MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

13—9 Area-based Estimation of Stereo Disparity using Hierarchical Windows

Atsushi Osa * Hiroshi Yamashita Hidetoshi Miike

Faculty of Engineering
Yamaguchi University

Abstract

We propose a new solution to stereo matching which is combined the area-based stereo method using hierarchical windows with the pixel-based stereo methods using global optimization. The area-based method has a problem that the best size of area is unknown. Since the proposed method changes the sizes of windows hierarchically, the local minima problem and a-priori knowledge of the size can be removed. The proposed method brings better robustness in the analysis of stereo images having periodic texture and textureless areas of images, and a scene having abrupt disparity changes, compared to the ordinary stereo method using hierarchical multi-resolution approach.

1 Instructions

A lot of algorithms have been proposed to evaluate the depth of the three-dimensional shape from stereo images. Wei et al. classified the current stereo algorithms into three classes: area-based, feature-based, and pixel-based[1]. The area-based stereo methods are based on the assumption that the disparities within a neighborhood of a pixel in each image are constant, thus the intensity distribution within the area can be used to find the corresponding pixel. An extension of the window brings a reduction of error of matching, however, the disparity map becomes low resolution. It is not clear how the solution is affected by the size of window. The feature-based stereo methods utilize signed zero-crossings, gradient peaks, line segment, and so on. The pixel-based stereo methods perform the matching by use of the intensity at a single pixel. The earliest formulations of the methods are based on minimizing the left and right intensity differences, subjected to a smoothness constraint using global optimization. To avoid slow convergence and local minima problem, the pixel-based methods often use hierarchical approaches that parameterize the disparity function by the superimpositions of hierarchical Gaussians smoothing filters and multi-resolutions[2].

In this study, we propose a new solution to the stereo matching problem. The method combines the pixel-based stereo methods using global optimization with the area-based stereo method using hierarchical windows. In this method, there is no need to know an ideal window size. A dense disparity map can be obtained. Adoption of the hierarchical windows brings higher robustness for objective stereo images compared to multi-resolution scheme that is the ordinary hierarchical approach.

*Address: 2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611 Japan. E-mail: osa@kde.yamaguchi-u.ac.jp

2 Method

We suppose that $f(x, y)$ and $g(x, y)$ are the image functions of the left and right cameras, respectively. For simplicity, we assume corresponding points in the two images to lie on the same horizontal scan lines (epipolar constraint). Our cost function is described by the follow equations.

$$E(u) = P(u) + \lambda S(u), \tag{1}$$

$$P(u) = \int_x\int_y \Bigg\{ \sum_{a=-\frac{\delta a}{2}}^{\frac{\delta a}{2}} \sum_{b=-\frac{\delta b}{2}}^{\frac{\delta b}{2}} \Big[f\left(x+u(x,y)+a, y+b\right) - g\left(x+a, y+b\right)\Big]^2 \Bigg\} dxdy, \tag{2}$$

$$S(u) = \int_x\int_y \left[\left(\frac{du}{dx}\right)^2 + \left(\frac{du}{dy}\right)^2 \right] dxdy, \tag{3}$$

where $u(x, y)$ is the disparity map to be determined, $\delta s = \delta a \times \delta b$ indicats the size of a local area (window size), and λ denotes a suitable weighting factor. In general, the conventional area-based method is designed to minimize the cost function directly on each window. Here, under the assumption that the disparity $u(x, y)$ is constant in a local area, we try to combine the pixel-based stereo methods using global optimization with the area-based stereo method. Inserting Eq.(1) into the Euler-Lagrange differential equation yields the following equations:

$$\sum_a \sum_b \left\{ \left[f(x+u(x,y)+a, y+b) - g(x+a, y+b) \right] \times \frac{d f(x+u(x,y)+a, y+b)}{dx} \right\} - \lambda \Delta u = 0. \quad (4)$$

We can solve the Eq.(4) by Gauss-Seidel method.

On the other hand, the conventional hierarchical approach parameterizes the disparity function by the superimpositions of hierarchical Gaussians smoothing filters and multi-resolutions. The multi-resolution scheme decreases calculation cost, however, it brings an information drop of image. In this study, the resolutions of images are kept constant at every level of hierarchies, but the sizes of windows are changed hierarchically. The evaluated disparity map at the last level is used as the initial result at the next level.

3 Experimental images and analysis

In this section, we compare the performance of the proposed method with one of the conventional pixel-based method. March's method [3] with adoption of the ordinary hierarchical processes [2] is picked out as the conventional method. Experimental data are provided from JISCT database [4] and stereo images of computer vision and pattern recognition group in Bonn university [5].

The proposed method introduced four levels of hierarchy, in which variable size of windows like 9×9 pixels, 5×5 pixels, 3×3 pixels, and 1×1 pixel were adopted. The control parameter λ was kept constant, 2000. The number of iteration at each level was 100. In the ordinary method, we introduced four levels of hierarchy, in which variable resolutions of images like 1/8, 1/4, 1/2, and 1/1 (original). The parameter λ was kept constant, 2000. The number of iteration at each level was 100.

A stereo matching method using a hierarchical approach was proposed in recent years, which was called intensity and gradient based matching method with a hierarchical Gaussians[1]. For the comparison of the results, the disparity maps of images 1 and 2 are photocopied from Wei's paper [1]. Note that the intensity of the disparity map is not the same scale as the results of our method and the ordinary method. Wei et al. [1] didn't present the result of the experimental stereo images 3 (corridor).

Figure 1 shows the stereo images 1 (pentagon) [4] and the disparity maps obtained by the respective methods. The images had resolution of 512×512 pixels. The brightness was quantified into 256 steps. Those stereo images had tiny disparity, in which the maximum depth was about only 7 pixels. At a glance, the disparity map of the proposed method and the ordinary method were almost coincident with each other. By detailed comparison, however, we found that the error of the proposed method in disparity map was smaller than that of the ordinary method. The resolution of disparity map estimated by the proposed method seems to be greater than that of Wei's method. Figure 2 shows a mean disparity changes respect to iterations for the stereo images 1. The curve represents convergent characteristic of estimated disparity. The mean disparity changes in the ordinary method seem to converge, but the mean disparity changes in the proposed method did not still converge. It is possible to increase accuracy of disparity map when the iteration is increased in the proposed method.

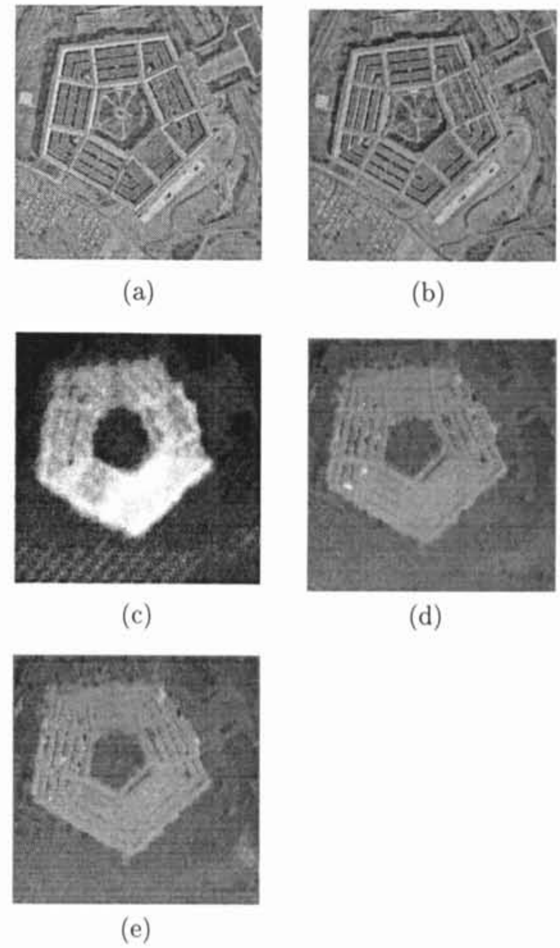

Figure 1: The stereo images 1 (pentagon). (a) Left image. (b) Right image. (c) Wei's disparity. (d) Disparity obtained by March's method + the ordinary hierarchical approach. (e) Our disparity map.

Figure 3 shows the stereo images 2 (shoe) [4] and the results of the analysis. The resolution of the images was 512×512 pixels. The brightness was quantified into 256 steps. The disparity of the floor having periodic structure was 1 pixel, and the area of foot and shoe had also a constant disparity of 24 pixels. On the area of foot and shoe, Wei's disparity map had a distribution of depth values, in which the edge of foot and shoe was blurred. In the ordinary method and the proposed method,

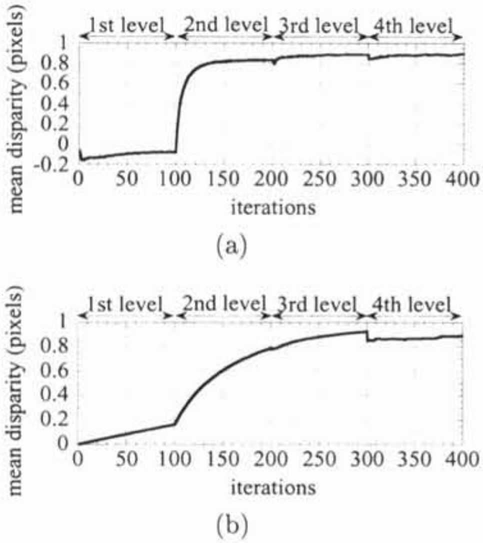

Figure 2: Mean disparity changes respect to iteration for stereo images 1. (a) March's method + the ordinary hierarchical approach. (b) The proposed method.

the foot and shoe areas of those results had almost a constant disparity. However, there was large mismatching area around the foot and shoe in the ordinary method. Figure 4 shows the changes of mean disparity against to the number of iterations for the stereo images 2. In the proposed method, the mean disparity changes did not well converge. It is possible to increase accuracy of disparity, when the iteration is increased in the proposed method.

The stereo images 3 (corridor) [5] were synthetic data (Fig.5(a)(b)). We could obtain theoretical disparity map (Fig.5(c)). The original images had resolution of 256×256 pixels. In the course of analysis, we made double resolution images from original to adjust the size to the stereo images 1 and 2. The brightness was quantified into 256 steps. The images 3 included many textureless areas. Figure 6 shows the results at four different hierarchical levels estimated by the ordinary method. There was a lot of mismatching area from the early hierarchical level. Figure 7 shows the results at four different hierarchical levels estimated by the proposed method. In spite that the spatial resolution of the disparity map was low, rough depth information was well recovered. The final errors estimated by the ordinary method and by the proposed methods are 6.06 and 4.03 pixels, respectively.

4 Summary

We can summarize the characteristics of the proposed method as follows. First, since we adopted the hierarchical windows approach, we obtained dense and accurate disparity map with the proposed method. Second, the ordinary hierarchical approach using multi-resolution has a faster convergence in

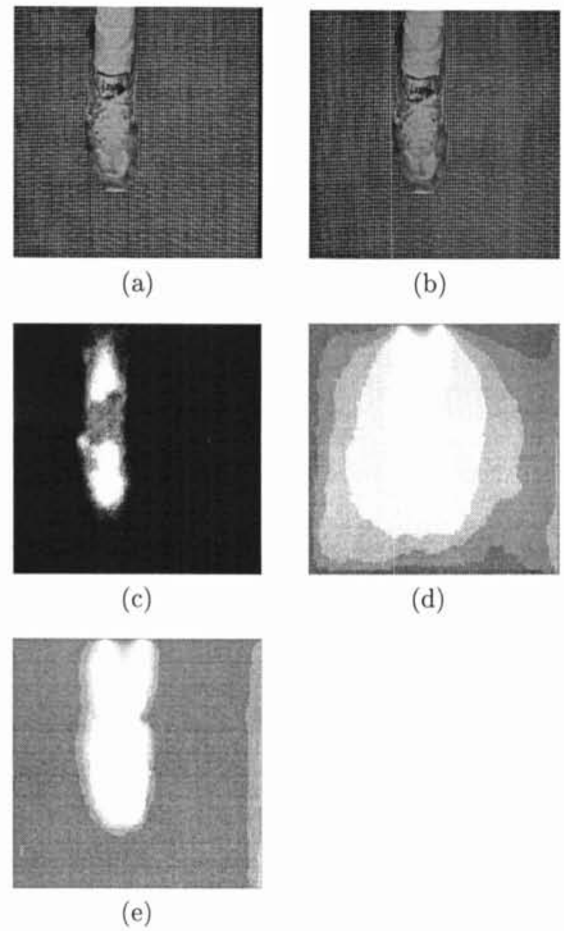

Figure 3: The stereo images 2 (shoe). (a) Left image. (b) Right image. (c) Wei's disparity. (d) Disparity obtained by March's method + the ordinary hierarchical approach. (e) Our disparity map.

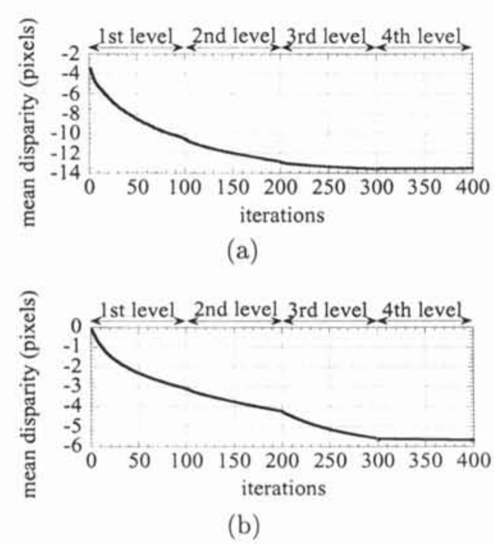

Figure 4: Mean disparity changes for stereo images 2. (a) March's method + the ordinary hierarchical approach. (b) The proposed method.

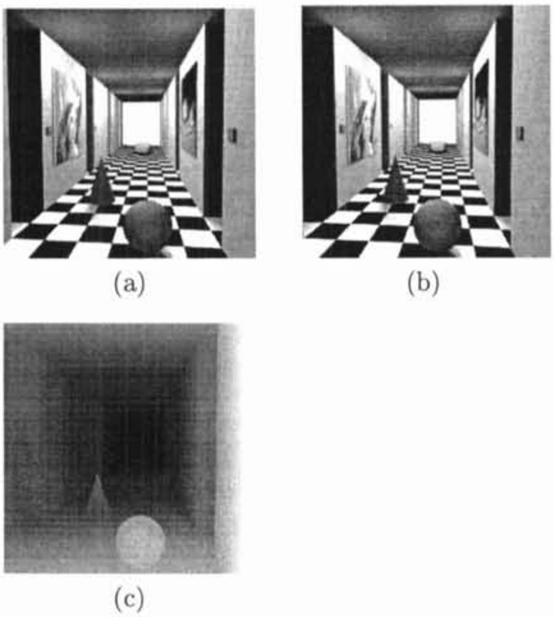

Figure 5: The stereo images 3. (a) Left image. (b) Right image. (c) Theoretical disparity map.

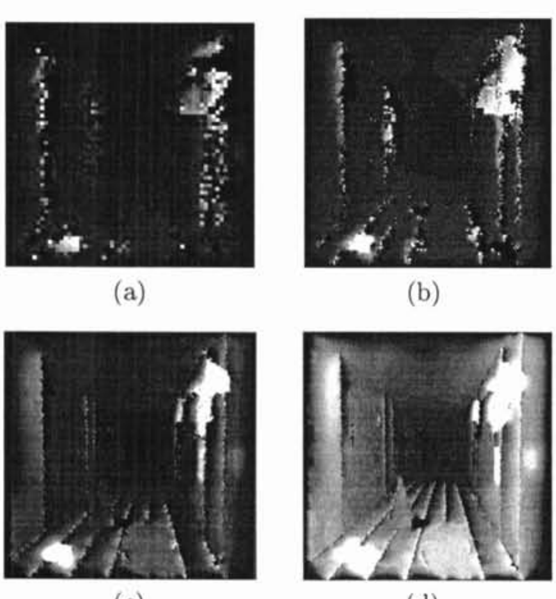

Figure 6: Disparity maps at four different hierarchical levels by March's method + the ordinary hierarchical approach. (a) 1st level, (b) 2nd level, (c) 3rd level, and (d) 4th level.

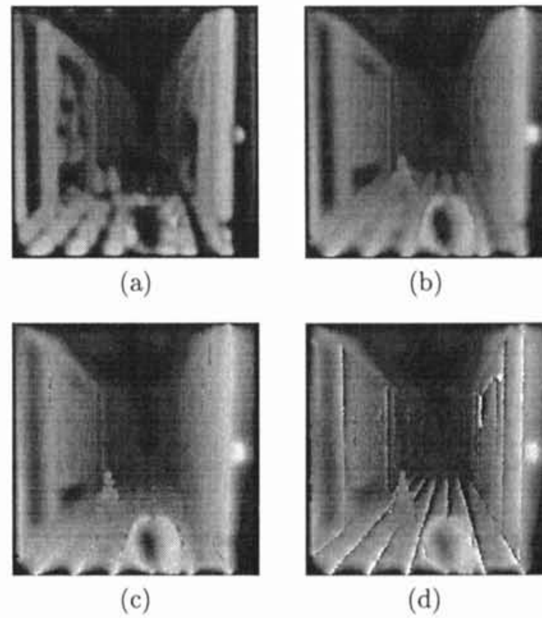

Figure 7: Disparity maps at four different hierarchical levels by the proposed method. (a) 1st level, (b) 2nd level, (c) 3rd level, and (d) 4th level.

the analysis than that of the proposed. However, the proposed method brings better robustness in the analysis of stereo images, having periodic texture, textureless images and a scene having abrupt disparity changes.

References

[1] G.Q. Wei, W. Brauer, and G. Hirzinger, "Intensity and gradient-based stereo matching using hierarchical Gaussian basis functions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.20, no.11, Nov. 1998.

[2] N. Yokoya, "Surface reconstruction from stereo images preserving discontinuities by multiscale regularization," Trans. of IEICE(D-II), vol.J76-D-II, no.8, pp.1667-1675, 1993. (in Japanese)

[3] R. March, "Computation of stereo disparity using regularization," Pattern Recognition Letters, vol.8, no.3, pp.181-188, 1988.

[4] R.C. Bolles, H.H. Baker, M.J. Hannah, "The JISCT stereo evaluation," Proc. DARPA Image Understanding Workshop, pp. 263-260.

[5] http://www-dbv.cs.uni-bonn.de/stereo_data/