Entire Shape Recovery Employing Virtual See-through Cameras

K. Mori, T. Sone, J. K. Tan, H. Kim, S. Ishikawa

Kyushu Institute of Technology
1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka 804-8550, Japan
(Tel: 81-93-884-3183; Fax: 81-93-884-3183)
(ishikawa@ctl.kyutech.ac.jp)

Abstract: The present paper describes a novel technique for simultaneous entire shape recovery of an object by use of multiple cameras surrounding the object. The rear shape is transformed to frontal view information and it is merged with the frontal view information into an entire shape information to which factorization is applied. This camera system is named a virtual see-through camera system, since the frontal cameras see through the rear shape as if the object is transparent. Experimental results show satisfactory performance of the technique.

Keywords: shape recovery, 3-D modeling, see-through, factorization, computer vision

I. INTRODUCTION

Three-dimensional modeling of our living environment is of great importance in simulating our various activities and planning better life. When we want to buy furniture, we are able to see its fitness in a house by looking into a virtual house equipped with the furniture. Creating a new building should be examined its appearance and fitness in the area where it is to be built by modeling the area and the building in a virtual space. These are the examples of useful applications of a virtual reality technology, in which various 3-D object models are created and employed.

Hence 3-D object modeling techniques employing cameras [1] have been studied rigorously to date in the computer vision community. The technique employing stereo cameras has long been used as a standard technique. This has become much popular since the proposal of the DLT [2] by which one doesn't have to acquire inner as well as outer camera parameters directly. Instead it employs a 3-D calibration tool and acquires the projective relation between the 3-D space and the image plane of an observing camera. The factorization method [3] further simplifies the 3-D modeling strategy. In the method, a single mobile camera takes images of a rigid object from multiple orientations and the method recovers the 3-D shape of the commonly observed part of the object along with recovering the camera motion, although it contains linear approximation in imaging by the camera.

The issue focused in the present paper is 3-D modeling of the entire shape of an object. Normally this is achieved by a set of cameras surrounding the object concerned. In an optical motion capture system, many calibrated cameras are placed fixed in a large studio and capture video images of an actor playing some action from multiple orientations. The basic recovery technique is stereo vision. Every pair of stereo cameras recovers the shape of the object within its view. The shape is partial in most cases, since the fixed camera pair cannot observe an opposite side of the object. The resultant partial 3-D shapes are merged into an entire shape, yielding a 3-D model of the object. In this procedure, registration technique [4] must be employed in order to connect the partial shapes one by one with the least connection errors in a 3-D space. This requests the system strict camera calibration. Simultaneous entire shape recovery without the registration would be the best for reducing computational load.

One also notices by the above explanation that it is not very simple to apply the technique to various kinds of 3-D object modeling. The image taking cameras need be calibrated in advance. How can it then be applied to the 3-D modeling of a street performer who won't wait for the advance camera calibration? How can it be applied to 3-D modeling of a statue around which many people gather in a marketplace? Thus one notices the advantage of a 3-D modeling technique based on image-based camera calibration.

In the present paper, a novel 3-D modeling technique is presented which performs simultaneous entire shape recovery based on image-based camera calibration. A similar technique was invented by some of the authors [5]. It gave restriction on the camera arrangement, however, that a facing cameras need colinearity with their principal axes. The restriction imposed on the proposed technique is that at least three points need be shared among observing cameras around the object concerned. These points in captured images are employed for camera calibration. This simpler constraint can be realized often easily when every camera is set so that it looks down the object. Thus it results in more flexible arrangement of the surrounding cameras compared to [5].

The idea of simultaneous recovery is explained in the following. The proposed technique applies the factorization to the 3-D shape recovery. The factorization [3] claims that, once a point, actually its 2-D image coordinates, is included in a measurement matrix, it recovers the 3-D location by factorizing the measurement matrix. Therefore the locations of rear points on an object are transformed to the locations where the frontal cameras would observe them if the object is transparent, and the transformed points are included in the measurement matrix defined by the frontal points, which yields an entire measurement matrix. By factorizing this matrix, all the specified points that spread over the object recover their 3-D locations and hence the entire 3-D modeling is done. Since the above procedure can be described that frontal cameras see through rear shape of an object as well as its frontal part, they are referred to *virtual see-through cameras*. Formulation of the present technique and experimental results are given in the following with discussion and conclusions.

II. PROCEDURE

The cameras surrounding an object are categorized into frontal cameras C^F and rear cameras C^R. Since the present technique recovers point location, the points for the recovery are specified on an object. They are classified into three sets. The point commonly observable from the both cameras C^F and C^R is denoted by P^{FR} which is an element of a set $G(P^{FR})$: The point observable only from the frontal cameras is denoted by P^{F-} which is an element of a set $G(P^{F-})$: The point observable only from the rear cameras is denoted by P^{-R} which is an element of a set $G(P^{-R})$. Then the set of points which the frontal cameras C^F observe is given by $G(P^F)= G(P^{FR}) \cup G(P^{F-})$, whereas the set of points the rear cameras C^R observe is given by $G(P^R)= G(P^{FR}) \cup G(P^{-R})$. See **Fig. 1** for example.

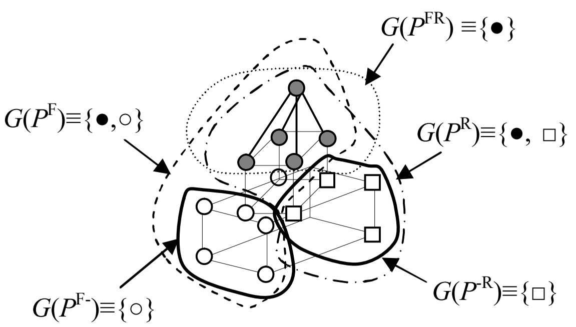

Fig. 1. Categorized points on an object. The points commonly observable from frontal and rear cameras, observable only from frontal cameras, and observable only from rear cameras are denoted by ●, ○, and □, respectively.

The application of the factorization with respect to the frontal cameras and to the points in the set $G(P^F)$ provides the following equation;

$$W^F=M^F S^F \tag{1}$$

Here W^F is a measurement matrix defined by the *xy*-coordinates of the points in the set $G(P^F)$, S^F is a shape matrix containing their 3-D coordinates, and M^F is an orientation matrix giving the orientations of the frontal cameras.

Matrix S^F in Eq. (1) can be decomposed into

$$S^F=\{S^{FR}|S^{F-}\} \tag{2}$$

Here S^{FR} is the shape matrix giving the 3-D coordinates of the points in the set $G(P^{FR})$ and S^{F-} is that giving the 3-D coordinates of the points in the set $G(P^{F-})$.

Employing the matrix S^{FR}, the orientation matrix of rear cameras denoted by M^R satisfies the following relation;

$$W^{FR}=M^R S^{FR} \tag{3}$$

where W^{FR} is the measurement matrix with respect to the points in $G(P^{FR})$. Since W^{FR} is known in an experiment and S^{FR} is given by Eq. (2), we have

$$M^R= W^{FR}[S^{FR}]^T(S^{FR}[S^{FR}]^T)^{-1} \tag{4}$$

The points in the set $G(P^{F-})$ are then projected virtually onto the image planes of the rear cameras by

$$V^{F-}=M^R S^{F-} \tag{5}$$

By this projection, the rear cameras have the measurement matrix W^R that contains all the projected points;

$$W^R=\{W^{FR}|V^{F-}|W^{-R}\} \tag{6}$$

On the other hand, the following holds with the set $G(P^{-R})$;

$$W^{-R}=M^R S^{-R} \tag{7}$$

Since W^{-R} is known and M^R is given by Eq. (4), we have

$$S^{-R}=([M^R]^T M^R)^{-1} [M^R]^T W^{-R} \tag{8}$$

The shape matrix S^{-R} gives the 3-D coordinates of the points in the set $G(P^{-R})$. They are projected virtually onto the image planes of the frontal cameras by

$$V^{-R}=M^F S^{-R} \tag{9}$$

Then we have the measurement matrix of the form

$$W^F=\{W^{FR}| W^{F-}|V^{-R}\} \tag{10}$$

Finally, from Eqs. (6) and (10), we have an overall measurement matrix W of the form (See also **Fig. 2**.)

$$W = \begin{pmatrix} W^F \\ \cdots\cdots \\ W^R \end{pmatrix} \tag{11}$$

In this way, all the selected points on object O have been projected onto the image planes of all the

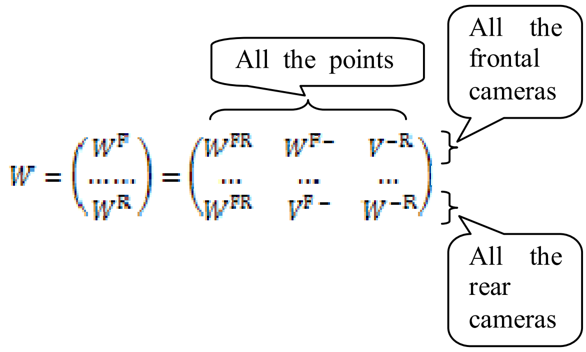

Fig. 2. Matrix *W* containing all the point information with all the surrounding cameras.

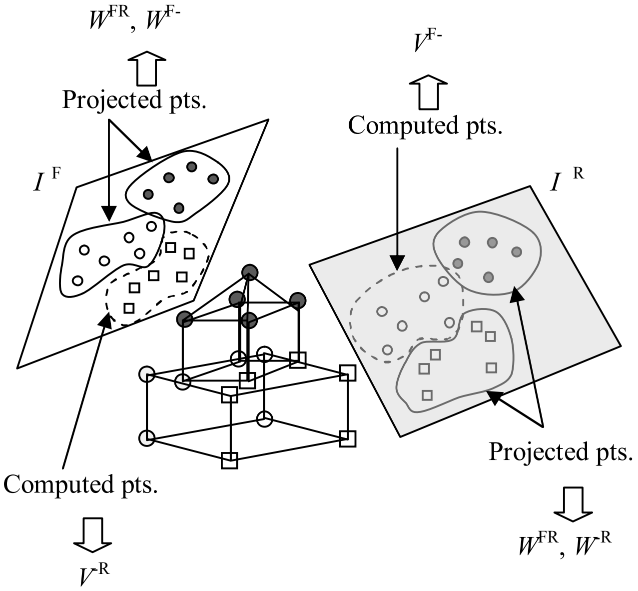

Fig. 3. Virtual see-through camera. Observable points in a scene are projected directly onto the image plane of a camera, whereas unobservable points are computed their projected locations in the image plane and included in it. The points □ cannot be seen from frontal cameras and their locations on the frontal image planes I^F are computed. In the same way, the points ○ are not seen from rear cameras and their locations on the rear image planes I^R are computed.

surrounding cameras. In another expression, all the points on object *O* have been observed by all the virtual see-through cameras. See **Fig. 3** for the idea.

The matrix *W* of Eq. (11) is factorized as

$$W=MS \tag{12}$$

providing the final orientation matrix *M* of all the cameras and the shape matrix *S* of all the points on the object. Matrix *S* gives the 3-D point model of object *O*.

It is noted that Eq. (11) represents the geometrical constraints of all the points on object *O* in the orientation of all the cameras. From the factorization shown by Eq. (12), a 3-D model that satisfies the constraints best in the sense of the LMS errors is obtained.

III. EXPERIMENTAL RESULTS

An experiment was conducted to examine the performance of the proposed technique. Instead of employing multiple cameras, as shown in **Fig. 4**, a single digital video camera and a turntable was used for taking images of a polyhedron from multiple orientations. The turntable was given turns manually. The angles of the turns were not very strict as the values are not employed in the recovery calculation. The employed factorization assumed weak perspective projection with the camera imaging. This linear projection well approximates the imaging by a lens when the thickness of the object concerned is 1/10 or less of the distance from the camera to the object [6].

Figure 5 shows the result of the entire 3-D recovery of a polyhedron. The polyhedral images were taken from 4 to 6 orientations. The recovery errors were calculated with respect to the side lengths and the angles the adjacent sides make at the vertices. The result is shown in **Table 1**. The average recovery error is 1.66% with the side lengths, where it is 2.27% with the angles. The recovery errors decrease monotonically according to the increase of camera observation orientations. As the geometrical constraints increase, more exact recovery is realized.

Fig. 4. Experimental setup: A digital video camera and a turntable on which an object is placed. The turntable is given turns manually for multiple views by the camera.

Table 1. Relative recovery errors of a polyhedron with respect to side lengths and angles at vertices.

No. of camera orientations	Side length	Angle
4	1.85	2.57
5	1.79	2.42
6	1.68	2.14
7	1.54	2.12
8	1.44	2.09

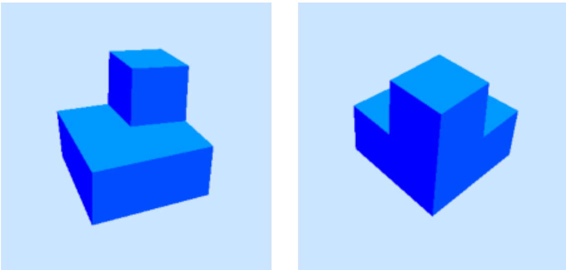

Fig. 5. Recovered polyhedron rendered from its 3-D point model: Frontal view (the left) and rear view (the right).

V. CONCLUSIONS

A technique was presented for recovering entire shape of an object by use of multiple surrounding cameras. The idea was to merge the rear shape information with the frontal shape information, and vice versa. The procedure was described as the information acquisition of the opposite side of an object by a virtual see-through camera. The shape recovery achieved satisfactory performance by taking geometrical constraints in multiple orientations into account simultaneously. This is the most important issue in the technique. The simultaneous consideration of the multiple geometrical constraints may result in less distortion with the recovered shape.

The presented technique can also be applied to non-rigid object recovery [7,8] by arranging multiple cameras around it, instead of a single camera and a turntable. The camera arrangement is more flexible in the technique compared to [8], and it may even be employed outdoors, since camera calibration is done by use of acquired images, i.e., image-based camera calibration.

The presented shape recovery is the recovery of the 3-D location of a point in a 3-D space and it provides a 3-D model of an object in the form of a point set. Hence graphical operation follows the present technique to make a solid model. This is a little disadvantage compared to a back projection technique [9], where a volumetric 3-D model is directly obtained without choosing points on an object, although concave parts cannot recover by the technique.

Those points specified on an object were categorized into three classes, i.e., the class of a point observable commonly from frontal as well as rear cameras, the class of a point observable only from frontal cameras, and the class of a point observable only from rear cameras. This categorization was done manually. It isn't necessarily be automated, since the present subject is 3-D modeling and not automatic recognition. Machine categorization and human examination may, however, realize a time saving and efficient modeling system.

The authors have been developing a man-machine interface system employing a camera mounted on the display of a computer. The system watches a user in front of a computer by the camera and recognizes his/her hand gestures by which a 3-D object in the display moves and rotates in a 3-D way. The present technique is going to be employed for providing the system with 3-D models of objects.

REFERENCES

[1] Gonzalez, R. C., Wintz, P.: *Digital Image Processing*, Addison Wesley, 1987.

[2] Abdel-Aziz, Y. I., Karara, H. M.: "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry", *Proc. Symp. Close-Range Photogrammetry*, pp.1-18, 1971.

[3] Tomasi, C., Kanade, T.: "Shape from motion under orthography: A factorization method", *Int. J. Comput. Vision*, Vol.9, Vo.2, pp.137-154. 1992.

[4] Cheng, Y., Medioni, G.: "Object modeling by registration of multiple range images", *Image and Vision Computing*, vol.10, no.3, pp.145-155, 1992.

[5] Tan, J. K., Ishikawa, S.: "Recovering entire shape of a deformable object employing a single measurement matrix", *Proc.1999 IEEE Int. Conf. on Image Processing*, CD-ROM, 27PP5A.4, 1999.

[6] Sato, J.: Computer Vision – Geometry of Vision, Corona Publishing, 1999. (in Japanese)

[7] Tan J. K., Ishikawa S.: "Deformable shape recovery by factorization based on a spatiotemporal measurement matrix", *Computer Vision and Image Understanding*, Vol. 82, No. 2, pp.101-109, 2001.

[8] Tan, J. K., Ishikawa, S., Hirokawa, S.: "A fully 3-D modeling method of non-rigid objects from multiple views", *J. Inst. Image Electron. Eng. Japan*, Vol.32, No.4, pp.483-487, 2003. (in Japanese)

[9] Uchinoumi M., Tan J. K., Ishikawa S.: "A simple structured real-time motion capture system employing silhouette images", *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics*, pp.3094-3098 (Oct., 2004).