Recognition of Shape and Generation of Free Viewpoint Image from Image Database

Haruki KAWANAKA[1]*, Yuji IWAHORI[2], Nobuaki SADO[3] and Kenji FUNAHASHI[1]

[1]*Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan*
[2]*Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi 487-8501, Japan*
[3]*TOHO GAS Information System Co., Ltd., 19-18 Sakurada-cho, Atsuta-ku, Nagoya 456-0004, Japan*
**E-mail address: haruki@center.nitech.ac.jp*

(Received November 30, 2004; Accepted February 10, 2005)

Keywords: Free Viewpoint Image, Silhouette, Principal Component Analysis, Pose Recognition

Abstract. For the purpose of generating a free viewpoint image in the scene, the previous studies need many actual cameras in the stadium. These studies take large scale environment with much cost, including actual camera environment such as a motion capture system. In this paper, a low cost and general approach is developed to generate a virtual image at another viewpoint from multiple viewpoint image database. This multiple viewpoint image database is designed for a soccer player and used to generate a virtual scene.

1. Introduction

The demands to generate the virtual scene from another viewpoint have been increased. For example, in order to show a scene of a soccer game from various viewpoints, some approaches use many cameras at the various locations or a camera with the function of pan-tilt-zoom. However, they show only the allocated camera images. They are not free viewpoint images. On the other hand, it is desired to generate a virtual image of the scene from the free viewpoint. To realize the generation, two main approaches have been proposed.

One is the image based rendering (LEVOY and HANRAHAN, 1996; NAEMURA *et al.*, 1998). The feature of this approach is that it does not need information of shape and reflectance of an object directly. Instead, it needs a large amount of image data of the scene with many cameras. The necessary condition is that the object is fixed essentially. From a practical points, it is difficult to be applied to the reconstruction of the scene such as variation from hour to hour. The other is the model based rendering (SATO and IKEUCHI, 1994; IKEUCHI and SATO, 2001). Necessary drawing techniques have been developed for this approach. As a result, it is available to render the complex shape including the texture. To increase the reality and reconstruct in-depth coverage, the specific installations and devices are used to make the modeling with much costs.

Recently, an approach to generate a virtual image at another view point has been proposed so that the soccer game can be seen from an intermediate viewpoint (OHTA, 2002). It requires large scale environment with many camera settings and installations. It takes much cost to realize the required environment. Such a system is available at restricted stadium, i.e., the application is restricted at only that stadium. Another research reproduces the sports scene using a motion capture (TAN *et al.*, 2000). This is also restricted at the indoor space. The motion capture requires the special wear and several markers as the feature points to be tracked from multiple cameras. From this reason, it is also difficult to be used in the actual playing game.

In this paper, more general approach without the special and expensive environment is proposed. A database is used to obtain a player's pose (shape) in a scene. The database consists of player images taken from multiple viewpoints. The appropriate pose image of each player is determined from the image database and used to generate a virtual image of the scene. Then each pose image is synthesized at the position onto the virtual scene. Here, the position of each player is provided by the trajectory system, which consists of three video cameras at one viewpoint.

2. Generation of Virtual Image from Image Database

2.1. Data acquisition from soccer player trajectory recording system

The soccer player trajectory recording system has been developed at our lab. The goal of this recording system is obtaining the trajectory of each player in real time. The system

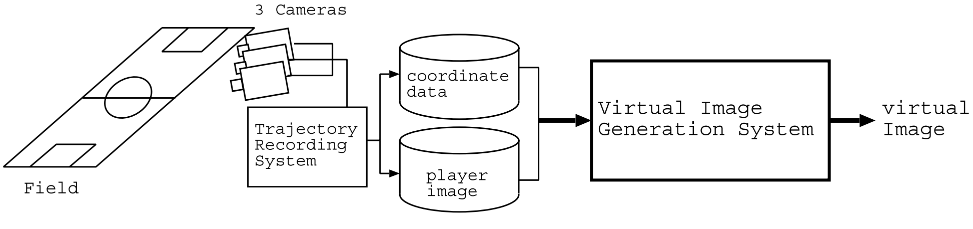

Fig. 1. System configuration.

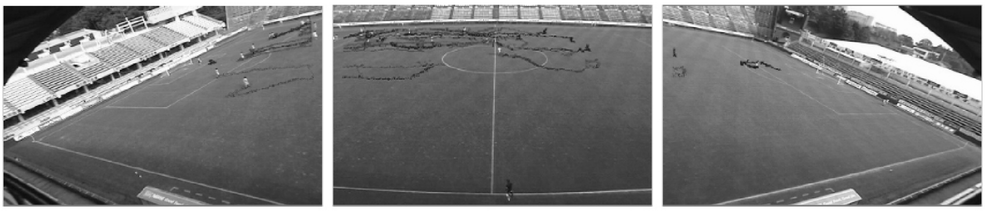

Fig. 2. Three images for the trajectory recording.

configuration is shown in Fig. 1. Three video cameras, which are connected to three PCs, are used to obtain the trajectory of each player in the whole soccer field. These three cameras are set at the same viewpoint. Figure 2 shows an example of three images which show each trajectory of each player. Proposed virtual image generation system uses its partial image and the corresponding coordinate of each player through the trajectory recording system.

2.2. Creation of image database

The image database was created using CG modeling software. An example is shown in Fig. 3. Various motions such as “running”, “walking”, “shot”, “pass”, “heading”, “trapping” etc. are designed and created. Around 200 poses are made for the variety of motion. For each pose of motion, eight images from eight viewpoints are created according to the rotation with every 45 degree.

To eliminate several factors such as the condition of light source, skin color, hair and uniform of each player, wearing shirts, shorts and socks, it is necessary to decrease the data size and to save the search time of the image database. From these points, image database is given using the silhouette as the invariant feature, which depends on the difference of pose and does not depend on the property of each player (see Fig. 4).

Here, the image size for each pose is normalized before obtaining the eigen vector of each silhouette image. That is, the rectangle region which surrounds the silhouette of each pose, is extracted. The target image is normalized so that the extracted region just touches with the square with keeping its aspect ratio. Let the pixel values of this normalized image be designated as $\vec{x}$ (Eq. (1)). $\vec{x}$ represents the raster scan, i.e., one dimensional expansion of the normalized image as

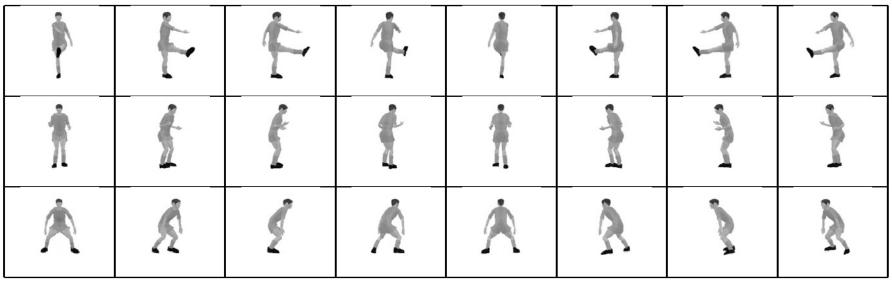

Fig. 3. CG model images.

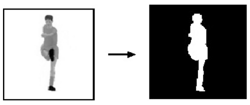

Fig. 4. Silhouette of a player.

$$\vec{x} = \left[x_1, x_1, \ldots, x_N\right]^T \tag{1}$$

where N represents the number of pixels. T means the transposition of matrix. Let P be the kinds of pose from 8 viewpoints, then the possible poses are expressed by a set of vectors as $\{{}_1\vec{x}^1, \ldots, {}_1\vec{x}^8, \ldots, {}_P\vec{x}^1, \ldots, {}_P\vec{x}^8\}$.

2.3. *Feature vector extraction*

Principal component analysis (PCA) is applied to compress vectors for the feature vector extraction. Let $\vec{m}$ be the average vector of all sample vectors (Eq. (2)).

$$\vec{m} = \frac{1}{8P}\sum_{i=1}^{P}\sum_{j=1}^{8} {}_i\vec{x}^j. \tag{2}$$

Let $\vec{X}$ be a set of vectors as

$$\vec{X} \equiv \left[{}_1\vec{x}^1 - \vec{m}, \ldots, {}_1\vec{x}^8 - \vec{m}, \ldots, {}_P\vec{x}^1 - \vec{m}, \ldots, {}_P\vec{x}^8 - \vec{m}\right] \tag{3}$$

$$\vec{C} \equiv \overrightarrow{XX}^T$$

$$\lambda_k \vec{e}_k = \vec{C}\vec{e}_k \quad (k = 1, \ldots, N)$$

where $\vec{m}$ means the average vector among sample vectors, $\vec{C}$ is a covariance matrix of $\vec{X}$. Here, eigen vectors in the eigen space are obtained as the base vectors ($\vec{e}_1$, $\vec{e}_2$, ..., $\vec{e}_M$), which correspond to the M eigen values of (λ_1, λ_2, ..., λ_M), where $M \le N$. When the proportion of total variance by M eigen values (i.e., M principal components) becomes over 90%, the principal components become effective sufficiently.

Each pose from each viewpoint is projected onto a point $\vec{f}$ in the eigen space as f

$$ {}_i\vec{f}^j = \left[\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_M\right]^T \left({}_i\vec{x}^j - \vec{m}\right). \tag{4}$$

The coordinate in the eigen space, the normalized image, and the corresponding viewing direction of the camera, are registered to each sample in the database. The pose ID is also uniquely registered so that 8 samples of the same pose with different viewpoints can be easily discriminated.

2.4. *Recognition of pose of player*

Let $\vec{y}$ be a vector obtained from the normalized silhouette of a player in a given image. This vector $\vec{y}$ is also projected into the eigen space according to the following equation.

$$\vec{g} = \left[\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_M\right]^T (\vec{y} - \vec{m}).$$

Here, the distance between $\vec{g}$ and each sample ${}_i\vec{f}^{\,j}$, is used for the recognition. When $\vec{g}$ is given, ${}_i\vec{f}^{\,j}$ which minimizes the distance among the whole image data set, is determinated as the most similar sample.

$$d = \min_{i,j} \left\| {}_i\vec{f}^{\,j} - \vec{g} \right\|. \tag{5}$$

2.5. *Pose recognition and synthesis of another view point image*

The most similar pose is determined from the image database, and the corresponding ID can be obtained through the matching process in the eigen space. The determined pose image is adjusted to be fit with the field coordinates of each player, which is acquired from the trajectory recording system. The corresponding virtual image from another viewpoint is put on the viewpoint coordinates transformed into the location of virtual viewpoint. The geometric adjustment with the original image is done to make a virtual image for mixed reality.

3. Experiments

3.1. *Experimental results*

An example of actual original image is shown in Fig. 5. For this image, the pose recognition is provided for each player. The database for pose recognition was generated from the image dataset which consists of a total of 2080 images, that is, 260 different poses from 8 viewing directions. Each image is normalized to 64 × 64 pixels. PCA is applied to this image set with the condition that the proportion of the total variance becomes more than 90%. Eigen vector is extracted as 286 dimensional vector which compresses the original vector. The results of this pose recognition are shown in Fig. 6.

For the quantitative evaluation of the result, the similarity s was measured using

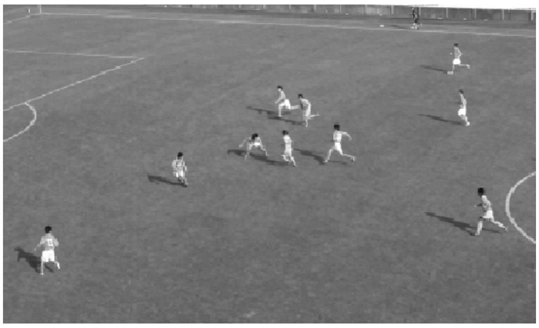

Fig. 5. Original image.

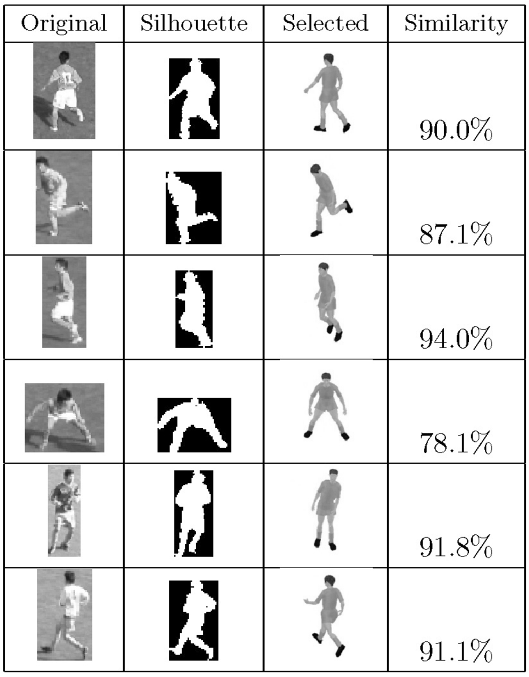

Fig. 6. Matching result from database.

$$s = \left(1 - \frac{\sum_{i=1}^{N} \left(x_i - y_i \right)}{N} \right) \times 100$$

where x_i represents the pixel value (0 or 1) of the silhouette in an original image, while y_i represents the pixel value of the silhouette in the matching result. N is the number of pixels. The value of N is 4096 in this experiment. The obtained results keeps the high accuracy as shown in Fig. 6.

A CG image of the stadium is shown in Fig. 7. A virtual stadium was created using the OpenGL. The most similar pose image of each player is put from the database. The virtual image with the same viewpoint generated for the whole players is shown in Fig. 8. The original image and the generated result of virtual image is almost similar. Finally, a virtual image from another viewpoint generated by this approach is shown in Fig. 9. It is shown that the generation of virtual image from quite different viewpoint can be realized. This result shows one frame image, but it is also possible to generate an animation by connecting each frame image sequentially. The generated virtual image gives reasonable impact as an application for the mixed reality.

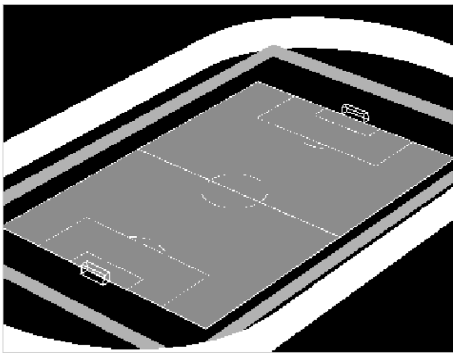

Fig. 7. A soccer field CG image.

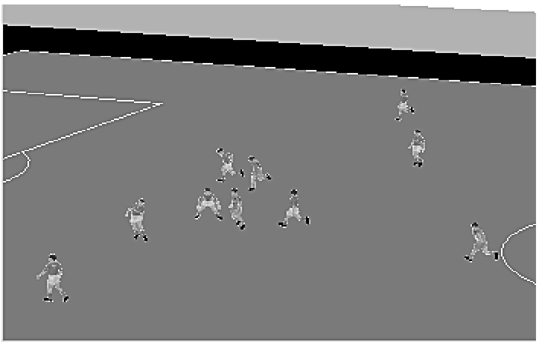

Fig. 8. Virtual image with the same viewpoint as original image.

3.2. Processing time and performance

The processing time and performance in the experiment are described. This experiment was performed using a PC with Athlon1.2 GHz 256 MB. First as the preprocessing, it takes 0.07 seconds to log and to label of each player image per one frame, then it takes 0.02 seconds per one player for both the normarization and the vectorization for an image. Then for pose recognition, it takes around 2 seconds to determine the nearest pose using 286 dimensional feature vectors from the image database with the total of 2080 poses. The average of the similarity of the pose recognition ratio is 89.4%, while the minimum similarity is 78.1%. These values are evaluated with the compressed 286 dimensional feature vectors. Finally, it takes 0.03 seconds per each player to paste another viewpoint image into the field CG image with the regulation of the image size. Through the whole processings, the rendering is done with around 20 seconds to generate another viewpoint image of one scene which includes ten players as shown in Fig. 9.

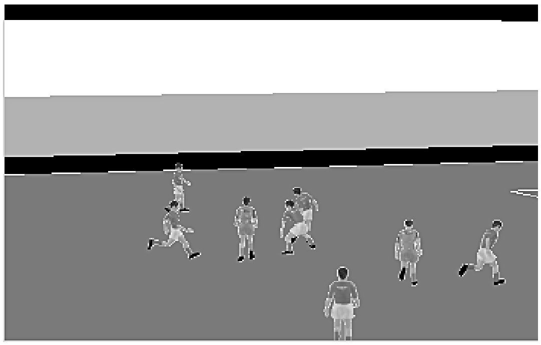

Fig. 9. Virtual image with different viewpoint from original image.

4. Conclusion

In this paper, a new approach is proposed to generate a virtual image from another viewpoint. The approach uses multiple image database and the PCA is applied for the pose recognition of each player. Another image with a different viewpoint can be generated with the recognized pose image through the proposed approach. Although the previous approaches use the large scale system, the entire approach is quite simple and generates the acceptable virtual scene from the designed multiple viewpoint image datasets.

To increase the reality for the generated virtual image, more kinds of the detailed pose image are used would improve the result more. While the large scale database results in taking costs for time and memory, which remains as the future subjects.

The authors would like to thank Prof. Yukio Sato of Nagoya Institute of Technology for his warmest and useful discussion. Iwahori's research is supported by the Kayamori Foundation of Information Science Advancement, Chubu University Grant, and the Grant-in-Aid for Scientific Research (Japan Society for the Promotion of Science) No. 16500108. Funahashi's research is supported by the Hori Information Science Promotion Foundation.

REFERENCES

IKEUCHI, K. and SATO, Y. (2001) *Modeling from Reality*, Kluwer Academic Press.

LEVOY, M. and HANRAHAN, P. (1996) Light field rendering, *Proc. ACM SIGGRAPH 96*, 31–42.

NAEMURA, T., KANEKO, M. and HARASHIMA, H. (1998) Ray-based rendering for virtual light sources, *The Institute of Image Information and Television Engineers*, 1328–1335.

OHTA, Y. (2002) Development of a 3D video stadium by a large-scale virtualized reality technology, *Meeting on Image Recognition and Understanding*, 341–348.

SATO, Y. and IKEUCHI, K. (1994) Temporal-color space analysis of reflection, *Journal of Optical Society of America*, A, 11–11, 2990–3002.

TAN, J. K., ISHIKAWA, S. and HAYASHI, K. (2000) A 3-d motion recovery technique for group sports employing uncalibrated video cameras, *IAPR Workshop on Machine Vision Applications*, 447–450.