

Journal of Zhejiang University SCIENCE
 ISSN 1009-3095
 http://www.zju.edu.cn/jzus
 E-mail: jzus@zju.edu.cn



Developing rigid constraint for the estimation of pose and structure from a single image

WEI Bao-gang (魏宝刚)^{†1}, LIU Yong-huai (刘永怀)²

(¹State Key Laboratory of CAD & CG, Zhejiang University, Hangzhou 310027, China)

(²Department of Computer Science, University of Wales, Aberystwyth Ceredigion SY23 3DB, Wales, UK)

[†]E-mail: wbg@zju.edu.cn

Received Sept. 30, 2003; revision accepted Dec. 10, 2003

Abstract: Pose and structure estimation from a single image is a fundamental problem in machine vision and multiple sensor fusion and integration. In this paper we propose using rigid constraints described in different coordinate frames to iteratively estimate structural and camera pose parameters. Using geometric properties of reflected correspondences we put forward a new concept, the reflected pole of a rigid transformation. The reflected pole represents a general analysis of transformations that can be applied to both 2D and 3D transformations. We demonstrate how the concept is applied to calibration by proposing an iterative method to estimate the structural parameters of objects. The method is based on a coarse-to-fine strategy in which initial estimation is obtained through a classical linear algorithm which is then refined by iteration. For a comparative study of performance, we also implemented an extended motion estimation algorithm (from 2D-2D to 3D-2D case) based on epipolar geometry.

Key words: Structural constraints, 3D-2D problem, Rigid transformation

Document code: A

CLC number: TP273

INTRODUCTION

Pose and structural parameters analysis of 2D images sets find application in many areas such as object recognition, motion estimation, navigation planning, and structural analysis of 3D objects (Huang and Netravali, 1994), image communication, and image coding (Mitiche and Aggarwal, 1986). Normally, given one set of 3D object point data and their corresponding 2D projective image point data, the 3D-2D calibration problem is referred to as the estimation of the camera parameters (position and orientation) in relation to the object's coordinate frame and, conversely, the estimation of

the structure of the object in camera centered coordinate frame.

To solve the 3D-2D problems, many methods had been proposed such as techniques based on conservation of distance between feature points before and after a rigid motion (Mitiche and Aggarwal, 1986), triangular geometry (Linnainmaa *et al.*, 1987), iterative least squares method (Haralick *et al.*, 1989; Carceroni and Brown, 1997; Lowe, 1991; Yamane *et al.*, 1996), pose first algorithm (Faugeras and Hebert, 1986), model based genetic algorithm (Toyama *et al.*, 1998), symmetry plane based algorithm (Hattori *et al.*, 1998), iterative weak perspective algorithm, iterative weak paraperspective algorithm, and non-linear algorithm (Dornaika and Garcia, 1999), among many others. Such methods, when applied to real world

* Project (No. M603228) supported by the Natural Science Foundation of Zhejiang Province, China

applications, present a number of common limitations, such as lack of efficiency, high sensitivity to noise (Tsai and Huang, 1984), and multiple solutions (Huang and Netravali, 1994; Quan and Lan, 1999). Careful analysis of these algorithms indicated that they were mainly based on a number of geometric invariants described in a single coordinate frame. There is thus, justification for investigating invariant properties as constraints for calibration when the two sets of data (before and after a transformation) are described by two different coordinate frames.

We investigated this case starting from Chasles' screw motion theory, which says that any rigid non-pure translational planar displacement can be represented as a single rotation around its pole. However, for spatial displacements (the 3D case), there is no similar known property. Our investigation into the geometric properties of reflected correspondence vectors resulted in a formalisation of constraints based on the concept of the reflected pole of a transformation, for both planar displacements and spatial displacements. This novel and useful concept describes relationships existing between a set of points or feature vectors defined in one coordinate frame before the motion and their reflected correspondences defined in another coordinate frame after the motion. Essentially, for a planar displacement, all points and their reflected correspondences must lie on concentric circles centered at the reflected pole of that planar displacement. For a spatial displacement, all points and their reflected correspondences must lie on the surface of concentric spheres centered at the reflected pole of that spatial displacement. The important aspect of such properties is that they provide perfect constraints on the corresponding points described by different coordinate frames before and after the motion and can be used to refine structural estimation as described in this paper.

Based on such rigid constraints, we first propose to estimate structural parameters through an iterative method. Once the structure of the object in camera centered coordinate frame has been estimated, the original 3D-2D pose estimation is transformed into a 3D-3D motion estimation

problem. As a result, a number of existing accurate and robust motion estimation algorithms can be used. Given that a number of researchers demonstrated that the constraint least squares method (CLS) (Arun *et al.*, 1987; Umeyama, 1991) is the most accurate and robust motion estimation method for image data corrupted by noise without outliers (Eggert *et al.*, 1997; Matei *et al.*, 1998), we thus, adopted this algorithm for camera pose estimation.

In order to obtain a good initial estimation of structural parameters for the iterative algorithm, we adopted a classical linear algorithm which is solved using the total least squares method (TLS) (Chaudhuri and Chatterjee, 1991) rather than the least squares method as used in the Pose First Linear Algorithm (PFLA) described in (Rodrigues and Liu, 1999). Even though the PFLA algorithm is sensitive to noise, it had been demonstrated capable of providing a good initial estimation, so that we expect that the PFLA with TLS will also behave well. For a comparative study of performance, we also implemented a 3D-2D extended motion estimation algorithm (Extended Tsai and Huang Algorithm, ETHA) based on epipolar geometry that was initially proposed in (Tsai and Huang, 1984) to solve 2D-2D problems. This algorithm was chosen because epipolar geometry represents a unique linear solution for all 3D-3D, 3D-2D, and 2D-2D motion calibration problems and has the advantage of simplicity of implementation. Experimental results based on both synthetic data and real images have shown that the proposed iterative algorithm is more robust and accurate than both the ETHA and classical linear algorithms.

The rest of this paper is organized as follows: the theoretical foundations for the proposed algorithm is described in Section 2, the iterative algorithm is proposed in Section 3, and experimental results are presented in Section 4. Finally, some conclusions are drawn in Section 5.

THEORETICAL FOUNDATION

A rigid body transformation can be represented by the following relationship:

$$\mathbf{p}' = \mathbf{R}\mathbf{p} + \mathbf{t} \quad (1)$$

where \mathbf{R} and \mathbf{t} represent, respectively, the rigid body rotation matrix and translation vector, \mathbf{p} is a point described in one coordinate frame before the transformation and \mathbf{p}' its corresponding point described in another coordinate frame after the transformation. We can represent this transformation as (\mathbf{R}, \mathbf{t}) . Given a correspondence pair $(\mathbf{p}, \mathbf{p}')$, its reflected correspondence (\mathbf{RC}) is defined as: $\mathbf{RC} = (\mathbf{p}, -\mathbf{p}') = (\mathbf{p}, \mathbf{p}'')$, leading to the definition of the reflected correspondence vector \mathbf{RCV} as: $\mathbf{RCV} = \mathbf{p} - \mathbf{p}''$. Since the field of view normally cannot exceed 180° , we assume in this paper that the rotation angle of the transformation is constrained as $0 \leq \theta < \pi$.

From Chasles' screw motion theory (Ball, 1900), the pole \mathbf{e} of a rigid non-pure translational planar transformation is equidistant to any correspondence $(\mathbf{p}, \mathbf{p}')$ subject to this transformation:

$$\|\mathbf{p} - \mathbf{e}\| = \|\mathbf{p}' - \mathbf{e}\|$$

since $\mathbf{I} - \mathbf{R}$ is invertible. This means that any rigid non-pure translational planar transformation can be represented as a single rotation around the pole \mathbf{e} . This useful property had been widely used to analyze non-pure translational planar motions in mechanics and kinematics (McCarthy, 1990). However, there is no similar property for rigid non-pure translational spatial transformations since $\mathbf{I} - \mathbf{R}$ is singular. In an earlier work (Rodrigues and Liu, 1999) we investigated the screw motion concept in which any rigid non-pure translational spatial displacement can be effected by a rotation around an axis and a translation along this axis. Projecting all correspondences $(\mathbf{p}, \mathbf{p}')$ on a plane perpendicular to the screw axis, a counterpart of the pole \mathbf{e} on this plane can be found. We have formalized a number of geometric properties that can be used as constraints to motion estimation.

In this work, we focus our investigation on the analysis of the geometrical properties of reflected correspondences $(\mathbf{p}, \mathbf{p}'')$. We put forward the following theorem on general displacement that can be used as rigid constraint to motion estimation problems both in 2D and in 3D:

Theorem 1 For any rigid transformation, there must exist one and only one point \mathbf{c} which is equidistant to any reflected correspondence $(\mathbf{p}, \mathbf{p}'')$ subject to this transformation:

$$\|\mathbf{p} - \mathbf{c}\| = \|\mathbf{p}'' - \mathbf{c}\| \quad (2)$$

Proof of sufficiency Assuming that the rotation angle θ is defined as $0 \leq \theta < \pi$, if the rigid body transformation (\mathbf{R}, \mathbf{t}) is known, then the point \mathbf{c} can be constructed as:

$$\mathbf{c} = -(\mathbf{I} + \mathbf{R})^{-1}\mathbf{t} \quad (3)$$

which is equivalent to: $\mathbf{c} + \mathbf{RC} = -\mathbf{t}$. Thus we have:

$$\|\mathbf{p}'' - \mathbf{c}\| = \|-\mathbf{R}\mathbf{p} - \mathbf{t} - \mathbf{c}\| = \|-\mathbf{R}\mathbf{p} + \mathbf{R}\mathbf{c}\| = \|\mathbf{p} - \mathbf{c}\| \quad (4)$$

This shows that the point \mathbf{c} is equidistant to any reflected correspondence $(\mathbf{p}, \mathbf{p}'')$.

Proof of necessity If there is another point \mathbf{c}' which is equidistant to any reflected correspondence $(\mathbf{p}, \mathbf{p}'')$, then we have: $\|\mathbf{p} - \mathbf{c}'\| = \|\mathbf{p}'' - \mathbf{c}'\|$ and $\|\mathbf{p} - \mathbf{c}'\| = \|\mathbf{p}'' - \mathbf{c}'\|$. As a result, we have: $(\mathbf{p}, -\mathbf{p}'')^\top (\mathbf{c} - \mathbf{c}') = 0$. Because $(\mathbf{p}, \mathbf{p}'')$ is an arbitrary reflected correspondence, then \mathbf{c}' and \mathbf{c} must be equal: $\mathbf{c}' = \mathbf{c}$.

The point \mathbf{c} is called the reflected pole of a rigid transformation. When $\mathbf{R} = \mathbf{I}$ which corresponds to a pure translational transformation where \mathbf{I} is the identity matrix, we have $\mathbf{c} = -\mathbf{t}/2$. The reflected pole \mathbf{c} synthesises the transformation information rotation matrix \mathbf{R} and translation vector \mathbf{t} . As a result, it bridges the points described in different coordinate frames before and after the motion. This is a very interesting and useful property which can be used to calibrate structural parameters.

Theorem 1 shows that, for a planar motion, the perpendicular bisector lines of all reflected correspondence vectors must intersect at the reflected pole \mathbf{c} of that planar motion and, for a spatial motion, the perpendicular bisector planes of all reflected correspondence vectors must intersect at the reflected pole \mathbf{c} of that spatial motion. As a result, all reflected correspondences subject to a planar displacement must lie on a circle centered at the reflected pole \mathbf{c} of that planar displacement. Similarly, all reflected correspondences subject to a spatial displacement must lie on a sphere centered

at the reflected pole \mathbf{c} of that spatial displacement. This property is more general than the property derived from screw motion analysis, since the screw motion must be categorized as planar or spatial depending on whether $\mathbf{I}-\mathbf{R}$ is invertible or not. In addition, while a pure translational motion is a singular case of the screw motion analysis, Theorem 1 is true either for pure translational motion or non-pure translational motion and it does not matter whether the motion is described in 2D or 3D.

DESCRIPTION OF THE ALGORITHM

From Eq.(1), the relationship between a 3D object point \mathbf{p} described in the object centered coordinate frame and its corresponding 2D perspective image point $\mathbf{P}'=(X', Y')^T$ described in camera centered coordinate frame can be expressed as:

$$z' \begin{pmatrix} \mathbf{P}' \\ 1 \end{pmatrix} = \mathbf{R}\mathbf{p} + \mathbf{t} \quad (5)$$

where the classical pin hole camera model is employed assuming that the focal length f of the camera is equal to 1 for convenience of computation; without loss of generality, z' represents the depth of point \mathbf{p} in camera centered coordinate frame, $(\mathbf{P}'^T, 1)^T$ represents the homogeneous coordinate of image point \mathbf{P}' , and \mathbf{R} and \mathbf{t} represent the orientation and position of the camera in object centered coordinate frame. In the rest of this paper, we use d to represent the depth z' for conciseness of notation.

The algorithm described below is called Coarse-to-Fine Geometric Algorithm (CFGa). The constraint to satisfy is given by Theorem 1: if the depth of every point is exactly known, then Eq.(2) is exactly right. However, in practice, due to noise corrupting image data, the equality in Eq.(2) does not hold. Moreover, 3D-2D estimation adds an extra layer of uncertainty, as depth information from 2D images is not available. The CFGa strategy then, is to use a linear algorithm for a first approximation of motion parameters. The algorithm then refines the initial structural estimation through iteration leading to a refined estimation of

pose parameters and then back to a refined estimation of structural parameters. Thus, given two sets of 3D-2D correspondences $(\mathbf{p}_i, \mathbf{P}'_i)$ where $(i=1, 2, \dots, n | n \geq 6)$ the steps in the CFGa algorithm are described as follows.

Step 1: Initial estimation of pose, structure, and reflected pole

For a good initial estimation of the depth of each point and reflected pole, we use a classical linear algorithm, Pose First Linear Algorithm (PFLA) (Faugeras and Hebert, 1986; Rodrigues and Liu, 1999). The PFLA estimation is solved by the total least squares method (TLS) (Chaudhuri and Chatterjee, 1991) instead of least squares method. Once the pose parameters rotation matrix $\hat{\mathbf{R}}$ and translation vector $\hat{\mathbf{t}}$ have been estimated, the initial depth of each point is estimated as:

$$d_i = \frac{\mathbf{R}_1 \mathbf{p}_i + t_x}{X_i}$$

where \mathbf{R}_1 is the first row of the estimated rotation matrix $\hat{\mathbf{R}}$ and t_x is the first component of the estimated translation vector $\hat{\mathbf{t}}$. Once the depth of every point is approximately known, the reflected pole \mathbf{c} can be initially estimated as defined by Eq.(3). All estimates are then refined by iteration. Initialize iteration number $k=0$, set $\mathbf{c}=\mathbf{c}^{(k)}$, $d_i=d_i^{(k)}$ and set a desired accuracy threshold ρ for $\|\mathbf{c}^{(k)}-\mathbf{c}^{(k-1)}\|$, and set the number M of maximum iterations.

Step 2: Optimise the estimation of the reflected pole

Set iteration number $k \rightarrow k+1$. In order to optimise the estimation of the reflected pole $\mathbf{c}^{(k-1)}$, we know from Eq.(2) that:

$$(d_i^{(k-1)})^2 L_i + 2d_i^{(k-1)} \mathbf{P}'_i^T \mathbf{c}^{(k-1)} = l_i - 2\mathbf{P}'_i^T \mathbf{c}^{(k-1)}$$

where $L_i=\mathbf{P}'_i^T \mathbf{P}'_i$ and $l_i=\mathbf{p}_i^T \mathbf{p}_i$. As a result, we can construct the following objective function that optimises the estimation of the reflected pole \mathbf{c} :

$$F = \min_{\mathbf{c}} \sum_{i=1}^n ((d_i^{(k-1)})^2 L_i + 2(d_i^{(k-1)} \mathbf{P}'_i^T + \mathbf{p}_i)^T \mathbf{c} - l_i)^2 \quad (6)$$

Then the updated reflected pole $\mathbf{c}^{(k)}$ can be estimated in the least squares sense as:

$$\mathbf{c}^{(k)} = \mathbf{A}^{-1}\mathbf{b} \tag{7}$$

where $\mathbf{A} = 2\sum_{i=1}^n (d_i^{(k-1)}\mathbf{P}'_i + \mathbf{p}_i)(d_i^{(k-1)}\mathbf{P}'_i + \mathbf{p}_i)^T$ and $\mathbf{b} = \sum_{i=1}^n ((d_i^{(k-1)})^2 L_i - l_i)(d_i^{(k-1)}\mathbf{P}'_i + \mathbf{p}_i)$.

Step 3: Refining the estimation of pose and structure

Once a calibrated pole $\mathbf{c}^{(k)}$ has been obtained, the depth of every point can be refined using the following procedure. From Eq.(2), it is known that:

$$d_i^2 L_i + 2d_i \mathbf{P}'_i^T \mathbf{c}^{(k)} + (\mathbf{c}^{(k)})^T (\mathbf{c}^{(k)}) - \|\mathbf{p}_i - \mathbf{c}^{(k)}\|^2 = 0$$

Therefore, we have:

$$d_{i,1} = \frac{-f + \sqrt{\Delta}}{e} \quad \text{and} \quad d_{i,2} = \frac{-f - \sqrt{\Delta}}{e}$$

where $e = L_i$, $f = 2\mathbf{P}'_i^T \mathbf{c}^{(k)}$, $g = (\mathbf{c}^{(k)})^T (\mathbf{c}^{(k)}) - \|\mathbf{p}_i - \mathbf{c}^{(k)}\|^2$ and $\Delta = f^2 - eg$.

In order to clarify these multiple solutions, compute the distances:

$$\begin{aligned} s_1 &= \|d_{i,1}\mathbf{P}'_i - \bar{\mathbf{p}}'\| \\ s_2 &= \|d_{i,2}\mathbf{P}'_i - \bar{\mathbf{p}}'\| \\ s &= \|\mathbf{p}_i - \bar{\mathbf{p}}\| \end{aligned} \tag{8}$$

where $\bar{\mathbf{p}}'$ and $\bar{\mathbf{p}}$ are the centroids of the two sets of 3D points which can be computed as:

$$\bar{\mathbf{p}}' = \frac{1}{n} \sum_{i=1}^n d_i^{(k-1)} \mathbf{P}'_i \quad \text{and} \quad \bar{\mathbf{p}} = \frac{1}{n} \sum_{i=1}^n \mathbf{P}_i$$

If $\|s_1 - s\| < \|s_2 - s\|$ then $d_i^{(k)} = d_{i,1}$. Otherwise $d_i^{(k)} = d_{i,2}$.

Once the depth of every point has been updated, compare the reflected poles:

If $\|\mathbf{c}^{(k)} - \mathbf{c}^{(k-1)}\| < \rho$ or $k > M$, then the algorithm outputs the depth of every point, otherwise go to Step 2. In the experiments described in the next Section we used $\rho = 0.01$ and $M = 30$.

Step 4: Pose re-estimation

The result after the structure $\{\hat{d}_1, \hat{d}_2, \dots, \hat{d}_n\}$ of the object has been estimated is that the original 3D-2D camera pose estimation problem has been changed into a 3D-3D motion estimation problem where a number of accurate and robust algorithms can be applied. Because it had been demonstrated

that the constraint least squares method (CLS) (Arun et al., 1987; Umeyama, 1991) is the most accurate and robust algorithm (Eggert et al., 1997; Matei et al., 1998) when image data are corrupted by noise without outliers, this algorithm is adopted by the authors for re-estimation of the 3D-2D camera pose parameters rotation matrix $\hat{\mathbf{R}}$ and translation vector $\hat{\mathbf{t}}$.

Step 5: Structural re-estimation

Once the pose parameters rotation matrix $\hat{\mathbf{R}}$ and translation vector $\hat{\mathbf{t}}$ are determined, the structural parameters can be re-estimated as:

$$\hat{d}_i = \frac{1}{3} \left(\frac{\mathbf{R}_1 \mathbf{p}_i + t_x}{X_i} + \frac{\mathbf{R}_2 \mathbf{p}_i + t_y}{Y_i} + \mathbf{R}_3 \mathbf{p}_i + t_z \right)$$

where \mathbf{R}_2 and \mathbf{R}_3 are the second and third rows of matrix $\hat{\mathbf{R}}$ and t_y and t_z are the second and third components of translation vector $\hat{\mathbf{t}}$.

EXPERIMENTAL RESULTS

In order to validate and provide better understanding of the performance of the CFGA and PFLA algorithms, they were implemented and compared with an algorithm called Extended Tsai and Huang Algorithm (ETHA). The ETHA algorithm is an extended version (from 2D-2D to 3D-2D) proposed in (Rodrigues and Liu, 1999) a well known calibration procedure originally proposed by (Tsai and Huang, 1984) for 2D-2D problems based on epipolar geometry. Apart from the obvious reason of comparing algorithms based on fundamentally different geometrical concepts, the choice of the ETHA algorithm has also been justified by the fact that epipolar geometry represents a unique linear solution for all 3D-3D, 3D-2D, and 2D-2D motion calibration problems. Furthermore, it has the advantage of simplicity of implementation. A series of experiments using synthetic data and real images were carried out as described below.

Results with synthetic data

We randomly selected 30 points with uniform

distribution from a cube $[10,40] \times [10,40] \times [10,40]$. A rotation axis \mathbf{h} was randomly generated with uniform distribution from a cube $[1,3] \times [1,3] \times [1,3]$. The set of 30 points were then rotated from 3 to 81 degrees at 3 degree interval around the fixed rotation axis \mathbf{h} followed by a constant translation vector \mathbf{t} randomly generated with uniform distribution from a cube $[5,25] \times [5,25] \times [5,25]$. We thus, know the selected points and their correspondences

$$(\mathbf{p}_i, \mathbf{R}') = ((x_i, y_i, z_i)^T, (x'_i, y'_i, z'_i)^T)$$

where $(i=1,2,\dots,30)$, $z_i > 1$ and $z'_i > 1$ with the object being in front of the camera assuming that the focal length of the camera is equal to 1 for convenience of computation without loss of generality. We also know the pose parameters rotation matrix \mathbf{R} and translation vector \mathbf{t} . All points \mathbf{P}_i are then projected on the plane $z'=1$ and, as a result, we have also accurate perspective image data $\mathbf{P}'_i = (X'_i, Y'_i, 1)^T$. Once the data are generated, we can then apply the algorithms so that their performances can be objectively compared.

In order to simulate real world noise contaminated data, Gaussian random noise was added to the coordinates of each correspondence with mean equal to 0 and standard deviation $\sigma_1=0.002$ in one series of experiments and $\sigma_2=0.004$ in another series of experiments. Finally, we applied the proposed CFGA algorithm and PFLA and ETHA

methods to estimate the structure and pose parameters. The parameters of interest were the depth \hat{d}_i of each point, the rotation axis $\hat{\mathbf{h}}$, the rotation angle $\hat{\theta}$, and the translation vector $\hat{\mathbf{t}}$.

We define the relative calibration error of the rotation axis $\hat{\mathbf{h}}$ as $e_h = \|\hat{\mathbf{h}} - \mathbf{h}\|$, the relative calibration error of rotation angle $\hat{\theta}$ as $e_\theta = (\hat{\theta} - \theta)/\theta$, the relative calibration error of the translation vector $\hat{\mathbf{t}}$ as: $e_t = \|\hat{\mathbf{t}} - \mathbf{t}\| / \|\mathbf{t}\|$, and the relative calibration error of depth \hat{d}_i of each point as $e_d = (\hat{d}_i - z'_i)/z'_i$. Experimental results with synthetic data are described in Figs.1, 2, 3, and 4 and Table 1. In all figures, the solid lines correspond to lower levels of noise and the dashed lines correspond to higher level of noise. In order to provide better visualization of calibration errors, we fixed the relative calibration error to 25% if its actual value is larger than 25%.

Figs.1, 2, 3, and 4 and Table 1 show that the CFGA algorithm is the most accurate and robust of the compared algorithms for noisy image data. The CFGA's relative calibration errors of the rotation axis, translation vector, and structure are 2.6708%, 3.1134%, and 0.3542% respectively, while the same parameters are 5.8278%, 6.3832%, and 0.4127% for the PFLA algorithm and 5.0905%, 15.4135%, and -1.5198% for the ETHA algorithm

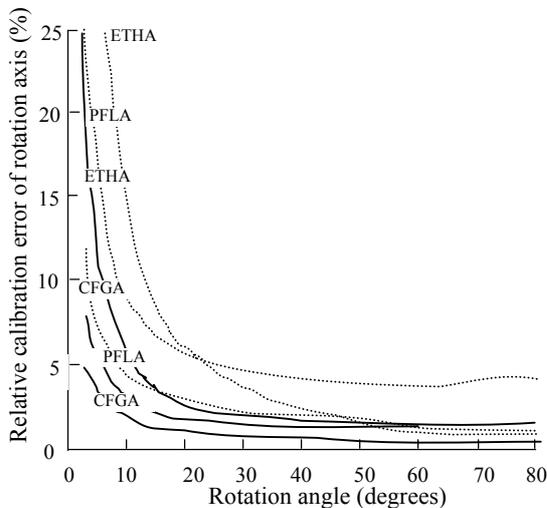


Fig.1 Relative calibration error of rotation axis over a range of rotation angles

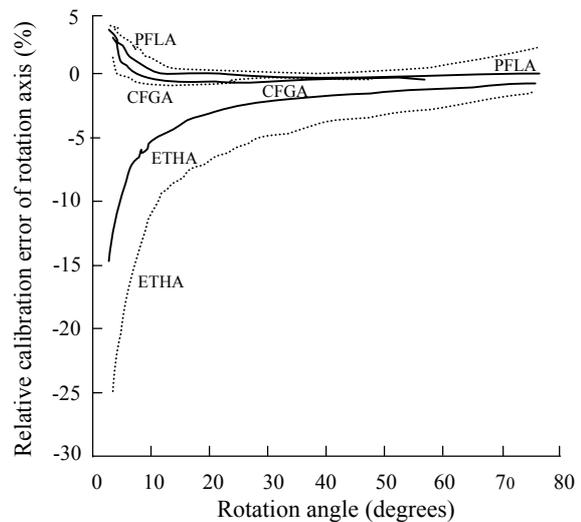


Fig.2 Relative calibration error of the rotation angle over a range of rotations

when image data are corrupted by higher level of noise as shown in this paper. More importantly, the results showed that the accuracy of the CFGA algorithm is at least a factor of 2 better than the accuracy of the PFLA for the calibration of rotation axis and translation vector. This shows that the refinement of the classical linear algorithm as proposed is effective and yields satisfactory calibration results. This improvement in accuracy is mainly because the CFGA algorithm makes full use of the rigid constraints between points described in different coordinate frames before and after the motion as described by Theorem 1. In addition, due

to the good behavior of the constraint least squares method in combating noise, the PFLA algorithm does not take these rigid constraints into consideration. Similarly, the ETHA algorithm just considers the constraint of co-planarity resulting in a significantly inferior accuracy when compared to the CFGA algorithm.

Table 1 shows that the calibration of translation vector is more sensitive to noise than the calibration of the rotation parameters rotation axis and rotation angle as pointed out by Tsai and Huang (1984).

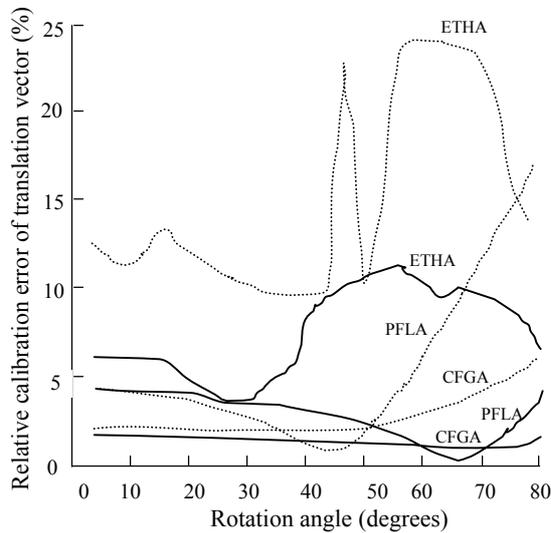


Fig.3 Relative calibration error of the translation vector over a range of rotation angles

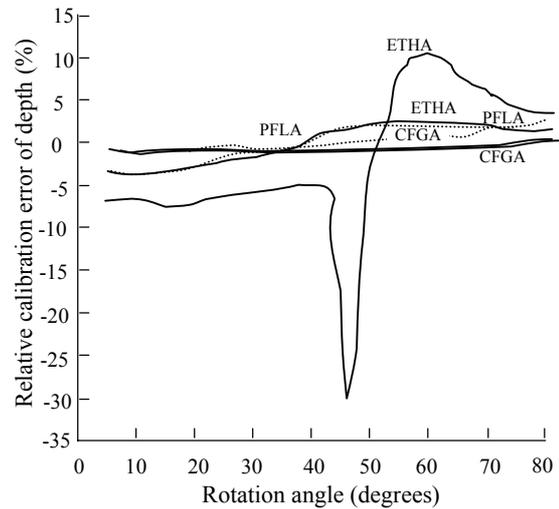


Fig.4 Relative calibration error of the structure (depth estimation) over a range of rotation angles

Table 1 Comparative errors for synthetic data corrupted by two levels noise (σ_1, σ_2)^{*}

Noise	Meas.	Algorithm	\hat{h}	$\hat{\theta}$	\hat{t}	\hat{d}
σ_1	Meam	PFLA	2.1361	-0.0782	3.1236	-0.3451
		CFGA	1.3006	0.1426	1.6465	-0.3240
		ETHA	2.6464	-2.7757	7.7536	0.2976
	σ	PFLA	1.3380	0.7923	1.2717	0.4547
		CFGA	0.8767	0.3151	0.2225	0.4129
		ETHA	4.5219	3.0074	2.5654	2.3756
σ_2	Meam	PFLA	5.8278	0.2250	6.3832	0.4127
		CFGA	2.6708	0.3856	3.1134	0.3542
		ETHA	5.0905	-5.7341	15.4135	-1.5498
	σ	PFLA	4.4655	1.2120	5.5864	1.2194
		CFGA	2.1759	0.6885	1.5313	1.1022
		ETHA	6.5809	5.4615	5.7271	8.4562

^{*}The mean and standard deviation σ of the calibration errors in percentage for rotation axis \hat{h} , rotation angle $\hat{\theta}$, translation vector \hat{t} , and depth estimation \hat{d}

Results with real images

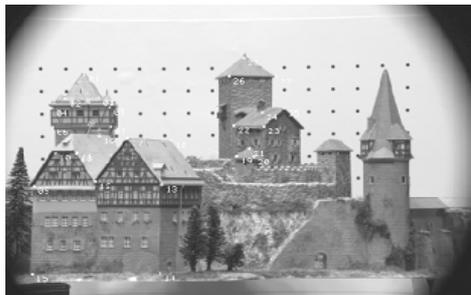
The image data as depicted in Fig.5 were downloaded from the image database of Calibrated Imaging Laboratory at Carnegie Mellon University. The three algorithms were applied to the data and experimental results are summarized in Tables 2 and 3.

Tables 2 and 3 show that the CFGA algorithm is the most accurate for the calibration of rotation angle and translation vector, while the CFGA's accuracy for the calibration of the rotation axis is similar to both PFLA and ETHA algorithms. Table 3 shows that the calibration of rotation parameters rotation axis and rotation angle for all algorithms is poor. This can be explained as follows. From Table 2 and real images Fig.5, it is known that the camera motion is nearly pure translational which results in all off diagonal elements of the reference matrix to be small, rendering the estimation of rotation parameters rotation axis and rotation angle very sensitive to noise. This means that a small difference of calibrated parameters can cause large relative calibration errors. Therefore, the relative calibration error of the translation vector is a reasonable measurement in this case as a measure of compari-

son of which algorithm is the most accurate. From this viewpoint, it is known that the CFGA algorithm is the most accurate among the CFGA, PFLA, and ETHA algorithms.

SUMMARY AND CONCLUSIONS

In this paper, after our investigation of geometric properties of reflected correspondences, we put forward the new concept of reflected pole of a rigid transformation. The reflected pole has useful geometric properties that can be used as constraint for calibration of transformation parameters. For a planar transformation (the 2D case) all reflected correspondences (\mathbf{p}, \mathbf{p}'') must lie on concentric circles centered at the reflected pole of that planar transformation and, for a spatial transformation (the 3D case) all reflected correspondences must lie on the surface of concentric spheres centered at the reflected pole of that spatial transformation. This analysis is thus, more general than the screw motion analysis as it is also valid for pure translational motions. A pure translational motion is a singular case of the screw motion analysis where the motion



(a)



(b)

Fig.5 Real castle images: (a) 3D description; (b) projective image

Table 2 Calibrated rotation matrix \hat{R} and translation vector \hat{t}

Method	\hat{R}			\hat{t}
Ref	0.9999	0.0111	0.0076	-566.7595
	-0.0111	0.9999	-0.0006	-516.2556
	-0.0076	0.0005	0.9999	1767.1555
PFLA	0.9990	0.0111	0.0081	-566.3399
	-0.0128	1.0025	-0.0001	-516.9812
	0.2268	-0.0901	0.9679	1695.0774
CFGA	0.9986	-0.0158	-0.0510	-540.2255
	0.0157	0.9999	-0.0008	-528.2741
	0.0510	0.0000	0.9987	1717.2342

Table 3 Relative calibration error of rotation axis \hat{h} , rotation angle $\hat{\theta}$, and the translation vector \hat{t}

Methods	Error \hat{h} (%)	Error $\hat{\theta}$ (%)	Error \hat{t} (%)
PFLA	170.4583	1193.6572	3.7421
CFGA	188.8157	295.2693	3.0005
ETHA	196.9993	355.2521	17.4422

must then be classified as planar or spatial.

Based on such rigid constraint, we then show how it can be applied to the calibration of transformation parameters. We proposed and demonstrated an iterative method to refine structural estimation. The method is based on a coarse-to-fine strategy where an initial estimation is obtained through a classical linear algorithm. The method first transforms the 3D-2D camera pose estimation problem to a 3D-3D motion estimation problem and the constraint least squares method is then applied to camera pose estimation. For a comparative study of performance, we also implemented an extended 3D-2D pose estimation procedure (Rodrigues and Liu, 1999) based on a well known 2D-2D motion estimation algorithm (Tsai and Huang, 1984). Experimental results based on both synthetic data and real images showed that the overall performance of the proposed iterative algorithm is superior to both the classical linear algorithm and the extended algorithm. This demonstrates that the proposed method is accurate and effective for calibration of structural and pose parameters and, because of its generality, robustness and simplicity, the method is useful for real world calibration problems.

References

- Arun, K.S., Huang, T.S., Blostein, S.D., 1987. Least-squares fitting of two 3-D point sets. *IEEE Trans. PAMI*, **9**:698-700.
- Ball, R.S., 1900. *A Treatise on the Theory of Screws*. Cambridge University Press.
- Carceroni, R.L., Brown, C.M., 1997. Numerical Methods for Model-based Pose Recovery. Technical Report 659, Computer Science Department, The University of Rochester, Rochester, New York.
- Chaudhuri, S., Chatterjee, S., 1991. Performance analysis of total least squares methods in three-dimensional motion estimation. *IEEE Trans. Robotics and Automation*, **7**:707-714.
- Dornaika, F., Garcia, C., 1999. Pose estimation using point and line correspondences. *Real Time Imaging*, **5**:215-230.
- Eggert, D.W., Lorusso, A., Fisher R.B., 1997. Estimating 3-D rigid body transformations: a comparison of four major algorithms. *MVA*, **9**:272-290.
- Faugeras, O.D., Hebert, M., 1986. The representation, recognition, and locating of 3-D objects. *Int. J. of Robotics Research*, **5**:27-52.
- Haralick, R.M., Joo, H., Lee, C.N., Zhuang, X., Vaidya V.G., Kim, M.A., 1989. Pose estimation from corresponding point data. *IEEE Trans. PAMI*, **9**:1426-1441.
- Hattori, K., Matsumori, S., Sato, Y., 1998. Estimating Pose of Human Face Based on Symmetry Plane Using Range and Intensity Images. Proc. of Int. Conf. Pattern Recognition, p.1183-1187.
- Huang, T.S., Netravali, A.N., 1994. Motion and structure from feature correspondence: a review. *Proceedings of the IEEE*, **82**:252-268.
- Linnainmaa, S., Harwood, D., Davis, L.S., 1987. Pose determination of a three-dimensional object using triangular pairs. *IEEE Trans. PAMI*, **10**:634-647.
- Liu, Y., Rodrigues, M.A., 1999. Geometric Understanding of Rigid Body Transformations. Proc. IEEE Int. Conf. on Robotics and Automation, p.1275-1280.
- Lowe, D.G., 1991. Fitting parameterized three-dimensional models to images. *IEEE Trans. PAMI*, **13**:441-450.
- Matei, B., Meer, P., Tyler, D., 1998. Performance Assessment by Resampling: Rigid Motion Estimators. IEEE CS Workshop on Empirical Evaluation of Computer Vision Algorithms, Santa Barbara, California, p.72-95.
- McCarthy, J.M., 1990. *Introduction to Theoretic Kinematics*. The MIT Press.
- Mitiche, A., Aggarwal, J.K., 1986. A Computational Analysis of Time-varying Images. In: T.Y. Yuong, K.S. Fu, eds., *Handbook of Pattern Recognition and Image Processing*, Academic Press.
- Quan, L., Lan, Z., 1999. Linear N-point camera pose determination. *IEEE Trans. PAMI*, **21**:774-780.
- Rodrigues M.A., Liu, Y., 1999. A Comparative Study of Some 3D-2D Computer Vision Algorithms. Proc. the ISCA 14th Int. Conf. on Computers and Their Applications (CATA'99), Cancun, Mexico, p.170-173.
- Toyama, F., Shoji, K., Miyamichi, J., 1998. Model-based Pose Estimation Using Genetic Algorithm. Proc. of Int. Conf. Pattern Recognition, p.198-201.
- Tsai, R.Y., Huang, T.S., 1984. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Trans. PAMI*, **6**:13-27.
- Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point pattern. *IEEE Trans. PAMI*, **13**:377-380.
- Yamane, S., Izumi, M., Fukunaga, K., 1996. A method of model based pose estimation. *Trans. IEICE D-II*, **J79-D-II**, **2**:165-173.