MVA'98 IAPR Workshop on Machine Vision Applications, Nov. 17-19, 1998, Makuhari, Chiba, Japan

8—8 An Efficient Technique for Motion Recovery Based on Multiple Views

Joo Kooi Tan*, Shinji Kawabata*, Seiji Ishikawa*

Department of Mechanical and Control Engineering
Kyushu Institute of Technology

Abstract

The factorization method is a technique for recovering 3-D shape of an undeformable object without using camera parameters. The technique is extended to an efficient method in this paper for recovering 3-D motion of a deformable object, *i.e.*, a human. The proposed system is composed of three fixed video cameras which take video images of a human motion. Three obtained image sequences are analyzed to yield measurement matrices at individual sampling times, and they are merged into a single measurement matrix to which the factorization is applied and the 3-D human motion is recovered at a time. Experimental results were satisfactory.

1. Introduction

Three-dimensional human motion recovery is an important as well as attractive subject of study. Applications may as well be expected to biological motion analysis, to behavioral or athletic science, or even to rehabilitation.

Various computer vision techniques[1] give basic ideas of recovering algorithms, all of which suffer from strict camera calibration. A novel technique employing factorization[2] recovers 3-D shape of an undeformable object in the uncalibration state. It can therefore offer an easy instrumentation in realizing a 3-D undeformable shape recovery system. As for deformable objects, however, they are still out of the scope in the application of the technique[2,3].

In this paper, the factorization is employed for recovering human motions. In principle, repeated use of the factorization in the time sequence results in 3-D shape recovery of deformable objects[4]. It does not, however, provide numerically stable results. Since recovering procedure is repeated independently, it causes independent recovery errors and the recovered motion often looses smoothness in the time lapse. In this paper, two algorithms for motion recovery employing factorization are presented and their performances are compared experimentally. One of the two algorithms, originally appeared in [5], is shown to give an efficient way for motion recovery compared with the other.

2. Two Algorithms of Motion Recovery Based on Factorization

In the proposed motion recovery system, we set F digital video cameras at arbitrary location around a subject. Suppose a subject's motion is taken and digital video camera f $(f=1,2,\ldots,F)$ produces an image stream $I_f(t)(t=1,2,\ldots,T)$. Let a feature point on image $I_f(t)$ be denoted by $(x_{fp}(t), y_{fp}(t))$ $(p=1,2,\ldots,P)$. Note that, for simplicity, the number of feature points P is fixed with respect to the parameters f and t in this particular study.

What should be prepared for the shape recovery is the measurement matrix of the size $2F\times P$ given at sampled time t from F image frames, *i.e.*,

$$\tilde{W}(t)=\begin{pmatrix} \cdots\cdots \\ \tilde{x}_{fp}(t) \\ \cdots\cdots \\ \tilde{y}_{fp}(t) \\ \cdots\cdots \end{pmatrix} \tag{1}$$

where,

$$\begin{aligned} \tilde{x}_{fp}(t) &= x_{fp}(t)-\frac{1}{PT}\sum_{t=1}^{T}\sum_{p=1}^{P}x_{fp}(t)\ , \\ \tilde{y}_{fp}(t) &= y_{fp}(t)-\frac{1}{PT}\sum_{t=1}^{T}\sum_{p=1}^{P}y_{fp}(t) \end{aligned} \tag{2}$$

* Address : Sensuicho 1-1,Tobata, Kitakyushu 804-8550, Japan.
E-mail : etheltan@is.cntl.kyutech.ac.jp
E-mail : ishikawa@is.cntl.kyutech.ac.jp

Note that the center of PT feature points are taken into account in the definition of $\tilde{x}_{fp}(t)$ and $\tilde{y}_{fp}(t)$.

If we take the world origin at the centroid of $\mathbf{s}_p(t)$ $(p=1,2,\ldots,P\ ;\ t=1,2,\ldots,T)$, $\sum_t \sum_p \mathbf{s}_p(t)=0$ and we have

$$\tilde{x}_{fp}(t)=(\boldsymbol{i}_f,\mathbf{s}_p(t)),$$
$$\tilde{y}_{fp}(t)=(\boldsymbol{j}_f,\mathbf{s}_p(t))$$

where $(\boldsymbol{i}_f,\boldsymbol{j}_f,\boldsymbol{k}_f)$ is the coordinate system of camera f and $\boldsymbol{k}_f=\boldsymbol{i}_f\times\boldsymbol{j}_f$. Hence we have

$$\tilde{W}(t)=\begin{pmatrix}\cdots\cdots\\ \boldsymbol{i}_f{}^T\\ \cdots\cdots\\ \boldsymbol{j}_f{}^T\\ \cdots\cdots\end{pmatrix}\begin{pmatrix}\cdots & \mathbf{s}_p(t) & \cdots\end{pmatrix}\equiv MS(t)\ . \quad (3)$$

Here T denotes transpose.

Equation(3) indicates that factorization can be applied to $\tilde{W}(t)$ given by Eq.(1) with respect to $t=1,2,\ldots,T$. This algorithm is referred to as *Algorithm R* hereafter. It is advantageous that all the recovered feature points $\mathbf{s}_p(t)$ $(p=1,2,\ldots,P\ ;\ t=1,2,\ldots,T)$ have the common origin at $\sum_t\sum_p \mathbf{s}_p(t)=0$. It is, however, likely that the obtained respective matrices M at $t(t=1,2,\ldots,T)$ in Eq.(3) may not exactly coincide with each other due to numerical errors.

Instead of employing Eq.(3), we define the measurement matrix in the form of

$$\tilde{W}=(\tilde{W}(1)\,|\,\tilde{W}(2)\,|\cdots|\,\tilde{W}(T))$$
$$=\left(\begin{array}{c|c|c|c}\cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\ \tilde{x}_{fp}(1) & \tilde{x}_{fp}(2) & \cdots\cdots & \tilde{x}_{fp}(T)\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\\ \tilde{y}_{fp}(1) & \tilde{y}_{fp}(2) & \cdots\cdots & \tilde{y}_{fp}(T)\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots\end{array}\right) \quad (4)$$

where $\tilde{x}_{fp}(t)$ and $\tilde{y}_{fp}(t)$ are defined by Eq.(2). Then, substituting Eq.(3) into Eq.(4), we have

$$\tilde{W}=M(S(1)\,|\,S(2)\,|\cdots|\,S(T))$$
$$\equiv MS \quad (5)$$

This equation tells us that factorization can be applied to $\tilde{W}$ given by Eq.(4) only once, which provides us with matrices M and S in Eq.(5) and we obtain PT recovered feature points $\mathbf{s}_p(t)$ $(p=1,2,\ldots,P\ ;\ t=1,2,\ldots,T)$ at a time. This algorithm is referred to as *Algorithm S*.

Motion recovery by *Algorithm S* is likely to be numerically stable compared with the recovery by *Algorithm R* because of a single application of factorization. Computation time, on the other hand, seems larger in the former than in the latter, since most of the computation in factorization is devoted to singular value decomposition(SVD) and SVD takes much time for a matrix of a large size like $\tilde{W}$ in Eq.(4).

3. Experimental Results

In the experiment, two kinds of human motion are used to evaluate the algorithms discussed in **2**. A subject who is standing up and sitting down on the chair in an indoor environment, and a subject who is walking in an outdoor environment are recovered their motions by *Algorithm R* and *Algorithm S*. In the performed experiment, three digital video cameras (*i.e.*, $F=3$) are placed at suitable positions where all the feature points are visible during the motion. The motion images are taken by the video cameras and they are fed into a personal computer via a video capture board. Each sequence of images are then forwarded to a workstation and the three images at each sampled time are analyzed on the display to find corresponding feature points $((x_{fp}(t),y_{fp}(t))$ in Eq.(2). Fifteen feature points (*i.e.*, $P=15$) are attached on a subject's body. Their correspondence is found one by one manually using a mouse. The standing up and sitting down on the chair motion is taken video during 6.5 seconds. On the other hand, the walking motion is taken video during 4.0 seconds and the respective video image is sampled every 0.1 second (*i.e.*, $T=65$ and $T=40$, respectively). All the video cameras are synchronized by turning on a flashlight in the beginning of a subject's motion.

Figure 1 and **Fig. 4** show three views of the subject in motion at a certain time. Each of the image size is 640×480 pixels. Both of the motions were successfully recovered by *Algorithm S* as well as *algorithm R*. The computation times by Sparc Station 20 from feeding $\tilde{W}(t)$ into the program till reporting S in Eq.(5) are shown in **Table 1** and **Table 2** respectively.

All the recovered feature points by *Algorithm S* and part of the recovered motion in the time sequence are

Table1: The computation time by *Algorithm R*

Standing and sitting down	Walking
0.799968 sec	0.400048 sec

Table2: The computation time by *Algorithm S*

Standing and sitting down	Walking
6.149754 sec	2.349906 sec

illustrated in **Fig.2**(a) and (b) for standing and sitting down motion, whereas the result of walking motion is shown in **Fig.5**(a),(b).

4. Discussion and Conclusion

We have proposed two algorithms employing factorization for motion recovery and performed recovery of the human motion mentioned in the former section by the two algorithms. Precision of the recovery was examined in the following way.

Let us denote the $x\,y$-coordinates of feature point $\mathbf{s}_p$ at time t on the image plane of camera f by $\tilde{\boldsymbol{m}}_{fp}(t)=[\tilde{x}_{fp}(t),\tilde{y}_{fp}(t)]^T$ and the recovered feature point $\mathbf{s}_p$ projected onto the same image plane by $\hat{\boldsymbol{m}}_{fp}(t)=[\hat{x}_{fp}(t),\hat{y}_{fp}(t)]^T$. To obtain the latter, $S_f=R_fS=(\boldsymbol{i}_f,\ \boldsymbol{j}_f,\ \boldsymbol{k}_f)^T S$ need be calculated with respect to the obtained S in Eq.(5). Then the distances of $\tilde{\boldsymbol{m}}_{1p}(t)$ and $\hat{\boldsymbol{m}}_{1p}(t)$ from the origin are given by

$$\begin{aligned}\tilde{\ell}_{1p}(t)&=\sqrt{\tilde{\boldsymbol{m}}_{1p}^T(t)\cdot\tilde{\boldsymbol{m}}_{1p}(t)}\\ \hat{\ell}_{1p}(t)&=\sqrt{\hat{\boldsymbol{m}}_{1p}^T(t)\cdot\hat{\boldsymbol{m}}_{1p}(t)}\end{aligned}\quad , \tag{6}$$

respectively. Employing them, the relative recovery error, $e_p(t)$, is defined by

$$e_p(t)=\left|\frac{\hat{\ell}_{1p}(t)}{\hat{\ell}_{11}(1)}-\frac{\tilde{\ell}_{1p}(t)}{\tilde{\ell}_{11}(1)}\right|\Big/\frac{\tilde{\ell}_{1p}(t)}{\tilde{\ell}_{11}(1)}\ . \tag{7}$$

The values of $e_p(t)$ with respect to $PT=975$ feature points are depicted in **Fig.3**, in which (a) corresponds to the result obtained by the application of *Algorithm S*, whereas (b) corresponds to *Algorithm R*. As is clearly seen, *Algorithm S* has less recovery errors than *Algorithm R*.

However, **Fig.6** ($PT=600$ feature points) shows that both of the algorithms have less recovery errors compared with Fig.3. We consider that even *Algorithm R* recovered the motion accurately in this case because of longer distances between three video cameras and the subject. However there occurred certain lack of stability in the recovered motion. In order to obtain smooth motion, transformation of some camera coordinate systems needed at several sampled times. On the contrary, such transformation is not necessary in the application of *Algorithm S*.

As a conclusion, *Algorithm S* based on Eqs.(2),(4),(5) has better performance compared with *Algorithm R* based on Eqs.(1),(2),(3) in point of numerical stability and precision of recovery. Although *Algorithm S* consumes computation time more than *Algorithm R*, the latter needs post-processing such as the coordinates transformation in order to recover the original motion. The reason why the former takes time is mostly due to the singular value decomposition of a large-size matrix $\tilde{W}$ in Eq.(4).

Algorithm S based on a single use of factorization is effective in motion recovery. Numerical instability of *Algorithm R* needs further investigation. The project is proceeding to realization of a system with more video cameras to eliminate dead angles for feature points on the subject in the variety of motions.

Part of this work is supported by The Sasakawa Scientific Research Grant from The Japan Science Society.

References

[1] B. K. P. Horn, *Robot vision*, MIT Press, Cambridge, MA(1986).

[2] C. Tomasi, T. Kanade, "Shape and motion from image streams under orthography: A factorization method", *Int'l J. Computer Vision*, **9**, 2, 137-154 (1992).

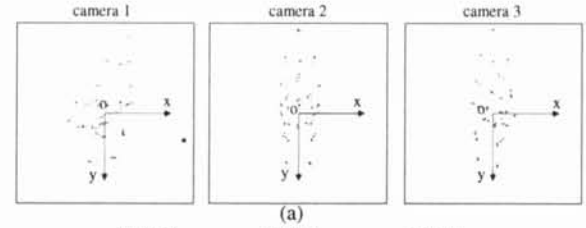
[3] T. Kanade, T. Morita, "Three-dimensional shape and motion recovery from image streams", *J. Inst. Electron., Inform., and Commun. Eng.*, **80**, 5, 479-487 (1997). (in Japanese)

[4] J. K. Tan, S. Ishikawa, *et al.*, "Recovering human motion by factorization", *Proc. 3rd Sympo. on Sensing via Image Inform.*, 111-114, Yokohama (June, 1997).

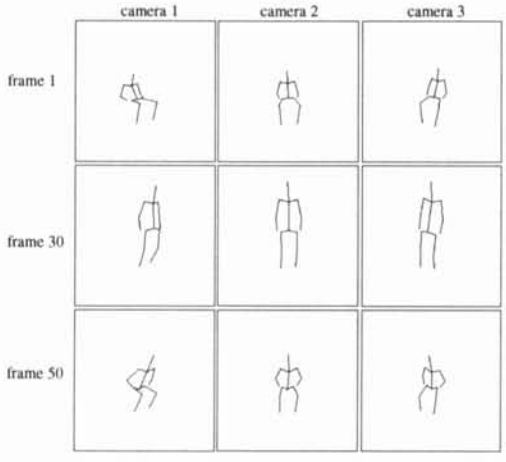
[5] J. K. Tan, S. Kawabata, S. Ishikawa, "Recovering shape of deformable objects by factorization", *J. Inst. Image Inform. and Tele. Eng.*, **52**, 3,406-408 (1998).(in Japanese)

(a) (b) (c)

Fig.1 Three views of a subject taken by three video cameras at a certain time:
(a) The left view,(b) the center view, and (c) the right view.

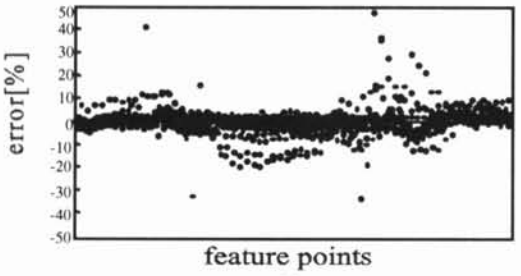


(a)

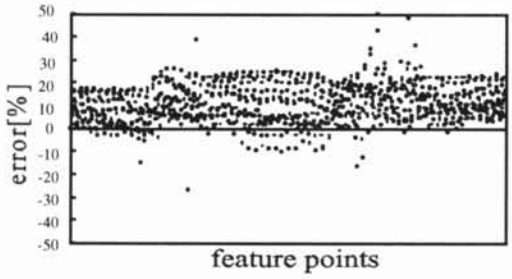


(b)

Fig. 2 Recovered motion employing *Algorithm S*:
(a) All the recovered feature points at each camera, and
(b) part of the motion sequence.



(a)

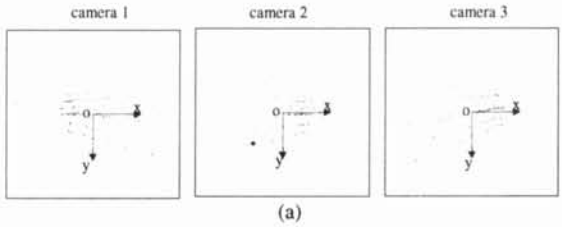


(b)

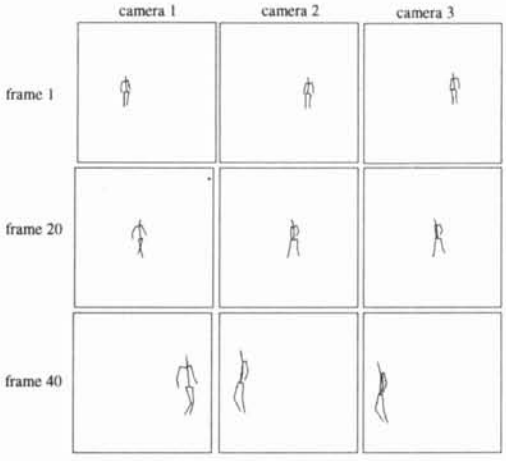
Fig.3 Errors of all the recovered feature points :
(a) *Algorithm S*, and (b) *Algorithm R*.

(a) (b) (c)

Fig.4 Three views of a subject taken by three video cameras at a certain time:
(a)The left view,(b) the center view, and (c) the right view.

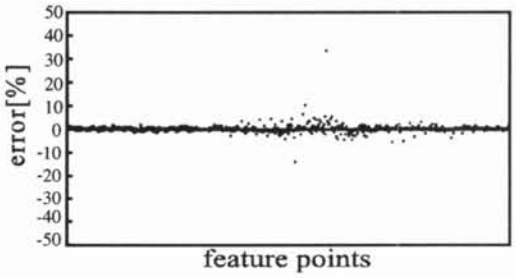


(a)

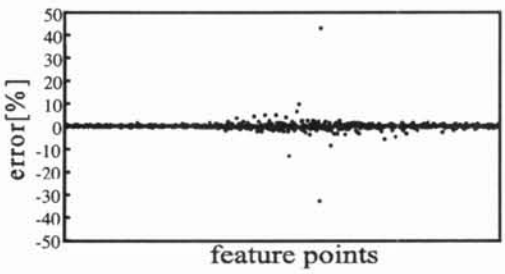


(b)

Fig. 5 Recovered motion employing *Algorithm S*:
(a) All the recovered feature points at each camera, and
(b) part of the motion sequence.



(a)



(b)

Fig.6 Errors of all the recovered feature points :
(a) *Algorithm S*, and (b) *Algorithm R*.