# Fast Window Based Stereo Matching for 3D Scene Reconstruction

Mohammad Mozammel Chowdhury and Mohammad AL-Amin Bhuiyan
Department of Computer Science and Engineering, Jahangirnagar University, Bangladesh

**Abstract:** *Stereo correspondence matching is a key problem in many applications like computer and robotic vision to determine Three-Dimensional (3D) depth information of objects which is essential for 3D reconstruction. This paper presents a 3D reconstruction technique with a fast stereo correspondence matching which is robust in tackling additive noise. The additive noise is eliminated using a fuzzy filtering technique. Experimentations with ground truth images prove the effectiveness of the proposed algorithms.*

**Keywords:** *Stereo correspondence, disparity, window cost, stereo vision, fuzzy filtering, 3D model.*

*Received February 21, 2010; accepted October 24, 2010; published online March 1, 2012*

## 1. Introduction

In computer and robot vision, stereo correspondence matching plays an important role. With the help of stereo correspondence algorithms stereo vision systems can be automatized. Stereo correspondence techniques are useful for machine vision applications to determine Three-Dimensional (3D) depth information of objects. They are required in applications like 3D reconstruction, robot navigation and control, autonomous vehicles, 3D growth monitoring, stereo endoscopy, and so on.

Several stereo correspondence algorithms have been developed to find matching pairs from two image sequences [1, 2, 3, 7]. But they exhibit very high computational cost. The extremely long computational time required to match stereo images is the main obstacle on the way to the practical application of stereo vision systems. However, stereo vision applications are often subject to strict real-time requirements. In applications, such as robotics, where the environment being modeled is continuously changing, these operations must also be fast to allow a continuous update of the matching set, from which 3D information is extracted.

Approaches to the stereo correspondence matching can be broadly classified into two categories: 1). Feature-based, 2). Intensity-based matching techniques. In the feature-based approaches, features such as, edge elements, corners, line segments, curve segments etc., are first extracted from the images, and then the matching process is applied to the attributes associated with the detected features. The main problem of this approach is that edge elements and corners are easy to detect, but may suffer from occlusion; line and curve segments require extra computation time, but are more robust against occlusion. In the intensity-based matching, the matching process is applied directly to the intensity profiles of the two images. In this method, as the two cameras are placed on the same horizontal baseline, the corresponding points in both images must lie on the same horizontal scanline. Such stereo configurations reduce the search for correspondences from two dimensions (the entire image) to one-dimension. Unfortunately, like all constrained optimization problems, whether the system would converge to the global minima is still an open problem.

An alternative approach in intensity-based stereo matching, commonly known as the window-based method, where matching is established on those regions in the images that are interesting; for instance, regions that contain high variation of intensity values in the horizontal, vertical, and diagonal directions. After the interesting regions are detected, a simple correlation scheme is applied in the matching process; a match is assigned to regions that are highly correlated in the two images. The problem associated with this window-based approach is that the size of the correlation windows must be carefully chosen. If the correlation windows are too small, the intensity variation in the windows will not be distinctive enough, and many false matches may result. If they are too large, resolution is lost, since neighboring image regions with different disparities will be combined in the measurement [11].

To overcome the limitations of these stereo correspondence algorithms this research proposes a fast window based stereo correspondence method for 3D scene reconstruction. In addition, a major task in the field of stereo vision is to extract information from stereo images corrupted by noise. For this purpose, a fuzzy filtering technique has been implemented.

This paper is organized as follows: Section 2 describes the basic principle of stereo correspondance estimation and 3D reconstruction. Section 3 then describes our proposed method for noise filtering. Section 4 presents the process of determining stereo disparity. The algorithm for calculating window cost and disparity is illustrated in section 5. In section 6, we have presented our proposed model for 3D scene reconstruction. Section 7 shows the experimental results and discussions. Finally, section 8 concludes the paper.

## 2. Stereo Correspondence and 3D Reconstruction

The overall stereo vision process for 3D reconstruction is illustrated in Figure 1.

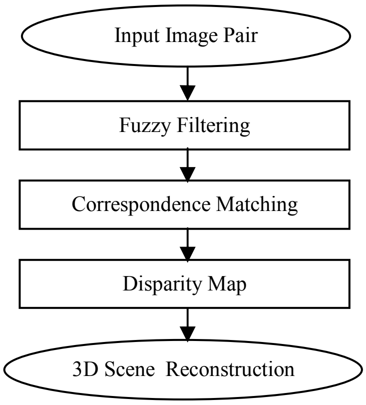



Figure 1. Fundamental steps employed in stereo vision.

There are two problems associated with stereo vision:

1. The correspondence problem-in which features corresponding to the same entities in 3D are to be matched across the image frames.
2. The reconstruction problem-in which 3D information is to be reconstructed from the correspondences.

In stereo vision, two images of the same scene are taken from slightly different viewpoints using two cameras, placed in the same lateral plane. For most pixels in the left image there is a corresponding pixel in the right image in the same horizontal line. Finding the same points in two images is called correspondence matching and is the fundamental computation task underlying stereo vision. The difference in the coordinates of the corresponding pixels is known as disparity, which is inversely proportional to the distance of the objects from the camera. Using the disparity, the depth can be defined by the following equation:

$$z = \frac{bf}{d} \tag{1}$$

where, $d$ is the disparity, $z$ is the distance of the object point from the camera (the depth), $b$ is the base distance between the left and right cameras, and $f$ is the focal length of the camera lens. Figure 2 shows that the two images of an object are obtained by the left and right cameras observing a common scene. This pair of stereo images allows us to obtain the 3D information about the object. Once we have obtained a distance map of the scene, we can then measure the shape and volume of objects [12, 13].

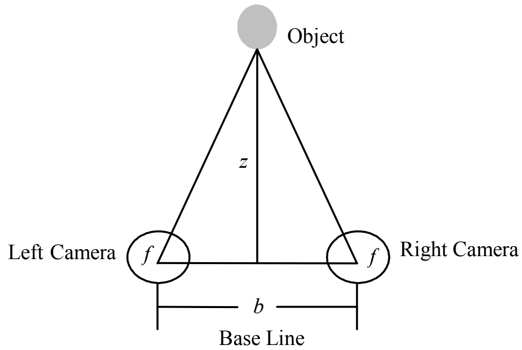



Figure 2. Stereo vision using two cameras placed in the same lateral plane.

## 3. Noise Elimination with Fuzzy Filtering

To extract information from stereo images corrupted by noise, we have employed a special fuzzy filtering method. This filter employs fuzzy rules for deciding the gray level of a pixel within a window in the image. This is a variation of the Median filter and Neighborhood Averaging filter with fuzzy values.

The algorithm of fuzzy filtering includes the following steps:

1. The gray values of the neighborhood pixels ($n \times n$ window) are stored in an array and then sorted in ascending or descending order.
2. Fuzzy membership value is assigned for each neighbor pixels. This step has the following characteristics:
   a. A Π-shaped membership function is used.
   b. The highest and lowest gray values get the membership value 0.
   c. Membership value 1 is assigned to the mean value of the gray levels of the neighborhood pixels.
3. We consider only $2 \times k+1$ pixels ($k/2<=n^2$) in the sorted pixels, and they are the median gray value and $k$ previous and forward gray values in the sorted list.
4. The gray value that has the highest membership value will be selected and placed as output.

Figure 3 shows a test image with additive noise and the corresponding output image after Fuzzy filtering. The memberships function for the gray levels in the test image is shown in Figure 4.

a) Original image.

b) Noised image.

c) Fuzzy filtered image.

Figure 3. Test image, noisy image and its corresponding output image after fuzzy filtering (using a 3×3 mask).

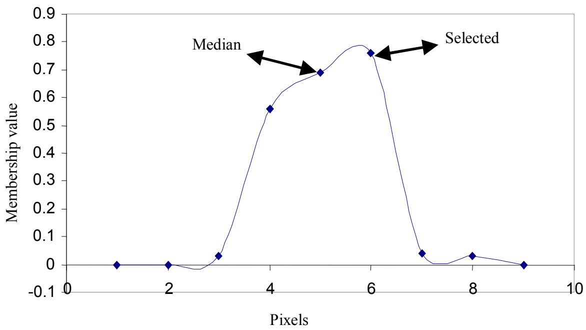



Figure 4. Memberships function for the gray levels in the test image.

## 4. Estimation of Stereo Disparity

Stereo correspondence or disparity is conventionally determined based on matching windows of pixels by using Sum of Square Differences (SSD), Sum of Absolute Differences (SAD), or normalized correlation techniques [4, 5, 6, 15]. To determine the correspondence of a pixel in the left image using SSD, SAD or normalized correlation techniques, the window costs are computed for all candidate pixels in the right image within the search range. The pixel in the right image that gives the best window cost is the corresponding pixel of the left image.

The basics of stereo correspondence matching are as follows:

*For each epipolar line*
*For each pixel in the left image*
*compare with every pixel on same epipolar line in right image*
*pick pixel with minimum match* cost

Window-based stereo matching technique is widely used due to its efficiency and ease of implementation. However, Barnard and Fishler [2] point out a problem in the selection of a window with fixed size and shape. Many researchers proposed adaptive window methods using windows of different shapes and size depending on local variations of intensity and disparity [10, 14]. But in adaptive window algorithms, the computation time is relatively higher than that of fixed window algorithms [8, 9, 16, 17]. To overcome this problem and to achieve a substantial gain in accuracy with less expense of computation time we have proposed a fast and very simple algorithm. Some applications, like autonomous vehicle and robot navigation, virtual reality and stereo image coding in 3D-TV, require a very fast estimation of dense stereo correspondence. It is aimed that this pruning proposal will be useful in such situations for speedy determination of dense disparity.

## 5. Window Cost and Disparity Estimation Algorithm

With a view to estimate dense disparity, for every pixel in the left image our goal is to find the corresponding pixel in the right image. We assume that the stereo images are rectified, which means that the corresponding epipolar lines are horizontal and on the same height. In window-based algorithms, to determine the correspondence of a pixel in the left image the window costs, which are the SSD or the SAD or the normalized correlation values, need to be computed for all candidate pixels in the right image within the search range by the following equation.

$$W_C(x,y,d)=\begin{cases}\sum_{i=1}^{w_x}\sum_{j=1}^{w_y}\left[f_L(x+i,y+j)-f_R(x+i+d,y+j)\right]^2, & for\ \ SSD\\ \sum_{i=1}^{w_x}\sum_{j=1}^{w_y}\left|f_L(x+i,y+j)-f_R(x+i+d,y+j)\right|, & for\ \ SAD\\ \dfrac{\sum_{i=1}^{w_x}\sum_{j=1}^{w_y}f_L(x+i,y+j)f_R(x+i+d,y+j)}{\sqrt{\sum_{i=1}^{w_x}\sum_{j=1}^{w_y}f_L^2(x+i,y+j)\sum_{i=1}^{w_x}\sum_{j=1}^{w_y}f_R^2(x+i+d,y+j)}}, & for\ \ Correlation\end{cases} \quad (2)$$

where $f_L(x, y)$ and $f_R(x, y)$ are the intensities of the pixel at a position $(x, y)$ the left and right images, respectively, $W_C(x, y, d)$ is the window cost of a pixel at position $(x, y)$ in the left image with disparity $d$, $w_x$ and $w_y$ are the window width and height, respectively.

The pixel in the right image that gives the best window cost, i.e., the minimum SSD or SAD value or the maximum correlation value indicates the corresponding pixel of the pixel in the left image. The computation time for disparity estimation of a pixel depends on image size and disparity search range. In direct search, it requires to compute the window costs for all candidate pixels within the search range, $-d_{max}$ to $+ d_{max}$. The conventional disparity estimation algorithm using direct search is described bellow:

*For each pixel (x, y) in the left image,*
*For* $d'=-d_{max}$ *to* $+d_{max}$ *do*
*Calculate* $W_c(x,y,d')$.
*End For*
*Find best* $W_c(x,y,d) \in W_c(x,y,d')$.
*Disparity of* $(x,y)=d$.

A fast technique for disparity estimation has been implemented by excluding unlikely correspondences. This pruning mechanism is based on the stereo matching constraint that the corresponding pixels should be close in color or intensity. To determine the correspondence of a pixel in the left image we just compute the window cost for candidate pixels in the right image within the search range if their intensity differences with respect to the pixel of the left image are less than a threshold value, $\delta$. The proposed pruning algorithm includes the following steps:

*For each pixel (x, y) in the left image,*

1. *For* $d'=-d_{max}$ *to* $+d_{max}$ *do*
   1.1. *If* $|f_L(x,y)-f(x+d',y)|<threshold$ *then*
   *Calculate* $W_C(x,y,d')$.
2. *Find best* $W_C(x,y,d) \in W_C(x,y,d')$.
3. *Disparity of (x,y)=d.*

Estimating the corresponding points, we can compute the disparity map. The disparity map then can be converted to a 3D map of the scene (i.e., recover the 3D structure) if the stereo geometry is known.

## 6. Reconstruction of 3D Scene

The estimated stereo disparity fields can be converted into dense depth information by camera geometry. As a result, the 3D model of the real scene is reconstructed from the original images and disparity information. To reconstruct the 3D image we have to compute the *z* coordinate of each point in the left camera frame. For a 3D point *(x, y, z)* we take *x* and *y* from the first stereo image and *z* is computed from the gray level intensity at *(x, y)* of the depth map. Figure 5 illustrates the stereo imaging which involves in obtaining two separate image views of an object point *P*. The distance between the centers of the two camera lenses *b* is called the base line. Our objective is to find the coordinates $(x_p, y_p, z_p)$ of the point *P* having image points $(x_L, y_L)$ and $(x_R, y_R)$ in the left and right camera frame, respectively.

Let's recover the position of *P* from its projections $P_L$ and $P_R$:

$$z_p = \frac{bf}{x_L - x_R} \tag{3}$$

$$x_p = \frac{b.x_L}{x_L - x_R} \tag{4}$$

$$y_p = \frac{b.y_L}{x_L - x_R} = \frac{b.y_R}{x_L - x_R} \tag{5}$$

$x_L - x_R = d$ is the disparity (i.e., the distance between the corresponding points in the two images when the images are superimposed). Therefore,

$$x_p = \frac{b.x_L}{d},\ y_p = \frac{b.y_L}{d}, z_p = \frac{bf}{d} \tag{6}$$

Using equation 6, we can easily obtain the 3D coordinates of the point-cloud of the object, which will help to reconstruct the 3D scene.

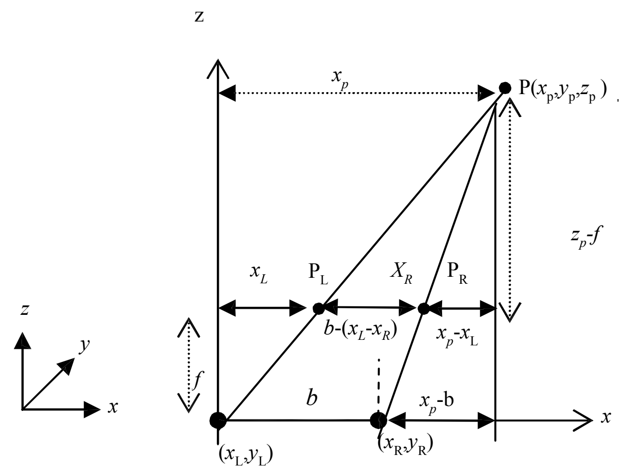



Figure 5. Model of the stereo imaging technique.

## 7. Results and Discussions

We have reconstructed the 3D scene from the test stereo images with a fast and robust stereo correspondence matching technique. The effectiveness of the stereo correspondence algorithm has been justified for different types of images of different resolutions with simple and complex background.

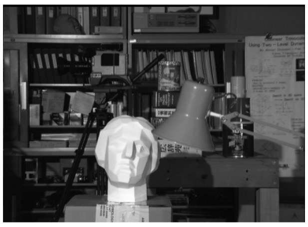
In order to demonstrate the effectiveness of the algorithm, we present the processing results from synthetic and real image pairs, including ones with ground-truth values for quantitative comparison with other methods. Experiments were carried out on a Pentium IV 2.1GHz PC with 512MB RAM. The algorithm has been implemented using Visual C++. The dense disparity map is obtained using some standard stereo images (Tsukuba Head) which dense ground truth is known. Figure 6 illustrates the gray scale stereo images including the ground truth, and estimated disparity image. The ground truth image is histogram equalized for visualization purpose. The reconstructed 3D scene is shown in Figure 7.

Disparities of the left image are estimated by SSD method without and with considering a threshold value for different window sizes. Table 1 summarizes the obtained disparity estimation results for the Tsukuba Head image using SSD method. The accuracies represent the percentage of correct disparities (i.e., same value as that of ground truth). It is also found that a threshold value of $\delta$*=30* is a good choice on the basis of a trade-off between accuracy and computation time. Figure 8 illustrates the runtime snapshot of disparity estimation results. The computational time versus window size graph is shown in Figure 9 which reveals that the computational cost increases with the size of the window. Therefore, a window of size 3×3 performs better results than any other window size, although some variable window size has been used for reducing the computational cost for better disparity estimation.

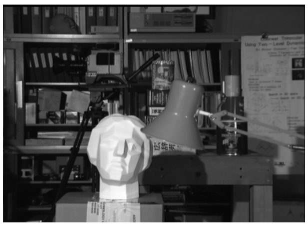
Table 1. Disparity estimation accuracy (in %) for different threshold values.

| Threshold Value (δ) | Window Size | Correct Matching Accuracy (%) | Reduction of Computation Time (%) |
|---|---|---|---|
| δ=0 | 7×7 | 67.0 | 0 |
| | 11×11 | 72.3 | |
| δ=20 | 7×7 | 65.9 | 44 |
| | 11×11 | 70.1 | |
| δ=25 | 7×7 | 66.1 | 42 |
| | 11×11 | 70.5 | |
| δ=30 | 7×7 | 66. 3 | 39 |
| | 11×11 | 71.9 | |

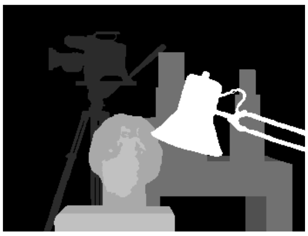
Table 2 shows the comparison of our proposed method with some other methods [1, 2, 3, 4, 5]. A window of size 7×7 size is applied for all methods. Experimental results predict that our proposed method provides a reduction of 45% of computational time and increase of 20% accuracy of correct matching.



a) Left image.



b) Right image.



c) Ground truth image.

d) Estimated disparity image.

Figure 6. A standard stereo image pair (Tsukuba Head) with ground truth.

We use a simple method for the reconstruction of the 3D scene of the test image pairs from the estimated 3D coordinates and original image information. The focal length of the stereo cameras used in this simulation is 35mm, the baseline distance is 15cm and pixel size is 0.1165mm. Table 3 shows the parameters used in the reconstruction process.

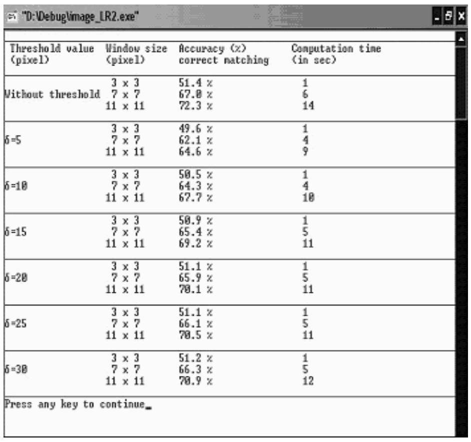
Figure 7. Reconstructed 3D image.

Table 2. Computational time reduction (%) and increase in correct matching with proposed pruning method.

| Disparity Estimation Method | Window Size | Computation Time Reduction(%) | Correct Matching Accuracy Increase (%) |
|---|---|---|---|
| Conventional Window Based Method [1-3] | 7×7 | 0 | 0 |
| Adaptive Window Method [4] | 7×7 | 21 | 13 |
| Average Window Method [5] | 7×7 | 33 | 17 |
| Proposed Method | 7×7 | 45 | 20 |

Table 3. Parameters used in 3D reconstruction.

| Parameter | Values |
|---|---|
| Window Size | 3×3 |
| Threshold level | δ=30 |
| Search range for disparity estimation | K=2 |
| Focal length of the cameras | F=35mm |
| Base distance | B=15cm |



"D:\Debug\image_LR2.exe"

| Threshold value (pixel) | Window size (pixel) | Accuracy (%) correct matching | Computation time (in sec) |
|---|---|---|---|
| Without threshold | 3 x 3 | 51.4 % | 1 |
| | 7 x 7 | 67.0 % | 6 |
| | 11 x 11 | 72.3 % | 14 |
| δ=5 | 3 x 3 | 49.6 % | 1 |
| | 7 x 7 | 62.1 % | 4 |
| | 11 x 11 | 64.6 % | 9 |
| δ=10 | 3 x 3 | 50.5 % | 1 |
| | 7 x 7 | 64.3 % | 4 |
| | 11 x 11 | 67.7 % | 10 |
| δ=15 | 3 x 3 | 50.9 % | 1 |
| | 7 x 7 | 65.4 % | 5 |
| | 11 x 11 | 69.2 % | 11 |
| δ=20 | 3 x 3 | 51.1 % | 1 |
| | 7 x 7 | 65.9 % | 5 |
| | 11 x 11 | 70.1 % | 11 |
| δ=25 | 3 x 3 | 51.1 % | 1 |
| | 7 x 7 | 66.1 % | 5 |
| | 11 x 11 | 70.5 % | 11 |
| δ=30 | 3 x 3 | 51.2 % | 1 |
| | 7 x 7 | 66.3 % | 5 |
| | 11 x 11 | 70.9 % | 12 |

Press any key to continue_

Figure 8. Runtime snapshot of disparity estimation accuracy (in %) measurements.

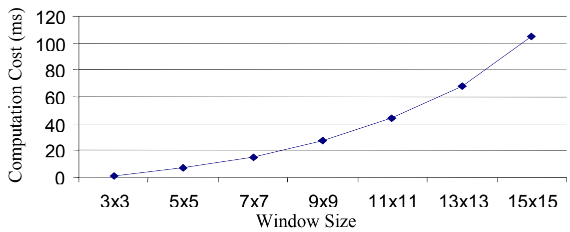




Figure 9. Computational cost vs. window size.

## 8. Conclusions

A 3D reconstruction technique from stereo image pair is proposed. We have reconstructed 3D information of real images from the estimated disparity fields and camera parameters. A fast stereo vision system has been developed that analyzes grayscale or color images to estimate the disparity map for 3D scene reconstruction. To deal with additive noise, a robust and novel approach capable of fast estimation of stereo correspondence is proposed. A fuzzy rule-based filtering is employed for noise elimination which can remove the non interesting points from the stereo images. A total of about 20 images are used to investigate the effectiveness of the proposed algorithms. Among them only a few points in the images are found false. Experimental results demonstrate that the success rate of more than 95% is achieved. The main reason behind the failure of those images in finding disparity image is the mismatch of epipolar geometry. The main target of this research is to build a robust and fast stereo vision system for real-time applications. Our next plan is to develop a complete 3D model from multi-view images.

## References


[1] Barnard S., "Stereo Matching by Hierarchical, Microcanonical Annealing," *in Proceedings of the International Joint Conference on Artificial Intelligence*, Italy, pp. 832-835, 1987.

[2] Barnard S. and Fishler M., "Stereo Vision," *in Proceedings of Encyclopedia of Artificial Intelligence*, New York, pp. 1083-1090, 1987.

[3] Calin G. and Roda V., "Real-Time Disparity Map Extraction: A Dual Head Stereo Vision System," *Latin American Applied Research*, vol. 37, no. 1, pp. 21-24 2007.

[4] Chowdhury M., Mondal A., and Bhuiyan M., "3D Imaging using Stereo Vision," *in Proceedings of the 7th International Conference on Computer and Information Technology*, pp. 307-312, 2004.

[5] Chowdhury M. and Bhuiyan M., "A New Approach for Disparity Map Determination," *Daffodil International University Journal of Science and Technology*, vol. 4, no. 1, pp. 9-13, 2009.

[6] Fua P., "A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features," *Machine Vision and Applications*, vol. 6, no. 1, pp. 35-49, 1993.

[7] Fusiello A., Trucco E., and Roberto V., "Efficient Stereo with Multiple Windowing," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, USA, pp. 858-863, 1997.

[8] Hannah M., "A System for Digital Stereo Image Matching," *Photogrammetric Engineering and Remote Sensing*, vol. 55, no. 12, pp. 1765-1770, 1989.

[9] Hoff W. and Ahuja N., "Surface From Stereo: Integrating Feature Matching, Disparity Estimation and Contour Detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 11, no. 2, pp. 121-136, 1989.

[10] Hu T. and Huang M., "A New Stereo Matching Algorithm for Binocular Vision," *International Journal of Hybrid Information Technology*, vol. 3, no. 1, pp. 1-8, 2010.

[11] Intille S. and Bobick A., "Disparity-Space Images and Large Occlusion Stereo," *in Proceedings of the European Conference on Computer Vision*, Sweden, pp. 179-186, 1994.

[12] Kanade T. and Okutomi M., "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 16, no. 9, pp. 920-932, 1994.

[13] Lmaati E., Oirrak A., Kaddioui M., Ouahman A., and Sadgal M., "3D Model Retrieval Based on 3D Discrete Cosine Transform," *The International Arab Journal of Information Technology*, vol. 7, no. 3, pp. 264-270, 2010.

[14] Matondang M. and Haron H., "Hybrid Computing Algorithm in Representing Solid Model," *The International Arab Journal of Information Technology*, vol. 7, no. 4, pp. 411-419, 2010.

[15] Muhlmann K., Maier D., Hesser J., and Manner R., "Calculating Dense Disparity Maps from Color Stereo Images an Efficient Implementation," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, pp. 30-36, 2001.

[16] Uddin M., Chowdhury M., Shioyama T., and Mondal M., "Fast Window Based Approach for Stereo Matching," *Journal of Science*, vol. 27, pp. 145-154, 2004.

[17] Veksler O., "Stereo Matching by Compact Windows Via Minimum Ratio Cycle," *in Proceedings of the IEEE International Conference on Computer Vision*, Vancouver, pp. 540-547, 2001.


**Mohammad Mozammel Chowdhury** is working as an Associate Professor in the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. He received his BSc degree in Electronics and Computer Science and MS degree in Computer Science and Engineering from the same university. He has published more than 25 articles in reputed journals and conference proceedings. His research interest includes: image processing, computer vision, robotics, 3d scene reconstruction, computer graphics and machine intelligence.

**Mohammad AL-Amin Bhuiyan** is a Professor in the Department of Computer Science and Engineering, Jahangirnagar University, Savar, Dhaka, Bangladesh. He got the PhD degree in Electrical engineering from Osaka City University, Japan, he did the post doctoral research from National Institute of Informatics, Japan and University of Hull, UK. His research interest includes: robotics, image processing, computer vision, face recognition, human-robot symbiosis.