13—9 Reconstruction of Measurement Matrices for Recovering Shape and Motion from Long Image Sequences

Ying-jieh Huang*
Information and Communication R&D Center
Ricoh Co., Ltd.

Hirobumi Nishida†
Software Research Center
Ricoh Co., Ltd.

Abstract

The factorization method for recovering 3D information from image sequences assumes that all feature points selected on the first frame can be tracked throughout. This assumption is violated, however, if some features disappear or new ones are introduced later. In this paper, based on the factorization method with the paraperspective projection model, we present a method for estimating image coordinates of occluded feature points in order for the estimated locations to approximate closely their perspective projections. The estimation accuracy is evaluated with synthetic data and some results are presented for 3D information extraction from real image sequences.

1 Introduction

Many handy devices for acquiring digital images and videos such as digital still/video cameras have recently appeared in markets. These digital media have significant advantages of being easily edited, modified, and manipulated, over traditional analog media. Therefore, technical needs have been increasing for the reuse of digital media and the extraction of useful information from them. In particular, the recovery of 3D object shapes from an image sequence has been an important research subject in computer vision, as well as in such application areas as robot vision, autonomous vehicles, 3D shape input through video cameras, model-based image/video coding, and 3D modeling.

Tomasi and Kanade proposed a *factorization method* [4] for the robust and efficient estimation of shape and motion from image sequences. Based on the orthographic projection model, this method formalizes the problem as solving a set of linear equations in terms of the shape and motion parameters. The solution obtained by this method is quite stable and accurate compared with other methods. Furthermore, Poelman and Kanade [3] developed a factorization method based on the paraperspective projection model, which approximates the perspective projection more closely, keeping the formalization linear in terms of the parameters for estimation.

The factorization method assumes that all feature points selected on the first frame of the image sequence can be tracked throughout. This assumption is violated if some features disappear or new ones are introduced later. Such situations can often happen in a long image sequence taken from the camera moving around the object. Tomasi and Kanade [4] cope with the problem of feature occlusion by reconstructing the projection of the feature point onto the image plane, as if the object were transparent. This approach is based on a partial estimation of shape and motion with the factorization method applied to a subset of frames and features. The projection of the occluded feature point is estimated from the partial estimation by the least-square criteria. The drawback of this approach is that the estimation is computed in terms of the projection model employed by the factorization method, i.e., the orthographic or paraperspective model. Therefore, the estimated location can be biased considerably from the perspective projection, which is the real camera model, and therefore, the accuracy of recovery of the whole object shape and camera motion can be degraded if such inaccurate local estimations are incorporated into the recovery process.

In this paper, based on the factorization method with the paraperspective projection model [3], we present a method for reconstructing the projection of occluded feature points onto the image plane in order for the estimated locations to approximate closely their perspective projections. The proposed method virtually enables the accurate tracking of locations of all the feature points throughout the image sequence, even when a large motion is allowed for the camera, some feature points disappear due to occlusion, and new feature points are introduced. Consequently, based on the feature correspondences among image frames, the object shape and the camera motion can be recovered accurately from a long image sequence.

This paper is organized as follows: In Section 2, the paraperspective projection model is outlined along with the factorization method. In Section 3, we develop algorithms for estimating image coordinates of occluded feature points. In Section 4, some experimental results are presented for recovering shape and motion from video image sequences and sets of

*Address: 3-2-3 Shin-Yokohoma, Kohoku-ku, Yokohama, 222-8530, Japan. E-mail: huang@ic.rdc.ricoh.co.jp

† Addess: 1-1-17 Koishikawa, Bunkyo-ku, Tokyo 112-0002, Japan. E-mail: hn@src.ricoh.co.jp

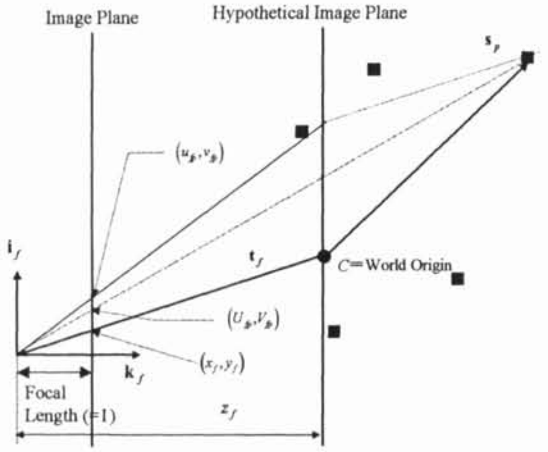

Fig. 1 Paraperspective projection model.

still images. Section 5 is the conclusion.

2 Paraperspective Projection Model and Factorization Method

We outline the parapersepective projection model [2], which is a linear projection model taking into account the scaling effect and the position effect. As shown in Fig. 1, the world origin is placed at the object's center of mass C. Let $\mathbf{s}_p$ be the 3D coordinates of the feature point p, $\mathbf{t}_f$ be the location of the camera's focal point in the frame f, $\mathbf{i}_f, \mathbf{j}_f \in R^3$ be the orthonormal vectors spanning the image plane f, and $\mathbf{k}_f = \mathbf{i}_f \times \mathbf{j}_f$ be the direction of the optical axis. In the paraperspective projection model, a point p whose location is $\mathbf{s}_p$ is observed in the frame f at image coordinates (u_{fp}, v_{fp}), where

$$u_{fp} = \mathbf{m}_f \cdot \mathbf{s}_p + x_f, \quad v_{fp} = \mathbf{n}_f \cdot \mathbf{s}_p + y_f \qquad (1)$$

Here, if we assume unit focal length for the camera,

$$x_f = \frac{(-\mathbf{t}_f) \cdot \mathbf{i}_f}{z_f}, \quad y_f = \frac{(-\mathbf{t}_f) \cdot \mathbf{j}_f}{z_f}$$
$$\mathbf{m}_f = \frac{\mathbf{i}_f - x_f \mathbf{k}_f}{z_f}, \quad \mathbf{n}_f = \frac{\mathbf{j}_f - y_f \mathbf{k}_f}{z_f}, \qquad (2)$$
$$z_f = (-\mathbf{t}_f) \cdot \mathbf{k}_f,$$

The perspective projection of point p onto the frame f is given by (U_{fp}, V_{fp}), where

$$U_{fp} = \frac{\mathbf{i}_f \cdot (\mathbf{s}_p - \mathbf{t}_f)}{z_{fp}}, \quad V_{fp} = \frac{\mathbf{j}_f \cdot (\mathbf{s}_p - \mathbf{t}_f)}{z_{fp}}, \qquad (3)$$
$$z_{fp} = z_f + \mathbf{k}_f \cdot \mathbf{s}_p$$

From the Taylor expansion of (3) about the point

$$z_{fp} \approx z_f, \qquad (4)$$

we can show that the paraperspective projection model is an first order approximation of the perspective projection under the condition that

$$|\mathbf{s}_p|^2 / z_f{}^2 \cong 0 \qquad (5)$$

Now, we consider a matrix $\mathbf{W}$ whose element w_{fp} is the image coordinates of the feature point f in the image frame p:

$$\mathbf{W} = \begin{bmatrix} u_{11} & \cdots & u_{1p} \\ \vdots & u_{fp} & \vdots \\ u_{F1} & \cdots & u_{FP} \\ v_{11} & \cdots & v_{1p} \\ \vdots & v_{fp} & \vdots \\ v_{F1} & \cdots & v_{FP} \end{bmatrix} \qquad (6)$$

This matrix is constructed by automatically tracking each feature point over the frames of video image sequences [1] or finding point correspondences among sets of still images. We define *the measurement matrix* $\mathbf{W}^*$ as

$$\mathbf{W}^* = \mathbf{W} - [U_1 \cdots U_f V_1 \cdots V_f]^T [1 \ \cdots \ 1], \qquad (7)$$

where

$$\frac{1}{P}\sum_{p=1}^{P} u_{fp} = U_f, \quad \frac{1}{P}\sum_{p=1}^{P} v_{fp} = V_f \qquad (8)$$

It can be shown that the rank of $\mathbf{W}^*$ is at most 3 regardless of F and P [4], and therefore, we can factorize it into motion $\mathbf{M}$ and shape $\mathbf{S}$:

$$\mathbf{W}^*_{(2F \times P)} = \mathbf{R}_{(2F \times 3)} \mathbf{S}_{(3 \times P)}. \qquad (9)$$

3 Estimating Image Coordinates of Occluded Feature Points

The assumption that all elements of the measurement matrix are known is violated anyway in real image sequences because of the occlusion of feature points or the tracking ability limitation of the feature tracker. To estimate the image coordinates of these occluded feature points, a sub-matrix is constructed for each occluded feature point from the measurement matrix [4]. All the elements in the sub-matrix are known except for the two corresponding to coordinates of the occluded feature point. The sub-matrix is then factorized in the row-wise extension manner or in the column-wise manner [4] to get partial 3D information. Under the projection model assumed, the missing elements then can be found by projecting the 3D information onto the frame image. Note that the projection model used in estimating the coordinates of occluded points is a linear approximation of the real perspective projection model, and the sub-matrix is rather smaller than the measurement matrix. The condition of approximation may no longer be satisfied and the redundancy may not be enough in a small sub-matrix. Though constructing a larger sub-matrix is possible, it is costly to estimate a lot of unknown elements.

Another method described in [3] to cope with the problems of feature points' occlusion is using weighted factorization by assigning a confidence value to each element of the measurement matrix, say zero for occluded points. It becomes to solve a nonlinear weighted least squares problem, and the iterations required for solving the system increase rapidly when the proportion of zero elements is large in the measurement matrix.

We note here that the expansion of (3) about the point of (4) must also satisfy the following.

$$\mathbf{k}_f \cdot \mathbf{s}_p \approx 0 \tag{10}$$

It means that the projection components of feature points along an optical axis must be small enough. Furthermore, if feature points lie on *hypothetical image plane* that is parallel to the image plane (Fig. 1), the paraperspective projections of these points coincide with their perspective projections. This gives us a hint for constructing a *good* sub-matrix to estimate the occluded feature point more precisely.

It is impossible anyway to find out the points directly from the measurement matrix that satisfy (10) since no 3D information is available at that time. However, if all the elements in the sub-matrix are concentrated in a small area within a frame, the dot product in (10) can be assumed reasonably to be small enough. Under the assumption that motion of feature points between frames is small, the projection of the occluded point at the previous frame will be used for constructing a sub-matrix. To guarantee that those known elements have almost the same depth, the image velocities of these points are also checked. If the sub-matrix constructed is unable to estimate the occluded projection, the size of sub-matrix will be increased and the estimation is done again.

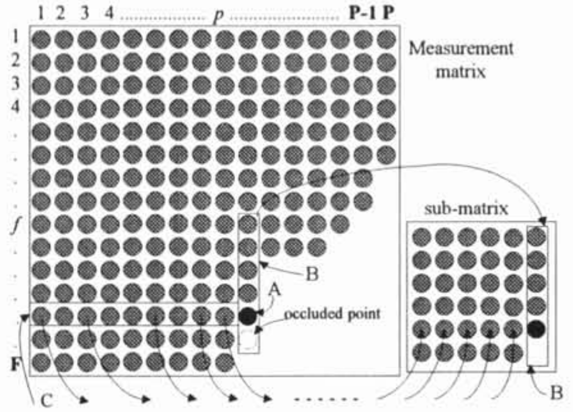

Fig. 2 Constructing a sub-matrix for estimating the occluded point projection from the measurement matrix.

The algorithm of constructing a sub-matrix for each occluded feature point is summarized as below and illustrated in Fig. 2 (only one half of measurement matrix is drawn for simplification).

Before estimating the occluded feature points, the measurement matrix is sorted so that known elements are concentrated at the upper-left part of the matrix. For each column including occluded point in the measurement matrix:

1) Find the first occluded point and register its projection coordinates ***A*** on previous frame.
2) Find out the elements in the row ***C*** whose projection coordinates are close to ***A***, and permute them according to their distances from ***A***.
3) Construct the sub-matrix by using columns including elements found in 2) with the same number of rows as ***B***.
4) Factorize the sub-matrix in the row-wise extension or in the column-wise extension [4] to get partial 3D information about the occluded point.
5) If the factorization succeeds, go to 7). Otherwise check whether the size of the sub-matrix can be increased or not. If it can, increase the size and go to 4). If it can not be increased any more, leave this column unsolved.
6) Swap this column with the next column including occluded points. Go to 1).
7) Find out the projection coordinates of the occluded point using results of 4)
8) Find the next occluded point in this column, then go to 2). If this column is totally estimated, swap this column with the next column including the occluded points.

4 Experimental Results

We validated our proposed estimation method quantitatively with synthetic data, and some results of 3D recovery from a real image sequence will be shown.

4.1 Analysis with Synthetic Data

To evaluate the estimation of occluded feature points, a synthetic measurement matrix was generated and the fill fraction (fraction of known elements) was varied from 1.0 to 0.65. The coordinates in the measurement matrix were perturbed with additive noise for modeling the imprecision of feature tracking. The elements of the measurement matrix were generated by using given 468 3D points and projecting them onto 60 frames with the given camera motion at a distance of 60 times the maximum length of 3D points under the perspective projection model. Each result represents the average error over 5 runs, using a different seed for each random noise level.

In Figs. 3, 4 and 5, the errors at fill fraction of 1 correspond to the tracking errors in the measurement matrix. Fig.3 shows that under lower level of noise ($\sigma \leq 0.1$), the errors of the measurement matrix increase slightly when fill fraction decreases. However, under higher level of noise ($\sigma \geq 0.5$), the errors decrease conversely when fill fraction decreases. This implies that the constructed sub-matrix works fine even though its elements include higher level of noise. There is a little rebound of error when fill fraction is down to 0.65, since it becomes hard to keep the density of feature points in the sub-matrix high under lower fill fraction in the measurement matrix.

In Figs. 4 and 5, the average errors of shape and rotation, which are results of factorization applied to the measurement matrix in Fig. 3 as a function of fill fraction, are shown respectively. The same trends of error decreasing at lower fill fraction can also be observed.

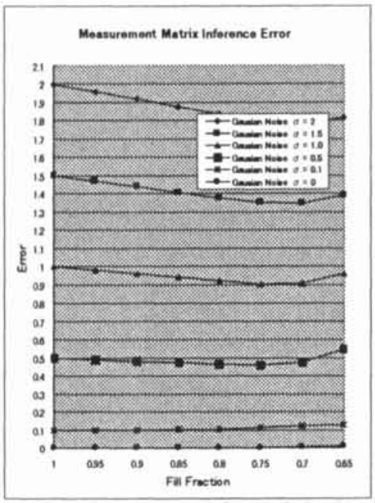

Fig.3 The estimation error of occluded points in measurement matrix.

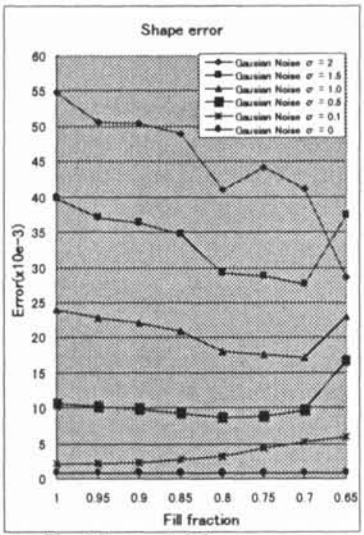

Fig. 4 Recovered shape error.

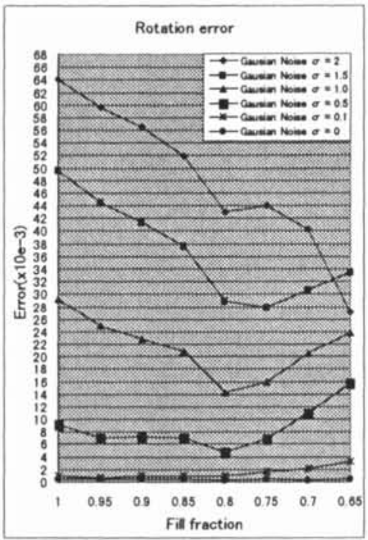

Fig. 5 Recovered rotation error.

4.2 Shape Recovered from Real Image Sequences

A MOAI model (180×70×50 mm) placed on a turntable was imaged by a commercial digital video camera (Sony DCR-VX9000) at a distance of about 3 meters. The first frame of the image sequence is shown at Fig. 6(a), and the result of selecting 3000 feature points using [1] is shown at Fig. 6(b). The fill fraction is 0.94 after the feature points were tracked throughout 60 frames. Two different views of recovered shape are shown at Fig. 6(c), (d).

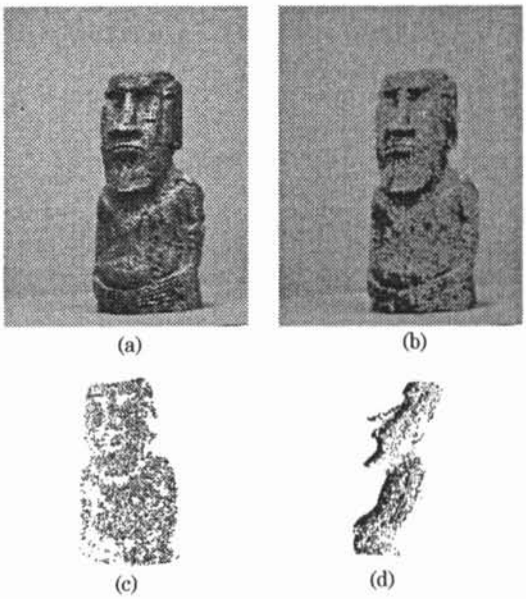

Fig. 6 First frame of the image sequence (a), feature selection results (b), and two views of recovered shape (c, d).

5 Conclusions

Based on the factorization method with the paraperspective projection model, we have presented a method for estimating image coordinates of occluded feature points in order for the estimated locations to approximate closely their perspective projections. The proposed method virtually enables the accurate tracking of locations of all the feature points throughout the image sequence, even when a large motion is allowed for the camera. Consequently, based on the feature correspondences among image frames, the object shape and the camera motion can be recovered accurately from a long image sequence. Through the evaluation of estimation accuracy with synthetic data, a significant improvement has been observed by incorporating the proposed method. Some results have also been presented for 3D information extraction from real image sequences.

References

[1] B.D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. Seventh Int'l Conf. Artificial Intelligence*, 1981.

[2] Y. Ohta, K. Maenobu, and T. Sakai, "Obtaining surface orientation from texels under perspective projection," *Proc. Seventh Int'l Joint Conf. Artificial Intelligence*, pp. 746—751, 1981.

[3] C.J. Poelman and T. Kanade, "A paraperspective factorization method for shape and motion recovery," *CMU-CS-93-219*, 1993

[4] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *International Journal of Computer Vision*, vol. 9, 1992, pp. 137—154.