MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28–30, 2000, The University of Tokyo, Japan

12—2

A 3-D Motion Recovery Technique for Group Sports Employing Uncalibrated Video Cameras

Joo Kooi Tan, Seiji Ishikawa, Ken-ya Hayashi
Department of Mechanical and Control Engineering
Kyushu Institute of Technology
Sensuicho 1-1, Tobata, Kitakyushu 804-8550, JAPAN

Abstract

Human motion recovery and modeling is a demanding technique in various areas such as sports, dancing, virtual reality, video game producing, *etc*. But the people in this area focus their efforts more on motion recovery of a single person than a group of persons. In this paper, a technique is described for recovering 3-D motions of a group of persons who are engaged in a certain activity. More than three video cameras are placed apart in front of the persons concerned and their motions are captured. Commonly visible feature points which spread over them are tracked on the video image streams to yield an extended measurement matrix. This matrix receives singular value decomposition and a shape matrix is finally obtained which contains the 3-D locations of all the chosen feature points. The technique is applied to 3-D motion recovery of the players in a mini-basketball game performed in a gymnasium and satisfactory results are obtained. The present technique may have application to, *e.g.*, formation evaluation of the players in a group sport as well as form analysis of individual players.

1 Introduction

Human motion recovery and modeling is a demanding technique in various areas such as a sport, dancing, virtual reality, video game producing, *etc*. In these areas, however, people direct their efforts more to a single person's motion recovery than a group of persons'. Motion recovery of a group of persons in a sport or dancing, for example, may give us lots of advantages for their formation evaluation as well as individual's motion analysis. For this purpose, existent motion recovery techniques such as a stereo camera system are not very convenient for use. Usually a group sport is performed in an athletic field or in a gymnasium where camera calibration is tough work. A motion recovery technique based on uncalibrated video cameras may contribute a lot particularly to motion recovery and analysis of field activities.

A 3-D shape recovery technique of non-rigid objects has been proposed[1,2]. The technique does not need calibration of the video cameras employed for the object's image capture, since it is an extension of a factorization method[3]. Moreover, another advantage of the technique is that it can be applied to any entities irrespective of single or multiple, only if feature points are chosen on them. This leads us to the application of the technique to motion recovery of group sports.

The present paper puts emphasis on applicability of the non-rigid shape recovery technique[1,2] to motion recovery in the environment where camera calibration is tough work. As one of the case studies, the present technique is applied to 3-D motion recovery of the players in a mini-basketball game in a gymnasium. The non-rigid shape recovery technique is outlined in the paper and experimental results are shown with discussion.

2 Measurement Matrix for Non-Rigid Objects[1,2]

In front of a non-rigid object (or non-rigid objects), F video cameras are placed and take images of the deformation. Video camera f produces image stream $I_f(t)$ $(t=1,2,\ldots,T)$. A feature point on the object at time t is denoted by $\boldsymbol{s}_p(t)$ (a 3×1 column vector). Let the

location of feature point $s_p(t)$ be denoted by $(x_{fp}(t), y_{fp}(t))$ on the image plane of video camera f at time t. After having found correspondence of the feature points $s_p(t)$ (p=1,2,…,P_t) among the images $I_f(t)$ (f=1,2, …,F), the following $2F \times P_t$ matrix $W(t)$ is obtained;

$$W(t) = \begin{pmatrix} x_{11}(t) & x_{12}(t) & \cdots & x_{1P_t}(t) \\ x_{21}(t) & x_{22}(t) & \cdots & x_{2P_t}(t) \\ \cdots & \cdots & \cdots & \cdots \\ x_{F1}(t) & x_{F2}(t) & \cdots & x_{FP_t}(t) \\ y_{11}(t) & y_{12}(t) & \cdots & y_{1P_t}(t) \\ y_{21}(t) & y_{22}(t) & \cdots & y_{2P_t}(t) \\ \cdots & \cdots & \cdots & \cdots \\ y_{F1}(t) & y_{F2}(t) & \cdots & y_{FP_t}(t) \end{pmatrix}. \quad (1)$$

The matrices $W(t)$(t = 1,2,…,T) are collected into a single $2F \times Q$ matrix W as

$$W = \big(W(1) \mid W(2) \mid \cdots \mid W(T)\big). \quad (2)$$

Here Q is the number of all the feature points during observation, *i.e.*,

$$Q = \sum_{t=1}^{T} P_t \ . \quad (3)$$

The extended measurement matrix of the size $2F \times Q$ is defined by

$$\tilde{W} = W - \frac{1}{Q} W \cdot E \ . \quad (4)$$

Here E is a $Q \times Q$ matrix whose entries are all unity. The entries of matrix $\tilde{W}$ are denoted by $(\tilde{x}_{fp}(t), \tilde{y}_{fp}(t))$.

All the P_t (t=2,3,…,T) feature points are projected onto the 3-D space at a certain sampled time and there the origin O is defined at the centroid of these feature points. If orthographic projection is assumed in the imaging through a camera lens, we have

$$\begin{aligned} \tilde{x}_{fp}(t) &= (\boldsymbol{i}_f, \boldsymbol{s}_p(t)) \\ \tilde{y}_{fp}(t) &= (\boldsymbol{j}_f, \boldsymbol{s}_p(t)) \end{aligned}, \quad (5)$$

where $\boldsymbol{i}_f$, $\boldsymbol{j}_f$ and $\boldsymbol{k}_f \equiv \boldsymbol{i}_f \times \boldsymbol{j}_f$ are unit vectors defining the lens coordinate system of video camera f. Substitution of Eq.(5) into the components of matrix $\tilde{W}$ in Eq.(4) results in the decomposition of the matrix $\tilde{W}$ into two matrices in the form

$$\tilde{W} = MS \ , \quad (6)$$

where M is a $2F \times 3$ matrix whose f's low contains $\boldsymbol{i}_f^{\mathrm{T}}$ and the (f+F)'s low contains $\boldsymbol{j}_f^{\mathrm{T}}$ (f= 1,2,…,F), and S is a $3 \times Q$ matrix of the form

$$\begin{aligned} S = \big(&\boldsymbol{s}_1(1), \ \boldsymbol{s}_2(1), \ \cdots, \ \boldsymbol{s}_{P_1}(1) \ \mid \\ &\boldsymbol{s}_1(2), \ \boldsymbol{s}_2(2), \ \cdots, \ \boldsymbol{s}_{P_2}(2) \ \mid \ \cdots \ \mid . \\ &\boldsymbol{s}_1(T), \ \boldsymbol{s}_2(T), \ \cdots, \ \boldsymbol{s}_{P_T}(T)\big) \end{aligned} \quad (7)$$

Matrix S gives 3-D coordinates of all the Q feature points. Thus the matrix provides with the 3-D shape and motion of the object concerned.

The matrix decomposition indicated in Eq.(6) is done by applying singular value decomposition (SVD) to matrix $\tilde{W}$. This is realized when $rank(\tilde{W}) = 3$ holds[3]. Orthonormality of the vectors $\boldsymbol{i}_f$ and $\boldsymbol{j}_f$ are also employed for uniquely decomposing the matrix $\tilde{W}$.

3 Motion Recovery of Basketball Players

In order to show the performance of the present technique, a 3 players vs. 3 players mini-basketball game is taken video images by three video cameras (*i.e.*, F=3) placed apart on the floor in a gymnasium. In order to make feature points tracking simpler, the players are asked to put on dark training wears, and small white balls and white bands are attached on their bodies. Feature points are also specified at the center of a basketball, a basketball ring, and a floor.

Out of several minutes of the captured video

images, the motion during 3 seconds is chosen and recovered. Since the sampling time interval is 0.1 second, T=30. The total number of the feature points is 1749(=Q). Matrix $\tilde{W}$ is therefore a 6×1749 matrix. It is decomposed by SVD and finally shape matrix S is obtained.

The image sequence employed for the motion recovery is partly shown in **Fig.1**. In the figure, (a) player 4 keeps the ball for a while and (b) he tries to pass the ball to player 5. (c) Player 3 jumps to interrupt the pass but he fails. (d) Player 5, having received the ball, shoots the ball. The recovered motion is illustrated in **Fig.2**, in which 6 players are shown. The goal board is indicated by a gray rectangle, whereas the basketball by a small square. The recovered feature points are connected by straight lines where possible and the bodies are described by filled rectangles for better visibility. Correspondence of the scenes with the video images of Fig.1 is given as follows: (a) corresponds to the scene of (1,2), (b) to (3,3), (c) to (1,4), and (d) to (3,5). Here (i, j) means the ith column of the jth row in Fig.2.

4 Discussion and Conclusions

A technique was shown for recovering human motions in a group sport employing uncalibrated video cameras. Since calibration is not necessary with the settled video cameras, the technique can not only be employed indoors but also be employed outdoors with much simplicity in video images capturing. This is one of the main advantages of the proposed technique over existent ones.

The present technique can be applied to motion analysis of the players in a group sport. For example, trajectory of each player, trajectory of a ball, an overall formation of the players, *etc.* can be obtained from the recovered motion. Velocities and the values of acceleration can also be computed with respect to specified motions. Thus a user can evaluate or interpret players' actions.

As one of the practical problems in multiple objects recovery, occlusion of rear objects often occurs, as is observed in Fig.2. One way of overcoming this issue is to employ surrounding video cameras[4]. This remains for further study.

Since the present technique assumes orthographic projection, it may results in certain recovery errors. The amount of the errors is empirically estimated as approximately 4% in average. It may be sufficient for human motions recovery and analysis in many sports though.

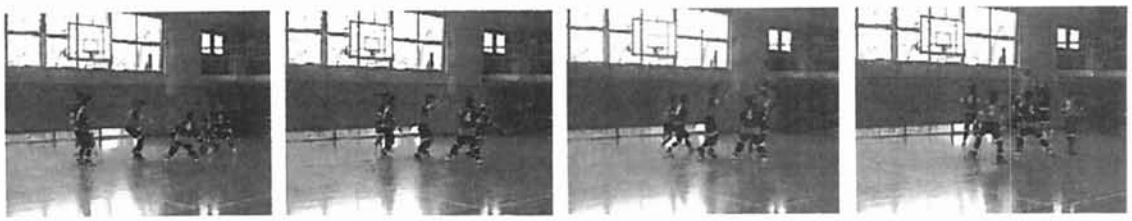
References

[1] J. K. Tan, S. Kawabata, S. Ishikawa, "Recovering shape of deformable objects by factorization", *J. Inst. Image Inform. and Tele. Eng.*, **52**, 3, 406-408 (1998). (in Japanese)

[2] J. K. Tan, S. Ishikawa, "Extracting 3-D motions of individuals at work by uncalibrated multiple video cameras", *Proc. 1999 IEEE Int. Conf. on Systems, Man and Cybernetics*, III-487-490(Oct., 1999).

[3] C. Tomasi, T. Kanade, "Shape and motion from image streams under orthography: A factorization method", *Int. J. Computer Vision*, **9**, 2, 137-154(1992).

[4] J. K. Tan, S. Ishikawa, "A method of modeling deformation of an object employing surrounding video cameras", *Int. Archives of Photogram. and Remote Sensing*, XXXIII, Part B5, 802-808(July, 2000).



(a) (b) (c) (d)

Fig. 1. Video images of a mini-basketball game.

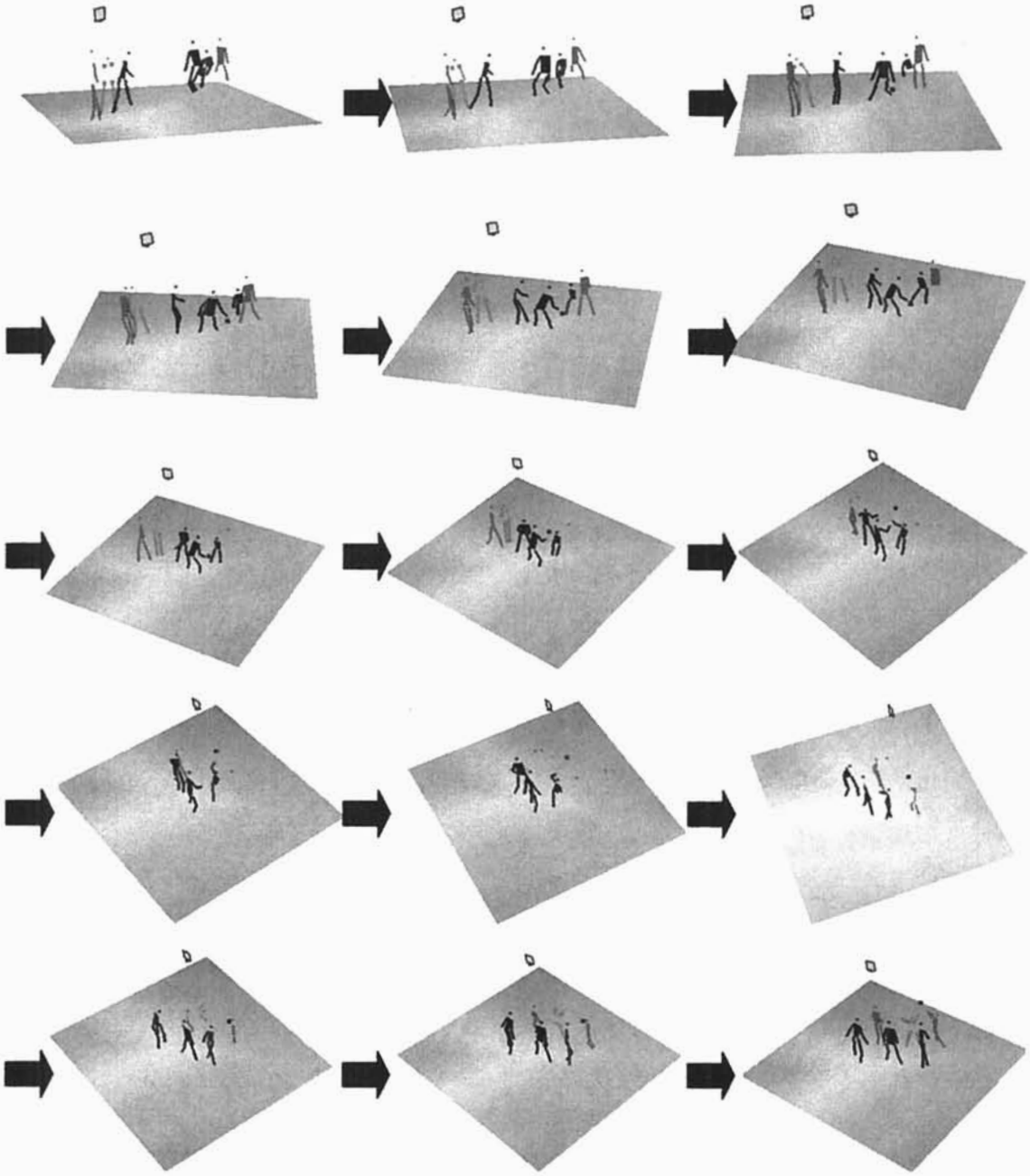


Fig. 2. The recovered motion. The scene proceeds as indicated by arrows.