MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara- ken New Public Hall, Nara, Japan

5—2

A Remote Motion Capture System Based on Mobile Cameras

Joo Kooi Tan[1], Kouhei Tou[1], Seiji Ishikawa[1]
Kyushu Institute of Technology,
Department of Mechanical and Control Engineering

Abstract

This paper describes and demonstrates a general technique for recovering 3-D shape of non-rigid objects in a distant or remote place. Two important stages are included in the system; a transfer stage and a recovery stage. A transfer stage includes two mobile cameras connected with two mobile computers and a wireless LAN at a remote place for capturing as well as transferring human motion images to a recovery stage. A computer in a recovery stage takes the role of recovering and representing the human motion in a 3-D way. The present technique is applied to a human walking and sitting down motion at a remote place. Experimental results are shown and discussion is given.

1 Introduction

Wireless or mobile communication systems using networks have already become one of the most important as well as useful tools in a new generation of communication technologies [1]. Because of the enabled multimedia transmission and Internet access, a large amount of image data at a distant place can be transferred in a high-speed between computer and computer via a network system they link. As a result, such remote image data can be employed in various ways through image processing or image analysis creating huge benefit in our life. By this reason, imaging technology has become much more popular than ever.

In this paper, we keep our purpose on imaging and communication technology application; how to establish a remote motion capture system to recover 3-D shape of distant non-rigid objects under an arbitrary environment.

Three-dimensional shape recovery techniques of non-rigid objects have been quite remarkably developed in recent years. Particularly human motion recovery has many potential applications such as creating a human body shape [2], modeling characters in video games or in a virtual reality space, motion analysis in various sports, traditional dances or skills preserving in an electronic museum, *etc.* Stereo vision is, as is well known, a popular technique for performing such 3-D shape recovery. But it always necessitates camera calibration, which is not very convenient particularly for outdoor environment. Alternatively motion recovery employing magnetic sensors is also a common technique. It restricts motion of the subject, however. Non-contact techniques based on optical measurement are obviously better for wider use. The present paper offers a non-contact optical technique without ordinary camera calibration.

We have already proposed a shape recovery technique of 3-D non-rigid objects based on multiple uncalibrated cameras [4,5]. It employs factorization [3] with an extended measurement matrix, which contains spatio-temporal information on the object's deformation. Since the technique necessitates cameras to be fixed around the object concerned, it can only deal with motions/movements in a limited space. To overcome this disadvantage, shape recovery employing multiple mobile cameras is proposed in the paper.

The general idea of our approach is to introduce a 3-D recovery technique under a remote environment. In this paper, we describe the technique and shows a practical system of a human motion recovery based on the technique.

2 A Remote Motion Capture System

The process of a proposed remote motion capture is composed of two stages as illustrated in **Fig.1**. The first stage called a transfer stage captures human motions at a remote place. The second stage called a recovery stage tracks corresponding feature points between captured images transferred from the first stage. By factorization, orientations of the cameras are calculated first employing the tracked rigid feature points (or rigid points). Thus factorization is employed for weak camera calibration. The camera orientations are then used to compute directly the 3-D coordinates of the non-rigid feature points (or non-rigid points) specified on the non-rigid objects from their projected positions. Finally, recovery computation and representation of the human motion in a three-dimensional way is performed.

2.1 Motion Capturing – Mobile Computers

At the first basic step of the system, synchronization between the mobile computers are executed by using the command 'Net Time' of windows system OS. In the proposed system illustrated in **Fig.1**, clocks of client mobile computers are synchronized with a clock of a host mobile computer through a wireless network system (IEEE802.11b) in order to synchronize the start of image capture by mobile cameras, which are connected with the mobile computers. In addition, for acquiring motion images, we set the system to a specified time to control the mobile cameras starting the motion capturing. Obviously, some data such as the starting time for capturing, the number of the frame, and frame/image intervals must be set before running the synchronization among the cameras.

On the other hand, two mobile computers that are connected to the cameras sample the captured images (IEEE1394 Interface) in a specified interval and store the sampled images in the memory of the respective computers. Here, if the number of the acquired images reaches the specified number, the sampled images begin to be pre-

[1] Address: Sensuicho1-1, Tobata, Kitakyushu 804-8550, Japan. E-mail: **{etheltan, ishikawa}@is.cntl.kyutech.ac.jp**

served automatically in the hard disk of the computers, and the system stops the acquirement procedure of images. The stored images are then transferred to the recovery stage, *i.e.*, to a computer in a remote room through a network system (IEEE802.3) including a wireless LAN (Access Point).

2.2 Recovery Technique – Mobile Cameras

As a part of the recovery stage, F video cameras are employed and move freely around the object concerned to capture its video images. The mobile cameras move around the non-rigid objects under the constraint that they also capture several rigid points commonly visible from the cameras during observation.

The present technique contains two steps. Concerning the first step, rigid points are used to yield two matrices; matrix M and matrix S according to the literature [3]. In the second step, matrix M obtained from the first step is employed for calculating the 3-D positions of those feature points on non-rigid objects. It should be noted that the camera calibration employed in the present technique is not an ordinary calibration. It does not use the landmarks whose 3-D coordinates are known in advance. Instead it employs 2-D locations of the rigid points projected on the captured image sequences. Therefore the technique does not need the world coordinate system.

An image capturing strategy in the present technique is shown in **Figure 2**. Here, a set of F mobile cameras C_f (f=1,2,…,F; F≥2) takes images of a non-rigid object. Suppose that the F mobile cameras take l (l=1,2,…,L; L≥2) locations and these cameras remain in the same position for time T_l at the l'th location. If sampling interval is denoted by Δt, the sampling is done at every k (k=1,2,…, $T_l/\Delta t \equiv K_l$). Then according to [4,5], *a spatio-temporal measurement matrix* W_l at location l is defined by

$$W_l=\begin{pmatrix} W_{0l}^{R1} & W_{1l}^{N1} & \cdots & W_{Kl,l}^{N1} \\ W_{0l}^{R2} & W_{1l}^{N2} & \cdots & W_{Kl,l}^{N2} \\ \vdots & \vdots & \cdots & \vdots \\ W_{0l}^{RF} & W_{1l}^{NF} & \cdots & W_{Kl,l}^{NF} \end{pmatrix}. \quad (1)$$

Here matrix W_{0l}^{Rf} contains the xy coordinates of chosen rigid points on the image plane of the f'th (f=1,2,…,F) camera, whereas W_{kl}^{Nf} (k=1,2,…,K_l) contains the xy coordinates of the non-rigid points on a non-rigid object. They are all two rows matrices: The first row contains the x coordinates of the feature points on respective image planes of the cameras, whereas the second row holds the y coordinates.

Matrices W_l (l=1,2,…,L) are merged into a single matrix W and represented in the form of

"Shown in the final page" (2)

The matrix given by Eq.(2) is sorted out into rigid and non-rigid matrices. The rigid matrices $W_{0.}^{R}$ are then shifted to a single column (the first column, for example) as shown in Eq.(3);

"Shown in the final page" (3)

Note that this shift can be performed on condition that all the cameras observe identical rigid points during observation, irrespective of the cameras' location l.

The first column in Eq.(3) is further collected into a single matrix W^R defined by

$$W^R=\begin{pmatrix} W_{01}^{R1} \\ \vdots \\ W_{01}^{RF} \\ W_{02}^{R1} \\ \vdots \\ W_{02}^{RF} \\ \vdots \\ W_{0L}^{R1} \\ \vdots \\ W_{0L}^{RF} \end{pmatrix} \quad (4)$$

According to the literature [3], Eq.(4) is factorized into a camera orientation matrix and a shape matrix after having translated the world origin to the centroid of all the rigid points registered in W^R. Then the following description holds;

$$\tilde{W}^R=\begin{pmatrix} \tilde{W}_{01}^{R1} \\ \vdots \\ \tilde{W}_{01}^{RF} \\ \tilde{W}_{02}^{R1} \\ \vdots \\ \tilde{W}_{02}^{RF} \\ \vdots \\ \tilde{W}_{0L}^{R1} \\ \vdots \\ \tilde{W}_{0L}^{RF} \end{pmatrix}=\begin{pmatrix} M_1^1 \\ \vdots \\ M_1^F \\ M_2^1 \\ \vdots \\ M_2^F \\ \vdots \\ M_L^1 \\ \vdots \\ M_L^F \end{pmatrix}\cdot S^R, \quad (5)$$

where $\tilde{W}_{0l}^{Rf}$ denotes a translated measurement matrix at location l of camera f. Thereby M_l^f denotes a camera orientation matrix of camera f at location l and matrix S^R contains the 3-D coordinates of all the chosen rigid points.

Let us now consider the non-rigid matrices W_{kl}^{Nf} (k=1,2,…,K_l; l=1,2,…,L; f=1,2,…,F) in Eq.(3), *i.e.*, the second stage of our recovery technique. The matrix W_{kl}^{Nf} is also translated with respect to the centroid of the rigid points to yield $\tilde{W}_{kl}^{Nf}$. Then the following relation holds;

$$\begin{pmatrix} \tilde{W}_{k1}^{Nf} & & & \\ & \tilde{W}_{k2}^{Nf} & & \\ & & \ddots & \\ & & & \tilde{W}_{kL}^{Nf} \end{pmatrix} = \begin{pmatrix} M_1^f & & & \\ & M_2^f & & \\ & & \ddots & \\ & & & M_L^f \end{pmatrix}\begin{pmatrix} S_{k1}^N & & & \\ & S_{k2}^N & & \\ & & \ddots & \\ & & & S_{kL}^N \end{pmatrix} \quad (6)$$

Here S_{kl}^{Nf} contains the 3-D coordinates of all the non-rigid points observed at the k'th sampling time from camera f at location l, but they are unknown yet.

If Eq.(6) is expressed componentwise, we have

$$\tilde{W}_{kl}^{Nf} = M_l^f \cdot S_{kl}^N \quad . \tag{7}$$

Then

$$S_{kl}^N = M_l^{f^+} \cdot \tilde{W}_{kl}^{Nf} \tag{8}$$

holds for $l\,(l = 1, 2, ..., L)$. Here $M_l^{f^+}$ is a pseudo-inverse matrix of M_l^f. From Eqs.(5) and (8), all the points recover their 3-D positions. Therefore, the final result is written in the form of

$$S = \left(S^R \mid S_1^N \mid S_2^N \mid ... \mid S_L^N \right) , \tag{9}$$

where

$$S_l^N = \left(S_{1l}^N \mid S_{2l}^N \mid \cdots \mid S_{Kl,l}^N \right) . \tag{10}$$

3 Experimental Results

In the performed experiment, as illustrated in **Fig.2,** two mobile cameras (*i.e.* C_L and C_R) each connected with a mobile computer and a wireless LAN are employed to recover the motion of a human in a distant building (SVBL of KIT). One hundred images were sampled from each of the image streams with the interval of 0.1 second, where the images were directly stored into the memories of the connected note-PCs and then they were transferred through a wireless LAN and the internet to the lab in another building.

Eleven rigid points are specified on the floor and on the rear desk, *etc*., whereas 17 non-rigid points (small white balls of 25mm ϕ) are put on the subject. The moving video camera is controlled its trajectory so that it captures the rigid points all the time during observation. All the 28 points were tracked on the video images in the recovery stage to yield the matrix of Eq.(2). Some video images of the subject's motion are shown in **Fig.3**. The result of 3-D recovery is depicted in **Fig.4**. In both figures, the time proceeds as indicated by arrows.

4 Discussion and Conclusions

The paper proposed a remote motion capture system for 3-D shape recovery employing multiple mobile cameras. The system is able to recover 3-D shape of a non-rigid object like a human in motion as well as a rigid object. In the experiment, the proposed technique was applied to the motion recovery of a person in a distant building employing two cameras. The two cameras captured his motion in front of the desk in a distant research lab. The images were directly stored into the memories of the connected note-PCs and they were then transferred through a wireless LAN and the internet to the lab in another building. From the sampled image data, his 3-D motion recovered successfully. The two-camera system greatly reduced the amount of video image processing, particularly feature points tracking and finding correspondence between two images. These are indeed tough work in a multiple-camera system.

The result of the human motion recovery was almost satisfactory. Numerical estimation of the recovery errors is now under investigation. We have been working on 3-D motion recovery based on an extended measurement matrix and factorization [4] and have obtained about 4% of recovery errors under the assumption of orthographic projection with respect to camera imaging. The amount of 4% is therefore our present expectation with respect to the recovery error in this experiment.

We have already proposed a motion capture system employing mobile cameras [6]. In the system, more than three cameras are fixed on a mobile frame and the frame follows a person in motion to capture his images. It is a calibration free system, but the cameras must be fixed on the frame. On the other hand, simpler calibration with the employed cameras is necessary in the present motion capture system, but this system can be realized employing unconstrained cameras.

One of the main advantages of the proposed system is that ordinary camera calibration is not necessary for the recovery. It only employs rigid points in images for the calibration. This allows the motion capture be employed in broader areas of study particularly related to human motions modeling and/or analysis. Mobility of the camera system is another advantage over existent motion capture techniques, in which fixed cameras are always employed like a stereo vision system [7], for example. This mobility will also contribute to wide use of the proposed motion capture system in future.

References

[1] T. Iida, "Wireless communications R&D in the science and technology policy in Japan", *IEICE Trans. on Electronics*, **E85-C**, 3, 419-427, 2002.

[2] M. Minoh, "Model centered framework for three dimensional human body modeling", *Trans. on Computer Vision and Image Media*, **43**, SIG 4(CVIM 4), 10-23, 2002. (in Japanese)

[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, "Human motion recovery by the factorization based on a spatio-temporal measurement matrix", *Computer Vision and Image Understanding*, **82**, 2, 101-109, 2001.

[5] J. K. Tan, S. Ishikawa, "Recovering 3-D motion of team sports employing uncalibrated video cameras", *IEICE Trans. on Information and Systems*, **E84-D**, 12, 1728-1732, 2001.

[6] T. Tabusa, J. K. Tan, S. Ishikawa, "Three-dimensional recovery and analysis of a motion in a large field", *Imaging & Visual Computing*. (to appear)

[7] R. C. Gonzalez, P. Wintz, "*Digital Image Processing*", Addison-Wesley Publishing Co., 1986.

$$
W = \begin{pmatrix}
W_{01}^{R1} & W_{11}^{N1} & \cdots & W_{K1,1}^{N1} & & & & & & & & & \\
W_{01}^{R2} & W_{11}^{N2} & \cdots & W_{K1,1}^{N2} & & & & & & & & & \\
\vdots & \vdots & \cdots & \vdots & & & & & & & & & \\
W_{01}^{RF} & W_{11}^{NF} & \cdots & W_{K1,1}^{NF} & & & & & & & & & \\
 & & & & W_{02}^{R1} & W_{12}^{N1} & \cdots & W_{K2,2}^{N1} & & & & & \\
 & & & & W_{02}^{R2} & W_{12}^{N2} & \cdots & W_{K2,2}^{N2} & & & & & \\
 & & & & \vdots & \vdots & \cdots & \vdots & & & & & \\
 & & & & W_{02}^{RF} & W_{12}^{NF} & \cdots & W_{K2,2}^{NF} & & & & & \\
 & & & & & & & & \ddots & & & & \\
 & & & & & & & & & W_{0L}^{R1} & W_{1L}^{N1} & \cdots & W_{KL,L}^{N1} \\
 & & & & & & & & & W_{0L}^{R2} & W_{1L}^{N2} & \cdots & W_{KL,L}^{N2} \\
 & & & & & & & & & \vdots & \vdots & \cdots & \vdots \\
 & & & & & & & & & W_{0L}^{RF} & W_{1L}^{NF} & \cdots & W_{KL,L}^{NF}
\end{pmatrix} \tag{2}
$$

$$
W' = \begin{pmatrix}
W_{01}^{R1} & W_{11}^{N1} & \cdots & W_{K1,1}^{N1} & & & & & & \\
W_{01}^{R2} & W_{11}^{N2} & \cdots & W_{K1,1}^{N2} & & & & & & \\
\vdots & \vdots & \cdots & \vdots & & & & & & \\
W_{01}^{RF} & W_{11}^{NF} & \cdots & W_{K1,1}^{NF} & & & & & & \\
W_{02}^{R1} & & & & W_{12}^{N1} & \cdots & W_{K2,2}^{N1} & & & \\
W_{02}^{R2} & & & & W_{12}^{N2} & \cdots & W_{K2,2}^{N2} & & & \\
\vdots & & & & \vdots & \cdots & \vdots & & & \\
W_{02}^{RF} & & & & W_{12}^{NF} & & W_{K2,2}^{NF} & & & \\
\vdots & & & & & & & \ddots & & \\
W_{0L}^{R1} & & & & & & & W_{1L}^{N1} & \cdots & W_{KL,L}^{N1} \\
W_{0L}^{R2} & & & & & & & W_{1L}^{N2} & \cdots & W_{KL,L}^{N2} \\
\vdots & & & & & & & \vdots & \cdots & \vdots \\
W_{0L}^{RF} & & & & & & & W_{1L}^{NF} & \cdots & W_{KL,L}^{NF}
\end{pmatrix} \tag{3}
$$

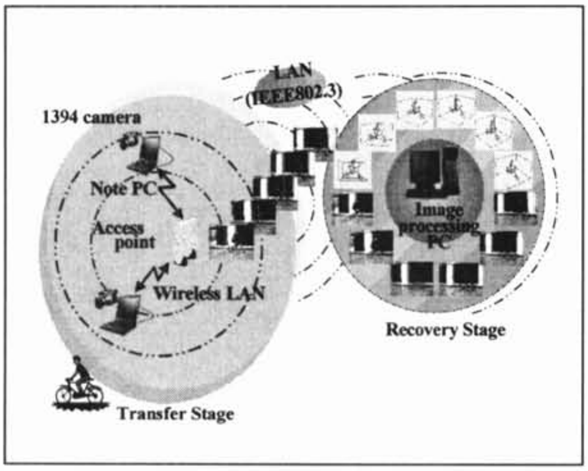
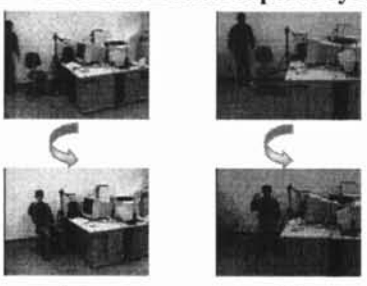


Fig.1 A remote motion capture system.

Fig.2 Image capture by two cameras.



(a) (b)

Fig.3 A video image of a person in motion; (a) camera C_L, and (b) camera C_R.

Fig.4 Recovered motion.