Acquisition of Sharp Depth Map from Multiple Cameras

Submission Paper to Special Issue on 3D Video Technology,
Signal Processing: Image Communication

- Author: Jong-Il Park and Seiki Inoue
- Affiliation: ATR Media Integration & Communications Research Labs.
- Address: 2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan
- Tel: +81-774-95-1466 Fax: +81-774-95-1408
- Email: {pji,sinoue}@mic.atr.co.jp

List of Symbols

Z: depth (Z value of an object point in camera-centered coordinate)

F: focal length

L: baseline stretch

$I_i(\mathbf{x})$: intensity at a point $\mathbf{x}$ on the image plane of image i.

$e(\mathbf{x}, d)$: sum of squared-difference of intensity with disparity d at a point $\mathbf{x}$.

d: disparity

Number of Pages: 18

Number of Figures: 13

Key Words: stereo matching, depth map, occlusion, hierarchical method

Acquisition of Sharp Depth Map from Multiple Cameras

Abstract

We present a method to estimate a dense and sharp depth map using multiple cameras. A key issue for obtaining sharp depth map is how to overcome the harmful influence of occlusion. Thus, we first propose an occlusion-overcoming strategy which selectively use the depth information from multiple cameras. With a simple selection technique, we resolve the occlusion problem considerably at a slight sacrifice of noise tolerance. Another key issue in area-based stereo matching is the size of matching window. We propose to use a hierarchical estimation scheme that attempts to acquire a sharp depth map such that edges of the depth map coincide with object boundaries on the one hand, reduce noisy estimates due to insufficient size of matching window on the other hand. Owing to the unique property of occlusion-overcoming strategy, we can utilize full benefit of hierarchical schemes. We show the method can produce a sharp and correct depth map for a variety of images.

1 Introduction

Recently, computer vision technology is widely used for achieving high degree of freedom and efficiency in video content creation. Thus, it is even said to be a kind of media technology [11].

We are developing a video component database in order to realize a flexible and versatile framework for video content creation [5]. It is based on the layered representation of video where a video sequence is regarded as a spatio-temporally ordered set of video components [17]. Video components are stored with various property information such as camera work, key words, depth, and the like in the database. We can freely select some video components from the database and enjoy arranging them in a spatio-temporal domain to make a new video and/or creating new video expressions by exploiting the given property information.

Among the information, one of the most important one would be depth. It is Z value of the camera-centered coordinate of the corresponding object point for each pixel, where Z axis is set to optical axis. Depth information corresponds to the spatial part in the spatio-temporal description of a scene. Thus, it takes a crucial role in making natural-looking videos and/or creating various video expressions with high degree of freedom using the video component database. Virtualized Reality [7], Z-keying for video composition [8], 3D special video effects [13], and arbitrary view generation [15] are typical application using the depth information. Moreover, we can automatically generate multi-layer description of a scene using depth information [14].

For such application, dense and sharp depth map is strongly required. Here, "sharp" means that object boundary and/or depth discontinuity should be correct. Correctness of depth map in shape is sometimes more important than precision of depth value. In this paper, how to get such dense and sharp depth map is the main theme. The proposed method is a hierarchical scheme consisting of an occlusion-overcoming disparity estimator which exploits stereo images from 5 cameras. Considering hardware feasibility, we confine the method to a signal-level processing. All the processing is localized such that a parallel implementation can be achieved.

This paper is organized as follows. After a brief review on related work in Section 2, we give details of the proposed method in Section 3. Experimental results are presented in Section 4.

2 Related Work

Stereo matching is a useful method in obtaining depth map from image. There are two approaches in stereo matching. One is area-based method and the other is feature-based method. Area-based method is, in general, used for obtaining a dense depth map [1]. However, it is well-known that the area-based stereo matching faces several problems such as lack of texture, occlusion, photometric change, repetitive pattern, and so on [1][2].

A considerable amount of effort has been exerted to cope with such problems in com-

puter vision [2][10]. Almost methods to obtain dense depth map are computationally expensive or iterative. Among some exception is multiple-baseline stereo matching [18][12]. It demands more cameras but alleviates the problems of lack of texture and repetitive texture without much increase of computational complexity. Recently, a real-time depth mapper has been developed on the basis of multiple-baseline method [8].

However, little attention has been paid to clearing the occlusion problem in stereo matching. In fact, occlusion is one of the main culprits to prevent from obtaining correct depth map in shape. In two-view stereo, occlusion problem is unavoidable and it is impossible to get correct match. Only some appropriate interpolation can fill such area based on some assumption and knowledge [1]. From the standpoint of correct match, multiple view(more than two) can give a clue to resolve the occlusion problem. When an area is occluded in an image from a camera, another camera located at a different position has a chance to observe the area and thus it can give a correct match. Kanade *et al.*'s depth mapper does not seem to explicitly exploit this property although they mentioned the occlusion problem a little [8]. Recently, Nakamura *et al.* extensively studied the occlusion problem [9]. Using eye array camera, they analyze occlusion patterns quantitatively and propose a disparity estimation scheme which is capable of detecting occlusion, selecting a proper mask for correct match, and thus preventing from mismatch. However, it demands at least 9 cameras and furthermore, it does not provide a strategy for controlling the effect of matching-window size.

A very important issue underlying the area-based matching is the size of matching window [6][10]. It should be large enough to include enough intensity variation for characterizing an area. But it should be small enough to avoid projective distortion. Toward resolving such a dilemma, two approaches have been proposed. One is to use a locally adaptive window [6]. It searches for a window that produces the estimate of disparity with the least uncertainty for each pixel of an image. Considerable improvement is obtained from the aspects of smooth surface and sharp disparity edges. However, the problem is that it is iterative. The other approach is to use hierarchical schemes [2][10][3][4]. By restricting searching range of fine resolution to the estimates of coarse resolution, the trade-off problem can be considerably alleviated. In this approach, the local distribution of estimates around depth discontinuities is of significant importance. If it is not clearly differentiated between objects, we cannot expect further improvement of estimate with hierarchical schemes.

In this paper, we propose an occlusion-overcoming multi-view stereo matching technique implemented on hierarchical schemes focused on getting correct depth map around discontinuities. The proposed occlusion-overcoming technique not only alleviates occlusion problem, but also produces abruptly changing estimates in the vicinity of depth discontinuities. The abrupt transition of estimates near depth discontinuities is the very desirable property in the hierarchical estimation schemes. Two kinds of hierarchical implementation are presented. One is fine-to-fine approach which requires large amount of computation. The other is a coarse-to-fine method which alleviates computational burden considerably. We will explain the details in the following section.

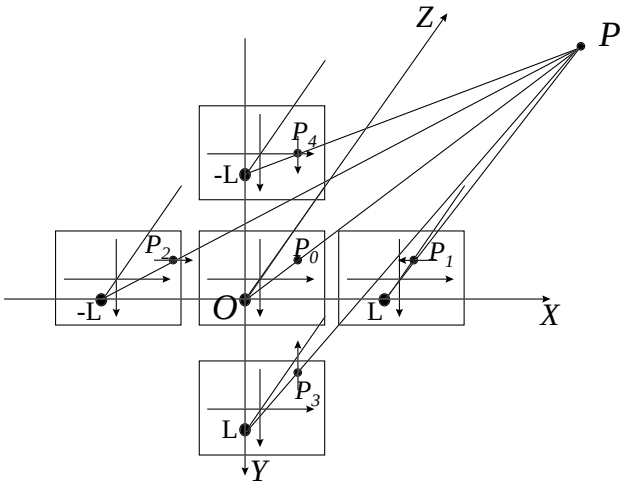

Figure 1: Projection geometry of camera system.

3 Depth Estimation

3.1 Configuration of 5 Camera System

The proposed configuration of multiple camera system is shown in Fig. 1. We put a camera at the center and a total of 4 cameras of the same specification to each direction of upper, lower, right, and left, separated by the same distance L.

Under the projection geometry of Fig. 1, an object point $P = (X, Y, Z)$ is projected to the point $p_0 = (x_0, y_0)$ on the image plane of the center(=base) camera, where $x_0 = F\frac{X}{Z}$ and $y_0 = F\frac{Y}{Z}$, and F is the focal length. It is also projected to $p_i = (x_i, y_i)$ on image plane of each inspection camera C_i, $i = 1, 2, 3, 4$, where

$$x_i = F\frac{X - D_{i,x}}{Z}, \qquad y_i = F\frac{Y - D_{i,y}}{Z}. \tag{1}$$

Here, the baseline stretch $D_i = (D_{i,x}D_{i,y})$'s are $D_1 = (L, 0)$, $D_2 = (-L, 0)$, $D_3 = (0, L)$, and $D_4 = (0, -L)$. In the configuration, the true disparity d_t of the object point P is $d_t = \frac{FL}{Z} = |p_i - p_0|$, for all i. Thus, by estimating the disparity, we can obtain the depth of an object point. In this paper, we estimate a disparity map. It is converted to depth map using the above relationship before application.

The camera configuration is based on the assumption that when a camera cannot give a correct match for a pixel because of occlusion, another camera located at the other side can give a good one. This holds good for almost occluding cases[1]. In this sense, we may say, the more the number of cameras, the more chance to obtain a good match we

[1]The configuration cannot cope with strongly concave area.

have. However, if we consider practical implementation, the number of cameras should be kept small. The least number of cameras to cope with such kind of occlusion can be considered to be 3. But, if we locate the 3 cameras collinearly, any dedicated schemes cannot be free from noisy estimates in the areas without intensity variation along the epipolar line. If they are not collinear, there will be less chance of overcoming occlusion problem. Thus, we set the number of cameras to 5 which can reasonably cope with the occlusion problem.

Another important issue is the baseline stretch. Long baseline gives accurate estimates but suffers from photometric distortion and severe occlusion. Since we are aiming not at 3D reconstruction of an object but at extracting object profiles of a complex scene, we give more weight on occlusion problem. Thus, we maintain the baseline stretch much shorter than the distance of the nearest object from the cameras (less than 1/20 throughout the experiments).

Optical axes of the 5 cameras are set to be parallel. The possible digression from the camera mapping model in the Fig.1 can be compensated for by camera calibration [19]. All the cameras should be synchronized in order to cope with moving objects. What we are to acquire is the depth map of the image from the center camera. Other cameras work as sensors in this sense. We now explain how we reduce possible bad matches and obtain a good one around occlusion area in the followings.

3.2 Occlusion-Overcoming Multi-View Matching

There are two types of errors around depth discontinuities in stereo matching. The presence of occluding boundaries within the matching window tends to confuse the matcher and often gives an erroneous depth estimate, as Dhond and Aggarwal pointed out[2]. This cannot be avoided with normal block-based matching. The other type of error is due to occlusion. This can happen without any discontinuity within the matching window. We will deal with these problems in this subsection.

Matching is performed at each pixel position with window overlapping. We use the sum of squared-difference(SSD) as a matching measure. At a point $\mathbf{x}$ on the image plane of the base camera C_0, the matching measure is calculated at each displacement d for each camera C_i by

$$e_i(\mathbf{x}, d) = \sum_{\mathbf{b} \in W} [I_0(\mathbf{x} + \mathbf{b}) - I_i(\mathbf{x} + \mathbf{b} + \mathbf{d}_i)]^2 \tag{2}$$

where I_i is the intensity and W is a matching window. The disparity should be the same for all cameras in the camera geometry such that $d = |\mathbf{d}_i|$ for all i.

A straightforward implementation of multiple-baseline stereo [8] using the camera configuration in Fig. 1 would be

$$\hat{d}(\mathbf{x}) \;=\; arg \min_{d} \sum_{i=1}^{4} e_i(\mathbf{x}, d). \tag{3}$$

It gives a good result if there is no discontinuity of depth. However, when there is a discontinuity of depth near the matching window for a pixel, we cannot expect all of

the matching data from the 4 directions gives us useful information. On the contrary, some data, especially from the direction of occluded area, affect the estimation harmfully. They should be eliminated for a good estimation. We illustrate typical example [2] of such phenomenon in Fig. 2 where the e_i curves for a typical point around object boundary are shown (the size of matching window = 7x7 [pixels]). Due to the influence of occlusion, that is, the bad observation data from lower and left cameras, the matching tends to produce undesirable result($\hat{d} \approx 19$) where the true disparity is $d_t \approx 6$. We see why the matching based on eq.(3) cannot be successful in such area.

If we can eliminate such bad observations during the matching, a considerable improvement can be expected. Thus, we devise a simple selection method.

One of the right and the left cameras and one of the upper and the lower cameras are supposed to give a correct match. Thus, we can assume that at least one of the 2 cameras located symmetrically is not influenced by occlusion and thus gives us correct matching data. In the areas where the above assumption does not hold, rare as it is, for example, inside of strongly concave area, we cannot be free from erroneous estimates.

At each displacement, we compare the difference $e_1(\mathbf{x}, d)$ from the right camera with $e_2(\mathbf{x}, d)$ from the left camera and discard the large one. Similarly, we do the process for images from the lower and the upper cameras. Then, the two data are summed. The summation provides more chance to get a correct match in the light of multiple-baseline stereo. By repeating the selection and summation of data along the epipolar lines, we get a 1-D curve for the estimation. We consider the displacement which gives the minimum of the curve as the disparity of the pixel. In short, the estimation scheme can be described by

$$\hat{d}(\mathbf{x}) \;=\; arg\min_{d} \sum_{i=1}^{2} \{min[e_1(\mathbf{x}, d), e_2(\mathbf{x}, d)] + min[e_3(\mathbf{x}, d), e_4(\mathbf{x}, d)]\}, \tag{4}$$

where $min[]$ represents the minimum of elements in the bracket.

The above procedure is equivalent to find the most proper camera set consisting of 3 cameras among the possible combination of camera configuration and apply multiple-baseline stereo. Note that the proposed scheme can deliberatively avoid the cases where 3 cameras are collinear. In such cases, the matcher sometimes fails to give a correct match for the areas with no intensity variation along the epipolar line.

As we see in the Fig. 3, a considerable improvement is achieved around depth discontinuity by the above scheme at the slight sacrifice of noise tolerance as is expected [12]. The loss of noise tolerance will be compensated for by a hierarchical scheme in the following subsection. In the disparity maps, we can observe two kinds of distortion. One is from occlusion. We see many noisy estimates around object boundaries in the left disparity map while no such estimates in the right disparity map. The other is *boundary overreach*. As Cochran *et al.* pointed out, the more strongly textured surface tends to leak into the less textured region. We see the disparity of higher texture tends to reach

[2] The image is from the multi-view image database of Univ. of Tsukuba. The baseline stretch is 8 [mm], the focal length is 10 [mm], the size of CCD is 1/3 [inch], and the image size is 640x480 [pixels]. The distance to the nearest building is 33 [cm].

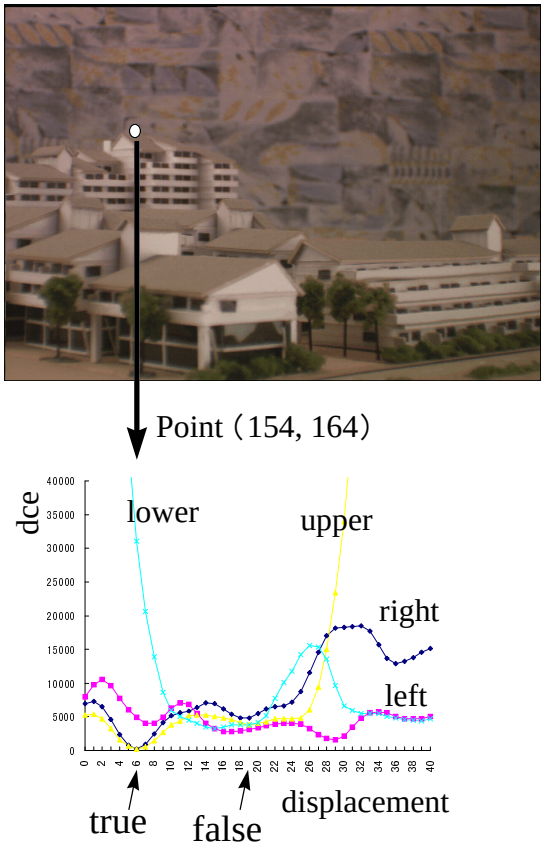

Figure 2: Illustration of the influence of occlusion in stereo matching from multiple cameras.

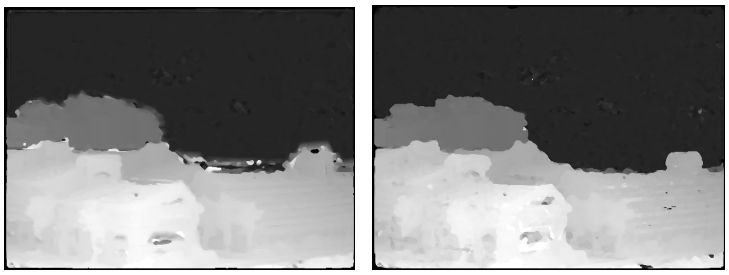

Figure 3: Disparity map from multi-camera matching. Left map is obtained without occlusion-overcoming strategy and right map with the strategy. Matching window size is 7x7.

over the true edge of disparity and out to lower texture area in both of the disparity maps. The difference is that the disparity map by the proposed occlusion-overcoming strategy shows clear and abrupt change around discontinuity, which comes from the nonlinear property of the strategy. Moreover, the amount of the boundary overreach is roughly half of the size of the matching window as we see in the disparity map with edge map overwritten in Fig.4. The size of matching window is 15x15 [pixels] and the size of the leaks is less than 7 pixels. We will discuss the matter in more detail using the results of synthetic images in the next section. The abrupt change of estimates and the limited boundary overreach are useful properties in constructing a hierarchical scheme.

Smaller matching window is favorable in the aspect of reducing boundary overreach. But, smaller matching window gives us less reliable results of estimation. There is a trade-off between reliability and geometric correctness of estimation with respect to the size of matching window. Thus, we propose to use hierarchical schemes in order to cope with the trade-off problem as follows.

3.3 Hierarchical Estimation Scheme

The proposed hierarchical method is based on the observation that, when we use the occlusion-overcoming strategy, the correct disparity for boundary overreach area exists near(within half of the size of the matching window) the point in the disparity map in most cases. We first explain a fine-to-fine implementation and then we present an efficient implementation using a resolution pyramid.

3.3.1 Fine-To-Fine Implementation

Figure 5 illustrates the concept of the fine-to-fine hierarchical method.

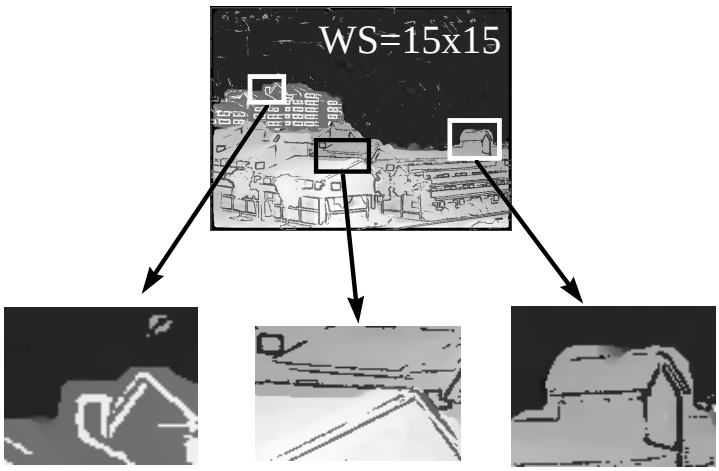

Figure 4: Boundary overreach. The amount of boundary overreach is about half the size of the matching window.

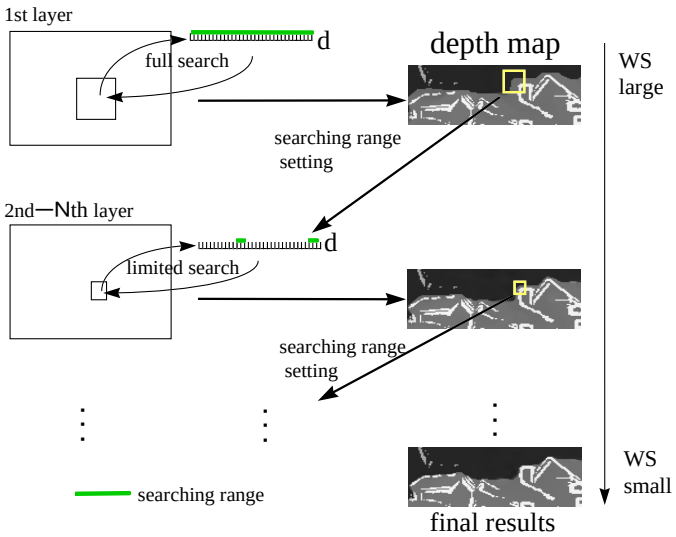

Figure 5: Illustration of the fine-to-fine hierarchical scheme. Searching range is restricted to the estimates in the matching window of previous(=large matching window) layer except for the 1st layer.

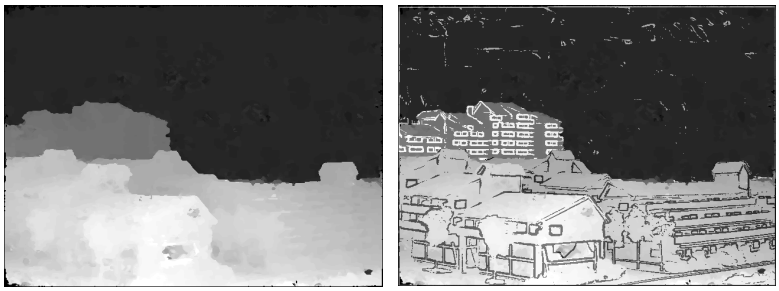

Figure 6: Disparity map obtained by the fine-to-fine hierarchical method [left]. Edge map is overwritten [right].

We start by obtaining a disparity map with a large matching window using the occlusion-overcoming strategy. We assume the true disparity value of a pixel exists within the matching window of the pixel in the disparity map of the 1st layer. Now, we reduce the size of the matching window by half. Then, the disparity is estimated by the occlusion-overcoming technique except that the searching range is restricted to a set consisting of the disparity values of the upper layer within the window of the upper layer at the position. This restriction of searching range is based on the observation on boundary overreach. The procedure is repeated until the last layer where the size of matching window is 3x3 [3].

When there is no disparity discontinuity around a point(within the matching window of the point), we don't need to estimate the disparity of the points again in the successive layers. Instead, we just enhance the resolution of the disparity value to sub-pixel accuracy by quadratic fitting and no more update in the successive layers. Some noisy estimates are eliminated by an edge-preserving order-statistics filter [20].

We see very sharp and correct boundaries in the disparity map of Fig. 6, which can be seen more clearly in the magnified disparity map of Fig. 7. It is obtained by using 3-layer(15x15 to 7x7 to 3x3) hierarchy.

It is possible to reach far lower complexity of matching by analyzing the matching procedure and eliminating the redundancy as is indicated in [8], since the method is based on a regularly overlapping matching.

[3]We set the least window size to 3x3 because 1-pixel matching shows extremely noisy results. As a result, 1-pixel extension inevitably appears at object boundaries. It is observed through many experiments that 1-pixel extension does not seem to cause noticeable artifacts in many application. If we pursue higher quality, it can be eliminated by postprocessing, for example, edge-based directional postprocessing [16].

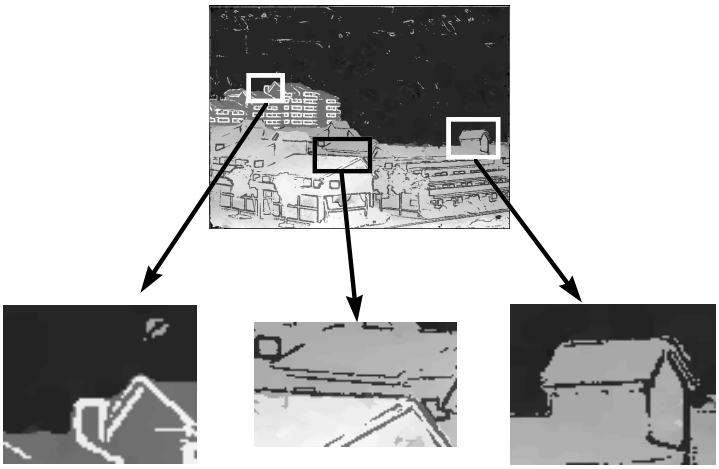

Figure 7: Magnified disparity maps showing how well the proposed method can overcome the occlusion problem and the boundary overreach.

3.3.2 Coarse-to-Fine Implementation

Fig.8 shows the concept of the proposed coarse-to-fine hierarchical method.

Resolution pyramid is first constructed by low-pass filtering and decimation. We use a 4-layer pyramid. The size of matching window is set to 3x3[pixels] through the layers. Searching is performed to a pixel accuracy at each layer.

We start from the most coarse layer. Using the occlusion-overcoming strategy, we acquire the disparity map of the first layer. One pixel distance at this layer corresponds to 8 pixels at the finest layer. Thus, the disparity map is coarse in both spatial resolution and accuracy. The amount of computation is thus reduced substantially.

Now we move to the 2nd layer. The spatial resolution and the accuracy are twice those of the previous layer. We repeat stereo matching by the occlusion-overcoming strategy. But the searching range is restricted to a set consisting of the estimates in the matching window at the corresponding position of the previous layer and their neighboring disparities (See Fig.8).

We repeat the same processing as of the 2nd layer for the successive layers. After obtaining the disparity map of the last layer, post-filtering to eliminate undesirable errors is applied and we get the disparity map.

In Fig.9, we show the disparity map obtained by the coarse-to-fine hierarchical method. We see quite a sharp disparity map is obtained by the method.

The computational burden is considerably reduced when we compare with the fine-to-fine hierarchical method. The quality of the disparity map is not much degraded.

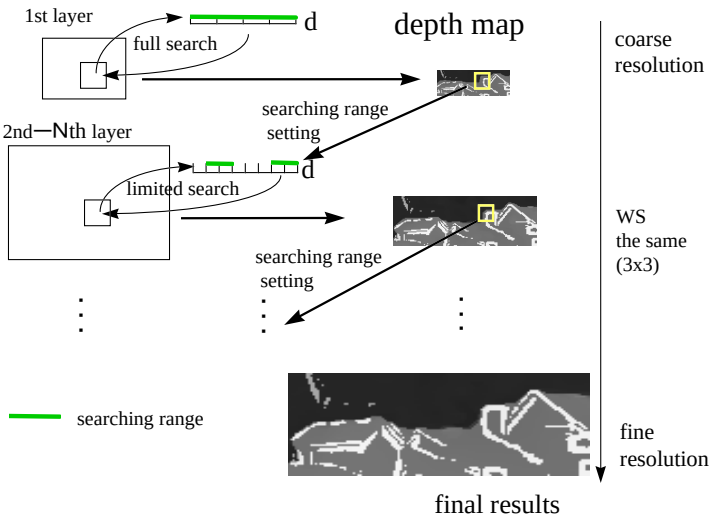

Figure 8: Illustration of the coarse-to-fine hierarchical scheme. Searching range is restricted to the estimates in the matching window of the previous layer except for the 1st layer.

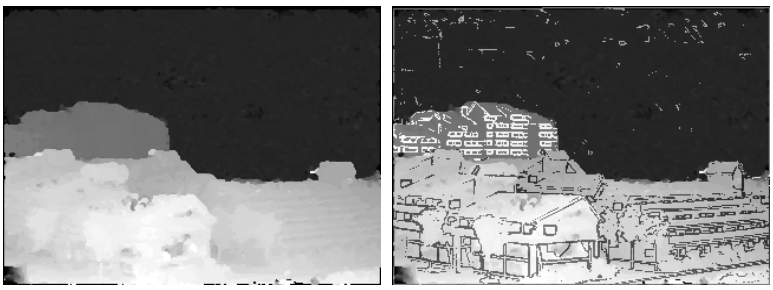

Figure 9: Disparity map obtained by the coarse-to-fine hierarchical method [left]. Edge map is overwritten [right].

4 Experimental Results

4.1 Synthetic Images

We have tested the proposed method using synthetic images with 64[pixels]x64[lines] and 8-bit gray-scale resolution. Each object is characterized by average intensity and texture. Textures and noises are added by i.i.d. Gaussian random number generator.

Figure 10 shows a result. The standard deviation of all the texture signals is 5.0 and that of noise signal is 3.0. The average intensity of each area is 178, 128, and 78, from center to boundary of the image, respectively. In this way, we simulate a realistic situation where the intensities within an object are highly correlated. The true disparities of each area is 7, 2, and 0 [pixels], respectively. We see the occlusion-overcoming strategy shows clear boundary. The amount of overreach is exactly half the size of matching window. Thus, a hierarchical implementation using the strategy shows a good result. However, without the strategy, boundaries are smoothed and thus, a hierarchical implementation does not seem to be of benefit to the result. These results are consistent with our assertion in Sec. 3.2.

Throughout many experiments with varying the parameters of synthetic images, we observed that

- the occlusion-overcoming strategy produces sharp disparity map due to its nonlinearity,
- the boundary overreach of the occlusion-overcoming method is less than half of the size of the matching window,
- the noise immunity is slightly lowered by introducing the occlusion-overcoming method when we fix the size of the matching window,
- performance of the hierarchical implementation with the occlusion-overcoming strategy is consistently superior to that without the method except unrealistic situations. In the unrealistic situations, the performance is almost the same.

Here, unrealistic situations refer to when the noise level is extremely high compared with the signal level and the correlation of the intensities within an object is very small.

4.2 Real Images

We have tested the proposed algorithm using a variety of indoor images with 640[pixels]x480[lines] and 8-bit gray-scale resolution. They include some eye-array images in the image database of University of Tsukuba and some indoor images shot in our laboratory (Fig. 11). We use single camera with varying positions and thus simulate multiple cameras to validate the method.

We show some of the results obtained by the hierarchical methods in Fig. 12 and Fig. 13.

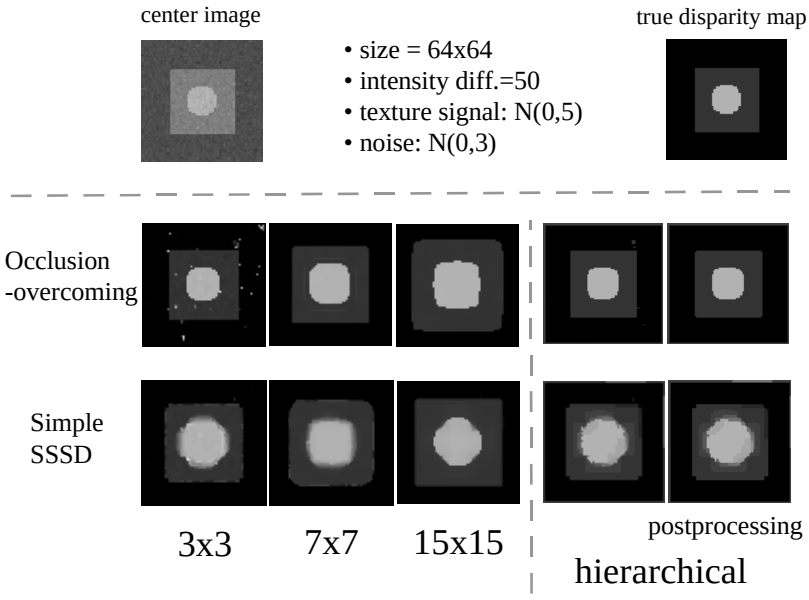

Figure 10: Simulation results with synthetic images.

Fig. 12 is the disparity maps of "Santa" image in the database of Univ. of Tsukuba. The image is shot by SONY 3CCD Video Camera(XC-003) mounted on a robot arm. The focal length is 10[mm], the CCD size is 1/3[inch], and the baseline stretch is 20[mm]. The distance to the nose of the stuff is 75[cm]. The left maps are the results of fine-to-fine hierarchical method. Searching range is set to 50 pixels and 4-layer hierarchy(31x31 to 15x15 to 7x7 to 3x3) is used. The right maps are the results of coarse-to-fine hierarchical method. Searching range at the top layer is set to 8 pixels. We see the object edges exactly coincide with disparity edges and the surfaces of disparity map are very smooth in both of the disparity maps. In the no texture areas, for example, the foot of the stuff, we see undesirable errors, which is the fundamental limitation of the area-based matching.

Fig. 13 is the disparity maps of "Lab" image shot by SONY 3CCD video camera(DXC-930) in our laboratory. The focal length is 7.5[mm], the CCD size is 1/2[inch], and the baseline stretch is 50[mm]. The distance to the nearest object(foot of small bear) is 75[cm]. Estimation parameters are the same as that of the former image. The left maps are the results of fine-to-fine hierarchical method. The right maps are the results of coarse-to-fine hierarchical method. We can confirm the methods produce sharp and correct disparity maps.

With the above configuration, the matching accuracy of 1 pixel corresponds to 1.5, 8.9, 34 [cm] for objects located at 75, 188, 375 [cm] from camera, respectively. We

Figure 11: Images used in the experiment. “Santa” image [upper]. “Lab” image [lower].

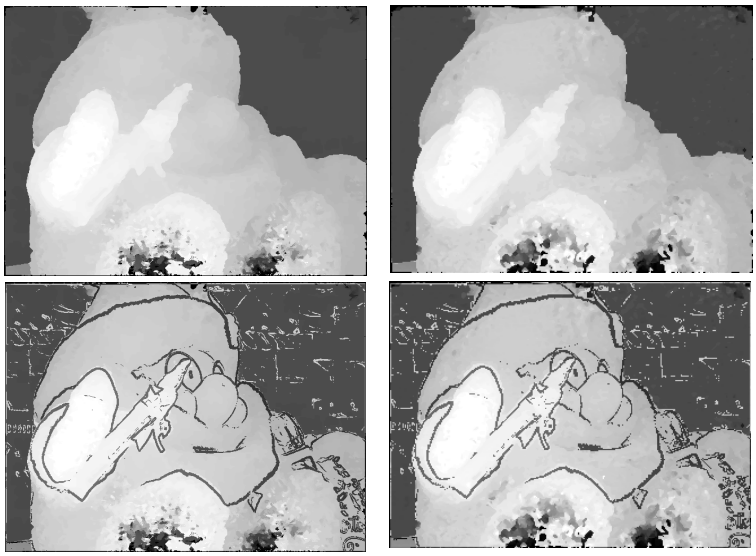

Figure 12: Disparity maps of "Santa" image with fine-to-fine hierarchical method [left] and with coarse-to-fine hierarchical method [right]. Edges are overwritten in the lower maps.

can cope with large objects by increasing the baseline stretch according to the size and distance of the objects. The accuracy is considered to be sufficient for our application since the matter is not accuracy but the exact shape of objects in a complex scene.

Throughout many experiments, we can confirm that both of the methods works very well for a variety of scenes. It achieves the correctness of disparity map around object boundaries. The coarse-to-fine approach shows slightly inferior performance to those of fine-to-fine approach as is expected. However, the difference is not substantial compared with the gap of the performance between the hierarchical methods in this paper and some methods without occlusion-overcoming/hierarchical strategy.

5 Concluding Remarks

We have investigated on obtaining a sharp and dense depth map from multiple cameras. The method is based on a simple selection of disparity information from 5 cameras implemented on a novel hierarchical estimation scheme. By using occlusion-overcoming strategy, we have reduced the harmful effects of occlusion considerably. Furthermore, based on the unique property of boundary overreach of the strategy, we have constructed a hierarchical estimator. A fine-to-fine implementation and a coarse-to-fine implementation are presented. We have confirmed that both of the implementation achieve shape-correctness of depth map on the one hand, alleviate the problem of lack of texture on

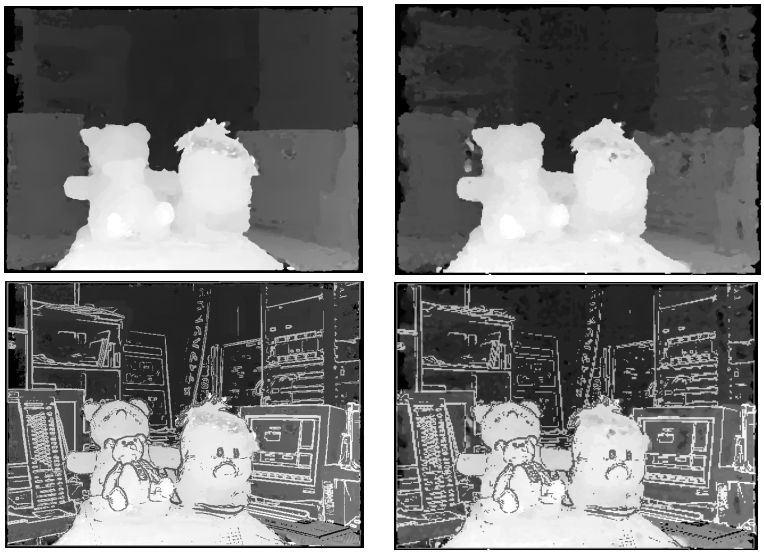

Figure 13: Disparity maps of "Lab" image with fine-to-fine hierarchical method [left] and with coarse-to-fine hierarchical method [right]. Edges are overwritten in the lower maps.

the other hand. The coarse-to-fine implementation reduces computational burden substantially but shows slightly inferior performance compared with the fine-to-fine one. The computation of the proposed method is regular and parallelizable and thus it is well suited for a hardware implementation.

Since the obtainable performance of the estimation depends on images, it would not be unusual for the obtained one not to come up to the desired one. In such cases, some interactive interface should be prepared to fill the gap of the quality. Thus, we are currently developing a user-friendly interface for post-processing of the estimation.

Moreover, we are developing application techniques using the depth map obtained [14], which include arbitrary view generation for image-based rendering, Z-key method for 3D video composition, automatic multi-layer description of a scene.

We have focused on only the spatial characteristics of a scene in this paper such that all of the processing is executed by the frame. Temporal property has not been considered. In order to obtain more satisfactory results, we may need to develop an integrated approach of spatial and temporal property of a scene. Therefore, a paradigm of motion and structure from multiple-baseline stereo would be very interesting as a future work.

References

[1] S.D.Cochran and G.Medioni, "3-D surface description from binocular stereo," *IEEE Trans. PAMI*, vol.14, no.10, pp.981-994, Oct. 1992.

[2] U.Dhond and J.Aggarwal, "Structure from stereo: A review," *IEEE Trans. System, Man, and Cybernetics*, vol.19, no.6, pp.1489-1510, Nov./Dec. 1989.

[3] W.E.L.Grimson, "A computer implementation of a theory of human stereo vision," *Phil. Trans. Royal Soc. London*, vol.B292, pp.217-253, 1981.

[4] M.J.Hannah, "Bootstrap stereo," *Proc. ARPA Image Understanding Workshop*, pp.201-208, College Park, MD, Apr. 1980.

[5] S.Inoue, "Mental image expression by media integration - COMICS," *Proc. of 1st International Workshop on New Video Media Technology*, pp.47-52, Seoul, Korea, March 1996.

[6] T.Kanade and M.Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Trans. PAMI*, vol.16, no.9, pp.920-932, Sept. 1994.

[7] T.Kanade *et al.*, "Virtualized Reality: Concepts and early results," *Proc. IEEE Workshop on Representation of Visual Scenes*, pp.69-76, June 1995.

[8] T.Kanade *et al.*, "A stereo machine for video-rate dense depth mapping and its new applications," *Proc. IEEE CVPR'96*, pp.196-202, San Francisco, June 1996.

[9] Y.Nakamura *et al.*, "Occlusion detectable stereo - Occlusion patterns in camera matrix," *Proc. IEEE CVPR'96*, pp.371-378, San Francisco, June 1996.

[10] V.S.Nalwa, *A Guided Tour of Computer Vision*, Addison-Wesley, 1993.

[11] Y.Ohta, "Computer vision as media technology," *Proc. Image Sensing Symposium*, pp.265-270, 1996 (in Japanese).

[12] M.Okutomi and T.Kanade, "A multiple-baseline stereo," *IEEE Trans. PAMI*, vol.15, no.4, pp.353-363, April 1993.

[13] J.Park *et al.*, "Extraction of depth information for scene description and its application," *ITE'96*, pp.112-113, Nagoya, Japan, July 1996 (in Japanese).

[14] J.Park and S.Inoue, "Image expression based on disparity estimation from multiple cameras," *Proc. 3rd Joint Workshop on Multimedia Communications*, 7-1, Taegu, Korea, Oct. 1996.

[15] J.Park and S.Inoue, "New view generation from multi-view image sequence," *Technical Report of IEICE*, IE96-121, pp. 91-98, Sapporo, Japan, Feb. 1997 (in Japanese).

[16] D.Scharstein, *View Synthesis Using Stereo Vision*, Ph.D. Thesis, Cornell Univ., Jan. 1997.

[17] M.Shibata *et al.*, "Scene describing method for video production," *ITEJ Tech. Report*, vol.16, no.10, pp.19-24, Jan. 1992 (in Japanese).

[18] R.Tsai, "Multiframe image point matching and 3-D surface reconstruction," *IEEE Trans. PAMI*, vol.5, no.2, pp.159-174, March 1983.

[19] R.Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol.RA-3, No.4, pp.323-344, Aug. 1987.

[20] K.H. Yang, S.G. Lee, and C.W. Lee, "Image Restoration of Noisy Images Using OS Filters with Adaptive Windows," *J. Korean Insti. of Telematics and Electronics*, vol. 27, no. 1, pp.112-119, Jan. 1990 (in Korean).