由物體的線特徵及運用假說測試和自我推演技巧來可靠地計算物體姿勢*
Reliable Determination of Object Pose from Line Features By Hypothesis Test* and Bootstrap Techniques

張欽圳　　蔡文祥
Chin-Chun Chang and Wen-Hsiang Tsai†
交通大學資訊科學系
Department of Computer and Information Science
National Chiao Tung University,
Hsinchu, Taiwan 300, ROC

摘要

一個穩定的電腦視覺系統所採用的演算法必須要能確保其輸出品質。在這篇論文中，我們提出一個能夠自動地確保由物體的線特徵所得到的物體姿勢品質的方法。這個方法是基於統計假說理論(hypothesis test)以及使用自我推演(bootstrap)技巧。為了確保品質，我們提出兩個測試函數。其中一個函數能夠在計算物體姿勢之前就過濾掉不好的資料。另一個則是在得到物體姿勢後用來評估所求得的精確度是否到達要求。若是仍無法確定其品質，則再用自我推演技巧進行更進一步的評估。實驗結果顯示第一個測試函數能夠偵測出品質不好的輸入資料，並且此方法可用來確保所估計的物體姿勢之品質。

Abstract

To develop a reliable computer vision system, the algorithm employed in the system must ensure the quality of its output. In this paper, a reliable approach to estimating a 3D object pose from 2D-to-3D line correspondences is proposed. An algorithm is developed to integrate the abilities of automatically evalutaing the quality of input data as well as that of the estimated object pose into a single system. To ensure the quality of the estimated result, two simple test functions are defined based on the statistical hypothesis testing. The first test function can filter out poor input data priori to computing the object pose. After estimating the object pose, the second test function can decide whether the estimation result is accurate or not if it has significant evidences to make the decision; otherwise, an evaluation function is called for a further evaluation, in which algorithm, based on the bootstrap technique is designed to estimate the standard error of the estimated object pose. Experimental results show that the first test function can detect input with low qualities or erroneous line correspondences and the proposed method always yields reliable estimated results.

*This work was supported by National Science Council under Grant NSC-86-2213-E-009-114 .

†To whom all correspondences should be addressed.

I. Introduction

Determining the orientation and position of an object relative to a camera is an essential and important problem in computer vision. Since determining object poses can be regarded as an early process of a computer vision system, the quality of the estimated results must be known and ensured before using them for subsequent processes.

In general, the position and orientation of an object is estimated using the relation between the 3D structure of the object and its 2D perspective projection image. Methods to solve the pose estimation problem using line correspondences can be classified into the direct method [1, 2, 3], and the iterative method [4, 5, 6, 7, 8, 9] which takes all line correspondences into account based on the concept of minimizing the sum-of-squares error [5, 6, 7], and robust statistics like the M-estimator [4] to deal with outliers like erroneous line correspondences or observed lines corrupted by serious image noise. No matter which approach used, it is necessary to know the accuracy of estimated results before using them. Consequently, a method to know whether the estimated results are accurate or not is required.

Poor input data leads to instability of the pose estimation algorithm and a poor estimated object pose gives rise to poor performance of a computer vision system. Traditionally, to know whether the estimated result is good or not, an one-dimensional test function is usually defined [12], and then the estimated result is classified to be good if the value of the test function of the estimated result is smaller than a pre-defined threshold value; otherwise, the estimated result is classified to be bad. However, this simple method has two drawbacks. First, the estimated result cannot be known to be bad before computing the object pose even if the input data is significantly poor. Second, this method measures the error of the estimated result by mapping the original six-dimensional parameters space (three for orientation and three for position) to a one-dimensional test function. This mapping loses valuable information, leading to making wrong decision, especially when the value of the test function of the estimated result is near the threshold value. In this study, to overcome the first drawback of the traditional method, the test is decomposed into two stages: the test function H_{pre} of the first stage detects significantly poor input data from which it is almost impossible to estimate accurate object poses, and the test function H_{post} of the second stage judges the quality of the estimated result. To overcome the second drawback, a reject option [13] is introduced to the test function H_{post} so that if the estimated result is within the reject region of H_{post}, then an evaluation function will be called for a further evaluation using the standard error of the object pose estimated by the bootstrap method [14].

In the following sections, we first formulate the problem of estimating object poses from line correspondences as one of minimizing an error function; in addition, the relation between the value of the error function and the agreement between the estimated object pose and the constraint equations is analyzed by the *reduced chi-square test* [15] in Section 2. In Section 3, in order to design the test function H_{pre}, two lower bounds of the error function are derived. How to apply the bootstrap method to evaluate the accuracy of the estimated pose is also illustrated in this section. To show how effective our method is, the proposed method is tested by computer simulations and by the use of real images in Section 4. Concluding remarks are included in the last section.

II. Problem Formulation

Let $\overline{L}_i, i = 1, 2, ..., N$, be N 3D lines of an object in an object coordinate system, $o_o - x_o - y_o - z_o$. The line $\overline{L}_i$ going through a point p_i with direction d_i can be described by

$$\overline{L}_i : \lambda d_i + p_i, \tag{1}$$

where λ is a scalar. Let L_i denote line $\overline{L}_i$ in the camera coordinate system $o_c - x_c - y_c - z_c$. The transformation from the object coordinate system to the camera coordinate system can be described by a rotation matrix R and a translation vector t, which represent the orientation and the position of the object with respect to the camera, respectively,

$$p_c = Rp_o + t, \tag{2}$$

where p_o and p_c represent the coordinates of a 3D point in the object coordinate system and a camera coordinate system, respectively. According to Eqs. (1) and (2), L_i can be described by

$$L_i : \lambda R d_i + R p_i + t, \tag{3}$$

and the 2D perspective projection of L_i is an image line l_i which can be described by

$$x_c \cos\theta_i + y_c \sin\theta_i + c_i = 0 \text{ and } z_c = f, \tag{4}$$

in the camera coordinate system where f is the focal length. From the relation between the 3D model lines $\overline{L}_i$, $i = 1, 2, ..., N$, and their 2D perspective projections l_i, $i = 1, 2, ..., N$, the pose of an object with respect to the camera can be estimated [2, 5, 6, 7, 9].

A. Definition of error function

As shown in Fig. 1, L_i , l_i and the origin o_c of the camera coordination system are on a common plane π_i, called the interpretation plane [9], with a unit normal vector n_i expressed as

$$n_i = \left(1 + \left(\frac{c_i}{f}\right)^2\right)^{-\frac{1}{2}} \left[\begin{array}{ccc} \cos\theta_i & \sin\theta_i & \frac{c_i}{f} \end{array} \right]^t \tag{5}$$

in the camera coordinate system. Since π_i is orthogonal to the direction vectors of the lines lying on π_i according to Eq. (3), n_i is orthogonal to Rd_i. In addition, $Rp_i + t$ is a point on π_i. Thus, two types of constraint equations can be obtained as follows:

$$n_i^t R d_i = 0, \ i = 1, 2, ..., N, \tag{6}$$

and

$$n_i^t (R p_i + t) = 0, \ i = 1, 2, ..., N. \tag{7}$$

Eq. (6) is called an *orientation-constraint equation* and Eq. (7) a *position-constraint equation* [2]. In the least sum-of-squares error sense, we can obtain an error function $E(R, t)$ to measure the degree of agreement between the estimated pose and the observed lines as follows:

$$E(R,t) = E_O(R) + E_P(R,t), \tag{8}$$

where

$$\begin{aligned} E_O(R) &= \sum_{i=1}^{N} \left(\frac{n_i^t R d_i}{\sigma_i}\right)^2, \\ E_P(R,t) &= \sum_{i=1}^{N} \left(\frac{n_i^t (R p_i + t)}{\sigma_i'}\right)^2, \end{aligned} \tag{9}$$

σ_i and σ_i', i = 1, 2, ..., N, are used for weighing each of the constraint equations. Now, the problem can be clearly stated as follows. (1) Is it possible to obtain an accurate pose of an object with respect to the camera from Eq. (8) ? (2) Is the estimated pose good enough ? To answer the two questions, we must define some evaluation functions.

B. Relation between error function and qualities of input data and estimated result

To use the error function to test the quality of the observed image lines and the estimated pose, we must define what poor quality is. A simple and straight-forward definition of poor quality is that the relative errors coming from the estimated object pose and the observed image lines of the object exceed some pre-specified tolerable limits. In this study, the tolerable limits are defined to be the relative error of the rotation matrix δ_R, the relative error of the translation vector δ_t, and the relative error of each of the unit normal vectors of the interpretation planes $\delta_{\Delta n}$ because the estimated rotation matrix, the estimated translation vector, and the observed image lines all contribute to the value of the error function $E(R, t)$. Thus, $E(R, t)$ can act as an indicator of the agreement between the observed image lines and the estimated object pose according to the pre-specified tolerable limits.

Suppose that the measurement error of each term in Eq. (9) are all approximate normal distributions, and σ_i and σ_i' are the standard deviations of $n_i^t R d_i$ and $n_i^t (R x_i + t_i)$, respectively. To test the degree of the agreement between the observed image lines and the constraint equations, which is measured in a sum of squares form, the *chi-square test* [12, 15] is an appropriate method. Since, in $E(R,t)$ there are $2N$ measurements and six parameters to be computed from the $2N$ measurements, a function χ^2 defined by

$$\chi^2 = \frac{1}{2N-6} E(R,t) \tag{10}$$

can be regarded as a chi-square with one degree of freedom [15]. Eq. (10) is also called *the reduced chi-square* [15]. Additionally, two functions χ_O^2 and χ_P^2 for testing the degree of agreement between the observed image lines and the orientation-constraint equations and the position-constraint equations can be similarly defined as follows:

$$\begin{aligned} \chi_O^2 &= \frac{1}{N-3} E_O(R), \\ \chi_P^2 &= \frac{1}{N-6} E_P(R,t). \end{aligned}$$

In practice, however, σ_i and σ_i', $i = 1, 2, ..., N$, are unknown and only the *a priori* information available is the tolerable limits, δ_R, δ_t, and $\delta_{\Delta n}$ which are specified by users. Thus, the relation between the specified tolerable limits and the standard deviations of $n_i^t R d_i$ and $n_i^t (R x_i + t_i)$, $i = 1, 2, ...,$ N must be figured out in order to use the reduced chi-square test.

C. Determination of σ_i and σ_i', $i = 1, 2, ..., N$

By properly assuming the probability distributions of the errors of the observed lines, the estimated rotation matrix $R^{\#}$, and the estimated translation vector $t^{\#}$, we can obtain an upper bound of the variance σ_i^2 expressed only in terms of δ_R and $\delta_{\Delta n}$ as follows (the derivation is omitted):

$$\sigma_i^2 \le \frac{9\delta_R^2}{26} + \frac{\delta_{\Delta n}^2}{13},$$

as well as an upper bound of the variance $\sigma_i'^2$ expressed in terms of δ_R, δ_t, and $\delta_{\Delta n}$ in the following:

$$\sigma_i'^2 \le \frac{9\delta_R^2 \|p_i\|_2^2}{26} + \frac{\delta_{\Delta n}^2 \left(\|p_i\|_2 + \|t^*\|_2\right)^2}{13} + \frac{\delta_t^2 \|t^*\|_2^2}{13}, \tag{11}$$

where $\|\cdot\|_2$ represents the 2-norm of a vector. Since $\|t^*\|_2^2$ is unknown, we may replace $\|t^*\|_2^2$ by its upper bound $\max \|t^*\|_2^2$, the longest distance between o_c and o_o, and obtain

$$\sigma_i'^2 \le \frac{9\delta_R^2 \|p_i\|_2^2}{26} + \frac{\delta_{\Delta n}^2 \left(\|p_i\|_2 + \max\|t^*\|_2\right)^2}{13} + \frac{\delta_t^2 \left(\max\|t^*\|_2^2\right)}{13}. \tag{12}$$

However, in the right-hand side of Eq. (11), the numerator of the third term represents the tolerable limit for the absolute translation error; thus, if the tolerable limit of the translation error is specified by the absolute translation error instead of the relative translation error, then we will have

$$\sigma_i'^2 \le \frac{9\delta_R^2 \|p_i\|_2^2}{26} + \frac{\delta_{\Delta n}^2 \left(\|p_i\|_2 + \max\|t^*\|_2\right)^2}{13} + \frac{\delta_t'^2}{13} \tag{13}$$

where $\delta_t'^2$ represents the tolerable limit for the absolute translation error. That is, if the tolerable limit for the error of the translation vector is specified by the absolute translation error, then Eq. (13) instead of Eq. (12) should be used.

For convenience, the pre-specified tolerable limits can be represented by $(\delta_R, \delta_t, \delta_{\Delta n})$. Representing the desired quality by the three terms has at least two advantages over the method which can only specify the quality of the observed image lines (i.e. $(0, 0, \delta_{\Delta n})$). The first advantage is flexibility because we can consider the value of the error function coming from the errors of the three sources: the estimated rotation matrix, the estimated translation vector, and the observed image lines, individually. The second advantage is intuition because we can directly specify the desired quality of the estimation result.

III. Evaluation of Input Quality and Accuracy of Estimated Object Pose

A. Lower bounds of error function

Writing the error function $E(R,t)$ in a matrix form, we can obtain

$$E(R,t) = \|Ar\|_2^2 + \|Br + Ct\|_2^2,$$

where $\|\cdot\|_2$ represents the 2-norm of a vector and

$$A = \begin{bmatrix} \frac{n_{11} d_1^t}{\sigma_1} & \frac{n_{12} d_1^t}{\sigma_1} & \frac{n_{13} d_1^t}{\sigma_1} \\ \vdots & \vdots & \vdots \\ \frac{n_{(N-1)1} d_{N-1}^t}{\sigma_{N-1}} & \frac{n_{(N-1)2} d_{N-1}^t}{\sigma_{N-1}} & \frac{n_{(N-1)3} d_{N-1}^t}{\sigma_{N-1}} \\ \frac{n_{N1} d_N^t}{\sigma_N} & \frac{n_{N2} d_N^t}{\sigma_N} & \frac{n_{N3} d_N^t}{\sigma_N} \end{bmatrix} \tag{14}$$

$$B = \begin{bmatrix} \frac{n_{11} p_1^t}{\sigma_1'} & \frac{n_{12} p_1^t}{\sigma_1'} & \frac{n_{13} p_1^t}{\sigma_1'} \\ \vdots & \vdots & \vdots \\ \frac{n_{(N-1)1} p_{N-1}^t}{\sigma_{N-1}'} & \frac{n_{(N-1)2} p_{N-1}^t}{\sigma_{N-1}'} & \frac{n_{(N-1)3} p_{N-1}^t}{\sigma_{N-1}'} \\ \frac{n_{N1} p_N^t}{\sigma_N'} & \frac{n_{N2} p_N^t}{\sigma_N'} & \frac{n_{N3} p_N^t}{\sigma_N'} \end{bmatrix} \tag{15}$$

$$C = \begin{bmatrix} \frac{n_1}{\sigma_1'} & \cdots & \frac{n_{N-1}}{\sigma_{N-1}'} & \frac{n_N}{\sigma_N'} \end{bmatrix}^t \tag{16}$$

$$r = \begin{bmatrix} R_{11} & R_{12} & R_{13} & R_{21} & R_{22} & R_{23} & R_{31} & R_{32} & R_{33} \end{bmatrix}^t.$$

In this study, the vector r is called the nine-dimensional vector associated with the matrix R. Taking a partial derivative of E with respective to t and setting the derivatives to zero, we obtain

$$t = -C^+ B r \tag{17}$$

where $C^+ = (C^t C)^{-1} C^t$. Consequently, for any rotation matrix R and translation vector t, we have that

$$E(R,t) \ge E\left(R, -C^+ B r\right).$$

Arranging $E(R, -C^{+}Br)$ into a compact form, we obtain a lower bound of the error function $E(R,t)$, $E(R,t) \geq r^t F r$ where

$$F = A^t A + B^t \left(I - CC^{+}\right) B \tag{18}$$

with I being the $N \times N$ identity matrix. It is easy to check that F is a positive semi-definite matrix. For convenience for subsequent discussions, let $\Xi(R,F) = r^t F r$.

Let $\alpha_1, \alpha_2, ..., \alpha_9$ be the unit eigenvectors of F and $\lambda_1, \lambda_2, ..., \lambda_9$ be the corresponding eigenvalues, respectively, where $\lambda_9 \geq \lambda_8 \geq ... \geq \lambda_1 \geq 0$ because F is a positive semi-definite matrix. From the Rayleigh-Ritz theorem [17], we can simply obtain a lower bound of $\Xi(R,F)$: $\Xi(R,F) \geq 3\lambda_1$; however, this lower bound is too loose to use in practice.

Let the nine-dimensional vector of M_i be α_i, K_i be the rotation matrix closest to M_i in the Frobenius matrix norm, k_i be the nine-dimensional vectors associated with K_i, $i = 1, 2, ..., 9$. According to [16], the rotation matrices K_i can be computed using singular value decomposition $K_i = U_i \text{diag}(1, 1, \det(U_i V_i)) V_i^t$, where $U_i S_i V_i^t$ is the singular value decompositions of M_i, respectively, with S_i being a diagonal matrix $\text{diag}(s1_i, s2_i, s3_i)$ where $s1_i \geq s2_i \geq s3_i \geq 0$. We can obtain a lower bound LB_1 of $\Xi(R,F)$ not smaller than $3\lambda_1$ in the following (the derivation is omitted):

$$LB_1 = \text{trace}(S_1)^2 \lambda_1 + \min\left\{3 - \text{trace}(S_1)^2, \text{trace}(S_2)^2\right\} \lambda_2 +$$

$$\max\left\{3 - \text{trace}(S_1)^2 - \text{trace}(S_2)^2, 0\right\} \lambda_3.$$

If the rank of the matrix F^* is eight, then the nine-dimensional vector associated with R^* is the eigenvector of F^* corresponding to the zero eigenvalue. By using the perturbation theory of eigenvalues and eigenvectors [13, 16], we can obtain an approximate lower bound LB_2 of $\Xi(R,F)$ as follows (the derivation is omitted):

$$LB_2 = 3\lambda_1 + \left(6 - 2\sqrt{3}\,\text{trace}\left(S_1\right)\right) \lambda_2.$$

Because $E(R,t)$ is not smaller than $\Xi(R,F)$, the largest one of LB_2 and LB_3 denoted as LB can be used as a lower bound of $E(R,t)$. Since these lower bounds are derived from the eigenvalues and eigenvectors of F, and the number of line correspondences is at least eight to make the rank of F to be eight. Thus, to use these lower bounds, the number of line correspondences must be larger than seven.

B. *Definitions of test functions: H_{pre} and H_{post}*

In order to distinguish the error function defined at different tolerable limits, the functions $E(R,t)$, $\Xi(R,F)$, χ^2, χ_O^2, and χ_P^2 are denoted by $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$, $\Xi_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,F)$, $\chi^2_{(\delta_R,\delta_t,\delta_{\Delta n})}$, $\chi^2_{O(\delta_R,\delta_t,\delta_{\Delta n})}$, and $\chi^2_{P(\delta_R,\delta_t,\delta_{\Delta n})}$. Because LB is a lower bound of $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$ which can be obtained before to estimate the object pose, we can foresee the ultimate quality of the estimated result via this lower bound. Thus, the function H_{pre} proposed in this study for testing the quality of the input data is defined as follows:

$$H_{pre} = \begin{cases} \text{unacceptable} & \text{if } \frac{LB}{2N-6} \geq v_1; \\ \text{acceptable} & \text{otherwise,} \end{cases} \tag{19}$$

where v_1 is a pre-defined threshold value. The function H_{post} proposed for testing the agreement between the estimated result and the constraint equations can be defined as follows:

$$H_{post} = \begin{cases} \text{acceptable} & \text{if } \chi^2_{O(\delta_R,0,0)}, \chi^2_{P(\delta_R,\delta_t,0)} \leq v_2 \\ & \text{and } \chi^2_{(\delta_R,\delta_t,\delta_{\Delta n})} < v_1 \\ \text{unacceptable} & \text{if } \chi^2_{(\delta_R,\delta_t,\delta_{\Delta n})} \geq v_1 \text{ ;} \\ \text{uncertain} & \text{otherwise,} \end{cases} \tag{20}$$

where $\chi^2_{(\delta_R,\delta_t,\delta_{\Delta n})}$, $\chi^2_{O(\delta_R,0,0)}$ and $\chi^2_{P(\delta_R,\delta_t,0)}$ can be computed after computing the best rotation matrix and translation vector by minimizing $E_{(0,0,\delta_{\Delta n})}(R,t)$ and v_2 is a predefined threshold value.

C. *Error estimation by the bootstrap method*

Using the test functions mentioned in the previous section, we can classify the quality of the estimated result to be good or bad if we have significant evidences. However, if we have no enough information to make that decision, then we must find other evidences to judge whether or not the quality of the output reaches the desired accuracy. In this study, to estimate the error of the estimated object pose, an algorithm based on the bootstrap method are proposed.

Let $n_i^{\#}$, $i = 1, 2, ..., N$, be the unit normal vectors of the interpretation planes when the orientation and position of the object are $R^{\#}$ and $t^{\#}$. Suppose that the errors between n_i, $i = 1, 2, ..., N$, and their corresponding noise-free versions, n_i^*, $i = 1, 2, ..., N$, as well as the errors (called residuals) $\Delta n_i^{\#}$ between n_i and $n_i^{\#}$, $i = 1, 2, ..., N$, are identically independent random variables. In addition, the shape of the probability distribution of the error between n_i and $n_i^{\#}$ is assumed to be similar to that of the error between n_i and n_i^*. Thus, by repeatedly computing the amount of the pose perturbed away from the estimated object pose when $n_i^{\#}$, $i = 1, 2, ..., N$, are corrupted by $\Delta n_i^{\#}$, $i = 1, 2, ..., N$, the bootstrap standard errors (BSE) with respect to $R^{\#}$ and $t^{\#}$ can be estimated. Specifically, to know the difference between the ith bootstrap replication $(R_i^{\$}, t_i^{\$})$ and $(R^{\#}, t^{\#})$, $R_i^{\$}$ can be defined to be equal to $R^{\#}$ multiplied by a rotation matrix R such that $\Xi_{(0,0,\delta_{\Delta n})}(R^{\#}R, F_i^{\$})$ is minimized where the matrix $F_i^{\$}$ is generated by Eq. (18) using the ith bootstrap sample. From Euler's theorem [16], we know that a rotation matrix R can be regarded as rotating around a unit vector $l = [\ l_1 \quad l_2 \quad l_3\ ]^t$ by some angle ψ, i.e., the Euler's representation [16]; in addition, R can be expanded as a Taylor series with respect to ψ at zero degree to the first-order term as follows:

$$R \cong I + \begin{bmatrix} 0 & -l_3 & l_2 \\ l_3 & 0 & -l_1 \\ -l_2 & l_1 & 0 \end{bmatrix} \psi. \tag{21}$$

Thus, $\Xi_{(0,0,\delta_{\Delta n})}(R^{\#}R, F_i^{\$})$ becomes

$$\Xi_{(0,0,\delta_{\Delta n})}(R^{\#}R, F_i^{\$}) = (Hl' + r^{\#})^t F_i^{\$} (Hl' + r^{\#}), \quad (22)$$

where l' denotes ψl and

$$H = \begin{bmatrix} 0 & R_{13}^{\#} & -R_{12}^{\#} & 0 & R_{23}^{\#} & -R_{22}^{\#} & 0 & R_{33}^{\#} & -R_{32}^{\#} \\ -R_{31}^{\#} & 0 & R_{11}^{\#} & -R_{23}^{\#} & 0 & R_{21}^{\#} & -R_{33}^{\#} & 0 & R_{31}^{\#} \\ R_{12}^{\#} & -R_{11}^{\#} & 0 & R_{22}^{\#} & -R_{21}^{\#} & 0 & R_{32}^{\#} & -R_{31}^{\#} & 0 \end{bmatrix}$$

By minimizing $\Xi(R^{\#}R, F_i^{\$})$, we can obtain that

$$l' = -(H^t F_i^{\$} H)^+ H^t F_i^{\$} r^{\#}, \quad (23)$$

and the squares error between the ith bootstrap replication and the estimated object pose can be computed as follows:

$$\begin{aligned} \|R_i^{\$} - R^{\#}\|_F^2 &= 2\left\|l'\right\|_2^2, \\ \|t_i^{\$} - t^{\#}\|_2^2 &= \left\|-C_i^{\$+} B_i^{\$}\left(r^{\#} + l''\right) - t^{\#}\right\|_2^2, \end{aligned} \quad (24)$$

where $l'' = [\ 0 \quad -l_3' \quad l_2' \quad l_3' \quad 0 \quad -l_1' \quad -l_2' \quad l_1' \quad 0\ ]^t$. The steps of the algorithm are described as follows.

Algorithm 1. Using the bootstrap method to estimate standard errors.

Input. Observed image lines, n_i represented by $[\ \cos\theta_i \sin\phi_i \quad \sin\theta_i \sin\phi_i \quad \cos\phi_i\ ]^t$, $i = 1, 2, ..., N$, model lines L_i, $i = 1, 2, ..., N$, and $R^{\#}$ and $t^{\#}$.

Output. Bootstrap standard errors for the rotation matrix and the translation vector.

Step 1. Generate $n_i^{\#}$ which can be represented by $[\ \cos\theta_i^{\#} \sin\phi_i^{\#} \quad \sin\theta_i^{\#} \sin\phi_i^{\#} \quad \cos\phi_i^{\#}\ ]^t$, $i = 1, 2, ..., N$, by transforming the 3D model lines with the rotation matrix $R^{\#}$ and the translation vector $t^{\#}$ and project them onto the image plane. Each of $n_i^{\#}$, $i = 1, 2, ..., N$, is adjusted to make the inner product of n_i and $n_i^{\#}$ to be nonnegative because it is reasonable to assume that the error between the two vectors is not larger than ninety degrees. Thus, the residuals, $\Delta n_i^{\#} = [\ \Delta\theta_i^{\#} \quad \Delta\phi_i^{\#}\ ]$, $i = 1, 2, ..., N$, can be produced by $\Delta\theta_i^{\#} = \theta_i - \theta_i^{\#}$, $\Delta\phi_i^{\#} = \phi_i - \phi_i^{\#}$, $i = 1, 2, ..., N$.

Step 2. Perform Step 2.1 to Step 2.3 w times to generate w square error terms: $(u_1^{\$}, v_1^{\$})$, $(u_2^{\$}, v_2^{\$})$, ...,$(u_w^{\$}, v_w^{\$})$.

Step 2.1. Randomly produce a sample of bootstrap error terms, $(\Delta n_{p(1)}^{\#}, \Delta n_{p(2)}^{\#}, ..., \Delta n_{p(N)}^{\#})$ which is a permutation of $\Delta n_1^{\#}$, $\Delta n_2^{\#}$, ..., $\Delta n_N^{\#}$ to generate the ith bootstrap sample: $(n_1^{\#} \oplus \Delta n_{p(1)}^{\#},\ n_2^{\#} \oplus \Delta n_{p(2)}^{\#},\ ...,\ n_N^{\#} \oplus \Delta n_{p(N)}^{\#})$ where $n_j^{\#} \oplus \Delta n_{p(j)}^{\#}$ denotes the vector

$$\begin{bmatrix} \cos\left(\theta_j^{\#} + \Delta\theta_{p(j)}^{\#}\right) \sin\left(\phi_j^{\#} + \Delta\phi_{p(j)}^{\#}\right) \\ \sin\left(\theta_j^{\#} + \Delta\theta_{p(j)}^{\#}\right) \sin\left(\phi_j^{\#} + \Delta\phi_{p(j)}^{\#}\right) \\ \cos\left(\phi_j^{\#} + \Delta\phi_{p(j)}^{\#}\right) \end{bmatrix}.$$

Step 2.2. Use the ith bootstrap sample to generate the matrix $F_i^{\$}$ by Eq. (18), and the other necessary matrices $A_i^{\$}$, $B_i^{\$}$, and $C_i^{\$}$ based on the tolerable limits $(0, 0, \delta_{\Delta n})$.

Step 2.3. Compute the square error term: $(u_i^{\$}, v_i^{\$})$ using Eq. (24).

Step 3. Compute bootstrap standard error ξ_R for the rotation matrix and that ξ_t for the translation vector as follows:

$$\xi_R = \left(\frac{1}{w}\sum_{i=1}^{w} u_i^{\$}\right)^{\frac{1}{2}}, \xi_t = \left(\frac{1}{w}\sum_{i=1}^{w} v_i^{\$}\right)^{\frac{1}{2}}.$$

Step 4. Stop.

Because the noise of the input data may be caused by erroneous line correspondences, in order to balance the occurrence of each residual errors, a sample of the bootstrap error terms is generated by a permutation of the residual errors in Step 2.1. This scheme is called the *permutation bootstrap* [14]. Algorithm 1 is automatic and easy to implement; however, it needs to perform several times of Monte Carlo simulations. In this study, we find 20 bootstrap replications are enough to estimate the bootstrap standard errors.

IV. Experimental Results

A. Computer simulation

The model lines are randomly generated within a 50 cm^3box. The x-component and the y-component of a translation vector are uniformly generated from the range [-50 cm , 50 cm], and the z-component is uniformly generated from the range [50 cm, 150 cm]. The generation of a rotation matrix is subject to no restriction. The focal length is 0.8 cm. To analyze the noise effect, n_i is perturbed by a noise vector Δn , $n_i' = n_i + \rho\Delta n$, where n_i' denotes the vector of n_i after adding noise, ρ is a scalar which controls the noise level, and Δn is a noise vector of which the elements are uniformly generated from the range [-1, 1].

In order to examine the goodness of the lower bound LB resulting from various numbers of line correspondences and noise levels, we do some experiments when noise levels are: 0.005, 0.01, 0.025, and 0.05, and the numbers of line correspondences are 8, 10, 12, and 14. To know the performance of our method under different tolerable limits, four sets of tolerable limits are designed, two of them numbered 1 and 2 being for the case that the tolerable limit of the estimated translation vector is specified by the relative translation error, and the others numbered 1′ and 2′ for the case that the tolerable limit of the estimated translation vector is specified by the absolute translation error. 32,000 random trials are executed (2000 trials for each combination of the number of line correspondences, the sets of tolerable limits and the noise levels). The minimum of $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$ is computed by using the actual rotation matrix R^* as the initial guess. From Fig. 2, we can see that the correlation coefficients between LB and the minimum of $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$ are greater than 0.6. This means that LB is highly linearly correlative to the minimum of $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$. In addition,the relative error between LB and the minimum of $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$ which is defined as $\frac{\left|\text{minimum of } E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t) - LB\right|}{\text{minimum of } E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)}$, is also shown in Fig. 2. It reveals that LB is at least half of the minimum of $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$, in average. Fig. 3 (a) shows the curve of the operating characteristics [13] of only using $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$ to evaluate the goodness of estimated results; Fig. 3 (b) shows the result of using H_{pose} with at most ten percent rejection rate. Obviously, a significant improvement is obtained. The numbers of line correspondences of Fig. 3 and Fig. 4 are eight.

B. Real images

Fig. 4 (a) shows the image of a magic cube which is used to test the performance of the proposed method. On the cube, twenty one straight lines shown in Fig. 4 (b) are detected. The pose of the cube which is computed by using the twenty one straight lines is regarded as the ground true solution. By randomly choosing eight lines from the twenty one lines, a test sample can be formed. 500 test samples are generated. The quality of the estimated pose is specified by the tolerable limit sets 1, 2, $1'$, and $2'$.

In this experiment, $v_1 = 3$, and $v_2 = 0.5$. In addition, when the estimated result is within the reject region of H_{post}, the BSEs are compared to the tolerable limits. If the BSEs are smaller than the tolerable limits, then the estimated result will be regarded to be good; otherwise, the estimated result will be regarded to be bad. Table 2 shows the experimental result. Moreover, Table 2 also shows a similar experiment except that each of test samples contains one erroneous line correspondences. In the case of normal data, the experimental results associated with the tolerable limit set 1 and $1'$ are not good because the relative errors of the estimation results are around the tolerable limits. That is, the quality of the estimation results is hard to discriminate in this case. However, the experimental results associated with the tolerable limit set 2 and $2'$ are satisfiable. In the case of containing one erroneous line correspondence, the test function H_{pre} has a good performance especially when the tolerable limit sets are 1 and $1'$ because the quality of the test samples is poor with respect to the tolerable limits.

V. Conclusions

In this paper, we propose a method to ensure the quality of estimated object poses. To estimate an object pose, an error function is defined based on the least-squares sense; in addition, the relation between the error function and the tolerable limits for the qualities of the input data and the estimated result is discussed. By using the derived lower bounds of the error function, a test function H_{pre} was defined to eliminate poor input data so that unnecessary computation can be avoided. After estimating the object pose, another test function H_{post} is defined to eliminate inaccurate estimated results. Since the two test functions H_{pre} and H_{post} are very simple, they need little computation load and can be easily to embedded in an existing algorithm to eliminate poor input data or estimated results. To know the standard error of the estimated result, an algorithm based on the bootstrap method is introduced. Because the distribution of the noise of input data is hard to know in advance, especially when the input data is corrupted by erroneous line correspondences, the bootstrap method is found to be very suitable for this application.

References

[1] M. Dhome, M. Richetin, J. T. Lapresté, and G. Rives, "Determination of the attitude of 3-D objects from a single perspective view," *IEEE Trans. on Patt. Anal. and Machine Intell.*, vol. 11, no. 12, pp. 1265–1278, 1989.

[2] S. Y. Chen and W. H. Tsai, "A systematic approach to analytic determination of camera parameters by line features," *Pattern Recognition*, vol. 23, no. 8, pp. 859–877, 1990.

[3] R. M. Haralick, C. N. Lee, K. Ottenberg, and M. N. Nölle, "Review and analysis of solutions of the three point perspective pose estimation problem," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 331–356, 1994.

[4] R. M. Haralick, H. Joo, C. N. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim, "Pose estimation from corresponding point data," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1426–1445, 1989.

[5] Y. Liu, T. S. Huang, and O. D. Faugeras, "Determination of camera location from 2-D to 3-D line and point correspondences," *IEEE Trans. on Patt. Anal. and Machine Intell.*, vol. 12, no. 1, pp. 28–37, 1990

[6] R. Kumar and A. R. Hanson, "Robust methods for estimating pose and a senitivity analysis," *CVGIP: Image Understanding*, vol. 60, no. 3, pp. 313–342, 1994.

[7] T. Q. Phong, R. Horaud, A. Yassine, and P. D. Tao, "Object pose from 2-D to 3-D point and line correspondences," *International Journal of Computer Vision*, vol. 15, pp. 225–243, 1995.

[8] D. F. Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code," *International Journal of Computer Vision*, vol. 15, pp. 123–141, 1995.

[9] C. N. Lee and R. M. Haralick, "Statistical estimation for exterior orientation from line-to-line correspondences," *Image and Vision Computing*, vol. 14, pp. 379–388, 1996.

[10] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[11] T. S. Huang and A. N. Netravali, "Motion and structure from feature correspondences: a review," *Proceedings of the IEEE*, vol. 82, no. 2, pp. 252–268, 1994.

[12] B. K. P. Horn, "Relative orientation," *International Journal of Computer Vision*, vol. 4, pp. 59–78, 1990.

[13] K. Fukunaga, *Introduction to Statistical Oattern Recognition*, Academic Press, Inc., Boston, Second edition, 1990.

[14] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, New York, 1993.

[15] J. R. Taylor, *An Introduction to Error Analysis*, Oxford University Press, Mill Valley, 1982.

[16] K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, New York, 1993.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1985.

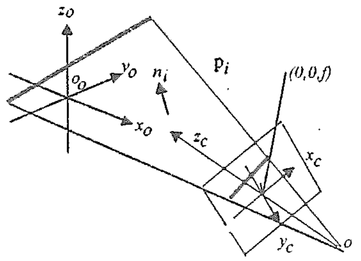
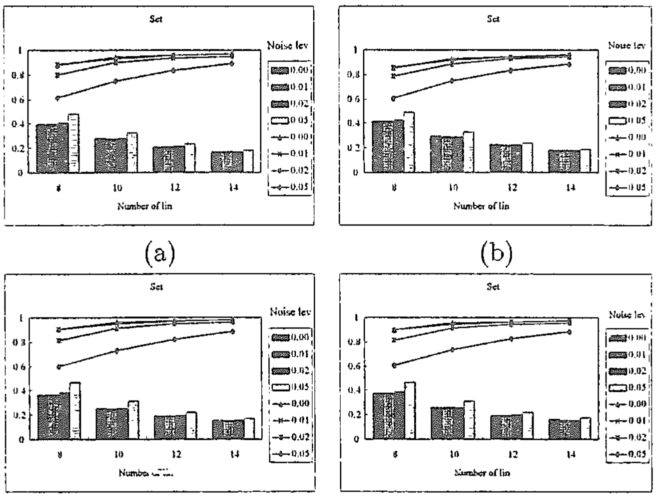


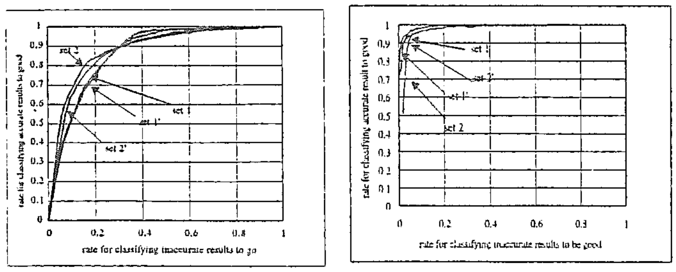
Figure 1 – An illustration of the relation between the object coordinate system and the camera coordinate system.



(a) (b)

(c) (d)

Figure 2 – Illustrations of the relation between LB and the minimum of $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$ versus various numbers of line correspondences, tolerable limits, and noise levels where the line graphs and the bar graphs represent, respectively, the average correlation coefficient and the average relative difference between LB and the minimum of $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$ versus various numbers of line correspondences and noise levels: (a) is for the tolerable limit set 1; (b) is for the set 2; (c) is for the set 1$'$; (d) is for the set 2$'$.

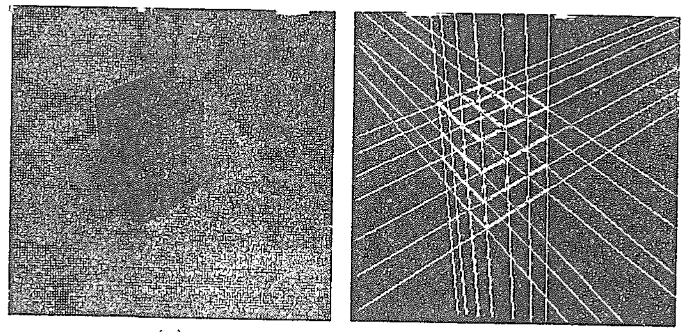


(a) (b)

Figure 3 – Illustrations of the operating characteristics of the rate of identifying accurate results to be good versus the rate of identifying inaccurate results to be good: (a) is for that only using $E_{(\delta_R,\delta_t,\delta_{\Delta n})}(R,t)$; (b) is for that using H_{post} with ten percent rejection rate.

Table 1. A list of tolerable limits.

set no.	δ_R	δ_t	$\delta_{\Delta n}$	set no.	δ_R	δ_t' (cm)	$\delta_{\Delta n}$
1	0.01	0.01	0.01	1$'$	0.01	1	0.01
2	0.025	0.03	0.025	2$'$	0.025	3	0.025



(a) (b)

Figure 4 – A real image used in this study: (a) is an image of a magic cubic; (b) is the result after performing straight line detection.

Table 2. Results of performance evaluation of the proposed method.

	normal data				one erroneous line correspondence			
tolerable limit set no.	1	2	1$'$	2$'$	1	2	1$'$	2$'$
number of good estimation results classified by H_{post} to be good	0	180	0	182	0	3	0	19
number of poor estimation results classified by H_{post} to be good	1	6	1	4	0	38	0	11
number of poor estimation results classified by H_{post} to be bad	73	0	69	0	166	226	158	241
number of good estimation results classified by H_{post} to be bad	29	0	36	0	0	0	1	0
the number of estimation results classified by BSE correctly	213	331	209	331	160	164	132	207
the number of estimation results classified by BSE incorrectly	183	3	184	3	0	54	14	22
number of poor test samples detected by H_{pre}	0	0	0	0	174	15	195	16
number of test samples which make H_{pre} to commit error	1	0	1	0	0	0	0	0
number of poor test samples	325	58	272	56	500	495	494	332
total test samples	500	500	500	500	500	500	500	500