VARIOUS IMAGE TAKING STRATEGIES FOR 3-D OBJECT MODELING BASED ON MULTIPLE CAMERAS

Joo Kooi Tan, Iku Yamaguchi*, Tomonori Tabusa**, Seiji Ishikawa*, Shunji Hirokawa****

*Department of Mechanical and Control Engineering, Kyushu Institute of Technology
Sensuicho 1-1, Tobata, Kitakyushu 804-8550, Japan
**Yuge National College of Technology, Yuge, Ehime 794-2593, Japan
***Department of Intelligent Machinery and Systems, Kyushu University, Fukuoka 810-8560, Japan

ABSTRACT

A method is described for obtaining 3-D object models, particularly human models, based on multiple cameras and a computer. Existing optical 3-D modeling systems employ calibrated cameras, which inevitably restricts the cameras at fixed positions, giving constraints to object motions. The proposed technique gives various ways of capturing images of object motions. It contains a mobile cameras system for taking images of an object moving widely by following it. It also provides a simultaneous entire shape recovery strategy by the employment of a surrounding cameras system. The proposed recovery technique puts its theoretical basis on the factorization method. The present paper focuses its attention on the issue of defining a measurement matrix having full entries. The technique is described and its performance is shown by the experiments on 3-D human motion modeling.

1. INTRODUCTION

Three-dimensional shape recovery techniques of non-rigid objects have quite remarkably developed in recent years. Particularly human motion recovery has many 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 always necessitates camera calibration, which is not very convenient particularly for outdoor use. Alternatively motion recovery employing magnetic sensors is also a common technique. It restricts motion range of the subject, though. Non-contact techniques based on optical measurement that can cover wide range motions are obviously better for extensive use. The present paper offers a non-contact optical technique with much more flexibility in image taking.

We have already proposed a shape recovery technique of 3-D non-rigid objects based on multiple uncalibrated cameras [2,3]. It employs a factorization method [1] with an extended measurement matrix [2] that 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 the motions/ movements in a limited space. This disadvantage can be overcome by the employment of multiple mobile cameras.

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 various image taking strategies drown from a general form of a measurement matrix and proposes a multiple mobile cameras system as a most convenient way of taking images of an object. The method is described and some experimental results are shown.

2. M-MATRIX — A GENERAL FORM

A measurement matrix is abbreviated as *an m-matrix* hereafter. Suppose that F cameras take images of an object by changing their locations L times. If we denote an m-matrix of the F cameras at location l by W_l, a general form of an entire m-matrix is given by the following formula, since camera orientations are generally different from each other among the L locations.

$$W = \begin{pmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_L \end{pmatrix}. \qquad (1)$$

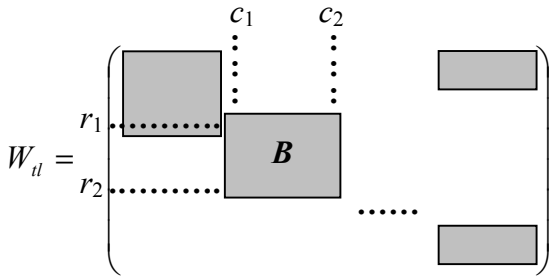
Here rows of matrix W correspond to camera orientations, whereas columns represent feature points on the object concerned. The remaining entries of W are all vacant.

Let us denote the observation time at location l by T_l and assume the sampling interval 1. Then an m-matrix W_l is written as follows;

$$W_l = \begin{pmatrix} W_{1l} & W_{2l} & \cdots & W_{T_l l} \end{pmatrix}. \qquad (2)$$

0-7803-8554-3/04/$20.00 ©2004 IEEE.

Matrix W_{tl} represents an m-matrix at sample time t and location l and its entries can be illustrated as follows;



. (3)

In the above matrix, sub-matrix B, for example, occupies the entries from the r_1'th row to the r_2'th row and from the c_1'th column to the c_2'th column. This means that, at sample time t, the cameras having the orientations corresponding to the rows from r_1 through r_2 observed the common feature points corresponding to the columns from c_1 through c_2. Thus the sub-matrix B is a full sub-matrix. The sub-matrix B, denoted by W_{tl}^{B}, contains the xy image coordinates of a feature point in the following form;

$$W_{tl}^{B} = \begin{pmatrix} x_{r_1,c_1}(t) & x_{r_1,c_1+1}(t) & \cdots & x_{r_1,c_2}(t) \\ x_{r_1+1,c_1}(t) & x_{r_1+1,c_1+1}(t) & \cdots & x_{r_1+1,c_2}(t) \\ \vdots & \vdots & \ddots & \vdots \\ x_{r_2,c_1}(t) & x_{r_2,c_1+1}(t) & \cdots & x_{r_2,c_2}(t) \\ y_{r_1,c_1}(t) & y_{r_1,c_1+1}(t) & \cdots & y_{r_1,c_2}(t) \\ y_{r_1+1,c_1}(t) & y_{r_1+1,c_1+1}(t) & \cdots & y_{r_1+1,c_2}(t) \\ \vdots & \vdots & \ddots & \vdots \\ y_{r_2,c_1}(t) & y_{r_2,c_1+1}(t) & \cdots & y_{r_2,c_2}(t) \end{pmatrix}_{l}^{B} \quad (4)$$

In this way, matrix W_{tl} of Eq.(3) represents the image coordinates of the feature points on an object commonly observable from a subset of cameras at location l, given a sample time t.

A general form of an m-matrix W of Eq.(1) cannot further be processed as it contains vacant entries within it. There are, however, some ways of employing the matrix W into 3-D shape/motion recovery. The following sections give some ideas on the use.

3. M-MATRIX FOR MOTION RECOVERY

The factorization technique originally proposed in [1] is the most simple case of Eq.(1), *i.e.*, L=1 in Eq.(1), T=1 in Eq.(2), and W_{11} of Eq.(3) is a single full matrix having the xy coordinates of the feature points on a rigid object. Then this simplest m-matrix W is decomposed into two matrices, M, a camera orientation matrix, and S, a shape matrix as

$$W = M \cdot S \quad (5)$$

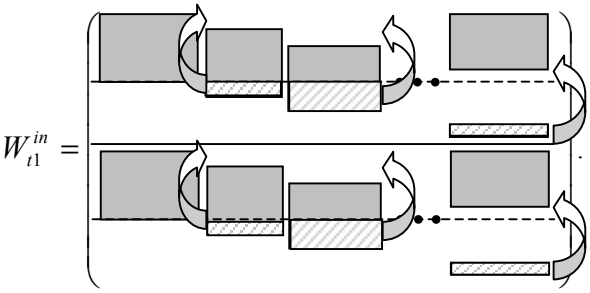
according to literature [1]. Since the matrix S contains the xyz coordinates of all the chosen feature points, the object recovers its shape. Note that the values in W are transformed into those values based on the coordinate system whose origin is the centroid of the feature points that spread on the object.

In the case of L=1 in Eq.(1), T≥2 in Eq.(2), and W_{t1} (t=1,2,…,T) of Eq.(3) are all full matrices, it can represent motion or deformation of the object concerned [2,3].

4. M-MATRIX FOR ENTIRE SHAPE RECOVERY

Let us consider the issue on acquiring an entire (or a front and rear) 3-D model of an object. Normally a registration technique [4] is employed for this purpose. The technique connects partial 3-D recovered shapes of the object into an entire 3-D model. This is a sequential procedure and therefore time consuming. But, if an even number of cameras, say 2F cameras, are placed around an object and if the light axes agree with every opposite pair of cameras, the entire 3-D model recover simultaneously by applying the factorization once to the defined m-matrix.

Normally there are very few feature points on an object visible from all the 2F cameras around the object. Since some successive cameras have commonly visible feature points, an initial matrix of W_{t1}, W_{t1}^{in}, has the style of

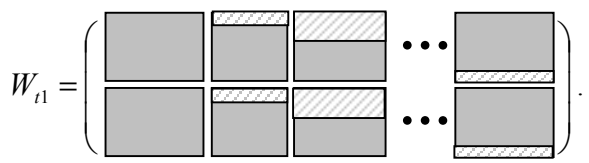


. (6)

Here each gray and hatched sub-matrix in the upper half contains the x coordinates of the feature points commonly visible from the camera orientations specified by the row numbers where the sub-matrix exists, whereas that in the lower half contains the y coordinates.

If we assume orthographic projection with respect to the lens imaging system of a camera, the lower part of the sub-matrices under the broken lines in Eq.(6), illustrated by hatched blocks, can be relocated to the part above the sub-matrix, as shown by the arrows, yielding the matrix W_{t1} given by Eq.(7).

Thus we have a full matrix W_{t1}. This matrix can be factorized as shown by Eq.(5), and the shape matrix S gives the 3-D coordinates of all the chosen feature points spread in front and at the rear of the object concerned.



. (7)

5. MOBILE CAMERAS SYSTEM I

The above measurement system places cameras at fixed positions like existent 3-D measurement systems commercially available. This of course restricts the motion of a person whose 3-D model is the present concern. A convenient way of taking images of such a moving object is to make the cameras move along with the object. To realize this, cameras are placed fixed on a mobile frame in this measurement system. Since the cameras are fixed on the frame, the camera orientations are relatively unchanged.

As an easy way of realizing this mobile camera system, the cameras are supposed to observe a certain number of specified feature points spread in front of the object concerned and track the feature points.

Since the frame can rotate and translate in a parallel way, the initial m-matrix is given by Eq.(1) and Eq.(2), where $L \geq 1$, $T \geq 1$, and matrix W_{tl} is a full matrix of the form

$$W_{tl} = \begin{pmatrix} \blacksquare & \blacksquare & \blacksquare & \bullet\bullet\bullet & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \bullet\bullet\bullet & \blacksquare \end{pmatrix}. \quad (8)$$

Since W_{tl} (t=1,2,…,T) are full matrices, W_l (l=1,2,…,L) are also full matrices. Matrix W of Eq(1) is not a full matrix, however, resulting in no further calculation.

To solve this difficulty, the world coordinate system is placed 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 can be represented as that all the sub-matrices W_l (l=1,2,…,L) are aligned in the same rows as follows;

$$W = \begin{pmatrix} W_1 & W_2 & \cdots & W_L \end{pmatrix}. \quad (9)$$

This provides a full m-matrix W. This matrix is factorized according to Eq.(5) and the obtained shape matrix S gives 3-D coordinates of the motion of the object relative to the cameras.

6. MOBILE CAMERAS SYSTEM II

The final case of 3-D shape recovery by multiple cameras is that mobile cameras take images of an object while moving independently. In this case, there is no way of producing a full m-matrix, since the camera orientations are all different through the camera movement from 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 must perform camera calibration using 3-D tools, the proposed technique doesn't employ such tools for the calibration. Instead it employs rigid points observed in the captured image streams. The procedure is described in the following.

At every position l (l=1,2,…,L), the cameras are required to observe identical feature points on static objects (referred to as *static feature points*) other than the feature points on non-rigid objects (referred to as *moving feature points*). 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} with respect to sub-matrix W_l of Eq.(2). Then W_l can be rewritten as

$$W_l = \begin{pmatrix} W_l^{\mathrm{R}} & W_l^{\mathrm{N}} \end{pmatrix}. \quad (10)$$

This is an alternative expression of Eq.(2), since W_l^{R} is the collection of those static feature points contained in W_{tl} (t=1,2,…,T) of Eq.(2). If $W_l^{\mathrm{R}} \equiv W_{0l}$, Eq.(2) can be rewritten as follows;

$$W_l = \begin{pmatrix} W_{0l} & W_{1l} & W_{2l} & \cdots & W_{Tl} \end{pmatrix}. \quad (11)$$

Here W_{tl} (t=1,2,…,T) include only the coordinates of moving feature points.

Employing Eq.(10), Eq.(1) is written in the form of

$$W = \begin{pmatrix} W_1^{\mathrm{R}} & W_1^{\mathrm{N}} & & & \\ & & W_2^{\mathrm{R}} & W_2^{\mathrm{N}} & \\ & & & & \ddots \\ & & & & W_L^{\mathrm{R}} \; W_L^{\mathrm{N}} \end{pmatrix}. \quad (12)$$

As W_l^{R} (l=1,2,…,L) contains identical static feature points, Eq.(12) is rewritten as

$$W = \begin{pmatrix} W_1^{\mathrm{R}} & W_1^{\mathrm{N}} & & & \\ W_2^{\mathrm{R}} & & W_2^{\mathrm{N}} & & \\ \vdots & & & \ddots & \\ W_L^{\mathrm{R}} & & & & \mathrm{W}_L^{\mathrm{N}} \end{pmatrix}. \quad (13)$$
$$\equiv \begin{pmatrix} W^R & W^N \end{pmatrix}$$

Employing Eq.(5), we have

$$W^{\mathrm{R}} = M \cdot S^{\mathrm{R}}, \quad (14)$$

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 matrix M by

$$M = \begin{pmatrix} M_1 & M_2 & \cdots & M_L \end{pmatrix}^{\mathrm{T}}. \quad (15)$$

Then we have

$$W_l^{\mathrm{N}} = M_l \cdot S_l^{\mathrm{N}}. \quad (16)$$

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

$$S_l^{\mathrm{N}} = M_l^{+} \cdot W_l^{\mathrm{N}}. \quad (17)$$

In this way, all the feature points registered in the m-matrix of Eq.(10) or Eq.(11) recover their 3-D positions. They are given in the form of S^{R} and S_l^{N} (l=1,2,..,L).

7. EXPERIMENTAL RESULTS

In the performed experiment, we aimed at the 3-D recovery of a human motion and its environment in a yard. A subject was instructed to move around in the yard. As shown in **Fig.1**, the subject's motion was taken images by two mobile cameras (*i.e.* C_L and C_R) that were connected to video transmitters, respectively. The video images were sent through the transmitter to two PCs in a distant room via the video receivers connected to the PCs. One-hundred images were sampled there from each of the image streams with the interval of 0.1 second for recovery calculation.

Twenty-four rigid points were specified on the three poles and on the two wires, whereas 17 non-rigid points (small white balls of 25mm ϕ) were put on the subject. The moving video camera was controlled its trajectory so that it captured the rigid points all the time during the observation. All the 41 points were tracked on the video images in the recovery stage to yield the matrix of Eq.(12). The feature points are specified manually in the initial images. They are then tracked using correlation calculation and manual check. Once the feature points are occluded, their locations can be inferred by the auto-regression technique, which is not employed here though, since observable feature points have been used in this particular experiment. Full automation of these procedures needs further investigation.

Some video images of the subject's motion are shown in **Fig.2**. The result of 3-D recovery is depicted in **Fig.3**. In both figures, the time proceeds as indicated by arrows.

8. DISCUSSION AND CONCLUSIONS

The paper proposed a 3-D object modeling system employing multiple cameras. The modeling technique is based on the factorization method that employs an m-matrix. The m-matrix needs to be a full matrix in order to be decomposed into two principal matrices. Therefore the paper has shown various ways of making full m-matrices that correspond to respective image taking strategies using multiple cameras.

In the experiment, two independently moving cameras system was employed and it performed 3-D modeling of an office environment including a human. The result was almost satisfactory. The recovery errors have been under investigation. We have been working on 3-D non-rigid objects recovery based on an m-matrix and factorization [2,3], and have obtained about 4% of the recovery errors under the assumption of orthographic projection with respect to camera imaging. This amount is therefore our present expectation with respect to the recovery error in this experiment.

Main advantages of the proposed 3-D object modeling techniques over others include that (i) camera calibration employing 3-D tools is not necessary for the recovery, that (ii) the entire motion as well as the frontal and the rear shape during observation time recovers simultaneously from a single m-matrix by applying factorization once to it, and that (iii) wide range motion can recover by the employment of mobile cameras systems. These facts may lead the proposed techniques to popular use in broader areas related to human motions modeling and/or analysis.

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

9. 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.

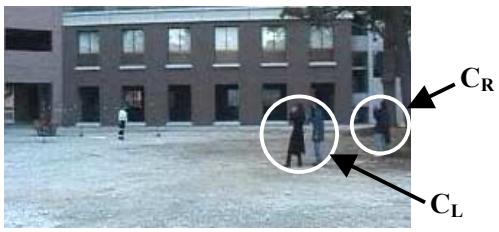


Fig.1 Mobile cameras system II.

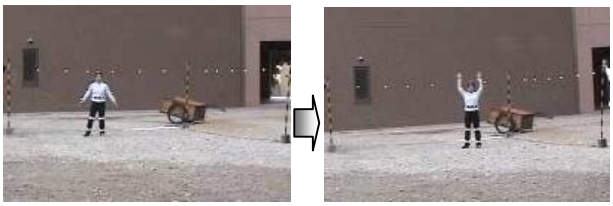


Fig.2 Video images of a person in motion.

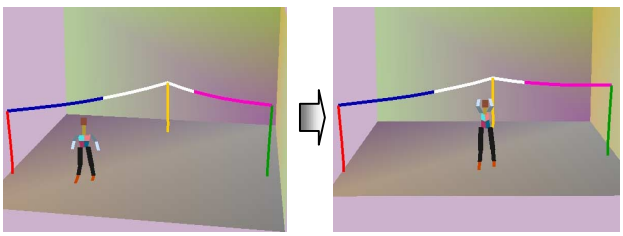


Fig.3 Recovered 3-D motion.