Mosaic-Based Global Vision System for Small Size Robot League

Yuji Hayashi[1], Seiji Tohyama[1], and Hironobu Fujiyoshi[1]

Dept. of Computer Science, Chubu University, Japan
yuji@vision.cs.chubu.ac.jp, sei@vision.cs.chubu.ac.jp, hf@cs.chubu.ac.jp
http://www.vision.cs.chubu.ac.jp/

Abstract. In the RoboCup F180 Small Size League, a global vision system using multiple cameras has been used to capture the whole field view. In the overlapping area of two cameras' views, a process to merge information from both cameras is needed. To avoid this complex process and rule-based approach, we propose a mosaic-based global vision system which produces high resolution images from multiple cameras. Three mosaic images, which take into account the height of each object such as our robots, opponent robots, and the ball on the field, are generated by pseudo corresponding points. Our system archives a position accuracy of better than 14.2 mm(mean: 4 mm) over a 4 × 5.5 m field.

1 Introduction

Recently, a global vision system using multiple cameras has been used in the RoboCup F180 Small Size League(SSL), since the field size was changed to 5,500 × 4,000mm to create more space. In the case of using two cameras as a global vision system, one is mounted over each half of the field to capture the image which has enough resolution for object recognition. However, there will be some problems with the use of two cameras. For example, in the overlapping area of both cameras' views, information from both cameras should be merged; however, this is considered a very complex process.

A possible solution to avoid this complex process is to employ a planar perspective transform known as "image mosaicking," which generates a high resolution image from two images. To obtain the mosaic image, a homography between a reference image and the other image is computed by correspondences. Although the mosaic image is generated from two camera images, the vision algorithm already in use in the SSL can be easily applied to the mosaic image without any changes. And any additional process such as merging information from both cameras is not needed. Since the mosaic image is registered as a planar image, there will be a blur around the object such as a robot, when the object has a height from the plane used for calculating the homography. This causes errors in the object identification, which processes regions to find the ball and robots and identifies our robots.

In this paper, we propose a mosaic-based global vision system, which generates mosaic images taking into account the height of each object. We will show that our system is capable of high accuracy in position estimation.

This paper is structured as follows. The second section points the problems in the use of a two camera system. The next section describes the proposed system in detail. Section 4 discusses the experimental results in position estimation and processing time. Section 5 concludes the paper.

2 Problems with a two camera system

In order to capture the image which has a pixel representiong 5 millimeters on the field, one is camera mounted over each half of the field as shown in Figure 1. Each camera image is processed by color segmentation and followed by object identification. Finally, the object's position is converted to real world coordinates for controlling the robots. After the vision processes described above, merging information from both cameras is required, because of the is overlapping area of the two camera images. To merge results from both cameras in the overlapping area, the following four methods are employed.

A. Updating recent result When a recent result is obtained from either camera, the final decision of the robot's position is updated.
B. Hard decision boundary The boundary for each camera is decided manually in advance the final decision of the robot's position is updated if the position is inside the boundary.
C. Hysteresis Final decision is taking over the results, while the robot is tracked by neither camera.
D. Fusion Estimated positions by both cameras are merged to world coordinates by weighting.

In method A, C, and D, robot1 and robot2 on the camera1 image shown in Figure 1 are not identified, because a part of makers on the robot is missing. In method B, the boundary requires continual adjustment when both cameras are mounted. To solve these problems, our approach generates a high resolution image from two cameras, and then processes the mosaic image by the vision algorithm which is used in a single camera system.

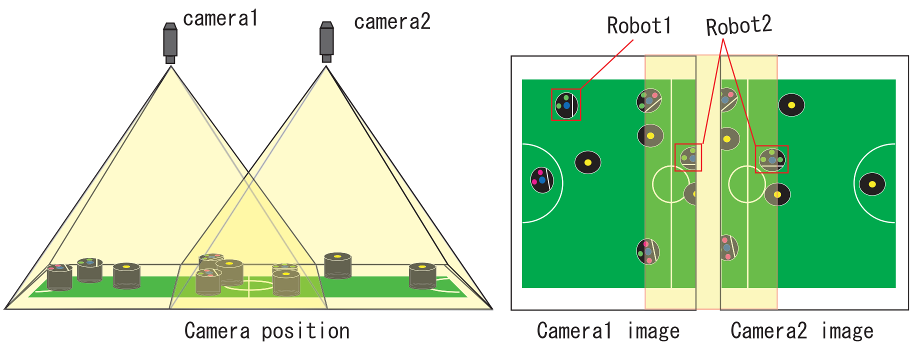

Fig. 1. Global Vision system using two cameras.

3 Mosaic-based global vision system

To obtain the mosaic image from both cameras, a homography is computed by correspondences between the two images. Since the mosaic image is registered as a planar image, there will be a blur on the image around the object which has a height from the plane used for calculating the homography. Our approach is generating three mosaic images which take into account the height of our robot, the opponent's robot and a ball. Each mosaic image is processed by color segmentation and object identification which usually are used in the global vision system of a single camera. By generating the mosaic image considering the height of each object, a highly accurate position estimate can be obtained.

3.1 Generating a mosaic image of the field plane

Figure 2(a) shows the relationship of projective geometry between both cameras and the base plane (field). The planar perspective transform based on a homography warps an image into another image using 8 parameters of the matrix $\boldsymbol{H}$ [7] [8]. The homography between the two images of a planar surface is expressed as

$$\boldsymbol{p'} = \boldsymbol{H_1 p_1} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \quad (1)$$

where $\boldsymbol{p_1} = [u_1, v_1, 1]^T$ on the camera1 coordinate and $\boldsymbol{p'} = [u', v', 1]^T$ on the mosaic image coordinate are corresponding points of two images. Masaic image coordinate $\boldsymbol{p'}$ corresponds to $\boldsymbol{P} = [x_W, y_W, z_W]^T$ on the world coordinate. The homography $\boldsymbol{H_2}$, which projects from the point $\boldsymbol{p_2} = [u_2, v_2, 1]^T$ on the camera2 to $\boldsymbol{p'}$ on the mosaic image, is computed inthe same way $\boldsymbol{H_1}$ was calculated. These relations are expressed as the following equations:

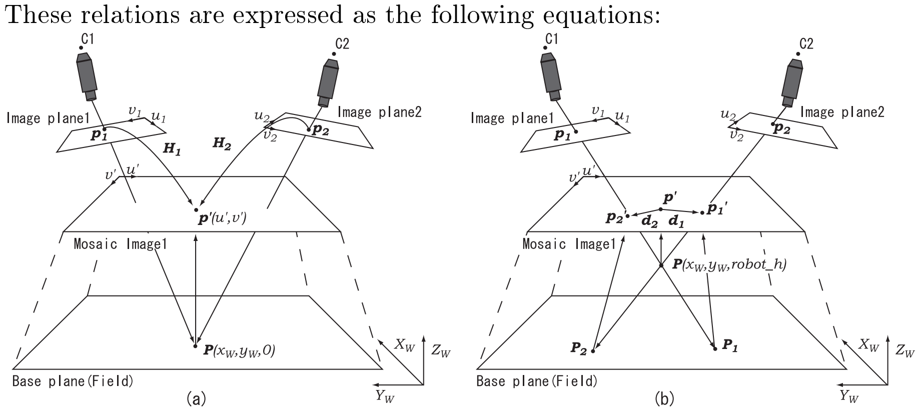

Fig. 2. Projective geometry between both cameras and a base plane.

$$\begin{aligned} \boldsymbol{p'} &= \boldsymbol{H_1 p_1} \\ \boldsymbol{p'} &= \boldsymbol{H_2 p_2} \\ \boldsymbol{P} &= \alpha \boldsymbol{p'}. \end{aligned} \quad (2)$$

Note that the mosaic image coordinate $\boldsymbol{p'}$ corresponds to the world coordinate $\boldsymbol{P}$ on the field by linear mapping. In our vision system, the mosaic image, in which a pixel represents 5 mm on the field, is generated using bilinear interpolation. Therefore, it is easy to obtain the world coordinate $\boldsymbol{P}$ from the mosaic image coordinate $\boldsymbol{p'}$ by

$$\boldsymbol{P}[\text{mm}] = 5[\text{mm/pixel}] \times \boldsymbol{p'}[\text{pixel}]. \tag{3}$$

The process for generating a mosaic image of the field is described as follows:

Step1 By Choosing the landmark points on the field plane (such as rectangle's corner), which are observed from camera1, corresponding points of $\boldsymbol{p_1}(u, v)$ on the camera1 image coordinate and $\boldsymbol{P}(X_w, Y_w, 0)$ on the world coordinate for the landmark are measured manually.
Step2 $\boldsymbol{P}$ is converted to the mosaic image coordinate $\boldsymbol{p'}$ by using equation (3).
Step3 The homography $\boldsymbol{H_1}$ is computed by using the correspondences of at least 4 points. The homography $\boldsymbol{H_2}$ is also computed in the same way $\boldsymbol{H_1}$ was calculated.
Step4 Blending is performed around the area where both images are overlapped. Finally, the mosaic image is generated using $\boldsymbol{H_1}$ and $\boldsymbol{H_2}$, from both camera images.

Figure 2(b) shows a mosaic image and both camera images. We can see that the ball on the field plane is very clear, but the markers on the top of the robot are not clear. Using the homography calculated from correspondences on the field plane, $\boldsymbol{P}$ viewed from camera1, which is located at a height of *robot_h* from the base plane (field), is projected to $\boldsymbol{P_1}$ on the world coordinate as shown in Figure 2(b). This causes errors denoted as $\boldsymbol{d_1}$ and $\boldsymbol{d_2}$ on the mosaic image as shown in Figure 2(b). For this reason, a blur shown in Figure 3(b) will be observed in the overlapping area on the mosaic image.

3.2 Generating a mosaic image of the virtual plane

Using the homography computed by correspondences on the field plane, pixels on the top of the robot on the camera image are not correctly projected to the mosaic image coordinate. In order to obtain the homography of any plane in 3D space, correspondences on the plane should be measured. However, it is impossible to measure the feature points on any plane in 3D space in a small amount of setup time. Our approach generates pseudo feature points on the image coordinate for the top of the robot, and the homography is computed as shown in Figure 4.

The process for generating a mosaic image taking into account a virtual plane is described as follows:

Step1 By choosing the landmark points on the field plane. Corresponding points of $\boldsymbol{p_1}(u, v)$ on the camra1 image coordinate and $\boldsymbol{P}(X_w, Y_w, Z_w)$ on the world coordinate are measured manually.

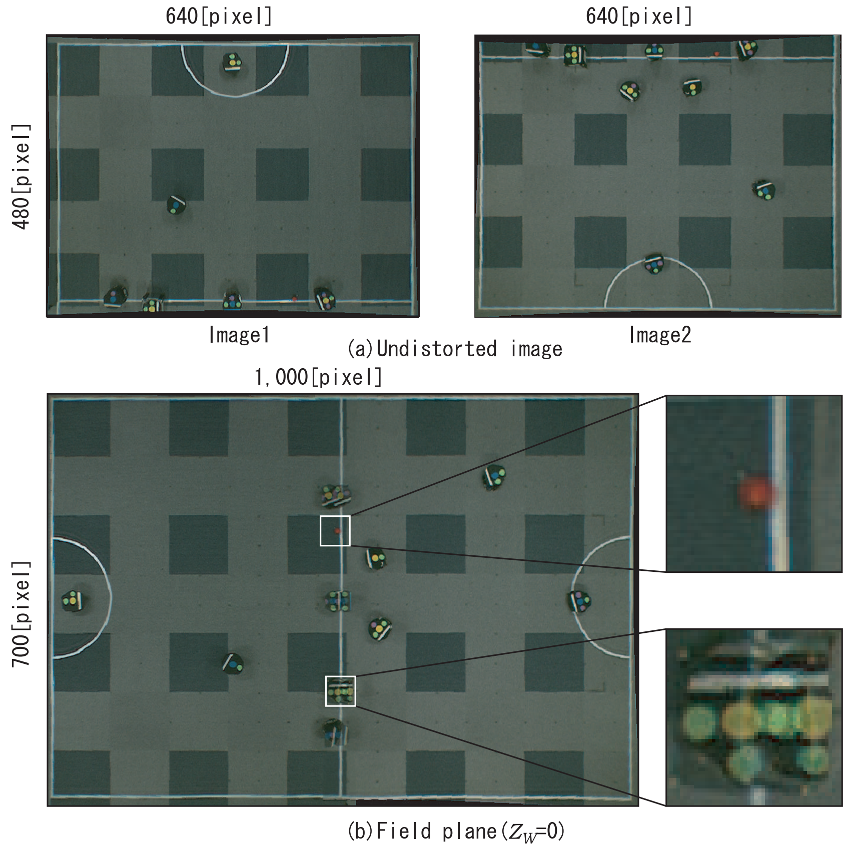

Fig. 3. Mosaic image of field plane.

Step2 Measuring the height of the robot($robot\_h$), and $\boldsymbol{P}$ is converted to the point $\boldsymbol{Q}(X_w, Y_w, robot\_h)$ on the virtual plane.

Step3 $\boldsymbol{Q}$ is projected to camera 1 by reverse projection using the intrinsic and extrinsic camera parameters. The pseudo corresponding point $\boldsymbol{q_1}$ is calculated by

$$\boldsymbol{q_1} = \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = \frac{1}{robot\_h'} \begin{pmatrix} fk_u & fs & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} (R \mid -RT) \begin{pmatrix} x_W \\ y_W \\ robot\_h \\ 1 \end{pmatrix} \quad (4)$$

where $\boldsymbol{R}$ is a rotation matrix, $\boldsymbol{T}$ is a translation matrix, f is a focal length, s is shearing factor, and k_u, k_v are unit length of each axis.

Step4 The homography $\boldsymbol{H_1'}$ is computed by using the correspondences of $\boldsymbol{q'}$ and $\boldsymbol{q_1}$. The homography $\boldsymbol{H_2'}$ is also computed by the same way of $\boldsymbol{H_1'}$.

Step5 Blending is performed around the area where both images are overlapped. Finally, the mosaic image taking into account the height of the object is generated using $\boldsymbol{H_1'}$ and $\boldsymbol{H_2'}$.

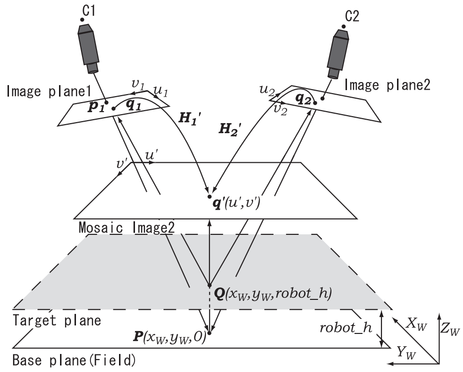

Fig. 4. Computation of homography of target plane.

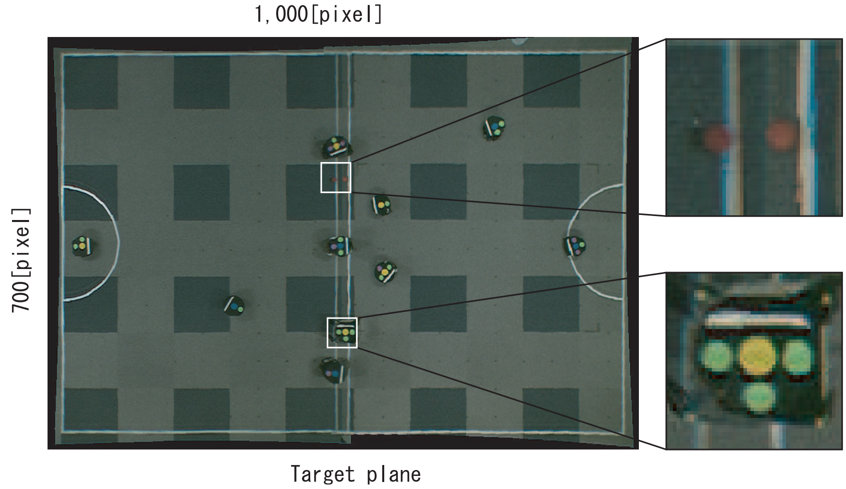

Fig. 5. Mosaic image of the virtual plane.

Figure 5 shows a mosaic image of the virtual plane which has a height of the robot ($robot\_h = 150$ mm). Although the ball on the field plane is not clear, we can not see any changes around the markers on the top of the robot. We also realized that the mosaic image of the virtual plane can absorb the change in the height within ± 2cm.

The global vision system we proposed has two merits. One of them to eliminate the necessity the process of merging the information from both cameras. The other is to facilitate the application of the vision algorithm which we have used commonly in single camera system such as CMVision[10] and robot identification method described in [9].

4 Experimental results

To determine the accuracy of the proposed method, we measured the robot's position as ground truth and compared it to the position given by our global vision system.

4.1 Configuration of vision system

Two cameras are mounted at a height of 3,000 mm, and each camera has a view of each area of 4,900 × 3,400 mm (overlapping area is 300 × 3,400 mm). Both cameras are calibrated using over 40 feature points on the field plane to estimate camera parameters.

4.2 Results

We evaluated the proposed method by location testing for 61 locations spread over the field. Table 1 shows results in regard to location of the robot (height is 150mm) by the two mosaic images of the virtual plane and the field plane. The mean of the position accuracy by the virtual plane (robot height) is 4mm. Note that a pixel represents approximately 5 millimeters on the field in our camera setting. This table shows that our vision system is able to correct to real world locations with a high degree of accuracy. The system achieves a position accuracy of better than 14.2 mm over a 4 m × 5.5 m field.

Figure 6 shows the distribution of the estimated positions. By the mosaic image of the field, it is clear that the error becomes larger regarding the distance from the optical center of the camera as shown in Figure 6(b). On the other hand, the mosaic image of the virtual plane is able to estimate real world locations very accurately, since the mosaic image was generated by taking into account the height of the robot.

Table 1. Error of the estimated positions [mm].

	Average		SD		Max	
	x	y	x	y	x	y
Field plane	43.5	34.3	24.9	20.7	95.7	77.1
Target plane	3.7	4.3	2.8	2.8	12.0	14.2

* SD : Standard deviation

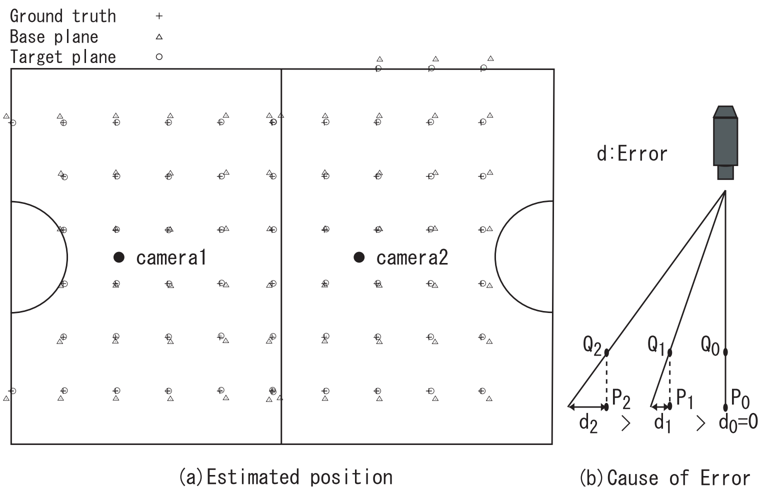

Fig. 6. Distribution of estimated positions.

4.3 Processing time

The proposed global vision system takes 81 ms for all processes including generating three mosaic images for our robots, opponent robots and field plan. (39 ms for generating three full mosaic images, 38 ms for color segmentation, and 4 ms for object identification). This processing time is not sufficient for controlling the robot in real-time using visual feed-back. To solve this problem, we generate mosaic images of only 30×30 pixels around each object position, which was detected in a previous frame, in order to run in real-time. The range of 30×30 pixels is about 15×15 cm in the real field. This search range works to reduce total processing time to 13 ms (generating three mosaic images takes 5.5 ms, color segmentation takes 5.5 ms, and object identification takes 1.1 ms for each process).

5 Conclusion

We proposed a mosaic-based global vision system using multiple cameras, which generates high resolution images taking into account the virtual plane of an object's height. Our system achieves a position accuracy of better than 14.2 mm(mean: 4 mm) over a 4×5.5 m field, and does not need any additional process of merging information from both cameras.

References

1. David Ball, Gordon Wyeth and Stephen Nuske, "A Global Vision System for a Robot Soccer Team", Proceedings of the 2004 Australasian Conference on Robotics and Automation (ACRA), Canberra, Australia, 2004.
2. Anna Egorova, Alexander Gloye, Cuneyt Goktekin, Achim Liers, Marian Luft,Raul Rojas, Mark Simon, Oliver Tenchio, and Fabian Wiesel, "FU-Fighters Small Size 2004", RoboCup 2004 Symposium, Small Size League Team Description, 2004.
3. Ng Beng Kiat, "LuckyStar 2004", RoboCup 2004 Symposium, Small Size League Team Description.
4. David Ball, Gordon Wyeth, "UQ RoboRoos 2004:Getting Smarter", RoboCup 2004 Symposium, Small Size League Team Description, 2004.
5. R. Y. Tsai. "A verstile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the Shelf TV Cameras and Lenses", In IEEE Journal of Robotics and Automation, Vol.RA-3, Num.4, pp. 323-344, 1987.
6. Z. Zhang. "A flexible new technique for camera calibration." IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330-1334, 2000.
7. R. Szeliski, "Image mosaicing for tele-reality applications", Rroc.IEEE Workshop on Applications of Computer Vision, pp.44-53, 1994.
8. Hartley A. and Zisserman A. "Multiple View Geometry in Computer Vision, Cambridge University Press", 2000.
9. Shinya Hibino, Yukiharu Kodama, Yasunori Nagasaka, Tomoichi Takahashi, Kazuhito Murakami, and Tadashi Naruse, "Fast Image Processing and Flexible Path Generation System for RoboCup Small Size League", RoboCup2002, pp.53-64, 2002.
10. James Bruce, Tucker Balch and Manuela Veloso, "Fast and Inexpensive Color Image Segmentation for Interactive Robots", In Proceedings of IROS-2000, Japan, October 2000.