Human Motion Recovery by Mobile Stereoscopic Cameras

Tan, J. K.*, Yamaguchi, I.*, Ishikawa, S.*, Naito, T.**, Yokota, M.**

*Department of Control Engineering, Kyushu Institute of Technology
Sensuicho 1-1, Tobata, Kitakyushu 804-8550, Japan
Email: {etheltan, ishikawa}@cntl.kyutech.ac.jp

**Department of Periodontology and Endodontology, Kyushu Dental College
Manazuru 2-6-1, Kokurakita, Kitakyushu 803-8580, Japan

Abstract

A method is described for reconstructing 3-D object models, particularly human models, based on multiple calibration-free mobile cameras and a computer. Existing optical 3-D modeling systems must employ calibrated cameras that are set at fixed positions. This inevitably gives constraints to the range of movement with object motions. In the proposed method, multiple mobile cameras take images of an object moving widely and its 3-D model is reconstructed from the obtained video image streams. Advantages of the proposed method include that the employed cameras are calibration-free and that the method claims at least two cameras, stereoscopic cameras, to be employed, which offers a simple system. The theory is described and the performance is shown by the experiment on 3-D human motion modeling. Precision of the 3-D model is also shown with discussion.

Keywords: Stereoscopic cameras, mobile cameras, motion capture, motion recovery, computer vision.

1. Introduction

Three-dimensional shape recovery techniques of deformable objects have attracted researcher especially in a computer vision community over the last decade. Particularly human motion recovery has become popular, since it has various potential applications such as creating a human model in video games or in a virtual reality space, motion analysis in various sports, traditional dances or skills preserving in an electronic museum, *etc*. Stereoscopic vision is, as is well known, a popular technique for performing such 3-D shape recovery. But it is not very convenient particularly for outdoor use because of camera calibration. Alternatively motion recovery employing magnetic sensors is also a common technique. It restricts motion range of the subject, however. Optical measurements with non-contact techniques that can capture wide range motions are obviously better for wide spread of the technique.

We have already proposed a shape recovery technique of 3-D non-rigid objects based on multiple calibration-free cameras [2,3]. It employs a factorization method [1] with an extended measurement matrix [2] that contains spatiotemporal information on the object's deformation. Since the technique necessitates cameras to be fixed around the object concerned, it can only deal with the motions/ movements in a limited space. This disadvantage can be overcome by the employment of multiple mobile cameras with much more flexibility in image taking.

The idea of our approach is to devise a way of creating a measurement matrix that should be a full matrix whose entries are all known. Once a full measurement matrix is given, it can be factorized into a camera orientation matrix and a shape matrix [1]. The shape matrix gives information on a 3-D shape/motion of the object concerned. This paper proposes a multiple mobile cameras system as a most convenient way of taking images of an object moving in a wide range. The method is described and some experimental results are shown. Precision of the created 3-D model is as well explained and discussion is given.

2. A measurement matrix for mobile cameras

The measurement system proposed in [2,3] places cameras at fixed positions like existent 3-D measurement systems commercially available. This inevitably restricts the motion of a person whose 3-D model is the present concern. An efficient way of taking images of such a moving object is to let cameras move along with the object. The idea of moving a camera is normally employed when static objects are concerned for 3-D shape recovery [1]. The technique presented in this paper allows camera motion to widely moving non-rigid as well as rigid objects.

We suppose that F cameras take images of an object by changing their locations L times. If we denote a measurement matrix of the F cameras at location l (l=1,2,…,L) by W_l, a general form of an overall measurement matrix W is given by the following;

$$W = \begin{pmatrix} W_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & W_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & W_L \end{pmatrix}. \tag{1}$$

Here matrices W_l (l=1,2,…,L) are placed at the diagonal positions of W and the remaining entries are all vacant. The matrix having this form cannot be further processed in terms of factorization [1].

It should as well be noted that cameras cannot be calibrated in this technique because of their mobility.

3. Mobile cameras system I

Ncan be done in the context of factorization method. To go further, we proposed a mobile frame system [5] in which cameras are placed fixed on a mobile frame. We placed the world coordinate system on the mobile cameras, instead of placing it on the ground where the mobile cameras move. Then the circumstance is equivalent to that the world moves around the cameras in place of the cameras moving in the world. This means that all the sub-matrices W_l (l=1,2,…,L) can be aligned in the same rows as follows;

$$W' = \begin{pmatrix} W_1 & W_2 & \cdots & W_L \end{pmatrix}. \tag{2}$$

Equation (2) provides a full measurement matrix and it can be factorized according to [2,3], and finally obtained shape matrix S giving 3-D coordinates of the motion of the object relative to the cameras.

4. Mobile cameras system II

In this section, we describe a technique [6] how to take images and recover shape of a widely moving object, employing independently moving cameras. In the technique, the camera orientations are all different through the camera movement from location l=1 to l=L. Therefore the cameras need calibration to obtain 3-D information of the object. But unlike existent 3-D recovery techniques that perform camera calibration using 3-D tools, the proposed technique doesn't employ such tools for the calibration. Instead, it employs rigid points that are observed in the captured image streams. The proposed procedure is described in the following.

At the position l (l=1,2,…,L), the cameras are required to observe common feature points on static objects (referred to as *static feature points* hereafter) other than the feature points on non-rigid objects (referred to as *moving feature points* hereafter). Let the sub-matrix containing static feature points be denoted by W_l^{R} and the sub-matrix containing moving feature points be denoted by W_l^{N}. Then the sub-matrix W_l at position l in Eq.(1) can be written as

$$W_l = \begin{pmatrix} W_l^{\mathrm{R}} & W_l^{\mathrm{N}} \end{pmatrix}. \tag{3}$$

Since matrix W_l^{R} (l=1,2,…,L) only contains common static feature points, they are collected into a single matrix W^{R} by

$$W^{\mathrm{R}} = \begin{pmatrix} W_1^{\mathrm{R}} \\ W_2^{\mathrm{R}} \\ \vdots \\ W_L^{\mathrm{R}} \end{pmatrix}. \tag{4}$$

Applying factorization to the above full matrix W^{R}, we have

$$W^{\mathrm{R}} = M \cdot S^{\mathrm{R}}, \tag{5}$$

where matrix S^{R} contains the 3-D coordinates of all the chosen static feature points. Camera orientations at location l (l=1,2,…,L) are also obtained by matrix M. Let us denote the camera orientation matrix M by

$$M = \begin{pmatrix} M_1 & M_2 & \cdots & M_L \end{pmatrix}^{\mathrm{T}}. \tag{6}$$

Then the 3-D coordinates of moving feature points are then calculated by

$$S^{\mathrm{N}} = M^{+} \cdot W^{\mathrm{N}}. \tag{7}$$

Here S^{N} is a matrix composed of S_l^{N} , M^+ is a matrix whose diagonal elements are M_l^+, and W^{N} is a matrix whose diagonal elements are W_l^{N}.

In this way, the measurement matrix of Eq.(1) can be decomposed as

$$W = M \cdot S \tag{8}$$

and all the feature points registered in matrix W recover their 3-D positions. They are given by

$$S = (S^{\mathrm{R}} \mid S^{\mathrm{N}}). \tag{9}$$

Precision of recovered 3-D shape is evaluated by projecting the obtained 3-D shape back onto the original video images [2]. We multiply the derived matrices M and S to obtain 2-D positions of recovered feature points on each image plane. Let us denote the result of the multiplication by $\hat{W}$, i.e.,

$$\hat{W} = M \cdot S. \tag{10}$$

The precision is then computed by evaluating the difference between W and $\hat{W}$.

5. Experimental results

In the performed experiment, a subject moved around in the yard and its motion was taken images by two mobile cameras, denoted by C_L and C_R, respectively. The cameras were connected to video transmitters and the video images were sent via the transmitter and the receivers to two PCs in a distant room. Sampling interval was 0.1 second and 100 images were sampled. Thus 10 seconds motion was processed for recovery. The chosen feature points are 24 for static feature points and 17 for moving feature points. In this particular experiment, the subject's motion was controlled so that the two cameras could observe the 24 static feature points during all the observation time.

Video images, obtained from cameras C_L and C_R, of the subject's motion are partly shown in **Fig.1**. The result of 3-D recovery is given in **Fig.2**. In both figures, the time proceeds as indicated by arrows. The precision of the recovery was computed and the recovery error was 1.75% for static feature points and 1.71% for moving feature points with respect to the left camera, and 1.72% for static feature points and 2.12% for moving feature points with respect to the right camera. Thus we have 1.8% of the average recovery error. **Figure 3** is the illustration of the precision of the experiment. The recovered images shown by broken lines are superposed onto the original image shown by solid lines. As we have achieved very small recovery errors, it is actually not easy to see the difference between them.

6. Conclusions

We have shown a technique for 3-D human motion recovery employing mobile cameras. In this technique, at least two cameras are used for the recovery. Therefore it can be called mobile stereoscopic vision by mobile stereoscopic cameras. Camera calibration is done using some representative points observed in images. This is completely different from existent techniques that use 3-D tools for the calibration. In this sense, our technique is easier to use especially for those who are not very familiar with shape/motion recovery.

The precision of the recovery or the recovery error was satisfactorily small, 1.8%, in the performed experiment, which is comparable to existent techniques based on fixed stereo cameras. At the moment, we employ orthographic projection as a lens model. Changing the model to week perspective projection, e.g., may result in better precision. This remains for the next step of the study.

Since the proposed mobile cameras system does not need ordinary camera calibration, it can be employed in remote places of interest where time compensation by tedious camera calibration may cause loosing important shots of the event of interest. In the proposed system, one only have to bring two cameras or more to such a place and just start taking images of the event there. If one sends the taken images back to the lab by a video transmitter or even internet, for example, the event can be modeled in a 3-D way in the lab. This means that a dynamic event at any distant place can be modeled in a 3-D way by the proposed technique. This pattern of employment of the technique may have various applications in future.

This study was partly supported by JSPS KAKENHI (14580450), which is greatly acknowledged.

References

[1] C. Tomasi, T. Kanade, "Shape and motion from image streams under orthography: A factorization method", *Int. J. Computer Vision*, Vol.9, No.2, pp.137-154, 1992.

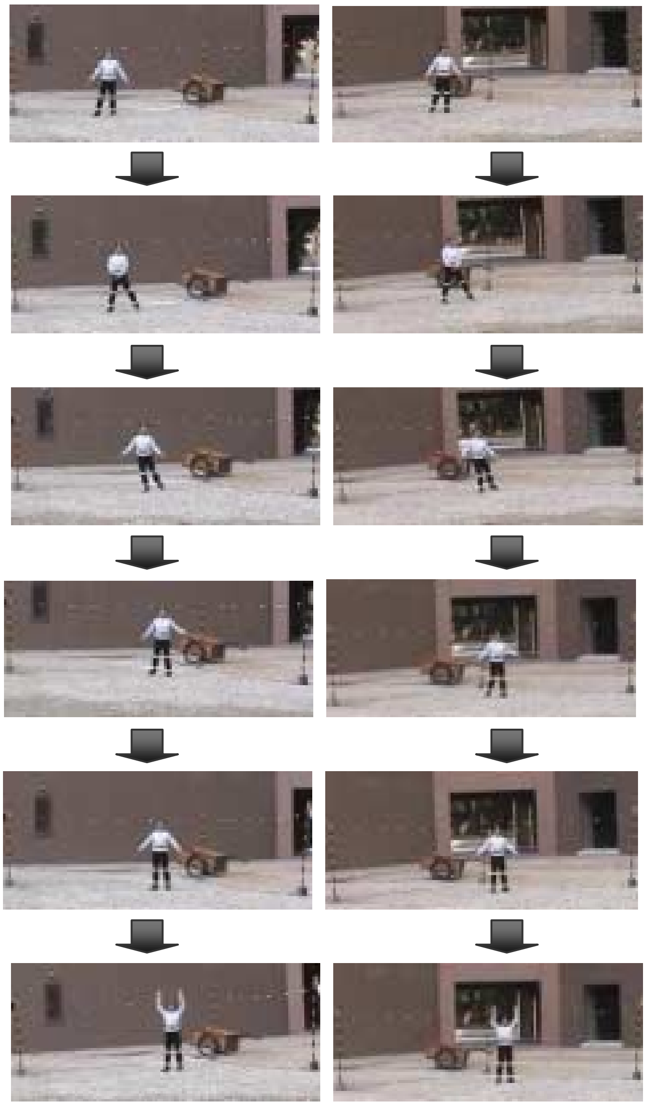
[2] J. K. Tan, S. Ishikawa, "Human motion recovery by the factorization based on a spatio-temporal measurement matrix", *Computer Vision and Image Understanding*, Vol.82, No.2, pp.101-109, 2001.

[3] J. K. Tan, S. Ishikawa, "Recovering 3-D motion of team sports employing uncalibrated video cameras", *IEICE Trans. on Information and Systems*, Vol. E84-D, No.12, pp.1728-1732, 2001.

[4] Y. Chen, G. Mediomi, "Object modeling by registration of multiple range images", *Image and Vision Computing*, Vol.10, No.3, pp.145-155, 1992.

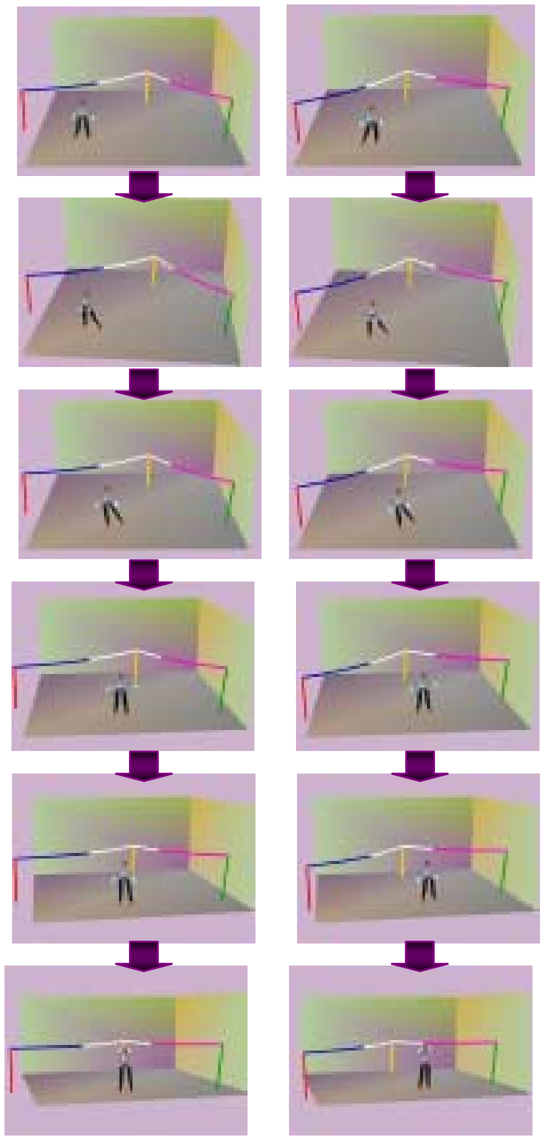
[5] T. Tabusa, J. K. Tan, S. Ishikawa, "Recovering human motions by mobile video cameras and factorization", *Proc. of Asian Conf. on Computer Vision*, 2002.

[6] J. K. Tan, I. Yamaguchi, T. Tabusa, S. Ishikawa, S. Hirokawa, "Various image taking strategies for 3-D object modeling based on multiple cameras", *Proc. of Int. Conf. on Image Processing*, 2487-2490, 2004.



(a) (b)

Fig. 1. Video images of a human in motion: (a) C_R, and (b) C_L .



(a) (b)

Fig. 2. Recovered 3-D motion: (a) C_R, and (b) C_L.

C_R

C_L

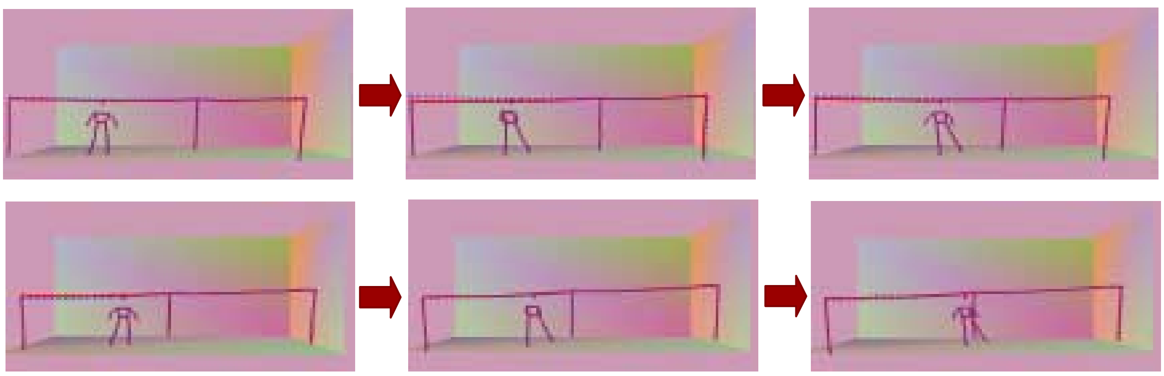


Fig. 3. Precision of the recovered 3-D model.