

A Hybridized Disparity Computation Fusing Disparity Space Image and Multi Resolution Image Segmentation

Mudasar Qadir, Youngsung Soh, Hadi Ashraf, and Intaek Kim

Abstract—3D information is getting more importance nowadays. There are several ways to get 3D information. One way to get 3D information involves the laser based depth estimators, which is a very costly process. Other way is to extract 3D information from 2D stereo image pairs by using stereo matching techniques. Stereo matching is a vast area of research. It can be classified into area (window) based approach, feature based approach, and optimization based approach. Area based approach generally generates dense disparity map with low accuracy and low computation time. Feature based method produces highly accurate and sparse disparity map with low computation time. Optimization based method produces dense disparity map with high accuracy and high execution time. By keeping these things in mind, we proposed a new hybrid way for disparity computation. The proposed method consists of three steps. They are disparity estimation, image segmentation, and disparity refinement. Since the ultimate goal of stereo matching is to obtain dense disparity map with high accuracy and low execution time, we select optimization based approach for disparity estimation step and for image segmentation step we select multi resolution image segmentation. At the end, disparity refinement is done by combining the result of both the previous steps. As there are several optimization techniques, we choose disparity space image (DSI) based dynamic programming (DP). We tested the proposed method on several stereo pairs and found that method produced reasonably good quality results.

Index Terms—Stereo matching, DSI, DP.

I. INTRODUCTION

Stereo vision is extensively used for getting disparity map and hence the depth information from stereo pairs. Stereo pairs are captured by using two stereo cameras arranged adjacently, where both cameras produce much similar 2D image pairs of the scene having some horizontal shift. Stereo vision technique helps in extracting 3D information from 2D information taken by stereo cameras, while other way to get 3D information involves the laser based depth estimators. 3D information is beneficial as it decreases the problem and restrictions associated with 2D.

Stereo matching plays a vital role in stereo vision. Its steps involve by taking one image as left and the other as right image to match the pixels from left image to right image. If the stereo image pair is already rectified, the horizontal distance between two matched pixels is termed as disparity; which then be expressed into depth.

Manuscript received December 5, 2013; revised February 8, 2014.

The authors are with the Myongji University, Yongin, 449-728, Korea (e-mail: mudasar.kalwar@gmail.com, sohan111@gmail.com, nothan111@gmail.com, kit@mju.ac.kr, soh@mju.ac.kr).

Categories of the stereo matching methods are; area (window) based methods, feature based methods, and optimization based methods. In area based methods usually matching is done by using fixed areas instead of features. In feature based methods, various features such as edges, lines, and points are used. The proposed methods for feature based matching include spatial relations or invariant descriptors, relaxation method, and wavelets. In optimization based methods, global optimization methods such as DP [1], energy minimization [2], and graph algorithms [3] are applied to compute and aggregate the matching cost.

Area based methods take less execution time to generate less accurate and dense disparity map, while feature based methods generally take less execution time to produce sparse disparity map with high accuracy, and optimization based methods generate highly accurate and dense disparity map while taking high execution time. Since the ultimate objective of stereo matching is to produce highly accurate and dense disparity map while taking less execution time, we choose hybrid method to compute the disparity where we combine DSI and the segmentation results produced from multi resolution image segmentation. Unlike previous methods we are producing dense disparity map with high accuracy.

This paper is organized as follows. Related work is presented in Section II while Section III explains the proposed method. Results are described in Section IV and Section V concludes the paper.

II. RELATED WORK

This section describes several DSI based stereo matching methods.

Bobick *et al.* [4] introduced a new method based on DSI in the existence of extensive occlusion. They reduced the sensitivity to occlusion cost and algorithm complexity by strong matching of feature points. They used ground control points that are helpful to get disparity jumps to avert transitory smoothness across depth discontinuities. They also used intensity edges as intensity edges coincide with depth edges in several cases. Their method only deals with parallel surfaces, besides that, their method generate fairly good results, but lack behind in real time processing.

Tsai *et al.* [5] introduced divide and conquer DSI method by mixing up the feature and optimization based methods. DSI based method is joined with feature based method to generate dense disparity map, since feature based method usually produces sparse disparity map. They partitioned each row of image into parts by looking for prominent features; then each part is conquered by using DSI based method. Hence robust disparity is guided by

prominent features and DSI based method generates dense disparity map between features. This method works well for low textured area as strong features are used but with high execution time.

Otha *et al.* [6] acquired disparity map by using DP based method for inter-scan line search and intra-scan line search. To match the connected edges, inter-scan line matching was used; while for matching every edge given the connected edges, intra-scan line was used. Thus the method worked well for the short connected edges, but it took long processing time for those images that contain long connected edges.

Baha *et al.* [7] proposed DSI based DP method, which involved DSI construction and disparity map refinement. For DSI construction, Recalculation of the sum of aggregate cost for each pixel inside the window is avoided by using adaptive windows. They showed the comparison of many traditional feature based, window based and optimization based methods. Their algorithm was taking low execution time as compared to others, but producing many matching errors at depth boundaries.

Kim *et al.* [1] proposed DSI based method where they reduced the execution time for cost matrix calculation by using varying search regions. They did full search to compute cost matrix at the line in the center and restricted search within a few lines up and down for all other scan lines. In this way computation time of cost matrix for DP is reduced. They suggested various neighborhood configurations for cost computation of each element in DSI. Beside all the efforts to reduce the execution time, still their method is not suitable for real time applications as, for instances, computing disparity on 512×512 stereo image pair takes almost 9 seconds.

Tao *et al.* [8] proposed color segmentation based stereo vision. Their main emphasis was on achieving the goals in obtaining reliable depth at depth boundaries and obtaining correct depth in textureless regions. They computed the depth using global matching approach. For textureless regions, color segmentation based approach was used where they assume that the same color region corresponds to same depth.

III. THE PROPOSED METHOD

As stated in Section I, we propose a hybrid method consisting of three steps. They are disparity estimation, image segmentation, and disparity refinement. For disparity estimation we choose DSI based DP method that consists of DSI construction, DP, and disparity computation. In image segmentation, we use multi resolution image segmentation. In the end we propose a novel way to refine the disparity by fusing the results of both disparity estimation and image segmentation. As disparity estimation and image segmentation steps are independent so they can be performed independently. The proposed method assumes that images are already rectified. Block diagram of the proposed method is shown in Fig. 1. Detailed description of each step is given below.

A. Disparity Estimation

Mehmood *et al.* [9] gave a good review of DSI based DP method. The first step is the construction of DSI in

which each corresponding row of left image and right image is used to construct DSI. For $M \times N$ image, N DSI's of size $M \times M$ are constructed. In second step, DP consisting of two parts is performed. In the first part, matching cost is aggregated to compute cost matrix. In the second part, backtracking is performed to search for the minimum cost path. The third step of the method computes disparities from each DSI and makes a whole disparity map.

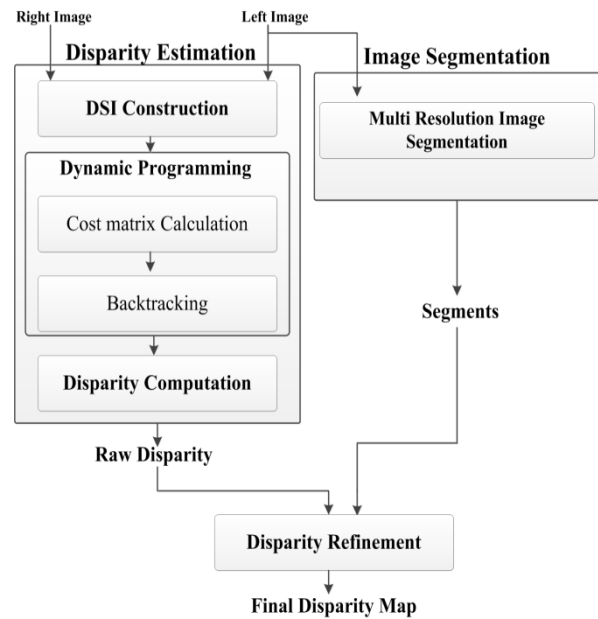


Fig. 1. Block diagram of the proposed method.

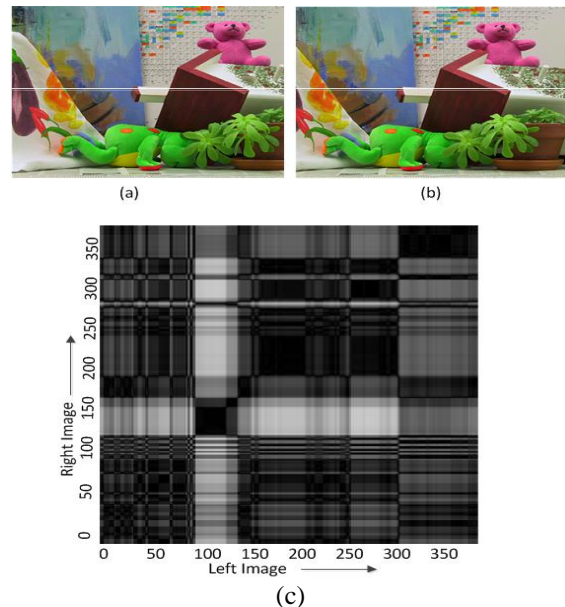


Fig. 2. DSI of one scan line for stereo pair (a) left image, (b) right image, and (c) constructed DSI.

DSI represents the matching error between the row of left image and the corresponding row of right image and can be obtained as in (1),

$$DSI_n(x_L, x_R) = |I_L(x_L, n) - I_R(x_R, n)| \quad (1)$$

where $I_L(x_L, n)$ is the pixel value of x_L position of the n^{th} scan line in the left image and $I_R(x_R, n)$ is that of the right image. The process of making DSI for one scan line is shown in Fig. 2. Here a specific row indicated in white is

used to construct DSI. We use the convention that one specific row from left image is arranged horizontally from left to right and corresponding row from right image is arranged vertically from bottom to top to compute the absolute differences of matching element pairs. So lower left corner becomes the DSI origin.

DP is used to find the minimum cost path. The first stage of finding the minimum cost path is to calculate the cost matrix C . We compute the cost of each element as in (2).

$$C(x_L, x_R) = \min\{C(x_L-1, x_R), C(x_L-1, x_R-1), C(x_L, x_R-1)\} + DSI(x_L, x_R) \quad (2)$$

where $C(x_L, x_R)$ is the cost at location (x_L, x_R) and $DSI(x_L, x_R)$ is the DSI value at the same location. Three immediate previous neighbors of location (x_L, x_R) are used to compute the cost $C(x_L, x_R)$ as shown in Fig. 3. The neighbor with lowest cost is selected and is added to DSI value to get the final cost. Here we do not consider occlusion cost to remove the effect of its heuristic features. Along with the computation of cost matrix, direction information at every point explaining where the lowest value comes from is also stored in direction matrix D for backtracking.

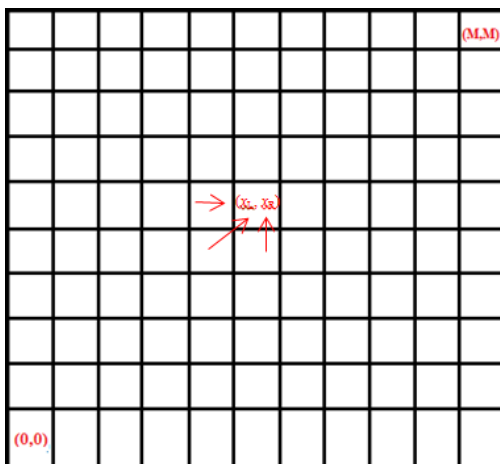


Fig. 3. Computation of cost matrix.

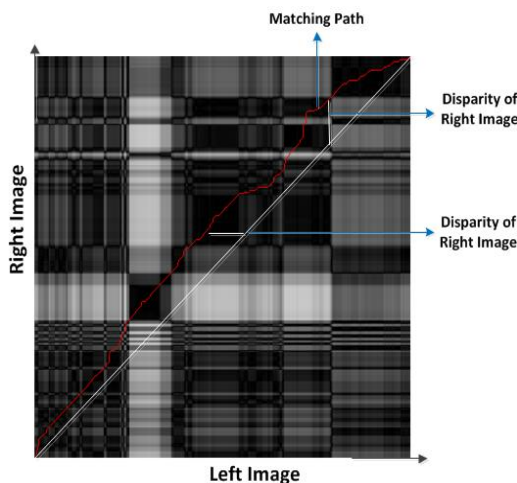


Fig. 4. DSI image with matching path and disparity

Direction matrix D was stored along with the computation of cost matrix. We find matching path by tracking back the saved direction from upper right corner to lower left corner of D . Fig. 4 depicts this step. The white line from upper

right to lower left is called the zero disparity line. Matching path is shown in red. The vertical distance between the zero disparity line and matching path is the disparity of left image. And the horizontal distance between the zero disparity line and matching path is the disparity of right image.

B. Image Segmentation

In this step, left image is divided into the segments. Segmentation is a mechanism of dividing the image into regions, having more or less homogenous properties. Burt *et al.* [10] proposed an iterative multi resolution linking method to find the connected components. They used linked pyramid structure for getting efficient computational properties. Their method is based on feature computation and segmentation. They adopted two computational strategies. In the first, they used multiple level of resolution to perform all processing; the advantage of this to ensure that processing is performed for any given portion of given image. In the second, iterative manner is used for segmentation and image properties reoccupation; this ensures that after completion of all the iterations the final segments of image are obtained accurately.

There are three main steps in their algorithm, which are computation of the Gaussian pyramid, segmentation by pyramid linking, and linked pixels averaging. Steps two and three are of iterative nature and are performed until a stable segmentation outcome of image is reached.

C. Disparity Refinement

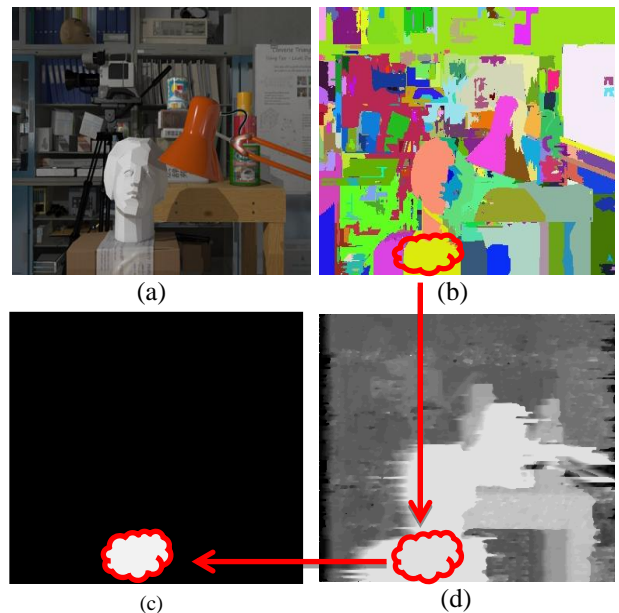


Fig. 5. Disparity Refinement (a) original image, (b) segmented image, (c) new disparity map, and (d) raw disparity map.

In this step we utilize the result of disparity estimation and image segmentation. From disparity estimation step we get the raw disparity map which may have disparity noise especially at the boundaries of the objects. DSI based method is not able to generate disparity jumps at depth boundaries since the computation of cost matrix does not allow big jumps and this results in smooth transition of disparities at depth discontinuities. To overcome this deficiency, we make use of image segmentation results. As

in [8] we assume that all points inside the same segment are experiencing the same depth. Thus our refinement step is that, for each segment, we sum all raw disparity values in the segment, take average, and use it as disparity values for all the points inside the segment. Fig. 5 illustrates this process. Fig. 5(a) and (b) show original image and segmented image respectively. Fig. 5(c) and (d) are part of new disparity map and raw disparity map produced from DSI based DP method respectively. In this figure we just choose to show the refinement process for one segment highlighted with red boarder. First we get the pixel locations of a segment shown in Fig. 5(b), sum up the values all the corresponding pixels in raw disparity map, compute the average of summation, and assign it to same location pixels in new disparity map. This process is applied to every segment present in the segmented image; at the end new disparity map will be turned into refined disparity map.

IV. EXPERIMENTAL RESULTS

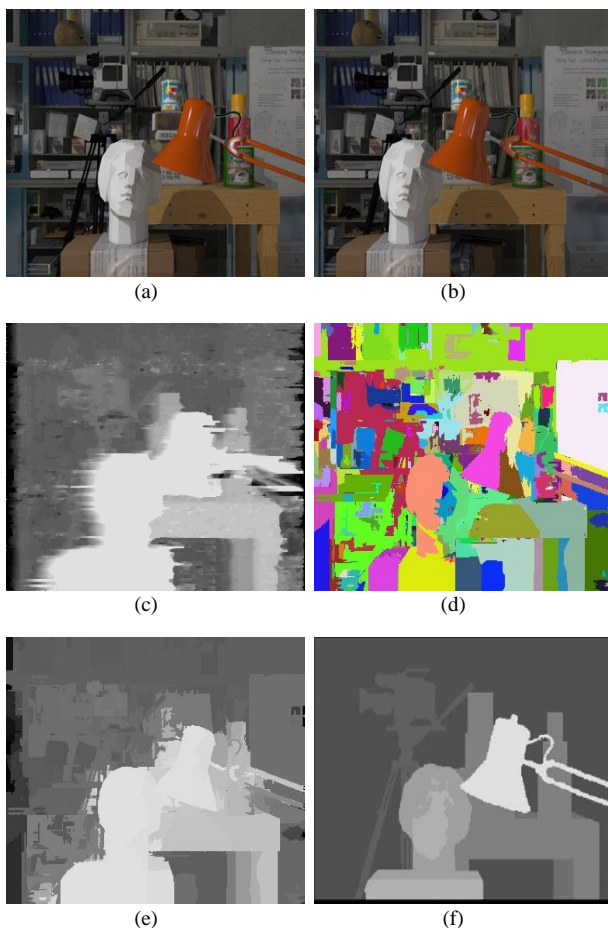


Fig. 6. Disparity results for Tsukuba pair (a) left image, (b) right image, (c) raw disparity, (d) segmented colored image, (e) final disparity after refinement, and (f) ground truth.

In this section we compare the proposed method with other conventional methods. Test image pairs are from Middlebury stereo vision page. We run the proposed algorithm for different stereo images. We used OpenCV function to segment the image. OpenCV function has one parameter for maximum level of parameters and other for error threshold. For the segment clustering, both of these parameters show much influence on number of segments.

Therefore both of the parameters should be selected very carefully, to produce good disparity results. Fig. 6 shows the disparity results produced by the proposed algorithm. The brighter the value, the closer the object is to the camera. Fig. 6(a) and (b) are left and right images of Tsukuba pair respectively. Fig. 6(c) and (d) are the raw disparity computed by DSI based DP method and segmented image respectively. We colored the segmented image in order to show the segments. Fig. 6(e) and (f) shows the final disparity result after refinement and ground truth respectively. As can be seen in raw disparity and final disparity results, lamp, human sculpture, desk, canes behind lamp, and camera shows reasonably good depth (disparity). Fig. 6(f), the refined version, exhibits better results where disparity noise is almost completely removed. In refined disparity clear borders of lamp, lamp stand, sculpture, and desk can be seen clearly while in unrefined disparity, noise on borders of objects are observed.

Similar Observations can be made from Fig. 7. Fig. 7(a) and (b) are left and right images of Teddy pair respectively. Fig. 7(c) and (d) are raw disparity computed by DSI based DP based method and colored segmented image respectively. Segmented image is colored to show the segments. Fig. 7(e) and (f) shows the final disparity results after refinement and ground truth respectively.

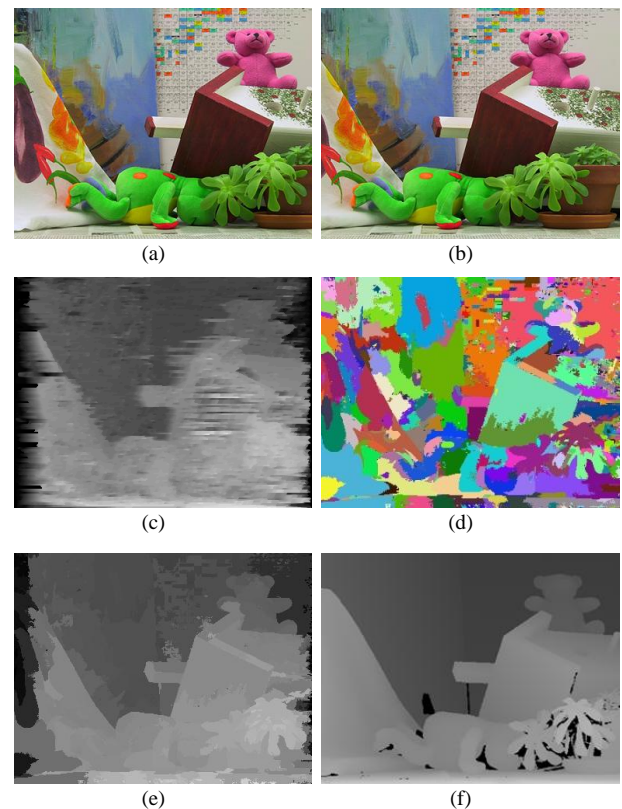


Fig. 7. Disparity results for Teddy pair (a) left image, (b) right image, (c) raw disparity, (d) segmented colored image, (e) final disparity after refinement, and (f) ground truth.

V. DISCUSSION

Stereo matching is a necessary step in stereo vision, which is a useful way to obtain depth information. We proposed a new hybrid method consisting of three steps, disparity estimation, image segmentation, and disparity refinement. For disparity estimation we select the DSI based

DP method and for image segmentation we select multi resolution image segmentation method. DSI based DP method consists of three parts. They are DSI construction, DP, and disparity computation. We suggested the new way to refine disparity by utilizing the results produced from disparity estimation and image segmentation results. As the disparity estimation and image segmentation parts can be run in parallel, our method is not putting any overhead for execution time.

Though the proposed method is not taking any extra time and is showing reasonably good results in terms of dense disparity, disparities at some places of the image are not accurate enough. We need to have more precise way of obtaining segments. In this work we assume parallel surfaces whereas, in real life, there may be non-parallel (or even curved) surfaces. We need to devise a way to overcome this limitation and this is intended for further research.

ACKNOWLEDGMENT

This work (Grants No. C0005448) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded by Korea Small and Medium Business Administration in 2012.

REFERENCES

- [1] C. H. Kim, H. K. Lee, and Y. H. HA, "Disparity Space Image based stereo matching using optimal path searching," *Image and Video Communications and Processing*, vol. 5022, pp. 752-760, 2003.
- [2] J. Sun, Y. Li, S. B. Kang, and H. Y. Shum, "Symmetric stereo matching for occlusion handling" in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, vol. 2, pp. 399-406.
- [3] O. Veksler, "Extracting dense features for visual correspondence with graph cuts", in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp.689-694, 2003.
- [4] A. F. Bobick, "Large occlusion stereo," *Int'l J. Computer Vision*, vol. 33, no. 3, pp. 181-200, 1999.
- [5] C. J. Tsai and A. K. Katsaggelos, "Dense disparity estimation with a divide-and-conquer disparity space image technique," *IEEE Trans. Multimedia*, vol. 1, pp. 18-29, Mar. 1999.
- [6] Y. Ohta and T. Kanade, "Stereo by intra- and inter- scanline search using dynamic programming," *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 7, Mar. 1985.
- [7] N. Baha and S. Larabi, "Real-Time disparity map computation based on disparity space image," in *Proc. the 13th International Arab conference on Information technology*, 2012, pp. 10-13.
- [8] H. Tao and H. S. Sawhney, "Global matching criterion and color segmentation based stereo," *IEEE Works on Applications of Computer Vision*, pp. 246-253, 2000.

- [9] A. Mahmood, Y. Soh, and I. Kim, "A new parallel implementation of DSIBased disparity computation using CUDA," in *Proc. 3rd Int'l conference on Information Computer Application*, 2014.
- [10] P. J. Burt and T. H. Hong, and A. Rosenfeld, "Segmentation and estimation of image region properties through cooperative hierarchical computation," *IEEE Tran. on SMC*, vol. 11, no.12, pp. 802-809, 1981.



Mudasar Qadir was born in Ghotki, Pakistan on Dec. 24, 1989. He got BS in computer science in 2011 from FAST-NU University in Peshawar, Pakistan. He entered a master course in information and communication engineering in Myongji University in 2013.

He worked in different software companies as a software engineer from Jul. 2011 to Feb. 2013.

His current interest of research includes object tracking, stereo vision and parallel algorithms for image processing.



Youngsung Soh was born in Seoul, Korea on Mar. 4, 1956. He got BS in electrical engineering in 1978 from Seoul National University in Seoul, Korea. He obtained MS and PhD in computer science from the University of South Carolina in Columbia, South Carolina, USA in 1986 and 1989, respectively.

He served in the Korean army from June 1980 to Sept. 1982. He worked in Systems Engineering Research Institute in Korea as a senior researcher from Sept. 1989 to Feb. 1991. He joined Myongji University in Korea from Mar. 1991 and is currently a full professor in the Dept. of Information and Communication Engineering.

His current interest of research includes object tracking, stereo vision, and parallel algorithms for image processing.

Prof. Soh is a member of Korea Information Processing Society and Korea Signal Processing Systems Society.



Hadi Ashraf was born in Lahore, Pakistan on Dec. 8, 1988. He got BS in electrical engineering in 2010 from Govt. University in Lahore, Pakistan. He entered a master course in information and communication engineering in Myongji University in 2012.

He joined a software company in Pakistan in Aug. 2010 and worked as a software engineer till April 2012.

His current interest of research includes object tracking, stereo vision and parallel algorithms for image processing.



Intaek Kim was born in Seoul, Korea in 1960. He received BS and MS in electronics engineering from Seoul National University in Seoul, Korea in 1980 and 1984 respectively. He obtained PhD in electrical engineering from Georgia Institute of Technology in Atlanta, Georgia, USA in 1992.

He worked for Goldstar central research lab from 1993 to 1995 as a senior engineer and joined Myongji University from 1995. He is now a professor in the Dept. of Information and Communication Engineering. His recent publications deal with the area of face recognition, hyperspectral image and MR imaging.

His research interest includes pattern recognition, image processing and smart grid area.

Prof. Kim is a member of Korean Institute of Electronics Engineer.