# Hierarchical stereo matching using feature extraction of an image

*Tae-June Kim*, Ji-Sang Yoo**

Department of Electronic Engineering
Kwangwoon University
Seoul, Korea
*mirrine@kw.ac.kr, *jsyoo@kw.ac.kr

## ABSTRACT

In this paper a hierarchical stereo matching algorithm based on feature extraction is proposed. The boundary (edge) as feature point in an image is first obtained by segmenting an image into red, green, blue and white regions. With the obtained boundary information, disparities are extracted by matching window on the image boundary, and the initial disparity map is generated when assigned the same disparity to neighbor pixels. The final disparity map is created with the initial disparity. The regions with the same initial disparity are classified into the regions with the same color and we search the disparity again in each region with the same color by changing block size and search range.
The experiment results are evaluated on the Middlebury data set and it show that the proposed algorithm performed better than a phase based algorithm in the sense that only about 14% of the disparities for the entire image are inaccurate in the final disparity map. Furthermore, it was verified that the boundary of each region with the same disparity was clearly distinguished.

**Keywords:** Stereo Image, Matching, Disparity

## 1. Introduction

Recently, research on 3D images processing has been active all over the world. Japan has already succeeded in commercializing 3D broadcasting from the beginning of this year by BS11. Research in this area is also being carried out actively at MPEG and standardization of the multi-view video coding (MVC) technology is completed. In addition, research on the free-viewpoint television (FTV) is also being carried out actively. The reason why large corporations, universities and research organizations in many nations are interested in this area is because the media is gradually advancing toward realistic broadcasting due to consumer demand and business demand. The goal of realism in an image is to give the user an immersive feeling as if the person is inside the image that is being viewed. While there are many ways for people to experience 3D, the representative method is the one due to binocular disparity. Because human eyes are separated by 6.5 cm on the average, it is possible to recognize the distance to an object. In order to experience 3D using binocular disparity, stereo images that can be used in 3D monitors are needed. A stereo image refers to the left and right images that correspond to the left and right eye of a human being. If a 3D image is viewed through a 3D monitor, the user will experience depth due to binocular disparity.

Two images of a stereo image have a high correlation because they are obtained from adjacent viewpoints. The process of finding such a correlation is generally called stereo matching. The result of finding similar pixels that are highly correlated between two images and marking the distance between the pixels is called a disparity map and it displays important information about correlation between the two images. Such information can be used for intermediate view reconstruction (IVR) and actual depth information can be extracted by using the camera information that was captured. In addition, it is also possible to segment specific information from the disparity map and synthesize it with a new image. The disparity map contains important meaning for a stereo image like this.

Generally, region based, feature based and probability based algorithms are used for stereo matching. Region based refers to a method in which a matching window of a specific size is used to find a corresponding matching window that has pixels with small errors in a different view image. While this is very simple and efficient, the results are not satisfactory for occluded and homogeneous regions. The results also vary according to the size of the matching window. For the feature based method, the matching point is searched with extracted features rather than with an entire image as in the region based method. A feature can be an edge or boundary in an image. The matching point obtained with a feature fills the disparity through interpolation and extrapolation. Error propagation due to such a method affects the accuracy of the disparity map. Most of the recent best matching methods use a probability based algorithm like belief propagation. This method generates a disparity map by considering the disparity of the surrounding pixels and selecting one with the optimum probability. While accuracy is high, it takes too much time due to high complexity.

In the proposed algorithm, initial disparity map is generated by applying interpolation and extrapolation to the disparity extracted with an extracted feature. The initial disparity map value and the color regions are segmented. The final disparity map is generated by detecting the occluded regions and the mismatches from the image that has been segmented according to the disparity value and the color regions by approaching the block matching in a hierarchical manner.

## 2. Proposed Algorithm

An algorithm that is proposed in this paper is composed of two steps. The first step is for generating the initial disparity map and the second step is for generating the final disparity map. Fig. 1 shows the block diagram for the stereo matching process proposed in this paper. In the initial disparity map, the disparity is spread according to autocorrelation for the entire feature-extracted image. The final disparity map is generated by preventing the propagation of errors that can occur from hierarchical matching by considering the occluded regions and mismatches.

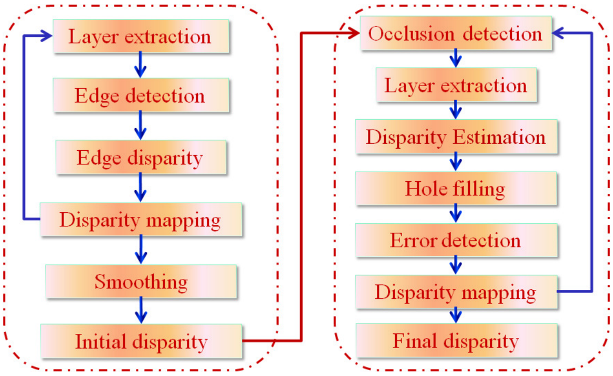


Fig. 1: The proposed stereo image matching process

### 2.1 Initial disparity map generation

There are many ways to extract the features that exist in the corner and boundary of an image. In this paper, the Sobel operator is used to extract the image boundaries that are used as the features. Fig. 2(a) and (b) shows an original image and the boundary disparity map which is extracted only on the image boundaries.

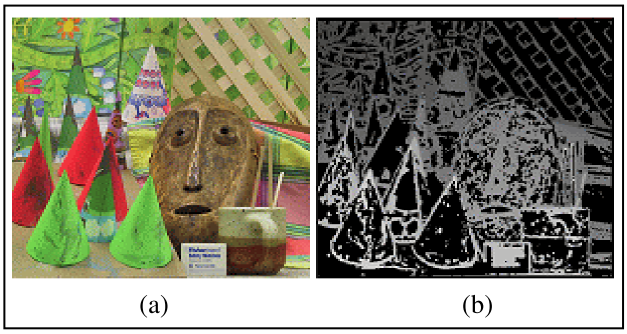


Fig. 2: (a) Original image (b) Boundary disparity map

In a homogeneous region of an image, there is relatively less information to estimate the disparity. Therefore, disparity estimated on the boundary will have a higher reliability. However, since the image boundary takes only a small part of an image, we have to find a method to estimate disparity in the surrounding regions. Since a discontinuity in disparity mostly occurs on an image boundary and since most of the disparity is similar in homogeneous regions of an image such as ones with the same color, spreading the disparity estimated around the feature points, to the surrounding neighbor pixels will result in an initial disparity map that is sufficiently reliable. If spreading is carried out in the image plane, spreading can occur in unwanted surrounding regions because the pixel value of surrounding regions is similar. To prevent this problem and to carry out spreading with high reliability, the initial disparity map is generated by segmenting the image into four components: R, G, B and white.

Fig. 3: Initial disparity map

By performing a series of processes for each channel separately, the effect from other channels is minimized. Spreading the boundary disparity map from each color region guarantees a high reliability. Even though the homogeneous color region has similar disparity value in the initial disparity map, there exist some errors. Therefore, we need a process to get the disparity values smoothed in homogeneous color regions. In this paper, an averaging filter is applied within each homogeneous region excluding boundary points. Fig. 3 shows an initial disparity map that provides initial values for region segmentation to determine the final disparity map.

### 2.2 Final disparity map generation

In order to generate a disparity map with high reliability, detection of occluded and mismatched regions is very important. The occluded regions need to be detected because if an inaccurate disparity value exists in those regions, inaccurate disparity can be assigned in the next hierarchical level. An occluded region occurs on a discontinuity between a large disparity and small disparity. Eq. (1) is the formula to detect an occluded region. The difference between two neighboring disparity values becomes the range of an occluded region

$$Length = Dis(x, y) - Dis(x+1, y) \qquad (1)$$

where Dis(x,y) is the current disparity and Dis(x+1,y) is the neighboring disparity in a disparity map.

Usually, a hole appear at those points from small disparity to large disparity and an occluded region appear in reverse, that is, from a large disparity to a small disparity. If a negative length occurs, it is not judged to be an occluded region since a hole has occurred. It is judged to be an occluded region only for a positive length. For the calculated length, an amount equal to the length from the discontinuous point for the large disparity and small disparity is judged to be the occluded region and is reset to the value of '0' on the disparity map.

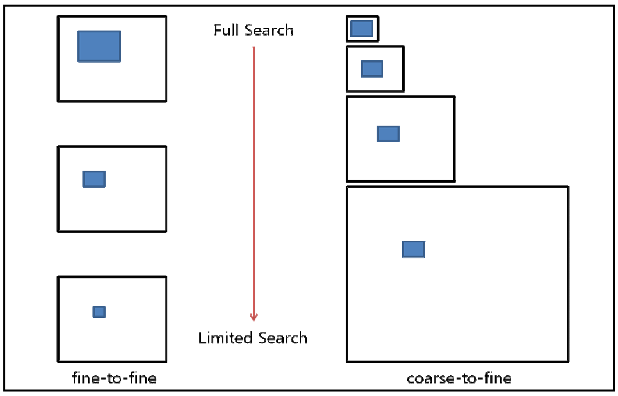

Fig. 4: A fine-to-fine and coarse-to-fine comparison

Fig. 4 shows a comparison between the fine-to-fine and coarse-to-fine methods which are representative hierarchical approach methods [1]. In the fine-to-fine method, the size of the matching window is reduced each time the hierarchy grows and the disparity is deduced while the search range is limited. In the coarse-to-fine method, the size of the matching window is not changed and the disparity is deduced while the image resolution is changed in a pyramid form. In this paper, a method similar to the fine-to-fine hierarchical approach is used to generate the final disparity map.

The hierarchical approach is not executed for occluded and mismatched regions. Therefore, after each hierarchical process is completed, holes such as occluded and mismatched regions are still remaining. An occluded region usually occurs at the boundary of an object. Therefore, a region that is determined to be an occluded region is filled with the background disparity with the assumption that it has a similar disparity to that of the background. For the mismatched regions, we can apply the same process to find their disparity.

$$CrossChecking = \; | O(x + Dis(x,y), y) - R(x,y) |$$

$$Dis(x,y)\begin{cases} Dis(x,y), CrossChecking < Tolerance \\ 0 \qquad ,else \end{cases} \tag{2}$$

where R(x,y) is a reference image pixel value, O(x,y) is an original image pixel value, and Dis(x,y) is the disparity value at (x,y).

We still need to check reliability of the obtained disparity after hole processing. Cross checking in Eq. (2) is used to check whether it is reliable or not. After Cross checking, we have uncertain disparities be judged to be holes again and assign zero disparity value to them. Since boundary regions of a given image usually contain inaccurate disparities, we judge them to be holes along with the holes generated through Cross Checking and assign zero disparity values. We regenerate a highly reliable disparity map for these holes by using Eq. (3) that is based on a Spiral Search for Spreading. This is so called, post processing in this paper. The three processes consisting of occluded region detection, hierarchical approach and post processing are all carried out at one hierarchy level.

$$\begin{cases} Valid \quad , \sum_{r,g,b} | R(x+\Delta x, y+\Delta y) - R(x,y) | < Tolerance \\ Invalid, else \end{cases} \tag{3}$$

$$(-SearchRange \le \Delta x, \Delta y \le SearchRange)$$

where R(x,y) is a reference image pixel value, $\Delta$x and $\Delta$y are the distance from x and y, respectively, r, g, and b indicate the range of red, green and blue regions with an identical disparity.

The reliability of obtained disparity map is improved as these whole processes are repeated up to the final hierarchy level.

## 3. Experiment Results

Middlebury stereo images were used as test images in this paper and their characteristics are given in Table 1.

Table 1: Experimental image size and the disparity range

| Test Image | Size | Disparity Range |
|---|---|---|
| Cones | 450x375 | 0-59 |
| Teddy | 450x375 | 0-59 |
| Tsukuba | 384x288 | 0-15 |
| Venus | 434x384 | 0-19 |

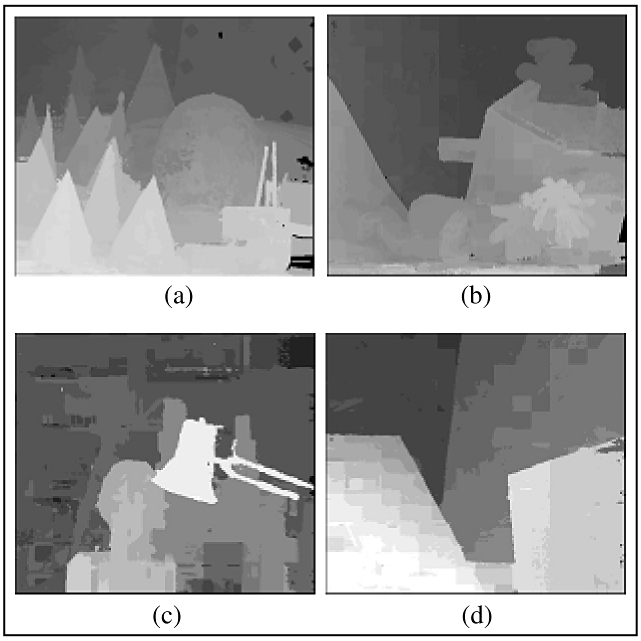
(a) (b) (c) (d)

Fig.5: The final disparity map for each experimental image

(a)Cones, (b)Teddy, (c)Tsukuba, (d)Venus

## 4. Conclusion

We evaluated the performance of the proposed algorithm on Middlebury data set. Inaccurate disparities accounted for an average of 14% of the experimental images in the four images and 3~5 small errors occurred in the comparisons with Ground truth. There was almost no difference in performance from dynamic programming and it showed better performance compared to PhaseBased [2] that was announced in 2007. However, disparity was found to become blocked in regions where the image color was homogeneous. This is a general problem that usually occurs in the hierarchical approach and needs to be solved in the future. This problem can be solved if reliability is improved by analyzing the disparity distribution pattern of neighboring pixels with high reliability.

## ACKNOWLEDGEMENT

This research was supported by the Ministry of Knowledge Economy, Korea, under the Information Technology Research Center support program supervised by the Institute of Information Technology Advancement.(Grant number IITA-2008-C1090-0801-0017)

## REFERENCES

[1] Jong-Il Park, Seiki Inoue, “Hierarchical Depth Mapping from Multiple Cameras,” ICIAP '97: Proceedings of the 9th International Conference on Image Analysis and Processing, Florence , ITALIE (17/09/1997) 1997, vol. 1310, pp. 685-692

[2] S. El-Etriby, A. Al-Hamadi, and B. Michaelis. “Dense stereo correspondence with slanted surface using phase-based algorithm,” Industrial Electronics, 2007. ISIE 2007. IEEE International Symposium on 4-7 June 2007 Page(s):1807-1813

[3] Michael Bleyer and Margrit Gelautz, “A Layered Stereo Algorithm using Segmentation and Global Visibility Constraits,” ICIP '04. Page(s):2997 - 3000 Vol. 5

[4] Brown, M.Z, Burschka, D, Hager, G.D, “Advances in computational stereo,” Pattern Analysis and Machine Intelligence, IEEE Transactions on Vol. 25, Issue 8, Aug. 2003 Page(s):993 - 1008

[5] Scharstein, D, Szeliski, R, Zabih, R, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” Stereo and Multi-Baseline Vision, 2001. (SMBV 2001). Proceedings. IEEE Workshop on 9-10 Dec. 2001 Page(s):131 - 140

[6] Fusiello, A, Roberto, V, Trucco, E, “Efficient stereo with multiple windowing,” Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on 17-19 June 1997 Page(s):858 - 863