline{M}$

and $\tilde{I}(x+x',y+y') = I(x+x',y+y') - \tilde{I}(x',y')$

Where

$\overline{M}$ is an average intensity of M in the model selected window.

$\tilde{I}$ is the average intensity of I in the selected window of the search image.

It is known that this correlation coefficient gives better results (especially in terms of robustness) than other possible expressions. This is due to the fact that it is not too sensitive to illumination changes.

A. CONSTRUCTION OF DISPARITY MAP WITH RIGHT IMAGE PLANE AS REFERENCE

Both the right and left image planes were divided in equal patches, each of which was of size 20*20. First of all, the right image plane was taken as a reference. The first patch of the first row of patches was correlated number by number with all the first row of patches. The patch number was noted with which it gave maximum correlation. The disparity of the first patch of the right image plane was obtained by,

$$d_{r1}(k,j) = 20 \times (i-j) \qquad (9)$$

Where 'i' is the patch no of left image plane's Patch (in k-th row of patches) which gave the max correlation with j-th patch of right image Plane (in the same row of patches).

This value of disparity was placed at the 1st Point (i.e. $d_{r1}(1,1)$) of the disparity matrix. Similarly the disparities of all the patches were calculated and they were placed at same position in the disparity matrix, which was the patch number of the right image plane.

B. CONSTRUCTION OF DISPARITY MAP WITH LEFT IMAGE PLANE AS REFERENCE

The same procedure as explained in A, was applied taking left image plane as reference i.e. the disparity matrix of left image plane was made by finding the disparities of all the patches of the left image plane by,

$$d_{l1}(k,j) = 20 \times (j - i) \tag{10}$$

C. CONTRUCTION OF drr1

Now from these two disparity maps, a third disparity map was made. The 1st disparity value of the 1st row of patches of dr1 was picked up; it was divided by 20 to get a column number of dl1, (Take i1). In ‘drr1’, the 1st row was filled by same values of ‘dr1’ up to ‘n-i1’, where ‘n’ is the total no. of columns. The rest of the row from ‘n-i1’ was filled by the values of ‘dl1’. Similarly all the rows of ‘drr1’ were filled.

V. FUZZY REASONING MODULE

DEPTH MEASUREMENT USING FUZZY

Now by using the expert knowledge (that we obtained from experimental data), we used Mamdani’s Fuzzy model, for calculating the distances of the objects from the cameras. For this purpose we had to find out the sizes of the objects, and we assigned membership functions for certain ranges of disparities and sizes of objects, and these membership values corresponded to the distances of the objects from the cameras. For this purpose edges detection techniques were exploited and disparity maps, ‘dr2’, ‘dl2’ and ‘drr2’ were made by the same procedure as described in section 2, after edges detection of both the images by suitable edge detector.

A. EDGES DETECTION

We exploit edges detection in order to separate the different objects belonging to the images. We experimented on three types of edges detection techniques in order to get the best one for our work. We used prewit and canny edge detectors and a first level ‘Haar’ wavelet transform. We removed the approximation details after taking the wavelet transform of both the images, and then taking the absolute value of the reconstructed image of rest of the details, we obtained the edges of the objects. After this process, we increased the dynamic range of the images by setting the values at the lower range of intensities (‘0.04’, in scaled double case) to zero, and intensity values above 0.1 to ‘1’ (for scaled double values b/w 0 and 1). At some places it gave more information about the edges but it also enhanced the noise in the image, so we chose the canny edge detector for our work. The idea underlying this filter is simple and effective: if we observe any image containing different regions we can see that edges can be modeled as a crossing between two different intensity levels (we are processing grey levels images).To cater for the noise, we used a median filter in our process. For this reason, before applying the canny edge detector we pre-process the images with a median filter that reduces noise on raw images.

Remaining procedure:

‘drr2’ consisted of zeros at all the pts, except where the edges were present. The sizes of objects were obtained by searching for nonzero values (that would surely be the disparities only), along the rows and the columns. As the disparities are already known, by knowing the size of the objects, we obtained their nearest distances from the cameras. Now as we knew the positions of the objects, we searched for the disparities in the same area occupied by objects in drr1. Now again using Mamdani’s Fuzzy model we now assigned the membership functions for certain depths and layers in objects (i.e. for multilayered surfaces), and these membership values corresponded to changes required in the intensity values for the 3-d reconstruction.

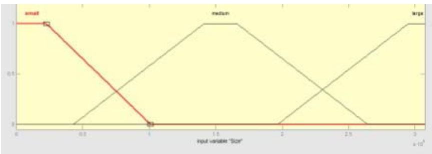
VI. 3-D RECONSTRUCTION

The 3-d reconstruction was done by observing the disparity maps, those feature points which were perceptible in left image plane and imperceptible in the right image plane had high disparities(almost equal to n), in the disparity map of drr1. Those feature points which were perceptible in right image plane and imperceptible in the left image plane had high disparities in the disparity map of drr2. These feature points were given low intensity values, and this change in intensity values (with variations in pixels) depended on the membership values assigned to a certain depth in object’s surface (which was obtained by variations in disparities).

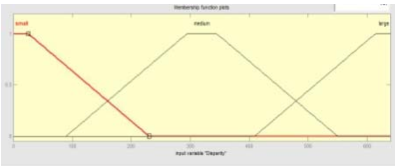
1st of all, the 1st disparity value is taken and in resulting image’s matrix, right image plane’s intensity values from drr1/2 to drr1 are mapped, then the rest of 1st the row of patches is filled by taking the mean of the intensity values of both the images, then disparity at patch no. ‘1’,(i.e., of dl1) is taken and then the rest of the 1st row(of resulting image) is mapped with intensity values of right image plane starting from ‘1’ to ‘dl/2’,similarly all the rows of patches in the resulting image’s matrix, are filled. Now filling all these patches special care is taken for all the feature points which are present in one image plane and missing in other image plane, and for patches of same object with varying disparities. Their intensity values are mapped by the fuzzy rules.

VII. RESULTS

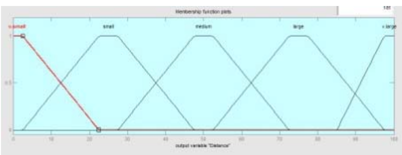
Pictures were taken of multilayered scenes, with baseline distance of 10cm.



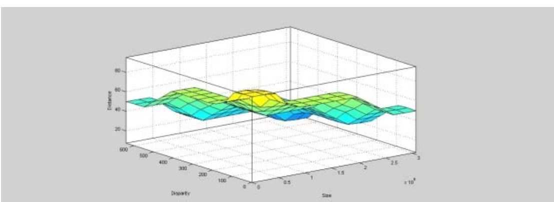
Membership functions for size



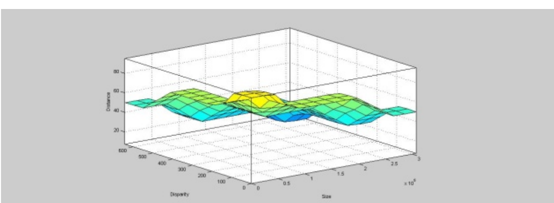
Membership functions for disparity



Membership functions for distance



Surface plot for experiment 1



Surface plot for experiment 2

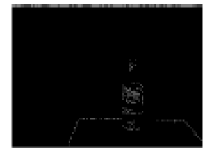
EXPERIMENT 1

Right Image

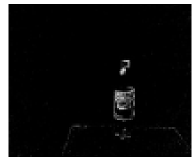
Right Image Plane after using canny edge detector

Edge detection of Right Image Plane using wavelet transform

Left Image Plane



Left Image Plane After using canny edge detector



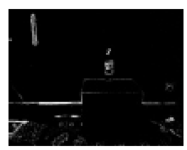
Edge detection of Left Image Plane using wavelet transform

3D reconstructed Image

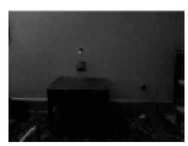
EXPERIMENT 2

Left Image Plane

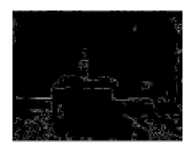
Left Image Plane after using canny edge detector



Edge detection of Left Image Plane using wavelet transform



Right Image Plane



Right Image Plane after using canny edge detector

Edge detection of Right Image Plane using wavelet transform

3D reconstructed Image

VIII. CONCLUSION AND FUTURE WORK

We have used the algorithm for cameras with parallel axis symmetry. Because of the involvement of slight rotation and translation along the Y-axis the results that we obtained, were not very precise for 3D reconstruction of images. We experimented on scenes that are pictured from a short distance because of the less expert knowledge available to us. We obtained distances using Fuzzy Logic for these scenes. In future we want to extend this work for scenes that are pictured from a large distance to get the distances of objects that are far away from us and to get the 3D coordinates of objects moving with very high speeds.

REFERENCES

[1] M. Zayed and J. Boonaert, 'Obstacles detection from disparity properties in a particular Stereo vision system configuration' (Published).

[2] Ding yuan and Ronald chung, 'Correspondence free stereo vision for the case of arbitrarily positioned cameras', in Proceedings of 2003 lEEE international Conference on Robotics & Automation, Taipei, Taiwan, September 14-19, 2003

[3] A. J. Lacey, N. A. Thacker, P. Courtney and S. Crossley, 'The evolution of TINA stereo vision subsystem' (Published).

[4] Sasakthi S. Abeysinghe and loganathan krishanthan, 'Three dimensional motion tracking using stereo vision (Published).

[5] Yasamin Mokri and Monsour Jamzad, 'Omni stereo vision system for an autonomous robot using neural networks' (Published).

[6] Y. Nishimoto and Y. Shirai 'A parallel matching algorithm for stereo vision' (Published).

ISBN:978-988-98671-6-4