International Journal of Scientific & Engineering Research, Volume 5, Issue 2, February-2014 705
ISSN 2229-5518

Efficient Approach for Single Camera 3D Images System

Asst. Prof. Dr. Dhafer R. Zaghar* , Dr. Ekbal Hussein**, M.Sc. Student Hala A.A. Numan**

* AL-Mustansiriyah University – Engineering Collage – Computer and Softwear Dept. -Baghdad - Iraq.
** Technology University - Computer Engineering Dept. – Baghdad - Iraq.

Abstract— The 3D techniques today has a wide range of applications, especially in medical and art, so it is one of the most promising technologies and important for the future. The error in disparity map and highly computational cost and execution time are the major obstacles in the applications of stereo vision to extract 3D. Catadioptric stereo system has newly acquired an increasing importance in 3D. The two captured images can be easily obtained at the same time without the need of exact multi-camera synchronization and calibration. Additionally, catadioptric system required 10 Degree Of Freedom (DOF), while stereo cameras have 16 DOF.

This paper proposed an algorithm to find the 3D information of scene by using mirror reflections of that scene. It will use the geometry of catadioptric stereo with two planar mirrors and single camera to show the epipolar constraints and estimate the fundamental matrix and the relative orientation. The proposed algorithm has three additional points to improve the previous works. First point is the calibration process obtaining intrinsic parameters; second point it will speed up execution time by using Sobel filter with vertical edge detection. Third point it will use the adaptive block size with vertical edge detection matching technique to find the disparity map. The significance of this work is reduction of the execution time and the error of the catadioptric stereo vision algorithm.

Key words: stereo vision, catadioptric, disparity mapping, epipolar constraint, fundamental matrix, essential matrix, adaptive block matching, Sobel filter.

1 Introduction

Optical systems combine the refracting (lens) and reflecting (mirror) elements are called catadioptric systems, these systems used to satisfy the stereo vision, where the stereo view can be captured by a single camera (catadioptric stereo) and using two or more mirrored surfaces. The catadioptric system can provide the third dimension by grabbing left and right images simultaneously, like as human's binocular vision [1].

Stereoscopic vision has an important role in many fields, such as in medicine, industrial robotics. Standard stereo vision are consist of two pinhole cameras and are generally expensive and have many difficulties in calibration, synchronization. To stand up these problems, several works in the robotics and computer vision literature have recently proposed the use of catadioptric vision sensors [2], which, combining both refracting (lens) and reflecting (mirrors) elements, so depth map of 3D object can provide in simple, compact, low cost and high speed, with small error, in order to achieve real-time performance in future depending on software solution rather than hardware solution that requires high cost and complicated design.

Several planar catadioptric sensors have been designed and implemented. 3D reconstruction of object had been obtained by using single planar mirror and pinhole camera, disparity map of the scene calculated by using mirror reflection of a sense with single camera and standard pixel matching algorithm, geometry and calibration of catadioptric system also explained in [3]. Systematic study of the imaging geometry of planar catadioptric stereo vision sensors presented and multiple view geometry for the state of static and moving cameras proposed in [4]. Image rectification technique was applied to map the epipolar lines in the original image into the horizontally aligned lines in the rectified image and computed the rotation angles of the single camera corresponding to the planar mirror [5].Gaussian noise is added to the image corners of the chessboard to verify the performance of the established stereo system from single camera and two planar mirror[6]. In this paper binocular vision hardware module has been implemented using one camera and two planar mirrors to imitate human vision. The proposed vision algorithm has been implemented and tested with real stereo images to illustrate its performance, and a comparison with traditional stereo cameras will be introduced.

2. System Design

The first step in hardware design of a catadioptric system is choosing the type and number of mirrors used. Since the aim of the work is to decrease the complexity of the system, while maintaining low cost and high speed in execution time, this design uses inexpensive multiple planar mirrors with dimensions (20 cm × 10 cm) facing the camera at distance (d1=20 cm) from left mirror and (d2=10 cm) from right mirror as shown in Figure (1), the position of object P is 1m from camera.

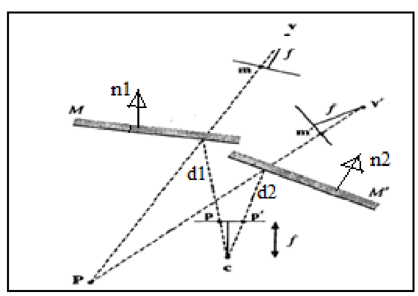


Figure (1): Stereo image formation with a single camera and two planar mirrors.

ISSN 2229-5518

Selection the angle between mirrors has an impact on the field of view (FOV) and the accuracy of disparity map. Small angle making object observed only in one mirror and this don't providing two views, while large angle appear part of object and this not suitable in matching, so the practical acceptable angle between mirrors is (175°> α >155°) and the optimal angle in this system is 165° as shown in Figure (2).

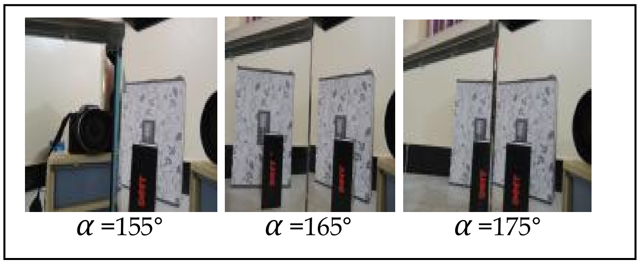


Figure (2): Different angles between mirrors

By mirror reflections of the scene, the reflected object from two planar mirrors and apparent by single camera is equivalent to two corresponding virtual cameras where the orientation and the focal length of these two virtual cameras is the same as the real camera. The relation between real camera and virtual cameras can be expressed by:

$$v = D^{(1)}c \qquad (1)$$

$$v` = D^{(2)} \qquad (2)$$

$$\text{where} \quad D^{(1)} = \begin{bmatrix} I - 2d_1 n_1{}^T & 2d_1 n_1 \\ 0^T & 1 \end{bmatrix} \qquad (3)$$

$$D^{(2)} = \begin{bmatrix} I - 2d_2 n_2{}^T & 2d_2 n_2 \\ 0^T & 1 \end{bmatrix} \qquad (4)$$

Where ($D^{(1)}, D^{(2)}$) are reflection transformation of two mirrors respectively, (n_2, n_2) are the normal vectors of the two mirror planes, (d_1, d_2) the distance from the real camera to the mirror planes.

The software design of proposed system as shown in Figure (3) consist of applying two stages. The first stage is the initial preparing stage, it used as initial process of capturing the image to compute intrinsic and extrinsic parameters of virtual cameras. While the second stage is running the catadioptric system, implementing when capturing each image, including rectification images to simplified matching process, computing disparity map by using Sobel Adaptive Block Matching Algorithm (SABMA) and finally by the triangulation the application is achieve.

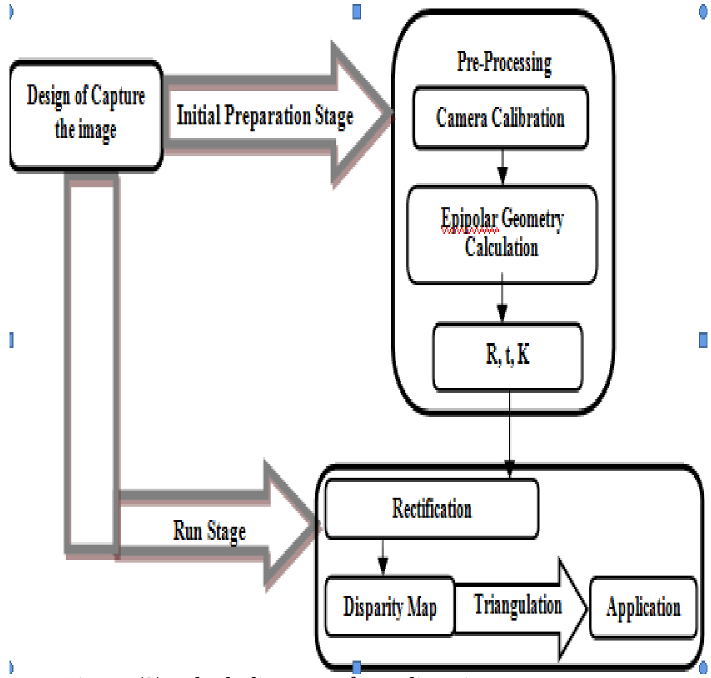


Figure (3): Block diagram of catadioptric stereo system.

3 Initial Preparation Stage

This stage is applying two processes before implementing image rectification, the first process is extract intrinsic camera parameters by camera calibration process and the second process is extract extrinsic parameters by planar motion depending on model of camera used. This stage was applied only one time at beginning capture the first image for display disparity map.

3.1 Intrinsic camera parameters

It is a necessary process in 3D computer vision in order to determine the relationship between the 3D position of a point in the world and the 2D pixel coordinates of its image in the camera which described in perspective transformation matrix. Using the camera calibration Toolbox[8] obtained the intrinsic parameters that define the perspective transformation from 3D object coordinates in the camera world coordinate system to the 2D camera image coordinate system as:

$$K = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (5)$$

Where f_x , f_y (pixels) denote the focal length of the camera along the x and y directions, s is a skew factor, (u_0, v_0) (pixels) are the principal point coordinates.

ISSN 2229-5518

In this paper used 9 images of planar pattern taken from different orientations, this chess board pattern consist of 7x5 squares each of which has the size of 30x30 millimeters. The result of ATLAB intrinsic parameters for the camera is shown in Table (1).

Real Camera		
Focal length	Horizontal	α = 850.300(pixel)
	Vertical	β = 821.537(pixel)
Principal Point	Horizontal	u_0 = 634.23 (pixel)
	Vertical	v_0 = 453.44(pixel)
Skew	s =0.0000 (angle of pixel axes = 90 degrees)	
Distortion	**kc** =[-0.0000 0.0105 0.0072 0.0329 0.0000]	

Table (1):Camera intrinsic parameters

3.2 The Extrinsic parameters of virtual cameras

The results obtained from implementing the first previous stage changing the system to new virtual form consist of two cameras instead of single camera and two planar mirrors. This stage will deal with this new virtual system constantly.

Extrinsic parameters $[R|t]$ define the transformation from the 3D object world coordinate system to the 3D camera world coordinate system by a rotation matrix R and a translation vector t:

$$[R|t] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \quad (6)$$

The relationship between two virtual cameras defined by epipolar geometry which encoded by the fundamental matrix (3×3) from planar motion, as shown in Figure (4), for uncalibrated camera and the essential matrix (3×3) for calibrated camera.

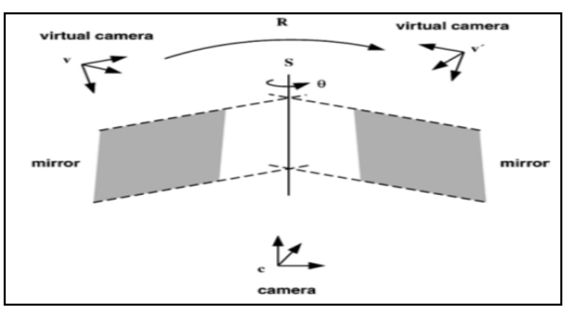


Figure (4) planar motion limitation

These two planes intersection at screw axis S which is remain stationary if mirrors are rotating and had been identical in two mirrors(m), as indicated in Figure (5), due to all epipolar lines have to intersect on m, so the virtual cameras are related by a pure rotation about S axis and this limited the motion between virtual camera to lie in the plane perpendicular to S. This limitation reducing the DOF of fundamental matrix from 7 in traditional stereo cameras to 6 in catadioptric systems because additional limitation arises from planar motion is the determinate of symmetric part of fundamental matrix is zero.

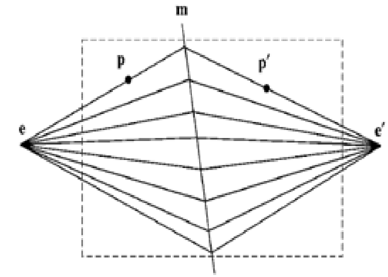


Figure (5) The epipolar geometry associated with planar motion

To calculate fundamental matrix from planar motion first needing to estimate initial fundamental matrix, equation (7), by normalized eight point algorithm[9] . By selecting 20 corresponding points in 3D world from left image and right image as in Figure (6) to obtain the homogenous coordinates for these points that used in algorithm to estimate initial fundamental matrix displayed in table (2).

$$p'^T F p = 0 \quad (7)$$

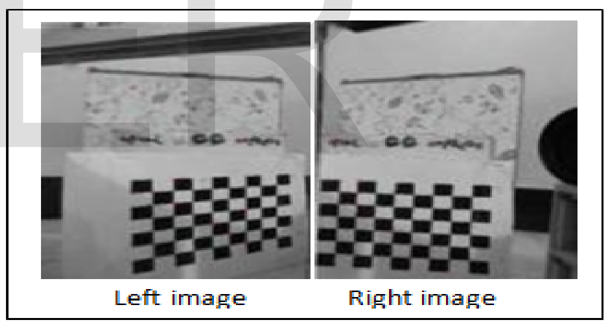


Figure (6) Selecting corresponding points in two images

ISSN 2229-5518

Points from left image			Points from right image		
X	Y	Z	X	Y	Z
53	86	1	28	110	1
102	113	1	66	128	1
156	122	1	113	131	1
260	105	1	230	93	1
80	186	1	9	195	1
99	200	1	23	208	1
221	204	1	138	210	1
264	194	1	199	201	1
143	270	1	47	274	1
122	293	1	27	293	1
204	284	1	110	297	1
144	312	1	46	315	1
238	301	1	153	321	1
165	331	1	66	337	1
223	323	1	131	340	1
187	347	1	86	359	1
204	364	1	108	381	1
124	406	1	22	397	1
222	379	1	127	406	1
204	393	1	148	428	1

Table (2) Corresponding points in two images

Every 20 corresponding points in left and right images give one linear equation in the unknown entries of f then write unknown parameters of fundamental matrix f into vector and the rest into matrix of equations:

$$\begin{bmatrix} xx' & xy' & x & yx' & yy' & y & x' & y' & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{20}x_{20}' & x_{20}y_{20}' & x_{20} & y_{20}x_{20}' & y_{20}y_{20}' & y_{20} & x_{20}' & y_{20}' & 1 \end{bmatrix} f = 0 \quad (8)$$

$$A\,f = 0 \quad (9)$$

From the least singular value of A, initial fundamental matrix is constructed.

$$A = UDV^T \quad (10)$$

$$F_{initial} = [\,0.0000 \;\; -0.0000 \;\; -0.0001$$
$$-0.0000 \;\; 0.0000 \;\; 0.0118$$
$$0.0030 \;\; -0.0087 \;\; -0.7946\,]$$

The left epipole e obtained from the third column of V corresponding to the null singular value of $F_{initial}$ while right epipole e' from the third column of U corresponding to the null singular value of $F_{initial}$.

$F_{initial}$.

$$F_{initial} = UDV^T \quad (11)$$

e =[-0.9652 -0.2616 -0.0008]
e` =[-0.9993 0.0378 0.0006]

From positive and negative eigenvalues (λ_1, λ_2) and eigenvector (n_1, n_2) of symmetric part of $F_{initial}$, m was calculated as[7]:

Where $F_s = (F + F^T)/2$ (12)

$$F_s = \begin{bmatrix} 0.0000 & -0.0000 & 0.0015 \\ -0.0000 & 0.0000 & 0.0016 \\ 0.0015 & 0.0016 & -0.7946 \end{bmatrix}$$

$$m = \sqrt{\lambda_1}\, n_1 + \sqrt{-\lambda_2}\, n_2 \quad (13)$$
if $\lambda_1 > 0$

or

$$m = \sqrt{\lambda_1}\, n_1 - \sqrt{-\lambda_2}\, n_2 \quad (14)$$
if $\lambda_1 < 0$

m = [0.0039 -0.0002 -0.8914]

From e, e\`, m, fundamental matrix from planar motion had been calculated.

$$F = [e']_x\, [m]_x\, [e]_x \quad (15)$$

F = [0.0000 -0.0000 -0.0000
-0.0000 0.0000 0.0042
0.0010 -0.0030 -0.2655]

Multiplying each point in left image by the calculated F, obtaining the corresponding epipolar line in right image as indicated in Figure (7)to facility matching process.

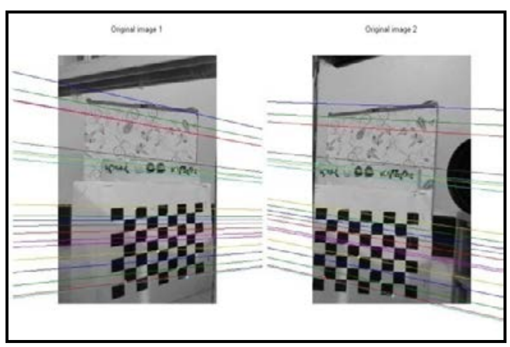


Figure (7) Epipolar lines

Extrinsic parameters of virtual cameras can be derived from the essential matrix when intrinsic parameters K are known, (equation 5), and identical for two virtual cameras as:

$$E = K'^T\, F\, K \quad (16)$$

E = [0.0000 0.0000 0.0000

ISSN 2229-5518

0.0000 0.0000 3.4505
0.8503 -2.4646 0.9129]

If assuming the first virtual camera matrix is P = [I| 0], the second virtual camera matrix, P′=[R| t], can be computed from decomposition of E into the product SR of a skew symmetric matrix and a rotation matrix as[4] :

$$E = [t]_x R = SR \quad (17)$$

$[t]_x$ is asymmetric matrix defined as:

$T \times R = [t]_x R$ and $\times$ denotes the cross product

By using SVD: $S = UZU^T$ $R = UWV^T$ or UW^TV^T

where $W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

The result of extrinsic parameters for virtual cameras see in table (3) which indicate the rotation matrix and translation vector of right virtual camera with respect to left virtual camera. It denote that the right virtual camera is about 844.17 to the right, 37.48 above, and 298.32 behind of the left virtual camera.

Table (3): The result of virtual cameras extrinsic parameters.

	Virtual Left camera		Virtual Right camera
Rotation matrix	$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	Rotation matrix	$R = \begin{bmatrix} 0.4944 & -0.0603 & 0.8671 \\ 0.0080 & 0.9979 & 0.0649 \\ -0.8692 & -0.0252 & 0.4939 \end{bmatrix}$
Translation vector	$t = [0.00 \quad 0.00 \quad 0.00]$	Translation vector	$t = [-844.17 \quad -37.48 \quad 298.32]$

5. Image Rectification

The rectification process applying the warping each image plane to ensure that the corresponding epipolar lines become parallel to the image axes and lie on the same scanline depending on parameters obtained from previous section. The important advantage of rectification is to reduce a 2D search problem to a 1D search problem as shown in Figure (8).

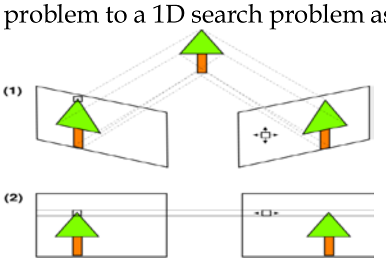


Figure (8): The rectification process.

Rectification process needed to produced two new Perspective Projection Matrices (PPMs) for two virtual cameras from rotating the old ones around their optical centers until focal planes become coplanar. The matrix R, which gives the virtual camera's pose, is the same for both new PPMs. It will be specified by means of its row vectors:

$$R = \begin{bmatrix} r_1^T \\ r_2^T \\ r_3^T \end{bmatrix} \quad (18)$$

By applying Fusiello's algorithm [10] [11]., the row vectors of R denoted to X, Y, and Z axes and can be calculated as :

1) The new X axis is parallel to the baseline and denote the direction of the epipole , this ensures that epipolar lines will be parallel in the rectified:

$$r_1 = \frac{t}{||t||} \quad (19)$$

where t translation between two virtual cameras.

2) The new Y axis is orthogonal to X and obtained as the cross product of t and the original left optical axis $[0, 0, 1]^T$ followed by a normalization to unit length:

$$r_2 = \frac{1}{\sqrt{t_x^2 + t_y^2}}[-t_y^2, t_x^2, 0]^T \quad (20)$$

3) The new Z axis orthogonal to XY:

$$r_3 = r_1 \times r_2 \,. \quad (21)$$

The rectification on real right and left images shown in Figure (9).

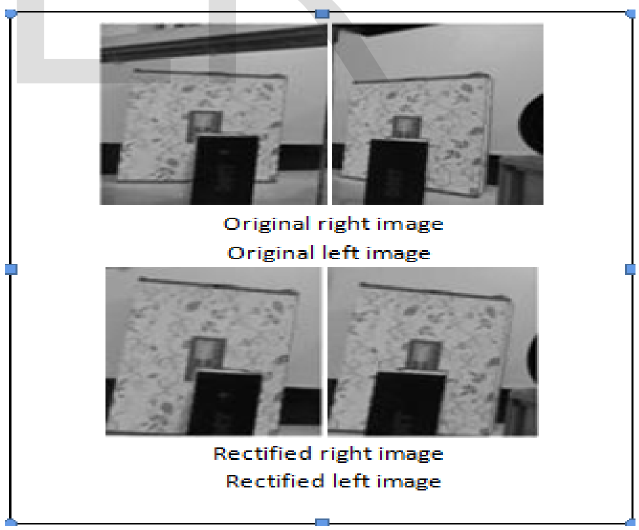


Figure (9) Rectification of real images.

6. Disparity Map

The catadioptric systems provide two images to the same scene at the same time with few differences between them shifting one image over top of the other image within a specific search range called disparity range to find corresponding points, the amount of shifting between two corresponding points called "disparity" which is depend on object distance. The set of disparities of all points called "disparity map" [12].

ISSN 2229-5518

The key step to obtain accurate depth information is by finding a detailed and accurate disparity map. To minimize computations, most real time systems use image brightness as a measure of similarity. The higher disparity of object pixel means that the object is closer to the camera and appears brighter in disparity map, while the less disparity means the object is far from the camera and appears darker [13].

The standard catadioptric stereo system based on pixel-by-pixel (full search) with matching cost function Sum-of-Absolute Difference (SAD), equation (22), which is one of the simplest of similarity measures, keeping the data size small and fast in implementation [7].

$$SAD(x,y) = \sum_{j=0}^{n} \sum_{i=0}^{m} \left| I_L\,(x+i, y+ j) - I_R\,(x+i\,, y+ j+d) \right| \quad (22)$$

where I_L, I_R are the intensity of pixel in left and right images respectively, m× n is the window size, d is the disparity value. The block matching algorithm SMA will implemented by MATLAB.

To reduce the error and speed up the calculations of the disparity map in catadioptric stereo this paper proposed **adaptive window** with **edge filter detection (Sobel filter)** to discriminate the borders of the objects in the scene in order to calculate the true disparity value for each object, so this method gives a **great reduction** in execution time.

7. Implementation and Experimental Results

The stereo vision hardware used in this paper, is implemented using two planar mirrors and single camera, connected to PC (P4, core i3, 2.10 GHz processor and 4GB RAM) with USB (Universal Serial Bus) cable.

After computing the fundamental matrix and extract R matrix and t vector, applying these parameters in rectification transform using the method described previously in section (5), this transform is used to warp each incoming image at run time then applying the proposed algorithm of adaptive block matching with [9x9] as large window size and [3x3] as sub window size to produce disparity map as shown in Figure (11) .

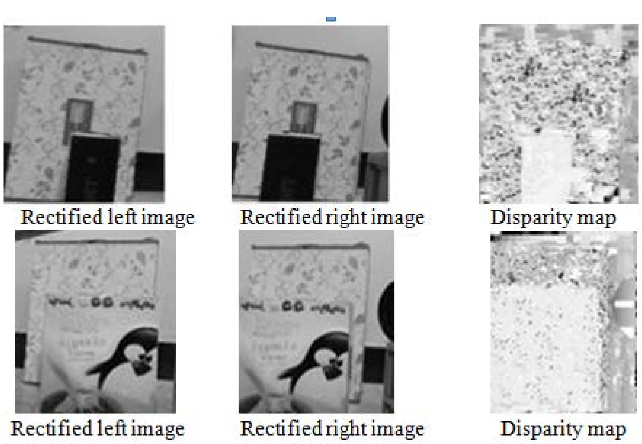
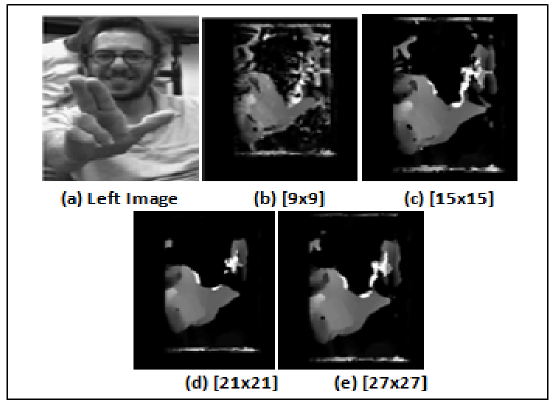
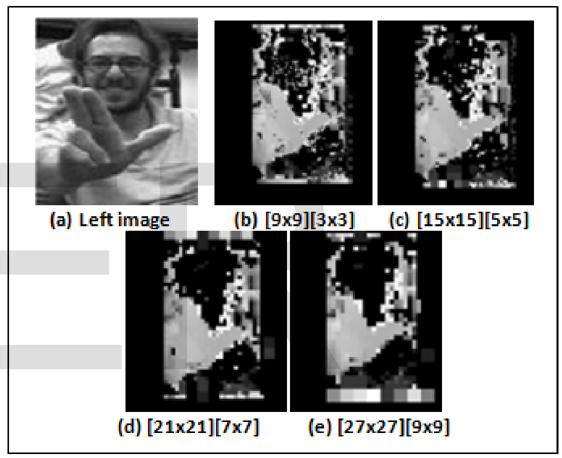


Figure (10): Disparity map for real images

The standard matching algorithm (SMA) and Sobel Adaptive Block Matching (SABMA) implemented by MATLAB for different block size is shown in Figure (11).



(A) SMA



(B) SABMA

Figure (11): Disparity map by different algorithms

A reduction in execution time of SABMA as compared with SMA shown in Table (4) and Figure (12), this reduction because the searching by Block-by-Block is speed up as compared at searching by Pixel-by-Pixel.

Table (4): The execution time comparison of image pair between (SMA) and (SABMA).

Large block size	Execution time (Sec.) of SMA	Execution time (Sec.) of SABMA	Execution time Percentage of SABMA with respect to SMA
[9×9]	18.18	0.24	1.320%
[15×15]	36.87	0.20	0.54 %
[21×21]	60.03	0.18	0.29 %
[27×27]	87.74	0.17	0.19 %

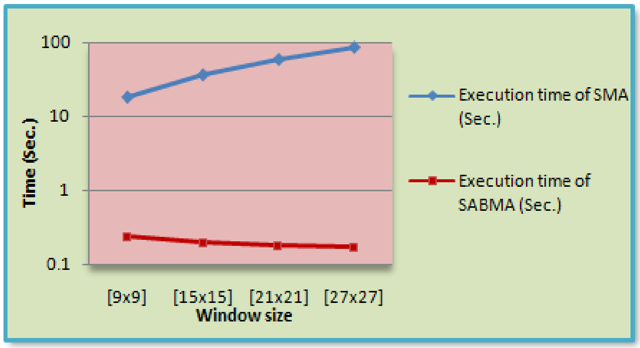


Figure (12): The execution time comparison of image pair between (SMA) and (SABMA).

8. Conclusions

The main important points that are concluded from this work are:

The minimization of the angles between mirrors(α) decrease the field of view, so the object is not observed in two mirrors because the relationship between shared field of view of two virtual cameras and angle between mirrors, where shared field of view = 2 α.

The device used to capture stereo images is also significant as the matching algorithm therefore we need to elaborate the quality of image while maintaining low computational cost.

Sobel filter is used to decrease the execution time because knowing the border of object is enough to recognize it without need to object's details but the value of threshold associated with error and accuracy of the disparity map, when using high value (Threshold>30), the time needing for execution decrease noticeably but more error also occur because more edges removed may be belong to other object and this caused not dividing the block matching, also when using low value (Threshold<10), the time needing for execution increase and the disparity map appear noisy because multi matching process.

References:

[1] Bo Hu, "Cast Shadows in Augmented Reality Systems", PhD thesis, Department of Computer Science, The College Arts and Sciences, University of Rochester, Rochester, New York, pp 43-52, 2005.

[2] Gian Luca Mariottinia, Stefano Scheggib , Fabio Morbidia a, Domenico Prattichizzob, "Planar mirrors for image-based robot localization and 3-D reconstruction", journal, a Department of Computer Science and Engineering, University of Texas at Arlington, 500 UTA oulevard, Arlington, TX 76019, USA. b Department of Information Engineering, University of Siena, Via Roma 56, 53100 Siena, Italy, 2011.

[3] Joshua Gluckman and Shree K. Nayar, "A Real-Time Catadioptric Stereo System Using Planar Mirrors", Department of Computer Science, Columbia University, New York, NY 10027, 1999.

[4] Gian Luca Mariottini, Stefano Scheggi, Fabio Morbidi and Domenico Prattichizzo, "Catadioptric Stereo with Planar Mirrors: Multiple-View Geometry and Camera Localization", Springer, 2010.

[5] Lingfeng XU, Oscar C. AU, Wenxiu SUN, Yujun LI, Sung Him CHUI, Chun Wing KWOK, "Image Rectification For Single Camera Stereo System", IEEE, Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, 2011.

[6] Yu Yan, Bingwei He, "Single Camera Stereo with Planar Mirrors", School of Mechanical Engineering, Fuzhou University,Fuzhou 350108, P.R. China, 2013.

[7] J. Gluckman and S.K. Nayar., "Catadioptric Stereo using Planar Mirrors", Int. J. Computer Vision, Department of Computer Science, Columbia University, New York, NY 10027,44(1):65–79, 2001.

[8] Bouguet J-Y., "Camera Calibration Toolbox for MATLAB", Available online: *http://www.Vision.Caltech.edu/bouguetj/Calib-doc/index.html*, 2010.

[9] Joshua Gluckman and Shree K. Nayar, "Planar Catadioptric Stereo: Geometry and Calibration", Department of Computer Science, Columbia University, New York, NY 10027, 1999.

[10] Fusiello A., Trucco E., and Verri A., "A Compact Algorithm for Rectification of Stereo Pairs", Machine Vision and Applications, Springer-Verlag, pp. 16-22, 2000.

[11] Chiang M., Lin H., and Hou C., "Development of a Stereo Vision Measurement System for 3D Three-Axial Pneumatic Parallel Mechanism Robot Arm", Article in Sensors Journal, doi: 10.3390/s110202257, pp. 2257-2281, 2011.

[12] Vatansever M., "3D Reconstruction Using a Spherical Spiral Scan Camera", M.Sc. Thesis, Computer Engineering, Izmir Institute of Technology, Izmir, 2006.

[13] Khaleghi B., Ahuja S., and Jonathan Wu Q. M., "An Improved Miniaturized Embeded Stereo Vision System (MESVS-II)", IEEE, Computer Vision and Systems Laboratory, 2008.

[14] Saba Ali Sadeq, "Stereo Vision for 3D Measurement in Robot Systems ", Msc. Thesis, Electrical Eng. Dept., AL-Mustansiriyah univ., 2012.