Paper

Improved Pre-focusing Method and Its Application to Three-Dimensional Space Scanning for Detecting An Object

TAKUMA SASAKI^{*†} Member, HIROSHI HANAIZUMI[†] Non-member

(Received January 23, 2018, revised July 5, 2018)

Abstract: A Three-Dimensional Scanning Method (TDSM) was proposed to detect an object or a person in a three-dimensional Region Of Interested (ROI) without the stereo matching process. The method TDSM included two processes; ROI scanning in the field of view of the multiple cameras and the object detection in the ROI. The former was realized by the theoretically derived two sets of coefficients for perspective transformations; one canceled disparities on a plane between the left and the center images, and the other those on another plane between the right and the center. The ROI was formed at the intersection of the two planes. The latter was performed by detecting residual disparities between the left images and the right images. One-dimensional optical flow was used for the detection. The method TDSM was successfully applied to detect a person walking in the field of view of the multiple cameras.

Keywords: Perspective transformation, Optical flow, Person detection

1. Introduction

Various types of autonomous cars have currently been attracted attention, developed and tested on actual roads. They always sense their surroundings to reach safely their destinations. Many types of sensors such as infrared lasers, cameras, and millimeter-wave radars [1] were used in the sensing process. Computer systems were also used to understand their surroundings and to determine the next actions. In fact, TOYOTA has already developed a practical Pre-Crash Safety System [2]. Camera-based systems have also been developed, *i.e.*, SUBARU's Eyesight and HONDA's Intelligent Night Vision system [3]. These systems commonly obtained 3D road information not only to correct control but also to detect pedestrians entering the street [4] for avoiding car-human accidents. The systems, however, needed a time-consuming stereo matching process [5]~[7] for getting the 3D information. There were at least two approaches to solve the problem. One was to use a sub-system, such as GPGPU, to speed up the processing [8]. The other was to adopt a smart method to avoid the stereo matching process. The former solution was trivial and expensive, though the latter low-cost and flexible. The authors have proposed a such method named as the prefocusing method (PFM) [9] to detect a pedestrian at a prefixed ROI. In the PFM, we needed calibrations for each prefocused ROI each time. A smart method for avoiding the calibration has been required.

In this study, we propose a spatial scanning method TDSM for random positioning of ROIs. Once a camera cal-

ibration is performed, TDSM automatically generates sets of coefficients for the perspective transformation to produce the ROI at any point in the field of view of the multiple cameras. Disparities between the left and right camera images are evaluated by using an optical flow technique [?]. We find an object or a person in the ROI when no disparity is detected in the ROI.

2. Methodology

The proposed method TDSM theoretically derives coefficients for pre-focusing in the 3D ROI for any point in the field of view of multiple cameras, and determines whether the object exists there.

2.1 Creation of ROI Suppose 3 cameras are linearly located, and assume their optical axes are parallel with no relative rotation. Each camera has already calibrated and they have their own coefficients for three to two dimensional (3D-2D) perspective transformation as

$$\begin{cases} p = \frac{a_{11}x + a_{12}y + a_{13}z + a_{14}}{a_{19}x + a_{110}y + a_{111}z + 1} \\ q = \frac{a_{15}x + a_{16}y + a_{17}z + a_{18}}{a_{19}x + a_{110}y + a_{111}z + 1} \end{cases} \quad (1)$$

$$\begin{cases} s = \frac{a_{c1}x + a_{c2}y + a_{c3}z + a_{c4}}{a_{c9}x + a_{c10}y + a_{c11}z + 1} \\ t = \frac{a_{c5}x + a_{c6}y + a_{c7}z + a_{c8}}{a_{c9}x + a_{c10}y + a_{c11}z + 1} \end{cases} \quad (2)$$

$$\begin{cases} u = \frac{a_{r1}x + a_{r2}y + a_{r3}z + a_{r4}}{a_{r9}x + a_{r10}y + a_{r11}z + 1} \\ v = \frac{a_{r5}x + a_{r6}y + a_{r7}z + a_{r8}}{a_{r9}x + a_{r10}y + a_{r11}z + 1} \end{cases} \quad (3)$$

respectively, where, a_{li} ($i = 1 \sim 11$) means coefficients for the 3D-2D perspective transformation from a point (x, y, z)

^{*} Corresponding Author: takuma.sasaki.3h@stu.hosei.ac.jp

[†] Graduate School of Computer and Information Sciences, Hosei University
3-7-2, Kajino, Koganei, Tokyo, 184-8584, Japan

in the global coordinates to a pixel (p, q) in the image captured by the left camera, $a_{ci}(i = 1 \sim 11)$, (s, t) , and $a_{ri}(i = 1 \sim 11)$, (u, v) for the center and the right ones, respectively.

When these cameras C_l , C_c and C_r observe the same rectangular on a plane A at a distance, their images have some disparities among them due to parallax. Suppose a 2D-2D perspective transformation from (p, q) to (s, t) as

$$\begin{cases} s = \frac{b_{11}p + b_{12}q + b_{13}}{b_{71}p + b_{78}q + 1} \\ t = \frac{b_{14}p + b_{15}q + b_{16}}{b_{71}p + b_{78}q + 1} \end{cases} \quad (4)$$

where, coefficients $b_{li}(i = 1 \sim 8)$ are calculated from the four pair of vertices. Using the transformation, we can register the left camera image onto the center one. This means that the transformation (4) removes the relative disparities of an object on the plane above. Thus, cameras C_l and C_c are prefocused on the plane A.

We set another plane B and apply the same process to cameras C_r and C_c with using a 2D-2D perspective transformation

$$\begin{cases} s = \frac{b_{r1}u + b_{r2}v + b_{r3}}{b_{r7}u + b_{r8}v + 1} \\ t = \frac{b_{r4}u + b_{r5}v + b_{r6}}{b_{r7}u + b_{r8}v + 1} \end{cases} \quad (5)$$

where, $b_{ri}(i = 1 \sim 8)$ are coefficients for the 2D-2D perspective transformation from (u, v) to (s, t) . The transformation (5) cancels relative disparities of the object on the plane B as same as (4) does.

Now, we have two sets of pre-focused cameras C_l and C_c on the plane A, and C_r and C_c on B. Suppose planes A and B have an intersecting line, and suppose an object set on the line is observed by all cameras at a time. Let's I_l , I_c and I_r be images captured by cameras C_l , C_c and C_r , respectively, and I_{lc} and I_{rc} be transformed images from I_l by (4) and from I_r by (5), respectively. The object observed in the image I_{lc} and that in I_c have no disparity between them, because the object is on the plane A. Similarly, the object observed in I_{rc} and that in I_c have also no disparity between them, because the object is also on the plane B. In this situation, the object in I_{lc} and that in I_{rc} have no disparity between them. Only object set on the intersecting line of the planes A and B achieves the situation. Thus, the ROI is created on the intersecting line of the planes A and B.

2.2 Scanning the ROI In the proposed method TDSM, the ROI is set on the intersecting line. For scanning the ROI, we have to produce two virtual planes A and B so that their intersecting line is on the point we want to observe. The procedures for performing the pre-focusing using the virtual plane are as follows;

- Step 3: Calculate the position of the four surrounding points in the images I_l and I_c by using (1) and (2).
- Step 4: Calculate the coefficients $b_{li}(i = 1 \sim 8)$ by using the four corresponding point pairs.
- Step 5: Select another four points on the virtual plane B using the same manner as step 2.
- Step 6: Calculate the position of the four surrounding points in the images I_r and I_c by using (3) and (2).
- Step 7: Calculate the coefficients $b_{ri}(i = 1 \sim 8)$ using the same manner as step 4.
- Step 8: Produce images I_{lc} and I_{rc} using the above coefficients

The object is detected at the focusing point when we have no disparity in the ROI area between I_{lc} and I_{rc} . The disparity is evaluated by using an optical flow technique.

3. Experiments and discussion

A web camera, HD WEBCAM C270 (Logicool), was used in the experiment. The image size was 1280×720 . We used two laser range finders, GLM70000 Professional (BOSCH), for determining 3D position a point in global coordinate system.

3.1 Camera calibration We installed three cameras, and the center camera was taken as the origin of the global coordinate system. Figure 1 shows a measurement environment of the 3D coordinates. Three cameras were arranged at intervals of 10 cm. We measured 3D coordinates of a corner point on a checkerboard plane using two laser range finders with baseline length of 233 cm as shown in Fig. ?? . Changing the position of the checkerboard, we observed many points on it with measuring distances from the positions A and B, respectively. The 3D coordinates of the point $K(x = x_k, z = z_k)$ were obtained from the length of three sides of the triangle ABK. Avoiding rank reduction in calculating the coefficients in the perspective transformation, we selected 19 independent points. The coefficients of

Figure 1: Measurement of global coordinates.

Step 1: Select the focusing point for the ROI.

Step 2: Select another four points on the virtual plane so that they surround the focusing point.

Table 1: The coefficients of the 3D-2D perspective transformation for each camera.

	Left camera	Center camera	Right camera
a_1	-8.326	-8.552	-9.278
a_2	-3.536	-3.564	-3.421
a_3	-4.589	-4.870	-5.298
a_4	698.0	776.5	843.0
a_5	0.008145	0.002491	0.07777
a_6	6.940	7.236	7.938
a_7	-2.478	-2.561	-2.701
a_8	-670.3	-705.6	-790.2
a_9	0.000057	0.000186	0.000157
a_{10}	-0.004808	-0.00486	-0.004894
a_{11}	-0.006897	-0.007097	-0.007537

the 3D-2D perspective transformation for each camera were calculated from them. Thus, the camera calibrations were done. The coefficients are listed in Table 1.

3.2 Scanning the pre-determined ROIs We pre-determined nine focusing points Q_1 to Q_9 on an elliptical path, and scanned there by changing the coefficient set of the perspective transformation as shown in Fig. 2. There were persons at points Q_1 , Q_3 and Q_5 , and no one existed in the other six places. We theoretically calculated nine sets of the coefficients $b_{li}(i = 1 \sim 8)$ in (4) and $b_{ri}(i = 1 \sim 8)$ in (5) for pre-focusing at the ROIs. One of them are listed in Table 2. Figure 3(a) shows the color composite image between the originally observed images, in which the left image is assigned to red and the right image is assigned to green and blue. We can see that disparities of the objects depend on

Figure 2: Pre-determined scanning points.

their distances from the camera. Figure 3(b) shows the result of focusing at the point $Q_1(x = 0, z = 300)$ using the theoretically generated coefficients of the 2D-2D perspective transformation. Only the center person at the point Q_1 had no disparities, though the other two persons at another position had disparities.

To determine whether the person exists in the focused ROI, we evaluated the disparities between the transformed stereo pair images by using the 1D optical flow. In the TDSM, three cameras were installed at the same height so that stereo pair images had only horizontal disparities. Therefore, we calculated the 1D optical flow. We detected the person when the optical flow was positive and less than 1. Figure 4 shows the result of the evaluating disparities. White areas were evaluated as ones where the condition above was satisfied. In Fig. 4, we can see that the center person at the pre-focused point Q_1 was well extracted, but the other two persons at points Q_3 and Q_5 were not extracted. Figures 5 show some of the scanning results where dispar-

(a) Disparities between the original stereo pair images.

(b) Disparities between the focused images.

Figure 3: Color composite images before and after focussing.

Table 2: Theoretical generated sets of coefficients for pre-focusing.

Focusing points (x, z) [cm]						
	$Q_1(0, 300)$		$Q_3(-50, 400)$		$Q_5(100, 500)$	
	b_l	b_r	b_l	b_r	b_l	b_r
b_1	0.9953	1.005	1.000	1.011	0.9991	1.022
b_2	-0.002249	0.03120	-0.008870	0.03422	-0.007305	0.04084
b_3	31.06	-21.93	13.54	-20.75	17.68	-18.16
b_4	0.001879	-0.004132	0.0009	-0.004902	0.001149	-0.006587
b_5	1.008	1.037	1.003	1.038	1.004	1.038
b_6	-1.276	-4.688	0.3406	-4.897	-0.04154	-5.355
b_7	0.000023	0.000019	0.000020	0.000016	0.000021	0.000011
b_8	-0.000015	0.000038	-0.000013	0.000037	-0.000013	0.000033

Figure 4: The result of disparity evaluation.

Figure 6: A watching window for evaluating ROI size.

(a) Results focused at the point Q_2 . (b) Ones focused at the point Q_3 . (c) Ones focused at the point Q_5 . (d) Ones focused at the point Q_8 .

Figure 5: Scanning results at pre-determined ROIs.

ities after the focusing are indicated in the left-hand side and person detected on the right. In Figs. 5(b) and (c), persons at the focusing points Q_3 and Q_5 were well extracted, respectively, though no person was detected at Q_2 and Q_8 as shown in (a) and (d). White areas seen in Figs. 5(a) and (d) were considered as noises due to a periodic pattern in clothes.

3.3 Determination of the ROI size To design the person detection system using the proposed method TDSM, we experimentally obtained the relation between the camera interval and the size of the ROI with extension in both x and z directions (see Fig. 2). We defined the extension as an area in which disparities between images captured by the left and the right cameras, respectively, were less than one pixel. The disparities were evaluated by the 1D optical flow. The camera intervals were 10 cm, 20 cm and 35

cm. We prefocused at a fixed point ($x = 0, z = 400$ cm). For measuring the 1D optical flow, we used a random dot pattern on a plane. We changed its distance from 350 cm to 450 cm with 10 cm steps. In order to evaluate disparities accurately, a watching window was set on the random dot pattern as shown in Fig. 6. The ROI was set along the vertical line crossing the center of the window. The 1D optical flow were evaluated all over the window, and were cumulated vertically for getting the stable distribution. The ROI extension was determined from the position at which the optical flow exceeded one pixel in length. We repeated the procedures with changing the distance and obtained the extension in the depth direction. Figures 7 show the distribution of the 1D optical flows for camera intervals of 10, 20 and 35 cm, respectively. The vertical axes show a kind of goodness of the focusing, where smaller disparities give us larger values. Areas in which the disparity is within 1 pixel length are indicated by yellow tiles. From our definition, the disparity is less than 1 pixel, we obtained the size of the ROI as $117 \times 20 \text{ cm}^2$ for camera interval of 10 cm, $100 \times 10 \text{ cm}^2$ for interval of 20 cm, and $37 \times 10 \text{ cm}^2$ for interval of 35 cm, respectively, where, the actual depth size for camera interval of 35 cm was estimated as less than 10 cm. Considering that the optical flow can be measured more accurately, the ROI is smaller than our estimation. A new definition of the ROI will be required. Estimated ROIs shown in Figs. 7 may have some distortions in shape. We consider that the distortion comes from the miss-arrangement in camera directions.

4. Conclusions

In this paper, we proposed a three-dimensional spatial scanning method, TDSM, to detect an object or a person existing at any 3D region by using multiple cameras without a stereo matching process. In the proposed method TDSM, a pre-focusing was performed by applying an appropriate 2D-2D perspective transformation to a stereo pair images so that the object at the pre-determined region had no disparity between the stereo pair images. To extract the object, the TDSM evaluated the disparities by using an 1D optical flow technique. TDSM performed a camera calibration using a 3D-2D perspective transformation to freely scan the 3D space. Once the calibration was performed, the TDSM theoretically generated sets of coefficients for the prefocusing at any ROI. In addition, we experimentally obtained a relation between the camera distance and the ROI size. A

Figure 7: Estimated ROI for some camera intervals.

future area of the study includes development of more complete 3D scanning system using multi-camera system and its application to a pedestrian tracking system.

References

- [1] S. Tokoro, K. Kuroda, A. Kawakubo, K. Fujita and H. Fujinami, "Electronically scanned millimeter-wave radar for pre-crash safety and adaptive cruise control system", Proc. IEEE, Intelligent Vehicles Symposium, pp.304-309, 2003.
- [2] S. Tokoro, K. Kuroda, A. Kawakubo, K. Fujita and H. Fujinami, "Electronically scanned millimeter-wave radar for pre-crash safety and adaptive cruise control system", Proc. IEEE, Intelligent Vehicles Symposium, pp.304-309, 2003.
- [3] T. Tsuji, H. Hattori, M. Watanabe and N. Nagaoka, "Development of Night-Vision System", IEEE Transactions on Intelligent Transportation Systems, Vol.3, No.3, pp.203-209, 2002.
- [4] M. Kasahara and H. Hanaizumi, "A Dynamic Background Subtraction Method For Detecting Walkers Using Mobile Stereo-Camera", The 2013 International Conference on Image Processing, Computer Vision, and Pattern Recognition, IPCV, pp.123-128, 2013.
- [5] M. Bertozzi, E. Binelli, A. Broggi and M. D. Rose, "Stereo Vision-based approaches for Pedestrian Detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops, CVPR Workshops, DOI: 10.1109/CVPR.2005.534, 2005.
- [6] S. Nedeveschi, S. Bota and C. Tomiuc, "Stereo-Based Pedestrian Detection for Collision-Avoidance Applications", IEEE Transactions on Intelligent Transportation Systems, Vol.10, No.3, pp.380-391, 2009.
- [7] A. Zarshenas, M. Mesmakhosroshahi and J. Kim, "Fast Depth Estimation using Spatio-temporal Prediction for Stereo-based Pedestrian Detection", Visual Communications and Image Processing, VCIP, DOI:10.1109/VCIP.2015.7457838, 2015.
- [8] C. D. Pantilie and S. Nedeveschi, "SORT-SGM: Subpixel Optimized Real-Time Semiglobal Matching for Intelligent Vehicles", IEEE Transactions on Vehicular Technology, Vol.61, No.3, pp.1032-1042, 2012.
- [9] T. Sasaki and H. Hanaizumi, "A Pre-Focusing Method For Detecting Pedestrians Running Out Into The Street Using A Mobile Stereo Camera", IEEE International Conference on Intelligent Systems and Image Processing, ICISIP2016, pp.331-336, 2016.
- [10] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", Proceeding of Imaging Understanding Workshop, pp.121-130, 1981.

Takuma Sasaki (Member) was born in Fukushima, Japan in 1994. He received B.Sc. degree in Computer and Information Sciences from Hosei University in 2016. He is currently a master course student of graduate school of Computer and Information Sciences in Hosei University. His research interest is image processing. He is a member of IEICE.

Hiroshi Hanaizumi (Non-member) was born in 1956. He received D.Eng. degree in mathematical engineering and instrumentation physics from the University of Tokyo in 1987. He is currently a professor in the Faculty of Computer and Information Sciences at Hosei University. His current research interests are medical image processing, and person recognition using face image and remote sensing. He is a member of IEICE, SICE, IEEE, and others.