A general recursive linear method for the perspective-n-point problem

De Xu [a,b], **You Fu Li** [b*], **Min Tan** [a]

[a] The Key Laboratory of Complex System and Intelligence Science,

Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P. R. China

[b] Department of Manufacturing Engineering and Engineering Management,

City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong

Email: sdxude@yahoo.com, meyfli@cityu.edu.hk, tan@compsys.ia.ac.cn

(*Corresponding author)

Abstract

In this paper, a new method for solving the perspective-n-point (PnP) problem is developed. With an auxiliary point, the linear method for the special case of four coplanar points is extended to find the coarse solutions for the general P3P problem. A recursive least square algorithm with a forgetting factor is introduced to find all the accurate solutions for the P3P problem. Then the algorithm is extended to the general PnP problem. The solution stability issues are investigated for the P3P, P4P and P5P problem respectively. Furthermore, a pattern is designed to ensure unique solution for the PnP problem. Experiments are performed to verify the effectiveness of the proposed method.

Keywords: Perspective-n-point problem, pattern design, three-dimensional sensing, pose estimation, visual positioning, recursive least square, solution distribution, solution stability.

1 Introduction

The Perspective-n-Point (PnP) problem was presented by Fischler and Bolles [1]. They defined it as to find the lengths of the line segments joining the center of perspective (CP) to each of the control points according to the relative positions of n control points and the angle to every pair of control points from the perspective point. This problem is also known as position and pose estimation with given points, which is to determine the position and orientation of the camera with respect to a scene object from n correspondent points [2].

As pointed out by Hu and Wu [3], there are some differences between the two definitions. The definition in [2] is stricter than that in [1]. If the transformation matrix is determined, the lengths of the line segments joining CP to each of the control points can also be obtained. Therefore, the definition in [2] is more popular for visual control. In this paper, we will adopt this definition to deal with the PnP problem.

The PnP problem is very important to many applications [4] such as automated cartography, robotics and automation, computer vision, computer animation [5], and photogrammetry [6]. It has been attracting researchers' attentions in the computer vision community. As fewer points provide more flexible in applications, many researchers have put their efforts into resolving the P4P and P3P problem.

In [2] an analytic solution method for the P4P problem was provided, which converted the problem into solving a biquadratic polynomial equation with one unknown variable. Although the equation can be solved with iterative method, the efficiency is low. Holt and Netravali [7] proposed a method to find the solution of transformation matrix for the P4P problem, in which three views were employed. However, the problem of multiple solutions was not addressed. Abidi and Chandra [8] presented an algorithm for pose estimation based on six lines and four points of a tetrahedral object. Quan and Lan [9] developed a linear method to solve the P4P and P5P problem. Hu and Wu [3] proved that there are 4 solutions at most for the P4P problem based on Horaud's definition, and 5 solutions at most based on the definition in [1]. Therefore, the method in [9] is not complete because it can only provide one solution.

The major methods in direct solutions to the P3P problem were reviewed in [10]. All methods involved high order nonlinear equations, which might result in unstable solutions. The authors analyzed the stability of previous methods and presented an analytical method to produce stable solutions. Wolfe and Mathis [11] discussed the multiple-solution issue for the P3P problem, and gave a geometric explanation of the camera triangle configurations that cause one, two, three, or four solutions. Alter [12] presented an approach to compute 3-D pose from 3 points using weak perspective method, which is based on the distances between the matched model and image points. DeMenthon and Davis [13] investigated the approximate solutions of the P3P problem. Comparison with the weak perspective method, para-perspective and ortho-perspective approximations produce lower errors for off-center image features. Juang [14] provided a method to deal with the P3P problem using biquadratic polynomial equation without sinusoidal function, i.e. no more use of the cosines law in order to improve computing accuracy. Moriya and Takeda [15] reported an application example of P3P algorithm used in an active visual system to estimate the relative rotation of cameras. In the system, there were two cameras whose translation was fixed and known. Two views with known translation were available. This kind of configuration simplified the P3P problem. However, it is just a special case for the P3P problem. Gao and Chen [16] decomposed the equation system for the P3P problem into triangle sets and developed geometric criteria for the solution classification and the number of solutions. They also gave the criteria for the P3P problem to have one, two, three, and four solutions based on algebraic approach that converted the P3P problem into a nonlinear equation and two linear inequalities. Nister [17] took the P3P problem as finding the intersections between a ruled quartic surface and a circle, and derived an eighth order polynomial equation whose roots correspond to the solutions.

Up to now, almost all of the methods for finding the solutions of the P3P problem involve solving high order nonlinear equations. There is no effective algorithm that can be extended to all PnP problem such as n=3, 4, 5. Another problem is that the existence of multiple solutions in visual positioning based on PnP has limited the applications of most of the solution methods.

This work attempts to develop a general linear method for finding all the solutions for the PnP problem. We design a pattern to ensure the uniqueness of the solution for the PnP problem. The rest of the paper is arranged as follows. Section 2 describes the frame assignments of the visual system and the camera model. In Section 3, the linear method to solve the coplanar P4P problem is discussed. A new method to solve the P3P problem is developed in Section 4. With an auxiliary point, the linear method for the special case of four coplanar points is extended to find the coarse solutions for the general P3P problem. Then a recursive least square algorithm with a forgetting factor is introduced to find all the accurate solutions for the general P3P problem. The solution distributions are also investigated. The algorithm is extended to the general PnP problem in Section 5. The solution stability for the P3P, P4P and P5P problem are studied. In Section 6, a pattern to ensure a unique solution for the PnP problem is designed. The experiments are presented in Section 7 to verify the effectiveness of the proposed method. Finally, the paper is concluded in Section 8.

2 System model

2.1 Coordinates frame assignments

Assume that there are three points whose positions are known in the world frame. The coordinates of a point P_i are denoted as P_i=[x_{wi}, y_{wi}, z_{wi}] in the world frame. The point, whose horizontal coordinate on the image captured by the camera is smallest, is chosen as the origin of the reference frame. The vector from the origin to the point with the largest horizontal coordinate on the image is assigned to X-axis direction of the reference frame, i.e. the X_r axis. From the remaining point to the vector, a line perpendicular to the vector can be drawn, with an intersection point between this line and the vector. The direction from the intersection point to the remaining point is assigned to the Y_r axis of the reference frame. The Z_r axis is determined with the right hand rule according to the X_r and Y_r axes. Without losing generality, we assume that the origin of the reference frame is point P_1, and X_r is the direction from point P_1 to P_2, as shown in Fig.1. In addition, the camera frame is assigned as follows. Its origin is assumed to be at the optical centre of the camera. Its Z-axis, i.e. Z_r, is from the camera to the scene along its optical axis. Its X-axis is selected as the horizontal direction of its imaging plane from left to right, and Y-axis as the vertical direction of its imaging plane from top to bottom.

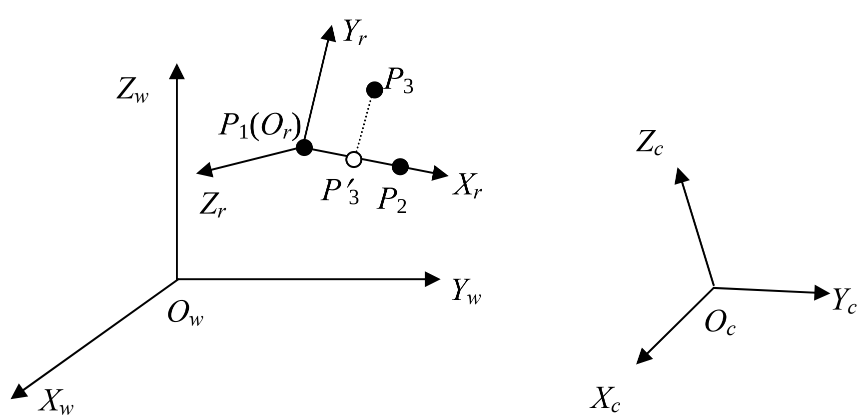


Fig.1 The world frame and the camera frame

2.2 The camera model

The pinhole model is adopted to describe the camera, assuming negligible lens distortion. The four parameters intrinsic model of a camera can be given as

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & u_0 \\ 0 & k_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c / z_c \\ y_c / z_c \\ 1 \end{bmatrix} = M_{in} \begin{bmatrix} x_c / z_c \\ y_c / z_c \\ 1 \end{bmatrix} \tag{1}$$

where $[u, v]$ are the coordinates of a point in an image, $[u_0, v_0]$ denote the image coordinates of the camera's principal point, $[x_c, y_c, z_c]$ are the coordinates of a point in the camera frame, M_{in} is the intrinsic parameter matrix, and $[k_x, k_y]$ are the magnification coefficients from the imaging plane coordinates to the image coordinates.

Assume that the camera frame is denoted as C, and the world frame as W. The transformation from C to W is known as the extrinsic parameters for the camera.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} {}^c n_{wx} & {}^c o_{wx} & {}^c a_{wx} & {}^c p_{wx} \\ {}^c n_{wy} & {}^c o_{wy} & {}^c a_{wy} & {}^c p_{wy} \\ {}^c n_{wz} & {}^c o_{wz} & {}^c a_{wz} & {}^c p_{wz} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = {}^c M_w \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{2}$$

where $[x_w, y_w, z_w]$ are the coordinates of a point in the world frame, and ${}^cM_w$ is the extrinsic parameter matrix of the camera, i.e. the transformation from C to W. In ${}^cM_w$, ${}^c n_w = \begin{bmatrix} {}^c n_{wx} & {}^c n_{wy} & {}^c n_{wz} \end{bmatrix}^T$ is the direction vector of the X-axis, ${}^c o_w = \begin{bmatrix} {}^c o_{wx} & {}^c o_{wy} & {}^c o_{wz} \end{bmatrix}^T$ is that of the Y-axis, ${}^c a_w = \begin{bmatrix} {}^c a_{wx} & {}^c a_{wy} & {}^c a_{wz} \end{bmatrix}^T$ is that of the Z-axis for the frame W expressed in the frame C, and ${}^c p_w = \begin{bmatrix} {}^c p_{wx} & {}^c p_{wy} & {}^c p_{wz} \end{bmatrix}^T$ is the position vector.

3 The linear method for four points on a plane

Assume that there are four points P_1 to P_4 on the plane $X_rO_rY_r$. The reference frame is shown in Fig.1. The coordinates of the points are denoted as $P_{ri}=[x_{ri}, y_{ri}, 0]$ in the reference frame, which can be calculated according to their positions in the world frame.

From equation (1), the coordinates of a point in the camera frame satisfy the following equation:

$$\begin{bmatrix} x_{ci}/z_{ci} \\ y_{ci}/z_{ci} \\ 1 \end{bmatrix} = M_{in}^{-1} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} 1/k_x & 0 & -u_0/k_x \\ 0 & 1/k_y & -v_0/k_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \tag{3}$$

Applying P_{ri} to equation (2), then

$$\begin{cases} x_{ci} = {}^c n_{rx} x_{ri} + {}^c o_{rx} y_{ri} + {}^c p_{rx} \\ y_{ci} = {}^c n_{ry} x_{ri} + {}^c o_{ry} y_{ri} + {}^c p_{ry} \\ z_{ci} = {}^c n_{rz} x_{ri} + {}^c o_{rz} y_{ri} + {}^c p_{rz} \end{cases} \tag{4}$$

Applying equation (4) to equation (3), we have

$$\begin{cases} x_{ri}\,{}^c n_{rx} + y_{ri}\,{}^c o_{rx} - x_{1ci} x_{ri}\,{}^c n_{rz} - x_{1ci} y_{ri}\,{}^c o_{rz} + {}^c p_{rx} - x_{1ci}\,{}^c p_{rz} = 0 \\ x_{ri}\,{}^c n_{ry} + y_{ri}\,{}^c o_{ry} - y_{1ci} x_{ri}\,{}^c n_{rz} - y_{1ci} y_{ri}\,{}^c o_{rz} + {}^c p_{ry} - y_{1ci}\,{}^c p_{rz} = 0 \end{cases} \tag{5}$$

where

$$\begin{cases} x_{1ci} = x_{ci}/z_{ci} = (u_i - u_0)/k_x \\ y_{1ci} = y_{ci}/z_{ci} = (v_i - v_0)/k_y \end{cases} \tag{6}$$

For n known points, we have n groups of equation set (5), i.e. $2n$ equations. These equations can be rewritten as

$$A_1 H_1 + A_2 H_2 = 0 \tag{7}$$

where A_1 is a $2n\times3$ matrix, A_2 is a $2n\times6$ matrix,

$$A_1 = \begin{bmatrix} x_{r1} & 0 & -x_{1c1}x_{r1} \\ 0 & x_{r1} & -y_{1c1}x_{r1} \\ \vdots & \vdots & \vdots \\ x_{rn} & 0 & -x_{1cn}x_{rn} \\ 0 & x_{rn} & -y_{1cn}x_{rn} \end{bmatrix}, A_2 = \begin{bmatrix} y_{r1} & 0 & -x_{1c1}y_{r1} & 1 & 0 & -x_{1c1} \\ 0 & y_{r1} & -y_{1c1}y_{r1} & 0 & 1 & -y_{1c1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{rn} & 0 & -x_{1cn}y_{rn} & 1 & 0 & -x_{1cn} \\ 0 & y_{rn} & -y_{1cn}y_{rn} & 0 & 1 & -y_{1cn} \end{bmatrix},$$

$$H_1 = \begin{bmatrix} {}^c n_{rx} & {}^c n_{ry} & {}^c n_{rz} \end{bmatrix}^T, H_2 = \begin{bmatrix} {}^c o_{rx} & {}^c o_{ry} & {}^c o_{rz} & {}^c p_{rx} & {}^c p_{ry} & {}^c p_{rz} \end{bmatrix}^T.$$

H_1 is just the vector ${}^c n_r$ and is a unit vector. With the constraint $\|H_1\| = 1$, the solutions of H_1

and H_2 can be obtained with linear method [18]. Then the extrinsic parameters matrix ${}^cM_r$, i.e. the transformation from the camera to reference frame, is obtained, in which the third column can be given by the cross product of the first and second column.

For four known points on a plane, A_1 is an 8×3 matrix, and A_2 is an 8×6 matrix. The solutions of H_1 and H_2 can be uniquely determined. More details about this can be found in the Appendix. A sufficient condition for the linear solution method of four points on a plane is given as Theorem 1.

Theorem 1: For four coplanar points, *iff*

(1) their positions are known in the world frame,

(2) and any three points among them are non-collinear,

then the distances from each point to the camera perspective center can be computed using a linear method with one view of a camera with known intrinsic parameters.

In addition, the extrinsic parameters of the camera in the world frame can be calculated via

$$ {}^cM_w = {}^cM_r \, {}^rM_w \tag{8} $$

where ${}^rM_w$ is the transformation from the world to reference frame, i.e. the position and pose of the world frame expressed in the reference frame; ${}^cM_r$ is the transformation from the camera to reference frame.

4 The recursive linear method for the P3P problem

4.1 The recursive linear method

As we know, it is not sufficient to use three known points to solve the extrinsic parameters from equation (7) with a linear method. Here, an auxiliary point with the average positions of the three points is introduced as the fourth known point. Because it is an invisible point, its image coordinates can not be directly found in the image. The average image coordinates of the three points are assigned to it instead. The four points satisfy the conditions of Theorem 1. Therefore, the extrinsic parameters matrix ${}^cM_r$ can be computed with the linear method in Section 3, which will be taken as initial values in following method. The results would introduce errors due to the inaccuracy of the image coordinates of the auxiliary point.

Equation (9) is deduced from the inner product constraint in the rotation matrix of ${}^cM_r$. For any of the three known points, equation (9) should be satisfied.

$$\begin{cases} {}^cn'_{rx}\,{}^cn'_{rx}+{}^cn'_{ry}\,{}^cn'_{ry}+{}^cn'_{rz}\,{}^cn'_{rz}-{}^co'_{rx}\,{}^co'_{rx}-{}^co'_{ry}\,{}^co'_{ry}-{}^co'_{rz}\,{}^co'_{rz}=0 \\ {}^cn'_{rx}\,{}^co'_{rx}+{}^cn'_{ry}\,{}^co'_{ry}+{}^cn'_{rz}\,{}^co'_{rz}=0 \end{cases} \tag{9}$$

where ${}^cn'_r={}^cn_r/{}^cp_{rz}$, ${}^co'_r={}^co_r/{}^cp_{rz}$, ${}^cp'_{rx}={}^cp_{rx}/{}^cp_{rz}$, ${}^cp'_{ry}={}^cp_{ry}/{}^cp_{rz}$.

According to equation (5) and (9), the parameters matrixes are defined as given in (10) to (12).

$$\varphi_i^T=\begin{bmatrix} x_{ri} & 0 & -x_{1ci}x_{ri} & y_{ri} & 0 & -x_{1ci}y_{ri} & 1 & 0 \\ 0 & x_{ri} & -y_{1ci}x_{ri} & 0 & y_{ri} & -y_{1ci}y_{ri} & 0 & 1 \\ k\,{}^cn'_{rx} & k\,{}^cn'_{ry} & k\,{}^cn'_{rz} & -k\,{}^co'_{rx} & -k\,{}^co'_{ry} & -k\,{}^co'_{rz} & 0 & 0 \\ 0 & 0 & 0 & k\,{}^cn'_{rx} & k\,{}^cn'_{ry} & k\,{}^cn'_{rz} & 0 & 0 \end{bmatrix} \tag{10}$$

$$Y_i=\begin{bmatrix} x_{1ci} & y_{1ci} & 0 & 0 \end{bmatrix}^T \tag{11}$$

$$\Theta^T=\begin{bmatrix} {}^cn'_{rx} & {}^cn'_{ry} & {}^cn'_{rz} & {}^co'_{rx} & {}^co'_{ry} & {}^co'_{rz} & {}^cp'_{rx} & {}^cp'_{ry} \end{bmatrix}^T \tag{12}$$

where k is a factor to reduce the errors in pose via enhancing the role of pose errors.

The initial value P_N for the Recursive Least Square (RLS) method is evaluated via formula (13). The four points including the three known points and the auxiliary point are employed to calculate P_N. The extrinsic parameters computed with the four points are also introduced as the initial values in the calculation of P_N through matrix φ_i.

$$P_N = (\Phi_N^T \Phi_N)^{-1} \tag{13}$$

where

$$\Phi_N = \begin{bmatrix} \varphi_1^T & \varphi_2^T & \varphi_3^T & \varphi_4^T \end{bmatrix}^T \tag{14}$$

The RLS method with a forgetting factor is introduced to calculate the accurate extrinsic parameters. This is described via the following

$$K_{N+1} = P_N \varphi_{N+1} (\rho^2 I + \varphi_{N+1}^T P_N \varphi_{N+1})^{-1} \tag{15}$$

$$\Theta_{N+1} = \Theta_N + K_{N+1} (Y_{N+1} - \varphi_{N+1}^T \Theta_N) \tag{16}$$

$$P_{N+1} = (P_N - K_{N+1} \varphi_{N+1}^T P_N) / \rho^2 \tag{17}$$

where ρ is the forgetting factor.

With formulas (10), (12), (15), (16) and (17), Θ is recursively calculated according to the position and image coordinates of each of the three known points until convergence is reached.

After the accurate Θ is obtained, the parameters are calculated in formula (18).

$$\begin{cases} {}^c n_r = {}^c p_{rz} \ {}^c n'_r, \quad {}^c o_r = {}^c p_{rz} \ {}^c o'_r, \quad {}^c a_r = {}^c n_r \times {}^c o_r \\ {}^c p_{rx} = {}^c p_{rz} \ {}^c p'_{rx}, \quad {}^c p_{ry} = {}^c p_{rz} \ {}^c p'_{ry}, \quad {}^c p_{rz} = 1 / \left\| {}^c n'_r \right\| \end{cases} \tag{18}$$

To improve the performance, a forgetting factor ρ is designed as a varying factor as given in (19) during the iterations. With the varying forgetting factor, the influence of the auxiliary point will quickly diminish, whereas the roles of the known points will remain.

$$\rho = \begin{cases} 0.02(j-1) + 0.6, \ j < 21 \\ 1, \ j >= 21 \end{cases} \tag{19}$$

Furthermore, an error function as given in (20) was designed to monitor the recursive process. Once the absolute value of *Err* is less than a given threshold, the recursive process will be stopped.

$$Err = \sum_{i=1}^{3}\left[(u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2\right] + k[(\left\|{}^c n_r\right\|^2 - 1)^2 + (\left\|{}^c o_r\right\|^2 - 1)^2 + (\left\|{}^c a_r\right\|^2 - 1)^2 + ({}^c n_r^T\,{}^c o_r)^2 + ({}^c o_r^T\,{}^c a_r)^2 + ({}^c a_r^T\,{}^c n_r)^2] \tag{20}$$

Here *Err* is the error index, $[\hat{u}_i, \hat{v}_i]$ are the estimated image coordinates of point P_i according to the calculated extrinsic parameters of the camera and the position of the point, $[u_i, v_i]$ are the measured image coordinates.

Here some case studies were conducted. The iteration in all simulations converged at several steps. The error curve of a case study with the method above is given in Fig.2. It can be seen that the recursive process converged very fast. The recursive algorithm was tried 20 times. The process converged at the fourth recursion. This indicates that the algorithm promises good performance for on-line computation. The RLS can be converged in several steps because the initial values are close to the real solutions and the forgetting factor enhances the roles of known points.

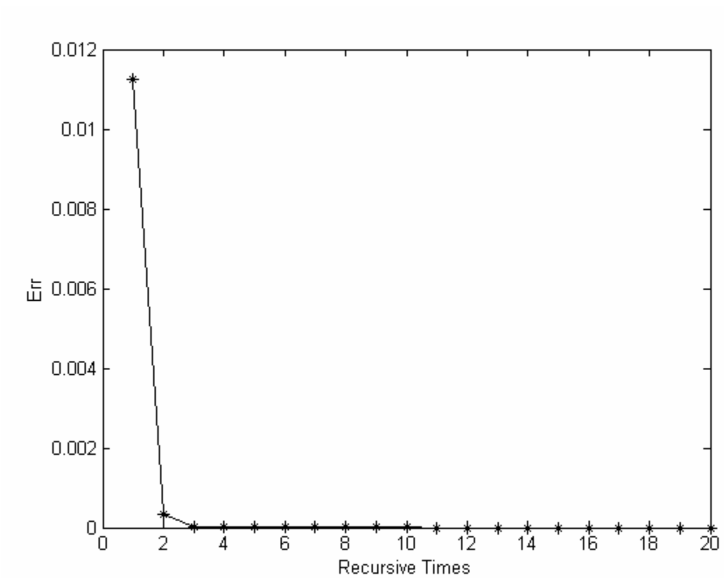
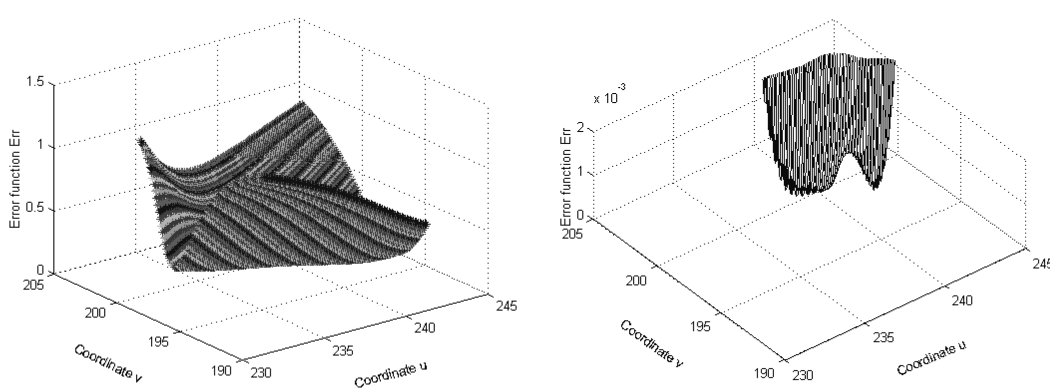


Fig.2 The error curve in RLS

4.2 Solution distribution

To investigate the solution distribution for the P3P problem, we evaluated the auxiliary point with the image coordinates of different image points in an area. The image point with the average image coordinates of the three known points was taken as the center of the area. The size of the area was assigned as 10×10 pixels. The solutions were computed with the method described in Section 3. An error function as given in formula (20) was employed to assess the solutions.

If the error *Err* is zero, then a solution is found. Therefore, the distribution of the error *Err* indicates the distribution of solutions. Since only the error distribution was investigated, the RLS method was not used in this simulation. With different relative relations between the world frame and the camera frame, there were one, two, three or four solutions found for the P3P problem. Fig.3 shows the error distribution for the P3P problem with two solutions. Fig.3 (a) and (b) show the same distribution, but in different scales. Fig.3 (a) is shown in the normal scale. Fig.3 (b) only shows the errors less than 10^{-3}. It can be seen from Fig.3 that the error function is smooth and there are two global minimum points corresponding to the two solutions. In addition, the solutions are distributed in a small area.



(a) Errors shown in normal　　　(b) Errors less than 10^{-3}

Fig.3 The solution error distribution with two solutions

4.3 The multiple-solution problem

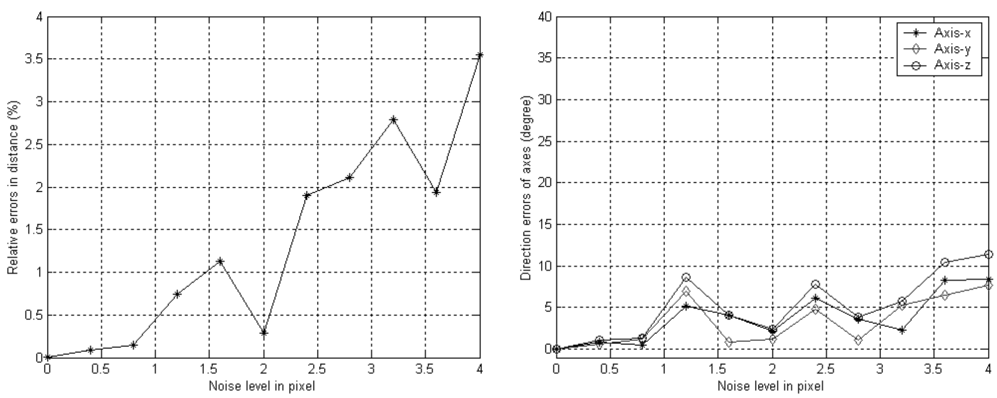
Previous researchers have proved that there exist multiple solutions for the PnP problem. For example, Wolfe and his collaborators [11] provided the geometric explanations of one, two, three, and four solutions for the P3P problem. In our method proposed above, it seems that only one solution is obtained. However, the initial values of P_N and Θ_N can be generated with different image offsets in a limited range, which are added to the image coordinates of the auxiliary point. For example, a circle or square can be drawn with 2 pixels radius at the auxiliary point as its center. The image offsets can then be selected from the edge of a circle or square. Since the error distribution is a smooth curved surface with several global minimum points corresponding to the solutions, all solutions can be found out using the RLS method in Section 4.1 with different initial values of P_N and Θ_N.

Combining the camera's intrinsic parameters with the positions of the three given points in the world frame, their coordinates in the camera frame were computed with the methods above while the image offsets were selected from the edge of a square. The techniques have been tested on random sets of points with image errors incorporated, the complete solutions can be found correctly. As a comparison, the same case was tested with the analytic method provided in [14]. The results turned out to be exactly the same as ours.

4.4 The solution stability

The solution stability of the method proposed was investigated for the P3P, P4P and P5P problem. In the simulations, the factor k was set to 10000 to enhance the role of pose errors in the process of the RLS. If the difference in the error *Err* between two iteration steps was less than 10^{-6}, then the recursive process was stopped. The number of recursive steps was limited to 40. The noise added to the image coordinates was random noise with a uniform distribution on the interval (0.0, N) where N is the noise level. The position errors were relative errors calculated from the actual and measured distances between the optical center of the camera and the origin of the world frame. The orientation errors were given in degrees showing the angles between the actual and measured directions of the axes of the world frame.

Fig.4 gives the position and orientation errors of the solutions for the P3P problem with our method with noise disturbance. Fig.4 (a) shows the relative position errors in percentage for different noise levels, and Fig.4 (b) gives the error curves of the axes' direction angle of the world frame expressed in the camera frame. It can be seen that the methods exhibit good stability against noise disturbance.



(a) Position errors with noise (b) Direction errors with noise

Fig.4 The solution errors for the P3P problem with noise

5 The recursive linear method for the PnP problem

Considering the general case that there are more than three known points, the method proposed in Section 4.1 can be extended to the Perspective-n-Point (PnP) problem. Any three points that are non-collinear can be selected to compute the rough extrinsic parameters. Combining an auxiliary point as described in Section 4.1, H_1 and H_2 can be solved from formula (7). Then the RLS method is given as follows.

$$\varphi_i^T = \begin{bmatrix} x_{ri} & 0 & -x_{1ci}x_{ri} & y_{ri} & 0 & -x_{1ci}y_{ri} & z_{ri} & 0 & -x_{1ci}z_{ri} & 1 & 0 \\ 0 & x_{ri} & -y_{1ci}x_{ri} & 0 & y_{ri} & -y_{1ci}y_{ri} & 0 & z_{ri} & -y_{1ci}z_{ri} & 0 & 1 \\ k\,{}^c n'_{rx} & k\,{}^c n'_{ry} & k\,{}^c n'_{rz} & -k\,{}^c o'_{rx} & -k\,{}^c o'_{ry} & -k\,{}^c o'_{rz} & 0 & 0 & 0 & 0 & 0 \\ k\,{}^c n'_{rx} & k\,{}^c n'_{ry} & k\,{}^c n'_{rz} & 0 & 0 & 0 & k\,{}^c a'_{rx} & k\,{}^c a'_{ry} & k\,{}^c a'_{rz} & 0 & 0 \\ 0 & 0 & 0 & k\,{}^c n'_{rx} & k\,{}^c n'_{ry} & k\,{}^c n'_{rz} & 0 & 0 & 0 & 0 & 0 \\ k\,{}^c a'_{rx} & k\,{}^c a'_{ry} & k\,{}^c a'_{rz} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{21}$$

$$Y_i = \begin{bmatrix} x_{1ci} & y_{1ci} & 0 & 0 & 0 & 0 \end{bmatrix}^T \tag{22}$$

$$\Theta^T = \begin{bmatrix} {}^c n'_{rx} & {}^c n'_{ry} & {}^c n'_{rz} & {}^c o'_{rx} & {}^c o'_{ry} & {}^c o'_{rz} & {}^c a'_{rx} & {}^c a'_{ry} & {}^c a'_{rz} & {}^c p'_{rx} & {}^c p'_{ry} \end{bmatrix}^T \tag{23}$$

where ${}^c n'_r = {}^c n_r / {}^c p_{rz}$, ${}^c o'_r = {}^c o_r / {}^c p_{rz}$, ${}^c a'_r = {}^c a_r / {}^c p_{rz}$, ${}^c p'_{rx} = {}^c p_{rx} / {}^c p_{rz}$, ${}^c p'_{ry} = {}^c p_{ry} / {}^c p_{rz}$, k is a factor to reduce the errors in pose through enhancing the role of pose errors.

The accurate value of Θ can be calculated using the RLS method with a forgetting factor as described in Section 4.1 according to the position and image coordinates of each known point.

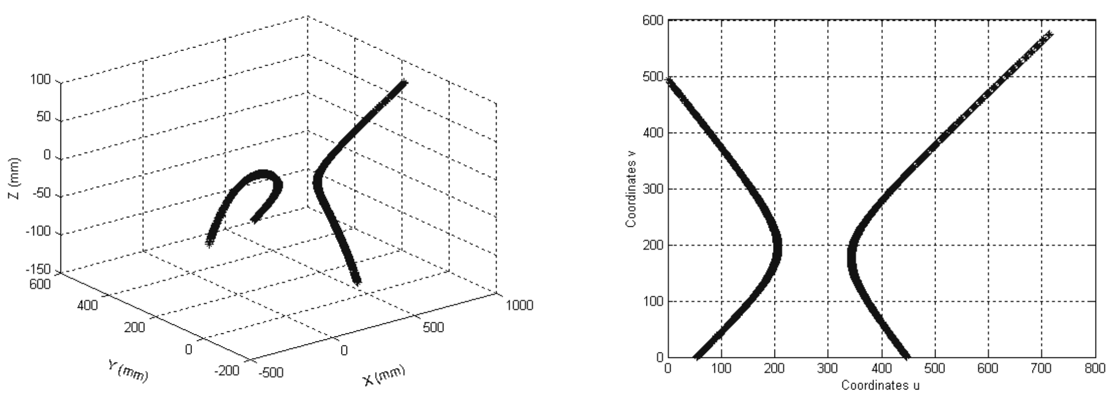
Similar to the method of dealing with the multiple-solution for the P3P problem, the initial values of P_N and Θ_N can be generated with different image coordinates of the auxiliary point, by adding an image offset in a limited range to the average image coordinates of the three known points. Then all solutions can be computed.

A set of known points and specified camera pose with two solutions were designed for case study to verify the proposed method. Although the results based on P3P, P4P and P5P were different in the simulations, they can be categorized into two groups corresponding to the two solutions. The results in each group were very close. This means that the general linear method proposed in this paper possesses the ability to find all solutions with high precision.

The stability for PnP problem with the above methods is investigated. The results show that the errors caused by random noise for the P3P and P4P problem are almost at the same level, regardless whether the points are coplanar or not for the P4P problem. The errors for the P5P problem are apparently smaller than those for the P3P or P4P problem. It means that the positioning results based on P5P are more robust than those for P3P or P4P with the same disturbance of random noise. In addition, the simulations show that the coplanar P4P has the best real-time performance among all the PnP problems.

6 Pattern design for unique solution

For *n* points, if the image coordinates of each point were the same in the two views of a camera, there would be two solutions in visual positioning based on PnP. To find the lower bound of the number of arbitrary points which can form unique solution, it is necessary to investigate the point distribution whose image coordinates in the two views are the same.



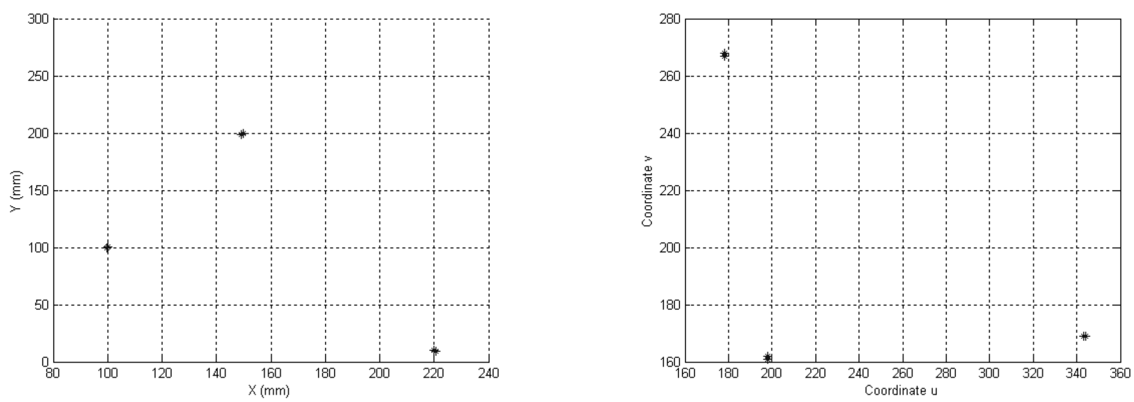
(a) The space points owning same image coordinates in two views

(b) The image coordinates in two views

Fig.5 The space points owning same image coordinates in two views and their image coordinates

Assume that the two views and the setup of the camera are given. For the same image coordinate in the two views, the corresponding space point can be calculated with stereo vision. For each view, a view line is formed between the optical center of the camera and the imaging point. If the two view lines intersect, then the space point has the same image coordinates in the two views. Fig.5 (a) shows the space points with the same image coordinates in the two views. Fig.5 (b) gives their image coordinates in the two views. We can find from Fig.5 that there are many points having the same image coordinates in the two views. In other words, no matter how many points there are, it is possible that there exist multiple solutions in the PnP problem.

To ensure unique solution for the PnP problem, additional constraints for the known space points are necessary. Generally, co-planarity is a simple and effective constraint. In Fig.5 (a), a section with z in the range [-0.1, 0.1] mm was extracted, which contains six points belonging to three groups. The points are shown in Fig.6 (a), and their image coordinates in the two views are displayed in Fig.6 (b). From Fig.6, a conclusion can be drawn that three points are not sufficient to ensure unique solution for the PnP problem.



(a) The points in a section

(b) Their image coordinates

Fig.6 The points in a section and their image coordinates in two views

Now, let us check the image coordinates of the centroid of the triangle formed by the three points. The image coordinates are [239, 200] and [242, 198] in the views as given. Obviously, the image coordinates of the triangle centroid are different although the coordinates of the three known points in the two views are the same. These four points constitute the smallest set, which can ensure the unique solution for the PnP problem. Generally, four coplanar points also provide one solution in most cases.

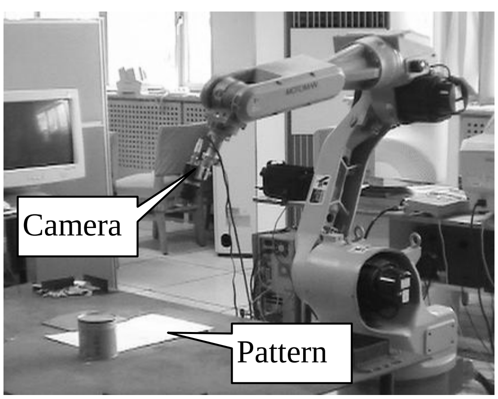
Based on the above analysis, a sufficient condition of the solution uniqueness for the PnP problem is given as Theorem 2. Its proof can be drawn from the discussions above and those in the Appendix.

Theorem 2: For four coplanar points, in which one is inside the triangle formed by other three, the solution for the PnP problem is unique.

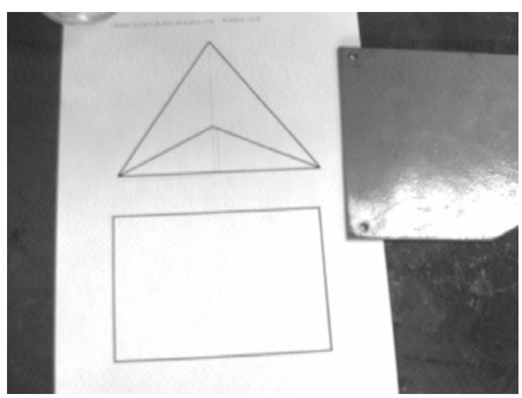
In a word, a four point pattern, in which one point is the centroid of the triangle formed by other three, is recommended for visual positioning based on PnP. This can provide unique solution, which is very important for robotics in model based visual control. Besides the four points in the pattern, more additional points are helpful for improving the positioning accuracy and robustness against noise.

7 Positioning Experiment

An experiment was designed to verify the effectiveness of the proposed pattern and methods. A calibrated camera was fixed on the end-effector of an industrial robot. The experiment scene is given in Fig.7 (a). The intrinsic parameters of the camera were given in formula (24). The pattern was given in Fig.7 (b). The coordinates of the four points in the triangle pattern were as shown in (25). The other points, such as ones in the rectangle, were used for verification of the proposed method. The image size was 768×576 in pixels. The unit of position was mm.



(a) Experiment scene



(b) A frame of image captured

Fig.7 Experiment with the proposed pattern

$$M_{in1}=\begin{bmatrix}921.19399 & 0 & 329.89242\\ 0 & 928.35038 & 320.01008\\ 0 & 0 & 1\end{bmatrix} \tag{24}$$

$$P_w=\begin{bmatrix}P_{w1} & P_{w2} & P_{w3} & P_{w4}\end{bmatrix}=\begin{bmatrix}0 & 133.4 & 63.5 & 63.5\\ 0 & 0 & 99.1 & 33.0\\ 0 & 0 & 0 & 0\end{bmatrix} \tag{25}$$

An image was captured and the imaging points of corners of the triangle and rectangle were extracted. The imaging points in the triangle pattern were employed to calculate the extrinsic parameters matrix ${}^{c}M_{r}$ via the methods described in Section 3 and 5. With the change of the position and pose of the end-effector, new images of the pattern were captured and ${}^{c}M_{r}$ were recalculated. Combining the new result of ${}^{c}M_{r}$ with the first one, the positions of the corner points in the reference frame could be obtained with stereovision. Then the lengths of lines formed by any group of two points were calculated, which are listed in Table 1. It can be seen that the measured results are very close to the actual ones. This indicates that the positioning results, i.e. the extrinsic parameters matrix ${}^{c}M_{r}$, via the proposed methods with the pattern have satisfactory accuracy.

Table 1 The visual measurement results based on the positioning method with the proposed pattern

Number of lines formed by any group of two points	1	2	3	4	5	6	7	8
Measured length (mm)	133.1	126.2	66.8	91.6	133.4	96.7	133.2	96. 0
	136.5	128.1	69.4	92.2	137.8	95. 2	134.6	99.9
Actual length (mm)	133.4	121.3	66.0	90.1	133.4	93.6	133.4	93.6

8 Conclusions

In this paper, a general recursive linear method for solving the PnP problem is developed. A point with the average coordinates of three known points is taken as an auxiliary point. With the four coplanar points including the auxiliary one, the coarse solutions for the general P3P problem can be computed with a linear method. Taking them as the initial values, the recursive least square algorithm with a varying forgetting factor is introduced to find all the accurate solutions for the general P3P problem. In the process of recursion, only the known points are taken into account. The method is extended to solve the general PnP problem. The RLS can converge in several steps because the initial values are close to the real solutions and the forgetting factor enhances the roles of known points.

Case studies were performed for the P3P, P4P and P5P problem. The case for the P3P problem was tested against an analytic method too and the results with both methods turned out to be exactly the same. As a distinct feature of our method, a cost function can be used to monitor the solution process. The multiple-solution problem is discussed for the PnP problem. The solution distribution is investigated, and the graphic explanations of the solution in image space are presented. The solution stability of the method is investigated for the P3P, P4P and P5P problem. A pattern to ensure unique solution for the PnP problem is proposed. This pattern consists of four coplanar points, in which one point is the centroid of the triangle formed by other three. The simulation results show that the positioning results based on P5P are more robust than those based on P3P or P4P in the presence of random noise disturbance. In addition, the experiments confirmed that the visual positioning based on PnP using the proposed pattern has a unique solution and it promises good real-time performance. The effectiveness of proposed method is also verified by the experiments.

Acknowledgement

The work described in this paper was supported by a grant from the Research Grants Council of Hong Kong [Project No. CityU117605]. The authors would like to thank the National High Technology Research and Development Program of China for supporting this work under grant 2002AA422160. We would also like to thank the National Key Fundamental Research and Development Project of China (973, No.2002CB312200) for their support.

References:

1. M. A. Fishler and R. C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartomated Cartography, Comm. ACM, vol. 24, no. 6, pp. 381-395, 1981.
2. R. Horaud, B. Conio, And O. Leboulleux, An Analytic Solutions for the Perspective 4-Point Problem, Computer Vision, Graphics, Image Processing, vol. 47, no.1, pp.33-44, Jul. 1989.
3. Z. Y. Hu and F. C. Wu, A Note on the Number of Solutions of the Noncoplanar P4P Problem, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp.550-555, Apr. 2002
4. X. S. Gao, X. R. Hou, J. L. Tang, and H. F. Cheng, Complete Solution Classification for the Perspective-Three-Point Problem, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 8, pp. 930-943, Aug. 2003
5. C. Su, Y. Xu, H. Li, and S. Liu, Application of Wu’s Method in Computer Animation, Proc. Fifth Int’l Conf. CAD/CG, vol. 1, pp. 211-215, 1997.
6. J. S. C. Yuan, A General Photogrammetric Method for Determining Object Position and Orientation, IEEE Trans. Robotics and Automation, vol. 5, no. 2, pp. 129-142, 1989.
7. R. J. Holt and A. N. Netravali, Uniqueness of Solutions to Three Perspective Views of Four Points, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, no. 3, pp. 303-307, Mar. 1995.
8. M. A. Abidi and T. Chandra, A New Efficient and Direct Solution for Pose Estimation Using

Quadrangular Targets: Algorithm and Evaluation, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, no. 5, pp. 534-538, May 1995.

9. L. Quan and Z. Lan, Linear N-Point Camera Pose Determination, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 8, pp. 774-780, Aug. 1999.

10. R. M. Harallick, C. N. Lee, K. Ottenberg, and M. Nolle, Analysis and Solutions of the Three Point Perspective Pose Estimation Problem, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 592–598, Maui, Hawaii, Jun. 1991.

11. W. J. Wolfe, D. Mathis, C. Weber, and M. Magee, The Perspective View of Three Points, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 13, no. 1, pp. 66-73, Jan. 1991.

12. T. D. Alter, 3-D Pose from 3 Points Using Weak-Perspective, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 8, pp. 802-808, Aug. 1994.

13. D. DeMenthon and L. S. Davis, Exact and Approximate Solutions of the Perspective-Three- Point Problem, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 14, no. 11, pp. 1100-1105, Nov. 1992.

14. J. G. Juang, Parameter Estimation in the Three-Point Perspective Projection Problem in Computer Vision, Proceedings of the IEEE International Symposium on Industrial Electronics, vol. 3, pp. 1065-1070, Jul. 1997.

15. T. Moriya and H. Takeda, Solving the rotation-estimation problem by using the perspective three-point algorithm, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol.1, pp. 766 – 773, Jun. 2000.

16. X. S. Gao and H. F. Chen, New Algorithms for the Perspective-Three-Point Problem, Journal of Computer Science and Technology, vol. 16, no. 3, pp. 194-207, May, 2001.

17. D. Nister, A Minimal Solution to the Generalised 3-Point Pose Problem, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol.1, pp. 560 - 567, Jun. 2004.

18. D. Faugeras and G. Toscani, The Calibration Problem for Stereo, In Proceedings of IEEE Computer

Society Conference on Computer Vision and Pattern Recognition, pp.15-20, Los Alamitos, CA, Jun. 1986.

Appendix

${}^{c}p_{rz}$ is non-zero. Dividing equation (5) by ${}^{c}p_{rz}$, then

$$\begin{cases} x_{ri}\,{}^{c}n'_{rx} + y_{ri}\,{}^{c}o'_{rx} - x_{1ci}x_{ri}\,{}^{c}n'_{rz} - x_{1ci}y_{ri}\,{}^{c}o'_{rz} + {}^{c}p'_{rx} = x_{1ci} \\ x_{ri}\,{}^{c}n'_{ry} + y_{ri}\,{}^{c}o'_{ry} - y_{1ci}x_{ri}\,{}^{c}n'_{rz} - y_{1ci}y_{ri}\,{}^{c}o'_{rz} + {}^{c}p'_{ry} = y_{1ci} \end{cases} \tag{A1}$$

here ${}^{c}n'_{r} = {}^{c}n_{r} / {}^{c}p_{rz}$, ${}^{c}o'_{r} = {}^{c}o_{r} / {}^{c}p_{rz}$, ${}^{c}p'_{rx} = {}^{c}p_{rx} / {}^{c}p_{rz}$, ${}^{c}p'_{ry} = {}^{c}p_{ry} / {}^{c}p_{rz}$.

For four points known on a plane, equation (A1) is rewritten as

$$CH' = D \tag{A2}$$

where C is a $2n\times8$ matrix, D is a $2n\times1$ matrix,

$$C = \begin{bmatrix} x_{r1} & 0 & -x_{1c1}x_{r1} & y_{r1} & 0 & -x_{1c1}y_{r1} & 1 & 0 \\ 0 & x_{r1} & -y_{1c1}x_{r1} & 0 & y_{r1} & -y_{1c1}y_{r1} & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{rn} & 0 & -x_{1cn}x_{rn} & y_{rn} & 0 & -x_{1cn}y_{rn} & 1 & 0 \\ 0 & x_{rn} & -y_{1cn}x_{rn} & 0 & y_{rn} & -y_{1cn}y_{rn} & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} x_{1c1} \\ y_{1c1} \\ \vdots \\ x_{1cn} \\ y_{1cn} \end{bmatrix}$$

Without loss of generality, assume that the coordinates of four known points are P_{r1}=[0, 0, 0], P_{r2}=[x_{r2}, 0, 0], P_{r3}=[x_{r3}, y_{r3}, 0] and P_{r4}=[x_{r4}, y_{r4}, 0] in the reference frame. Submitting P_{r1} to P_{r4} to matrix C, then the determinant is computed as

$$\det(C) = \begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ x_{r2} & 0 & -x_{1c2}x_{r2} & 0 & 0 & 0 & 1 & 0 \\ 0 & x_{r2} & -y_{1c2}x_{r2} & 0 & 0 & 0 & 0 & 1 \\ x_{r3} & 0 & -x_{1c3}x_{r3} & y_{r3} & 0 & -x_{1c3}y_{r3} & 1 & 0 \\ 0 & x_{r3} & -y_{1c3}x_{r3} & 0 & y_{r3} & -y_{1c3}y_{r3} & 0 & 1 \\ x_{r4} & 0 & -x_{1c4}x_{r4} & y_{r4} & 0 & -x_{1c4}y_{r4} & 1 & 0 \\ 0 & x_{r4} & -y_{1c4}x_{r4} & 0 & y_{r4} & -y_{1c4}y_{r4} & 0 & 1 \end{vmatrix}$$

$$= x_{r2}^2 y_{r3} y_{r4} (x_{r4} y_{r3} - x_{r3} y_{r4})[(x_{1c2} - x_{1c4})(y_{1c3} - y_{1c4}) - (x_{1c3} - x_{1c4})(y_{1c2} - y_{1c4})] \quad \text{(A3)}$$

The value of det(C) is the product of five terms such as x_{r2}^2, y_{r3}, y_{r4}, $(x_{r4}y_{r3}-x_{r3}y_{r4})$ and $[(x_{1c2}-x_{1c4})(y_{1c3}-y_{1c4})- (x_{1c3}- x_{1c4})(y_{1c2}-y_{1c4})]$. $x_{r2}\neq 0$ since P_{r1} and P_{r2} are different points. For the second term, $y_{r3}=0$ means that P_{r1}, P_{r2} and P_{r3} are collinear. For the third term, $y_{r4}=0$ means that P_{r1}, P_{r2} and P_{r4} are collinear. The fourth term $(x_{r4}y_{r3}-x_{r3}y_{r4})=0$ indicates that P_{r1}, P_{r3} and P_{r4} are collinear. The fifth term $[(x_{1c2}-x_{1c4})(y_{1c3}-y_{1c4})- (x_{1c3}- x_{1c4})(y_{1c2}-y_{1c4})]=0$ only if P_{r2}, P_{r3} and P_{r4} are collinear.

It can be found from (A3) that det(C) is nonzero if any three points are not linearly correlated. In this case, rank(C)=8. There are eight independent equations for eight linear independent variables in (7). Hence, H_1 and H_2 can be solved.

List of Figure Captions

1. Fig.1 The world frame and the camera frame
2. Fig.2 The error curve in RLS
3. Fig.3 The solution error distribution with two solutions
4. Fig.4 The solution errors for the P3P problem with noise
5. Fig.5 The space points owning same image coordinates in two views and their image coordinates
6. Fig.6 The points in a section and their image coordinates in the two views
7. Fig.7 Experiment with the proposed pattern

List of Table Captions

1. Table 1 The visual measuring results based on the positioning method with the proposed pattern