A Three-Dimensional Spatial Scanning Method for Detecting An Object Using Multiple Cameras

Takuma Sasaki* and Hiroshi Hanaizumi

Graduate School of Computer and Information Sciences, Hosei University, Kajino 3-7-2, Koganei, Tokyo 184-8584, Japan
*takuma.sasaki.3h@stu.hosei.ac.jp, hana@hosei.ac.jp

Abstract

A Three-Dimensional Scanning Method (TDSM) was proposed to detect an object or a person in a 3D Region Of Interested (ROI) without the stereo matching process. The method TDSM consisted of two processes; the 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 disparities between the left 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, pre-focusing method, optical flow.

1. Introduction

Various types of autonomous cars have currently been attracted attention and developed and tested on actual roads. They always sense their surroundings to safely reach their destination. In the sensing process, sensors such as infrared lasers, cameras, and millimeter-wave radars are used. Computer systems are also used to understand their outputs and to determine the next action. In fact, TOYOTA has already developed a practical Pre-Crash Safety System. Camera-based systems have also been developed, i.e., SUBARU’s Eyesight and HONDA’s Intelligent Night Vision system. These systems commonly obtain 3D road information not only to correct control but also to detect pedestrians entering the street for avoiding car-human accidents. The systems, however, need a time-consuming stereo matching process for getting 3D information. There are at least two approaches to solve the problem. One is to use a sub-system, such as GPGPU, to speed up the processing. The other is to adopt a smart method to avoid the stereo matching process. The former solution is trivial and expensive, though the latter low-cost and flexible. The authors have proposed a such method named as the pre-focusing method (PFM) to detect a pedestrian at a pre-fixed 3D ROI. In the PFM, we need calibrations for each pre-focused ROI. A smart method for avoiding the calibration has been required.

In this study, we propose a 3D spatial scanning method TDSM for random determination of ROIs. Once a camera calibration 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 a 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 perspective transformation as

\[
\begin{align*}
    p &= a_{11}x + a_{12}y + a_{13}z + a_{14} \\
    q &= a_{21}x + a_{22}y + a_{23}z + a_{24} \\
    s &= a_{31}x + a_{32}y + a_{33}z + a_{34} \\
    t &= a_{41}x + a_{42}y + a_{43}z + a_{44} \\
\end{align*}
\]

where, \(a_{ij} (i = 1 \sim 11)\) means coefficients for the perspective transformation from a point \((x, y, z)\) in the global coordinates to a pixel \((p, q)\) in the image captured by the left camera, \(a_{ij} (i = 1 \sim 11), (s, t), a_{ij} (i = 1 \sim 11)\) and \((u, v)\) for the center and the right cameras, respectively.

When these cameras \(C_l, C_c\) and \(C_r\) observe a rectangular on a plane A at a distance, their images have some disparities among them. Suppose a viewpoint transformation from \((p, q)\) to \((s, t)\) as

\[
\begin{align*}
    s &= b_{11}p + b_{12}q + b_{13} \\
    t &= b_{21}p + b_{22}q + b_{23} \\
\end{align*}
\]

where, coefficients \(b_{ij} (i = 1 \sim 8)\) 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 an object on the plane above has no disparity between the center image and one captured by the left camera then transformed by \((4)\). Thus, cameras \(C_l\) and \(C_c\) are pre-focused on the plane A.

We set another plane B and apply the same process to cameras \(C_r\) with a perspective transformation

\[
\begin{align*}
    s &= b_{31}u + b_{32}v + b_{33} \\
    t &= b_{41}u + b_{42}v + b_{43} \\
\end{align*}
\]

where, \(b_{ij} (i = 1 \sim 8)\) are coefficients for the perspective transformation from \((u, v)\) to \((s, t)\). They cancel disparities of an object on the plane B between the center camera image and one captured by the right camera then transformed by \((5)\).

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 \(I_k\) and \(I_e\) be transformed images from \(I_l\) by \((4)\) and from \(I_r\) by \((5)\), respectively. The object observed in the image \(I_k\) and that in \(I_e\) have no disparity between them, because the object is on the plane A. Similarly, the object observed in \(I_k\) and that in \(I_e\) have also no disparity between them, because the object is also on the plane B. In this situation, the object in \(I_k\) and that in \(I_e\) 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. The procedures for performing the pre-focused using the virtual plane are as follows;

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.
Step 3: Calculate the position of the four surrounding points in the images \(I_l\) and \(I_r\) by using \((1)\) and \((2)\).
Step 4: Calculate the coefficients \(b_i (i = 1 \sim 8)\) by using the four corresponding point pairs.
Step 5: Select another four points on the virtual plane B using the same manner as step2.
Step 6: Calculate the position of the four surrounding points in the images \(I_l\) and \(I_r\) by using \((3)\) and \((2)\).
Step 7: Calculate the coefficients \(b_i (i = 1 \sim 8)\) using the same manner as step 4.
Step 8: Produce images \(I_k\) and \(I_e\) using the above coefficients.

The object is detected at the focusing point when we have no disparity in the ROI area between \(I_k\) and \(I_e\). The disparity is evaluated by using an optical flow technique.

### 3. Experiments and discussion

We used a web camera, HD WEBCAM C270
(Logicool), as the stereo camera. The image size was 1280×720. We used a laser range finder, GLM70000 Professional (BOSCH), for measuring a point in global coordinates.

3.1 Camera calibration

We installed 3 cameras and the center camera was taken as the origin of the global coordinate system. The distance between the left camera and the center camera and the distance between the right camera and the center camera were 10 [cm]. A corner point of a checkerboard was measured using a laser range finder as shown in Fig.1. The coordinate $z_k$ and the coordinate $x_k$ of the point $K$ in the global coordinate were obtained from the length of three sides of the triangle ABP. The board was repositioned and corner points were measured at a total of 19 places. The coefficients of the perspective transformation for each camera are listed in Table 1.

Table 1. The coefficients of the perspective transformation for each camera.

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

3.2 Focusing the pre-determined ROI

We selected three focusing points and theoretically calculated three sets of the coefficients for pre-focusing at these regions. One of them are listed in Table 2. Fig.2(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. Three persons were at $(x = 0, z = 300)$, $(x = -50, z = 400)$ and $(x = 100, z = 500)$. We can see that disparities of the object depend on their distances from the camera. Fig.2(b) shows the result of focusing at $(x = 0, z = 300)$ using the obtained

Table 2. The coefficients of the perspective transformation.

<table>
<thead>
<tr>
<th>Focusing point $(x = 0, z = 300)$</th>
<th>$b_1$</th>
<th>$b_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_1$</td>
<td>0.9953</td>
<td>1.005</td>
</tr>
<tr>
<td>$b_2$</td>
<td>-0.002249</td>
<td>0.03120</td>
</tr>
<tr>
<td>$b_3$</td>
<td>31.06</td>
<td>-21.93</td>
</tr>
<tr>
<td>$b_4$</td>
<td>0.001879</td>
<td>-0.004132</td>
</tr>
<tr>
<td>$b_5$</td>
<td>1.008</td>
<td>1.037</td>
</tr>
<tr>
<td>$b_6$</td>
<td>-1.276</td>
<td>-4.688</td>
</tr>
<tr>
<td>$b_7$</td>
<td>0.000023</td>
<td>0.000019</td>
</tr>
<tr>
<td>$b_8$</td>
<td>-0.00015</td>
<td>0.000038</td>
</tr>
</tbody>
</table>

(a) Original stereo pair image (without focusing).

(b) Focused stereo pair images.

Fig. 1. Measurement of global coordinates.

Fig. 2. Color composite images before and after focused.
coefficients of the perspective transformation. We can see that only the person at the center had no disparity while the other persons still had disparities. Fig. 3 shows results of focusing at the other regions. Persons at different regions were respectively focused and had no disparities.

3.3 Extraction of the focused area

To detect the person at the focused region, we evaluated the disparity by using the optical flow technique. A threshold of the optical flow was 0 or more and less than 1. Fig.4 shows the result of the evaluation. White regions in Fig.4 were evaluated as focused areas. We can see that the optical flow detects the person at the focusing ROI. We compared the calculation cost of the proposed method TDSM with that of the template matching. Fig.5 shows the disparity map. To generate the map, we needed to perform the whole pixel matching. Furthermore, to detect the object from the disparity map, we needed to perform the template matching by using template images of various sizes. The calculation cost of the template matching was more expensive than that of the proposed method TDSM.

4. Conclusions

In this study, we proposed a 3D spatial scanning method TDSM to detect an object or a person in a 3D ROI without the stereo matching process. Once a camera calibration is performed, TDSM theoretically generates sets of coefficients for the perspective transformation to produce the ROI at any point in the field of view of the multiple cameras. To detect an object at the pre-determined ROI, TDSM evaluates disparities between pre-focused stereo pair image by using an optical flow technique. Future area of study includes a development of a system which tracks the pedestrian detected via TDSM.

References

3, No.3, pp. 203-209, 2002


