



PERGAMON

Pattern Recognition 34 (2001) 2573–2576

**PATTERN  
RECOGNITION**

THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Rapid and Brief Communication

# A robust stereo disparity estimation using adaptive window search and dynamic programming search

Chang Seob Park<sup>1</sup>, Hyun Wook Park\*

*Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, 373-1 Kusong-dong, Yuseong-gu, Taejeon 305-701, South Korea*

Received 2 November 2000; accepted 21 November 2000

## 1. Introduction

Most research efforts in stereo vision have been focused on an accurate disparity estimation. Several factors, such as perspective projection, intensity inconsistency between stereo image pair, and occlusion of objects, make correspondence problem of the disparity estimation difficult. To overcome these problems and to estimate an accurate disparity map, several schemes have been developed, which can be grouped into two categories such as area-based and feature-based approaches [1,2].

In area-based approaches, pixels or regions are used to measure the similarity between stereo pair. They yield dense disparity fields but tend to fail because of local ambiguities in the correspondence. Various estimation techniques have been proposed to overcome these problems [1,2]. Stereo matching by using correlation or sum of squared difference (SSD) is a basic technique that obtains a dense disparity map from stereo images. In the stereo matching, window size must be large enough to reduce noise-sensitive distortion. However, the position of maximum correlation or minimum SSD becomes ambiguous and blurred if the window size is too large. On the other hand, if the window size is too small, it gives poor disparity estimate when the signal-to-noise ratio of the stereo images is low. Therefore, the adaptive window method [3] had been proposed to overcome problems with fixed-size window. This paper provides a robust disparity estimation solution, which combines the adap-

tive window search and the dynamic programming search.

## 2. Proposed algorithm

### 2.1. Area-based disparity estimation by adaptive-size window

Kanade and Okutomi [3] developed an adaptive technique with different window size for each pixel, which selects an appropriate window by evaluating the local variation of the intensity and disparity. This method requires high computational costs for searching the disparity with the adaptive window iteratively. Our proposed adaptive algorithm is more efficient in computational costs than their algorithm since it does not use iterative process. The proposed algorithm firstly determines the search window size for each pixel, and then performs disparity estimate with the window.

Fig. 1 shows an example of window size decision. Basically, we select window size such that the area within the window has similar gray values. Variance of an object boundary is apt to change abruptly. If search window includes the abruptly change area, disparity errors can be produced because there is much depth difference. The thick lines in Fig. 1 are object boundaries, and point T is a target point for disparity estimation. A left image of stereo pair is diagonally divided into four parts centered at the target point. For the four divided parts, the window size is expanded from center to each direction ( $x^-$ ,  $x^+$ ,  $y^-$ ,  $y^+$ ). While the window size increases to each direction, if the local variance of a pixel within the window exceeds a certain threshold ( $VTH = 2$  in this paper), then we stop expanding the window to such direction. This means that a certain disparity discontinuous points such as an object boundary is encountered. Local variance,  $var_L(i, j)$ , of a left image,  $f_L(i, j)$ , is obtained as

<sup>1</sup> Also with Technical Research Institute, Korean Broadcasting System, Seoul, South Korea.

\* Corresponding author. Tel.: + 82-42-869-3466; fax: + 82-42-869-3410.

E-mail address: hwpark@athena.kaist.ac.kr (H.W. Park).

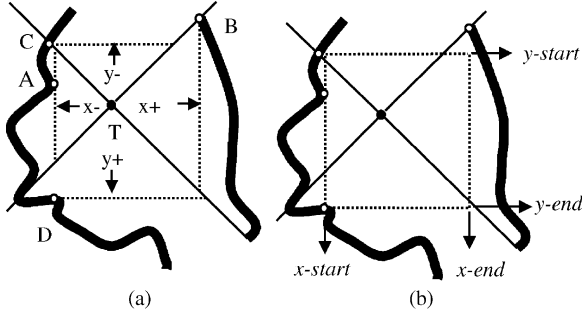


Fig. 1. Procedure to determine adaptive-size window: (a) window size growing, and (b) final window of the rectangle with broken lines.

follows:

$$var_L(i, j) = \frac{1}{4} \sum_{u=0}^1 \sum_{v=0}^1 (f_L(i+u, j+v) - avg_L(i, j))^2, \quad (1)$$

$$avg_L(i, j) = \frac{1}{4} \sum_{u=0}^1 \sum_{v=0}^1 f_L(i+u, j+v). \quad (2)$$

In Fig. 1, increase of window size in the  $x^-$  direction stops at the point A,  $x^+$  at the point B,  $y^-$  at the point C, and  $y^+$  at the point D. Thus, the points of A, B, C, and D define the  $x\_start$ ,  $x\_end$ ,  $y\_start$ , and  $y\_end$ , respectively, of a rectangular window (dotted rectangle in Fig. 1(b)). We confine the maximum window size, MAXW, for saving calculation time and the minimum window size, MINW, for efficient matching. We select MINW =  $5 \times 5$  and MAXW =  $41 \times 41$  empirically in this paper.

Disparity estimation is performed on both the intensity map and the variance map by using the adaptive-size window for more reliable estimation. The correlation function is used as a matching cost function as follows:

$$cost_f(i, j, d) =$$

$$\frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f_L(i, j) f_R(i-d, j)}{[\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f_L(i, j)^2]^{1/2} [\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f_R(i-d, j)^2]^{1/2}}, \quad (3)$$

where  $f_L(i, j)$  and  $f_R(i, j)$  are left and right images of stereo pair, respectively, window size is  $M \times N$ , and  $d$  is the disparity. The cost function for variance,  $cost_{var}(i, j, d)$ , are similarly calculated from local variance maps,  $var_L(i, j)$  and  $var_R(i, j)$ , of left and right images, respectively, as defined in Eq. (1).

If the matched result is in confidence with following three criteria, then the value of  $d$  is the valid disparity. However, the matched result does not satisfy with the criteria, the pixel point is reserved to dynamic programming search for more reliable estimation. Three criteria

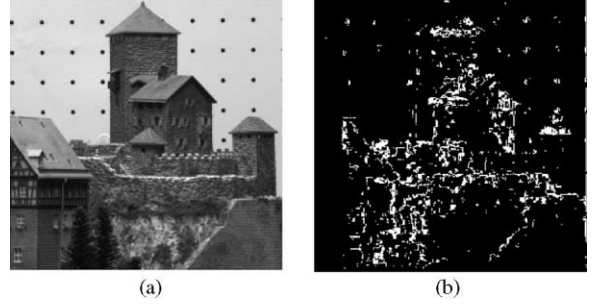


Fig. 2. Mask and disparity map: (a) original left image, (b) mask from the adaptive window search (white region corresponds to invalid disparity).

are given as follows:

- (1)  $max\_cost_f(i, j) > 0.9$ , where  $max\_cost_f(i, j)$  is the correlation of the most matched point between left and right images, i.e.,  $max\_cost_f(i, j) = Max_d cost_f(i, j, d)$ .
- (2)  $max\_cost_f(i, j) - 2^{nd}\_max\_cost_f(i, j) > THI$ , where  $2^{nd}\_max\_cost_f(i, j)$  is the second largest correlation between left and right images, and THI is set to 0.0001 in this paper.
- (3) Disparity values from stereo images and variance maps of the images are identical, i.e.,  $d_f = d_{var}$ , where  $d_f$  satisfies  $max\_cost_f(i, j) = cost_f(i, j, d_f)$ ,  $d_{var}$  satisfies  $max\_cost_{var}(i, j) = cost_{var}(i, j, d_{var})$ , and  $max\_cost_{var}(i, j) = Max_d cost_{var}(i, j, d)$ .

Fig. 2(b) shows an example of mask that is produced by adaptive window search. In this mask, the black regions correspond to the valid disparity from the adaptive window search, whereas the white regions correspond to invalid disparity. Therefore, the white regions have to be processed by dynamic programming search described in the following section.

## 2.2. Disparity estimation by dynamic programming

Dynamic programming finds a global matching optimum as the best path between two sequences of primitives, where one of them should be ordered. Dynamic programming has been proven to be a very effective method for solving edge-based stereo matching within a scanline [4]. It is based on the assumption that there are no edge reversals in the two corresponding scanlines. This is a realistic assumption because edge reversals rarely occur in natural images. When they do appear, even the human visual system has problems with identifying correct matches. Cox et al. [5] pointed out that several good paths could be found through matching space by using only the occlusion and ordering constraints. They found set of correspondences that

maximized the cost function subject to ordering and unique constraints. Our proposed algorithm uses the same cost function as Cox et al. and use the maximum likelihood, minimum horizontal plus vertical discontinuities (MLMHV) algorithm [5]. Also we adopt the dynamic programming algorithm with sub-scanline that is defined as unreliable disparity region from previous adaptive window search.

In the mask of Fig. 2(b), groups of consecutive invalid pixels ('XXXXXX' in Fig. 3) are selected along each scan line. Each group consists of the consecutive invalid pixels and their adjacent pixels having valid disparity values at the left- and right-hand side (' $d_1$ XXXXXX $d_2$ ' in Fig. 3). Considering the start and end disparity values in each group, disparity space image (DSI) of the left and right images is generated as shown in Fig. 3 by using cost function of Cox et al. [5]. The cost function,  $C(i, j)$ , is a mean square error (MSE) function as follows:

$$C(i, j) = \frac{(f_L(i, j) - f_R(i, j))^2}{4\sigma^2}, \quad (4)$$

where  $\sigma^2$  is a variance of white noise in image and it is set to 4 in our experiments.

The dynamic programming algorithm is applied to the DSI. The optimal path is obtained by tracing the path that has the minimum cost from upper-right corner to lower-left corner. In Fig. 3, the optimal path is denoted by the sequence of the gray pixels. In the optimal path, we can calculate actual disparity values by using location difference between corresponding pixels in left and right images. In conventional disparity estimation method, the dynamic programming is applied to whole horizontal scan line of left and right images so that its DSI is always square. However, our proposed method uses a part of scan line to generate DSI. Therefore, the number of pixels in the line segment of left image may be different from the number of corresponding pixels in right image as an

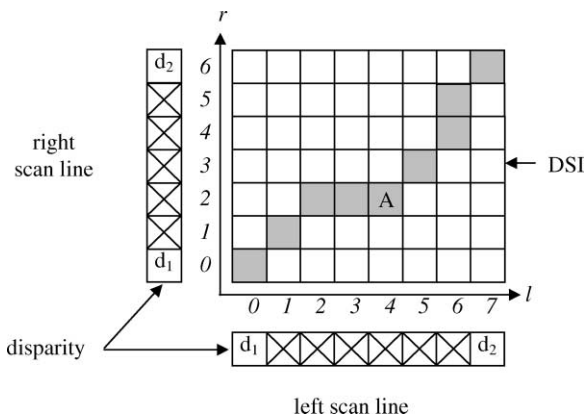


Fig. 3. An example of disparity space image (DSI) when  $d_2 = d_1 + 1$ .

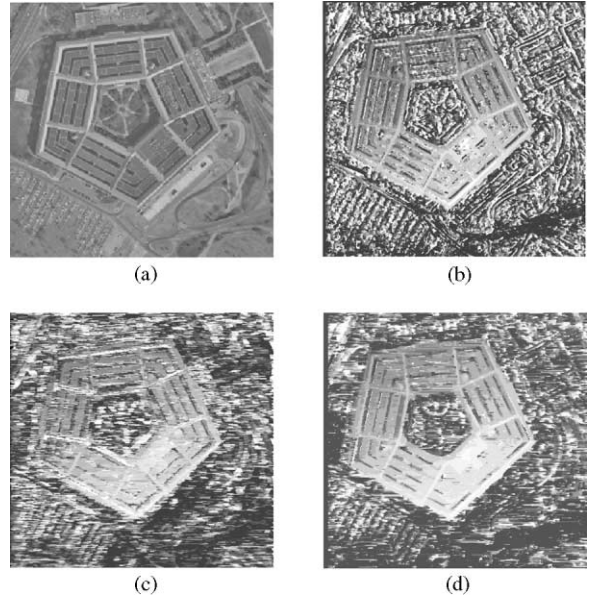


Fig. 4. The "pentagon" stereo image pair of (a) left image, (b) disparity map from  $5 \times 5$  window, (c) disparity map from MLMHV algorithm, and (d) disparity map from our proposed algorithm.

example of Fig. 3. Fig. 3 shows the case of  $d_2 = d_1 + 1$ . The real disparity value of the matched pixel  $p$  in DSI is calculated as follows:

$$\text{disparity}(p) = l - r + d_1, \quad (5)$$

where  $(l, r)$  is the location of a pixel  $p$  in DSI and  $d_1$  is the disparity value of the left most pixel. If one matched pixel ('A' in Fig. 4) has  $l = 4$ ,  $r = 2$  and  $d_1 = 5$ , then its real disparity value becomes 7. The final disparity map is obtained by filling the disparity values obtained by dynamic programming search at the invalid pixels of the adaptive window search.

### 3. Experiment results

Experiments are performed in Pentium PC using Visual C++ for selected natural images. In most cases, our algorithm outperforms the window or dynamic programming only methods. One result of Pentagon image is shown in Fig. 4, in which the proposed algorithm shows clear contour of the building and it shows less errors in the area of roads and parking lots around the building. Other methods show incorrect disparity shape in roof of building, or more noise around the building.

Since the quantitative evaluation of the disparity estimation is very difficult, it is usually performed by subjective visual evaluation. Even though it is not a direct

Table 1  
PSNR (dB) of the reconstructed left images from corresponding right images by using disparity maps

Image	5 × 5 window	MLMHV	Proposed
Shrub	36.6	37.5	36.9
Castle	29.1	29.1	30.4
Ball	16.9	15.1	16.1
Parkmeter	31.7	34.2	33.3
Pentagon	21.5	22.2	22.6
Face	19.7	23.7	23.8

evaluation, we performed a quantitative evaluation that compared the original left image and the reconstructed left image from the right image using the disparity map by measuring peak signal-to-noise ratio (PSNR) as follows:

$$PSNR = 10 \log_{10} \left( \frac{255 \times 255}{MSE} \right) \text{dB}, \quad (6)$$

where

$$MSE = \frac{1}{K_1 K_2} \sum_{i=0}^{K_1-1} \sum_{j=0}^{K_2-1} (f_L(i,j) - \tilde{f}_L(i,j))^2. \quad (7)$$

In Eq. (7),  $\tilde{f}_L(i,j)$  is the reconstructed left image from right image using obtained disparity map,  $K_1 \times K_2$  is

the image size, and the images have a maximum gray value of 255.

In the occluded region of the stereo image pair, we cannot reconstruct left image from right image correctly because there is no information of occluded region in one of the two images. The occluded region can be detected by dynamic programming, but the region is not specially treated for PSNR calculation in this paper. Therefore, there may be a discord between the PSNR and the subjective visual evaluation. PSNRs of the reconstructed left images are given in Table 1. In general, our method shows good results in most of the images.

## References

- [1] U.R. Dohnd, J.K. Aggarwal, Structure from stereo: a review, IEEE Trans. SMC 19 (6) (1989) 1489–1510.
- [2] E. Trucco, A. Verri, Introductory Techniques for 3-D Computer Vision, Prentice-Hall, Englewood Cliffs, NJ, 1998.
- [3] T. Kanade, M. Okutomi, A stereo matching algorithm with an adaptive window: theory and experiment, IEEE Trans. Pattern Anal. Mach. Intell. 16 (9) (1994) 920–932.
- [4] Y. Ohta, T. Kanade, Stereo by intra- and inter-scanline search using dynamic programming, IEEE Trans. Pattern Anal. Mach. Intell. 7 (2) (1985) 139–154.
- [5] I.J. Cox, S.L. Hingorani, S.B. Rao, B.M. Maggs, A maximum likelihood stereo algorithm, Comput. Vision Image Understanding 63 (3) (1996) 542–567.