IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 28, 2019, accepted August 29, 2019, date of publication September 2, 2019, date of current version October 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939020

Simultaneous Pose and Correspondence Determination Combining Softassign and Orthogonal Iteration

HANG DONG[1], CHANGKU SUN[1,2], BAOSHANG ZHANG[1,2], AND PENG WANG[1,2]
[1]State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China
[2]Science and Technology on Electro-Optic Control Laboratory, Louyang Institute of Electro-Optic Equipment, Luoyang 471009, China

Corresponding author: Peng Wang (wang_peng@tju.eud.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51875407, and in part by the Aeronautical Science Foundation of China under Grant 20175148006.

ABSTRACT Pose estimation with unknown correspondences between 3D object points and 2D image points is known as the simultaneous pose and correspondence determination problem in the field of computer vision. It currently is still diffcutlt to solve particularly with the appearance of occlusion and cluster. In this paper, we present a new iterative algorithm for the pose estimation of an 3D object without any additional 3D-2D corresondence infromation. Our method combines SoftAssign algorithm for derterming the correspondence and OI (orthogonal iteration) algorithm for computing the pose. An assignment matrix which describes the correspondence is first introduced to the objective function of OI algorithm, and the simultaneous pose and correspondence determination problem is formulated as that of minimizing the weighted object space collinearity error. The pose and correspondence are evolved from an initial pose to an optimum value by nesting the two algorithms into one deterministic annealing process. Simulation and experimental results demonstrate that the proposed method is computationally more efficient and more accurate than the state-of-art methods.

INDEX TERMS Pose estimation, orthogonal iteration, correspondence determination, SoftAssign.

I. INTRODUCTION

Pose estimation is widely used in virtual reality [1], unmanned air vehicle navigation [2], aircrafts docking [3] and robot control [4]. In industrial measurement, point-based pose estimation is computing the pose of an object through 3D known geometry and the corresponding 2D image, which is also known as the PnP (perspective-n-point) problem [5]. With the known 3D-2D correspondence, the PnP problem is aimed at computing the rotation and the translation of the object coordinates system with respect to the camera coordinates system. In recent years, numerous effective methods have been put forward to solve this problem [6]–[12].

One thing should be noticed that all the PnP algorithms work only if the 3D-2D correspondences are determined in advance. However, the premise is often difficult to achieve because the image points are usually hard to be distinguished from each other due to their similar 3D geometry features. For example, in order to estimate the pose of a helmet (Fig.1(a)), some LED points with known coordinates are mounted on it as the object feature points, and the image captured by an infrared camera from a certain angle is shown in Fig.1 (b). Without any additional information, it is difficult to find the one-to-one 3D-2D correspondences. Besides, the occlusion and cluster case should also be taken into considered since they are inevitable in actual measurement, which will make the correspondence determination more complicated:

1) Occlusion: occlusion happens when part of the object points is obscured or outside the field of view due to large rotation angle, so some object points are not detected as the image features of that object, as is shown in Fig.1(c).
2) Cluster: Noise and imperfect image extraction algorithm are the main reasons for cluster. The appearance of cluster will cause some features which do not belong to the object to be mistakenly detected as the image features. As is shown in Fig.1 (d), the points in the red circle are some interference points.

The associate editor coordinating the review of this article and approving it for publication was Baozhen Yao.

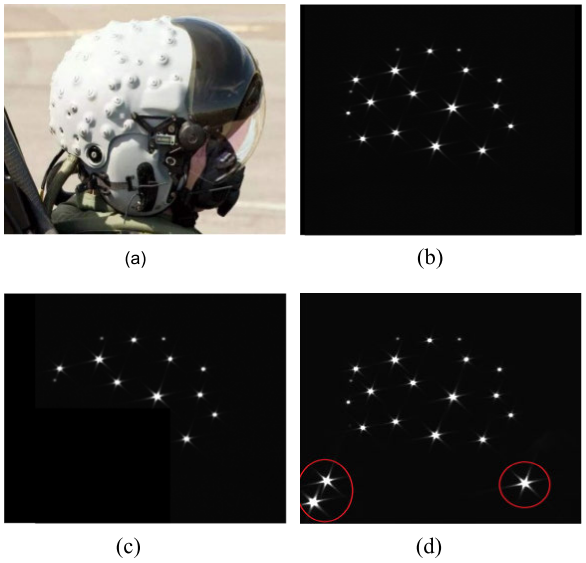


FIGURE 1. Confused correspondences between object points and image point. (a) An object (helmet) with some LED feature points mounted. (b) The image of the object. (c) Occlusion case (d) cluster case.

If correspondences are not well determined, it will lead to the invalidation of the subsequent PnP algorithm and the incorrect pose estimation result [13]. Therefore, pose problem and correspondence problem are closely connected, and one needs to design a method to estimate the pose with unknown correspondences, which is the simultaneous pose and correspondence determination problem [14]. When the initial pose is not far away from the true pose, the SoftPosit [12] algorithm is still the most efficient algorithm so far.However, it entails trying too many guesses before finding a proper initial pose, which will affect the speed of the algorithm [15]. Besides, SoftPosit cannot handle coplanar case because the scaled orthographic projection matrix in the embedded Posit algorithm will be singular if the geometry features of the object are coplanar. For these reasons, the goal of our work is to design a more accurate, faster and more widely used algorithm to estimate the pose of an object from non-corresponding points.

A. RELATED WORK

The RANSAC [5] which is based on hypothesize-and-test principle is a classic solution to the simultaneous pose and correspondence determination problem. In the RANSAC algorithm, random samples consisting of 3D object features and 2D image features with hypothesized correspondences are selected, and the pose is computed by the PnP methods. Then the object features will be projected back onto the image plane according to computed pose data. The correctness of the pose depends on whether the back-projected image is sufficiently similar to the original. If it is similar enough, the pose will be accepted, or another new correspondence will be hypothesized and the process will be repeated. A more effective solution similar to RANSAC is Blind PnP [15], which establishes different Gaussian mixture models (GMM) for different pose priors, reduces the incorrect potential matches rapidly and makes the search process more targeted. Other solutions are also proposed in [16]–[19] to improve the RANSAC. These hypothesize-and-test approaches have a high success rate, but at a heavy computation cost.

Different from the hypothesize-and-test methods, the EvoPose [20] which is based on Genetic Algorithm constructs an objective function of reprojection errors according to the perspective projection model, and evolves the candidate pose solution (the six-dimensional array composed of rotation and translation) by genetic operators. After some generations, good pose solutions emerge toward optimality. Xia *et al.* [21] enhances the evolutionary algorithms with a new efficient scheme: the candidate solution is evolved only when the offspring is better than the parent, so the survival probability of good pose offspring is increased, which will improve the efficiency. Generally, these Genetic algorithms are accurate when the object points are not very close to each other, but it takes too much time during the evolution process from one generation to another, and the perspective projection model is not robust to the image noise.

Some other methods which formulate the cost function by different projection model are also proposed [12], [22], [23]. Among them, SoftPosit [12] stands out because its success rate and efficiency. The SoftPosit algorithm integrates the iterative correspondence determination technique named Softassign [25], [26] and the iterative pose estimation algorithm named POSIT [27] into one single iteration loop.All potential matches are treated equally during the search process of the optimal pose instead of the hypothesize-and-test schedule. The pose and correspondence are determined simultaneously by minimizing a global objective function based on a weak perspective model in a deterministic annealing process. The time complexity of SoftPosit is $O(NM^2)$ (N and M are the number of 3D object points and 2D image points, respectively), which is still the most efficient so far.

Haoyin *et al.* [24] proposed SoftSI algorithm that combines two SVD (singular value decomposition)-based shape description theorems and a new proposed PnP algorithm named SI [24]. Their mehtod can elimate the bad initial values quickly accroding to the standard deviation of the translation vector at the first few iteration steps, which accelerate the process of finding a proper initial value for the iterative algorithm. The SoftSI algorithm is fast and robust to noise for the cases without occlusion and cluster.

Inspired by the work of SoftPosit, we propose a new pose estimation with unknown correspondence algorithm named SoftOI. For determining the correspondence, Softassign algorithm is also employed in our method; for computing the pose, different from SoftPosit, we combine the Softassign algorithm with another more accurate and efficient iterative pose estimation algorithm named OI (orthogonal iteration) [10]. The weihted space collinearity error is first introduced to describe the matching relations beteen the image points and the object points. During the deterministic annealing process, the pose and correspondence are determined alternately

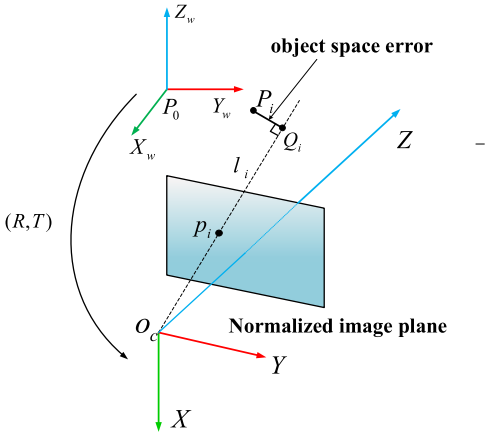


FIGURE 2. Geometry of camera model.

through minimizing the global weighted object space collinearity error at each iteration step. Experiment results shows that the proposed algorithm can make the simultaneous pose and correspondence determination problem be solved in a more accurate and efficient way even occlusion and cluster cases are taken into considertion. Besides, our mehtod has a broder application range in which both the coplanar and noncoplanar cases can be handled.

Our paper is organized as follows: The camera model and the OI pose estimation algroithm embedded in SoftOI are introduced briefly in Section III. Section IV explains the proposed simultaneous pose and correspondence determination algorithm (SofOI) in detail. The silmulations and experiment results of SotOI are showed in Section V.

II. POINTS WITH KNOWN CORRESPONDENCES

A. CAMERA MODEL

As is shown in Fig.2, given a set of non-collinear object points ${}^wP_i({}^wX_i, {}^wY_i, {}^wZ_i)^T (i = 1, 2, \ldots n)$ expressed in the object coordinate system $P_0 - X_wY_wZ_w$. The coordinates of the same points expressed in camera coordinate system $O_c - XYZ$ are $P_i(X_i, Y_i, Z_i)^T$, which can be represented by the following rigid transformation:

$$P_i = R \cdot {}^wP_i + T \tag{1}$$

where $R = (R_1, R_2, R_3)^T \in SO(3)$ and $T = (T_x, T_y, T_z)^T \in R^3$ are the rotation matrix and translation vector between $P_0 - X_wY_wZ_w$ and $O_c - XYZ$. The homogenous image point $p_i = (u_i, v_i, 1)^T$ is the projection of an object point P_i on the normalized image plane, and the line l_i that passes through O_c and p_i is called the line-of-sight of P_i. Q_i is the orthogonal projection of P_i on the line-of-sight with the expression:

$$Q_i = V_i(R \cdot {}^wP_i + T) \tag{2}$$

where

$$V_i = p_i p_i^T / p_i^T p_i \tag{3}$$

is the line-of-sight projection matrix which can project an object point orthogonally to line-of-sight defined by the corresponding image point. Theoretically, P_i should be on the line-of-sight defined by p_i, or in other words, P_i should coincide with Q_i. This fact can be expressed as follows:

$$R \cdot {}^wP_i + T = V_i(R \cdot {}^wP_i + T) \tag{4}$$

B. THE OI (ORTHOGONAL ITERATION) POSE ESTIMATION ALGORITHM

The distance between P_i and Q_i is called the object-space collinearity error, and the principle of OI algorithm is computing the pose by minimizing the sum of the squared object-space collinearity errors as:

$$\begin{aligned} E(R,T) &= \sum_{i=1}^{n} d_i^2 \\ &= \sum_{i=1}^{n} ||(E_3 - V_i)(R \cdot {}^wP_i + T)||^2 \end{aligned} \tag{5}$$

Note that (5) is quadratic in T, if R is fixed, the optimum value for T can be obtained by computing the partial derivatives of E with respect to T, and the expression of the translation vector T is:

$$\begin{aligned} T(R) &= -(\sum_{i=1}^{n}(E_3 - V_i)^2)^{-1} \sum_{i=1}^{n}(E_3 - V_i)R \cdot {}^wP_i \\ &= -(\sum_{i=1}^{n}(E_3 - V_i))^{-1} \sum_{i=1}^{n}(E_3 - V_i)R \cdot {}^wP_i \end{aligned} \tag{6}$$

So if the initial R^0 has been obtained by using a weak perspective approximation [30], one can compute the optimal R by the following iterative steps: First, assume that R^k is the kth estimation of R, then $T^k = T(R^k)$, $Q_i^k = V_i(R^k \cdot {}^wP_i + T^k)$, and the next estimation R^{k+1} can be determined by solving the problem as follows:

$$R^{k+1} = \arg\min_R \sum_{i=1}^{n} ||(R \cdot {}^wP_i + T) - Q_i^k||^2 \tag{7}$$

It is an absolute orientation problem from $\{{}^wP_i\}$ to $\{Q_i^k\}$, and which can be resolved with the SVD (singular value decomposition) method [10]. Define:

$$S = \sum_{i=1}^{n}(Q_i^k - \overline{Q^k})({}^wP_i - \overline{{}^wP_i})^T \tag{8}$$

where $\overline{Q^k}$ and $\overline{{}^wP_i}$ are the centroid of $\{Q_i^k\}$ and $\{{}^wP_i\}$ respectively. Let (U, Σ, V) be a SVD of S, that is $USV^T = \sum$. The solution to (7) is $R^{k+1} = VU^T$, then the estimation $T^{k+1} = T(R^{k+1})$ can be obtained according to (6), and this process is repeated until R is the same as that computed at the previous step.

III. POSE ESTIMATION WITH UNKNOWN CORRESPONDENCES

The global object function of OI algorithm proposed in (5) is the expression of the sum of the object-space collinearity errors between the object points and the orthogonal projections on the line-of-sights defined by the corresponding

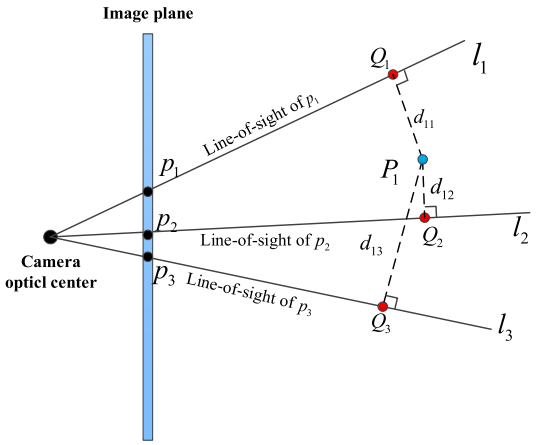


FIGURE 3. Schematic diagram of one object point and its orthogonal projections.

image points. For a set of image points $p_j(j = 1, 2, \ldots M)$ and object points $P_i(i = 1, 2, \ldots N)$ without known correspondences, an object point may match any one of the image points, so the pose problem will be confused, especially for the case of clutter or occlusion in which M is more or less than N.

To solve the pose estimation with unknown correspondences problem, the weights m_{ij} which determine the assignments between object and image points are introduced, and the simultaneous pose and correspondence can be formulated as the minimization of the new weighted object space collinearity error as:

$$\begin{aligned} E(R, T) &= \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}(d_{ij}^2 - \alpha) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}(||R \cdot {}^wP_i + T - Q_j||^2 - \alpha) \end{aligned} \quad (9)$$

where d_{ij}^2 is the square distance between an object point P_i and the orthogonal projection of P_i on the line-of-sight of p_j, that is $Q_j = V_j(R \cdot {}^wP_i + T)$. As is shown in Fig.3, Q_1, Q_2, Q_3 are the orthogonal projections of the object point P_1 (one object point in point sets $P_i(i = 1, 2, \ldots N)$)) on the line-of-sight of p_1, p_2, p_3 (three image points in point sets $p_j(j = 1, 2, \ldots N)$), respectively. d_{11}, d_{12}, d_{13} are the Euclidean distances between P_1 and Q_1, Q_2, Q_3. If p_2 is the corresponding image point of P_1, we call Q_2 the own orthogonal projection of P_1. The weight m_{ij} is equal to 1 if P_i and p_j is matched, or m_{ij} is equal to 0. Another way of thinking the matched relations is that m_{ij} is equal to 1 only if Q_j is exactly the own orthogonal projection of P_i. To avoid the trivial solution in which all the weights $m_{ij} = 0$ be the final solution to (9), a parameter α is introduced to encourage the matching from p_j to P_i. If $d_{ij}^2 - \alpha \leq 0$, p_j and P_i are potentially matched with each other. If $d_{ij}^2 - \alpha > 0$, p_j is definitely not the matched point of P_i. Note that if all the weights are all fixed (1 or 0) and $\alpha = 0$, the global objective function (9) is equivalent to (5).

In order to accelerate the algorithm, we unitize $p_j = (u_j, v_j, 1)^T$ to $p_j(\widetilde{u_j}, \widetilde{v_j}\widetilde{w_j})^T$ before computing the line-of-sight projection matrix V_i, where:

$$\widetilde{u_j} = u_j / \sqrt{u_j^2 + v_j^2 + 1^2} \quad (10)$$

$$\widetilde{v_j} = v_j / \sqrt{u_j^2 + v_j^2 + 1^2} \quad (11)$$

$$\widetilde{w_j} = 1 / \sqrt{u_j^2 + v_j^2 + 1^2} \quad (12)$$

so $p_j p_j^T$ is always equal to 1 and the global object function E can be rewritten as:

$$\begin{aligned} &E(R, T) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}(||R \cdot {}^wP_i + T - p_j^T(R \cdot {}^wP_i + T)p_j||^2 - \alpha) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}[(R_1 \cdot {}^wP_i + T_x - p_j^T(R \cdot {}^wP_i + T)\widetilde{u_j})^2 + \ldots \\ &\quad (R_2 \cdot {}^wP_i + T_y - p_j^T(R \cdot {}^wP_i + T)\widetilde{v_j})^2 + \ldots \\ &\quad (R_3 \cdot {}^wP_i + T_z - p_j^T(R \cdot {}^wP_i + T)\widetilde{w_j})^2 - \alpha] \end{aligned} \quad (13)$$

To sum up, the simultaneous pose and correspondence problem can be solved by an expectation-maximization strategy [28] as the following iteration steps:

1) Compute the correspondences (m_{ij}) according to a given pose parameters (R and T).
2) Compute the maximum likelihood estimation of pose parameters (R and T) according to the correspondences (m_{ij}) determined in step 1.
3) Update the pose parameters by the estimation results in step 2, and iteration process will be repeated until the parameters of both pose and correspondence converge.

A. CORRESPONDENCE DETERMINATION

Following the framework in [14], the correspondence determination aims at finding a $N \times M$ "0-1" assignment matrix $\mathbf{M} = \{m_{ij}\}$ which specifies the matching relations between N object points and M image points explicitly. An element m_{ij} with value 1 at the row i and column j in $\mathbf{M}$ indicates that the ith object point P_i is matched with the jth image point p_j. For the case of clutter and occlusion, a slack row $N+1$ and a slack column $M+1$ are added into $\mathbf{M}$. A value 1 at row $N+1$ and column j indicates that clutter occurs, and the image point p_j has not found any matched 1 among the object points; A value 1 at row i and column $M+1$ suggests that occlusion occurs, and there is no image point matched with the object point P_i. Since each object point can match only one image point at most, and vice versa, $\mathbf{M}$ must satisfy a constraint: the sum of weights in each row and each column are always adding up to 1. The "0-1" assignment $\mathbf{M}$ can be computed through an iterative Softassign method which combines the Sinkhorn's algorithm and deterministic annealing.

At the beginning of the iteration, $\mathbf{M}$ starts with a fuzzy matrix $\mathbf{M}_0 = \{m_{ij}^0\}$ in which each element initialized to:

$$\begin{aligned} m_{ij}^0 &= \exp(-\beta(d_{ij}^2 - \alpha)) \\ &= \exp[-\beta((R_1 \cdot {}^wP_i + T_x - p_j^T(R \cdot {}^wP_i + T)\widetilde{u}_j)^2 + \ldots \\ &\quad (R_2 \cdot {}^wP_i + T_y - p_j^T(R \cdot {}^wP_i + T)\widetilde{v}_j)^2 + \ldots \\ &\quad (R_3 \cdot {}^wP_i + T_z - p_j^T(R \cdot {}^wP_i + T)\widetilde{w}_j)^2 - \alpha)] \end{aligned} \tag{14}$$

and all of the slack row and slack column elements are set to a small constant. The parameter α is the maximum permissible squared object space error between an object point P_i and its own orthogonal projection, which is used to judge whether there is a potential matching between P_i and p_j. It must be set to a proper value for the reason that: on one hand, a overlarge α will cause the incorrect match between P_i and Q_j though they are far apart from each other, which will lead the Softassign to converge prematurely and be limited to a local optimum; on the other hand, if α is too small, P_i will not be assigned to match with Q_j though Q_j is the own orthogonal projection of P_i when image noise exits, and the Softassign cannot converge even when all the correct matches have be found. Thus, the parameter α should be set to the max maximum image noise mapped to the object space. Referring to the statistical theory in [12], we take $\alpha = 9.21 \times \sigma^2 \times (R_3 \cdot {}^wP_i + T_z)^2$ in which σ is the standard deviation of the image noise. Besides, the annealing parameter β needs to be initially set to a small value in order to ensure no correspondences will be ruled out immediately if the initial pose is not close to the real pose at the beginning of the iteration. However, it will cause a low convergence speed in the deterministic annealing process if β is too small. Based on experimental experience, β is initialized to $\beta_0 = 0.0005$. The distance d_{ij}^2 is obtained by the initial pose parameters, and $\mathbf{M}_0$ is processed concurrently by the two mechanisms:

First, a Sinkhorn technique [32] is employed in which each element of $\mathbf{M}_0$ is normalized alternatingly for several times through dividing the sum of that row or column elements respectively. This operation can be formulated by the equation:

$$\begin{cases} m_{ij}^{k+1} = m_{ij}^k / \sum_{i=1}^{N+1} m_{ij}^k, & 1 \le i \le N+1, 1 \le j \le M \\ m_{ij}^{k+1} = m_{ij}^{k+1} / \sum_{j=1}^{M+1} m_{ij}^{k+1}, & 1 \le i \le N, 1 \le j \le M+1 \end{cases} \tag{15}$$

where m_{ij}^k are the elements of the kth iterative matrix $\mathbf{M}_k$ and m_{ij}^{k+1} are that of $\mathbf{M}_{k+1}$. This process will keep running until $\mathbf{M}$ converges. Sinkhorn aims to make the resulting assignment matrix meet the constraint discussed previously: the sums of weights in each row and each column are equal to 1.

Then, the deterministic annealing [32] is applied during which the term β is updated according to $\beta = 1.05 \times \beta$ after the Sinkhorn is accomplished at each iteration step. The upper limit number of iteration in deterministic annealing is $\log_{1.05}(0.5/\beta_0)$ [12]. As the deterministic annealing goes on, the elements m_{ij} corresponding to the smaller d_{ij}^2 is going to converge to 1 while the other elements tend to converge to 0. The goal of deterministic annealing is to make the correspondence which has the minimum distance in each row and each column be assigned to the match.

B. POSE DETERMINATION

Assume that all the weights m_{ij} are obtained. Given a fixed R, the optimal value for T can be obtained by computing the partial derivatives of E with respect to T:

$$\begin{aligned} T(R) &= -[\sum_{j=1}^{M} m_j'(E_3 - V_j)]^{-1} \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}(E_3 - V_j)R \cdot {}^wP_i \\ &= F \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}(E_3 - V_j)R \cdot {}^wP_{ii} \end{aligned} \tag{16}$$

where $m_j' = \sum_{i=1}^{N} m_{ij}$. The optimal rotation matrix which minimizes the global objective function E can be obtained by solving the weighted absolute orientation problem from $\{{}^wP_i\}$ to $\{Q_j^k\}$ at each iteration step:

$$R^{k+1} = \arg\min_R \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}||(R^k \cdot {}^wP_i + T) - Q_j^k||^2 \tag{17}$$

By adding the weights m_{ij}, S can be rewritten as:

$$\begin{aligned} S &= \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}(Q_j^k - \overline{Q^k})({}^wP_i - \overline{{}^wP_i})^T \\ &= \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}Q_j^k({}^wP_i - \overline{{}^wP_i})^T \\ &\quad - \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}\overline{Q^k}({}^wP_i - \overline{{}^wP_i})^T \\ &= \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}Q_j^k({}^wP_i - \overline{{}^wP_i})^T \end{aligned} \tag{18}$$

where:

$$\overline{{}^wP_i} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}{}^wP_i}{\sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}} \quad \overline{Q^k} = \frac{\sum_{i=1}^{M} m_{ij}Q_i^k}{\sum_{i=1}^{M}\sum_{j=1}^{N} m_{ij}} \tag{19}$$

Let (U, Σ, V) be a SVD of S, the solution to (17) is $R^{k+1} = VU^T$, then $T^{k+1} = T(R^{k+1})$, and $Q_j^{k+1} = V_j(R^{k+1} \cdot {}^wP_i + T)$. As the iteration progresses, P_i will gradually get closer to the corresponding line-of-sight, and this process will be repeated until the pose is converged.

C. START AND TERMINATION OF SOFTOI

The SoftOI is an iterative method, and a proper initial pose that is not far away from the true pose has a higher probability to make the SoftOI successful. However, there is no

$$R = \begin{bmatrix} \cos\beta\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\alpha\sin\gamma & \sin\alpha\sin\gamma + \cos\alpha\sin\beta\cos\gamma \\ \cos\beta\sin\gamma & \cos\alpha\cos\gamma + \sin\alpha\sin\beta\sin\gamma & -\sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma \\ -\sin\beta & \cos\beta\sin\alpha & \cos\beta\cos\alpha \end{bmatrix}$$
$$T = (T_x, T_y, T_z) \tag{20}$$

additional information to judge whether an initial pose is close to the true pose or not. What is more, the objective function in (9) has many local optimal solutions; a single initial pose is not enough to obtain the global optimal value, so we need to try several different initial poses to find one which can lead the SoftOI to converge to the correct pose. Since a pose can be expressed by the three Euler angles and three translation parameters as shown at the top of this page (20), as shown at the top of this page, where α, β and γ are the yaw, pitch and roll angle, and x, y and z are the translation components in the three directions, we establish a pose candidates set in which each pose candidate is consist of two parts: a 3D Euler angle vector $EA = (\alpha, \beta, \gamma)$ and a 3D translation vector $T = (T_x, T_y, T_z)$. The Euler angle vectors are regularly generated at intervals of $\pi/6$ in the range of $-\pi$ to π, and the translation vectors are randomly generated by the pseudorandom number generator [33] in a 3D hypercube that can approximately cover the true translation. The total number of the candidates is 2197 $((2\pi/(\pi/6)+1)^3 \times 1)$, which means that the SosfOI will give up if a proper initial pose is not found within 2197 tries. With an initial pose chosen from the pose candidates set, the SoftOI algorithm keeps searching for the global optimal value until the termination criterion is met at the first time.

The way to judge whether the termination criterion has been met is whether there are enough correct correspondences between object points and image points have been found. However, in real imagery, not all correct matches will be found due to high image noise or imperfect image feature point extraction algorithm. Therefore, we set a threshold for the ratio of correct matches to $k = 0.9$ in actual measurement, which indicates the algorithm will be terminated if more than 90% of the detected object points have found the corresponding image points. As is mentioned in section IV A, there will be an approximate "1" in the ith row and jth column of the assignment matrix $\mathbf{M}$ if the ith object point and the jth image point are matched, the N_{match} (number of the correct matches) can be intuitively counted from $\mathbf{M}$ at each iteration. In addition, after about 30 iterations in the deterministic annealing process, the N_{match} tends to remain unchanged while the pose also tends to converged. So to make the algorithm efficient, the current search will also be early terminated if there is no significant increase on N_{match} after at most $\log_{1.05}(0.5/\beta_0)$ iterations [12] and be restarted with another new initial pose from the pose candidates set.

The threshold $k = 0.9$ is just an empirical but not perfect value, as it can be the case that an accurate pose be "ignored" because the ratio of N_{match} is less than 0.9 due to the high image noise. Conversely, the search may also be terminated with a wrong pose if the ratio of N_{match} is greater than 0.9, this is because too many clutter points are wrongly regarded as the image points if the clutter rate is high. However, these cases are not common, and the success rate of SoftOI in different conditions of clutter and occlusion rate will be shown in the simulations in the Section V.

D. SOFTOI ALGORITHM FLOW

Initial

a) List of object points ${}^wP_i = ({}^wX_i, {}^wY_i, {}^wZ_i)^T (i = 1, 2, \ldots N)$.
b) List of the normalized image points $p_j = (u_j, v_j, 1)^T (j = 1, 2, \ldots M)$.
c) The terms m_{ij} in the slack row and slack column are initialized to $\gamma = 1/(\max\{M, N\} + 1)$.
d) β is initialized to $\beta_0 = 0.0005$.
e) α is initialized to $9.21 \times \sigma^2 \times (R_3 \cdot {}^wP_i + T_z)^2$.

For K < 2197

a) Choose an initial pose (R^0, T^0) from the pose candidates sets
b) Compute the orthogonal projection points of P_i on the line of sights defined by p_j: $Q_j = V_j(R \cdot {}^wP_i + T)$

While $N_{match} < ceil(0.9N(1 - p_o))\&\&\beta < \beta_{end}$ $(\beta_{end} = 0.5)$

a) Compute the square distances:

$$\begin{aligned} d_{ij}^2 = {} & (R_1 \cdot {}^wP_i + T_x - p_j^T(R \cdot {}^wP_i + T)\widetilde{u}_j)^2 \\ & + (R_2 \cdot {}^wP_i + T_y - p_j^T(R \cdot {}^wP_{ii} + T)\widetilde{v}_j)^2 \\ & + (R_3 \cdot {}^wP_i + T_z - p_j^T(R \cdot {}^wP_i + T)\widetilde{w}_j)^2 \end{aligned}$$

b) Compute $m_{ij}^0 = \gamma \exp(-\beta(d_{ij}^2 - \alpha))$.
c) Normalize each row and column of $\mathbf{M}$ alternatingly for several times until $\|\mathbf{M}^k - \mathbf{M}^{k-1}\| \le 1$;
d) Compute the defined $F = -[\sum_{j=1}^{M} m_j'(E_3 - V_j)]^{-1}$, with $m_j' = \sum_{i=1}^{N} m_{ij}$.
e) Compute the defined $S = \sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}Q_j^k(P_i - \overline{P_i})^T$, then process S by SVD method: $USV^T = \sum$.
f) Compute the optical estimation of the rotating matrix R^{k+1} at the next iteration step $R^{k+1} = VU^T$.

g) Compute the optical estimation of the rotating matrix T^{k+1} at the next iteration step $T^{k+1} = F\sum_{i=1}^{N}\sum_{j=1}^{M} m_{ij}(E_3 - V_j)R^{k+1} \cdot {}^{w}P_i$

h) update $\beta = \beta_u \cdot \beta$, $\beta_u = 1.05$ is the updated parameter;

End while

If $N_{match} > 0.9N(1 - p_o)$

Output $\mathbf{M}$, R and T

Break;

End if

K = K + 1;

End for

Output:

a) The assignment matrix $\mathbf{M} = \{m_{ij}\}$ which explicitly specifies the matching relations between M image points and N object points.

b) Rotation matrix $R = (R_1, R_2, R_3)^T$.

c) Translation vector $T = (T_x, T_y, T_z)^T$.

IV. SIMULATIONS AND EXPERIMENTS RESULTS

We test the performance of our approach in the simultaneous pose and correspondence problem with Monte Carlo simulations in which the effects of occlusion and clutter are considered. In the simulations, 100 independent random trials are conducted with a virtual camera, of which the image size is 800 × 700 pixels and focal length $f = 800$.

First, N object points (vertices) which are randomly distributed in the range of $[-2, 2] \times [-2, 2] \times [4, 8]$ are generated. Next, the object is rotated to an arbitrary orientation and translated to arbitrary point according to a generated pose which can be denoted by R_{true} and T_{true}. Then the object is projected back to the image plane with according to the perspective projection model. Theoretically, the number of image points M detected on the image plane is equal to the number of object points N. However, N will be less than M if some of object points are occluded, and more than M if clutter occurs. Let P_o and P_c denote the ratio of occlusion and clutter respectively, to simulate occlusion and clutter, NP_o points must be deleted from the feature points and $N(1 - P_o)P_c/(1 - P_c)$ clutter points must be added on the image. Finally, Gaussian noise with standard deviation $\sigma = 1$ pixel is added to the image feature points in both x and y coordinates, and the correspondences between the object points and image points are disorganized. Each trial is performed with the combination of the three parameters: $N \in \{20, 30, 40, 50, 60, 70, 80\}$, $P_o \in \{0.2, 0.4, 0.6\}$, $P_c \in \{0.2, 0.4, 0.6\}$.

Fig.4 shows a recording of updated process of determining the correct pose and correspondence by our method. The blue filled point is the reprojection of object points and the red hollow dot is the detected image points (including the clutter points). The blue solid lines and red dotted lines which connect the object points and image points respectively are used to make the pictures easier to understand, and they are not used by the algorithm.

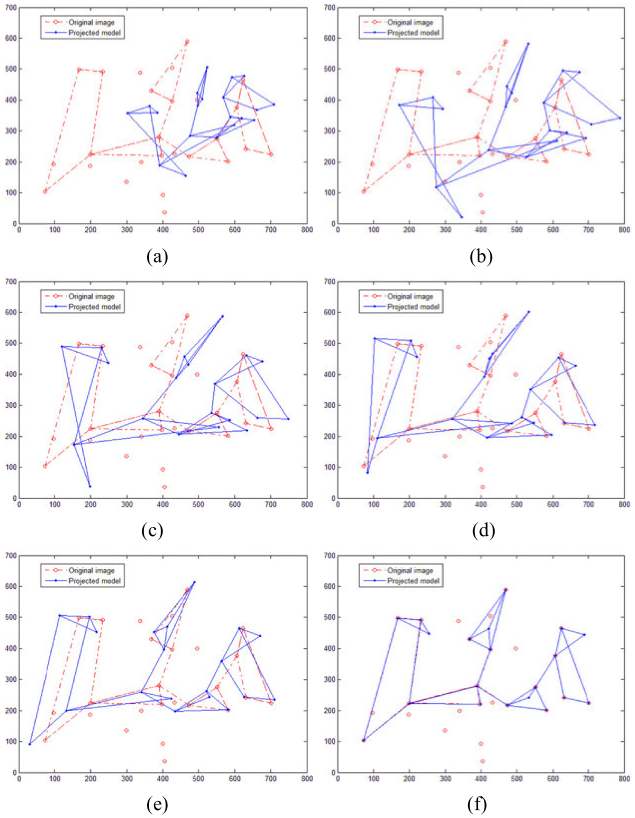


FIGURE 4. The recording of updated process of softOI. (a) The initial guess of the pose. (b)∼(f) are the 1th,5th,13th,20th,27th, 33th steps of the iteration. $P_o = 0.2$, $P_c = 0.4$.

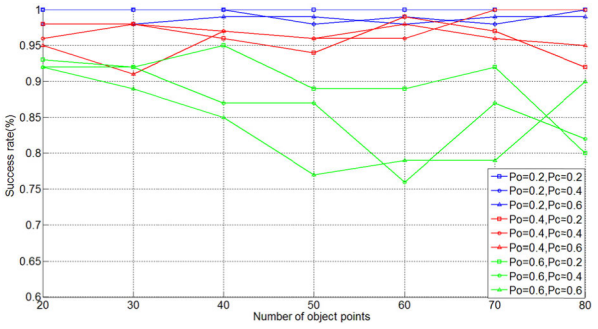


FIGURE 5. The success rate of SoftOI for different combinations of P_o and P_c with the number of object points from 20 to 80.

A. SIMULATION OF SUCCESS RATE OF SOFTOI

Since the correspondence and pose determination are optimized mutually, a good pose will be obtained if the correct correspondence is determined, and vice versa. In simulations, we consider a trial to be successful when over 90% of the correct correspondences have been found and the pose error is small enough. Since the number of pose candidates set is limited to 2197, a trial is regarded as a failure if it has not succeeded within 2197 initial poses. However, one cannot expect all the trials to be successful due to different levels of image noise and object points recognition rate. The success

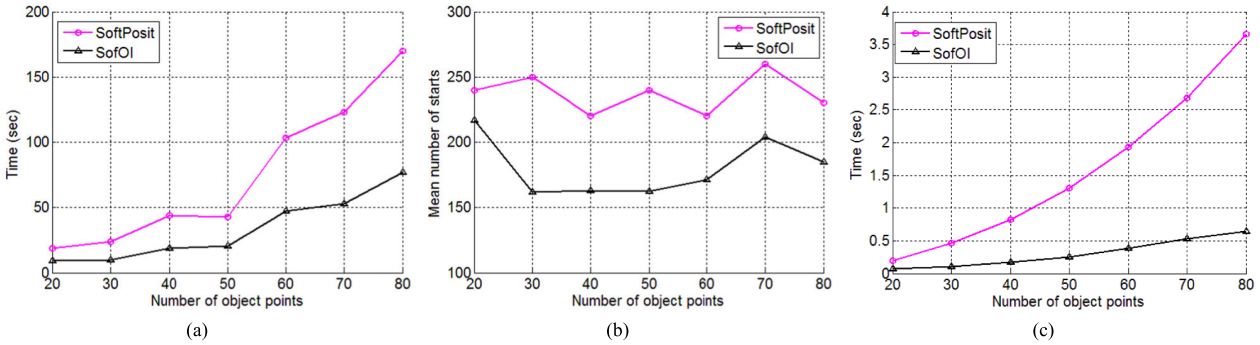


FIGURE 6. The efficiency comparison between SoftPosit and SoftOI. (a) the computational time comparison for different number of object points with unknown correspondences. (b) the mean number of random starts comparison for different number of object points. (c) the computational time comparison for different number of object points if an initial pose has been found.The experiments of SoftPosit and SoftOI are conducted under the circumstance of $P_o = 0.4$ and $P_c = 0.4$.

rates of SoftOI with different combinations of occlusion rate P_o and cluster rate P_c are shown in Fig.5. One hundred trials were conducted for each combination. As we can be from Fig.5, the success rate of SotfOI decreases with the increase in P_o and P_c. When the occlusion rate is small ($P_o = 0.2$), the success rate is high (over 97%). Even when the occlusion occurred seriously ($P_o = 0.6$), a good pose can be found with a probability of more than 75%. In addition, with the increasing number of object points, the success rate doesn't decline significantly when P_o is smaller than 0.6. It suggests that the SoftOI can keep a stable and good performance though the scene becomes complex.

B. SIMULATION OF RUN TIME

The complexity of SoftOI is $O(NM^2)$, which is comparable to that of SoftPosit. However, the running speed of SoftOI is almost two times faster than that of SoftPosit, as can be seen from Fig.6(a). The main reason is that the OI algorithm nested in our algorithm is global convergent [10], and it can obtain the correct pose even though the initial pose is not near from the true pose when the correspondences are fixed. Benefited from the global convergence in OI, our method has strong search ability for optimal value, which indicates that it needs less initial start poses in the whole searching process. Obviously, the less the initial start poses are needed, the faster the algorithm will be. Under the same conditions, the mean numbers of initial poses required in each successful trail are shown in Fig.6 (b). To find a proper initial pose from the pose candidates set, our method needs about average 57 attempts less than SoftPosit.

In addition, the most of the runtime of our method is spent on the searching for the proper initial pose, and the speed of the SoftOI will become much faster if a proper initial pose has been found. As can be seen from Fig.6 (c), the SoftOI requires about 0.6s for 80 points with unknown correspondences while the runtime of SoftPosit is about 3.6s.

C. SIMULATION OF ACCURACY

The accuracy of SoftOI algorithm is evaluated by the pose errors between the true pose (R_{true}, T_{true}) and the estimated pose (R, T). The rotation error is defined

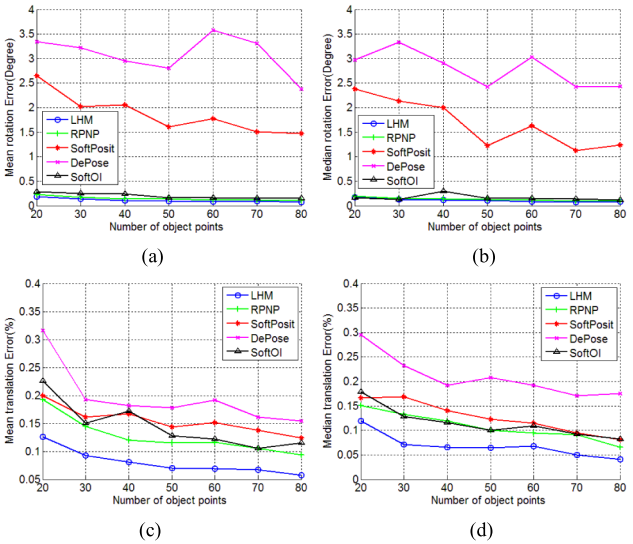


FIGURE 7. The comparison of pose estimation error for different number of points. (a),(b) the mean and median rotation errors of different methods. (c),(d) the mean and median translation errors of different methods. The experiments of SoftPosit, Depose and SoftOI are conducted under the circumstance of $P_o = 0.4$ and $P_c = 0.4$.

as: $e_{rot} = \max_{k=1}^{3} a\cos(dot(R_{ktrue}, R_k)) \times 180/\pi$, where R_k is the kth column of R, and the translation error is measured according to $e_{trans} = ||T_{true} - T||/||T|| \times 100$. The performance of SoftOI are compared with two other state of art simultaneous pose and correspondence determination algorithms (SoftPosit, DePose) and two PnP methods (RPnP[8], LHM[10]). Since the SoftOI, SoftPosit [14] and Depose [21] may not converge to the correct pose with a 100% probability, only the successful trails are considered here in order to make the accuracy experiment equitable. The simulation of the accuracy is conducted with 100 independent random trials and the pose estimation errors of different methods are plotted in Fig.7. Fig.7 (a) and Fig.7 (b) are the mean and median rotation errors comparison, while Fig.7 (c) and Fig.7 (d) are the mean and median translation errors comparison. As can be seen, the SoftOI has shown its better performance when compared with other simultaneous pose and correspondence determination algorithms. The rotation

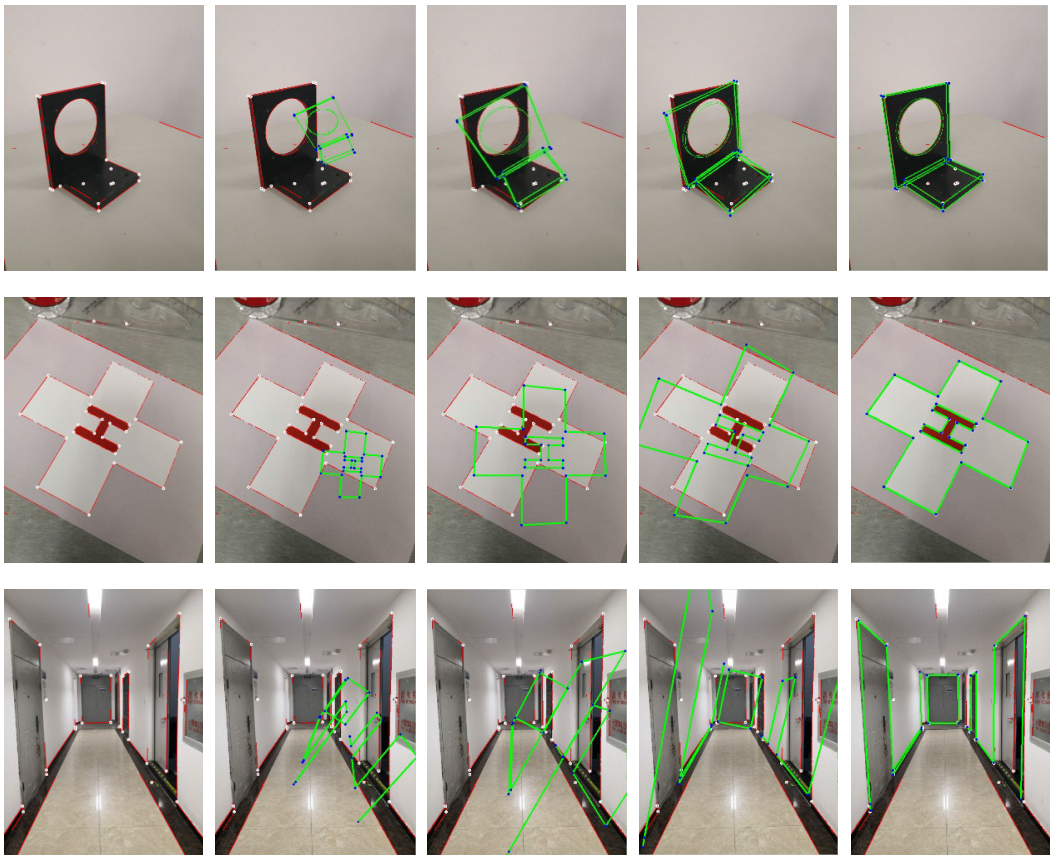


FIGURE 8. The applications of the SoftOI in some different simulated scenarios. (a) The reference images. (b) The projections of the CAD model according to the initial pose. (c) (d) Some sample projections in the iteration process. (e) Final results of SoftOI.

accuracy has been improved more than 1° when the SoftOI is employed, whereas the translation error has also decreased slightly. What is more, even compared with PnP methods, the rotation and translation accuracy of SoftOI are similar with RPnP and just about 0.05° and 0.7% lower than LHM when the number of object points is more than 50. However, one cannot expect that SoftOI can be superior to the PnP methods, this is because SoftOI aims to solve not only the pose problem but also the correspondence problem, and it is usually terminated with an approximate rather than an absolute "0-1" assignment matrix to ensure the speed of algorithm. Even though a rough "0-1" assignment is enough to determine the correspondences, it will bring interference from the incorrect potential matching points more or less during the pose refinement process. If more high accuracy is necessary, the pose can be recalculated by the PnP method after the correct correspondences have been determined by SoftOI.

D. EXPERIMENTS

We test our simultaneous pose and correspondence determination algorithm on real images. Fig.8 shows the applications of the SoftOI in some different simulated scenarios such as hand-eye robot work piece grabbing (the first row), helicopter automatic landing (the middle row) and indoor robotic vehicle navigation (the bottom row). In order to accomplish these works, the relative pose between the camera and the object must be determined.

For each scenery in (a), lines are detected by Hough transform [29], shown in red, and some of their intersections are extracted as the 2D image feature points (mark with white). The object feature points are obtained from the CAD model, and we connect their projections (marked with blue) with green lines to make it easier to understand. The correspondences between the image feature points and object feature points are not known, and the number of object points, the occlusion rate P_o and cluster rate P_c in each scenery are: 16, 21%, 25% (piece grabbing); 24, 8%, 15% (automatic landing); 22, 10%, 20% (robotic vehicle navigation). Each image in (b) shows the projection with an initial guess, and the projections according to the refined pose at some iteration step in SoftOI are shown in (c) and (d).As can be seen, the 3D models are getting closer to the object, which suggests that the pose and correspondences are evolving towards the true value. The final pose and correspondence determination results are shown in (e), in which the projections of 3D models have all matched with the object points. So far, the correspondence and pose have all been successfully determined

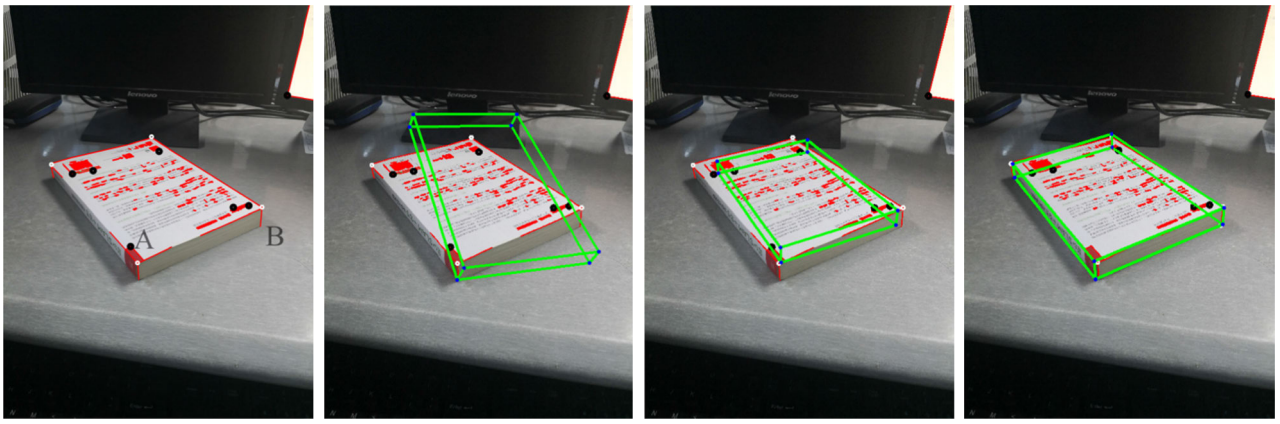


FIGURE 9. A failure case of a book pose recognition. (a) The reference images. (b) The projections of the CAD model according to the initial pose. (c) The wrong result of SoftOI. (d) The correct result of SoftOI.

by SoftOI within 84, 73, 93 iterations, respectively. Besides, the object points of the helicopter land mark in the second scenery (middle row) are coplanar, which cannot be solved by SoftPosit, so our algorithm has a broader range of applications.

Since the success rate of our method is not 100%, and there is a little probability of failure in the application of SoftOI if the occlusion rate and cluster rate are high. Figure 9 shows a failure case of a book pose recognition in which the 8 vertexes of the book are used as the object points. The P_c and P_o are 0.63 and 0.50, which means that there exists 7 cluster points (marked with black circle), and only 4 object points are correctly detected (marked with white circle). As can be seen from Fig.9 (c), the CAD models (marked with blue circle) have not matched with the vertexes of the book well in the final result, which suggests that the compute pose error is large. This is because some cluster points are wrongly detected as the object points, and the CAD model points are likely to get closer to the ''faked'' object points with the evolution of the pose if the cluster points are not far away from the ''true'' object points. However, the SoftOI considers an image point and an object point are matched if the distance between them is small enough, and terminates once 4 ($ceil(8 \times 0.5 \times 0.9)$) matches have been found, so the algorithm may fall into local optimum and converge to an incorrect pose. This situation can be improved by a better image point extract algorithm in which more object points can be detected and more cluster can be filtered out. For example, if we manually remove one cluster point at the position ''A'' and add one more vertex as the detected object point at position ''B'' in Figure 9(a), and the P_c and P_o become 0.54 and 0.37, then a good pose and a correct correspondence can be obtained by the SoftOI again, as is shown in Figure9 (d).

Besides, another factor that will cause a failure is that there is no proper initial pose be found in the pose candidates set, and the termination condition has not been satisfied although SoftOI is restarted 2197 times. Since an initial pose is composed of an Euler angle vector and a translation vector, there is a probability that the Euler angle vector is suitable but the translation vector is too far away from the true translation. However, the $N_{matches}$ will likely be less than the threshold due to large distance between object points and the corresponding image points if the initial translation vector have not evolved into an acceptable value within $\log_{1.05}(0.5/\beta_0)$ times, so the initial pose will be refused even if the initial rotation is proper. To solve this problem, we can appropriately reduce the Euler angle interval between the two adjacent candidates in the pose candidates set, and generate more candidates to increase the probability of finding a proper translation vector. However, it may cause a little loss of efficiency.

V. CONCLUSION

In this paper, we present a new simultaneous pose and correspondence determination algorithm in which occlusion and cluster cases are also taken into account. An assignment matrix that specifies the matching relations between object points and image points is introduced into OI algorithm. The Softassign correspondence determination algorithm and OI pose estimation algorithm are nested into one iteration loop, and the pose and the correspondence are determined alternately by minimizing the object function based on the weighted object space collinearity error and by applying the deterministic annealing technique. Compared with other state-of-art simultaneous pose and correspondence determination algorithms, our method has the advantage of higher accuracy, faster speed and wider applications. The SoftOI algorithm is well suited for complex pose estimation applications such as robot guidance, aircrafts docking and objects recognition where correspondences between image point and object point are not known in advance. In the future work, we will improve initial pose set establishment method to reduce the overall algorithm time consumption and make the SoftOI more efficient in the real-time measurement.

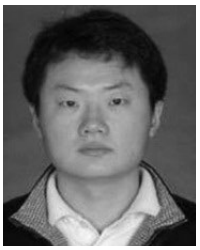
REFERENCES

[1] K. Diaz-Chito, J. M. D. Rincón, A. Hernández-Sabaté, and D. Gil, "Continuous head pose estimation using manifold subspace embedding and multivariate regression," *IEEE Access*, vol. 6, pp. 18325–18334, 2018.
[2] M. Strohmeier, T. Walter, J. Rothe, and S. Montenegro, "Ultra-wideband based pose estimation for small unmanned aerial vehicles," *IEEE Access*, vol. 6, pp. 57526–57535, 2018.
[3] H. J. Asl and J. Yoon, "Adaptive vision-based control of an unmanned aerial vehicle without linear velocity measurements," *ISA Trans.*, vol. 65, pp. 296–306, Nov. 2016.
[4] Y.-L. Kuo, B.-H. Liu, and C.-Y. Wu, "Pose determination of a robot manipulator based on monocular vision," *IEEE Access*, vol. 4, pp. 8454–8464, 2016.
[5] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[6] Y. Zheng, Y. Kuang, S. Sugimoto, K. Astrom, and M. Okutomi, "Revisiting the PnP problem: A fast, general and optimal solution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2344–2351.
[7] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," *Int. J. Comput. Vis.*, vol. 81, no. 2, pp. 155–166, 2009.
[8] S. Li, C. Xu, and M. Xie, "A robust O(n) solution to the perspective-n-point problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1444–1450, Jul. 2012.
[9] Y. Zheng, S. Sugimoto, and M. Okutomi, "Aspnp: An accurate and scalable solution to the perspective-n-point problem," *IEICE Trans. Inf. Syst.*, vol. E96.D, no. 7, pp. 1525–1535, 2013.
[10] C.-P. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Trans. pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 610–622, Jun. 2000.
[11] K. Zhou, X. Wang, Z. Wang, H. Wei, and L. Yin, "Complete initial solutions for iterative pose estimation from planar objects," *IEEE Access*, vol. 6, pp. 22257–22266, 2018.
[12] X. Guo, J. Tang, J. Li, C. Shen, and J. Liu, "Attitude measurement based on imaging ray tracking model and orthographic projection with iteration algorithm," *ISA Trans.*, to be published. doi: 10.1016/j.isatra.2019.05.009.
[13] H. Opower, *Multiple View Geometry in Computer Vision*, R. Hartley and A. Zisserman. Cambridge, U.K.: Cambridge Univ. Press, 2000, p. 624.
[14] P. David, D. Dementhon, R. Duraiswami, and H. Samet, "SoftPOSIT: Simultaneous pose and correspondence determination," *Int. J. Comput. Vis.*, vol. 59, no. 3, pp. 259–284, Sep. 2004.
[15] F. Moreno-Noguer, V. Lepetit, and P. Fua, "Pose priors for simultaneously solving alignment and correspondence," in *Proc. 10th Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2008.
[16] J. Sánchez-Riera, J. Östlund, P. Fua, and F. Moreno-Noguer, "Simultaneous pose, correspondence and non-rigid shape," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1189–1196.
[17] J. S. Beis and D. G. Lowe, "Indexing without invariants in 3D object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, pp. 1000–1015, Oct. 1999.
[18] D. Scaramuzza, "1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints," *Int. J. Comput. Vis.*, vol. 95, no. 1, pp. 74–85, 2011.
[19] M. Fenzi, R. Dragon, L. Leal-Taixé, B. Rosenhahn, and J. Ostermann, "3D object recognition and pose estimation for multiple objects using multi-prioritized RANSAC and model updating," in *Proc. Symp. German Assoc. Pattern Recognit.*, 2012.
[20] C. Rossi, M. Abderrahim, and J. C. Diaz, "EvoPose: A model-based pose estimation algorithm with correspondences determination," in *Proc. IEEE Int. Conf. Mechatronics Automat.*, Jul./Aug. 2006, pp. 1551–1556.
[21] J. Xia, X. Xu, and J. Xiong, "Simultaneous pose and correspondence determination using Differential Evolution," in *Proc. 8th Int. Conf. Natural Comput.*, 2012, pp. 703–707.
[22] P. David, D. DeMenthon, R. Duraiswami, and H. Samet, "Simultaneous pose and correspondence determination using line features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, pp. 1–8.
[23] X. Zhang, Z. Zhang, Y. Li, X. Zhu, Q. Yu, and J. Ou, "Robust camera pose estimation from unknown or known line correspondences," *Appl. Opt.*, vol. 51, no. 7, pp. 936–948, 2012.
[24] H. Zhou, T. Zhang, and W. Lu, "Vision-based pose estimation from points with unknown correspondences," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3468–3477, Aug. 2014.
[25] S. Gold, A. Rangarajan, C.-P. Lu, S. Pappu, and E. Mjolsness, "New algorithms for 2D and 3D point matching: Pose estimation and correspondence," *Pattern Recognit.*, vol. 31, pp. 1019–1031, Aug. 1998.
[26] S. Gold and A. Rangarajan, "A graduated assignment algorithm for graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 4, pp. 377–388, Apr. 1996.
[27] D. F. Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code," *Int. J. Comput. Vis.*, vol. 15, no. 1, pp. 123–141, Jun. 1995.
[28] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.
[29] M. Ekstrom, "Digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 4, pp. 484–486, Aug. 1980.
[30] R. Horaud, F. Dornaika, and B. Lamiroy, "Object pose: The link between weak perspective, paraperspective, and full perspective," *Int. J. Comput. Vis.*, vol. 22, no. 2, pp. 173–189, 1997.
[31] R. Sinkhorn, "A relationship between arbitrary positive matrices and doubly stochastic matrices," *Ann. Math. Statist.*, vol. 35, no. 2, pp. 876–879, 1964.
[32] P. N. Suganthan, E. K. Teoh, and D. P. Mital, "Pattern recognition by graph matching using the potts MFT neural networks," *Pattern Recognit.*, vol. 28, no. 7, pp. 997–1009, 1995.
[33] F. Koeune, "Pseudorandom number generator," *Stream Ciphers*, vol. 36, no. 1, pp. 57–67, 2009.

HANG DONG received the B.S. degree with the School of Automation, Shandong University of Science and Technology, in 2013. He is currently pursuing the Ph.D. degree with the Department of State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University. His current research interests include pose estimation, optical measurement technology, and vision navigation.

CHANGKU SUN received the M.S. degree from the Harbin Institute of Technology, in 1990, and the Ph.D. degree from the Saint Petersburg Precision Mechanics and Optics Institute, Russia, in 1994. He is currently a Professor with the Department of State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, and a Visiting Researcher with the Science and Technology on Electro-Optic Control Laboratory, Louyang Institute of Electro-Optic Equipment. His research interests include machine vision, optical measurement technology, inertial measurement technology, and visual and inertial fusion pose measurement.

BAOSHANG ZHANG received the B.S. and M.S. degrees from Northwestern Polytechnical University, in 2006 and 2009, respectively. He is currently pursuing the Ph.D. degree with the Department of State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University. He is currently an Engineer with the Science and Technology on Electro-Optic Control Laboratory, Luoyang Institute of Electro-Optic Equipment. His current research interests include visual localization and pose estimation.



PENG WANG received the B.S. and Ph.D. degrees from the Department of Precision Instruments Engineering, Tianjin University, China, in 2004 and 2008, respectively, where he is currently an Associate Professor with the Department of State Key Laboratory of Precision Measuring Technology and Instruments. He is also a Visiting Researcher with the Science and Technology on Electro-Optic Control Laboratory, Louyang Institute of Electro-Optic Equipment. His research interests include machine vision, optical measurement technology, inertial measurement technology, and visual and inertial fusion pose measurement.