SICE Annual Conference 2007
Sept. 17-20, 2007, Kagawa University, Japan

Multi-baseline Stereo by Maximizing Total Number of Interest Points

Tomokazu Sato and Naokazu Yokoya

Graduate School of Information Science, Nara Institute of Science and Technology, Japan
(E-mail: tomoka-s@is.naist.jp)

Abstract: This paper proposes a novel method for estimating depth without similarity measures such as SSD and NCC. Our idea for estimating a depth map is very simple; only counting interest points in images is integrated with the framework of multi-baseline stereo. Even by a simple algorithm, depth can be determined without computing similarity measures such as SSD and NCC that have been used for conventional stereo matching. The proposed method realizes robust depth estimation against occlusions with lower computational cost. Though a naive TNIP based method can realize fast and robust depth estimation, the accuracy of estimated depth is lower than one by SSSD based method because TNIP uses sparse data. In this paper, we also show that accuracy of depth estimation can be increased by combining TNIP based method and SSSD based method. In experiments, the validity and feasibility of our algorithm are demonstrated for both synthetic and real outdoor scenes.

Keywords: Image measurement, Multi-baseline stereo, Interest point.

1. INTRODUCTION

Depth map estimation from images is one of very important theme in computer vision, because depth information is used in a number of different applications such as 3-D reconstruction, surveillance, and new view synthesize. In this paper, we focus on the framework of multi-baseline stereo for moving camera system that is one of the standard methods for depth estimation from a large number of image input.

The original multi-baseline stereo method [1] was proposed by Okutomi and Kanade for multiple image input. The multi-baseline stereo has such a very good feature that an arbitrary number of images can be simultaneously used for depth estimation. This increases the accuracy of depth estimation and decreases the ambiguity in stereo matching. By recent development of camera calibration techniques, some researchers have employed the multi-baseline stereo framework for a freely moving video camera [2], [3], [4], [5], [6]. A freely moving video camera is suitable for 3-D modeling of a large scale environment because it easily makes a long distance baseline between cameras. However, there exist some problems in multi-baseline stereo method for a moving video camera as follows:

(1)Inaccurate depth estimation due to occlusions: When a point on an object where depth should be estimated is occluded by the other objects in a part of an input video, the occluder gives a negative score to the score function of the multi-baseline stereo: SSSD (Sum of SSD). This negative score prevents the algorithm from obtaining correct estimation of depth map around occlusions.

(2)High computational cost: However utilization of a large number of input images increases accuracy of depth estimation, it consume the large amount of memory and computational resources. Some patches for the occlusion problem may also increase the computational time.

To avoid these problems, we propose a new approach that estimates depths by only counting interest points as shown in Figure 1 that are corners and cross points of edges in video images. The framework of our depth estimation is basically the same as the original multi-baseline stereo except for a newly employed objective function: TNIP (Total Number of Interest Points). The idea is based on the assumption that the corners of objects and cross points of texture edges in the 3-D space (3-D interest points) will appear in video images as 2-D interest points at the projected positions of the 3-D interest points. By searching a depth that maximizes the total number of 2-D interest points under epipolar constraint, the depth can be determined as a position of a 3-D interest point.

Fig. 1 Examples of interest points.

By using the new objective function TNIP for depth estimation, the problems pointed out earlier can be solved; (1) the new score function TNIP is not significantly affected by occluders, (2) computational cost is drastically decreased because depth can be determined by only counting interest points. However we cannot estimate depths for pixels where any interest points exists, that is not a critical problem for 3-D modeling and some other applications because usually 3-D interest points contain corners of the 3-D model. In most cases, depth interpolation is sufficient. Also we should note that TNIP based method estimates the depth for the 3-D corners

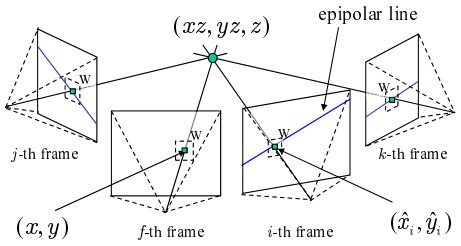


Fig. 2 3-D position of (x, y) with depth z and its projected line to each frame.

rather than the depth for the target pixel. Thus, the accuracy of depths given by raw TNIP function is little lower than that by SSSD, while TNIP drastically decreases computational cost. In this paper, to solve weakness of raw TNIP function, we also suggest the hybrid approach in which both SSSD and TNIP are used. In the hybrid approach, first, TNIP is used to roughly determine the depth for each interest point. For limited searching range determined by TNIP, depth is re-searched by SSSD.

The rest of this paper is structured as follows. First, the original multi-baseline stereo method for a moving video camera is briefly described in Section 2. In Section 3, the new score function TNIP for multi-baseline stereo is proposed. Each process for estimating a dense depth map is detailed in Section 4. Experimental results with simulation and a real scene show the validity and feasibility of the proposed method in Section 5. Finally, Section 6 describes conclusion and future work.

2. MULTI-BASELINE STEREO BY USING SSSD

In this section, firstly, coordinate systems of a general moving camera are defined. The principle of the multi-baseline stereo [1] using SSSD is then briefly summarized.

2.1 Definition of coordinate systems for moving camera

In the multi-baseline stereo, as shown in Figure 2, a depth z of a pixel (x, y) in the f-th frame is estimated by using images from the j-th to k-th frame ($j \leq f \leq k$). In the following, for simplicity, we assume that the focal length is 1 and lens distortion effect has already been corrected by known intrinsic parameters. In this case, a 3-D position of (x, y) with depth z is represented as (xz, yz, z) in the camera coordinate system of the f-th frame. The 3-D position (xz, yz, z) is projected to the position $(\hat{x}_i, \hat{y}_i)$ in the image of the i-th frame by the following expression.

$$\begin{pmatrix} a\hat{x}_i \\ a\hat{y}_i \\ a \\ 1 \end{pmatrix} = \mathbf{M}_{fi} \begin{pmatrix} xz \\ yz \\ z \\ 1 \end{pmatrix}, \tag{1}$$

where a is a parameter, $\mathbf{M}_{fi}$ denotes a 4×4 transformation matrix from the camera coordinate system of the f-th frame to the camera coordinate system of the i-th frame. In the multi-baseline stereo, as shown in Figure 2, the point $(\hat{x}_i, \hat{y}_i)$ is constrained on the epipolar line, which is the projection of the 3-D line connecting the position (xz, yz, z) and the center of projection in the f-th frame onto the i-th frame.

2.2 Depth estimation using SSSD

In the traditional multi-baseline stereo, depth z of pixel (x, y) is determined by using the similarity measure SSD. The SSD is computed as the sum of squared differences between two image patterns that have a certain size W. The SSD for (x, y) in the f-th frame and $(\hat{x}_i, \hat{y}_i)$ in the i-th frame is defined using image intensity I as follows.

$$SSD_{fixy}(z) = \sum_{(u,v)\subseteq W} \left\{ I_f(x+u, y+v) - I_i(\hat{x}_i+u, \hat{y}_i+v) \right\}^2. \tag{2}$$

To evaluate the error of the depth z for all the input images, the SSD is summed up as follows:

$$SSSD_{fxy}(z) = \sum_{i=j}^{k} SSD_{fixy}(z). \tag{3}$$

The depth z is determined for each frame so as to minimize the SSSD function. Generally, to find a global minimum of the SSSD, depth z should be searched for all the depth range along a 3-D line from a reference pixel (x, y).

If the pixel (x, y) in the f-th frame is occluded by other objects in the i-th frame, $SSSD_{fxy}(z)$ for the true depth z is increased by the occluder because $SSD_{fixy}(z)$ gives a large error. Thus, to obtain a correct depth at such an occluded part, some other computationally expensive extensions should be added to the original multi-baseline stereo. For example, a modified SSSD can be computed by summing up only lower halves of SSDs [5], [7]. However, there still remains the computational cost problem.

3. MULTI-BASELINE STEREO BY COUNTING INTEREST POINTS

In this section, a new score function TNIP is defined to estimate depth z of pixel (x, y) using the multi-baseline stereo framework. Generally, feature points in a 3-D space, such as corners of objects and cross points of texture edges, appear as 2-D feature points in images at projected positions of the 3-D feature points. Such a 2-D feature points can be easily detected by interest operators such as Harris's [8] and Moravec's [9] operators.

In this study, depth z is determined so as to maximize the TNIP score function that is defined as follows.

$$TNIP_{fxy}(z) = \sum_{i=j}^{k} \sum_{(u,v)\subseteq W} H_i(\hat{x}_i+u, \hat{y}_i+v). \tag{4}$$

$$H_i(u,v) = \begin{cases} 1 \text{ ; interest point exists at } (u,v) \text{ in } i\text{-th frame.} \\ 0 \text{ ; otherwise} \end{cases} \tag{5}$$

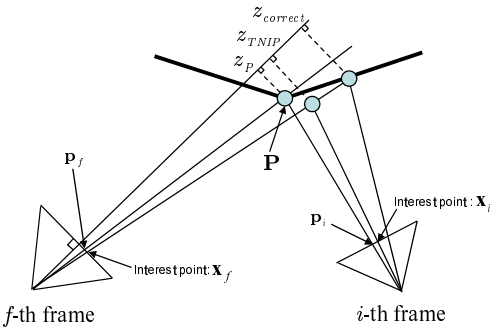


Fig. 3 Cause of estimation error for TNIP.

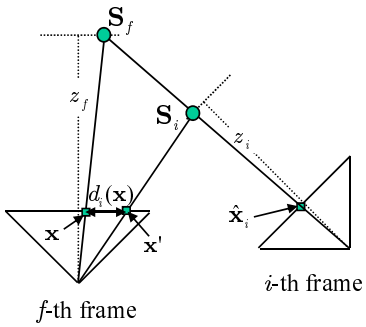


Fig. 4 Elimination of outliers.

The TNIP score represents the total number of interest points that exist in $(\hat{x}_i, \hat{y}_i)$ centered windows W for all the frames. Note that the size of W should be appropriately small because interest points are not detected at positions far from projected positions of (xz, yz, z) when there exists a feature point in 3-D space.

By using the TNIP instead of the SSSD function in the multi-baseline stereo, computational time can be drastically decreased because the time consuming process of comparing intensity patterns can be removed from the depth estimation. Moreover, the TNIP has another good feature that the TNIP is not significantly influenced by occluders because it counts only positive scores. These claims will be justified by experiments later.

4. DEPTH ESTIMATION FROM AN IMAGE SEQUENCE

This section describes the processes of depth estimation. In our method, after detecting the interest points in all the input images, depths of all the interest points are determined by the multi-baseline stereo framework with the TNIP score function. Next, outliers of estimated depths are then eliminated by using their confidences defined by considering the consistency among the results in multiple frames.

4.1 Depth estimation for interest points

Depths of all the interest points detected on video images are computed by maximizing the TNIP score function that is defined in Section 3. The depth z is searched along a 3-D line from each target pixel to find a maximum TNIP in a given range of depth. By repeating the estimation of depth z for all the interest points in the input image sequence, sparse depth data can be acquired. Note that any intensity images are not needed for TNIP based depth estimation. Only 2-D positions of interest points and camera parameters should be stored to compute TNIP. It means that our method needs only 1/8 memory space to compute depth by compared with SSSD, if 8 bit grayscale images are assumed to be used for SSSD. However TNIP based method can estimate depths with low computational cost, raw TNIP function has a weak point that accuracy for estimated depth is little worse than that by SSSD. The following paragraphs describe the reason and solution for this weakness.

As illustrated in Figure 3, TNIP based method estimates depth value by using interest points that are detected around projected position $\mathbf{p}_i (j \leq i \leq k)$ of the 3-D interest point $\mathbf{P}$. However, detected position $\mathbf{x}_i$ of interest point $\mathbf{P}$ in the i-th frame is not always coinside with the projected position $\mathbf{p}_i$ due to feature detection error. For the target frame f, there also exists detection error for the pixel $\mathbf{x}_f$. Thus, searching line for depth of $\mathbf{x}_f$ may not cross the 3-D point $\mathbf{P}$. As shown in Figure 3, in this case, although the correct depth for the pixel $\mathbf{x}_f$ is $z_{correct}$, TNIP function is maximized at z_{TNIP} that is near to the depth z_P of the 3-D position $\mathbf{P}$. This is caused by a characteristics of the TNIP function; TNIP based method estimates the depth for the 3-D corners rather than the depth for the target pixel. At the same time, SSSD based method does not have such a problem.

For this problem, one solution is to refine the estimated depth using SSSD. More concretely, after TNIP based depth estimation, limited range $(z_{TNIP} - Cl < z < z_{TNIP} + Cl)$ can be re-scanned by SSSD. If we employ this hybrid approach, however the advantage for little memory requirement is vanished, computational efficiency is maintained because searching range of depth for SSSD is much limited in refinement process. Moreover, because TNIP function is not significantly affected by occlusions, the hybrid approach can estimate robust and accurate depth.

4.2 Elimination of outliers

In this process, unreliable depths are eliminated by cross validation approach for multiple image input. Figure 4 illustrates an example setting of two cameras. The depth z_f of the pixel $\mathbf{x}$ is evaluated by consistency check of the estimated depths. First, for the i-th frame, the projected position $\hat{\mathbf{x}}_i$ of the 3-D position $\mathbf{S}_f$ that corresponds to the depth z_f is computed. Inversely, for the f-th frame, the projected position $\mathbf{x}'$ of the 3-D position $\mathbf{S}_i$ that corresponds to the depth z_i is computed. Consistency for the depth z_f and z_i is then evaluated by the distance $d_i(\mathbf{x}) = |\mathbf{x} - \mathbf{x}'|$ in the target frame. In this research, confidence $R(\mathbf{x})$ for the depth z_f of the pixel $\mathbf{x}$ is defined as followings using distance d_i.

$$R(\mathbf{x}) = \frac{\sum_{i=j}^{k} \{0; d_i(\mathbf{x}) > T, 1; d_i(\mathbf{x}) \leq T)}{k - j + 1}, \quad (6)$$

where T is a threshold for the distance d_i that judges whether the depth z_f is consistent with the depth z_i or not. The confidence $R(\mathbf{x})$ indicates a rate of consistent depth pairs. If all the depths are correctly estimated, $R(\mathbf{x})$ becomes maximum value 1. In our method, the depth z_f of the pixel $\mathbf{x}$ whose confidence $R(\mathbf{x})$ is lower than given threshold U is regarded as an outlier, and is deleted.

5. EXPERIMENTS

We have carried out two kinds of experiments. One is concerned with the comparison of SSSD, TNIP and HYBRID method in computer simulation. The other is conducted for depth map estimation for a real outdoor environment. For all the experiments in this section, Harris interest operator [8] is employed as an interest point detector.

5.1 Quantitative evaluation in computer simulation

In this section, first, configuration of the computer simulation is detailed. Next, to determine the best window size for SSSD, TNIP and HYBRID, preliminary experiment is carried out. Finally, by using the best window size for each method, accuracy and computational efficiency are compared.

5.1.1 Setup of simulation

In the experiment, two textured planes are located in a virtual environment, and a virtual camera takes an image sequence by moving the camera around these planes. Two kinds of texture patterns are used for the planes, as shown in Figure 5. The layout of the planes and the motion path of virtual camera are illustrated in Figure 6. Totally 91 input images, some of which are shown in Figure 7, are taken by the moving camera whose motion draws a quarter circle as illustrated in Figure 6. By the motion of the camera, the plane 1 is occluded by the plane 2 in after half of the input images, and both textures are apparently distorted by the camera motion. To take into account camera calibration errors about intrinsic and extrinsic camera parameters, the Gaussian noise with standard deviation σ is added to the projected positions of the 3-D points and these positions are sampled to pixels. The other parameters that are used for this experiment are shown in Table 1.

5.1.2 Determination of window size

In this preliminary experiment, the best size of window W in Eqs. (2) and (4) are determined by using the conditions described in the previous paragraph. To determine the size of window W, we have evaluated the rate of inaccurate depth based on the average re-projection error that is defined as follows.

$$E_p = \frac{1}{N}\sum_{i=1}^{N} |\hat{\mathbf{x}}_{ip} - \bar{\mathbf{x}}_{ip}| \tag{7}$$

where p is a pixel index and N is a number of images that are used for depth estimation and is set as 91 in this experiment. $\hat{\mathbf{x}}_{ip}$ is a projected position of the estimated depth z in the i-th frame for the pixel p, and $\bar{\mathbf{x}}_{ip}$ is the projected position of the ground truth. In this experiment, if E_p is over 1.0 pixel for the pixel p, the estimated depth for the pixel p is judged as inaccurate.

Figure 8 shows the rate of inaccurate depths for various window size and various noise level. As shown in this figure, although the best window size of TNIP is 3×3 pixels, the rate of inaccurate depth is greater than that of SSSD for 7×7 size that is the best size for SSSD. As described in Section 4.1 , it is difficult for TNIP to estimate high accurate depth due to detecting error of feature positions in the target frame.

On the other hand, in the HYBRID approach, depths are refined after TNIP based estimation using SSSD func-

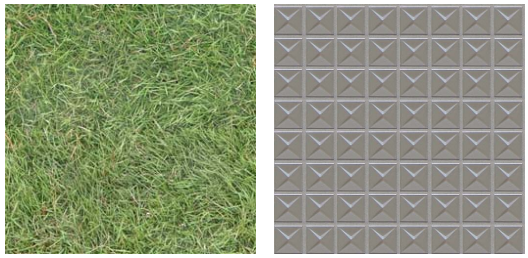
Table 1 Given parameters for simulation.

(a) for depth estimation.

Searching range of depth [mm]	3,000 - 35,000
Re-scanning range coefficient C	10

(b) for outlier elimination.

Threshold of distance T [pixel]	1.0
Threshold of confidence U	0.4



(a) plane 1 (b) plane 2

Fig. 5 Textures of planes.

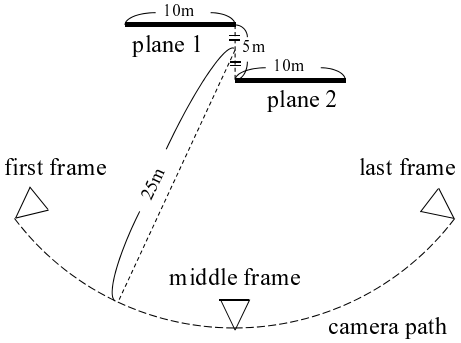
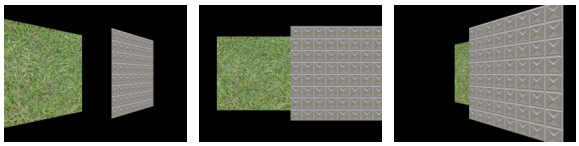


Fig. 6 Layout of planes and camera path in simulation.



(a) first frame (b) middle frame (c) last frame

Fig. 7 Sampled frames from 91 input images.

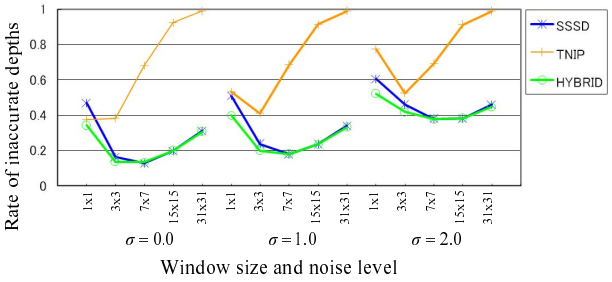


Fig. 8 Rate of inaccurate depths ($E \geq 1.0$pixel) for various window size and noise level.

tion. For the HYBRID approach, horizontal axis in Figure 8 indicates window size of SSSD in the refinement process, and window size of TNIP for initial estimate is fixed as 3×3 pixels. From this figure, it is confirmed that 7×7 pixels is the best size for HYBRID and the rate of inaccurate depths are almost same with SSSD.

Table 2 indicates the average time to estimate a depth of a single pixel using all the 91 input images with respect to different window sizes. The computation time is measured by using a PC (CPU: Pentium-4 Xeon 3.20GHz dual, Memory: 2GB). From this table, we can confirm that the computational costs of TNIP (3×3 window) and HYBRID (7×7 window) are about 9 times and 5 times cheaper than that of SSSD (15×15 window), respectively. Note that all the implementation of these methods are same except for the objective function.

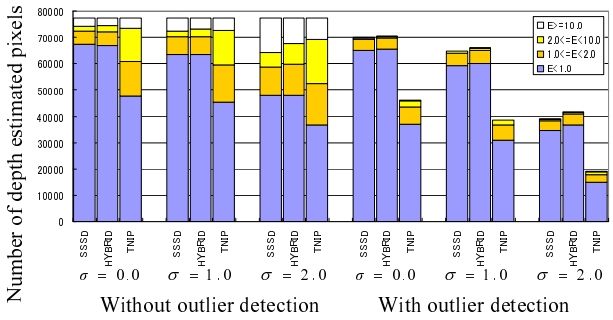
Table 2 Average computational time for estimating a depth of a single pixel [milli-seconds].

W	1×1	3×3	7×7	15×15	31×31
SSSD	13.6	25.2	**86.3**	353.9	1530
TNIP	9.0	**9.8**	11.2	13.0	21.2
HYB.	10.3	11.5	**16.7**	40.1	141.3

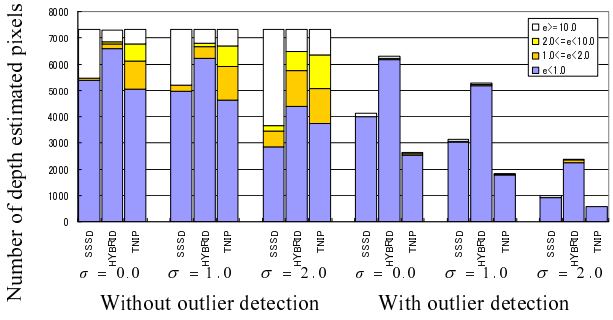
5.1.3 Accuracy comparison of TNIP, SSSD and HYBRID

In this experiment, TNIP, SSSD and HYBRID method are compared by using the best window size for each method. To analyze characteristic of each method for the occlusions, we divide all the image region (ALL) to occluded region (OCC) and other region (NOR). In this experiment, the region where the pixel in the target frame is occluded in more than half of reference images is categorized to OCC.

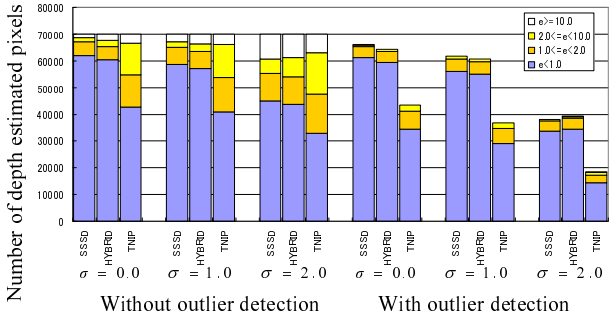
Figure 9 shows stacked bar chart that indicates breakdown of estimated errors. Vertical axis in this figure indicates number of estimated depths and each chart are separated by the level of average re-projection errors. For the horizontal axis, combination of outlier elimination (with or without), noise level σ and objective functions (SSSD, TNIP and HYBRID) are specified. As shown in Figure 9(a), before outlier elimination, the rate for accurate depths ($E < 1.0$pixel) for region ALL are almost same for SSSD and HYBRID, and that of TNIP is worse than them in this case. For large errors ($E \geq 10.0$pixels), although the rate is almost same level for all the methods



(a) All region: ALL



(b) Occluded region: OCC

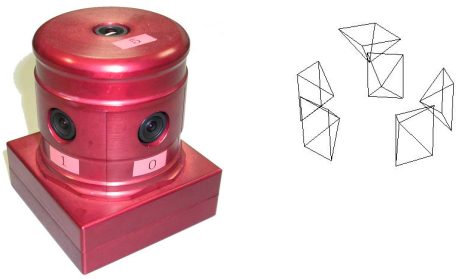


(c) Outside of occluded region: NOR

Fig. 9 Analysis of estimation errors for each region.

when noise level is low, if noise level is high ($\sigma = 2.0$), more large errors of TNIP are suppressed than others. After outlier elimination, for all the methods, most of inaccurate results ($E \geq 2.0$pixels) are eliminated. It shows effectiveness of outlier elimination described in Section 4.2.

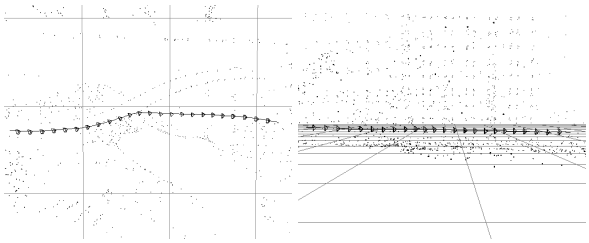
Figures 9(b)(c) show the error rate for occluded region (OCC) and other region (NOR). Note that scale of the vertical axis for OCC is different with other graphs because the occupancy rate of region OCC for region ACC is about 0.1. From the comparison of (b) and (c), we can confirmed that the error rate for SSSD in the region OCC is much higher than that in the region NOR. In contrast, the error rates of TNIP and HYBRID for the region OCC are not drastically changed from those for the region NOR. These results verify robustness of our proposed method for occluded region.



(a) appearance (b) view volume

Fig. 10 Omni-directional multi-camera system: Ladybug.

Fig. 11 Sampled frame of input image sequences.



(a) top view (b) side view

Fig. 12 Camera path of OMS used for input (29m).

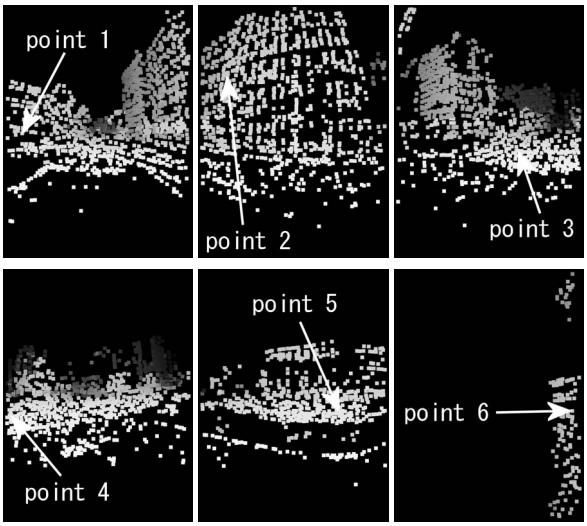


Fig. 13 Result of depth estimation for interest points in Figure 11.

5.2 Depth estimation in an outdoor environment

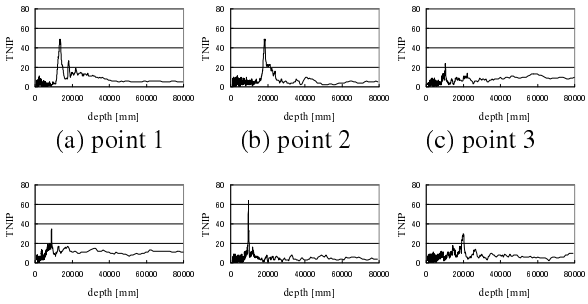
In this experiment, an outdoor environment is captured by an Omni-directional Multi-camera System (OMS): Ladybug [11]. Figure 10 shows an appearance and view volume of Ladybug. This camera system has six radially located camera units and takes synchronized six image sequences at 15fps (resolution of each camera: 768×1024 pixels).

First, the outdoor environment was captured by the OMS as 3,000 images (500 frames). Figure 11 shows a sampled frame of six input image sequences. Intrinsic camera parameters including geometric relations among fixed camera units are calibrated in advance by using a marker board and a 3-D laser measure [12]. Extrinsic camera parameters of the input image sequences are estimated using bundle adjustment by tracking both a small number of feature landmarks of known 3-D positions and a large number of natural features of unknown 3-D positions in input images across adjacent camera units [13]. Figure 12 illustrates the recovered camera path that is used as an input for depth estimation. The curved line and pyramids denote the motion path of a camera unit and its posture at every 20 frames, respectively. The length of the camera path is approximately 29m. The accuracy of estimated camera path is evaluated as 50mm about camera position and 0.07degree about camera posture [13].

By using input images and estimated camera parameters, omni-directional depth maps are actually estimated by the method HYBRID. After detection of interest points for all the frames of six input image sequences by Harris operator, depths of all the interest points are estimated using the TNIP score function. In this experiment, 1,750 interest points are detected on average in a single input image (10,500 points per frame). Interest points in the $(f-100)$-th to the $(f+100)$-th frames at every 2 frames (606 images,101 frames) are used to estimate depth data for the f-th frame. The size of window W was set as 7×7 according to the result of the computer simulation. The searching range to find a maximum TNIP in this stage is 1,000mm (near) to 80,000mm (far).

Next, low confidence depths are eliminated. The thresholds for T and U for outlier detection were set as 2.0 pixels and 0.3, respectively. In this experiment, about half of the estimated depths are rejected as outliers. Figure 13 shows the results of depth estimation for the images in Figure 11. In this figure, depth values are corded in intensity. Figure 14 indicates TNIP values of randomly selected six interest points in Figure 13. We can confirm from Figure 14 that each TNIP plot has a single apparent peak at a certain depth value and there are no other comparable peaks. This clearly shows that depth estimation can be easily achieved for these interest points.

Finally, omni-directional dense depth maps are generated using depth interpolation. Figure 15 shows a panoramic image that is generated from six input images



(a) point 1 (b) point 2 (c) point 3

(d) point 4 (e) point 5 (f) point 6

Fig. 14 TNIP scores for searching depth.

shown in Figure 11. Figure 16 shows the corresponding dense depth map. By comparing these figures, we can confirm that depth map is correctly computed for most parts of the input image. However, some incorrect depths are also observed around the boundaries between the buildings and the sky. These incorrect results are caused by depth interpolation over different objects. To improve the result, region information in input images should be considered for interpolation.

6. CONCLUSION

In this paper, a novel multi-baseline stereo for a moving camera has been proposed, where depth can be determined by only counting the number of interest points. However, raw TNIP function has a weakness that high accurate depth estimation is difficult due to feature detection error in the target frame. To solve weakness of raw TNIP function, we also proposed HYBRID approach that refines the estimated depth using SSSD function for limited searching range. Our method has be proven to be robust against occlusions, and the computational cost for the proposed method is also cheaper than the method based on the traditional SSSD function. In experiments, these claims have been justified by using both synthetic and real image sequences. In future work, estimated depth maps will be integrated to reconstruct a 3-D model of a large outdoor environment.

Fig. 15 Panoramic image generated from six images acquired by OMS.

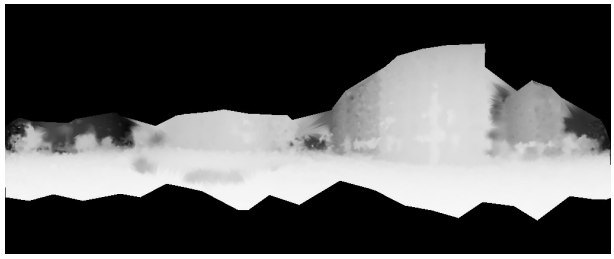


Fig. 16 Generated dense depth map.

REFERENCES

[1] M. Okutomi and T. Kanade: "A Multiple-baseline Stereo," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 4, pp. 353–363, 1993.

[2] S. B. Kang, J. A. Webb, C. Zitnick and T. Kanade: "A Multibaseline Stereo System with Active Illumination and Real-time Image Acquisition," Proc. Int. Conf. on Computer Vision, pp. 88–93, 1995.

[3] S. B. Kang and R. Szeliski: "3-D Scene Data Recovery using Omnidirectional Multibaseline Stereo," Int. Journal of Computer Vision, Vol. 25, No. 2, pp. 167–183, 1997.

[4] W. Zheng, Y. Kanatsugu, Y. Shishikui and Y. Tanaka: "Robust Depth-map Estimation from Image Sequences with Precise Camera Operation Parameters," Proc. Int. Conf. on Image Processing, Vol. II, pp. 764–767, 2000.

[5] T. Sato, M. Kanbara, N. Yokoya and H. Takemura: "Dense 3-D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo Using a Hand-held Video Camera," Int. Journal of Computer Vision, Vol. 47, No. 1-3, pp. 119–129, 2002.

[6] M. Okutomi, Y. Katayama and S. Oka: "A Simple Stereo Algorithm to Recover Precise Object Boundaries and Smooth Surface," Int. Journal of Computer Vision, Vol. 47, No. 1-3, pp. 261–273, 2002.

[7] S. B. Kang, R. Szeliski and J. Chai: "Handling Occlusions in Dense Multi-view Stereo," IEEE Conf. on Computer Vision and Pattern Recognition, Vol.1, pp. 103–110, 2001.

[8] C. Harris and M. Stephens: "A Combined Corner and Edge Detector," Proc. Alvey Vision Conf., pp. 147–151, 1988.

[9] H. Moravec: "Towards Automatic Visual Obstacle Avoidance," Proc. Int. Joint Conf. on Artificial Intelligence, p. 584, 1977.

[10] P. Heckbert Ed.: Graphics Gems IV, pp. 47–59, Academic Press, 1994.

[11] Point Grey Research Inc.: "Ladybug," http://www.ptgrey.com/.

[12] S. Ikeda, T. Sato and N. Yokoya: "High-resolution Panoramic Movie Generation from Video Streams Acquired by an Omnidirectional Multi-camera System," Proc. IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent System, pp. 155–160, 2003.

[13] T. Sato, S. Ikeda and N. Yokoya: "Extrinsic Camera Parameter Recovery from Multiple Image Sequences Captured by an Omni-directional Multi-camera System," Proc. European Conf. on Computer Vision, Vol. 2, pp. 326–340, 2004.