

An Extended Perspective Three Points Problem

Yong Liu, Claus B. Madsen, and Moritz Störring

Computer Vision and Media Technology Laboratory

Aalborg University, Denmark

{yliu,cbm,mst}@cvmt.dk

<http://www.cvmt.dk>

Abstract. In this paper, an extended P3P (Perspective Three Points) problem is investigated. It is formulated as fitting three moving vertices along their associated optical rays to a known triangle structure. This allows the three optical rays to come from one, two or three cameras respectively. The classical P3P problem for only single camera is considered as a special case of the extended P3P problem. An analysis on these different cases is given in a uniform way. Experiments with simulated data show the effectiveness of the approach.

1 Introduction

Pose estimation is an important task in model-based vision and photogrammetry. The problem may be stated as: For a known model and a calibrated camera, given N 2D-3D correspondences, find the respective pose between the object and the camera, which is addressed as Perspective N Points (PNP) problem. Classical techniques for parameter estimation such as least squares and iterative algorithms [1,2,3] can give an optimal solution when no mismatch is present. However, they have no internal mechanisms for detecting and rejecting gross errors. Contrast to the least squares methods which use as much of the data as possible for initial estimation, the RANSAC algorithm [4] uses as small an initial set as possible and scores this set by its consistent data, which was proved to be robust to outliers. Six pairs of 2D-3D matching points are minimal for a linear pose estimation. For less than six matches, nonlinear estimation is involved and a closed form of the solution is often proposed. See [4,5,6] for P4P problem and [4,7,3] for P3P problem. Three points are the minimal set from which a finite number of solutions can be derived. There are eight solutions in total, no more than four of them can be found to be feasible [4]. Although P3P problem has been addressed since 1841 by the German mathematician Grunert [8], the present P3P problem still persists in the single camera case. Now let us consider a configuration shown in the left of Figure 1, the calibrated camera triplet views a known triangle with its vertices on the sphere surface of a ball. Due to occlusion, every camera is able to capture only one of the vertices, a P1P problem for each camera. This configuration is not covered by the classical P3P problem. Our motivation is to extend the classical P3P problem from a single camera consideration to stereo pair and triplet considerations. We have

two reasons to do so. First, the extended configuration is more general than the classical one, which should be covered from a theoretical viewpoint. Second, there are also demands from applications. For example, in an augmented reality system involving a multi-user interface, each user wears a Head Mounted Display (HMD), of which the extrinsic parameters could be determined by an ultrasound or magnetic system. When the users are sitting around a table and viewing a Place Holder (PH) object on the table, the PH object may be partially occluded to some of the users. An extreme situation is as shown in the left of Figure 1. It is desirable to collect the non-occluded information from the different users to determine the pose of the PH object with respect to the users.

In this paper, the P3P problem is restated as fitting three points moving on three optical rays to a known triangle structure, an analysis on these different cases is given in a uniform way. It is trivial to verify that the classical P3P problem falls into a special case of the extended P3P problem.

The paper is organized as follows: A general formulation of P3P problem is presented in Section 2, an analysis on the extended P3P problem under different configurations is given in Section 3, the experiments in Section 4 are carried on simulated data. Finally, this paper is summarized in Section 5.

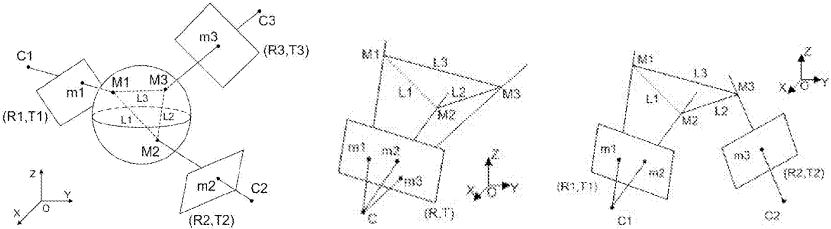


Fig. 1. The P3P problem in different situations

2 General Formulation of P3P Problem

In this paper we use the standard camera model, setting the principle point in the image center and focal length to 1, which is obtainable by a transformation using the calibration matrix for the image coordinates. In the left of Figure 1, let the euclidean transformation from the world coordinate system to the i -th ($i = 1, 2, 3$) camera be (R_i, T_i) . In the local camera coordinate system, the 3D coordinates of the optical center is $C_i = (0, 0, 0)^T$ and the image point $m_i = (u_i, v_i, 1)^T$, where (u_i, v_i) is its 2D coordinate. In the world coordinate system, these two points have their coordinates as $C_i^w = -R_i^T \cdot T_i$ and $m_i^w = R_i^T \cdot m_i - R_i^T \cdot T_i$. The three points M_i , m_i^w and C_i^w are collinear, so,

$$M_i = s_i \cdot R_i^T \cdot m_i - R_i^T \cdot T_i, \quad i = 1, 2, 3 \quad (1)$$

where s_i are three unknowns.

When s_i varies, the 3D point M_i is moving along its optical rays defined by its projection and the associated optical center. In order to make sure that M_i is located in front of its viewing camera, the intuitive constraint on s_i is that s_i should be positive. Other constraints come from the lengths of the triangle sides, say L_1, L_2 and L_3 , which can be formulated as:

$$\begin{aligned} |M_1 - M_2| &= L_1 \quad , & (2) \\ |M_2 - M_3| &= L_2 \quad , & (3) \\ |M_3 - M_1| &= L_3 \quad , & (4) \end{aligned}$$

where $|\cdot|$ is the 2-norm of a vector.

Let $R_i = (r_i^{p,q})$, $p, q \in \{1, 2, 3\}$ and $T_i = (t_i^1, t_i^2, t_i^3)^T$, Eqn. (1) is expanded as:

$$M_i = \begin{pmatrix} a_i^1 s_i + b_i^1 \\ a_i^2 s_i + b_i^2 \\ a_i^3 s_i + b_i^3 \end{pmatrix} \quad , \quad (5)$$

with

$$a_i^1 = r_i^{1,1} u_i + r_i^{2,1} v_i + r_i^{3,1} \quad , \quad (6)$$

$$a_i^2 = r_i^{1,2} u_i + r_i^{2,2} v_i + r_i^{3,2} \quad , \quad (7)$$

$$a_i^3 = r_i^{1,3} u_i + r_i^{2,3} v_i + r_i^{3,3} \quad , \quad (8)$$

$$b_i^1 = -r_i^{1,1} t_i^1 - r_i^{2,1} t_i^2 - r_i^{3,1} t_i^3 \quad , \quad (9)$$

$$b_i^2 = -r_i^{1,2} t_i^1 - r_i^{2,2} t_i^2 - r_i^{3,2} t_i^3 \quad , \quad (10)$$

$$b_i^3 = -r_i^{1,3} t_i^1 - r_i^{2,3} t_i^2 - r_i^{3,3} t_i^3 \quad . \quad (11)$$

Substitute Eqn. (5) into Eqn. (2), (3) and (4), we get

$$A_1 s_1^2 + B_1 s_2^2 + C_1 s_1 s_2 + D_1 s_1 + E_1 s_2 + F_1 - L_1^2 = 0 \quad , \quad (12)$$

$$A_2 s_2^2 + B_2 s_3^2 + C_2 s_2 s_3 + D_2 s_2 + E_2 s_3 + F_2 - L_2^2 = 0 \quad , \quad (13)$$

$$A_3 s_3^2 + B_3 s_1^2 + C_3 s_3 s_1 + D_3 s_3 + E_3 s_1 + F_3 - L_3^2 = 0 \quad . \quad (14)$$

The computation of A_1, B_1, \dots, F_1 in the first equation only involves $(a_i^1, a_i^2, a_i^3, b_i^1, b_i^2, b_i^3)$ for $i = 1, 2$. The coefficients of the second equation involve the case of $i = 2, 3$ and the third ones involve the case of $i = 3, 1$. In this way, the k -th coefficients ($k = 1, 2, 3$) are associated with indices (i, j) in $(a_i^1, a_i^2, a_i^3, b_i^1, b_i^2, b_i^3)$, can be expressed as:

$$A_k = a_i^{1^2} + a_i^{2^2} + a_i^{3^2} \quad , \quad (15)$$

$$B_k = a_j^{1^2} + a_j^{2^2} + a_j^{3^2} \quad , \quad (16)$$

$$C_k = -2a_i^1 a_j^1 - 2a_i^2 a_j^2 - 2a_i^3 a_j^3 \quad , \quad (17)$$

$$D_k = 2a_i^1(b_i^1 - b_j^1) + 2a_i^2(b_i^2 - b_j^2) + 2a_i^3(b_i^3 - b_j^3) , \quad (18)$$

$$E_k = -2a_j^1(b_i^1 - b_j^1) - 2a_j^2(b_i^2 - b_j^2) - 2a_j^3(b_i^3 - b_j^3) , \quad (19)$$

$$F_k = (b_i^1 - b_j^1)^2 + (b_i^2 - b_j^2)^2 + (b_i^3 - b_j^3)^2 . \quad (20)$$

These coefficients can be completely determined by the camera extrinsic parameters and image coordinates of the projections, and then the three quadratic equations with three unknowns Eqn. (12), (13) and (14) remain to be solved. There exist as many as $2 \times 2 \times 2 = 8$ solutions. In the next section we present an analysis of the solutions in the different cases.

3 Analysis of Different Cases

This extended P3P problem is about the minimal condition for pose estimation. We assume none of the vertices appears simultaneously in two or three images, otherwise its 3D position is obtainable by 3D triangulation, and then the information becomes redundant. The general configuration is as shown in the left of Figure 1. The minimal configuration for a single camera and a stereo camera are illustrated in the middle and right of Figure 1, respectively. In the stereo case, two image planes and their optical centers are coincident, or the two camera frames are coincident. In the single camera case, all three camera frames are coincident.

3.1 A Single Camera

In this case, all three vertices are projected on the same image plane, or we say the three camera frames are coincident. In Eqn. (9), (10), (11), b_i^k remains invariant when i runs over $i = 1, 2, 3$, which results in the vanishing of D_k , E_k and F_k in Eqn. (18), (19), (20). Taking the camera coordinate system as reference, we set $R = I$, the 3×3 Identity matrix and $T = (0, 0, 0)^T$. Let $d_i = (u_i^2 + v_i^2 + 1)^{1/2}$, which is the distance from the i -th image point to the optical center. Let $\cos\theta_i = (u_i u_j + v_i v_j + 1)/(d_i d_j)$, which is the cosine of the angle of two optical rays formed by two image points with the optical center respectively. By scaling s_i as $a = d_1 s_1$, $b = d_2 s_2$ and $c = d_3 s_3$, we rewrite the Eqn. (12), (13), (14) as:

$$a^2 + b^2 - 2ab \cos \theta_1 = L_1^2 , \quad (21)$$

$$b^2 + c^2 - 2bc \cos \theta_2 = L_2^2 , \quad (22)$$

$$a^2 + c^2 - 2ac \cos \theta_3 = L_3^2 , \quad (23)$$

here we see a, b, c are the distances of three vertices to the optical center. For these three equations a closed form solution is derived in [4]. Usually, there are four complex solutions and four real solutions. For the four real solutions, two of them are positive and the other two are just isomorphic negative ones, which are unfeasible since the 3D points are not in front of the camera. In some configurations there may exist four feasible solutions, see [4].

3.2 A Stereo Camera

As shown in the right of Figure 1, two points are projected in the first image plane and the third one is projected on the second image plane. Without loss of generality, we can take the first camera as the reference with $R_1 = I$, $T_1 = (0, 0, 0)^T$. It can be shown that the rotation and translation from camera 1 to camera 2 coordinate system can be expressed as $R_2 = R_2^o R_1^{oT}$ and $T_2 = -R_2^o R_1^{oT} T_1^o + T_2^o$ respectively, where R_1^o and T_1^o are the rotation and translation from the world coordinate system to camera 1 coordinate system, R_2^o and T_2^o are that to camera 2 coordinate system.

For this stereo camera situation, the symbolic computation in our computer is not fast enough for deriving an analytical solution. The way we deal with this situation is that once the image measurements are obtained, the coefficients in Eqn. (12)~(14) are determined. The numeric algorithms in Maple or Matlab are exploited to find the solutions.

3.3 A Triplet Camera

As shown in the left of Figure 1, for the triplet camera configuration, the equations (12), (13), (14) are three quadratic equations with the most general form among the different P3P cases. There exist as many as eight solutions. Usually, no more than four solutions are feasible. But in some special configurations, all the eight solutions can be in front of the cameras. As the stereo camera situation, our computer did not derived a closed form of solutions within two days computation, thus, we use some numeric algorithms in Maple or Matlab to derive the solutions for the measured image points.

4 Experiments

A. At first we present an experiment with simulated data. Three vertices of a cube with their coordinates as $M_1 = (1, 1, 1)^T$, $M_2 = (1, -1, -1)^T$ and $M_3 = (-1, 1, 1)^T$ form a triangle. Three cameras with a focal length 1 are simulated, their optical centers are randomly located between the sphere with radius 3 and the sphere with radius 8, and with their viewing directions facing to the origin $(0, 0, 0)^T$. The first test is done with all three vertex images coming from the first camera, i.e., the classical P3P problem. Then in the second test we collect two vertex images from the first camera and the third vertex image from the second camera, which means a stereo camera case. While in the third test, each camera of the triplet contributes one vertex image, a triplet camera case. For the two or three camera case, we take the first camera as the reference camera and calculate respective rotation and translation of the second and the third ones as presented in Section 3.2. The estimated triangle poses are compared with the ground truth. In order to simulate the measurement noise, we add gaussian noise with different deviations. Let $Imax = \max(|u_i|, |v_i|)$, $i = 1, 2, 3$ represent the maximum value of the absolute coordinates among all image points in the three images, then we set the signal noise ratio as $SNR(\text{dB}) = 20 \log \frac{Imax}{\sigma}$, where σ is

the deviation of the gaussian noise. The results of the estimated poses in terms of mean absolute angles errors are shown in Table 1, 2 and 3, where the three rotation angles are calculated from the rotation matrix as $\phi = -\sin^{-1}(R^{3,3})$, $\theta = \tan^{-1}(R^{3,2}/R^{3,3})$, $\psi = \tan^{-1}(R^{2,1}/R^{1,1})$. The experimental results show that pose estimations for the three different cases have similar level of accuracy, which is out of our intuitive expectation. The stereo and triplet camera case should has better accuracy than single camera case in that for the single camera case the three optical rays are closer to parallel. The analysis on the sensitivity of poses to the image noise will be the future work.

Table 1. Poses of the triangle with the single camera

SNR (dB)	$\Delta\theta$ (deg)	$\Delta\phi$ (deg)	$\Delta\psi$ (deg)	Δt_1	Δt_2	Δt_3
48	0.49	0.78	0.65	0.006	0.007	0.067
36	1.51	1.99	2.03	0.042	0.029	0.246
24	13.49	16.54	11.56	0.153	0.139	1.445
12	28.87	34.96	45.27	0.336	0.388	2.384

Table 2. Poses of the triangle with the stereo camera

SNR (dB)	$\Delta\theta$ (deg)	$\Delta\phi$ (deg)	$\Delta\psi$ (deg)	Δt_1	Δt_2	Δt_3
48	1.04	1.06	0.45	0.036	0.008	0.023
36	1.83	3.18	3.90	0.131	0.033	0.048
24	13.65	12.84	11.96	0.366	0.139	0.411
12	30.55	54.55	39.20	0.795	0.360	1.826

Table 3. Poses of the triangle with the triplet camera

SNR (dB)	$\Delta\theta$ (deg)	$\Delta\phi$ (deg)	$\Delta\psi$ (deg)	Δt_1	Δt_2	Δt_3
48	0.48	0.59	0.53	0.024	0.004	0.012
36	1.99	2.10	1.27	0.091	0.026	0.041
24	6.65	8.95	4.53	0.299	0.093	0.109
12	33.92	42.96	29.03	0.676	0.663	0.730

B. In practical, one typical application of minimal pose estimation is to deal with the problem with outliers. This test is to evaluate our algorithm when outliers are involved. We propose two step approach: 1) Use extended P3P for outlier

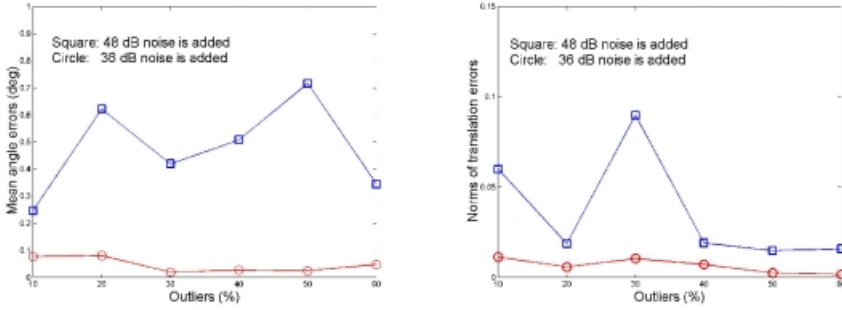


Fig. 2. The mean angle errors w.r.t outliers (left) and the translation errors w.r.t outliers (right)

removal. 2) Use standard PNP on the left inliers. In the experiment, 100 3D points are generated randomly around the origin with the coordinates ranges as $(X, Y, Z) \in [-1, 1]$. A camera is randomly generated in the exact same way as in the first test. For the projected 2D image points, by exchanging some indices, we get the wrong correspondence between 3D and 2D points. We set the percentages of outliers, denoted as p , from 10% to 60% in a step of 10%. In order to find all the inliers to carry out the pose estimation, the RANSAC method is exploited. Every time we choose 3 correspondences randomly, use our algorithm to calculate the rotation and translation parameters. No matter how many solutions there are, all the real and positive solutions are used to check all the correspondences. Each solution is scored by the number of its fitting correspondences. Repeat this procedure to N times, finally, the highest scored solution and its fitting correspondences can be used to derive an optimized solution by least square methods, see [1,2]. The number N can be derived as follows, if we repeat N times and the probability of all the three selected points to be inliers is larger than 0.99, the relation is formulated as, $1 - (1 - (1 - p)^3)^N > 0.99$. For the different p of outliers, the minimum number N is listed in Table 4. The results of the

Table 4. The minimum trial number vs. outliers percentage

p (%)	10	20	30	40	50	60
N	4	7	11	19	35	70

estimated angle errors and translation errors are shown in the left and right of Figure 2, in which the errors are represented as: $\Delta R = (|\Delta\theta| + |\Delta\phi| + |\Delta\psi|) / 3$, $\Delta T = \sqrt{(\Delta t_1)^2 + (\Delta t_2)^2 + (\Delta t_3)^2}$.

5 Summary and Conclusions

In this paper, we describe the P3P problem in a new viewpoint. The classical P3P problem is extended to the multiple camera cases. A formulation is proposed to solve this extended P3P problem in a uniform way for the case of single, stereo and triplet cameras. The experiments show the robustness of our approach to different noise level, as well as the evaluation for outliers.

Acknowledgments. This research is in part funded by the ARTHUR project under the European Commissions IST program (IST-2000-28559). This support is gratefully acknowledged.

References

1. David G. Lowe, "Fitting parameterized three-dimensional models to images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 5, pp. 441–450, May 1991.
2. Robert M. Haralick, Hyonam Joo, Chung-Nan Lee, XinHua Zhuang, Vinay G. Vaidya, and Man Bae Kim, "Pose estimation from corresponding point data," *IEEE Transactions on System, Man, and Cybernetics*, vol. 19, no. 6, pp. 1426–1446, December 1989.
3. Robert M. Haralick, Chung-Nan Lee, Karsten Ottenberg, and Michael Nolle, "Analysis and solution of the three point perspective pose estimation problem," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 592–598, 1991.
4. M. Fischler and R. Bolles, "Random sampling consensus: a paradigm for model fitting with application to image analysis and automated cartography," *Commun. Assoc. Comp. Mach.*, vol. 24, pp. 381–395, 1981.
5. Radu Horaud, Bernard Conio, and Oliver Leboulleux, "An analytic solution for the perspective 4-point problem," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 47, pp. 500–507, 1989.
6. Daniel F. Dementhon and Larry S. Davis, "Model-based object pose in 25 lines of code," *International Journal of Computer Vision*, vol. 15, pp. 123–141, 1995.
7. Daniel Dementhon and Larry S. Davis, "Exact and approximate solutions of the perspective-three-point problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1100–1105, November 1992.
8. J.A. Grunert, "Das pothenotische problem in erweiterter gestalt nebst bber seine anwendungen in der geodäsie," *Grunerts Archiv für Mathematik und Physik Band 1*, pp. 238–248, 1841, Band1.