Robotica (2002) volume 20, pp. 385–393. © 2002 Cambridge University Press
DOI: 10.1017/S0263574702004186 Printed in the United Kingdom

Adaptation of Lowe's camera pose recovery algorithm to mobile robot self-localisation

Omar Ait-Aider, Philippe Hoppenot and Etienne Colle

CEMIF – Complex Systems Group, University of Evry, 40 rue du Pelvoux, 91020 Evry Cedex (France)
E-mail: oaider@cemif.uniy-evry.fr; hoppenot@cemif.uniy-evry.fr; ecolle@cemif.uniy-evry.fr

(Received in Final Form: February 16, 2002)

SUMMARY
This paper presents an adaptation of Lowe's numerical model-based camera localisation algorithm to the domain of indoor mobile robotics. While the original method is straightforward and even elegant, it nonetheless exhibits certain weaknesses. First, due to an affine approximation, the method is not consistent with perspective projection especially when the dimensions of objects seen are large in comparison with their distances to the camera. Next, the non-linearity of equations makes convergence properties sensitive both to the initial solution estimate and to noise. By taking the specificity and exigency of the mobile robotics domain into account, a new formulation of this method is proposed in order to improve efficiency, accuracy and robustness in the presence of noisy data and variable initial conditions. According to this formulation, line correspondences are used rather than points, the number of degrees of freedom is reduced, the affine approximation is removed and rotation is uncoupled from translation. Test results with both synthetic and real images illustrate the improvements expected from theoretical modifications.

KEYWORDS: Lowe's camera; Pose recovery; Self-localisation; Mobile robotics.

1. INTRODUCTION

The problem of camera localisation relative to real-world objects using a single view arises in several types of applications, such as object recognition, hand-eye co-ordination and visual navigation. A wide range of methods for this type of camera pose recovery has been studied in the literature. They can all be grouped and designated under the term "model-based localisation" by virtue of sharing the basic principle of using *a priori* knowledge (a model) of the geometry of objects in the viewed scene. The location of a real-world feature on the image is constrained by projection rules and camera characteristics (intrinsic parameters) on one side and by the location of this feature relative to the camera (extrinsic parameters) on the other. In respecting the mathematical formalism of this assumption, a correspondence between a geometric feature of the 3D real world and its 2D projection on the image can be expressed in the form of an equation whose unknowns are the extrinsic parameters that contain the desired camera location. The problem then consists of establishing a sufficient number of 3D-2D correspondences to recover all of the parameters.

These methods however differ from one another in many aspects: the internal camera model, the kinds of features used for correspondences, the mathematical formalism used to express the location (position and orientation) of 3D objects, the computational technique, the number of unknowns, etc.

The majority of these methods use points or lines as features for the 2D-3D correspondences and are based on matching a model feature to its presumed projection obtained from image measurements. Each match yields an equation whose unknowns are functions of the translation vector and the rotational matrix between a real world-related frame and a camera-related frame.

Two main groups can be distinguished herein: analytical methods and numerical methods. An analytical solution consists of resolving a set of non-linear equations. The number of equations must be the same as the number of unknowns (generally 6 in all: 3 for translation and 3 for rotation). If additional equations are available, they are used to remove ambiguity due to a possible multiplicity of solutions. One of the first analytical methods was presented by Fischler and Bolles,[1] who recovered the camera location by computing the distances between the optical centre and three points of the modelled rigid object. The set of equations is then obtained using the distances between each couple of modelled points and the angles between the lines of sight for each image point. The system is thereby transformed into an eight-degree polynomial equation. The authors established that for three correspondences, up to eight solutions may be found. Dhome[2] gives another analytical method based on line correspondences. He first decomposed the global transformation between world frame and camera frame into two transformations by introducing an additional frame whose *xy* plane is the interpretation plane of one of the line segments. The first transformation is thus completely independent of the modelled object, and the author then computed two unknown angles. As with the previous method, the system gets transformed into an eight-degree polynomial equation. One problem encountered with these methods is the presence of multiple solutions. Quan[3] presents a linear method to identify a unique solution using four or five-point correspondences. Another problem is the presence of noise in image measurements within all practical applications. This noise exerts an effect on the accuracy of the recovered location. Dhome notes that his method must not be contrasted with a numerical method, but rather is to be used for the purpose of initialisation since

https://doi.org/10.1017/S0263574702004186 Published online by Cambridge University Press

it yields all of the model space attitudes compatible with the interpretation of the three lines.

In numerical methods, an error function expresses the distances between each image feature and the projection of the corresponding feature in the real world using the current camera location. The transformation is then corrected iteratively starting from the initial estimate of the location in a minimisation process, such as the least-squares method. This approach is better adapted to problems in which measurements are noisy and especially when the system of equations is over-determined (i.e. the number of correspondences is greater than the number of unknowns), yet convergence is not always guaranteed. Due to the non-linearity of perspective projection equations and the expressions of location as a function of extrinsic parameters (particularly for rotation), convergence depends on both the minimisation method chosen and the quality of the initial location estimates.

One numerical method has been presented by Lowe.[4–6] In expressing the error function, the distance between the projection of each point of the model and the corresponding point seen on the image is written as a function of the location parameters. To ensure the efficiency of the algorithm, the translation is expressed within the camera frame. In order to use Newton's optimisation method, Lowe needed to express the partial derivatives of the error function in each location parameter. To achieve this step, the translation is expressed within the camera frame and an affine approximation is derived of the distance from each point to the optical centre, by considering that this distance is the same for all points. In addition, the correction of each of the three rotational parameters at each iteration is considered small enough to assume that the three basic rotations are independent. These approximations provide an elegant linear system of equations, yet give rise to many convergence problems. Araujo and Carceroni[7] removed the affine approximation on the third component of the translation vector and showed, by means of experimental evaluation, that convergence performance is improved. The approach proposed by Liu[8] uses line correspondences. The error function expression is obtained by the scalar product of the director vector of each modelled segment and the vector normal to the interpretation plane of the corresponding line on the image. The rotation and translation are uncoupled. Two solutions are forwarded therein: a linear solution that requires more than nine correspondences, and a non-linear solution using at least three correspondences. Unfortunately, the latter solution is ineffectual when the unknown angles are larger than 30°. Phong and Horaud[9] improved this method by introducing both the unit quaternions to express rotation and a minimisation algorithm featuring better global convergence characteristics. Other authors have used simpler camera models, such as weak perspective, para-perspective and affine camera models,[10–12] with almost all being based on point correspondences. These models may be applied when the dimensions of the viewed objects are small in comparison with the distance from the optical centre.

This paper focuses on the application of camera localisation techniques in the domain of mobile robot self-localisation. After analysing the specificity and exigency of this domain, a new formulation of Lowe's algorithm is developed. The aim of this work is to obtain an algorithm for camera pose recovery offering improved performance (in terms of efficiency, robustness and accuracy). Within this formulation, the affine approximation of the original formulation is removed in order to better incorporate full-perspective effects. The number of degrees of freedom of the system is reduced depending on the specific domain of the mobile robotics context. The translation and rotation recovery is uncoupled. In Section 2, the original formulation is presented. Section 3 discusses modifications and improvements to the algorithm in order to adapt it to the context of mobile robot localisation. Simulations and experimental results are then displayed in Section 4 to illustrate the performance of the derived method in comparison with that of the original method.

2. LOWE'S ALGORITHM FORMULATION

Let's consider a co-ordinate system related to both a geometrically-modelled 3D environment and a set of points $\mathbf{P_i}(X_i, Y_i, Z_i)$ of a rigid object expressed in this frame. Let's include a second co-ordinate system related to the camera such that its origin O_c is the optical centre and the z-axis is normal to the image plane located a distance f (the focal lens) from O_c (see Figure 1).

In considering a pinhole model, the intrinsic camera parameters α_u, α_v, u_0 and v_0 can be obtained by means of calibration.[13–15]

If we assume that the transformation between the two frames is a rotation $\mathbf{R}$ and a translation $\mathbf{T}$, we can compute the estimated projection u_i, v_i of $\mathbf{P_i}$ on the image as follows:

$$(x_i, y_i, z_i) = \mathbf{R} \,.\, (\mathbf{P_i} + \mathbf{T_i}) \tag{1}$$

$$(u_i, v_i) = \left(\alpha_u \,.\, \frac{x_i}{z_i} + u_0,\ \alpha_v \,.\, \frac{y_i}{z_i} + v_0 \right) \tag{2}$$

In Equation (1), (x_i, y_i, z_i) specify the coordinates of $\mathbf{P_i}$ in the camera frame.

Let's now consider the set of corresponding pixels (u_{mi}, v_{mi}) measured from the image. The pose recovery problem

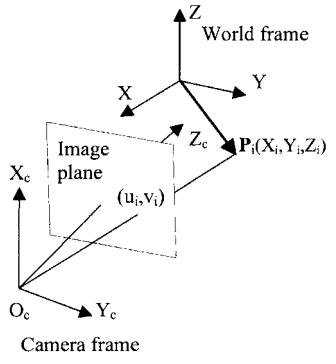


Fig. 1. Perspective projection.

https://doi.org/10.1017/S0263574702004186 Published online by Cambridge University Press

then consists of computing the optimal translation vector and rotational matrix that minimise the errors $e_{ui} = u_i - u_{mi}$ and $e_{vi} = v_i - v_{mi}$. With an initial estimate of $\mathbf{R_0}$ and $\mathbf{T_0}$ and assuming the local linearity of u_i, v_i as functions of the location parameters, we can apply Newton's method to iteratively reach the optimal **R** and **T** by computing at each step the correction for each location parameter. We must first express the Jacobian matrix of the partial derivatives of u_i and v_i with respect to these parameters. In order to achieve greater efficiency, Lowe reshaped the translation parameters as follows:

$$(x_i, y_i, z_i) = \mathbf{R} \,.\, \mathbf{P_i} \tag{3}$$

$$(u_i, v_i) = \left(\alpha_u \cdot \frac{x_i}{z_i + D_z} + u_0 + D_x, \alpha_v \cdot \frac{y_i}{z_i + D_z} + v_0 + D_y \right) \tag{4}$$

In Equation (3), (x_i, y_i, z_i) are the co-ordinates of $\mathbf{P_i}$ obtained by applying the rotation $\mathbf{R}^{-1}$ on the world frame. D_x and D_y specify the object location in the image plane and D_z the distance between the object and the camera's optical centre.

Newton's method does not require an explicit representation of individual rotational parameters, but merely a way both to modify the original orientation in mutually-orthogonal directions Ψ, θ and ϕ about the *x*, *y* and *z*-axes of the camera co-ordinate system and to calculate the partial derivatives of u_i and v_i with respect to rotational parameters Ψ, θ and ϕ. On this basis, Lowe elected to maintain the initial specification of **R** (i.e. a 3×3 matrix) and to combine it with an incremental rotation composed of the three basic rotational corrections. He obtained a simple form for the partial derivatives presented in Table I. Note that with this specification of rotation, the correction matrix is to be evaluated at each iteration once the correction of parameters has been calculated, thus reducing the algorithm efficiency.

Each model point matching an image point yields two linear equations of the following form:

$$\frac{\partial u_i}{\partial D_x} \Delta D_x + \frac{\partial u_i}{\partial D_y} \Delta D_y + \frac{\partial u_i}{\partial D_z} \Delta D_z + \frac{\partial u_i}{\partial \psi} \Delta \psi + \frac{\partial u_i}{\partial \theta} \Delta \theta + \frac{\partial u_i}{\partial \phi} \Delta \phi = E u_i \tag{5}$$

Table I. Partial derivatives of u_i and v_i with respect to location parameters where: $c = 1/z_i + D_z$.

	u_i	v_i
D_x	1	0
D_y	0	1
D_z	$-\alpha_u c^2 x_i$	$-\alpha_v c^2 y_i$
Ψ	$-\alpha_u c^2 x_i y_i$	$-\alpha_v c(z_i + y_i^2)$
θ	$-\alpha_u c(z_i + x_i^2)$	$\alpha_v c^2 x_i$
ϕ	$c x_i$	$C y_i$

$$\frac{\partial v_i}{\partial D_x} \Delta D_x + \frac{\partial v_i}{\partial D_y} \Delta D_y + \frac{\partial v_i}{\partial D_z} \Delta D_z + \frac{\partial v_i}{\partial \psi} \Delta \psi + \frac{\partial v_i}{\partial \theta} \Delta \theta + \frac{\partial v_i}{\partial \phi} \Delta \phi = E v_i \tag{6}$$

With at least three correspondences, the six location parameters can be recovered.

The problem with the previous formulation is that D_x and D_y are assumed to be approximately constant for all points of the viewed object, when in fact they depend on the distance of each point to the optical centre. Such an assumption is not consistent with perspective projection, especially if the dimensions of the object are not sufficiently small in comparison with its distance from the camera. Araujo and Carceroni[7] remove this affine approximation and propose a fully-projective formulation of Lowe's algorithm; they have rewritten Equations (1) and (2) as follows:

$$(u_i, v_i) = \left(\alpha_u \cdot \frac{x_i + Dx}{z_i + Dz} + u_0, \alpha_v \cdot \frac{y_i + Dy}{z_i + Dz} + v_0 \right) \tag{7}$$

The revised partial derivatives are presented in Table II.

3. ADAPTATION OF LOWE'S ALGORITHM TO A MOBILE ROBOTICS CONTEXT

In the application herein, a mobile robot moves in a partially-modelled 3D indoor environment, such as a flat. The model includes walls, the floor, the ceiling, windows, doors and some heavy pieces of furniture. The camera is mounted onto the robot's mobile base and takes perspective views from its current location. A set of features are extracted from the image and matched with those of the model. The pose recovery process is then applied.

To adapt Lowe's technique to the domain of mobile robot self-localisation, the specificities of this particular kind of application must be studied. Certain simplifications, such as reducing the number of degrees of freedom, can lead to decreasing the complexity of the system of equations and its non-linearity.[16–18] The context unfortunately generates greater exigency in terms of efficiency (real-time application), robustness (noisy data, ambiguities due to multiple solutions) and accuracy. In the following sub-section, modifications carried out on the original method are presented along with the set of factors influencing each choice.

Table II. Partial derivatives of u_i and v_i with respect to location parameters where: $[a\ b\ c] = [x_i + D_x\ y_i + D_y\ 1/z_i + D_z]$.

	u_i	v_i
D_x	$\alpha_u c$	0
D_y	0	$\alpha_v c$
D_z	$-\alpha_u a c^2$	$-\alpha_v b c^2$
Ψ	$-\alpha_u a c^2 y_i$	$-\alpha_v c(z_i + b c y_i)$
θ	$\alpha_u c(z_i + a c x_i)$	$\alpha_v b c^2 x_i$
ϕ	$-\alpha_u c y_i$	$\alpha_v c x_i$

https://doi.org/10.1017/S0263574702004186 Published online by Cambridge University Press

3.1. Quality of the initial estimation

The performance of non-linear function optimisation using efficient algorithms has been correlated with the quality of the initial solution estimation. The camera-localisation technique belongs to the family of absolute localisation methods, which in general are combined with dead-reckoning techniques to provide an estimate of the current robot location. This information however may be not available in all instances, e.g. when resetting the system, or may prove to be false due to the limitations of dead-reckoning. This observation means that the quality of the initial estimate is not always guaranteed in this kind of real application. A way to reduce the non-linearity of camera pose recovery equations must then be found. Due to the dimensions of model objects used in mobile robotics (e.g. wall junctions, doors, large furniture) in comparison with their distances from the camera, the use of a full-perspective model of the camera is unfortunately nearly mandatory. Moreover, this model is definitely non-linear. The approximation used in the original formulation of Lowe's algorithm must be removed because it does not adequately incorporate the perspective effects.

3.2. The number of degrees of freedom

As seen below, camera location is characterised by a translation vector $\mathbf{T}=[t_x\ t_y\ t_z]$ and a rotational matrix $\mathbf{R}$. $\mathbf{T}$ that represents the translation between the camera optical centre and the origin of the world frame. The rotation $\mathbf{R}$ is composed of the three Euler angles Ψ, θ and ϕ about the *x*, *y* and *z*-axes of the camera's co-ordinate system. $\mathbf{R}$ and $\mathbf{T}$ carry the camera frame onto the world frame. In general, indoor mobile robots operate in a 3D environment, yet their displacements are in a 2D horizontal space at a known and constant height from the ground. t_z and θ are thus assumed to be known and Ψ is zero. The system then becomes a 3DOF (degrees of freedom) system with 3 parameters (ϕ, t_x and t_y).

3.3. Making use of line correspondences

As discussed above, the majority of camera localisation methods are based on point or line correspondences. In this application, the image is first segmented into contours. Contours generally correspond to physical elements in the work space, such as edges constituted by intersections between surfaces of the flat. These edges tend to be straight segments. Lines are easier to extract from contour images and their characterisation by polygonal approximation is reliable even in the presence of noise. Partial occlusion (due to the view angle or the presence of non-modelled objects) does not affect line representation parameters. Furthermore, the extremities of the edges that could possibly be considered as point features are not always seen on the image due to the dimension of the flat edges in comparison with their distance to the camera. These reasons make it more prudent to use straight line correspondences. Thus, the 3D model can simply comprise a set of straight segments whose extremities have known co-ordinates in the world frame.

Formulation of the method. In light of the previous assumptions, a formulation of Lowe's algorithm using straight line correspondences is presented in this section. The translation is first expressed in the world frame rather than the camera frame, in which case the number of unknowns can be reduced (as shown above). The rotational matrix between camera frame and world frame is calculated as follows:

$$\mathbf{R}=\begin{pmatrix}\cos(\theta) & 0 & \sin(\theta)\\ 0 & 1 & 0\\ -\sin(\theta) & 0 & \cos(\theta)\end{pmatrix}.\begin{pmatrix}\cos(\phi) & -\sin(\phi) & 0\\ \sin(\phi) & \cos(\phi) & 0\\ 0 & 0 & 1\end{pmatrix}$$

$$=\begin{pmatrix}\cos(\theta)\cos(\phi) & -\cos(\theta)\sin(\phi) & \sin(\theta)\\ \sin(\phi) & \cos(\phi) & 0\\ -\sin(\theta)\cos(\phi) & -\sin(\theta)\sin(\phi) & \cos(\theta)\end{pmatrix}$$

The projection Equation (2) can then be rewritten as:

$$\begin{cases} u_i=\alpha_u.\dfrac{(X_i-t_x).\cos(\theta).\cos(\phi)-(Y_i-t_y).\cos(\theta).\sin(\phi)+(Z_i-t_z).\sin(\theta)}{-(X_i-t_x).\sin(\theta).\cos(\phi)+(Y_i-t_y).\sin(\theta).\sin(\phi)+(Z_i-t_z).\cos(\theta)}+u_0 \\ v_i=\alpha_v.\dfrac{(X_i-t_x).\sin(\phi)+(Y_i-t_y).\cos(\phi)}{-(X_i-t_x).\sin(\theta).\cos(\phi)+(Y_i-t_y).\sin(\theta).\sin(\phi)+(Z_i-t_z).\cos(\theta)}+v_0 \end{cases} \quad (8)$$

A straight line on the image plane is characterised by a slope and a perpendicular distance to the origin (see Figure 2). Let's consider a straight line of parameters ρ_{mi} and d_{mi} extracted from a segmented image. By projecting two arbitrary points $\mathbf{P_{1i}}(X_{1i}, Y_{1i}, Z_{1i})$ and $\mathbf{P_{2i}}(X_{2i}, Y_{2i}, Z_{2i})$ of the corresponding model line using an initial estimate of t_x, t_y and ϕ, we obtain two pixels $\mathbf{p_{1i}}(u_{1i}, v_{1i})$ and $\mathbf{p_{2i}}(u_{2i}, v_{2i})$ that form a straight line at a certain distance and with a certain slope differential with respect to the measured line. The respective distances from $\mathbf{p_{1i}}$ and $\mathbf{p_{2i}}$ to the image line are:

$$d_{1i}=\cos(\rho_{mi}).u_{1i}+\sin(\rho_{mi}).v_{1i}-d_{mi}$$

$$d_{2i}=\cos(\rho_{mi}).u_{2i}+\sin(\rho_{mi}).v_{2i}-d_{mi}$$

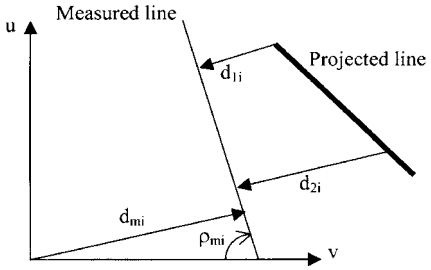


Fig. 2. Expression of the error function using line correspondences.

https://doi.org/10.1017/S0263574702004186 Published online by Cambridge University Press

Expressing d_{1i} and d_{2i} for each line correspondence with respect to the location parameters, an error function is obtained whose minimisation yields the sought optimal values t_x, t_y and ϕ. Replacing u_{1i}, u_{2i}, v_{1i} and v_{2i} by the expression in (9), the error function can be written as:

$$\left\{\begin{array}{l} \dfrac{\begin{array}{r}[\lambda_{1i}(X_{1i}-t_x)+\lambda_{2i}(Y_{1i}-t_y)].\cos(\phi)+[\lambda_{2i}(X_{1i}-t_x)\\ -\lambda_{1i}(Y_{1i}-t_y)].\sin(\phi)+[\lambda_{3i}(Z_{1i}-t_z)]\end{array}}{\begin{array}{r}-(X_i-t_x).\sin(\theta).\cos(\phi)+(Y_i-t_y).\sin(\theta).\sin(\phi)\\ +(Z_i-t_z).\cos(\theta)\end{array}} \\ \quad +d_{0i}-d_{mi}=d_{1i} \\ \dfrac{(X_i-t_x).\sin(\phi)+(Y_i-t_y).\cos(\phi)}{\begin{array}{r}-(X_i-t_x).\sin(\theta).\cos(\phi)+(Y_i-t_y).\sin(\theta).\sin(\phi)\\ +(Z_i-t_z).\cos(\theta)\end{array}} \\ \quad +d_{0i}-d_{mi}=d_{2i} \end{array}\right. \tag{9}$$

where:

$$\lambda_{1i}=\alpha_u.\cos(\theta).\cos(\rho_{mi}),\ \lambda_{2i}=\alpha_v.\sin(\rho_{mi}),$$
$$\lambda_3=\alpha_u.\sin(\theta).\cos(\rho_{mi}),\ d_{0i}=u_0.\cos(\rho_{mi})+v_0.\sin(\rho_{mi})$$

Uncoupling the translation and rotation. The system of Equations (9) is non-linear and contains multiple unknowns. Convergence properties are highly dependent upon the quality of the initial estimate of the solution vector. Many situations unfortunately arise in which the robot is "completely lost" in its environment and has no perception of its actual location. An approach to reducing the effects of nonlinearity is to find a way to uncouple some of the variables. One solution then would consist of first seeking to decrease the angles between each viewed and projected line pair and afterwards reducing the resulting perpendicular distances. The first step is achieved by minimising the difference between d_{1i} and d_{2i}. This is possible by restoring Lowe's approximation that considers that the denominator of u_i and v_i is constant.[4] The error function to minimize is obtained from subtracting the two Equations in (9):

$$\sigma_{1i}.\cos(\phi)+\sigma_{2i}.\sin(\phi)+\sigma_{3i}=d_{1i}-d_{2i} \tag{10}$$

where:

$$\sigma_{1i}=(A_i-C_i).(X_{1i}-X_{2i})+B_i.(Y_{1i}-Y_{2i})$$
$$\sigma_{2i}=B_i.(X_{li}-X_{2i})-(A_i-C_i).(Y_{1i}-Y_{2i})$$
$$\sigma_{3i}=D_i.(Z_{1i}-Z_{2i})$$

with:

$$A_i=\alpha_u.\mathrm{tg}(\theta).\cos(\rho_{mi})$$

$$B_i=\frac{\alpha_v.\cos(\rho_{mi})}{\sin(\theta)}$$

$$C_i=u_0.\cos(\theta)+v_0.\sin(\theta)-d_{mi}$$
$$D_i=A_i.\mathrm{atg}(\theta)+C_i.\mathrm{tg}(\theta)$$

The rotation and translation parameters are now uncoupled. An initial estimate of the solution can be found by analytically solving one of these equations. A numerical optimisation by means of least squares using Newton's method is then to be applied.

Given the optimal angle ϕ, translation recovery becomes a linear problem, i.e.:

$$\left\{\begin{array}{l}\xi_{1i}.t_x+\xi_{2i}.t_y+\xi_{3i}=d_{1i}\\ \xi_{1i}.t_x+\xi_{2i}.t_y+\xi_{4i}=d_{2i}\end{array}\right. \tag{11}$$

where

$$\xi_{1i}=(A_i-C_i).\cos(\phi)-B_i.\sin(\phi)$$
$$\xi_{2i}=(A_i-C_i).\sin(\phi)-B_i.\cos(\phi)$$
$$\xi_{3i}=((A_i-C_i).X_{1i}+B_i.Y_{1i}).\cos(\phi)$$
$$+(B_i.X_{1i}-(A_i-C_i).Y_{1i})\sin(\phi)$$
$$+D_i(Z_{1i}-t_z)+d_{0i}-d_{mi}$$
$$\xi_{4i}=((A_i-C_i).X_{2i}+B_i.Y_{2i}).\cos(\phi)$$
$$+(B_i.X_{2i}-(A_i-C_i).Y_{2i})\sin(\phi)$$
$$+D_i(Z_{2i}-t_z)+d_{0i}-d_{mi}$$

In the new formulation, the error function is expressed directly with respect to location parameters. The three basic component orientations are not considered as independent of one another and their mutual orthogonality is taken into account in the equations. This means that upon each function evaluation during an iterative minimisation process, the correction vector must simply be added to the actual parameter vector that reduces computational cost in comparison with the original formulation.

Another advantage of this new formulation is the removal of singularity for the one-point junction case. Lowe's method, as in the straight line-based methods of Liu,[8] Phong[9] or Dhome,[2] diverges when all lines used for correspondences intersect at the same point. In this case, the equations obtained are redundant; moreover, the system is not over-determined and hence not well-suited for numerical optimisation.

4. EVALUATION OF THE METHOD

The presented 3DOF method has been tested on both synthetic and real images in order to evaluate its performance in comparison with the fully-projective version of Lowe's algorithm developed by Araujo and Carceroni.[7] According to the authors, this method displays better performance than the original one. A new version using line correspondences of Araujo's algorithm was first developed and then the 3DOF method was compared to this version.

Test conditions correspond to those for indoor mobile robotics applications in terms of variations in pose parameters, noise on the data and the quality of initialisation.

Testing with synthetic images. Synthetic data allow for a statistical study with a large number of useful situations to help estimate location accuracy, convergence properties and the robustness of each method.

A model of a flat room has been built with a set of straight segments (see Figure 3). The camera simulator is then posed

https://doi.org/10.1017/S0263574702004186 Published online by Cambridge University Press

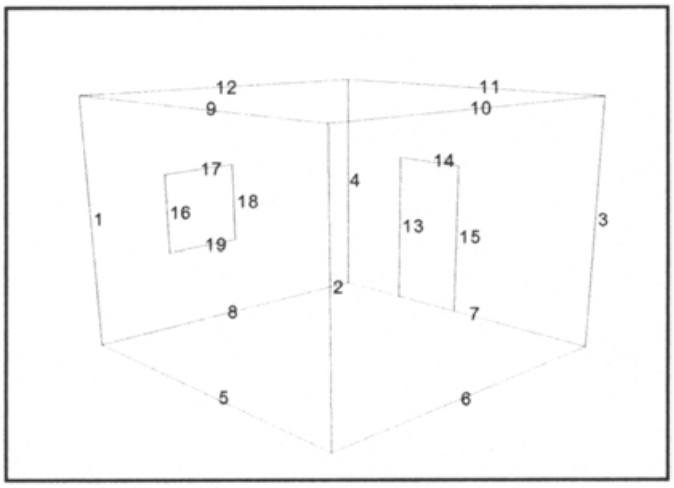


Fig. 3. An example of a 3D work space model.

with various reference orientations and positions, and an image of the model is simulated from each location. A set of 3D segments and synthetic line correspondences is obtained for each image. Both the 3DOF method and Araujo's method are then applied and the error between computed and reference poses gets calculated.

The reference locations were randomly generated by discarding those where the number of visible segments was less than 3. Locations are represented by a vector of the six position and orientation parameters $[t_x, t_y, t_z, \Psi, \theta, \phi]$, uniformly-distributed over the following intervals or values:

$$[0\text{m}, 4\text{m}] \text{ for } t_x \text{ and } t_y,$$

$$[0.5\text{m}, 1.5\text{m}] \text{ for } t_z,$$

$$[-15°, +15°] \text{ for } \theta,$$

$$[-180°, +180°] \text{ for } \phi,$$

$$\Psi = 0°.$$

1,500 different situations were ultimately selected, with the number of correspondences varying from 3 to 10.

To model imperfections of the intrinsic camera parameter set, the image segmentation and the polygonal line approximation errors, noise was added to image line parameters ρ_e and d_e. Measured parameters ρ_m and d_m have been obtained as follows:

$$\rho_m = \rho_e + 2 \,.\, \text{nl} \,.\, \delta\rho,$$

$$d_m = d_e + 5 \,.\, \text{nl} \,.\, \delta d$$

where the values of $\delta\rho$ and δd are uniformly distributed over the interval $[-1, +1]$ and nl defines the noise level. For each location, the two algorithms were executed with nl varying from 0 to 1, thereby introducing perturbations on parameters that vary from 0° to 2° for ρ_m and from 0 to 10 pixels for d_m.

In order to study the effects of initial estimation quality, algorithms were tested with randomly-generated vectors of initial location parameters. The difference between the

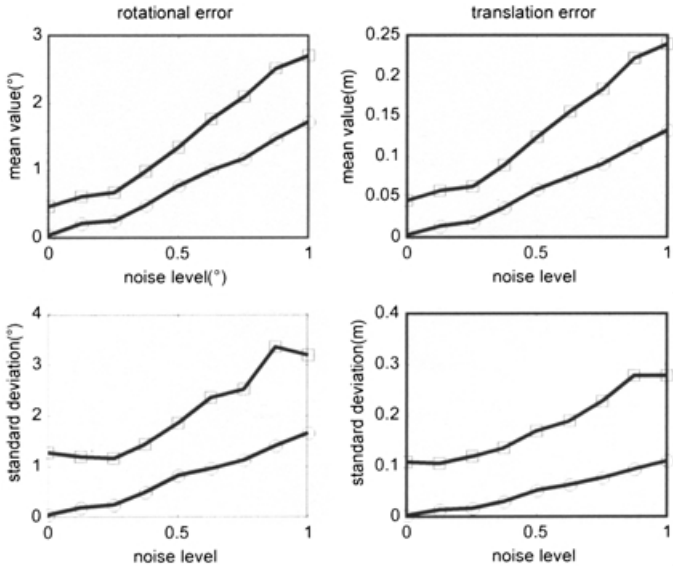


Fig. 4. Localisation accuracy, □: Araujo's method, ○: 3DOF method.

component of the initial parameter vector and the reference angle varied between 5° and 90°.

In all, more than 36,000 test runs were performed for each method.

Results presented in the following section indicate the rotational error ϵ_ϕ and translation error ϵ_T, as defined by:

$$\epsilon_\phi = |\phi_{\text{reference}} - \phi_{\text{computed}}|, \; \epsilon_T = \|T_{\text{reference}} - T_{\text{computed}}\|.$$

Analysis of results. The results shown below are obtained after eliminating the 1% of extreme cases. This step has been dictated by the fact that a method which, in the general case, shows good convergence properties can suddenly diverge for some singular cases and influence the analysis of results.

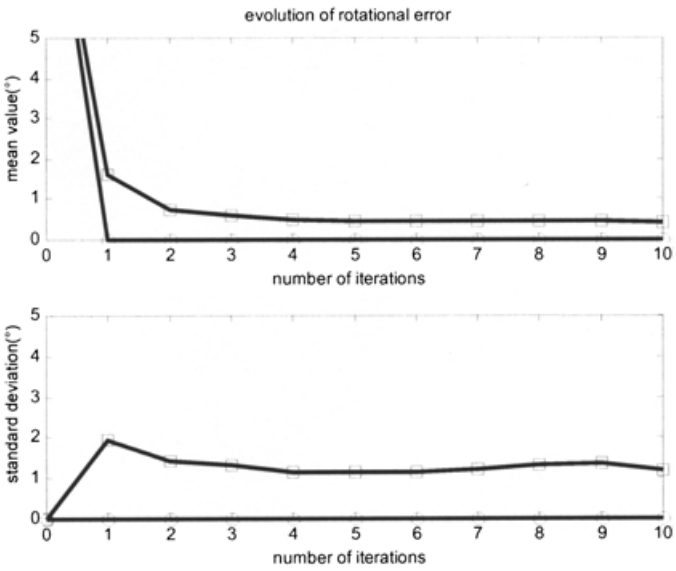


Fig. 5. Localisation error evolution, □: Araujo's method, ○: 3DOF method.

https://doi.org/10.1017/S0263574702004186 Published online by Cambridge University Press

4.1. Accuracy

The results in Figure 4 give mean values and standard deviations of the rotational and translation errors obtained with different noise levels. In general, they reveal that the 3DOF algorithm converges toward a better approximation of the actual location. The lower level of the standard deviation of these results indicates that this method is more reliable in the presence of noise. Practically speaking, the accuracy at a reasonable noise level satisfies the exigencies of mobile robotics, which are typically about 2° for rotation and 10 cm for translation.

4.2. Convergence speed

In Figure 5, the evolution of rotational error with respect to the number of iterations is analysed. The noise level was set at nl=0 and the initial orientation at an angle of 10° from the solution. For this simulation, it clearly appears that the 3DOF method provides a faster decrease in error and requires fewer iterations to reach an acceptable error level. This observation implies that if a compromise between real-time performance and accuracy were to be found by limiting the number of iterations in real applications, the 3DOF method would be of greater use.

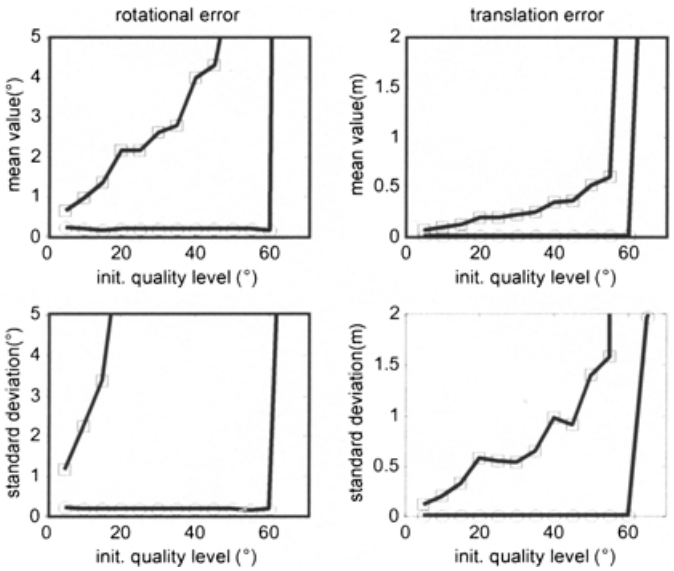


Fig. 6. Sensitivity to the quality of initialisation, □: Araujo's method, ○: 3DOF method.

4.3. Sensitivity to the quality of initial solution estimation

Figure 6 presents the influence of the quality of initialisation on each algorithm. Results obtained with nl=0 denote the rotational error mean value and standard deviation with respect to a rotational initialisation quality defined by: $\Delta\phi = |T_{reference} - \phi_{initialisation}|$. It appears that while the performance of Araujo's method progressively deteriorates, the 3DOF method remains reliable until a stable threshold around 60° has been reached. This feature makes the 3DOF method better adapted for real mobile robot localisation applications in which the initialisation error cannot be accurately defined, but may simply be overestimated. An error margin of 120° (–60° to +60° around the reference) can be considered as comfortable in a mobile robotics context.

4.4. Sensitivity to the number of line correspondences

Another interesting result is the evolution in localisation accuracy with respect to the number of available line correspondences. In practical situations, the number of extracted and correctly-matched lines varies with the visibility angle, light conditions, etc. Results in Figure 7 show the mean value and standard deviation of rotational and translation errors as a function of the number of correspondences for two distinct noise levels of nl=0 and nl=0.5. The accuracy of the 3DOF method is practically constant even as the number of correspondences decreases to the requested minimum. In contrast, Araujo's method is very sensitive to this number and almost becomes stable

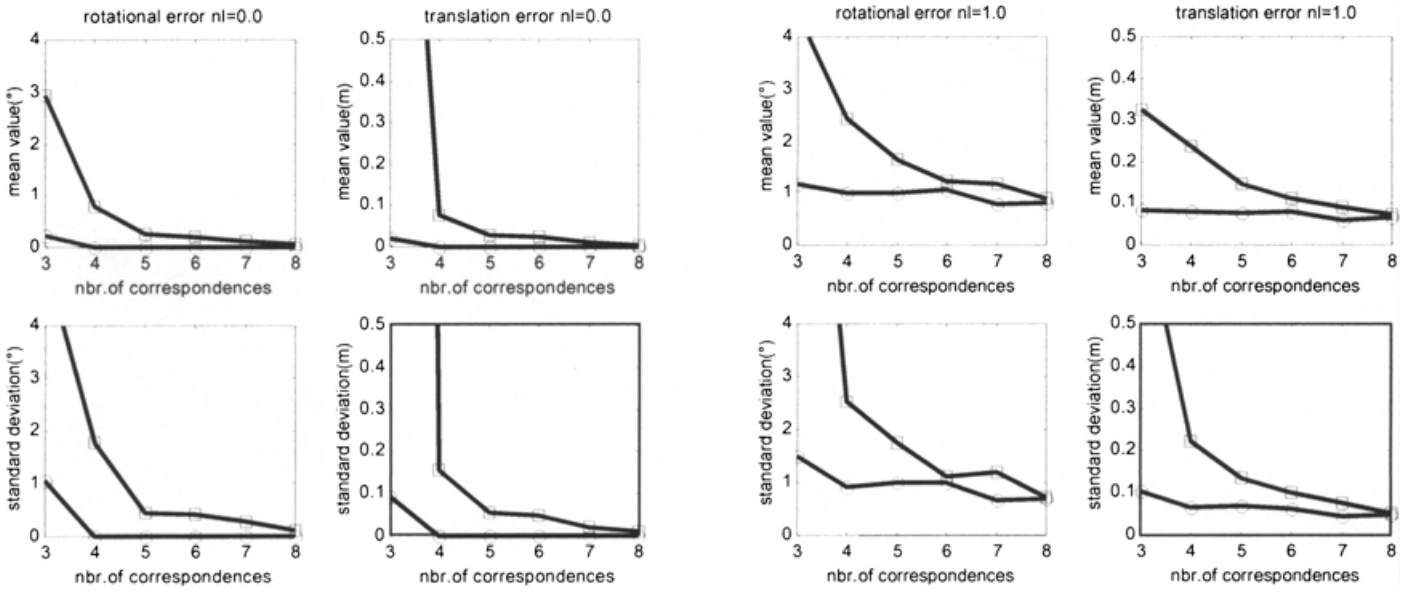


Fig. 7. Sensitivity to the number of correspondences, □: Araujo's method, ○: 3DOF method.

https://doi.org/10.1017/S0263574702004186 Published online by Cambridge University Press

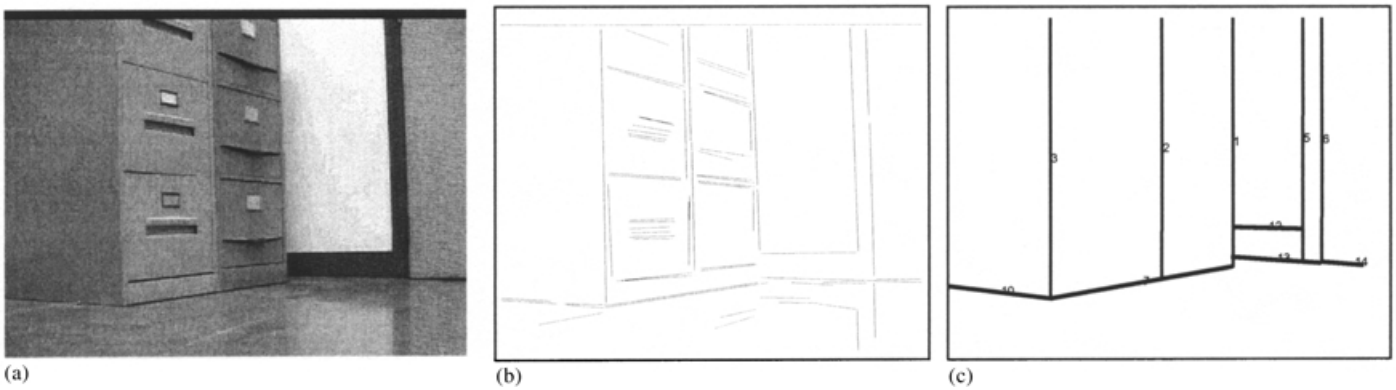


Fig. 8. Test with real images: a – original image; b – image lines; c – projected model lines.

Table III. Sample of results from tests on real images.

Image number	Rotational error (°)	Translation error (m)
1	0.20	0.00
2	0.30	0.00
3	0.46	0.00
4	0.50	0.01
5	0.50	0.02
6	0.53	0.02
7	1.56	0.05
8	1.60	0.05
9	1.80	0.07
10	2.10	0.08
11	2.80	0.08
12	2.90	0.11

only when the number of correspondences is greater than 7.

It should be noted that sharply better results can be obtained using minimisation methods with better convergence properties, such as Levenberg-Marquardt or trust region.

Testing with real images. The goal of this test is to confirm the applicability of Lowe's new formulation method, as presented herein, in a real application. The internal model of the SONY FCB-IX47 camera was first calculated by means of a calibration procedure. A 3D model of a portion of the laboratory was then established and the camera was posed at several locations. For each location, two images were taken without moving the camera. On the first image, a set of known points was added to the environment in order to refine the reference location estimated by manual measurements using a calibration procedure for calculating extrinsic parameters. The 3DOF method was then run using the second image. Sample results are presented in Table III. Figure 8a shows a sample of the images used in these tests. Lines extracted from image contours (Figure 8b) have been matched with lines estimated by model projections (Figure 8c).

The results in Table III show that the rotational and translation error distributions obtained serve to confirm the statistical results of the simulation. In comparison with simulation tests, nl can be situated between 0.25 and 0.75. From these initial results, the 3DOF method seems well adapted to mobile robotics in terms of accuracy.

5. CONCLUSION AND FURTHER WORK

A new formulation of Lowe's camera pose recovery method was implemented for the mobile robot self-localisation application. In light of the specificity of the domain, certain modifications were performed on the original formulation:

- straight line correspondences, rather than point correspondences, were used;
- equations were expressed in a way that allows reducing the number of degrees of freedom and using the full-perspective projection;
- the full-perspective projection model was used to remove affine approximations from the original formulation; and
- error functions were expressed directly with respect to location parameters.

Results show that these modifications considerably improve performance of the original method. The improvements are significant in that they add reliability and accuracy to this pose recovery algorithm for the automatic 2D-3D matching problem. The matching algorithms used for generating 2D-3D correspondence hypotheses do indeed call upon pose recovery procedures in their inner loops. One example would be prediction-verification methods[4,19] or focal extensions. The time savings are then multiplied by the number of calls on the pose recovery algorithm, with this number varying to a wide extent depending on the number of correspondences, geometric constraints and the initial solution estimate.

References

1. M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", *Communication of the ACM* **24**, No. 6, 381–395 (1981).
2. M. Dhome, M. Richetin, J.T. Lapreste and G. Rives, "Determination of the attitude of 3-D objects from single perspective view", *IEEE Trans. on Pattern Analysis and Machine Intelligence* **11**, No. 12, 1256–1278 (1989).
3. L. Quan and Z. Laun, "Linear N-point camera pose determination", *IEEE Trans. On Pattern Analysis and Machine Intelligence* **21**, No. 8, 774–780 (1999).

https://doi.org/10.1017/S0263574702004186 Published online by Cambridge University Press

4. D.G. Lowe, *Perceptual Organization and Visual Recognition* (Boston, MA, Kluwer, 1985), ch. 7.
5. D.G. Lowe, "Three-dimensional object recognition from single two dimensional images", *Artificial Intelligence* **31**, No. 3, 355–395 (1987).
6. D.G. Lowe, "Fitting parmetrized three-dimensional models to images", *IEEE Trans. on Pattern Analysis and Machine Intelligence* **13**, No. 5, 441–450 (1991).
7. H. Araujo, R. Carceroni and C. Brown, "A fully projective formulation for Lowe's tracking algorithm", *Technical Report 641* (Univ. of Rochester, 1996).
8. Y. Liu, T.S. Huang and O.D. Faugeras, "Determination of camera location from 2D to 3D line and point correspondences", *IEEE Trans. on Pattern Analysis and Machine Intelligence* **12**, No. 1, 28–37 (1999).
9. T.Q. Phong, R. Horaud and P.D. Tao, "Object pose from 2-D to 3-D point and line correspondences", *Int. J. of Computer Vision* **15**, 225–243 (1995).
10. D.F. Dementhon and L.S. Davis, "Model-based object post in 25 lines of code correspondences", *Int. J. of Computer Vision* **15**, 123–141 (1995).
11. R. Horaud, S. Christy and F. Dornaika, "Object pose: the link between weak perspective, para perspective and full perspective", *Technical Report RR-2356* (INRIA, 1994).
12. C.P. Lu, G.D. Hager and E. Mjolsness, "Fast and globally convergent pose estimation from video images", *IEEE Trans. on Pattern Analysis and Machine Intelligence* **22**, No. 6, 610–622 (2000).
13. R. Horaud and O. Monga, *Vision par Ordinateur, outils Fondatmentaux* (Hermes, Paris, France, 1993).
14. O.D. Faugeras, *Three Dimensional Computer Vision: a geometric viewpoint* (MIT Press, Boston, 1993).
15. P. Puget and T. Skordas, "Calibrating a mobile camera", *Image and Vision Computing* **8**, 341–347 (1990).
16. R. Talluri and J.K. Aggarwal, "Mobile robot self-location using model-image feature correspondence", *IEEE Trans. on Pattern Analysis and Machine-Intelligence* **12**, No. 1, 63–77 (1996).
17. P.S. Lee, Y.E. Shen and L.L. Wang, "Model-based location of automated guided vehicles in the navigation session by 3D computer vision", *Journal of Robotic Systems* **11**, No. 3, 181–195 (1994).
18. J. Borenstein, H.R. Everett, L. Feng and D. Wehe, "Mobile robot positioning: sensors and techniques", *Journal of Robotic Systems* **14**, No. 4, 231–249 (1997).
19. X. Pennec, "Toward a generic framework for recognition based on uncertain geometric features", *Vider: Journal of Computer Vision Research* **1**, No. 2, 57–87 (1998).

https://doi.org/10.1017/S0263574702004186 Published online by Cambridge University Press