Factorization Method Using Interpolated Feature Tracking via Projective Geometry

Hideo Saito [†,††]   Shigeharu Kamijima [†]
†Department of Information and Computer Science, Keio University
Yokohama-City, 223-8522, Japan
††PRESTO, Japan Science and Technology Corporation(JST)
saito@ozawa.ics.keio.ac.jp

Abstract

The factorization method by Tomasi and Kanade simultaneously recovers camera motion and object shape from an image sequence. This method is robust because the solution is linear by assuming the orthographic camera model. However, the only feature points that are tracked throughout the image sequence can be reconstructed, it is difficult to recover whole object shape by the factorization method. In this paper, we propose a new method to interpolate feature tracking so that even the loci of unseen feature points can be used as inputs of the factorization for object shape reconstruction. In this method, we employ projective reconstruction to interpolate untracked feature points. All loci of all detected feature points throuout the input image sequence provide correct reconstructed shape of the object via the factorization. The results of reconstruction are evaluated by the experiment using synthetic images and real images.

1 Introduction

Modeling of real objects play an important role for realizing applications of virtual reality. Shape recovery from an image sequence taken with unknown motion camera is one of methods for modeling of real objects. Tomasi and Kanade proposed a method for modeling from a handy video image sequence, which is called as the factorization method [10]. In the factorization method, they assume the orthographic projection so that the loci of feature points in the image sequence can robustly be separated into camera motion and shape of the object by a linear solution. However, the factorization method has the following problems for applying it into real application.

The first problem is that the error in the recovered shape and camera motion is caused by the approximation error to orthographic projection. Therefore, various improved factorization method by avoiding the orthographic projection have been proposed, such as using the para-perspective projection [6], or the perspective projection [1, 3]. However, those are still suffered by the approximation of the camera model.

The second problem is that the 3D position of a feature point can be recovered only if the locus of the feature point is tracked through all the input image sequence. In a real image sequence, it is difficult to track sufficient number of feature points to recover the object shape, because of the change of illumination conditions, occlusion of the feature

points, out of FOV of the feature points, etc. One solution for such problem is divide an input image sequence into some sub-sequences, so that more feature points can be tracked throughout the sub-sequences. Then the 3D positions recovered in the sub-sequences are integrated into one shape model, but the 3D recovery errors in the sub-sequences are accumulated. Shum et.al.[9] proposed to apply PCA for interpolating missing feature tracking locus. Although the PCA provides the most possible interpolation, the error caused by the missing data cannot removed completely.

Alternative solution for the second problem is to estimate the projected position of the feature point that cannot be tracked in the input image sequence. One approach is already mentioned by Tomasi and Kanade [10]. In this approach, camera positions in the image sequence are first estimated by applying the factorization method to the tracked feature points throughout the sequence, even though the number of the tracked points is small. Then the positions of other untracked feature points are estimated by using the camera position estimated by the first factorization. However, the estimated positions of the untracked feature points are affected by the approximation error of the orthographic projection.

In this paper, we propose a novel approach to solve the second problem of the factorization via projective reconstruction. In this method, the projected positions of untracked feature points are estimated by projective reconstruction, so that all feature points can be tracked in all frames in the input image sequence. Then the 3D positions of all feature points can be recovered from the complete loci of them by using the factorization method. Since the projective reconstruction employed in this method provides accurate projected point under the perspective projection camera model, the interpolated loci by the proposed method is not affected by the camera model approximation error. The experimental results for demonstrating the efficacy of the proposed method are shown in the section 5.

2   Factorization by Tomasi and Kanade

The factorization proposed by Tomasi and Kanade is a method for simultaneous recovery of both camera motion and object shape from the loci of feature points tracked in an image sequence taken with a moving camera.

In the 3D coordinate with the origin at the center of the set of feature points, the pose of the camera of f-th frame and the position of p-th feature point are represented as ($\mathbf{i}_f$ $\mathbf{j}_f$ $\mathbf{k}_f$), and $\mathbf{s}_p$, respectively. The projected position of the p-th feature point on the f-th frame is then indicated by (u_{fp}, v_{fp}). A measurement matrix $\mathbf{W}$ is defined from the loci of P feature points through F frames as the following Eq.(1).

$$
W = \begin{bmatrix}
u_{11} & \cdots & u_{1P} \\
\vdots & u_{fp} & \vdots \\
u_{F1} & \cdots & u_{FP} \\
v_{11} & \cdots & v_{1P} \\
\vdots & v_{fp} & \vdots \\
v_{F1} & \cdots & v_{FP}
\end{bmatrix}
\tag{1}
$$

Under the assumption of the orthographic projection of the camera, the modified measurement matrix $\mathbf{W}^*$, which is derived by subtracting an average of each row from every

value in the measurement matrix $\mathbf{W}$, can be represented by the inner product between camera pose matrix $\mathbf{M}$ and object shape matrix $\mathbf{S}$ as shown in Eq.(2).

$$\mathbf{W}^*_{(2F \times P)} = \mathbf{M}_{(2F \times 3)} \mathbf{S}_{(3 \times 2F)} \tag{2}$$

Accordingly, factorizing $\mathbf{W}^*$ into $\mathbf{M}$ and $\mathbf{S}$ provides the camera motion at every frame and the 3D position of every feature position. Since SVD is employed to factorizing $\mathbf{W}^*$, the recovery of the camera motion and the shape can robustly be computed.

3 Projective Reconstruction

A 3D projective transform can be represented by the following equation with the homogenous coordinate.

$$\lambda \tilde{\mathbf{X}}' = \mathbf{H}_p \tilde{\mathbf{X}} \tag{3}$$

where $\mathbf{H}_p$ indicates an arbitrary 4×4 matrix. Projective reconstruction [7, 13] recovers the shape of the object with ambiguity of arbitrary projective transforms. An object shape taken with two cameras can be recovered in terms of projective reconstruction, if the fundamental matrix between the two cameras is known. In our paper, we employ a projective reconstruction method based on SVD[8], of which the computation is relatively robust, for estimating projected position of untracked feature points.

For this projective reconstruction, fundamental matrix $\mathbf{F}_{12}$ between two base frames f_1 and f_2, which are selected from an input image sequence, is computed by corresponding 2D positions between two base frames that can be obtained from some of tracked feature points. The fundamental matrix gives the epipole $\mathbf{e}'$ of the second base frame f_2 as shown in the following equation.

$$\mathbf{F}_{12}^T \mathbf{e}' = 0 \tag{4}$$

Then, the projective space spanned with two base frame images is projected onto each base frame image by the following projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$.

$$\begin{aligned} \mathbf{P}_1 &= [\mathbf{I}0] \\ \mathbf{P}_2 &= [\mathbf{M}\mathbf{e}'] \end{aligned} \tag{5}$$

where

$$\mathbf{M} = -\frac{1}{\|\mathbf{e}'\|^2} [\mathbf{e}']_\times \mathbf{F}_{12}. \tag{6}$$

Therefore, the 3D position of a point in the projective space, $\mathbf{X} = [x, y, z, t]^T$, can be related to the projected 2D positions of the point on two base frame images, $\mathbf{m} = [u, v]^T$ and $\mathbf{m}' = [u', v']^T$, as the following equation.

$$\mathbf{A}\mathbf{X} = \mathbf{0}. \tag{7}$$

Here, let $\mathbf{P}^i$ represents i-th column vector of the projection matrix $\mathbf{P}$, and then the following equation is derived.

$$\mathbf{A} = \left[\mathbf{P}_1^1 - u\mathbf{P}_1^3, \mathbf{P}_1^2 - v\mathbf{P}_1^3, \mathbf{P}_2^1 - u'\mathbf{P}_2^3, \mathbf{P}_2^2 - v'\mathbf{P}_2^3 \right]^T \tag{8}$$

Accordingly, we can compute the position of the projective reconstruction x, y, z, t by applying SVD to $\mathbf{A}$ for computing the eigen vector corresponding to the minimum SVD value. The eigen vector is $\mathbf{X} = [x, y, z, t]^T$.

4 Proposed Method

The proposed method takes an input image sequence of an object that is captured by a handy video camera. The image sequence is manually divided into multiple sub-sequences with overlapped some frames. In the first frame of each sub-sequence, feature points are extracted and tracked though the sub-sequence by Kanade-Lucas-Tomasi Feature Tracker [5, 11]. Some of the extracted feature points cannot be tracked until the end frame of the sub-sequence because of illumination change, occlusion, out of the FOV, etc. The positions of untracked feature points for such cases are estimated by "intra sub-sequence interpolation". Then the feature loci obtained in each sub-sequences are integrated according to the feature points in the overlapped frames. The loci in un-overlapped frames are estimated by "inter sub-sequence interpolation". Finally, locus of every feature point is obtained throughout the input image sequence, so that complete measurement matrix can be obtained. The factorization method is once applied to the complete measurement matrix, and then the whole 3D shape can be reconstructed.

4.1 Feature Tracking for Sub-sequence

A handy video camera captures an object with the camera moving around the object in about $360°$, and an image sequence consist of F frames is obtained. The image sequence is divided into some sub-sequences with overlapped frames. The division into sub-sequences is performed by manual operation.

Feature points are detected in the first frame in every sub-sequence, and tracked in the sub-sequence. The tracking of some features are failed, and then the loci of such feature points are broken.

4.2 Intra Sub-sequence Interpolation

The positions of feature points that cannot be tracked are estimated by projective reconstruction of the untracked feature points.

Suppose that feature point represented by $\mathbf{s}_l$ is not tracked in f_l-th frame. Two base frames f_1 and f_2 for projective reconstruction of the point $\mathbf{s}_l$ are selected from the frames in which the point $\mathbf{s}_l$ can be tracked as shown in Fig. 1(a). In our experiment, we select the first frame and the 30th frame in each sub-sequence as a default selection.

In the projective space spanned by the base frames, 3D positions of all feature points that appear in both the base frames are projectively reconstructed based on the method described in Sec.3 (Fig.1(b)) In the projectively reconstructed points, some points are also tracked in f_l-th frame. Using such points, the projection matrix $\mathbf{P}_l$, which projects a point in the projective space onto l-th frame f_l, is computed according to the following equation (Fig.1(c)).

$$\lambda \mathbf{m}_i = \mathbf{P}_l \mathbf{s}_i \tag{9}$$

where $\mathbf{s}_i$ represents the projectively reconstructed points which are also tracked in f_l-th frame at position $\mathbf{m}_i$. λ indicates an arbitrary scaling factor. The computed projection matrix $\mathbf{P}_l$ provides the position of the untracked point $\mathbf{s}_l$ in f_l-th frame as shown in Fig.1(d)

According to this procedure, all untracked points can be interpolated in each sub-sequence.

(a)selection of base images

(b)projective reconstruction

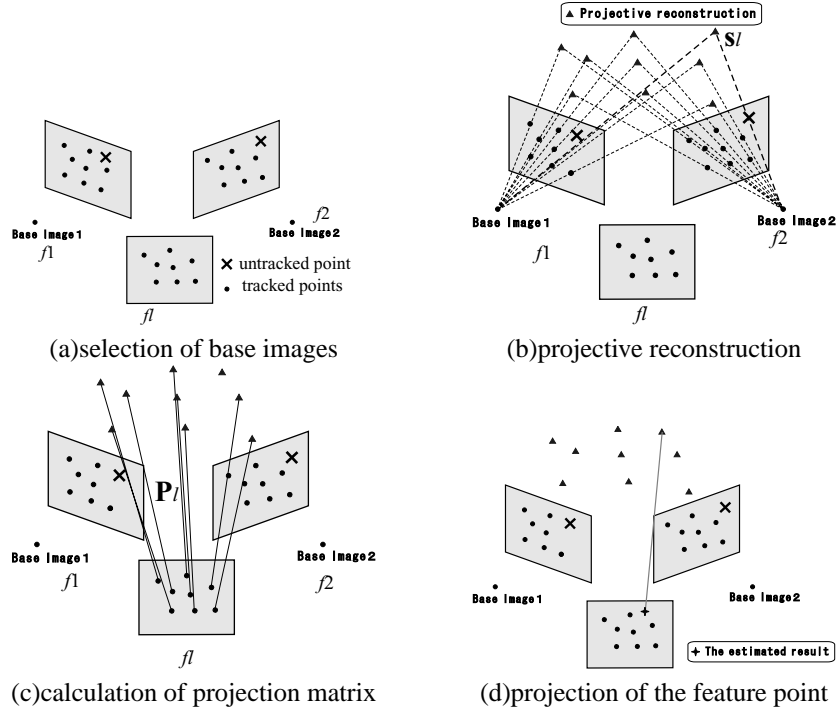(c)calculation of projection matrix

(d)projection of the feature point

Figure 1: Interpolation of projected position of untracked feature point.

4.3 Inter Sub-sequence Interpolation

Intra sub-sequence interpolation can generate a measurement matrix that includes all loci of feature points detected in the first frame of a sub-sequence. Inter sub-sequence interpolation integrates the measurement matrix of each sub-sequence into a measurement matrix that includes all loci of feature points detected in all of the sub-sequences.

Neighboring two sub-sequences share some overlapped frames. For integrating both sequences, two base frame images are selected from the overlapped frames. Then the feature points included in both sequences are projectively reconstructed in the projective space spanned by the base frame images. Since the projective space is shared by two sub-sequences, feature points in one sub-sequence can be projected onto frames in another sub-sequence. Accordingly, all the loci of feature points in both the sub-sequences can be estimated for all frames in both sub-sequences. Such procedure is also applied to all pairs of neighboring two sub-sequences, and then the complete measurement matrix that include all loci of the feature points detected in all sub-sequences.

4.4 Factorization and Shape Modeling

By applying the factorization to the complete measurement matrix, 3D positions of all feature points and camera motion of all frames in the sequence can be recovered. Delaunay triangulation generates a 3D polygon mesh model from the 3D feature points. Texture
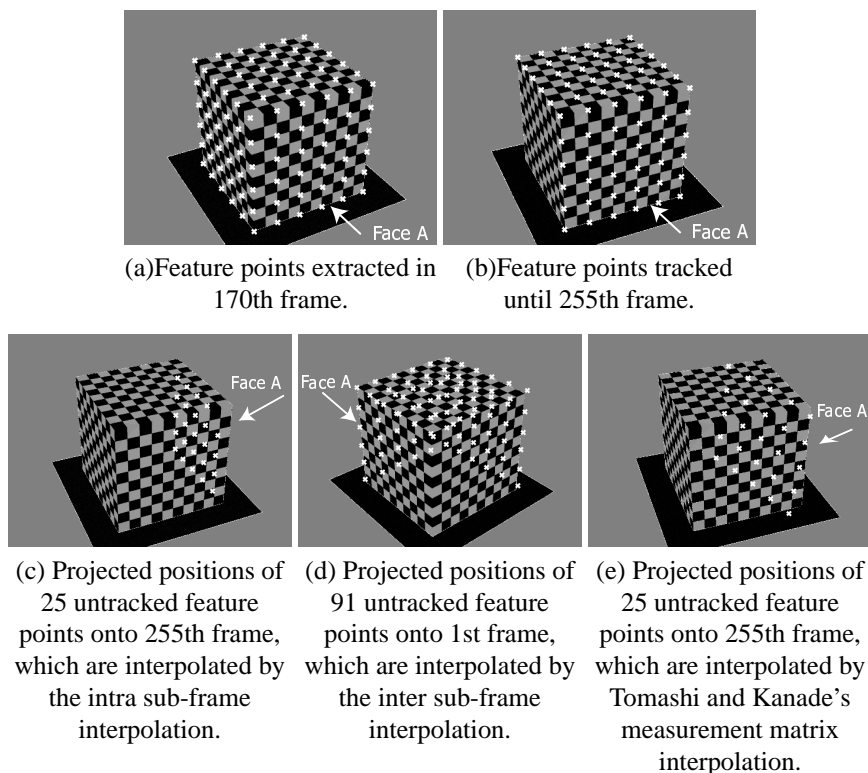
(a)Feature points extracted in 170th frame.

(b)Feature points tracked until 255th frame.

(c) Projected positions of 25 untracked feature points onto 255th frame, which are interpolated by the intra sub-frame interpolation.

(d) Projected positions of 91 untracked feature points onto 1st frame, which are interpolated by the inter sub-frame interpolation.

(e) Projected positions of 25 untracked feature points onto 255th frame, which are interpolated by Tomashi and Kanade's measurement matrix interpolation.

Figure 2: Projected position of feature points in the experiment whose input is a synthetic image sequence.

on every polygon in the model is rendered from the images of the sequence. Such texture mapping provides realistic 3D model that can be shown in a virtual space.

5 Experiments and Discussions

For demonstrating the efficacy of the proposed method, a computer-synthesized image sequence and a real image sequence are prepared as inputs of the proposed method.

5.1 Experiments with Synthesized Image Sequence

An image sequence of a virtual object is synthesized by computer graphics software "Pov-Ray" for demonstrating the accuracy in the interpolation. The image sequence consists of 360 frames with 480×360 pixels. The sequence is divided into 4 sub-sequences with 30 frames overlapping between neighboring sub-sequences. Feature points are the corner points of the checker pattern. In each sub-sequence, the first frame captures three planes of the object, and the last frame only captures two planes. This means that about third of the feature points detected in the first frame can not be tracked until the last frame.

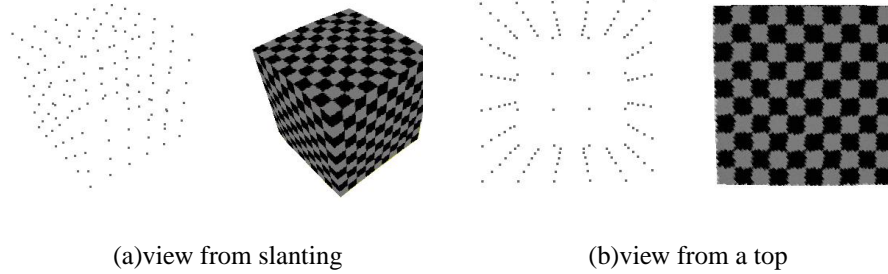(a)view from slanting   (b)view from a top

Figure 3: 3D feature points and 3D textured model reconstructed by the proposed method in the experiment whose input is a synthetic image sequence

Fig.2(a) shows 91 feature points extracted in the 170th frame that is the first frame in the sub-sequence #3. Until the 255th frame that is the last frame of the sub-sequence #3, 66 feature points shown in Fig.2(b) are tracked. Using the intra sub-frame interpolation, the positions of the untracked 25 points in the 255th frame are estimated as shown in Fig.2(c). Those 25 points are the corner points of the checker pattern on the unseen plane (face A) of the object in the 255th frame.

Using the inter sub-frame interpolation, the positions of 91 points extracted in the first frame of the sub-sequence #3 shown in Fig.2(a) are also estimated in the first frame of the sub-sequence #1 as shown in Fig.2(d). Although those 91 points are not seen in the 1st frame, inter sub-frame interpolation can estimate their positions.

Under the same condition as the intra sub-frame interpolation shown in Fig.2(c), the positions of the untracked points in the 255th frame are interpolated by using the inter-polation method by Tomashi and Kanade, which are shown in Fig. 2(e). In Fig. 2(e), the interpolated positions are not fit on the surface of the unseen plane (face A). This is because that the proposed method estimates the position of the untracked points by the projective reconstruction, while Tomashi and Kanade method estimates based on shape reconstruction under the orthographic approximation with the factorization.

Since we know the accurate projected positions of all the feature points because of the computer-synthesized image sequence, the interpolation error can be evaluated. The average errors of the intra sub-sequence interpolation (Fig.2(c)) and the inter sub-sequence interpolation (Fig.2(d)) are 2.84 pixels and 2.89 pixels, respectively. Considering that the error of feature point extraction and tracking is about 1.2 pixels, the error caused by the interpolation using the proposed method is relatively small.

In Fig.3 shows the recovered 3D feature points and 3D polygon mesh model with texture. The fact that the checker pattern can correctly be recovered on the reconstructed 3D model demonstrates the accuracy of the proposed method.

Fig.4 shows the 3D positions of the feature points and the pose and the direction of camera which are recovered by (a) the measurement matrix interpolation Tomashi and Kanade method, (b) simple integration of partial shapes which are reconstructed by applying the factorization to each sub-sequence independently, and (c) the proposed method. The 3D recovery errors in shape reconstruction for (a), (b), and (c) are 0.172, 0.052, and 0.016, respectively. The error value represents the averaged distance with the accurate

| | (a) Tomashi and Kanade interpolating method. | (b) Simple integration method. | (c) Proposed method. |
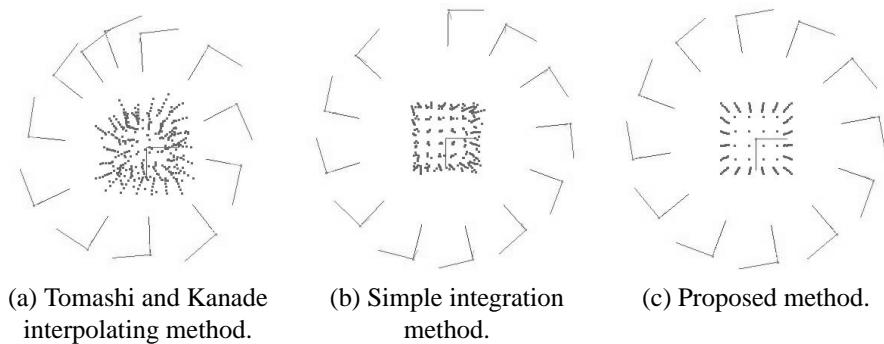
	(a)	(b)	(c)
3D recovery error	0.172	0.052	0.016

Figure 4: Comparison with the conventional method

shape. One unit of the length is defined by one edge of the object cube shape.

As shown in Fig.4(a), the interpolation method by Tomashi and Kanade includes larger error, which is caused by approximation of camera model. Their interpolation is based on the 3D position estimated by the factorization, in which the camera model is approximated to orthographic. This approximation causes error in interpolated position.

The simple integration of partial shapes accumulates the shape reconstruction error of each factorization. The factorization only minimumizes the error in each sub-sequence, but does not optimize for whole the sequence. The recovered camera motion also indicates the error accumulating behavior.

On the other hand, the advantage of the factorization, such that the estimation error in shape and camera motion can be minimumized by using the SVD, affects to whole the input image sequence in the proposed method. This is because that the projective reconstruction interpolates all untracked feature points in all frames in the sequence, so that the factorization can be applied to whole the sequence at once.

5.2 Experiments with Real Image Sequence

A handy camera captures the car miniature as shown in Fig.5 as an image sequence consists of 256 frame images with 480×360 pixel resolution. Using the proposed method, 3D positions of feature points on the object can be recovered as shown in Fig.6. From the feature points, 3D polygon mesh model with the texture can be recovered as shown in Fig.7.

6 Conclusion

In this paper, we propose a method for generate a measurement matrix that include loci of all extracted feature point throughout an input image sequence by intra and inter interpolation of sub-sequences based on the projective reconstruction. By applying the

factorization to the complete measurement matrix, shape of whole object and camera motion for all frames in the sequence can correctly be recovered. The experimental results demonstrate the accuracy in the shape and camera motion reconstruction of the proposed method.

The accuracy performance depends on the scheme for dividing input sequence into sub-sequences and the selection of base frames for the projective reconstruction. Therefore, optimizing the input sequence division and base frame selection will be one of future research topics.

References

[1] S. Christy and R. Horaud, "Euclidean reconstruction: from paraperspective to perspective," Proc.4th European Conference on Computer Vision,vol.2,pp.129–140,Apr,1996.

[2] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT PRess, 1993.

[3] J. Fujiki, T.Kurata, "Recursive factorization method for the paraperspective model based on the perspective projection," Proc. ICPR 2000, vol.1, pp.406-410, Aug.2000.

[4] M. Han and T. Kanade,"Perspective Factorization Methods for Euclidean Reconstruction," Carnegie Mellon University Technical Report, CMU-RI-TR-99-22, Aug.1999.

[5] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," International Joint Conference on Artificial Intelligence, pages 674-679, 1981.

[6] C.J. Poleman and T. Kanade,"A paraperspective factorization method for shape and motion recovery," IEEE Trans. Pattern Analysis and Machine Intelligence,vol.19,no.3,pp.206–218,1997.

[7] C. Rothwell and G.C surka and O. Faugeras,"A Comparison of Projective Reconstruction Methods for Pairs of Views," INRIA Research Report,no.2538,Apr,1995.

[8] Y. Seo and K.S. Hong,"Calibration-free augmented reality in perspective,"IEEE Trans. Visualization and Computer Graphics,vol.6,no.4,pp.346–359,Oct,2000.

[9] H. Shum and K. Ikeuchi and R. Reddy,"Principal component analysis with missing data and its application to polyhedral object modeling,"IEEE Trans. Pattern Analysis and Machine Intelligence,vol.17,no.9,pp.854–867,Sep,1995.

[10] C. Tomashi and T. Kanade,"Shape and motion from image streams under orthography: A factorization method," International Journal of Computer Vision,vol.9,no.2,pp.137–154,1992.

[11] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.

[12] G. Xu and Z. Zhang: *Epipolar Geometry in Stereo, Motion and Object Recognition*, Kluwer Academic Publishers, 1996.

[13] Z. Zhang,"Determining the epippolar geometry and its uncertainly: A review," INRIA Research Report,no.2927,July,1996.
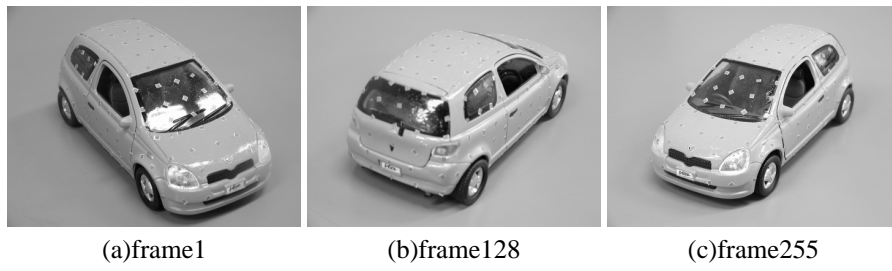
(a)frame1   (b)frame128   (c)frame255

Figure 5: The real image sequence of the experiment whose input is a real image sequence
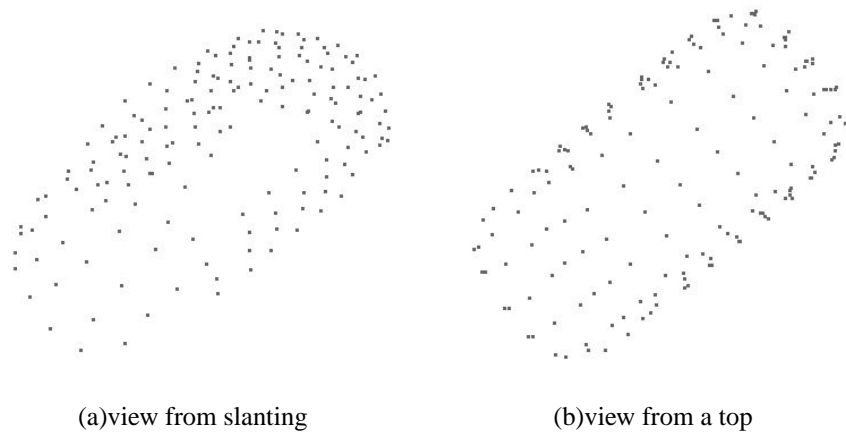

(a)view from slanting   (b)view from a top

Figure 6: 3D feature points model reconstructed by the proposed method in the experiment whose input is a real image sequence.

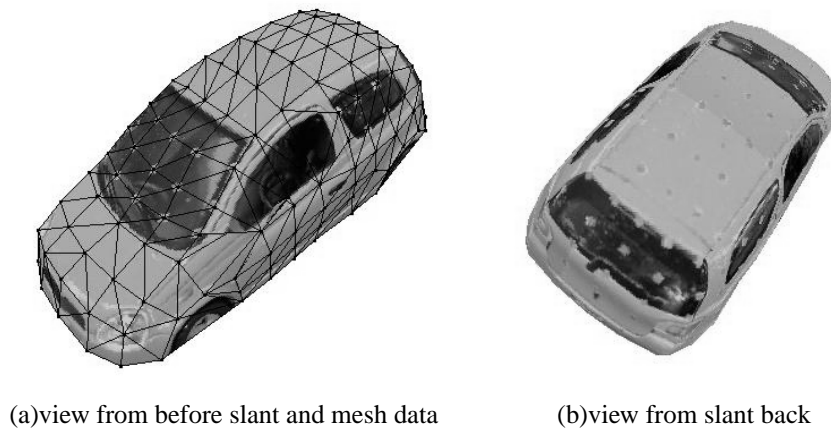(a)view from before slant and mesh data (b)view from slant back

Figure 7: 3D model reconstructed by the proposed method in the experiment whose input is a real image sequence