

# A Factorization Method Using 3-D Linear Combination for Shape and Motion Recovery

Kuo-chang Hwang  
Japan Systems Co., Ltd.  
2-31-24 Ikejiri, Setagaya, Tokyo, Japan  
kouk@japan-systems.co.jp

Naokazu Yokoya, Haruo. Takemura, Kazumasa Yamazawa  
Nara Institute of Science and Technology  
8916-5 Takayama, Ikoma, Nara, Japan  
{yokoya, takemura, yamazawa}@is.aist-nara.ac.jp

## Abstract

*This study proposes a new factorization method for shape and motion recovery. In past, the fourth greatest singular value of the measurement matrix was ignored. But when noise is larger enough so that the fourth greatest singular value can not be ignored, it would be difficult to get reliable results by using the traditional factorization method. In order to acquire reliable results, We start with adopting an orthogonalization method to find a matrix which be composed of three mutually orthogonal vectors. By using this matrix another matrix can be obtained. Then, the two expected matrices which represent shape of object and motion of camera/object, can be obtained through normalization. This study also conducts several experiments to discuss the feasibility of the proposed method.*

## 1. Introduction

Much attention has been put on computing the 3-D shape and motion from a long sequence of images during the last few years. Previous approaches for solving this problem usually consider: 1) whether the camera is calibrated or not, 2) whether a projective or an affine model is used [8][10][2].

Among them, Tomasi and Kanade [8] developed a robust and efficient method for accurately recovering the shape of object and motion of camera/object from a sequence of images under orthographic projection, called *factorization method*. The factorization method has been believed to be possible under linear approximations of imaging system and without camera calibration. Lately, the original factorization method have been extended to scaled orthographic, paraperspective projection [4] and perspective projection [9]. However, in most cases they all start with the three greatest singular values acquired by the Singular Value Decomposition (SVD) technique [5] to factor the measurement matrix into two matrices. Then, by normalizing the

two matrices the expected two matrices which represent object shape and camera motion is obtained.

Unfortunately, when the image noise is larger enough so that the fourth greatest singular value can not be ignored, the traditional factorization method might fail to reach the accurate solution. Another problem remains in the normalization procedure of the factorization method. It is because that sometimes the unknown invertible matrix might be difficult to get [6][1].

This study concentrates on factorization method from the Rank-Theorem perspective, and improves the step of factoring by the SVD technique. According to the Rank-Theorem, it would be possible to get three mutually orthogonal vectors from a measurement matrix [3]. Once the three mutually orthogonal vectors were identified as one matrix, it would be easy to take a form of *3-D linear combination equation* for obtaining another matrix whose elements are three coefficients of  $2F$  sets ( $F$  represents the number of frames). Then the normalization of the two matrices can help this study to recover object shape and camera motion, and by using the proposed method the problem of normalization can be easily solved (described in Section 4).

This paper presents a form of factorization under orthographic projection, although the form of factorization also can be easily extend to other projective models. This study also provides the field of *structure-from-motion* with three advantages as follows: 1) It can robustly recover object shape and camera motion even if the emergence of image noise; 2) It can effectively solve the problem of normalization; 3) Its computation is very fast due to a simple algorithm. In addition, this paper also presents a series of experiment to show the feasibility of the proposed method.

## 2. A Summary of Traditional Factorization Method and Problems Description

This section presents a summary of the traditional factorization method under orthographic projection, and points

out two problems: one is on the fourth greatest singular value (described in Section 2.3), and the other is on the normalization (described in Section 2.4).

## 2.1. Orthographic Projection

Under orthographic projection model, the projection  $(x_{fp}, y_{fp})$  of the  $p$ -th point  $\mathbf{s}_p = (s_{x_p}, s_{y_p}, s_{z_p})^T$  in 3D space onto image frame  $f$  is given as follows:

$$x_{fp} = \mathbf{i}_f^T \cdot (\mathbf{s}_p - \mathbf{t}_f), \quad y_{fp} = \mathbf{j}_f^T \cdot (\mathbf{s}_p - \mathbf{t}_f), \quad (1)$$

where  $\mathbf{t}_f = (t_{x_f}, t_{y_f}, t_{z_f})^T$  is the vector from the world origin to the origin of image frame  $f$ .  $\mathbf{i}_f$  and  $\mathbf{j}_f$  are a pair of unit vectors which represent x-axis and y-axis. These equations can be rewritten as:

$$x_{fp} = \mathbf{m}_f^T \cdot \mathbf{s}_p + t_{x_f}, \quad y_{fp} = \mathbf{n}_f^T \cdot \mathbf{s}_p + t_{y_f}. \quad (2)$$

## 2.2. Measurement Matrix

Suppose that  $P$  feature points were tracked over  $F$  frames of an image sequence were tracked, and their image coordinates  $\{(x_{fp}, y_{fp}) | f = 1, \dots, F, p = 1, \dots, P\}$  were collected into a single  $2F \times P$  measurement matrix  $W$ .

$$W = \begin{bmatrix} x_{11} & \dots & x_{1P} \\ \dots & \dots & \dots \\ x_{F1} & \dots & x_{FP} \\ y_{11} & \dots & y_{1P} \\ \dots & \dots & \dots \\ y_{F1} & \dots & y_{FP} \end{bmatrix}. \quad (3)$$

Equations (2) and (3) of all points and frames can now be combined into a single matrix equation  $W = MS + Te_p^T$ , where  $M$  is the  $2F \times 3$  motion matrix whose rows are  $\mathbf{m}_f^T$  and  $\mathbf{n}_f^T$ ,  $S$  is the  $3 \times P$  shape matrix whose columns consist of  $\mathbf{s}_p$  points, and  $T$  is the  $2F \times 1$  translation vector that collects the projections of camera translation along the image plane and  $\mathbf{e}_p = (1, \dots, 1)^T$ . Then a ‘‘registered’’ measurement matrix can be developed for which the translation vector is subtracted from  $W$  as follow:

$$\tilde{W} = W - Te_p^T = MS. \quad (4)$$

According to Rank-Theorem, the maximum rank of  $M$  and  $S$  is three. Thus, the maximum rank of  $\tilde{W}$  is also three.

## 2.3. Decomposition of Measurement Matrix with SVD

As previously discussed,  $rank(\tilde{W}) \leq 3$  is proved. Here, assume that  $rank(\tilde{W}) = 3$ . Hence, the three greatest singular values can be determined through the SVD technique,

and the fourth and its following singular values almost approaches to zero. Equation (5) displays this process.  $\tilde{W}$ , factoring it into a product of two matrices  $\hat{M}$  and  $\hat{S}$ .

$$\begin{aligned} \tilde{W} &= U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T \\ &\simeq U_1 \Sigma_1 V_1^T \\ &= \hat{M} \hat{S}, \end{aligned} \quad (5)$$

From Equation (5), motion matrix  $M$  and shape matrix  $S$  can be defined as follows:

$$\hat{M} = U_1 \Sigma_1^{\frac{1}{2}}, \quad \hat{S} = \Sigma_1^{\frac{1}{2}} V_1^T. \quad (6)$$

**First Problem:** When noise corrupts the images, the rank of  $\tilde{W}$  will no longer be three. Consider a problem when the fourth greatest singular value is not so small that  $\sigma_4 \simeq \sigma_3$ . Therefore, adopting the SVD algorithm cannot accurately or completely reconstruct the shape and motion. The solution of this problem will be shown in Section 3.

## 2.4. Normalization of $\hat{M}$ and $\hat{S}$

The decomposition of Equation (5) is determined as a linear transformation. Any non-singular  $3 \times 3$  matrix  $A$  and its inverse could be inserted between  $\hat{M}$  and  $\hat{S}$ . Their product should still equal to  $\tilde{W}$ . Thus the actual motion and shape are provided as follows:

$$M = \hat{M} A, \quad S = A^{-1} \hat{S}. \quad (7)$$

The correct  $A$  can be determined by using two geometry constraints that can be defined as follows :

$$\begin{aligned} |\mathbf{m}_f|^2 &= 1, & |\mathbf{n}_f|^2 &= 1, \\ \mathbf{m}_f \cdot \mathbf{n}_f &= 0. \end{aligned} \quad (8)$$

**Second Problem:** In the original papers [8][4], no details of the normalization procedure or criterion to be optimized were presented, and in [6][1] pointed out that many choices are possible for this normalization and a variety of results have been obtained depending on the choice. Indeed, at our knowledge, matrix  $A$  might not be obtained. The more detailed description of the reason and the solution can be found in Section 4.

## 3. Decomposition of Measurement Matrix Using 3-D Linear Combination

In Section 2.2, maximum rank of  $\tilde{W}$  be three was introduced. Let us consider the Rank-Theorem again. The rank of three means that there are three independent vectors. Each row of  $\tilde{W}$  is projected into a space which is constructed by the three mutually orthogonal vectors. This study attempts to find the three mutually orthogonal vectors  $\hat{S}$ , and then to solve  $\hat{M}$  by using the  $\hat{S}$  and  $\tilde{W}$ .

### 3.1. Orthogonalization

First, three orthogonal vectors must be found or built from  $\tilde{W}$ . A good method, the Gram-Schmidt orthogonalization method, can straightforwardly and efficiently find three mutually orthogonal vectors. Rewrite the elements of  $\tilde{W}$  to be  $\tilde{\mathbf{w}}_f = (\tilde{\mathbf{w}}_{f1}, \tilde{\mathbf{w}}_{f2}, \dots, \tilde{\mathbf{w}}_{fp})^T$ , and suppose that the rank of  $3 \times P$  matrix  $\tilde{S} = (\tilde{\mathbf{w}}_{k1}, \tilde{\mathbf{w}}_{k2}, \tilde{\mathbf{w}}_{k3})$  is three. For getting the three orthogonal vectors  $\hat{\mathbf{s}}_i, i = 1, 2, 3$ , the Gram-Schmidt orthogonalization method is extended as follows:

$$\hat{\mathbf{s}}_1 = \tilde{\mathbf{w}}_{k1}, \quad (9)$$

$$\hat{\mathbf{s}}_j = \tilde{\mathbf{w}}_{kj} - \sum_{i=2}^j \frac{\hat{\mathbf{s}}_{i-1}^T \cdot \tilde{\mathbf{w}}_{kj}}{\|\hat{\mathbf{s}}_{i-1}\|^2} \hat{\mathbf{s}}_{i-1}, \quad j = 2, 3. \quad (10)$$

When the three mutually orthogonal vectors were identified as one matrix, it would be easy to take a form of *3D linear combination equation* for obtaining another matrix whose elements are three coefficients of 2F sets (shown in Equation (12)).

$$\hat{m}_{fk} = \frac{\hat{\mathbf{s}}_k^T \cdot \tilde{\mathbf{w}}_f}{\|\hat{\mathbf{s}}_k\|^2}, \quad k = 1, 2, 3. \quad (11)$$

As a consequence, the rows of  $\tilde{W}$  are the product of the elements of  $\hat{\mathbf{s}}_i$  multiplied by  $\hat{m}_{fi}$  in each equation (frame) as following expression,

$$\tilde{\mathbf{w}}_f = \hat{m}_{f1}\hat{\mathbf{s}}_1 + \hat{m}_{f2}\hat{\mathbf{s}}_2 + \hat{m}_{f3}\hat{\mathbf{s}}_3, \quad (12)$$

and as previously described, by using the proposed algorithm (3DLC)  $\tilde{W}$  can be also separated into  $\hat{M}$  and  $\hat{S}$ , as good as Equation (5) obtained by the SVD technique.

In Section 2.3 a problem was pointed out that the fourth greatest singular value which is very close to the third greatest singular value ( $\sigma_4 \simeq \sigma_3$ ). Let us consider again, in this situation the rank of  $\tilde{W}$  will be over three. The solution of the problem is that even if the  $\sigma_4 \simeq \sigma_3$ , the elements of  $\tilde{W}$  can be also projected into a space which is constructed by  $\hat{S}$  through the proposed method.

### 3.2. Determination of $\hat{\mathbf{s}}_1$

So far, how to determine the first independent vector  $\hat{\mathbf{s}}_1$  was not described. The determination of  $\hat{\mathbf{s}}_1$  has many approaches. In general, there are three alternative ways for solving this problem as follows: 1) Based on the general rule, the feature points of the first frame are the most accurate with respect to all other frames. The first vector could be chosen. 2) Anyone of the exact vector  $\tilde{\mathbf{w}}_i$  could be chosen from  $\tilde{W}$ . 3) Choosing a new vector which is obtained by the average of  $\tilde{W}$  also could be considered. In our experiment, the first approach is employed.

## 4. Solving the Problems of Normalization

In Section 2.4, a summary of the solution of normalization has been described. However, Equation (8) is not sufficient enough for obtaining the invertible matrix  $A$ . Here, how to find  $A$  with geometry constraints will be redescribed.

**Geometry Constraint:** The matrix  $A$  can be solved by two geometry constrains, one is the length of unit vector  $\hat{\mathbf{i}}_f$  and  $\hat{\mathbf{j}}_f$ , and the other is the inner product of orthonormality of axes. Here, Let  $B$  be a symmetric matrix, then the following equations can be developed to satisfy these constraints.

$$\begin{aligned} \hat{\mathbf{m}}_f^T B \hat{\mathbf{m}}_f &= 1, & \hat{\mathbf{m}}_{f+F}^T B \hat{\mathbf{m}}_{f+F} &= 1, \\ \hat{\mathbf{m}}_f^T B \hat{\mathbf{m}}_{f+F} &= 0. \end{aligned} \quad (13)$$

From Equation (13), the matrix  $B$  can be developed as :

$$B = L\Lambda L^T. \quad (14)$$

Because  $W$  is not accurate, Equation (13) can not provide us with a solution of  $B$ . Instead, the conjugate gradient method [5] is adopted to reach a minimum value of the Equation (15) through iterations of acquiring  $B$  value.

$$G(B) = \frac{1}{2} \left( \sum_{f=1}^{2F} (\hat{\mathbf{m}}_f^T B \hat{\mathbf{m}}_f - 1)^2 + \sum_{f=1}^F (\hat{\mathbf{m}}_f^T B \hat{\mathbf{m}}_{f+F})^2 \right). \quad (15)$$

Then  $A$  can be determined as :

$$A = L\Lambda^{\frac{1}{2}}R, \quad RR^T = I. \quad (16)$$

According to Equation (16), the diagonal elements  $\lambda_i, i = 1, 2, 3$  of  $\Lambda$  must be positive, Indeed,  $\lambda_i$  might not be necessarily positive so that  $A$  cannot be solved. The reason is why a problem described in Section 2.4 was pointed out. Of course, the iterative process can be terminated, or  $B$  can be redefined to develop an approximate symmetric matrix by using non-linear algorithm. However it seems not a good idea. By using the proposed method, because the first vector  $\hat{\mathbf{s}}_1$  has 2F selection so this problem can be easily solved that all  $\lambda_i$  are positive. The solution of this problem is demonstrated as the following procedures:

1. Let  $j = 1$ .
2. Let  $\hat{\mathbf{s}}_1 = \tilde{\mathbf{w}}_j$ .
3. By using Equation (10),  $\hat{\mathbf{s}}_2$  and  $\hat{\mathbf{s}}_3$  can be solved.
4. For getting another matrix  $\hat{M}$ , Equation (15) can help us to check  $\lambda_i$  is positive or not.
5. If  $\exists \lambda_i \leq 0$  then goto (2) until all  $\lambda_i$  are positive.

In our experiment, the positive  $\lambda_i$  can be successfully obtained only through one or twice selection.

## 5. Experimental Results

### 5.1. Evaluation of Recovered Shape

The proposed factorization method is first performed with a real image sequence for shape recovery. Experimental target is acquired by hand-held video camera. For feature tracking, an algorithm is adopted based on [7]. At the beginning, 16 points are selected in the first frame, and these feature points are automatically tracked and kept throughout a sequence of 16 frames shown in Figure 1. The reconstructed shape is shown in Figure 2. The recovered shape is very similar to real object for visual comparison.

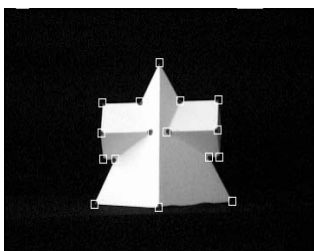


Figure 1. The 16 feature points selected on 1st frame.

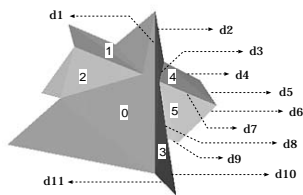


Figure 2. Reconstructed shape.

To evaluate the shape recovery performance quantitatively, the orientations of facets in relation to one another are shown in Table 1 which illustrates the comparison of the results of the three ways: measured angles on the actual model (real), estimated angles with traditional method (SVD) and estimated angles with the proposed method (3DLC).

Table 1. Quantitative evaluation of shape recovery.

| Facets | real b | estimation(SVD) | estimation(3DLC) |
|--------|--------|-----------------|------------------|
| 0-1    | 45°    | 46.47°          | 44.21°           |
| 3-4    | 45°    | 47.54°          | 45.58°           |
| 0-2    | 80°    | 82.08°          | 80.97°           |
| 3-5    | 80°    | 80.91°          | 81.96°           |
| 1-2    | 90°    | 88.68°          | 89.68°           |
| 4-5    | 90°    | 101.71°         | 97.02°           |
| 0-3    | 85°    | 91.77°          | 89.08°           |
| 1-4    | 0°     | 3.88°           | 5.78°            |
| 2-5    | 0°     | 4.04°           | 1.38°            |

The distance of the  $d_1, \dots, d_{11}$  (shown in Figure 2) are also measured. The error between real and estimated data

are evaluated by Equation (17).

$$\epsilon = \frac{1}{L} \sum_{i=1}^L \left\| \frac{Real_i}{\|Real\|} - \frac{Estimation_i}{\|Estimation\|} \right\|. \quad (17)$$

Through the normalization of  $(d_1, \dots, d_L)$ ,  $L = 11 \times 2$  sides  $-1 = 21$ , Using the proposed 3DLC method,  $\epsilon = 0.045$  can be obtained, and Using the traditional SVD method  $\epsilon = 0.042$  can be obtained. As a result, the error of estimated size was within the limit of 4.5 percent.

### 5.2. Evaluation of Recovered Motion

In relation to motion recovery, this experiment only focuses on y-axial rotation because my laboratory does not have much equipments for measuring other motion. Figure 3 was acquired by a static digital camera, and the target is rotated 10 degrees on each frame. The whole sequence is 6 frames. The result is shown in Figure 4 which illustrates the comparison between the measured rotation and computed y-axial rotation of object with the proposed method.

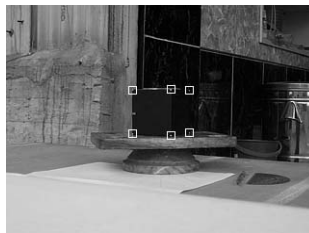


Figure 3. The 6 feature points selected on 1st frame.

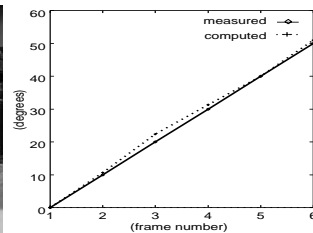


Figure 4. Measured and computed y-axial rotation.

### 5.3. Analysis of Synthetic Data with Noise

This study also demonstrates the robustness of the proposed method in the presence of noise. A situation is simulated that  $\sigma_4 = \sigma_3$ . For attempting to solve the first problem by using the proposed method, a comparison between the proposed 3DLC method and traditional SVD method will be displayed on two synthetic image sequences shown in Figure 5. One (a) is 16 feature points and whole sequence is 70 frames, and the other (b) is 80 feature points and whole sequence is 100 frames, and their the fourth greatest singular value are equal to the third.

In Figure 6, the results in left side were obtained with the SVD technique and the results in right side were obtained with the proposed 3DLC method. From Figure 6, it is observed that (a) could not recover shape completely, but (c) could recover, because that (c) has more points and

frames than (a). Although (c) could recover, several points were not observed on the same plane. By comparison, it is clearly shown that (b) and (d) could accurately recover the object shape.

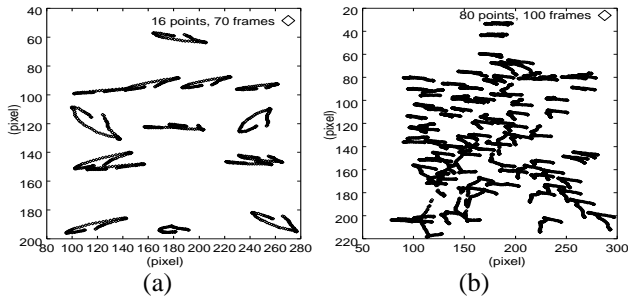


Figure 5. Trajectories of two synthetic data in which  $\sigma_4 = \sigma_3$ .

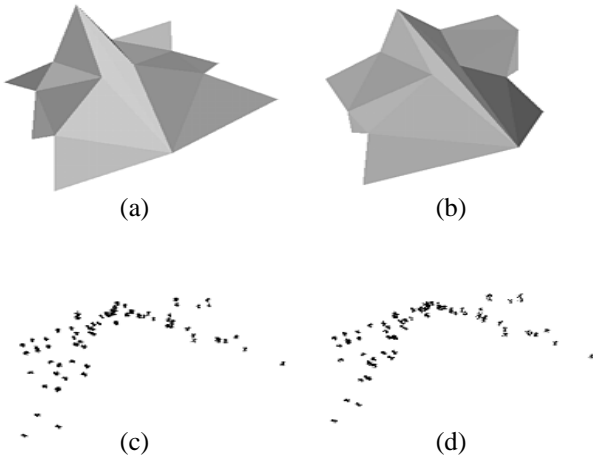


Figure 6. Reconstruction of object shape from synthetic data shown in Figure 5: (a) and (c) are with traditional method, and (b) and (d) are with the proposed method.

### 5.4. Computational Cost

We also show a comparison of efficiency between the proposed 3DLC algorithm and the SVD technique. For measuring computational cost, 20, 30, ..., and 100 feature points are automatically selected in a sequence of 100 frames (the first frame is also shown in Figure 1). The measured computational costs are shown in Figure 7. Here, the feature tracking time is excluded. In result, the proposed 3DLC algorithm is more efficient than SVD. Especially, when points increase, the difference between the SVD technique and the proposed algorithm becomes obvious.

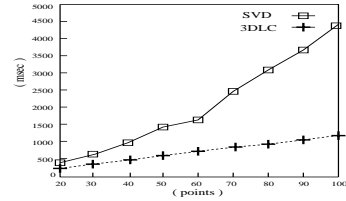


Figure 7. Comparison of computational cost.

## 6. Conclusions

In this paper, a new factorization method which uses 3-D linear combination to decompose the measurement matrix for shape and motion recovery was proposed.

From good point correspondences, the proposed method can not only recover the object shape accurately comparably to those by SVD technique, but also recover robust camera/object motion. It is also demonstrated that when noise is larger enough so that the fourth greatest singular value can not be ignored, the proposed method can still recover the shape robustly. Furthermore, the proposed algorithm is very simply and fast. As a consequence, the present work expands the application of the shape and motion recovery from unreliable data.

Future works include: (1) extension of the method to other projective models; (2) application of the method to 3D model construction of real complex scenes.

## References

- [1] K. Deguchi. 3-d shape reconstruction from endoscope image sequences by the factorization method. *IEICE Trans. Inf. & Syst.*, E79-D(9):1329–1336, 1996.
- [2] O. D. Faugeras. *Three-Dimensional Computer Vision*. The MIT Press, 1993.
- [3] K. Hwang. A new factorization method based on 3d linear combination for shape and motion recovery. *The 55th General Conf. of IPSJ*, 2(5AB-3), Sept. 1997.
- [4] C. J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *Tech. Rep. CMU-CS-93-219*, Dec. 1993.
- [5] W. H. Press. *Numerical Recipes in C*. Cambridge University Press, 1988.
- [6] T. Sasano and K. Deguchi. On the factorization method for shape from image streams. *IEICE Tech. Rep.*, Jan. 1994.
- [7] J. Shi and C. Tomasi. Good features to track. *IEEE CVPR*, pages 593–600, June 1994.
- [8] C. Tomasi and T. Kanade. The factorization method for the recovery of shape and motion from image streams. *Proc. Image Understanding Workshop*, pages 459–472, Jan. 1992.
- [9] B. Triggs. Factorization methods for projective structure and motion. *Proc. Computer Vision and Pattern Recognition*, pages 845–851, 1996.
- [10] S. Ullman. *The Interpretation of Visual Motion*. The MIT Press, Cambridge, MA, 1979.