Pose Estimation for Multiple Camera Systems

Wen-Yan Chang and Chu-Song Chen
Institute of Information Science, Academia Sinica, Taipei, Taiwan
{wychang, song}@iis.sinica.edu.tw

Abstract

Pose estimation of a multiple camera system (MCS) is usually achieved by either solving the PnP problem or finding the least-squared-error rigid transformation between two 3D point sets. These methods employ partial information of an MCS, in which only a small number of features in one or two cameras can be utilized. To overcome this limitation, we propose a new pose estimation method for an MCS that uses complete information of an MCS. In our method, we treat the MCS as a single generalized camera [7][14] and formulate this problem in a least-squared manner. An iterative algorithm is proposed for solving the least-squared problem. From the experimental results, it shows that the proposed method is accurate for pose estimation of MCS.

1. Introduction

In recent years, multiple camera systems (MCS) have received much attention in computer and robot vision. For example, a multi-camera network [6] was designed as a new "eye" to achieve better robustness for structure from motion. In [13], a GlobalAll camera array was presented, which can provide electronic pan-tilt-zoom effect for applications of vision-based intelligent room. In [11], a camera matrix was proposed for achieving better stereo matching and 3D reconstruction. In essence, MCS is referred to as a camera set consisting of at least two cameras. These cameras are mounted on some rigid objects and thus their relative positions and orientations are fixed.

This paper deals with the pose-estimation problem for MCS. Pose estimation, referred to as computing the rotation and translation of a visual sensor relative to known geometry, is a fundamentally important problem for computer and robot vision. In fact, almost all the existing pose-estimation methods were designed for a single camera but not a camera set, and it still lacks of systematic methods for solving their pose-estimation problem. In this paper, a generally effective method is proposed for pose estimation of MCS.

Consider a set of cameras with fixed and known translations and orientations to each other. Without lost of generality, we can set an MCS coordinate system to which all the cameras refer, and an illustration is shown in Fig. 1. The MCS pose-estimation problem is formally described below:
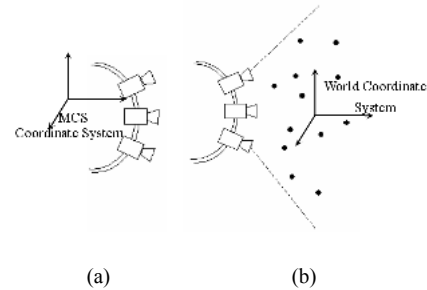

Fig 1. An illustration of a MCS. (a) A set of mounted cameras with known and fixed translations and orientations each other and a MCS coordinate system is defined in these cameras. (b) A MCS pose estimation problem is defined as finding the rigid transformation between the MCS coordinate system and the world coordinate system.

[MCS Pose Estimation] *If there are several (at least three) 3D points in the environment with known coordinates with respect to a world coordinate system, the MCS pose estimation problem is defined as finding the rotation and translation between the MCS and the world coordinate systems.*

2. Possible Ways for Pose Estimation of MCS and Their Limitations

In this section, we discuss some possible ways that can be used to solve the pose estimation problem of an MCS by applying existing methods. One way is to exploit the least-squared-error rigid transformation from two point sets, which was thoroughly investigated in the computer/robot vision community [1][5][8][10], as introduced in the following. Let P_1, P_2, ..., P_n be n points with known coordinates to a world coordinate system. Consider an MCS consisting of m cameras that takes m images, I_1, I_2, ..., I_m, respectively. Assume that there are n' points commonly appear in some m' images, where $2 \leq m' \leq m$ and $3 \leq n' \leq n$. Without lost of generality, let these n' points be P_1, P_2, ..., $P_{n'}$. Since the cameras have been calibrated and represented w.r.t. a unified MCS coordinate system and $m' \geq 2$, the m' images form a multi-view stereo image set. The 3D coordinates of all the n' points can therefore be computed in association with the MCS coordinate system via simple triangulation (or equivalently, ray intersection). Then, consider that these n' points also have known coordinates w.r.t. the world coordinate system. The MCS pose estimation-problem can thus be transformed as finding the rotation R and translation t such

that $Q_i = RP_i+t$, $i = 1, \ldots, n'$, where P_i and Q_i are the coordinates of these n' points w.r.t. the world and MCS coordinate systems, respectively.

Due to image noises in practice, we can not find (R, t) satisfying $Q_i = RP_i+t$ for all $i = 1, \ldots, n'$ but usually find the one with least sum of squared errors instead:

$$E = \sum_i \|Q_i-(RP_i+t)\|^2. \tag{1}$$

When $n' \geq 3$, the minimization E in (1) has closed-form solutions [1][5][8][10], and the MCS pose (associated with the world coordinate system) can then be estimated by solving (1).

However, there are two drawbacks for solving the MCS pose-estimation problem in this way. First, this approach can be used only when there exist at least three commonly visible points for some pair of images (i.e., $n' \geq 3$ and $m' \geq 2$). If all the pairs of images have less than three commonly visible points, this method can not be applied. Second, even when $n' \geq 3$ and $m' \geq 2$ are satisfied, only partial information of a MCS is used for estimating the pose because n' may be smaller than the number of world points, n. Intuitively speaking, information in association with the remained $n-n'$ points and $m-m'$ images are wasted.

Another way to solve the MCS pose-estimation problem via existing approaches is to use the solution of the perspective-n-point (PnP) problem [4][6][9][12][15]. In our case, given n'' points ($n'' \geq 3$) and their projecting points in some image I_i, $i = 1, \ldots, m$, the solution of the PnP problem is the rotation and translation between the i-th camera coordinate system and the world coordinate system. Since the MCS has been calibrated, the rotation and translation between the MCS and the i-th camera coordinate systems ($i = 1, \ldots, m$) is fixed and known. By composing the two pairs of rotation and translation above, the rotation and translation between the world and MCS coordinate systems can then be obtained.

However, solving the MCS pose-estimation problem in this way also suffers from the same drawbacks. First, the applicable condition is limited because such a method can not be used when no images have sensed three world points, even the total number of world points is indeed sufficient for determining the MCS pose. Second, the solution obtained in this way can not fully exploit the information useful for pose estimation because only a single image is used.

In this paper, we propose a method that exploits all the world points for pose estimation of an MCS. Our method can also be used when all images have sensed less than three world points or all pairs of images have less than three commonly sensed points, as long as the number of world points, n, is at least three. Details can be found in the next section.

3. The Proposed Method

Consider a world point P_i ($i = 1, \ldots, n$), P_i is visible (or sensible) in the image I_j means that there is a point p_{ji} in the image plane of the j-th camera satisfying that L_{ji}, the ray from O$_j$ to P_{ji}, passes through P_i, where O$_j$ is the focal center of the j-th camera, $j = 1, \ldots, m$, and L_{ji} is represented w.r.t. the world coordinate system. Let $\mathbf{A}$ be the set consisting of the world points that are visible in more than one image and $\underline{\mathbf{A}}$ be the set of world points that is visible in only one image. For each point P_i contained in $\mathbf{A}$, the number of rays associated with it is more than one, and its coordinate w.r.t. the MCS coordinate system can be computed from the intersection point of all the associated rays. On the contrary, the coordinate w.r.t. the MCS coordinate system can not be estimated for the points contained in $\underline{\mathbf{A}}$.

According to the definitions above, we discuss the MCS pose-estimation in the following two situations. First, consider a world point $P \in \mathbf{A}$. Assume its estimated coordinate w.r.t. the MCS coordinate system to be S. The objective function in the case of $P_i \in \mathbf{A}$ is defined as

$$E_A = \min\sum_i \|S_i-(RP_i+t)\|^2. \tag{2}$$

Second, we consider the situation when a world point $P \in \underline{\mathbf{A}}$. In this case, there is a ray L passing through it and this ray can be represented by a vector (c, v) w.r.t. the MCS coordinate system, where c is a point passed by this ray and v is the direction of this ray. Thus, the projection of P on L can be written as $Proj(P;L) = vv^T(RP+t-c)+c$. The distance between P and L is

$$\|Proj(P;L)-P\| = \|(vv^T-I)(RP+t-c)\|, \tag{3}$$

where I is the 3 by 3 identity matrix. The objective function in the case $P_i \in \underline{\mathbf{A}}$ is defined as

$$E_{\underline{A}} = \min\sum_i \|(v_i v_i^T-I)(RP_i+t-c_i)\|^2. \tag{4}$$

By combining the two situations above, the objective function minimized in our approach is $E = E_A+E_{\underline{A}}$:

$$E = \min[\sum_{P_i \in A, i=1,\ldots|A|} \| S_i - (RP_i + t)\|^2$$
$$+ \sum_{P_k \in \underline{A}, k=1,\ldots|\underline{A}|} \| (v_k v_k^T - I)(RP_k + t - c_k)\|^2],$$

where $|\mathbf{A}|$ and $|\underline{\mathbf{A}}|$ are the cardinalities of $\mathbf{A}$ and $\underline{\mathbf{A}}$, respectively. In the above, the points contained in $\mathbf{A}$ provide *point-to-point correspondence information*, and those contained in $\underline{\mathbf{A}}$ provide *point-to-line correspondence information*. Our algorithm use both kinds of

correspondence information for solving the MCS pose-estimation problem, as presented below:

Algorithm:

Step 0. Let $(\boldsymbol{R_0}, \boldsymbol{t_0})$ be the transformation initially given.

Step 1. Compute $P^*_i = \boldsymbol{R_0}P_i + \boldsymbol{t_0}$ for all $i = 1, ..., n$.

Step 2. Find the closest point P'_i of P^*_i.

   2.1. If $\boldsymbol{P}_i \in \mathbf{A}$, set $P'_i = S_i$ for all $i = 1, ...,n$, where S_i is the intersection point of the associated rays.

   2.2. If $\boldsymbol{P}_i \in \underline{\mathbf{A}}$, set $P'_i = v_i v_i^T(P^*_i - c_i) + c_i$ for all $i = 1, ...,n$, where vector (c_i, v_i) is associated with a ray $\boldsymbol{L}$ that passes through $\boldsymbol{P}_i$.

Step 3. Find the least-squared-error rigid transformation [1] between P^*_i and P'_i, $i = 1, ...,n$. That is, find $(\boldsymbol{R_{new}}, \boldsymbol{t_{new}})$ that minimizes $\sum_i \|(\boldsymbol{R_{new}}P^*_i + \boldsymbol{t_{new}} - P'_i)\|^2$.

Step 4. If $\boldsymbol{R_{new}}$ is close enough to the identity matrix and $\boldsymbol{t_{new}}$ is close enough to the zero vector, then **stop**; Else compose $(\boldsymbol{R_{new}}, \boldsymbol{t_{new}})$ and $(\boldsymbol{R_0}, \boldsymbol{t_0})$ by $\boldsymbol{R_0} \leftarrow \boldsymbol{R_{new}}\boldsymbol{R_0}$, $\boldsymbol{t_0} \leftarrow \boldsymbol{t_{new}} + \boldsymbol{R_{new}}\boldsymbol{t_0}$, and go to *step 1*.

By using this algorithm, all the world points can be employed. In fact, the proposed algorithm is inspired from both the iterative-closest-point algorithm [2] and NPnP problem [3]. The monotonic convergent property of these algorithms is ensured. Hence, similar convergent properties also hold for our algorithm.

In the initialization (*Step 0*) of our method, three strategies are suggested to obtain a good initial:

(1) If available, pick up three points from **A** randomly and calculate an initial rigid transformation via least-squared-error transformation [1].

(2) If more than three world points can be seen in one of the cameras in MCS, the associated three rays can be used to estimate an initial pose by solving the P3P problem [15].

(3) We also can regard a MCS as a non-perspective imaging device, and the method for initializing the non-perspective n point problem [3] can be adopted.

Among them, strategy 3 is applicable as long as the number of world points, n, is at least three. Hence, strategy 3 is employed in our experiments.

4. Experimental Results

In our experiment, a MCS consists of three CCD cameras with overlapped field-of-views is used, as shown in Fig. 2.

There are three images captured simultaneously in our MCS. The intrinsic model of this MCS is calibrated in advance, so that every pixel in the MCS image set corresponds to a ray w.r.t. the MCS coordinate system. We place this MCS in two positions of an indoor environment to construct a stereo pair, and some 3D fiducial marks are measured in advance for pose estimation. Then, the poses of the MCS are estimated for these two placements via the proposed method. In Fig. 3, two sets of MCS image captured from these two placements are shown and the red points in Fig. 3 are the fiducial marks for pose estimation.

According to epipolar geometry, the correspondence in one MCS images should lie on a line in the other MCS image set. In Fig. 3(a), we pick up 8 points in one MCS image set. If no errors occur, the corresponding epipolar lines will pass through the corresponding points in the other MCS image set. From Fig. 3(b), it shows that our method is very accurate so that these epipolar lines all pass through the corresponding points.

In addition, a stereo pair of MCSs can help us compute the 3D coordinate of any other point in this environment if its corresponding point is identified. We use this stereo pair to compute the coordinates of some 3D points. Table I lists the relative errors measured for some length ratios, where line 0 shown in Figure 4 serves as the unit length. The relative error is the ratio of the difference between real and estimated values to the real value.

5. Conclusion

In this paper, we propose a pose estimation method for MCS to improve the existing methods. In our method, all the fiducial marks can be used for pose estimation. To obtain a good initial estimate, some initialization strategies are suggested. Then, an iterative algorithm is proposed to obtain a least-squared solution of a MCS. From the experimental results, it shows that the proposed method is accurate.

References
[1] K. S. Arun, T. S. Huang and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. Pat. Ana. & Mach. Int*., Vol. 9, No. 5, 1987.

TABLE I.   Errors for some length ratios.

	Real ratio	Estimated ratio	Relative Error
Case 1	0.8430	0.8554	0.014709
Case 2	0.1109	0.0995	0.102795
Case 3	0.0693	0.0654	0.056277
Case 4	0.1109	0.1270	0.145175
Case 5	0.2219	0.2491	0.122577

[2] P. J. Besl and N. D. McKay, "A method for registration of 3D shapes," *IEEE Trans. Pat. Ana. & Mach. Int.*, Vol. 14, No. 5, 1992.

[3] C. S. Chen and W. Y. Chang, "On Pose Recovery for Generalized Visual Sensors," *IEEE Trans. Pat. Ana. & Mach. Int.* (pending publication).

[4] D. F. Dementhon and L. S. Davis, "Model-based object pose in 25 line of code," *Int. J. Computer Vision*, Vol. 15, pp. 123-141, 1995.

[5] O. D. Faugeras and M. Hebert, "A 3D recognition and positioning algorithm using geometrical matching between primitive surfaces," *Proc. Int. Joint Conf. Artificial Intelligence*, pp. 996-1002, 1983.

[6] C. Fermüller, et al, "Multi-camera networks: Eyes from eyes," *Proc. IEEE Workshop on Omnidirectional Vision*, pp. 11–18, 2000.

[7] M.Grossberg and S. Nayar, "A general imaging model and a method for finding its parameters," *Proc. Int. Conf. Computer Vision,* 2001.

[8] T. S. Huang, S. D. Blostein, and E. A. Margerum, "Least-squares estimation of motion parameters from 3D point correspondences," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 1986.

[9] R. M. Haralick, C. N. Lee, K. Ottenberg and M. Nolle, "Analysis and solutions of the three point perspective pose estimation problem," *Proc. IEEE Int. Conf. Computer Vision & Pattern Recognition*, 1991.

[10] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Am.*, Vol. 4, pp. 629-642, 1987.

[11] T. Kanade, P. Rander, and P. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes," *IEEE MultiMedia*, Vol. 4, No. 1, pp 34-47, 1997.

[12] C. P. Lu, G. D. Hager and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Trans. Pat. Ana. & Mach. Int.*, Vol. 22, No.6, 2000.

[13] M. Nicolescu and G. Medioni, "GlobeAll: Panoramic video for an intelligent room," *Proc. Int. Conf. Pattern Recognition*, pp. 823-826, 2000.

[14] R. Pless, "Using many cameras as one," " *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2003.

[15] J. S. C. Yuan, "A General Photogrammetric Method for Determining Object Position and Orientation," *IEEE Trans. Robotics and Automation*, Vol. 5, pp. 129-142, 1989.

Fig 2. The MCS used in our experiment that consists of three CCD cameras.
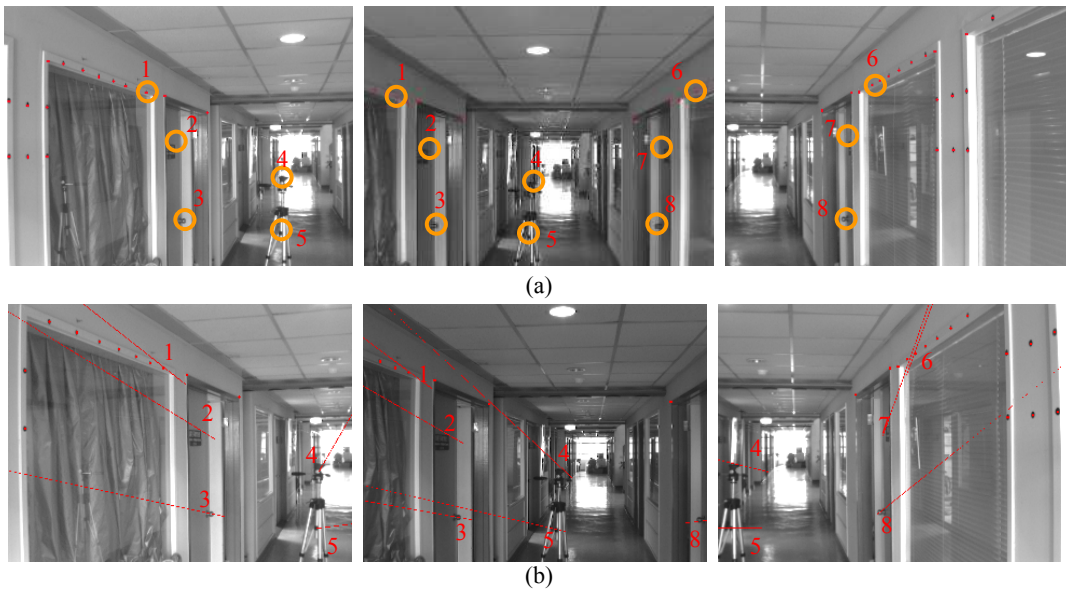

(a)

(b)

Fig 3. The image sets captured by our MCS from two placements in an indoor environment, the red points are the fiducial marks used for pose estimation. (a) Eight points selected from one MCS image set. (b) The corresponding epipolar lines are shown in the other MCS image set.

Fig 4. Experiment on the Length-ratio. The red lines are used for the length-ratio results, and line 0 is the unit length.