Reliable Determination of Object Pose from Line Features by Hypothesis Testing

Chin-Chun Chang and Wen-Hsiang Tsai, *Senior Member, IEEE*

Abstract—To develop a reliable computer vision system, the employed algorithm must guarantee good output quality. In this study, to ensure the quality of the pose estimated from line features, two simple test functions based on statistical hypothesis testing are defined. First, an error function based on the relation between the line features and some quality thresholds is defined. By using the first test function defined by a lower bound of the error function, poor input can be detected before estimating the pose. After pose estimation, the second test function can be used to decide if the estimated result is sufficiently accurate. Experimental results show that the first test function can detect input with low qualities or erroneous line correspondences and that the overall proposed method yields reliable estimated results.

Index Terms—3D-to-2D, line features, object poses, hypothesis testing, reject option, reliable estimated poses.

1 Introduction

Estimating the pose of an object, called the pose estimation problem or the 2D-to-3D problem [1], [2], is an important problem in computer vision. Related applications are broad, such as camera calibration and robot self-positioning. To act as an early process of a computer vision system, the quality of the estimated poses must be ensured for subsequent processes.

Generally, the pose of an object can be estimated from the relation between its 3D structure and perspective projection. Points and lines are the most popular features for their simplicity. Since two points form a line, the method using line features may be applied to the pose estimation problem with point features [3]. Thus, only line features are considered in this study. Methods to estimate the object pose using line features can be found in [4], [5], [3], [6], [7], [8]. No matter which method is used, it needs to know the quality of the estimated poses.

Usually, a 1D test function is designed to test the quality of the estimated result. Compared with other evaluation methods, such as the bootstrap method [9] and the error propagation method [10], this approach is simpler. Furthermore, if a lower bound of the test function is known, the quality of the estimated pose can be foreseen. However, this approach maps the original parameter space to a 1D test function and this mapping may lose valuable information and lead to a wrong decision.

In this study, two test functions are proposed to test the qualities of input data and estimated poses with respect to some specified quality thresholds. They are based on an error function defined by the relation between the line features of an object and the quality thresholds. The first one is defined by a lower bound of the error function; therefore, it can detect poor input. After estimating the pose, the second one can be used to qualify the estimated pose as acceptable or unacceptable. To avoid making crisp decision, it has a reject option [11] to label as "unreliable" those estimated poses which are hard to qualify.

The paper is organized as follows: In Section 2, an error function is defined and analyzed. In Section 3, a lower bound of the error function is derived and the proposed test functions are defined. In Section 4, the proposed method is tested by synthesized and real images; in addition, an example of applying the test function to the pose estimation problem with point features is included. Concluding remarks are in the last section.

2 Definition Of Error Function

2.1 Geometric Relation and Definition of Error Function

Let $\mathbf{L}_i, i = 1, 2, \cdots, N$, be 3D model lines of an object in an object coordinate system (OCS). The line $\mathbf{L}_i$ going through a point $\mathbf{p}_i$ with direction $\mathbf{d}_i$ can be represented by $\lambda\mathbf{d}_i + \mathbf{p}_i$, where λ is a scalar. The transformation from the OCS to a camera coordinate system (CCS) can be described by $\mathbf{p}_c = \mathbf{R}\mathbf{p}_o + \mathbf{t}$, where the rotation matrix $\mathbf{R}$ and the translation vector t specify the orientation and position of the object relative to a camera with six parameters to be estimated, and $\mathbf{p}_o$ and $\mathbf{p}_c$ are the coordinates of a 3D point in the OCS and CCS, respectively. Hence, $\mathbf{L}_i$ in the CCS, is $\lambda\mathbf{R}\mathbf{d}_i + \mathbf{R}\mathbf{p}_i + \mathrm{t}$. The perspective projection of $\mathbf{L}_i$ is an image line $\mathbf{l}_i$ described by $x_c \cos\theta_i + y_c \sin\theta_i + c_i = 0$ and $z_c = f$ in the CCS where f is the focal length of the camera. As shown in Fig. 1, $\mathbf{L}_i$, l_i, and the origin of the CCS are on a common plane π_i with a unit normal vector $\mathbf{n}_i = (1 + (\frac{c_i}{f})^2)^{-\frac{1}{2}}[\cos\theta_i \quad \sin\theta_i \quad \frac{c_i}{f}]^t$ in the CCS. Since $\mathbf{n}_i$ is orthogonal to the direction of the line lying on π_i and since $\mathbf{R}\mathbf{p}_i + \mathbf{t}$ is a point on π_i, we have two constraint equations $\mathbf{n}_i^t\mathbf{R}\mathbf{d}_i = 0$ and $\mathbf{n}_i^t(\mathbf{R}\mathbf{p}_i + \mathbf{t}) = 0$, called the orientation and position constraint equations, respectively. To measure the consistency of these constraints, we define an error function $E(\mathbf{R}, \mathbf{t})$ in the sum of squares error sense as follows:

$$E(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N}\left(\frac{\mathbf{n}_i^t\mathbf{R}\mathbf{d}_i}{\sigma_i}\right)^2 + \sum_{i=1}^{N}\left(\frac{\mathbf{n}_i^t(\mathbf{R}\mathbf{p}_i + \mathbf{t})}{\sigma_i'}\right)^2,$$

where σ_i and σ_i', $i = 1, 2, \cdots, N$, are some scalars used for weighting the constraint equations.

2.2 Relation among Error Function and Qualities of Input Data and Estimated Result

In general, if the errors of the estimated pose and the observed unit normal vectors are not greater than some prespecified thresholds, the qualities of them can be considered good. This consideration permits the error function to have some uncertainty around zero. In this study, three quality thresholds δ_{t}, $\delta_{\mathbf{R}}$, and $\delta_{\Delta\mathrm{n}}$ are defined to specify the allowed absolute error between the estimated and the actual translation vectors, the allowed relative error for the estimated rotation matrix, and that for the observed unit normal

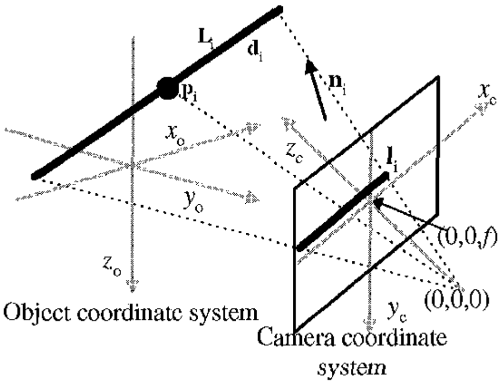


Fig. 1. Geometric relation between the camera and the object.

- *The authors are with the Department of Computer and Information Science, National Chiao Tung University, Hsinchu, Taiwan 300, Republic of China. E-mail: whtsai@cis.nctu.edu.tw.*

Manuscript received 2 Mar., 1998; revised 22 Mar., 1999.
Recommended for acceptance by K. Ikeuchi.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107630.

0162-8828/99/$10.00 © 1999 IEEE

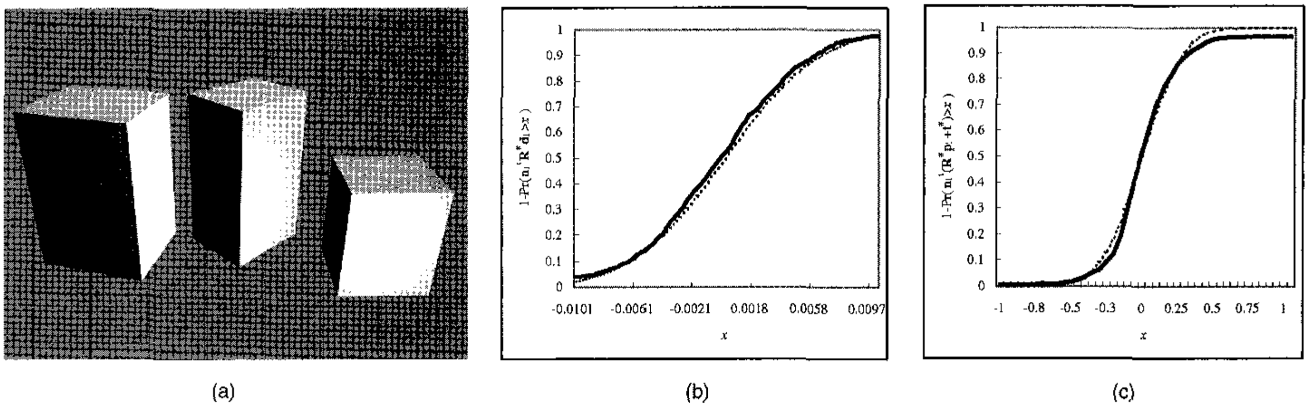


Fig. 2. (a) A synthesized image; (b) and (c) are cumulative distributions for $\mathbf{n}_1^t\mathbf{R}^{\#}\mathbf{d}_1$ and $\mathbf{n}_1^t(\mathbf{R}^{\#}\mathbf{p}_1 + \mathbf{t}^{\#})$, respectively, where the dotted curves in (b) and (c) are normal distributions with zero means and standard deviations 0.005 and $\sqrt{0.05}$, respectively.

vectors, respectively. Let rotation matrix $\mathbf{R}^{\#}$ and translation vector $\mathbf{t}^{\#}$, satisfying the requirement of the quality thresholds, denote an accurate estimated pose. If the distribution of $E(\mathbf{R}^{\#}, \mathbf{t}^{\#})$ is known, the quality of a pose can be tested by checking if the value of the error function of the pose is in the allowed range. Hence, the distribution of $E(\mathbf{R}^{\#}, \mathbf{t}^{\#})$ must be analyzed first.

In this study, the distributions of $\mathbf{n}_i^t(\mathbf{R}^{\#}\mathbf{p}_i + \mathbf{t}^{\#})$ and $\mathbf{n}_i^t\mathbf{R}^{\#}\mathbf{d}_i$ were found approximately normally distributed by computer simulations with the synthesized objects shown in Fig. 2a. Fig. 2b and Fig. 2c show the distributions of 1,000 samples. Each sample was produced from the lines fitting the edge pixels perturbed with Gaussian noise, and $\mathbf{R}^{\#}$ and $\mathbf{t}^{\#}$were obtained by a maximum likelihood estimator [8] with the ground true as an initial guess.

For each $i = 1, \cdots, N$, suppose that $\mathbf{n}_i^t\mathbf{R}^{\#}\mathbf{d}_i$ and $\mathbf{n}_i^t(\mathbf{R}^{\#}\mathbf{p}_i + \mathbf{t}^{\#})$ are normally distributed and σ_i and $\sigma_i^{'}$ denote their standard deviations, respectively. Since there are $2N$ measurements and six parameters to be estimated, $\frac{1}{2N-6}E(\mathbf{R}^{\#}, \mathbf{t}^{\#})$ denoted by $\hat{E}(\mathbf{R}^{\#}\mathbf{t}^{\#})$, can be regarded as a chi-square (χ^2) with one degree of freedom (denoted by $\chi^2(1)$) [12]. Furthermore, the *chi-square goodness-of-fit test* [13], [12] can be used to check if the residues in the orientation and position constraint equations are consistent with the uncertainty specified by σ_i and $\sigma_i^{'}$, $i = 1, \cdots, N$. So, a link between the test and the quality thresholds can be built by figuring out the relation among the thresholds, and σ_i and $\sigma_i^{'}$, $i = 1, \cdots, N$.

2.3 Determination of σ_i and $\sigma_i^{'}$, $i = 1, 2, \cdots, N$

In this study, the method employed to estimate object poses is assumed unknown. However, in general, the employed method tries to obtain a pose concentrating around the actual pose. In this study, normal distributions are used to describe the degree of concentration. So, the proposed test can be regarded as checking the degree of consistency between the qualities of the input and the estimated pose and those obtained by a "standard" method whose input and estimated pose are distributed in a desired manner. With some descriptions for the distributions, the upper bounds of σ_i^2 and $\sigma_i^{'2}$ in terms of δ_{R}, δ_{t}, and $\delta_{\Delta\mathrm{n}}$ can be obtained to be

$$\sigma_i^2 \leq \frac{9\delta_{\mathrm{R}}^2}{26} + \frac{\delta_{\Delta\mathrm{n}}^2}{13},$$

$$\sigma_i^{'2} \leq \frac{9{\delta_{\mathrm{R}}}^2\|\mathbf{p}_i\|_2^2}{26} + \frac{\delta_{\Delta\mathrm{n}}^2\left(\|\mathbf{p}_i\|_2 + \max\|\mathbf{t}^*\|_2\right)^2}{13} + \frac{\delta_{\mathrm{t}}^2}{13},$$

where $\|\cdot\|_2$ denotes the l_2 norm and $\max\|\mathbf{t}^*\|_2$ denotes the longest distance between the origins of the CCS and the OCS (for details, see Appendix A).

3 Testing Input Quality and Accuracy of Estimated Object Pose

3.1 Lower Bound of Error Function

Writing $E(\mathbf{R}, \mathbf{t})$ in a matrix form, we have $E(\mathbf{R}, \mathbf{t}) = \|\mathbf{A}\mathbf{r}\|_2^2 + \|\mathbf{B}\mathbf{r} + \mathbf{C}\mathbf{t}\|_2^2$, where

$$\mathbf{A} = \left[\frac{\mathbf{d}_1 \otimes \mathbf{n}_1}{\sigma_1} \quad \cdots \quad \frac{\mathbf{d}_{N-1} \otimes \mathbf{n}_{N-1}}{\sigma_{N-1}} \quad \frac{\mathbf{d}_N \otimes \mathbf{n}_N}{\sigma_N}\right]^t,$$
$$\mathbf{B} = \left[\frac{\mathbf{p}_1 \otimes \mathbf{n}_1}{\sigma'_1} \quad \cdots \quad \frac{\mathbf{p}_{N-1} \otimes \mathbf{n}_{N-1}}{\sigma'_{N-1}} \quad \frac{\mathbf{p}_N \otimes \mathbf{n}_N}{\sigma'_N}\right]^t,$$
$$\mathbf{C} = \left[\frac{\mathbf{n}_1}{\sigma'_1} \quad \cdots \quad \frac{\mathbf{n}_{N-1}}{\sigma'_{N-1}} \quad \frac{\mathbf{n}_N}{\sigma'_N}\right]^t,$$
$$\mathbf{r} = [R_{11}R_{12}R_{13}R_{21}R_{22}R_{23}R_{31}R_{32}R_{33}]^t,$$

in which $\mathbf{d} \otimes \mathbf{n} = [n_1\mathbf{d}^t \quad n_2\mathbf{d}^t \quad n_3\mathbf{d}^t]^t$ is the left direct product of two 3×1 vectors $\mathbf{d}$ and $\mathbf{n}$, and $\mathbf{r}$ is called the 9D vector associated with $\mathbf{R}$ and, thus, $\|\mathbf{r}\|_2 = \sqrt{3}$. By some manipulation, we have $E(\mathbf{R}, \mathbf{t}) \geq \mathbf{r}^t\mathbf{F}\mathbf{r}$ where $\mathbf{F}$ is positive semidefinite and equal to $\mathbf{A}^t\mathbf{A} + \mathbf{B}^t(\mathbf{I} - \mathbf{C}\mathbf{C}^+)\mathbf{B}$ with $\mathbf{C}^+ = (\mathbf{C}^tC)^{-1}\mathbf{C}^t$ and $\mathbf{I}$ the $N \times N$ identity matrix. Let $\Xi(\mathbf{R}, \mathbf{F}) = \mathbf{r}^t\mathbf{F}\mathbf{r}$, $\alpha_1, \alpha_2, \cdots, \alpha_9$ be the nine unit eigenvectors of $\mathbf{F}$ and $0 \leq \lambda_1 \leq \cdots \leq \lambda_9$ be the corresponding eigenvalues. Thus, $\Xi(\mathbf{R}, \mathbf{F})$ can be expressed as

$$\Xi(\mathbf{R}, \mathbf{F}) = \sum_{i=1}^{9} \lambda_i(\mathbf{r}^t\alpha_i)^2. \qquad (1)$$

From the Rayleigh-Ritz theorem [14], we have $\Xi(\mathbf{R}, \mathbf{F}) \geq 3\lambda_1$, but this bound is not tight enough.

Let $\mathbf{M}_i$ be the matrix associated with α_i and $\mathbf{K}_i$ be the rotation matrix closest to $\mathbf{M}_i$, in the l_2 matrix norm. We can have $\mathbf{K}_i = \mathbf{U}_i\mathrm{diag}(1, 1, \det(\mathbf{U}_i\mathbf{V}_i))\mathbf{V}_i^t$ and $\mathbf{U}_i\mathbf{S}_i\mathbf{V}_i^t$ is the singular value decomposition (SVD) [15] of $\mathbf{M}_i$, where $\mathrm{diag}(d_1, d_2, \cdots d_n)$ represents an $n \times n$ diagonal matrix and $\mathbf{S}_i = \mathrm{diag}(s1_i, s2_i, s3_i)$ with $s1_i \geq s2_i \geq s3_i \geq 0$. As shown in Appendix B, a lower bound LB_1 of $\Xi(\mathbf{R}, \mathbf{F})$ not smaller than $3\lambda_1$ can be obtained as follows:

$$LB_1 = tr(\mathbf{S}_1)^2\lambda_1 + \min\left\{3 - tr(\mathbf{S}_1)^2, tr(\mathbf{S}_2)^2\right\}\lambda_2 + \max\left\{3 - tr(\mathbf{S}_1)^2 - tr(\mathbf{S}_2)^2, 0\right\}\lambda_3.$$

Furthermore, in Appendix C, by using the perturbation theory of eigenvalues and eigenvectors [11], [15], an approximate lower bound LB_2 of $\Xi(\mathbf{R}, \mathbf{F})$ can be obtained as follows:

$$LB_2 = 3\lambda_1 + (6 - 2\sqrt{3}tr(\mathbf{S}_1))\lambda_2.$$

In this study, the larger of LB_1 and LB_2, denoted by LB, is used as a lower bound of $E(\mathbf{R}, \mathbf{t})$, and the number of line features is suggested to be at least eight.

TABLE 1
Quality Thresholds Used in this Study

set no.	$\delta_{\mathbf{R}}$	$\delta_{\mathbf{t}}(mm)$	$\delta_{\Delta\mathbf{n}}$
1	0.005	5	0.01
2	0.01	10	0.01
3	0.025	25	0.01
4	0.05	50	0.01

3.2 Definitions of Test Functions: H_{pre} and H_{post}

Let $E_{(\delta_{\mathbf{R}},\delta_{\mathbf{t}},\delta_{\Delta\mathbf{n}})}$ and $\hat{E}_{(\delta_{\mathbf{R}},\delta_{\mathbf{t}},\delta_{\Delta\mathbf{n}})}$ denote E and $\hat{E}$ with σ_i^2 and $\sigma_i'^2, i=1,\cdots,N$, determined by the way described in Section 2.3 with respect to a set of quality thresholds $(\delta_{\mathbf{R}},\delta_{\mathbf{t}},\delta_{\Delta\mathbf{n}})$, respectively. Since LB can be obtained before pose estimation, the function H_{pre} for testing the quality of the input data is defined as follows:

$$H_{pre} : \mathit{unacceptable} \text{ if } \frac{LB}{2N-6} > v; \mathit{acceptable} \text{ otherwise,}$$

where v is a predefined threshold value. After pose estimation, the function H_{post} for testing the quality of the estimated pose is defined as follows:

$$H_{post} : \mathit{unacceptable} \text{ if } \hat{E}_{(\delta_{\mathbf{R}},\delta_{\mathbf{t}},\delta_{\Delta\mathbf{n}})} > v;$$
$$\mathit{acceptable} \text{ if } \hat{E}_{(\frac{\delta_{\mathbf{R}}}{\kappa},\frac{\delta_{\mathbf{t}}}{\kappa},0)} \leq v; \mathit{unreliable} \text{ otherwise.}$$

With stricter quality thresholds $(\frac{\delta_{\mathbf{R}}}{\kappa},\frac{\delta_{\mathbf{t}}}{\kappa},0)$ where $\kappa > 1$ in the above test, the quality of the estimated pose is accepted in a conservative way. In general, the significance level of the $\chi^2(1)$ distribution is desired to be within 0.1 to 0.025, so v can be chosen from 3 to 5. In this study, $v = 3$. Let r_1 be the rate of rejecting the accurate estimated pose distributed in the way assumed in Section 2.3. The relation among v, r_1, and κ can be derived to be $r_1 \geq 2(\Phi(\sqrt{v}) - \Phi(\frac{\sqrt{v}}{\kappa}))$, where $\Phi(\cdot)$ denotes the *standard normal distribution (the derivation is omitted to* shorten this paper). Hence, κ is suggested to be not greater than three because v is chosen as 3 and r_1 is desired to be not greater than 0.5. In this study, we set $\kappa = 3$.

4 Experimental Results

First, some terminologies are defined as follows: Classifying the accurate estimated result to be unacceptable is called the type I error. Classifying the inaccurate estimated result to be acceptable is called the type II error. The error rate and the reject rate are defined by

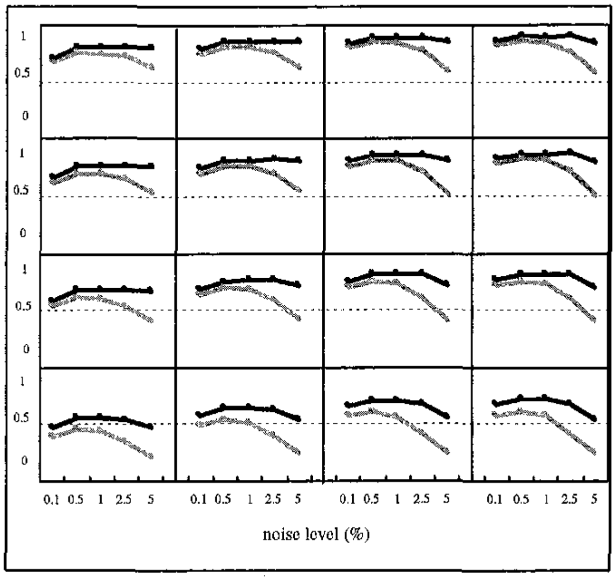
$$\text{error rate} = \frac{\text{the number of type I errors + the number of type II errors}}{\text{the number of test samples}},$$

$$\text{reject rate} = \frac{\text{the number of estimated results labeled unreliable}}{\text{the number of test samples}}.$$

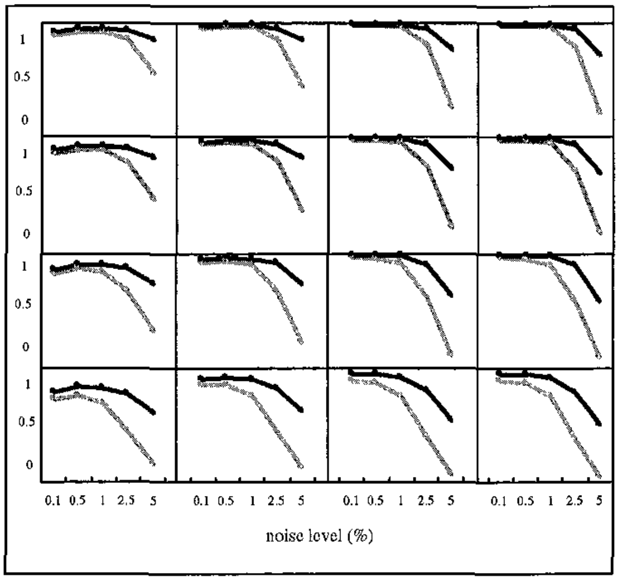
Four sets of quality thresholds were designed in Table 1. In this experiment, we did not use a good estimation method because this method is similar to the "standard" method. Alternatively, the method for obtaining pose parameters by defeating the test functions was preferred. Here, a method estimating the pose by minimizing $E_{(\delta_{\mathbf{R}},\delta_{\mathbf{t}},\delta_{\Delta\mathbf{n}})}(\mathbf{R},\mathbf{t})$ was adopted because the estimated pose can make H_{post} commit the type II error based on the fact that the estimated pose has the minimum of $E_{(\delta_{\mathbf{R}},\delta_{\mathbf{t}},\delta_{\Delta\mathbf{n}})}(\mathbf{R},\mathbf{t})$ but does not guarantee optimality.

4.1 Computer Simulation

To analyze the noise effect, $\mathbf{n}_i$ was perturbed by adding a noise vector $\rho\Delta\mathbf{n}$, where ρ is a scalar controlling the noise level. The elements of $\Delta\mathbf{n}$ were randomly generated from $[-1, 1]$. Eight-hundred-thousand random trials have been done with the numbers of line features 8, 10, 12, and 14 at noise levels



(a)



(b)

Fig. 3. Comparisons of LB and $3\lambda_1$ versus various numbers of line features, quality thresholds, and noise levels: (a) and (b) are for the average ratio and the average correlation coefficient, respectively, where from top to bottom are for the quality sets 1 to 4, and the numbers of line features from left to right are 8, 10, 12, and 14, respectively.

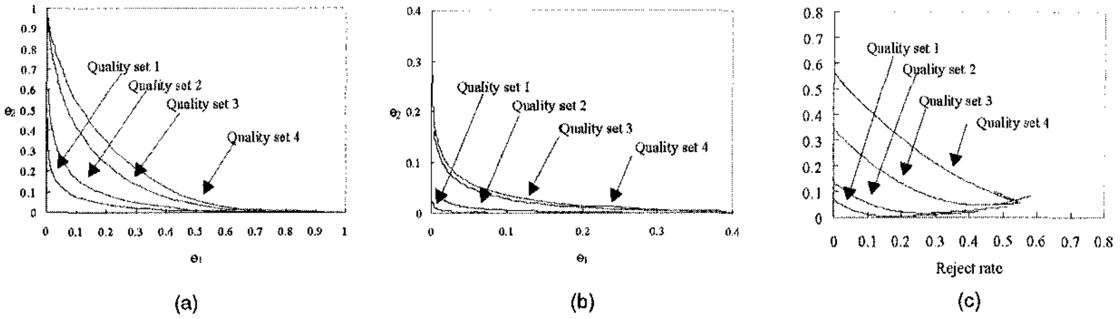


Fig. 4. (a) and (b) are operating characteristics curves of type I error (ε_1) versus type II error (ε_2) for H_{post}without and with a reject option, respectively;(c) is the error-reject curve for H_{post}.

TABLE 2
Experimental Results for the Synthesized Image, Cube Image and Printer Image

	synthesized image				cube image							
					no erroneous l.c.				one erroneous l.c.			
(i)	1	2	3	4	1	2	3	4	1	2	3	4
(ii)	0.64	0.94	0.99	1.0	0.103	0.033	0	0	0.642	0.491	0.086	0
(iii)	0.38	0.246	0.2	0.1132	0.892	0.566	0.139	0.12	0.999	0.989	0.904	0.859
(iv)	0.85	0.522	0.028	0.052	0.059	0.45	0.799	0	0	0.065	0.308	0.131
(v)	0.66	0.4	0.022	0.047	0.273	0.759	0.55	0.08	0.024	0.122	0.278	0.153
(vi)	0.0027	0.006	0.026	0.08	0.093	0.01	0.03	0.03	0	0.034	0.008	0.043

	printer image							
	no erroneous l.c.				one erroneous l.c.			
(i)	1	2	3	4	1	2	3	4
(ii)	0.01	0	0	0	0.838	0.989	0.553	0.213
(iii)	0.278	0.049	0.0017	0.0001	0.996	0.989	0.949	0.9
(iv)	0.644	0.207	0.002	0	1	0.907	0.396	0.433
(v)	0.702	0.227	0.0024	0	0.044	0.0512	0.064	0.104
(vi)	0.037	0.019	0.0017	0.0001	0	0.0005	0.008	0.013

The items (i) to (vi) represent the set of quality thresholds, ratio of the number of poor data detected by H_{pre} to the number of poor data to the number of test samples, rate of accurate estimated results labeled "unreliable", reject rate, and error rate, respectively.

0.001, 0.005, 0.01, 0.025, and 0.05 with respect to the four quality sets. Comparisons of LB and $3\lambda_1$ are given in Fig. 3. It shows that LB is highly correlative to the minimum of $E_{(\delta_{n},\delta_{t},\delta_{\Delta n})}(\mathbf{R},\mathbf{t})$ (denoted by E^{*}) and better than $3\lambda_1$, especially when the noise level is higher. Hence, LB is more suitable than $3\lambda_1$ to detect poor inputs. Fig. 4 shows some results of H_{post} tested by the samples with eight lines. The curves in Fig. 4 were generated by varying v from zero to ten at an interval of 0.1. The parts of the curves corresponding to v in the range $[3, 5]$ are indicated by thick curves. Fig. 4a and Fig. 4b illustrate the necessity for H_{post} to have a reject option and Fig. 4c shows that the reject region of H_{post} is proper.

4.2 Tests with Synthesized Image

Ten-thousand test samples were generated from the synthesized objects shown in Fig. 2a and a quarter of them were designed to contain an erroneous line correspondence (l.c.). Each sample contained 27 image lines obtained by fitting edge pixels corrupted by Gaussian noise with standard deviations 1, 2, or 3. Table 2 shows that the error rates are low and more than a half of poor inputs were detected. As shown in Fig. 6, most of the estimated poses have qualities within the reject regions specified by the quality sets 1 and 2, so the reject rates with respect to the two sets are high.

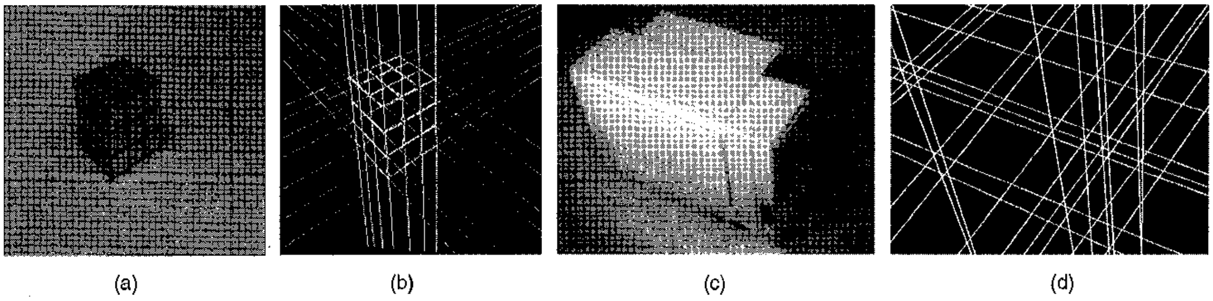


Fig. 5. Real images: (a) is an image of a cube; (c) is an image of a printer; and (b) and (d) are the results after performing straight line detection for (a) and (c), respectively.

4.3 Tests with Real Images

Fig. 5a shows an image of a cube. The pose of the cube estimated by Lee and Haralick's method [8] with well-detected lines was regarded as the ground true. In this experiment, to produce test samples with various qualities, some of the lines were not well-detected, as shown in Fig. 5b. The relative errors of the unit normal vectors formed by these lines were about 0.5 percent to 8 percent. Ten-thousand test samples were generated, each of which contained eight lines randomly chosen from these lines. As shown in Fig. 6, most of the estimated poses have relative orientation errors about 0.5 percent to 2 percent and absolute translation errors about 5 mm to 8 mm, and are close to the reject regions specified by the quality sets 2 and 3. Hence, as shown in Table 2, the reject rates with respect to the two sets are high. This indicates that the reject region of H_{post} is proper. Besides, the other 10,000 test samples, each of which contained an erroneous l.c., were produced. About half of poor inputs with respect to the quality sets 1 and 2 and 10 percent of poor inputs with respect to the quality set 3 were detected. This experiment was also done for the printer image. Since the differences among the line features of the printer image are more significant than those of the cube image, the experimental result for the printer image is better than that for the cube image.

4.4 An Example of the Test Function for 2D/3D Feature Point Correspondences (p.c.s)

Experiments of using H_{pre} to test if there exist erroneous p.c.s between the 3D model points and the image points have been conducted for the synthesized image and the printer image. The corners of the 3D objects in the two images were the model points. For each test sample, the image lines and the corresponding model lines were formed by connecting every pair of the visible corners and the corresponding pair of the model points. The quality thresholds were specified by the quality set 4. For each of the two images, 150 test samples with erroneous p.c.'s were produced in addition to 50 normal test samples with visible corners corrupted by Gaussian noise with a standard derivation of 2. Fig. 7 shows the value $\frac{LB}{2N-6}$ of every test sample where the horizontal line is the threshold of H_{pre}. It reveals that the test samples with erroneous p.c.s can be detected.

In summary, the experimental results show that H_{pre} can detect poor inputs, and the reject region of H_{post} is proper because the quality of the rejected estimated pose is close to the specified quality thresholds.

5 Conclusions

In this paper, a method has been proposed to test the quality of estimated poses with respect to the quality thresholds which can be specified straightforwardly by users for their applications. Two test functions, H_{pre} and H_{post}, have been defined. H_{pre} can be used to detect poor inputs. So, we can abandon poor inputs to avoid unnecessary computation or use some methods to deal with them. After pose estimation, H_{post} can be used to qualify estimated results as acceptable, unacceptable, or unreliable. The unreliable estimated result can be either used with less confidence or reevaluated by other methods. The test functions are more suitable for pose estimation methods having the properties that the resultant residues in the orientation and position constraints are approximately normally distributed and that the estimated poses concentrate around the actual pose. Since the two test functions are simple, they can be easily embedded in an existing algorithm.

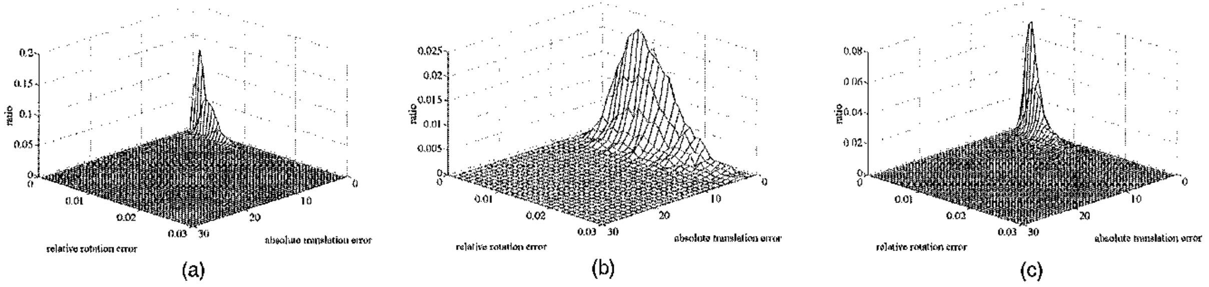


Fig. 6. Population of the estimated poses for the synthesized image and the real images:(a), (b), and (c) are for the synthesized image, the cube image, and the printer image, respectively, where 87, 90, and 99 percent of estimated poses are shown for the three images, respectively, and the rest are outside the region displayed.

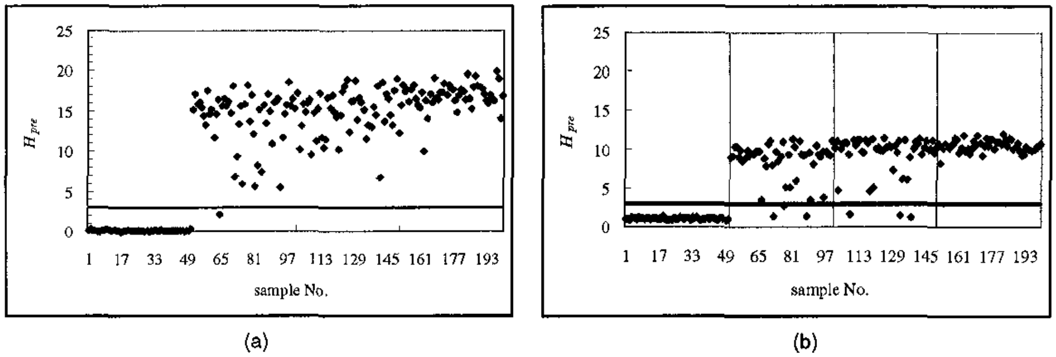


Fig. 7. Plots of the experimental results for testing if the test sample contains erroneous p.c.'s:(a) is for the synthesized image, and (b) is for the printer mage, where the four quarters of the test samples from left to right are the normal test samples, and the test samples with one, two, and three erroneous p.c.'s.

Appendix A
Determination of Standard Deviations σ_i and $\sigma_i^{'}, i = 1, 2, ..., N$

Let $\mathbf{n}_i$ be described by its noise-free version $\mathbf{n}_i^*$ plus a noise vector $\Delta\mathbf{n}_i$, $\mathbf{R}^*$ and $\mathbf{t}^*$ be the ground true rotation matrix and translation vector, respectively, and $\Delta\mathbf{R} = \mathbf{R}^{\#} - \mathbf{R}^*$ and $\Delta\mathbf{t} = \mathbf{t}^{\#} - \mathbf{t}^*$. For simplicity, we assume $\Delta\mathbf{R}$, $\Delta\mathbf{t}$, and $\Delta\mathbf{n}_i$, $i = 1, 2, \cdots, N$, are uncorrelated and have zero means. Since the distribution of $\Delta\mathbf{t}$ is desired to be homogeneous in all directions, the elements of $\Delta\mathbf{t}$ are assumed to be *independently and identically distributed (i.i.d.)* random variables (r.v.s) with a normal density $N(0, \sigma_{\mathrm{t}}^2)$. The same is for $\Delta\mathbf{n}_i$, $i = 1, 2, \cdots, N$, with $N(0, \sigma_{\Delta\mathrm{n}}^2)$. So, the variance of $\mathbf{n}_i^t\mathbf{R}^{\#}\mathbf{d}_i$ is

$$\begin{aligned}\sigma_i^2 &= \mathrm{E}[(\mathbf{n}_i^t\mathbf{R}^{\#}\mathbf{d}_i)^2] \\ &= \mathrm{E}[(\mathbf{n}_i^{*t}\Delta\mathbf{R}\mathbf{d}_i)^2] + \mathrm{E}[(\Delta\mathbf{n}_i^t\mathbf{R}^*\mathbf{d}_i)^2] + \mathrm{E}[(\Delta\mathbf{n}_i^t\Delta\mathbf{R}\mathbf{d}_i)^2]. \quad (2)\end{aligned}$$

By ignoring the right most term of (2), the least significant term, and using the fact that the spectral matrix norm is induced by the l_2 vector norm [14], we have

$$\sigma_i^2 \leq \mathrm{E}[s_1(\Delta\mathbf{R})^2] + s_1(\mathrm{Cov}(\Delta\mathbf{n})),$$

where $s_1(\cdot)$ denotes the largest singular value of a matrix.

Representing $\mathbf{R}^{\#}$ by $\mathbf{R}^*$ multiplied by a rotation matrix $\mathbf{R}^{'}$, we get $\Delta\mathbf{R} = \mathbf{R}^*(\mathrm{R}' - \mathbf{I}_{3\times3})$. Suppose that $\mathbf{R}^{\#}$ is close to $\mathbf{R}^*$. Thus, expressing $\mathbf{R}'$ by a Taylor's series up to the first-order term [15], we can obtain $\Delta\mathbf{R} = \mathbf{R}^*\mathbf{L}$, where $\mathbf{L}$ is an antisymmetric matrix in terms of l_1, l_2, and l_3, which are assumed i.i.d. r.v. s from $N(0, \sigma_{\mathrm{R}}^2)$. Hence, $\frac{\|\Delta\mathbf{R}\|_2^2}{2\sigma_{\mathrm{R}}^2}$ has the $\chi^2(3)$ distribution [16]. *Since* $\frac{\|\Delta\mathbf{R}\|_2}{\sqrt{3}}$ is desired to be not greater than δ_{R}, $\mathrm{Pr}\left(\frac{\|\Delta\mathbf{R}\|_2^2}{2\sigma_{\mathrm{R}}^2} > \frac{3\delta_{\mathrm{R}}^2}{2\sigma_{\mathrm{R}}^2}\right)$ must be small. At a 99.5 percent confidence interval for $\frac{\|\Delta\mathbf{R}\|_2^2}{2\sigma_{\mathrm{R}}^2}$, we can obtain $\sigma_{\mathrm{R}}^2 \leq \frac{3\delta_{\mathrm{R}}^2}{26}$. Since $s_1(\Delta\mathbf{R})^2 = \frac{\|\mathbf{L}\|_2^2}{2}$ (the proof is simple and thus omitted), we have $\mathrm{E}[s_1(\Delta\mathbf{R})^2] = 3\sigma_{\mathrm{R}}^2 \leq \frac{9\delta_{\mathrm{R}}^2}{26}$. Using a similar technique, we can obtain $s_1(\mathrm{Cov}(\Delta\mathbf{n})) \leq \frac{\delta_{\Delta\mathrm{n}}^2}{13.}$ Accordingly, we have $\sigma_i^2 \leq \frac{9\delta_{\mathrm{R}}^2}{26} + \frac{\delta_{\Delta\mathrm{n}}^2}{13}$. Upper bounds for $\sigma_i'^2, i = 1, 2, \cdots, N$, can be determined by a similar method.

Appendix B
Derivations of LB_1

From the SVD of $\mathbf{K}_i$, we have the following three results for $\mathbf{K}_i$: 1) $\|\mathbf{K}_i - \sqrt{3}\mathbf{M}_i\|_2^2 = (1 - \sqrt{3}s1_i)^2 + (1 - \sqrt{3}s2_i)^2 + (\det(\mathbf{U}_i\mathbf{V}_i) - \sqrt{3}s3_i)^2$, 2) $\mathbf{k}_i^t\alpha_i = s1_i + s2_i + \det(\mathbf{U}_i\mathbf{V}_i)s3_i$, and 3) $\mathbf{r}^t\boldsymbol{\alpha}_i \leq \mathbf{k}_i^t\boldsymbol{\alpha}_i$ where $\mathbf{k}_i$ and $\mathbf{r}$ are the 9D vectors associated with $\mathbf{K}_i$ and an arbitrary rotation matrix $\mathbf{R}$, respectively. Let $\mathbf{K}_i^{'}$ be the rotation matrix closest to $-\mathbf{M}_i$ in the l_2 matrix norm and $\mathbf{k}_{\mathrm{i}}^{'}$ be the associated 9D vector. The SVD of $\mathbf{K}_i^{'}$ can be expressed as $\mathbf{U}_i^{'}\mathbf{S}_i\mathbf{V}_i^{'t}$ where $\det(\mathbf{U}_i^{'}\mathbf{V}_i^{'}) = -\det(\mathbf{U}_i\mathbf{V}_i)$. The other three results for $\mathbf{K}_i^{'}$ can also be expressed by substituting $\mathbf{K}_i, \mathbf{k}_i, \mathbf{M}_i, \alpha_i$, and $\det(\mathbf{U}_i\mathbf{V}_i)$ in the above three with $\mathbf{K}_i^{'}, \mathbf{k}_i^{'}, -\mathbf{M}_i, -\alpha_i$, and $\det(\mathbf{U}_i^{'}\mathbf{V}_i^{'})$, respectively. Since $\det(\mathbf{U}_i\mathbf{V}_i)$ is equal to 1 or -1, we have

$$\min\left\{\left\|\mathbf{K}_i - \sqrt{3}\mathbf{M}_i\right\|_2^2, \left\|\mathbf{K}_i - \sqrt{3}(-\mathbf{M}_i)\right\|_2^2\right\} = 6 - 2\sqrt{3}\mathrm{tr}(\mathbf{S}_{\mathrm{i}}), \quad (3)$$

$$\max\left\{(\mathbf{k}_i^t\boldsymbol{\alpha}_i)^2, (\mathbf{k}_i^{'t}(-\boldsymbol{\alpha}_i))^2\right\} = \mathrm{tr}(\mathbf{S}_i)^2, \quad (4)$$

$\mathbf{r}^t\boldsymbol{\alpha}_i \leq \mathbf{k}_i^t\alpha_i$, and $\mathbf{r}^t(-\boldsymbol{\alpha}_i) \leq \mathbf{k}_i^{'t}(-\boldsymbol{\alpha}_i)$. That is, we have $(\mathbf{r}^t\boldsymbol{\alpha}_i)^2 = (\mathbf{r}^t(-\boldsymbol{\alpha}_i))^2 \leq tr(\mathbf{S}_i)^2$. In addition, $\|\mathbf{R}\|_2^2 = 3$ and $\mathrm{tr}(\mathbf{S}_i) \geq 1$, $i = 1, \cdots, 9$. Hence, LB_1 can be obtained.

Appendix C
Derivations of LB_2

Suppose that $\mathbf{R}^{\#}$ and $\mathbf{t}^{\#}$ minimize $\Xi(\mathbf{R}, \mathbf{F})$. Let $\mathbf{F}^{\#}$ be the noise-free $\mathbf{F}$ with the pose parameters $\mathbf{R}^{\#}$ and $\mathbf{t}^{\#}$. Let $\alpha_i^{\#}$ be the unit eigenvector of $\mathbf{F}^{\#}$ and $\lambda_i^{\#}$ be the corresponding eigenvalue, for $i = 1, 2, \cdots, 9$, respectively. Thus, the 9D vector of $\mathbf{R}^{\#}$ is either $\sqrt{3}\alpha_1^{\#}$ or $-\sqrt{3}\alpha_1^{\#}$. By regarding $\mathbf{F}$ as a perturbed $\mathbf{F}^{\#}$, from the perturbation theory of eigenvalues and eigenvectors [11], [15], the first-order approximations of the eigenvectors of $\mathbf{F}$ in terms of the eigenvectors and eigenvalues of $\mathbf{F}^{\#}$ are given as follows:

$$\boldsymbol{\alpha}_i \cong \boldsymbol{\alpha}_i^{\#} + \sum_{j=1, j\neq i}^{9}\left(\frac{\boldsymbol{\alpha}_j^{\#t}\mathbf{F}\boldsymbol{\alpha}_i^{\#}}{\lambda_i^{\#} - \lambda_j^{\#}}\right)\boldsymbol{\alpha}_j^{\#}, i = 1, 2, \cdots, 9. \quad (5)$$

Thus, from (1) and (5), and the fact that $\mathbf{F}$ is symmetric, $\Xi(\mathbf{R}^{\#}, \mathbf{F})$ can be expressed as

$$\Xi(\mathbf{R}^{\#}, \mathbf{F}) = 3\alpha_1^{\#t}\mathbf{F}\alpha_1^{\#} \cong 3\lambda_1 + 3\sum_{i=2}^{9}\lambda_i\left(\frac{\alpha_i^{\#t}\mathbf{F}\alpha_1^{\#}}{\lambda_1^{\#} - \lambda_i^{\#}}\right)^2. \quad (6)$$

Regarding $\sqrt{3}\mathbf{M}_1$ or $-\sqrt{3}\mathbf{M}_1$ as the perturbed $\mathbf{R}^{\#}$, from (6) and (7), we have

$$3\sum_{i=2}^{9}\left(\frac{\boldsymbol{\alpha}_i^{\#t}\mathbf{F}\boldsymbol{\alpha}_1^{\#}}{\lambda_1^{\#}-\lambda_i^{\#}}\right)^2 \cong \left\|\text{perturbed version of } \mathbf{R}^{\#}-\mathbf{R}^{\#}\right\|_2^2 \qquad (7)$$

$$\geq 6-2\sqrt{3}\mathrm{tr}(\mathbf{S}_1).$$

From $\lambda_9 \geq \cdots \geq \lambda_2$, and (6) and (7), we have the following result from which LB_2 can be obtained:

$$\Xi(\mathbf{R}^{\#},\mathbf{F}) \ \geq 3\lambda_1+3\lambda_2\sum_{i=2}^{9}\left(\frac{\boldsymbol{\alpha}_i^{\#t}\mathbf{F}\boldsymbol{\alpha}_1^{\#}}{\lambda_1^{\#}-\lambda_i^{\#}}\right)^2 \geq 3\lambda_1+(6-2\sqrt{3}\mathrm{tr}(\mathbf{S}_1))\lambda_2.$$

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their valuable comments and help. This work was supported by the Republic of China National Science Council under Grant NSC-86-2213-E-009-114.

REFERENCES

[1] R.M. Haralick and L.G. Shapiro, *Computer and Robot Vision*, vol. 2. Reading, Mass.: Addision Welsey 1993.

[2] T.S. Huang and A.N. Netravali, "Motion and Structure from Feature Correspondences: A Review," *Proc. the IEEE*, vol. 82, no. 2, pp. 252-268, 1994.

[3] Y. Liu, T.S. Huang, and O.D. Faugeras, "Determination of Camera Location from 2D to 3D Line and Point Correspondences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 28-37, Jan. 1990.

[4] M. Dhome, M. Richetin, J.T. Lapreste, and G. Rives, "Determination of the Attitude of 3D Objects from a Single Perspective View," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, pp. 1,265-1,278, Dec. 1989.

[5] 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.

[6] R. Kumar and A.R. Hanson, "Robust Methods for Estimating Pose and a Senitivity Analysis," *Computer Vision and Graphic Image Processing: 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 2D to 3D Point and Line Correspondences," *Int'l J. Computer Vision*, vol. 15, pp. 225-243, 1995.

[8] 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.

[9] K. Cho, P. Meer, and J. Cabrera,, "Performance Assessment through Bootstrap," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 11, pp. 1,185-1,198, Nov. 1997.

[10] S. Yi, R.M. Haralick, and L.G. Shapiro, "Error Propagation in Machine Vision," *Machine Vision and Applications*, vol. 7, pp. 93-114, 1994.

[11] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second ed. Boston, Mass.: Academic Press, 1990.

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

[13] B.K.P. Horn, "Relative Orientation," *Int'l J. Computer Vision*, vol. 4, pp. 59-78, 1990.

[14] R.A. Horn and C.R. Johnson, *Matrix Analysis*. New York, NY: Cambridge Univ. Press, 1985.

[15] K. Kanatani, *Geometric Computation for Machine Vision*. New York, NY: Oxford Univ. Press, 1993.

[16] G.G. Roussas, *A Course in Mathematical Statistics*. second ed. New York, NY: Academic Press, 1997.

Tracking Human Motion in Structured Environments Using a Distributed-Camera System

Q. Cai and J.K. Aggarwal, *Fellow, IEEE*

Abstract—This paper presents a comprehensive framework for tracking coarse human models from sequences of synchronized monocular grayscale images in multiple camera coordinates. It demonstrates the feasibility of an end-to-end person tracking system using a unique combination of motion analysis on 3D geometry in different camera coordinates and other existing techniques in motion detection, segmentation, and pattern recognition. The system starts with tracking from a single camera view. When the system predicts that the active camera will no longer have a good view of the subject of interest, tracking will be switched to another camera which provides a better view and requires the least switching to continue tracking. The nonrigidity of the human body is addressed by matching points of the middle line of the human image, spatially and temporally, using Bayesian classification schemes. Multivariate normal distributions are employed to model class-conditional densities of the features for tracking, such as location, intensity, and geometric features. Limited degrees of occlusion are tolerated within the system. Experimental results using a prototype system are presented and the performance of the algorithm is evaluated to demonstrate its feasibility for real time applications.

Index Terms—Tracking, human modeling, motion estimation, multiple perspectives, Bayesian classification, end-to-end vision systems.

1 INTRODUCTION

TRACKING human motion is of interest in numerous applications such as surveillance, analysis of athletic performance, and content-based management of digital image databases. Recently, growing interest has concentrated upon tracking humans using distributed monocular camera systems to extend the limited viewing angle of a single fixed camera [1], [2], [3]. In such a setup, the cameras are arranged to cover a monitored area with overlapping vision fields to ensure a smooth switching among cameras during tracking. We present a comprehensive framework for automatically tracking coarse human models across multiple camera coordinates and demonstrate the feasibility of an end-to-end person tracking system using a unique combination of motion analysis on 3D geometry in different camera coordinates with existing techniques in motion detection, segmentation, and pattern recognition. The nonrigidity of the human body is addressed by matching points of the middle line of the human image, spatially and temporally, using Bayesian classification schemes. The key to successful tracking in the proposed work relies on our unique method of 3D motion prediction and estimation from different perspectives. Experimental studies using a three-camera prototype system show its efficiency in computation and potential for real time applications.

The earliest work in this area is, perhaps, by Sato et al. [1]. They considered the moving human image as a combination of various blobs. All distributed cameras were calibrated in the world coordinate system, which corresponds to a CAD model of the

- *Q. Cai is with the Consulting Group, Realnetworks, Inc., 2601 Elliott Ave., Seattle, WA 98121. E-mail: qcai@real.com.*
- *J.K. Aggarwal is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084. E-mail: aggarwaljk@mail.utexas.edu.*

Manuscript received 16 Oct. 1997; revised 3 Sept. 1999.
Recommended for acceptance by R. Szeliski.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107749.

0162-8828/99/$10.00 © 1999 IEEE