NAIST-IS-MT9651128

Master's Thesis

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

Kuo-chang Hwang

February 13, 1998

Department of Information Systems
Graduate School of Information Science
Nara Institute of Science and Technology

Master's Thesis
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
MASTER of ENGINEERING

Kuo-chang Hwang

Supervisors: Naokazu Yokoya, Professor
Kunihiro Chihara, Professor
Haruo Takemura, Associate Professor

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

Kuo-chang Hwang

Abstract

This study proposes a new factorization method for shape and motion recovery. It adopts three mutually orthogonal vectors to factor a measurement matrix in order to get two matrices which represent object shape and camera/object motion. 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, this study starts with adopting an orthogonalization method to find a matrix with the three mutually orthogonal vectors. By using this matrix another matrix can be obtained. Then, the two expected matrices can be obtained through normalization. During the process of getting the two expected matrices, this study also conducts several experiments to discuss the feasibility of the proposed method.

Keywords:

factorization, orthogonalization, shape from motion, motion recovery, singular value decompostion

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9651128, February 13, 1998.

3 次元線形結合を用いた因子分解法による物体形状とカメラ運動の復元*

黄 國彰

内容梗概

動画像から物体の形状, あるいはカメラもしくは物体の運動の復元には多くの関心が寄せられてきた。その代表的な手法の一つとして Tomasi と金出が提案した因子分解法がある.

従来の因子分解法は, 殆んど特異値分解を用いて計測行列と呼ばれる画像内の特徴点の時系列データの最も大きな 3 つの特異値のみを見出し, 物体の形状とカメラの運動を表す行列に分解する. しかし, 画像のノイズが大きく, あるいは追跡の不確かさにより, 4 番目の特異値を無視できなくなる. その時, 特異値分解を用いると, 形状や運動を復元できなくなる場合がある. また, 計算コストの面では, 特異値分解を用いたため高くついた. 本論文では, それらの状況を想定し, 従来の因子分解法を改良し, その有効性を示す. Rank-Theorem によると計測行列の階数が 3 であるならば, 必ず 3 つの独立かつ直交するベクトルが存在するため, それらの線形結合で表現できる. 故に, 手法としては, 計測行列から直接 3 つの互いに直行ベクトルを見つけ, 1 つの行列として求める. そして, この行列と計測行列を用い, “3 次元線形結合” というアルゴリズムでもう 1 つの行列を求める. 最後に, この 2 つの行列を正規化すれば, 物体の形状とカメラもしくは物体の運動を復元できる. また, このアルゴリズムが簡単であるため, 従来の手法と比べて大幅の計算コストを削減できる. 本論文では, 提案手法の全容, その応用と有効性を検証する実験について述べる.

キーワード

因子分解法, 直交法, 運動による形状復元, 運動の復元, 特異値分解

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT9651128, 1998 年 2 月 13 日.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background

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 [1][2][11] [12][14].

Among them, Tomasi and Kanade [1][2] developed a robust and efficient method for accurately recovering the object shape and camera/object motion from a sequence of images under orthographic projection (shown in Figure 1), called *factorization method*. The factorization method has been believed to be possible under linear approximations of imaging system and without camera calibration.

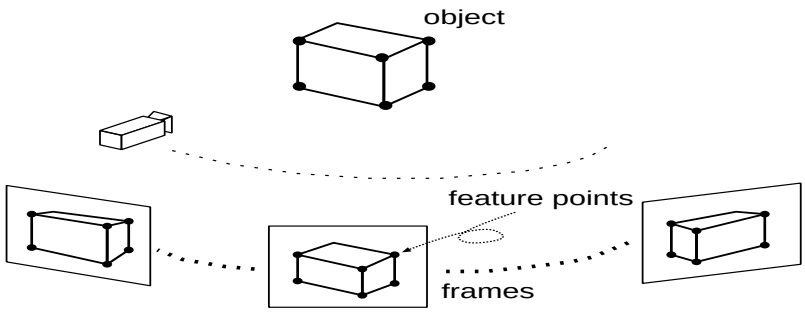


Figure 1. Image system.

Lately, the original factorization method have been extended to scaled orthographic, paraperspective projection by Poelman and Kanade [3][4]. It can be applied to a wider realm of situation than the original factorization method. Furthermore, Costeira [6] improved the original factorization method, and factorization method would be possible to segment and recover the motion and shape of multiple independent moving objects. Recently, Triggs [8] and Deguchi [7] improved the factorization method so that it can be fitted to the case of truly perspective images if *projective depths* are recovered. However, in most cases they all start with the three greatest singular values acquired by the Singular Value

Decomposition (SVD) technique [13] 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 [5][7]. Sasano and Deguchi introduced several methods which are possible for solving the problem of the normalization. At the same time, several methods, such as Powell's method, conjugate gradient method, or Quasi-Newton method [13], also have been known recently. However, it has been found that the different methods might come up with different results. Under this circumstance, the problem of the normalization still exist.

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 and independent vectors from a measurement matrix [10][20]. 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.

1.2 Paper Organization

The remainder of this paper is organized as follows. Section 2 briefly recall a summary of traditional factorization developed by Tomasi and Kanade and point out several problems. Section 3 shows how to improve the traditional factorization. Here, this study proposes a *3D linear combination* algorithm that use a simple and straightforward formulation of the orthogonalization method to decompose measurement matrix into two matrices. Section 4 redescribe the procedure of normalization in details, and shows the solution of the problem of normalization. Section 5 shows the complete algorithm for the 3DLC factorization method. Section 6 provides several experiment to prove the feasibility of the proposed method, and compare the traditional method with the proposed method. Section 7 shows a simple application of the proposed method for real/synthetic complex scene. Finally, Section 8 provides a conclusion and gives directions for future work.

2. A Summary of Traditional Factorization Method and Problems Description

This section presents a summary of the 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 nomalization (described in Section 2.4).

2.1 Orthographic Projection

Under orthographic projection model, shown in Figure 2, 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 by the following equations:

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

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.

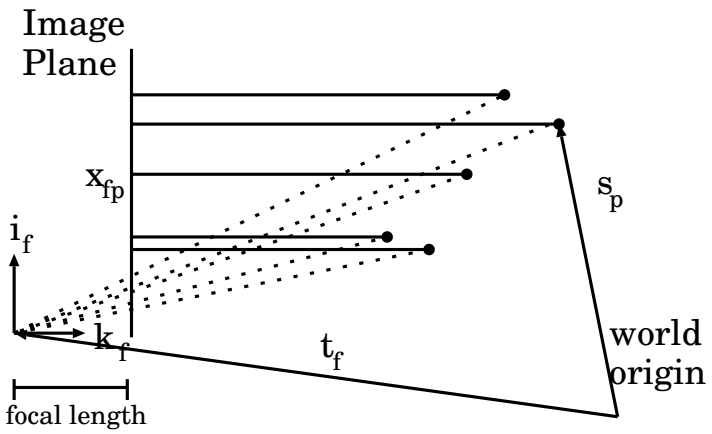


Figure 2. Orthographic projection in two dimensions.

These equations can be rewritten as:

$$x_{fp} = \mathbf{m}_f^T \cdot \mathbf{s}_p + t_{x_f}, \qquad y_{fp} = \mathbf{n}_f^T \cdot \mathbf{s}_p + t_{y_f}, \tag{2}$$

where

$$\mathbf{m}_f = \mathbf{i}_f, \qquad \mathbf{n}_f = \mathbf{j}_f, \tag{3}$$

$$t_{x_f} = -(\mathbf{t}_f^T \cdot \mathbf{i}_f), \qquad t_{y_f} = -(\mathbf{t}_f^T \cdot \mathbf{j}_f). \tag{4}$$

2.2 Measurement Matrix

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

$$W = \begin{bmatrix} x_{11} & \ldots & x_{1P} \\ \ldots & \ldots & \ldots \\ x_{F1} & \ldots & x_{FP} \\ y_{11} & \ldots & y_{1P} \\ \ldots & \ldots & \ldots \\ y_{F1} & \ldots & y_{FP} \end{bmatrix}, \tag{5}$$

Each column of the measurement matrix W represents the image trajectory of one feature point over all frames, and each row of the measurement matrix W contains the image coordinates x or y of all the feature points in each frame.

Equations (2) and (5) of all feature points and frames can now be combined into a single matrix equation as follow:

$$W = MS + T\mathbf{e}_p^T, \tag{6}$$

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 $\mathbf{e}_p = (1, \ldots, 1)^T$.

Then a "registered" measurement matrix can be developed for which the translation vector is subtracted from W as follow:

$$\tilde{W} = W - T\mathbf{e}_p^T = MS. \tag{7}$$

Hence, the measurement matrix $\tilde{W}$ is a product of two matrices which represent motion matrix M and shape matrix S. According to Rank-Theorem, the maximum rank of $M(2F \times 3)$ and $S(3 \times P)$ is three. Thus, the maximum rank of the measurement matrix $\tilde{W}$ is also three.

2.3 Decomposition of Measurement Matrix with SVD

Most traditional factorization methods use the SVD algorithm to decompose the measurement matrix $\tilde{W}$. Here, this subsection present a description of the SVD algorithm for decomposing $\tilde{W}$ into $\hat{M}$ and $\hat{S}$. Besides, the research questions of this study will also be expressed.

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 (8) displays this process. $\tilde{W}$, factoring it into a product of two matrices $\hat{M}$ and $\hat{S}$.

$$\begin{aligned} \tilde{W} &= U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T \\ &\simeq U_1 \Sigma_1 V_1^T \\ &= \hat{M} \hat{S}, \end{aligned} \tag{8}$$

where $\Sigma_1 = diag(\sigma_1, \sigma_2, \sigma_3)$ is a 3×3 diagonal matrix whose diagonal values are the greatest three singular values of $\tilde{W}$, and $\Sigma_2 = diag(\sigma_4, \ldots, \sigma_P)$ is a $(P-3) \times (P-3)$ diagonal matrix whose diagonal values are close to zero.

From Equation (8), motion matrix M and shape matrix S can be defined as follows:

$$\hat{M} = U_1 \Sigma_1^{\frac{1}{2}}, \qquad \hat{S} = \Sigma_1^{\frac{1}{2}} V_1^T. \tag{9}$$

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 (8) is determined as a linear transformation. Any non-singular 3×3 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:

$$M = \hat{M} A, \qquad S = A^{-1} \hat{S}. \tag{10}$$

The correct A can be determined using the fact that the rows of the motion matrix M represent the camera axes, and they must be of a certain form. Since $\mathbf{i}_f$ and $\mathbf{j}_f$ are unit vectors, a geometry constraint can be defined by Equation (3) as follow :

$$|\mathbf{m}_f|^2 = 1, \qquad |\mathbf{n}_f|^2 = 1, \tag{11}$$

and because they are orthogonal, another geometry constraint can be also defined as follow :

$$\mathbf{m}_f \cdot \mathbf{n}_f = 0. \tag{12}$$

Using Equations (11) and (12), the matrix A can be obtained.

Second Problem: In the original papers [2][3], no details of the normalization procedure or criterion to be optimized were presented, and in [5][7] 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 and mutually orthogonal vectors. Each row of $\tilde{W}$ is projected into a space which is constructed by the three mutually orthogonal vectors (as illustrated in Figure 3). This study attempts to find the three mutually orthogonal vectors $\hat{S}$, and then to solve $\hat{M}$ by using the $\hat{S}$ and $\tilde{W}$.

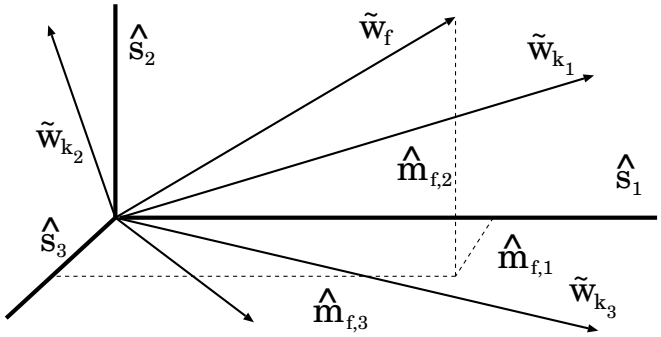


Figure 3. A space built by three mutually orthogonal vectors.

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 normal orthogonal vectors from a series poly-dimensional vectors. In this study, the Gram-Schmidt orthogonalization method is adopted to find three mutually orthogonal vectors.

Rewrite the elements of $\tilde{W}$ to be $\tilde{\mathbf{w}}_f = (\tilde{\mathbf{w}}_{f1}, \tilde{\mathbf{w}}_{f2}, \ldots, \tilde{\mathbf{w}}_{fp})^T$, and suppose that the rank of $3 \times P$ matrix $\tilde{S} = (\tilde{\mathbf{w}}_{k_1}, \tilde{\mathbf{w}}_{k_2}, \tilde{\mathbf{w}}_{k_3})$ is three. For getting the three orthogonal vectors $\hat{\mathbf{s}}_i, i = 1, 2, 3$, now the Gram-Schmid orthogonaliation method is extended as follows:

$$\hat{\mathbf{s}}_1 = \tilde{\mathbf{w}}_{k_1}, \tag{13}$$

$$\hat{\mathbf{s}}_2 = \tilde{\mathbf{w}}_{k_2} - \frac{\hat{\mathbf{s}}_1^T \cdot \tilde{\mathbf{w}}_{k_2}}{\|\hat{\mathbf{s}}_1\|^2}\hat{\mathbf{s}}_1, \tag{14}$$

$$\hat{\mathbf{s}}_3 = \tilde{\mathbf{w}}_{k_3} - \frac{\hat{\mathbf{s}}_1^T \cdot \tilde{\mathbf{w}}_{k_3}}{\|\hat{\mathbf{s}}_1\|^2}\hat{\mathbf{s}}_1 - \frac{\hat{\mathbf{s}}_2^T \cdot \tilde{\mathbf{w}}_{k_3}}{\|\hat{\mathbf{s}}_2\|^2}\hat{\mathbf{s}}_2, \tag{15}$$

and $\hat{\mathbf{s}}_2$ and $\hat{\mathbf{s}}_3$ can be reformulated as follow:

$$\hat{\mathbf{s}}_j = \tilde{\mathbf{w}}_{k_j} - \sum_{i=2}^{j} \frac{\hat{\mathbf{s}}_{i-1}^T \cdot \tilde{\mathbf{w}}_{k_j}}{\|\hat{\mathbf{s}}_{i-1}\|^2}\hat{\mathbf{s}}_{i-1}, \quad j = 2, 3. \tag{16}$$

According to the Rank-Theorem, the rank of $\hat{S} = (\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, \hat{\mathbf{s}}_3)^T$ is still three, and there are following properties:

$$\hat{\mathbf{s}}_1^T \cdot \hat{\mathbf{s}}_2 = 0, \qquad \hat{\mathbf{s}}_1^T \cdot \hat{\mathbf{s}}_3 = 0, \qquad \hat{\mathbf{s}}_2^T \cdot \hat{\mathbf{s}}_3 = 0. \tag{17}$$

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 (20)).

Let

$$\hat{M} = \begin{bmatrix} \hat{m}_{11} & \hat{m}_{12} & \hat{m}_{13} \\ \hat{m}_{21} & \hat{m}_{22} & \hat{m}_{23} \\ \vdots & \vdots & \vdots \\ \hat{m}_{2F,1} & \hat{m}_{2F,2} & \hat{m}_{2F,3} \end{bmatrix}, \tag{18}$$

and each element of $\hat{M}$ can be obtained as follow:

$$\hat{m}_{fk} = \frac{\hat{\mathbf{s}}_k^T \cdot \tilde{\mathbf{w}}_f}{\|\hat{\mathbf{s}}_k\|^2}, \quad k = 1, 2, 3. \tag{19}$$

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,

$$\tilde{\mathbf{w}}_f = \hat{m}_{f1}\hat{\mathbf{s}}_1 + \hat{m}_{f2}\hat{\mathbf{s}}_2 + \hat{m}_{f3}\hat{\mathbf{s}}_3, \tag{20}$$

and as previously described, by using the proposed algorithm (3DLC) measurement matrix $\tilde{W}$ can be also separated into $\hat{M}$ and $\hat{S}$, as good as Equation (8) obtained by the SVD technique.

So far, how to select k_i was not described yet. In order to reach the generalizability, this study supposes that $rank(W) \geq 3$. At first, $k_1 = i, k_2 = F + i$ are

chosen, where i is a fixed positive integer with $1 \leq i \leq F$, say $i = 1$. Then the k_3 is selected according to the following procedure.

Let

$$l_f = \|\mathbf{w}_f\| = \sqrt{\sum_{j=1}^{P} w_{fj}^2}, \tag{21}$$

$$\hat{\mathbf{s}}_1 = \mathbf{w}_{k_1}, \tag{22}$$

$$\tilde{m}_{f1} = \hat{\mathbf{s}}_1^T \mathbf{w}_f = \sum_{j=1}^{P} w_{fj} \hat{s}_{1j}, \tag{23}$$

$$\hat{\mathbf{s}}_2 = \mathbf{w}_{k_2} - \frac{\hat{m}_{k_2 1}}{l_{k_1}^2} \hat{\mathbf{s}}_1, \tag{24}$$

$$\hat{m}_{f2} = \hat{\mathbf{s}}_2^T \mathbf{w}_f = \sum_{j=1}^{P} w_{fj} \hat{s}_{2j}, \tag{25}$$

$$l_1' = l_{k_1}, \ \ l_2' = \|\hat{\mathbf{s}}_2\| = \sqrt{\sum_{j=1}^{P} \hat{s}_{2j}^2}, \tag{26}$$

write

$$u_f = \frac{\hat{m}_{f1}}{l_f l_1'} + \frac{\hat{m}_{f2}}{l_f l_2'}, \tag{27}$$

then the minimum value of u_f is chosen so that $k_3 = f$. Once k_1, k_2, k_3 were determined, the three orthogonal vectors are obtained by the equations (13), (14), and (15).

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.

$$\hat{\mathbf{s}}_1 = \tilde{\mathbf{w}}_1. \tag{28}$$

2. Anyone of the exact vector $\tilde{\mathbf{w}}_\mathbf{i}$ could be chosen from $\tilde{W}$.

$$\hat{\mathbf{s}}_1 = \tilde{\mathbf{w}}_{exact}. \tag{29}$$

3. Choosing a new vector which is obtained by the average of $\tilde{W}$ also could be considered.

$$\hat{s}_{1j} = \frac{1}{2F} \sum_{i=1}^{2F} \tilde{w}_{ij}, \qquad j = 1, \ldots, P. \tag{30}$$

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, Equations (11) and (12) are 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 $\mathbf{i}_f$ and $\mathbf{j}_f(= 1)$, and the other is the inner product of orthonormality of axes ($= 0$). Then the following equations can be developed to satisfy these constraints.

$$\begin{aligned} \|\hat{\mathbf{m}}_f^T A\| &= 1, \\ \|\hat{\mathbf{m}}_{f+F}^T A\| &= 1, \\ \hat{\mathbf{m}}_f^T A \cdot A^T \hat{\mathbf{m}}_{f+F} &= 0. \end{aligned} \tag{31}$$

Here, Let

$$B = AA^T = \begin{bmatrix} a & d & f \\ d & b & e \\ f & e & c \end{bmatrix}, \tag{32}$$

B is a symmetric matrix, then Equation (31) can be rewritten as:

$$\begin{aligned} \hat{\mathbf{m}}_f^T B \hat{\mathbf{m}}_f &= 1, \\ \hat{\mathbf{m}}_{f+F}^T B \hat{\mathbf{m}}_{f+F} &= 1, \\ \hat{\mathbf{m}}_f^T B \hat{\mathbf{m}}_{f+F} &= 0. \end{aligned} \tag{33}$$

From Equation (33), the matrix B can be developed as :

$$B = L \Lambda L^T, \tag{34}$$

where,

$$L = \begin{bmatrix} 1 & 0 & 0 \\ \frac{d}{a} & 1 & 0 \\ \frac{f}{a} & \frac{e - \frac{df}{a}}{b - \frac{d^2}{a}} & 1 \end{bmatrix}, \tag{35}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \tag{36}$$

$$\lambda_1 = a, \qquad \lambda_2 = b - \frac{d^2}{a}, \qquad \lambda_3 = c - \frac{f^2}{a} - \frac{(e - \frac{df}{a})^2}{b - \frac{d^2}{a}}. \tag{37}$$

Because the measurement matrix W is not accurate, Equation (33) can not provide us with a solution of B. Instead, the conjugate gradient method [13] is adopted to reach a minimum value of the Equation (38) through iterations of acquiring B value.

$$G(B) = \frac{1}{2}\left(\sum_{f=1}^{2F}(\hat{\mathbf{m}}_f^T B \hat{\mathbf{m}}_f - 1)^2 + \sum_{f=1}^{F}(\hat{\mathbf{m}}_f^T B \hat{\mathbf{m}}_{f+F})^2\right). \tag{38}$$

Then A can be determined as

$$A = L\Lambda^{\frac{1}{2}}. \tag{39}$$

Remark: According to Equation (39), the diagonal elements of Λ must be positive,

$$\Lambda = diag(\lambda_1, \lambda_2, \lambda_3), \;\; \lambda_i > 0. \tag{40}$$

Indeed, λ_i might not be necessarily positive so that the matrix 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 λ_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 Equations (14) and (15), $\hat{\mathbf{s}}_2$ and $\hat{\mathbf{s}}_3$ can be solved.
4. For getting another matrix $\hat{M}$, Equation (38) can help us to check λ_i is positive or not.
5. If $\exists\lambda_i \leq 0$ then goto (2) until all λ_i are positive.

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

So far, Equation (39) are not so sufficient to determine the matrices M and S completely. The reason for this is because the origin point of the world coordinate was only limited, but the directions of the axes of the world coordinate is not limited. If the axes of the world coordinate are clarified, then from W the matrices M and S can be determined completely. Thus, A can be redefined as follow :

$$A = L\Lambda^{\frac{1}{2}}R, \tag{41}$$

where R is a 3×3 normal orthogonal matrix, and $RR^T = I$, I is a 3×3 identical matrix. For example, the axes of the world coordinate are chosen such that

$$\mathbf{i}_1 = (1,0,0)^T, \qquad \mathbf{j}_1 = (0,1,0)^T. \tag{42}$$

Then Equation (10) can be write as:

$$(\hat{m}_{11}, \hat{m}_{12}, \hat{m}_{13})L\Lambda^{\frac{1}{2}}R = (1,0,0), \tag{43}$$

$$(\hat{m}_{F+1,1}, \hat{m}_{F+1,2}, \hat{m}_{F+1,3})L\Lambda^{\frac{1}{2}}R = (0,1,0). \tag{44}$$

Let $R = (\mathbf{r_1}, \mathbf{r_2}, \mathbf{r_3})$. Then from $RR^T = I$, $\mathbf{r}_1^T, \mathbf{r}_2^T$ can be obtained through following equations.

$$\begin{aligned} \mathbf{r}_1^T &= (1,0,0)R^T \\ &= (\hat{m}_{11}, \hat{m}_{12}, \hat{m}_{13})L\Lambda^{\frac{1}{2}}, \\ \mathbf{r}_2^T &= (0,1,0)R^T \\ &= (\hat{m}_{F+1,1}, \hat{m}_{F+1,2}, \hat{m}_{F+1,3})L\Lambda^{\frac{1}{2}}, \end{aligned} \tag{45}$$

and $\mathbf{r}_3$ satisfies

$$\mathbf{r}_1^T\mathbf{r}_3 = 0, \qquad \mathbf{r}_2^T\mathbf{r}_3 = 0, \qquad \|\mathbf{r}_3\| = 1. \tag{46}$$

These equations will yield two solutions, this reflects the difference between the directions of the z-axis of the world coordinate and the z-axis of the coordinate system which indicates the orientation of the camera corresponding to the first image.

5. Outline of the Complete Algorithm

Based on the development in the previous sections, now a new complete algorithm can be developed for the factorization of the measurement matrix $\tilde{W}$ derived from a sequence of images into motion matrix M and shape matrix S as defined in Equation (7).

1. Compute the matrix $\hat{S}$ which consist of three mutually orthogonal vectors as the following procedure:

   (a) Let $j = 1$.

   (b) Let $\hat{\mathbf{s}}_1 = \tilde{\mathbf{w}}_j$.

   (c) $\hat{\mathbf{s}}_2$ and $\hat{\mathbf{s}}_3$ can be solved through Equation (14) and (15).

2. Once matrix $\hat{S}$ is obtained, for fitting Equation (8), compute another matrix $\hat{M}$ through Equation (19).

3. Compute the symmetric matrix B by imposing the geometry constraints defined in Equation (33), and Equation (38) can help us to acquire B.

4. Check λ_i is positive or not through Equation (37). If $\exists \lambda_i \leq 0$ then goto (1b) and increases j until all λ_i are positive.

5. Determine the expected posture matrix R.

6. Compute the invertible matrix A through Equation (41).

7. Compute the motion matrix M and the shape matrix S by Equation (10).

6. Experimental Results

This study will present four sets of experiments to demonstrate how the proposed method works: 1) An analysis of recovered object shape with real images. 2) An evaluation of recovered object motion with real images. 3) A simulation with synthetic data whose the fourth greatest singular value is equal to the third, and a comparative analysis between the results of this method and the results of the traditional factorization method will be presented. 4) A comparison of computational cost between the proposed method and the traditional factorization method.

6.1 Evaluation of Recovered Object Shape

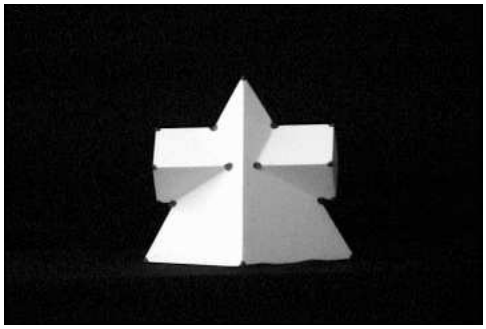
The proposed factorization method is first perform with a real image sequence. The requirement here is that every feature point must be visible in each frame, and can be tracked throughout all the images. Here, the occlusion problem is exclude. Experimental target shown in Figure 4 was acquired by hand-held video camera. For feature tracking, an algorithm is adopted based on [15]. 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 5.

The reconstructed shape is shown in Figure 6, where an algorithm of delaunay net is adopted to build a three dimensional model with 16 triangular patches [14][16][17][18] [19]. The recovered shape is very similar to real object for visual comparison.

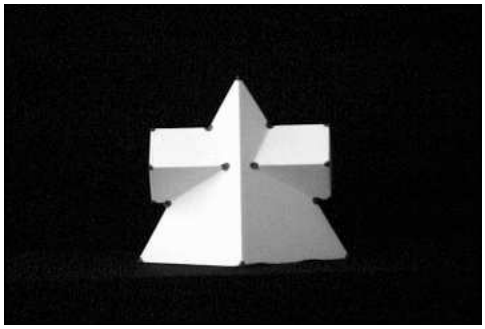
To evaluate the shape recovery performance quantitatively, the orientations of facets in relation to one another are shown in Table 1 which illustrates the comparison of the results of the three ways: measured angles on the actual model (real), estimated angles with traditional method (SVD) and estimated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

The distance of the $d1, \cdots, d11$ (shown in Figure 6) are also measured. The error between real and estimated data are evaluated by Equation (47).

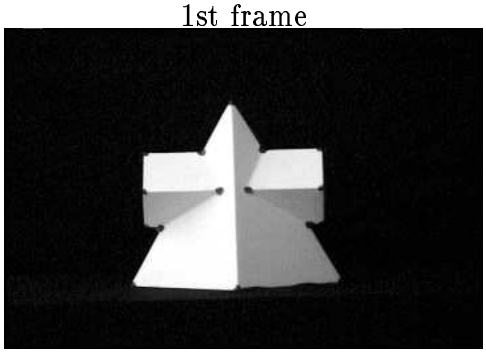
$$\epsilon = \frac{1}{L}\sum_{i=1}^{L}\left\| \frac{Real_i}{\|Real\|} - \frac{Estimation_i}{\|Estimation\|} \right\|. \tag{47}$$



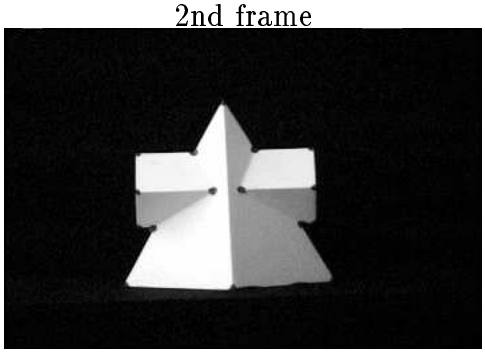
1st frame



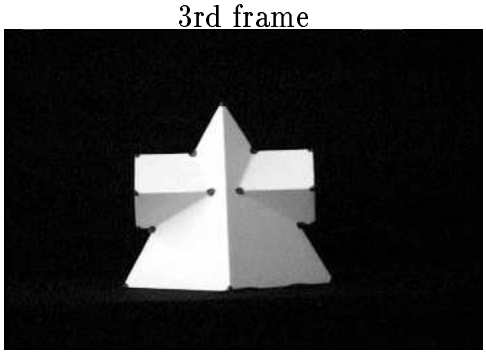
2nd frame



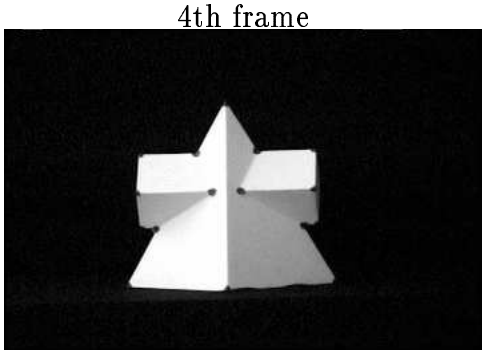
3rd frame



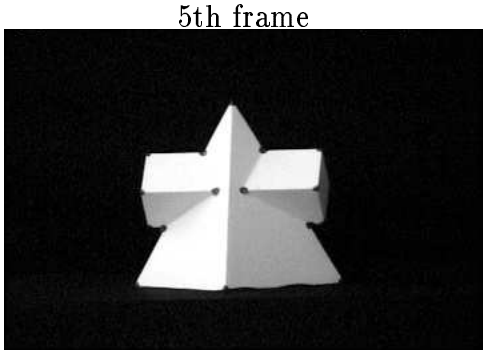
4th frame



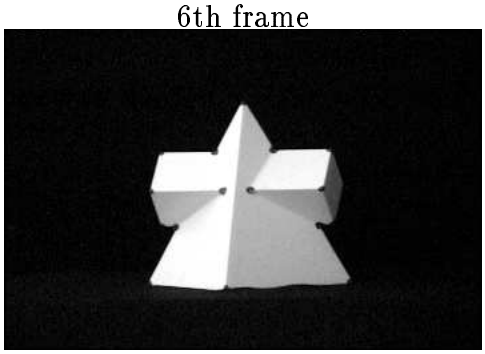
5th frame



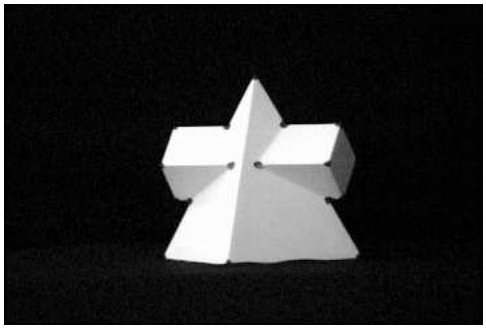
6th frame



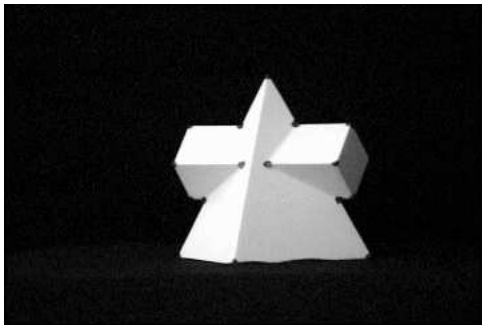
7th frame



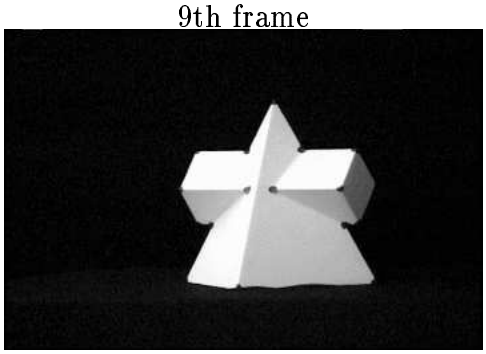
8th frame



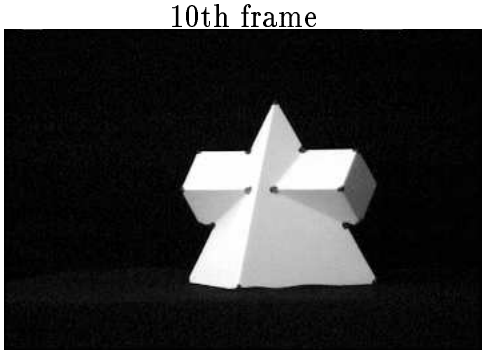
9th frame



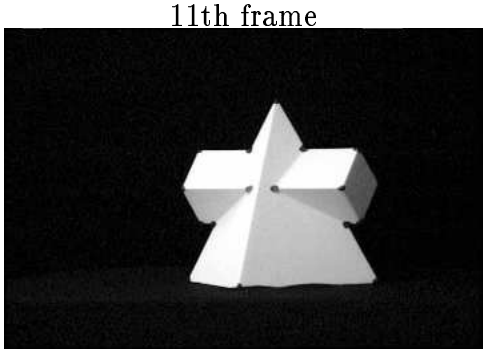
10th frame



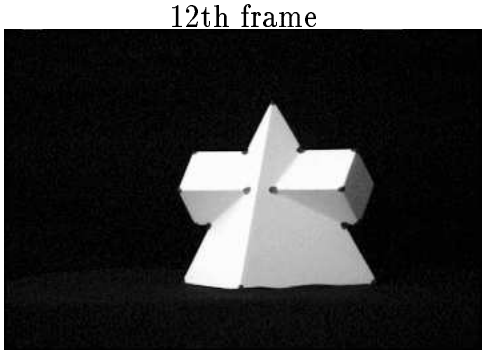
11th frame



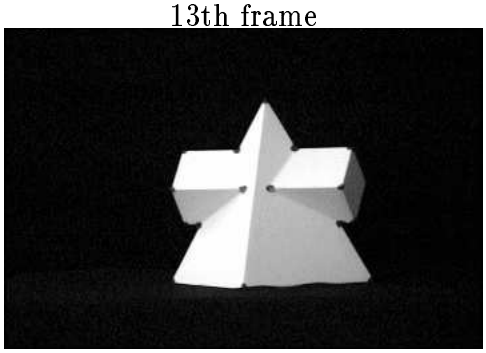
12th frame



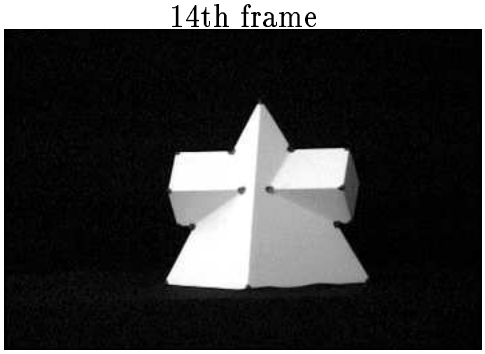
13th frame



14th frame

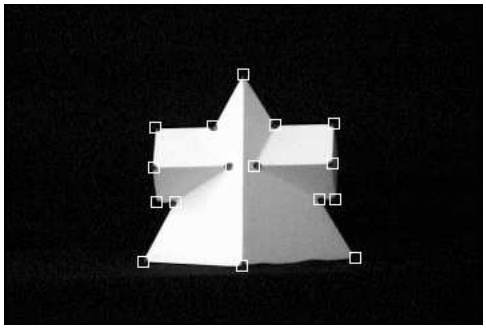


15th frame

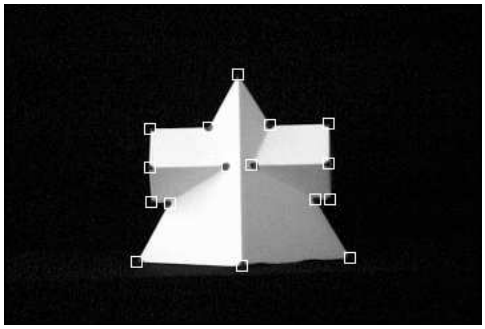


16th frame

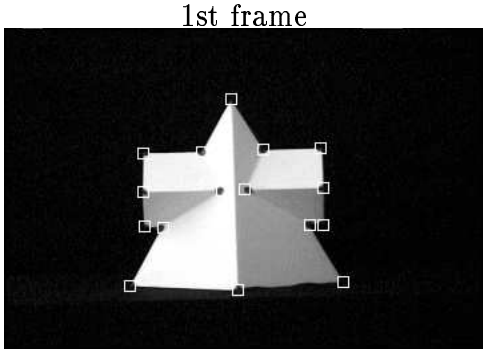
Figure 4. A sequence of 16 frames.



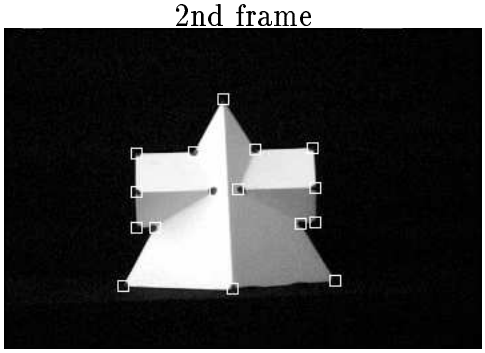
1st frame



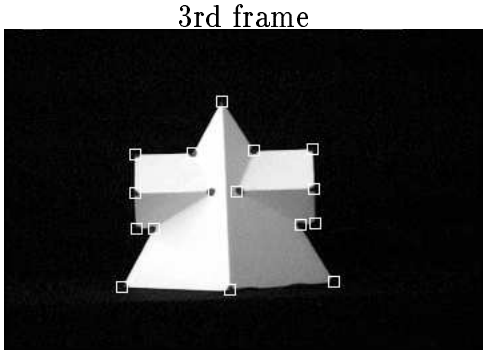
2nd frame



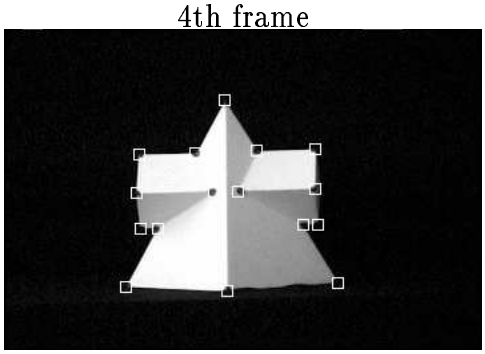
3rd frame



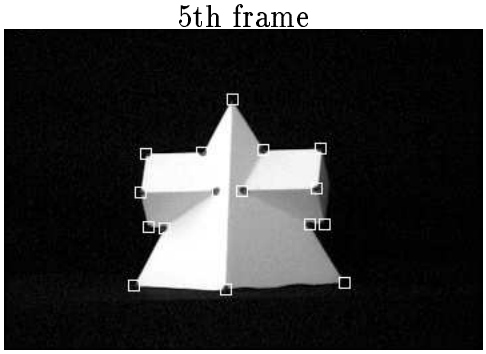
4th frame



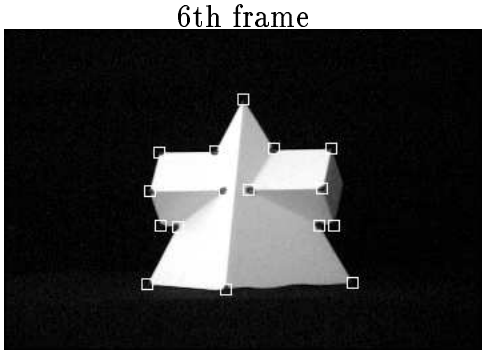
5th frame



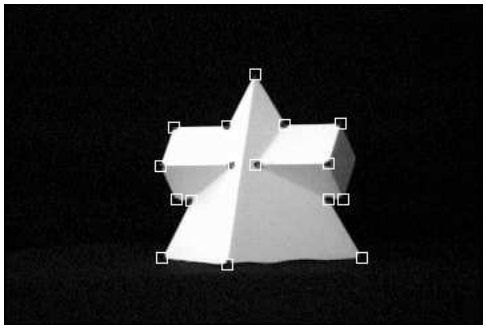
6th frame



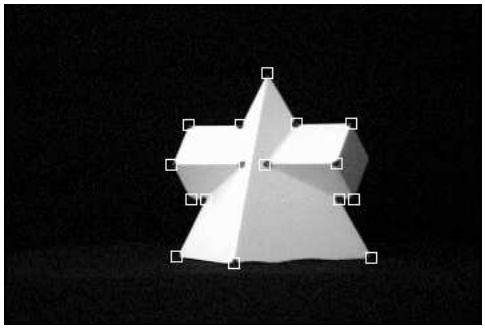
7th frame



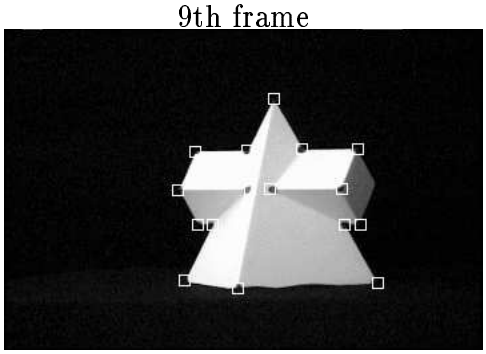
8th frame



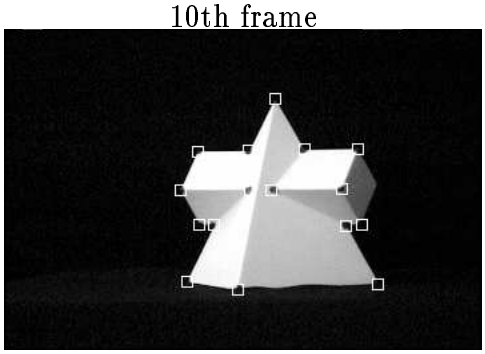
9th frame



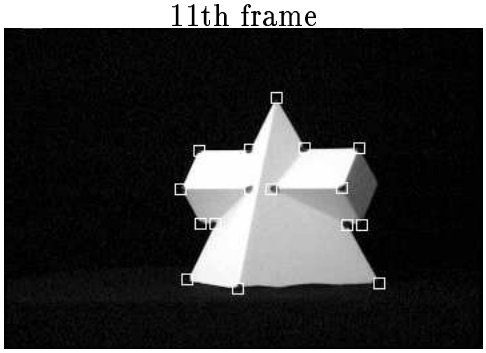
10th frame



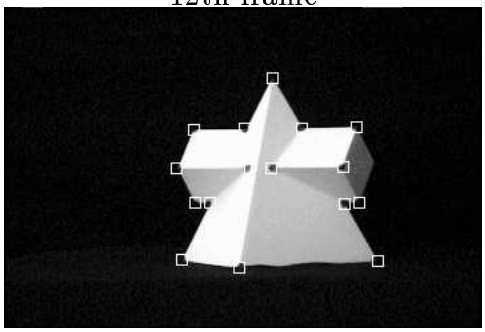
11th frame



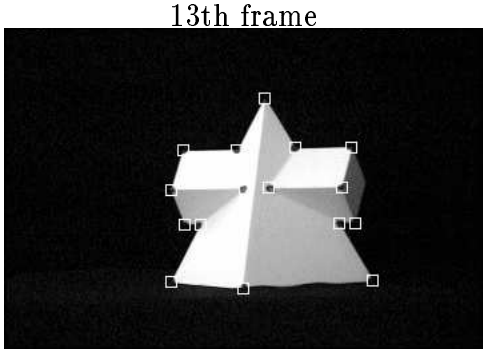
12th frame



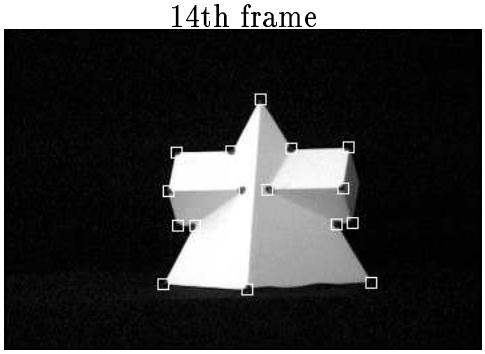
13th frame



14th frame



15th frame



16th frame

Figure 5. The 16 feature points selected by the automatic detection method.

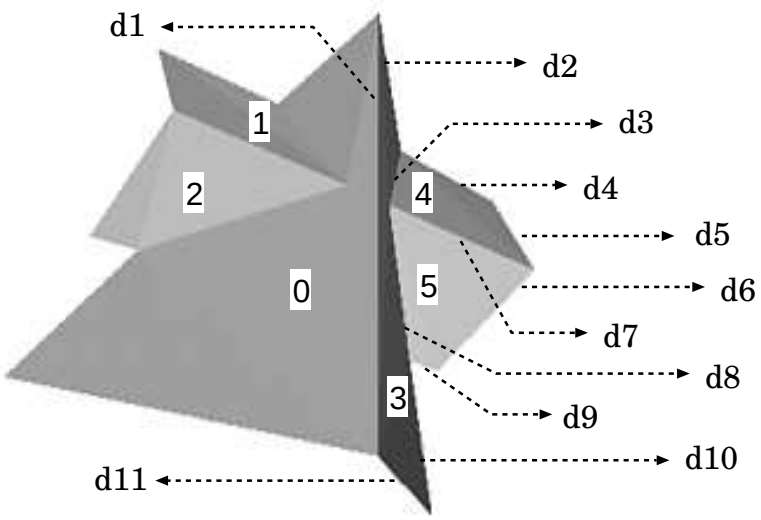


Figure 6. Reconstructed shape with delaunay net and 16 triangular patches.

Table 1. Quantitative evaluation of shape recovery. The proposed 3DLC is compared with traditional SVD in angles between reconstructed facets.

Facets	real	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°

Through the normalization of $(d1, \cdots, dL)$, $L = 11 \times 2$ 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. According to the results, the error of estimated size was within the limit of 4.5 percent.

6.2 Evaluation of Recovered Object Motion

Above experiments have been shown the feasibility of reconstruction 3D shape of object. In relation to motion recovery, this subsection a performance of reconstruction motion of object will be presented, and this experiment only focuses on y-axial rotation because my laboratory does not have much equipments for measuring other motion. Figure 7 was acquired by a static digital camera, and the target is rotated 10 degrees on each frame. Figure 8 shows 6 feature points selected by the automatic detection method, and whole sequence is 6 frames. To evaluate the motion recovery performance quantitatively, the result is shown in Figure 9 which illustrates the comparison between the measured rotation and computed y-axial rotation of object with the proposed method (3DLC).

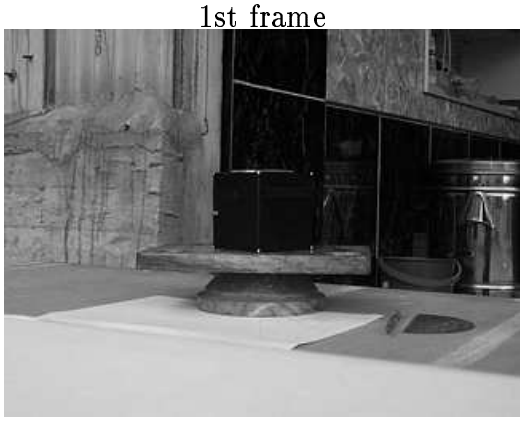
6.3 Analysis of Synthetic Data with Noise

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 3LDC method and traditional SVD method will be displayed on two synthetic image sequences shown in Figure 10. 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.

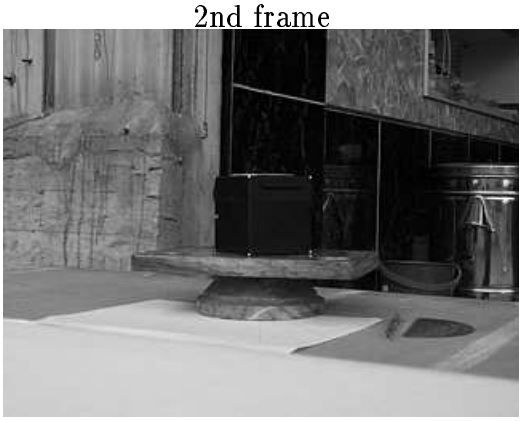
In Figure 11, the results in left side were obtained with the SVD technique and the results in right side were obtained with the proposed 3DLC method. From Figure 11, it is observed that (a) could not recover shape completely, but (c) could recover, because that (c) has more points and frames than (a). 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.

1st frame

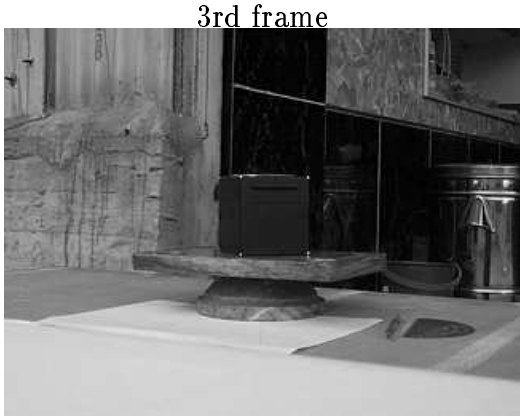
2nd frame



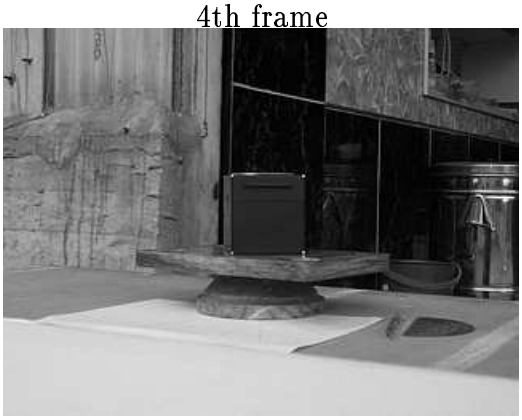
3rd frame



4th frame

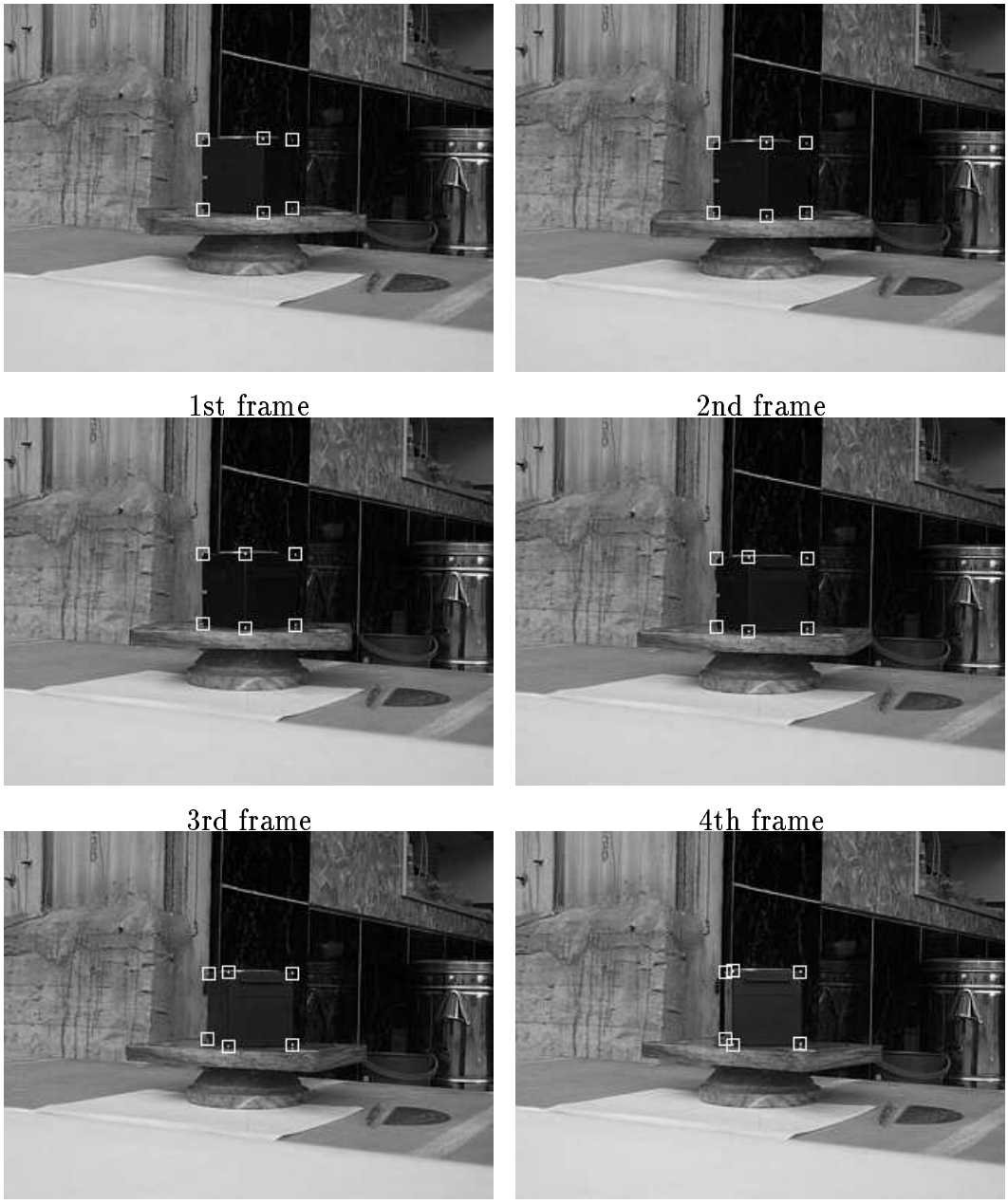


5th frame



6th frame

Figure 7. A sequence of 6 frames.



1st frame 2nd frame

3rd frame 4th frame

5th frame 6th frame

Figure 8. The 6 feature points selected by automatic detection method.

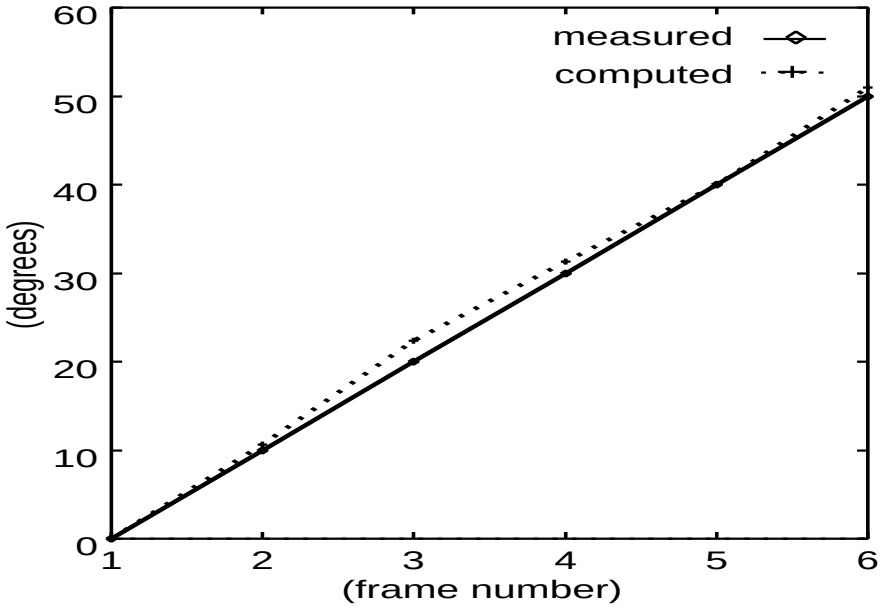
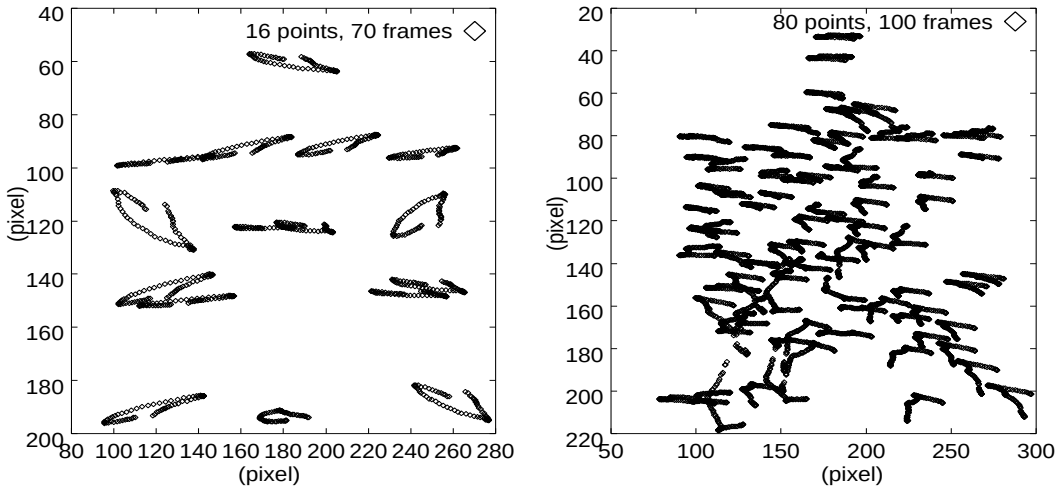


Figure 9. Measured and computed y-axial rotation of object.



(a) (b)

Figure 10. Trajectories of two synthetic data in which $\sigma_4 = \sigma_3$: (a) shows trajectories of 16 points over a sequence (70f), and (b) shows trajectories of 80 points over a sequence (100f).

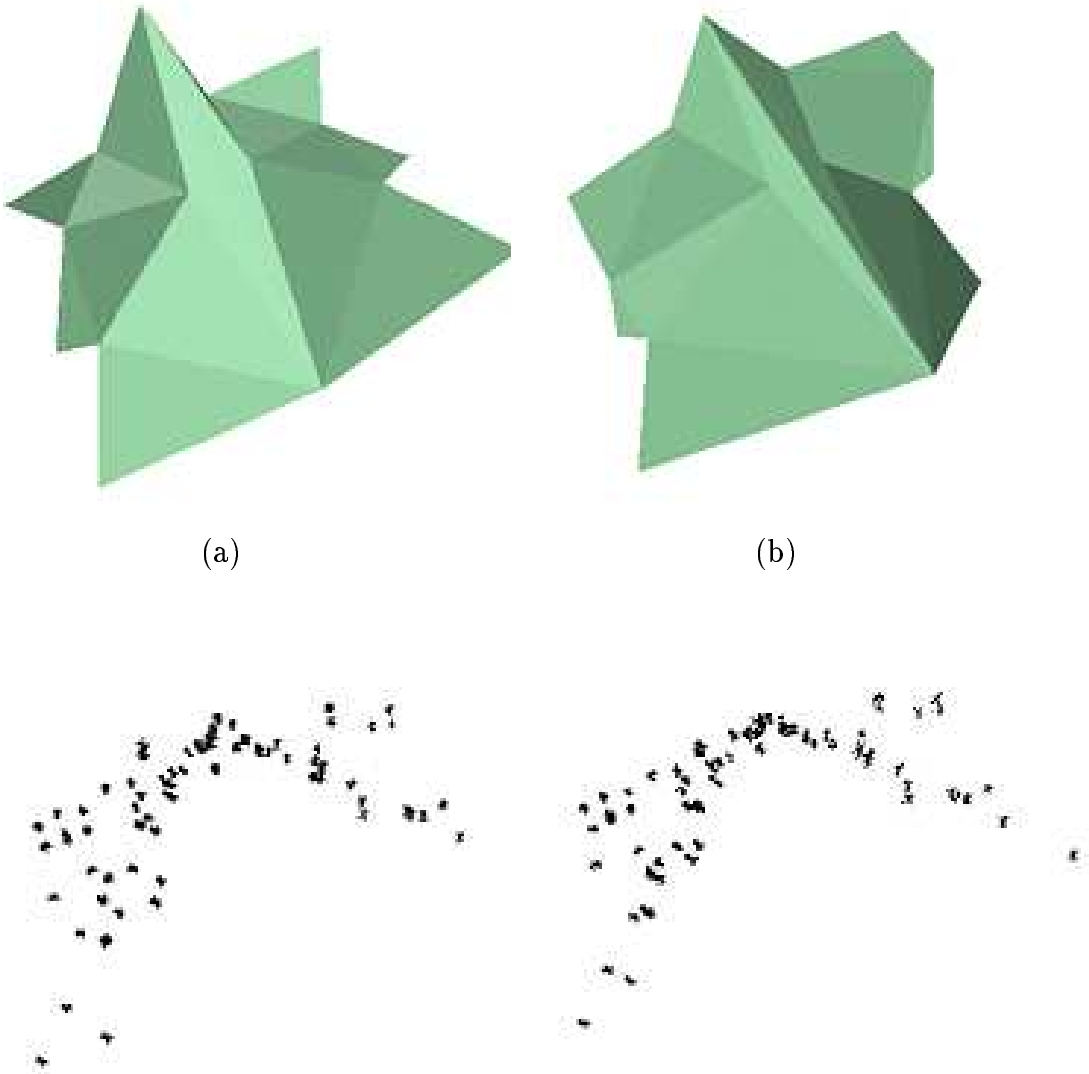


Figure 11. Reconstruction of object shape from synthetic data shown in Figure 10: (a) and (c) are with traditional method, and (b) and (d) are with the proposed method.

6.4 Computational Cost

Now this study shows a comparison of efficiency between the proposed 3DLC algorithm and the SVD technique. For measuring computational cost, $20, 30, \cdots$, and 100 feature points are automatically selected in a sequence of 100 frames (their 16 frames is also shown in Figure 4). The obtained computational costs are shown in Figure 12. Here, the feature tracking time is not included. 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.

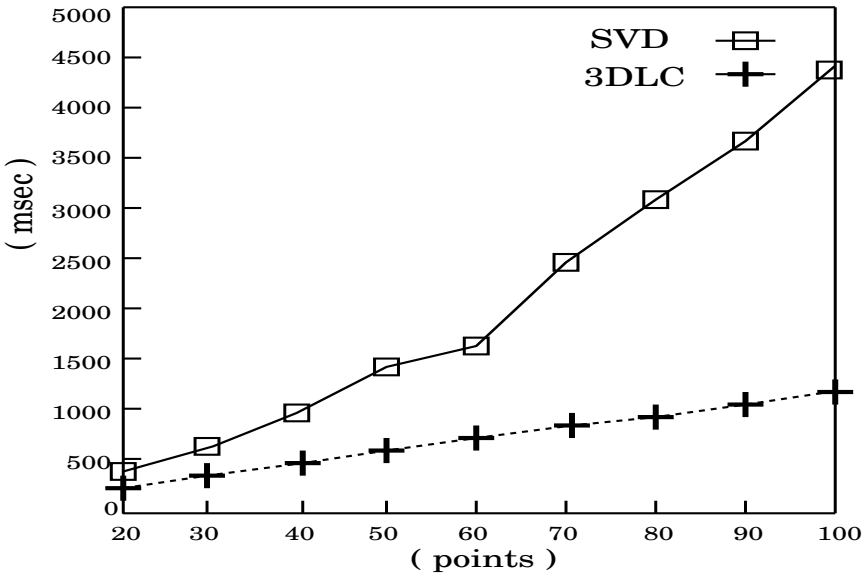


Figure 12. Comparison of computational cost.

7. An Application of the Proposed Method for Real/Synthetic Complex Scene

The factorization method has been applied on several provinces such as, 3-D shape modeling [17][19], reconstruction of terrain from an aerial image sequence [3], reconstruction of human stomach from an endoscope image sequence [7], etc. Recently, in the field of real complexness, the fusion of real and synthetic object on virtual environment has been paid much attention.

Here, this study attempts to build an application of using the proposed method for putting the recovered three dimensional shape data on virtual environment. In regard to getting 3-D shape data of real objet, contact digitizer can help us, but it is limited by location. By using factorization method, non-contact digitizer can be developed if exact data is not require. In Section 6.1, three dimensional feature points could be recovered. However, the recovered feature points are not complete model.

In order to modeling, a delaunay algorithm [14][16] is adopted to bulid a net that the points in relation to one another are linked. Then, imposing two constraints: the orientation of viewer and the texture information, ambiguous patches could be reduced and a complete model could be construct . A more detailed description of the method can be found in [17][19]. Once the reconstructed three dimensional model is built,

recovered 3D shape can be registered into virtual environment or CAD tool, and it could be manipulated, redesigned a new model, created a composed image or a complex scene (shown in Figure 14). The disposition of real/synthetic complex scene is illustrated in Figure 13.

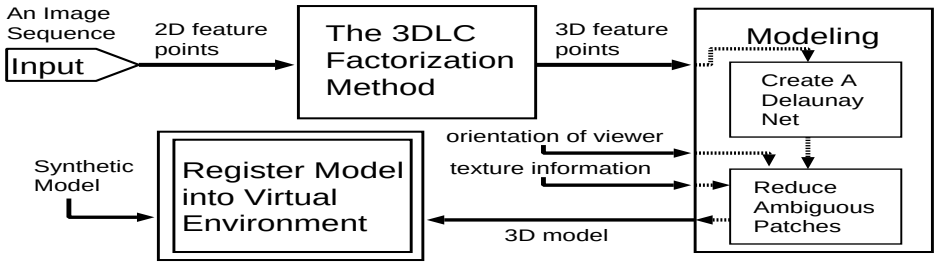


Figure 13. The disposition of real/synthetic complex scene.

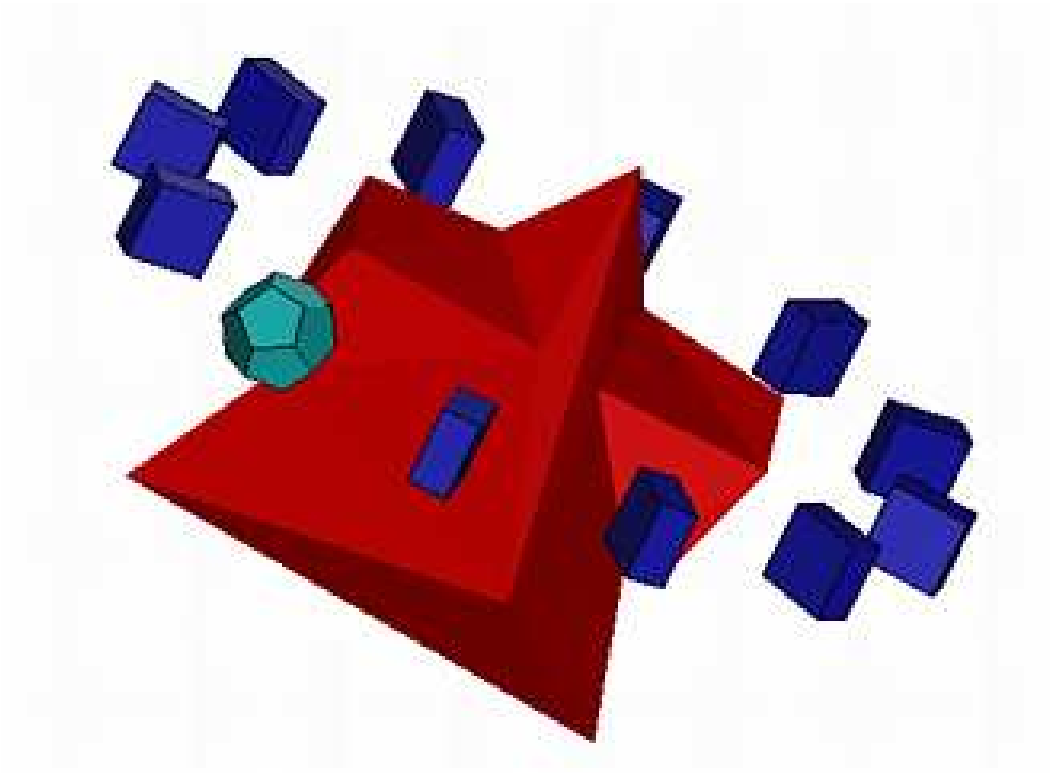


Figure 14. Recovered object and synthetic object on virtual environment.

8. Conclusion

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 recover the object shape accurately comparably to those by SVD technique, and 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. At the same time, the problem of the normalization described in Section 2.4 and 4 were solved by our method. As a consequence, the present work expands the application of the shape and motion recovery from unreliable data. On the other hand, an application was implemented for real/synthetic complex scene, and This system provide a new concept for virtual reality use.

Future works include: (1) extension of the method to other projective models such as scaled orthographic, paraperspective and perspective projections; (2) application of the method to 3D model construction of real complex scenes for real-time use.

Acknowledgements

The author wishes to thank Naokazu Yokoya, Professor, Kunihiro Chihara, Professor and Haruo Takemura, Associate Professor for their instruction, and thank goes to Kazumasa Yamazawa, Assistant and Hidehiko Iwasa, Assistant for their insightful comments. The author also wishes to thank Yuan Sheng Tang for pointing out the existence of the orthogonalization method and suggesting its applicability to factorization method. Additional thank goes to Nobuhiro Aihara, Hiromi Fukunaga and the members of Foundations of Software Laboratory for their helpful comments. Lastly, thank goes to my wife and my family for their props, and pray for repose of my mother soul.

References

[1] C. Tomasi and T. Kanade: Shape and Motion from Image Streams: a Factorization Method, *Technical Report CMU-CS-91-172*, Carnegie Mellon University, Pittsburgh, Sep. 1991.

[2] C. Tomasi and T. Kanade: The Factorization Method for the Recovery of Shape and Motion from Image Streams, *Proc. Image Understanding Workshop*, San Diego, pp.459–472, Jan. 1992.

[3] C.J. Poelman and T. Kanade: A Paraperspective Factorization Method for Shape and Motion Recovery, *Technical Report CMU-CS-93-219*, Carnegie Mellon University, Pittsburgh, Dec. 1993.

[4] T. Kanade, C.J. Poelman and T. Morita: A Factorization Method for Shape and Motion Recovery, *IEICE D-II*, Vol.J76-D-II, No.8, pp.1497–1505, 1993. (in Japanese).

[5] T. Sasano and K. Deguchi: On the Factorization Method for Shape from Image Streams, *IEICE Technical Report*, Jan. 1994, (in Japanese).

[6] J. Costeira and T. Kanade: A Multi-body Facorization Method for Motion Analysis, *Technical Report CMU-CS-TR-94-220*, Carnegie Mellon University, Pittsburgh, Sep. 1994.

[7] K. Deguchi et al.: 3-D Shape Reconstruction from Endoscpoe Image Sequences by The Factorization Method, *IEICE Trans. Inf. & Syst.*, Vol. E79-D, No.9, pp.1329–1336, Sep, 1996.

[8] B. Triggs: Factorization Methods for Projective Structure and Motion, *In Proc. Computer Vision and Pattern Recognition*, pp.845–851, 1996.

[9] K. Deguchi and B. Triggs: Factorization Method for Structure from Perspective Images, *Technical Report of IEICE*, PRMU96-139, pp.81–88, Jan. 1997, (in Japanese).

[10] K. Hwang et al.: A New Factorization Method Based on 3D Linear Combination for Shape and Motion Recovery, *The 55th General Conf. of IPSJ*,No.5AB-3, Sep. 1997, (in Japanese).

[11] C.V. Stewart, R.Y. Flatland and K. Bubna: Geometric Constraints and Stereo Disparity Computation, *Int. Journal of Computer Vision*, 20(3), pp.143–168, 1996.

[12] S. Ullman: *The Interpretation of Viusal Motion*, The MIT Press, Cambridge, MA, 1979.

[13] W.H. Press et al.: *Numerical Recipes in C*, Cambridge University Press, 1988.

[14] O.D. Faugeras: *Three-Dimensional Computer Vision*, The MIT Press, 1993.

[15] J. Shi and C. Tomasi: Good Features to Track, *IEEE CVPR*, pp.593–600, Jun, 1994.

[16] K.L. Clarkson, K.Mehlhorn, and R. Seidel: Four results on randomized incremental constructions, *Comp. Geom.: Theory and Applications*, pp.121–185, Jun, 1993.

[17] H. Saito: 3D Shape Modeling from Multiview Images, *The 10th Circuit & System Workshop*, pp.169–174, 1997, (in Japanese).

[18] H. Yamamoto, S. Uchiyama and H. Tamura: The Delaunay Triangulation for Accurate Three-Dimensional Graphic Model, *IEICE Thesis D-II*, Vol.J78-D-II, No.5, pp.745–753, 1995, (in Japanese).

[19] J. Nagai, H. Yoshida and H. Saito: 3D Model Reconstruction from Image Sequence, *Proc. 1997 IEICE General Conf.*, No.D–12–121, Mar. 1997, (in Japanese).

[20] K. Hwang, N. Yokoya, H. Takemura and K. Yamazawa: A Factorization Method Using 3-D Linear Combination for Shape and Motion Recovery, *12th ICPR Int. Conf. on Pattern Recongnition*, 1998, (application).

Appendix

A. Recovery under Scaled Orthographic Projection

Scaled orthographic projection models the scaling effect of perspective projection, but not the position effect. The scaled orthographic factorization method can be used when the object remains centered in the image, or when the distance to the object is large relative to the size of the object.

A.1 Scaled Orthographic Projection

Under scaled orthographic projection, object points are orthographically projected onto a hypothetical image plane parallel to the real image plane but passing through the object's center of mass **c**. This image is then projected onto the image plane using perspective projection (see Figure A.1).

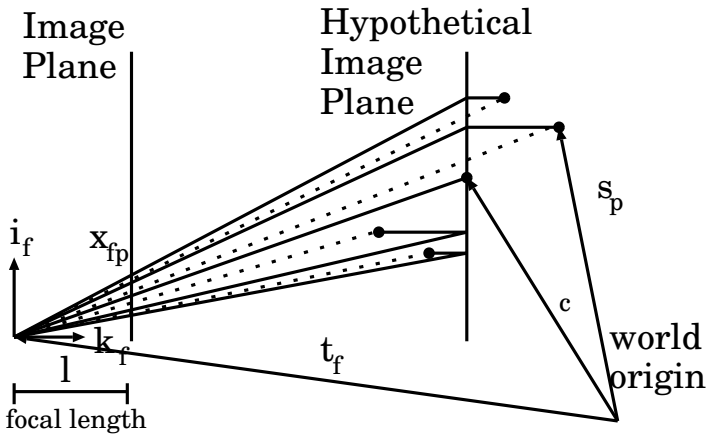


Figure A.1. Scaled orthographic projection in two dimensions.

The perspective projected points all lie on a plane parallel to the image plane so that their depth are $z_f = (\mathbf{c} - \mathbf{t}_f) \cdot \mathbf{k}_f$. Thus, the coordinate (x_{fp}, y_{fp}) is scaled

by the ratio of the focal length to the depth z_f.

$$x_{fp} = \frac{l}{z_f}(\mathbf{i}_f^T \cdot (\mathbf{s}_p - \mathbf{t}_f)), \qquad y_{fp} = \frac{l}{z_f}(\mathbf{j}_f^T \cdot (\mathbf{s}_p - \mathbf{t}_f)), \tag{48}$$

For simplicity, we assume unit focal length, $l = 1$. We fix world origin at the object's center of mass, and rewrite the above equations as:

$$x_{fp} = \mathbf{m}_f^T \cdot \mathbf{s}_p + t_{x_f}, \qquad y_{fp} = \mathbf{n}_f^T \cdot \mathbf{s}_p + t_{y_f}, \tag{49}$$

where

$$z_f = -\mathbf{t}_f^T \cdot \mathbf{k}_f, \tag{50}$$

$$t_{x_f} = -\frac{\mathbf{t}_f^T \cdot \mathbf{i}_f}{z_f}, \qquad t_{y_f} = -\frac{\mathbf{t}_f^T \cdot \mathbf{j}_f}{z_f}, \tag{51}$$

$$\mathbf{m}_f = \frac{\mathbf{i}_f}{z_f}, \qquad \mathbf{n}_f = \frac{\mathbf{j}_f}{z_f}. \tag{52}$$

A.2 Decomposition

Because Equation (49) is identical to Equation (2), the measurement matrix $\tilde{W}$ can still be written as $\tilde{W} = MS$, and using 3DLC algorithm can factor $\tilde{W}$ into the product of $\hat{M}$ and $\hat{S}$.

A.3 Normalization

For solving matrix A, we combine the two Equations (52) to impose the constraint as follow:

$$\mathbf{m}_f^T \mathbf{m}_f = \mathbf{n}_f^T \mathbf{n}_f \left(= \frac{1}{z_f^2}\right), \tag{53}$$

and because $\mathbf{m}_f$ and $\mathbf{n}_f$ are just scalar multiples of $\mathbf{i}_f$ and $\mathbf{j}_f$, we can still use the constraint that

$$\mathbf{m}_f^T \mathbf{n}_f = 0. \tag{54}$$

Once the matrix A has been found, the reconstructions of motion M and shape S are acquired by Equation (10).

B. Recovery under Paraperspective Projection

Paraperspective projection models not only the scaling effect of perspective projection, but also the position effect. As a consequence, it is a closer approximation to perspective projection than orthographic/scaled orthographic projection.

B.1 Paraperspective Projection

The paraperspective projection of an object onto an image, illustrated in Figure B.1, involves two steps: 1) An object point is projected along the direction of line connecting the focal point of the camera to the object's center of mass, onto a hypothetical image plane parallel to the real image plane and passing through the object's center of mass. 2) The point is then projected onto the real image plane using perspective projection.

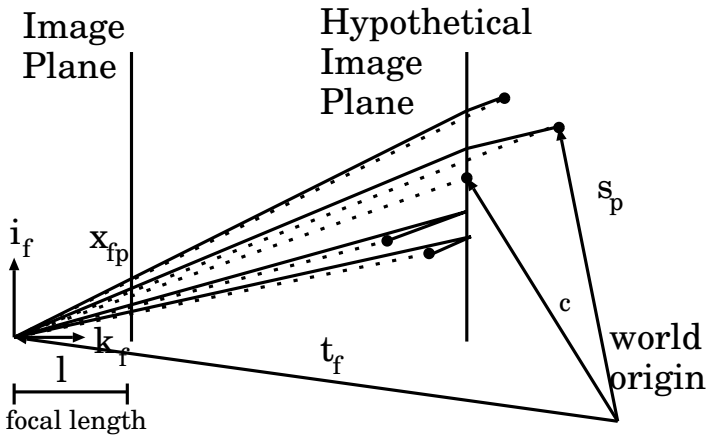


Figure B.1. Paraperspective projection in two dimensions.

Thus, the coordinate (x_{fp}, y_{fp}) can be defined as the following equations.

$$x_{fp} = \frac{l}{z_f}\{\left[\mathbf{i}_f^T + \frac{\mathbf{i}_f^T \cdot \mathbf{t}_f}{z_f}\mathbf{k}_f^T\right]\cdot\mathbf{s}_p - (\mathbf{t}_f^T \cdot \mathbf{i}_f)\}, \qquad y_{fp} = \frac{l}{z_f}\{\left[\mathbf{j}_f^T + \frac{\mathbf{j}_f^T \cdot \mathbf{t}_f}{z_f}\mathbf{k}_f^T\right]\cdot\mathbf{s}_p - (\mathbf{t}_f^T \cdot \mathbf{j}_f)\}. \tag{55}$$

For simplicity, we assume unit focal length, $l = 1$. We fix world origin at the

object's center of mass, and rewrite the above equations as:

$$x_{fp} = \mathbf{m}_f^T \cdot \mathbf{s}_p + t_{x_f}, \qquad y_{fp} = \mathbf{n}_f^T \cdot \mathbf{s}_p + t_{y_f}, \tag{56}$$

where

$$z_f = -\mathbf{t}_f^T \cdot \mathbf{k}_f, \tag{57}$$

$$t_{x_f} = -\frac{\mathbf{t}_f^T \cdot \mathbf{i}_f}{z_f}, \qquad t_{y_f} = -\frac{\mathbf{t}_f^T \cdot \mathbf{j}_f}{z_f}, \tag{58}$$

$$\mathbf{m}_f = \frac{\mathbf{i}_f - t_{x_f}\mathbf{k}_f}{z_f}, \qquad \mathbf{n}_f = \frac{\mathbf{j}_f - t_{y_f}\mathbf{k}_f}{z_f}. \tag{59}$$

B.2 Decomposition

Because Equation (56) is identical to Equation (2), the measurement matrix $\tilde{W}$ can still be written as $\tilde{W} = MS$, and using 3DLC algorithm can factor $\tilde{W}$ into the product of $\hat{M}$ and $\hat{S}$.

B.3 Normalization

For solving matrix A, from Equations (59), we impose the constraints as follow:

$$\frac{\mathbf{m}_f^T \mathbf{m}_f}{1 + t_{x_f}^2} = \frac{\mathbf{n}_f^T \mathbf{n}_f}{1 + t_{y_f}^2} = \frac{\mathbf{m}_f^T \mathbf{n}_f}{t_{x_f} t_{y_f}} \left(= \frac{1}{z_f^2} \right). \tag{60}$$

Once the matrix A has been found, the reconstructions of motion M and shape S are acquired by Equation (10).