Improving Pose Estimation Using Image, Sensor and Model Uncertainty

V. Caglioti‡, F. Mainardi‡,
M. Pilu§, D. G. Sorrenti†

‡Dipartimento di Elettronica e Informazione
Politecnico di Milano - ITALY

§Robotics and Vision Group - Department of Artificial Intelligence
University of Edinburgh - SCOTLAND

†Vision and Robotics Lab. - Dept. of Electronic&Electrical Engeneering
King's College - London - ENGLAND

Abstract

This work proposes a methodology for the analysis of the uncertainty in the localization of objects when considering uncertain image data, camera and object geometry parameters. The uncertainty is propagated through an extended static Kalman filter initialized with the parameters used for the localization and updated with new matched features obtained by back-projecting onto the image. At the end of the process, a better estimate of the object pose with its uncertainty is given along with a new estimate of the used uncertain object features and the camera parameters. The methodology is now in use in an object localization system.

1 Introduction

The role of uncertainty is very important where a measure of the position and orientation of a modeled object in space is required. Besides providing information on the reliability of the data, the reason for using the uncertainty of the sensory data is for improving the estimate of the parameters we want to measure. This improvement is obtained by combining pieces of information derived from different sensory data according to their uncertainty. Durrant-White [9] introduced a method for uncertainty propagation, even in presence of deterministic relations between different observed features. By this approach, however, the determination of the *a posteriori* covariance matrix is difficult. Studies on uncertainty propagation analysis have already appeared in the Computer Vision literature (e.g. [17] [10] [6], [13]). In many of the presented approaches, uncertainty in the image features has been regarded as the main source of uncertainty. However, we notice that uncertainty in both the projection parameters of the camera and in the object geometry is often more relevant than uncertainty in the image features.

In the present work we propose a methodology for the uncertainty propagation in the localization of modeled objects. The uncertainty sources considered include

BMVC 1994 doi:10.5244/C.8.78

(i) the image data, *(ii)* the projection parameters of the camera and *(iii)* the object geometry. The propagation from the last two uncertainty sources had not been studied in previous works. Uncertainty on the geometric features of the object model are not only due to unpredictable deviations from their nominal values in man-made objects, but they are also due to the adoption of simplified model, as in modeling a smoothed edge as sharp (see Fig. 1). We show that when all these uncertainty sources are taken into account, a better estimate of the object pose is obtained, both in term of estimate and variance. As in other studies, such as [7], the deviations from the nominal values are assumed to be normally distributed.

The technique adopted to analyse the uncertainty propagation is based on the Static Extended Kalman Filter. Equivalently, a non-linear weighted least squares technique (such as the Levenberg-Marquardt Method [12]) could have been used, with weights inversely proportional to the variance of the employed data.

A system that employs the proposed method has been implemented. Basically it aims to localize polyhedral objects from single images using a minimal sub-pattern matching algorithm developed by Caglioti [5]; it will be shown how the proposed method helps improve the pose estimate of the localized object in the presence of uncertainty on image features, object geometry and projection parameters.

The rest of the paper is organized into two main sections: Section 2 describes the methodology we use, and Section 3 gives a concrete example and some experimental results.

2 Uncertainty Propagation Analysis

In this section we first describe how we represent the uncertainty of the image, objects and camera parameters. Then we explain how the proposed methodology is well suited to the minimal sub-pattern matching problem. Finally we show how the uncertainty propagation is carried out for the general case in which image, object and camera parameters are uncertain.

2.1 Uncertainty Model of the Image Features

An image of the scene is taken by a calibrated camera and some features are extracted from it.

Let us call $\mathbf{p} \in \Re^m$ the vector of the parameters defining these features. Because of the nature of the image acquisition process, $\mathbf{p}$ is affected by uncertainty that can be described, under very general assumptions (see [11], [1], [7], [9], [8] and [10]) with a probability distribution function $f_g(\mathbf{p})$.

The foundation of our approach is based on representing uncertainty of $\mathbf{p}$ by its covariance matrix $\Lambda_\mathbf{p}$ with an assumed normal probability distribution $N(\overline{\mathbf{p}}, \Lambda_\mathbf{p})$. The advantage of this assumption is that there are well-known statistical tools for estimation and decision problems (e.g. Kalman filter and χ^2 test). The estimation of $\Lambda_\mathbf{p}$ can be carried out using specific techniques (e.g. [9],[1], [4],[16]).

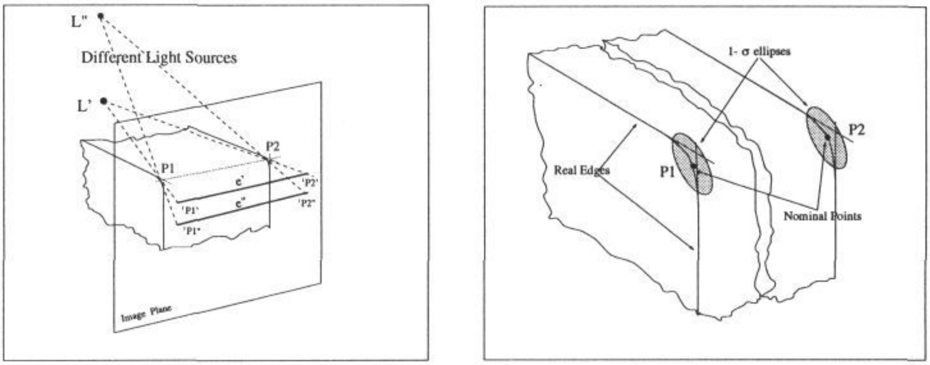

Figure 1: Effect of lighting on a smoothed edge (left). By considering as uncertain the straight edge in the polyhedral model (right) it is possible to account for this effect

2.2 Object Uncertainty

The problem of the propagation of the uncertainty of the *object* geometric features involved in the localization has not been studied previously. However we believe that by introducing the object model uncertainty, a deeper analysis and comprehension of the scene is obtained.

Let us consider an object as composed of certain features that can each be represented as a point $\mathbf{q}_i \in \Re^k$ in the k-dimensional space of the feature parameters. Some reasons for introducing object model uncertainty are the following:

1. Real objects in a scene normally exhibit small but unpredictable deviations from the nominal geometry especially for man-made and large scale industrial production objects.

2. In certain cases it is possible to account for the use of simplified object models by assigning uncertainty to the not properly modeled features. For example, Figure 1 shows the case of a smoothed edge modeled as sharp. Giving proper uncertainty to the edge in the polyhedral model (i.e. according to its curvature), we can allow for the unpredictable shift of the edge in the image due to changes of lighting conditions.

The approach we propose is to treat every unknown disturbance due to the object parametrization by considering the uncertain object parameters as normally distributed $N(\overline{\mathbf{q}}_i, \Lambda_{\mathbf{q}_i})$.

2.3 Uncertainty of Camera Calibration Parameters

Practically all calibration methods (e.g. [15]) involve the treatment of a fairly great number of observations and thus, from the Central Limit Theorem, all the estimated parameters are provided as Gaussian variables. In this paper we refer to the camera parameters as a vector $\mathbf{f} = N(\overline{\mathbf{f}}, \Lambda_f)$.

2.4 Localizing by Sub-patterns and Improving Hypothesis Confidence

Let $\mathbf{P} = \{p_i \in \Re^m\}$ and $\mathbf{Q} = \{q_j \in \Re^k\}$ be the sets of image and object features respectively. Following some criteria, clusters (patterns) of image features are put in correspondence with clusters of objects, and possible object position hypotheses are produced by a proper localization algorithm. Indicating by $\mathbf{y} \in 2^{\mathbf{P}}$ an *image pattern P*, by $\mathbf{e} \in 2^{\mathbf{Q}}$ an *object pattern* and by $\mathbf{f}$ the vector of sensor parameters, a localization algorithm L can be expressed as

$$L(\mathbf{y}, \mathbf{e}, \mathbf{f}) \Longrightarrow \begin{cases} \text{not\_matched} \\ \text{position vector } \mathbf{x} \in \Re^6 \end{cases}$$

In this paper we consider a generic L that uses only the 6 parameters strictly necessary to localize. Several efficient algorithms of this kind have been proposed (e.g. [3] or [5]) but all yield poor precision due to the small amount of information used.

The assumption of small uncertainty of image, objects and sensor parameters allow us to use nominal values for the generation of nominal object position hypotheses, that is $\overline{\mathbf{x}} = L(\overline{\mathbf{y}}, \overline{\mathbf{e}}, \overline{\mathbf{f}})$.

After the initial estimate of the position, the object features are projected onto the image; let us call this set of projected feature $\mathbf{P}^I = h(\mathbf{x}, \mathbf{Q}, \mathbf{f})$.

Further matches between real images features $\mathbf{p}_j \in 2^{\mathbf{P}}$ and projected features $\mathbf{p}_i^I \in 2^{\mathbf{P}^I}$ can be used to improve our knowledge of the object position. A proper statistical test $S(\cdot,\cdot)$ is applied to each pair $(\mathbf{p}_i^I, \mathbf{p}_j)$ to determine whether the correspondence is correct. As we shall see in the next section, after each k^{th} positive test, $(\mathbf{p}_j)_k$ is fed into an Extended Kalman filter that yields an improved estimate of the pose along with its uncertainty. Moreover, within our framework of uncertain objects, some objects and camera parameters could also be inserted in the state vector of the filtering process.

2.5 Uncertainty Propagation with Uncertain Images, Model and Calibration Parameters

The model of the visual observation can be expressed as:

$$\mathbf{y} = h(\mathbf{x}, \mathbf{e}, \mathbf{f}) + \mathbf{v} \ , \tag{1}$$

where

- $\mathbf{y}$: vector of image parameters (the observation);
- $\mathbf{e}$: vector of model parameters;
- $\mathbf{f}$: vector of camera parameters;
- $\mathbf{x}$: the object position-orientation vector;
- h: camera projection vector-like function (e.g. perspective transformation);
- $\mathbf{v}$: a small additive noise affecting the extraction of image features.

Once the correspondence hypothesis between an image pattern and a model one has been generated, a nominal value $\overline{\mathbf{y}}$, $\overline{\mathbf{x}}$, $\overline{\mathbf{e}}$ and $\overline{\mathbf{f}}$ of the corresponding uncertain vectors is available through L. Relation 1 can be linearized in proximity of the nominal value. If δ indicates the variation with respect to the nominal value, then:

$$\delta \mathbf{y} = \mathbf{J}\delta\mathbf{x} + \mathbf{H}\delta\mathbf{e} + \mathbf{K}\delta\mathbf{f} + \mathbf{v} \tag{2}$$

where $\mathbf{J} = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\overline{x},\overline{e},\overline{f}}$, $\mathbf{H} = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{e}}\right|_{\overline{x},\overline{e},\overline{f}}$ and $\mathbf{K} = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{f}}\right|_{\overline{x},\overline{e},\overline{f}}$.

Let $\mathbf{s}$ be the vector $\mathbf{s} = \left[\ \mathbf{x}\ \mathbf{f}\ \mathbf{e}\ \right]^T$. As above, we assume $\mathbf{s} = N(\mathbf{s}, \Lambda_s)$, as we intend to propagate the normal Gaussian uncertainty of image and camera parameters through the linearized inverse camera projection. Such a linear transformation combines Gaussian variables and yields Gaussian variables.

2.5.1 Initialization of the Covariance Matrix

Consider equation 2, linearized in the proximity of the nominal values of the uncertain parameters. Suppose that the minimal localization algorithm L has not yet been applied: both the covariance matrix Λ_e of the model geometry parameters and Λ_f of the camera parameters are given along with their respective estimates, while both the position estimate and the covariance are still unknown. Therefore we can apply a modified version of the Kalman Filter equations ([9],[14]) in which the a priori information matrix Λ_x^{-1} of the position parameters is supposed to be zero (i.e. no information about $\mathbf{x}$ is available).

By applying a matrix inversion Lemma (see [14]), the following a posteriori covariance matrix Λ_s of the state vector $\mathbf{s}$ is obtained:

$$\Lambda_s = \begin{bmatrix} \mathbf{J}^{-1}(\Lambda_y + \mathbf{H}\Lambda_e\mathbf{H}^T + \mathbf{K}\Lambda_f\mathbf{K}^T)(\mathbf{J}^T)^{-1} & -\mathbf{J}^{-1}\mathbf{K}\Lambda_f & -\mathbf{J}^{-1}\mathbf{H}\Lambda_e \\ -\Lambda_f\mathbf{K}^T(\mathbf{J}^T)^{-1} & \Lambda_f & \mathbf{0} \\ -\Lambda_e\mathbf{H}^T(\mathbf{J}^T)^{-1} & \mathbf{0} & \Lambda_e \end{bmatrix}$$

The estimate of the state vector $\mathbf{s}$, after the localization, is given by $\overline{\mathbf{s}} = \left[\ \overline{\mathbf{x}}\ \overline{\mathbf{f}}\ \overline{\mathbf{e}}\ \right]^T$.

The *a posteriori* covariance matrix shows that a correlation arises between $\mathbf{x}$ and both $\mathbf{f}$ and the initial value of $\mathbf{e}$, while no correlation between $\mathbf{f}$ and $\mathbf{e}$ is introduced. In addition, the covariance matrices of the $\mathbf{f}$ and $\mathbf{e}$ parameters are left unchanged: this means that no information on them is derived from the localization. This is due to the zero *a priori* information about $\mathbf{x}$: the whole of the information derived from the localization is "absorbed" by $\mathbf{x}$.

2.5.2 Exploiting Further Information

From the initial estimate of $\mathbf{s}$ we can calculate the expected image of the object using the h function of paragraph 2.4. In order to decide whether an image feature $\mathbf{p}_k$ matches one of the homologous object feature $\mathbf{q}_k$, it is compared to the projected feature whose expected value is given by $\overline{\mathbf{p}}_k^I = h(\overline{\mathbf{x}}, \overline{\mathbf{q}}_k, \overline{\mathbf{f}})$, where $\mathbf{x}$, $\mathbf{q}_k$, $\mathbf{f}$ are the current estimates of the position, the model features and camera parameters. As additional features are recognized to be part of the object image, the object position $\mathbf{x}$ and also both calibration parameters $\mathbf{f}$ and model geometry are updated. In general, all these parameters are *a posteriori* correlated.

Let $\mathbf{Q}_k = [\ \mathbf{J}_k \quad \mathbf{K}_k \quad \mathbf{H}_k\ ]$, where $\mathbf{J}_k$, $\mathbf{K}_k$, $\mathbf{H}_k$ are the jacobian matrices of $\mathbf{p}_k^I$ with respect to $\mathbf{x}$, $\mathbf{q}_k$, $\mathbf{f}$ respectively. From equation 2, the covariance matrix of $\mathbf{p}_k^I$ is:

$$\Lambda_{p_k^I} = [\mathbf{Q}_k] \cdot \Lambda_s \cdot [\mathbf{Q}_k]^T$$

A statistical test S can be used to decide if image features $(\mathbf{p}_i)_k$ match object features $(\mathbf{q}_j)_k$. An example of S is the Mahalanobis distance [14], which gives the following matching criterion:

$$\frac{1}{2}(\mathbf{p}_k - \mathbf{p}_k^I)(\Lambda_{p_k} - \Lambda_{p_k^I})^{-1}(\mathbf{p}_k - \mathbf{p}_k^I)^T \leq 1$$

Once a match has been found, the estimates of $\mathbf{x}$, $\mathbf{e}$, $\mathbf{f}$ are updated along with their covariance matrix. Suppose that there are N of these matches and that each $\mathbf{p}_k$ (that is the k^{th} observation) is statistically independent of the previous *k-1*; otherwise, a procedure for obtaining a $\mathbf{p}_k'$ independent from the past but containing all the information necessary to update the estimate is straightforward (e.g. [2]). The updating algorithm for the supplementary matched features is the following:

$$\text{for } k\text{=1 to } N \left\{ \begin{array}{c} \hat{\mathbf{s}} = \overline{\mathbf{s}} + \Lambda_s \mathbf{Q}_k^T (\mathbf{Q}_k \Lambda_s \mathbf{Q}_k^T + \Lambda_{p_k})^{-1} (\mathbf{p}_k - \mathbf{p}_k^I) \\ \hat{\Lambda}_s = \Lambda_s - \Lambda_s \mathbf{Q}_k^T (\mathbf{Q}_k \Lambda_s \mathbf{J}_k^T + \Lambda_{p_k})^{-1} \mathbf{Q}_k \Lambda_s \\ \mathbf{s} \longleftarrow N\ (\ \hat{\mathbf{s}}\ ,\ \hat{\Lambda}_s\ ) \end{array} \right.$$

By these equations, not only the object position $\mathbf{x}$ is updated, but the supplementary matches between model and image features yield also new information about both the calibration parameters and the model geometry.

The procedure is general, but it is easy to derive the case in which $\mathbf{f}$ or $\mathbf{e}$ or both are not updated. In particular it is obvious that an improvement of calibration parameters should be pursued only in the case the object is well known ($\Lambda_e = 0$). In the case $\Lambda_e \neq 0$ there is no reason for estimating new camera parameters.

3 Experimental Results

We describe a system developed to experiment with the methodology proposed in Section 2; both simulation and real-image experiments will be described.

3.1 Overview of the Experimental System

The system addresses the sub-case of localization of polyhedra in a complex scene acquired by a single calibrated camera using an algorithm developed by Caglioti [5]. Both image and camera parameters are considered uncertain (2.1 and 2.3) and a Gaussian uncertainty is given to the object models (see 2.2). The raw image is first processed by a Canny edge detector and later is segmented into straight edges whose uncertainties (due chiefly to spread and discretization of points) are represented by normal distributions around their nominal (ρ,ϑ) parameters (see, e.g., [10] [4] [17]); the method we used to actually compute this uncertainty has been described in [4].

At this point we create all possible pairings of image and object patterns that comply to certain qualitative constraints (e.g. concavity-convexity angles); in the present implementation, the number of these pairs is of the order of the number of model patterns.

Each correspondence hypothesis is translated by the localization algorithm [5] into a possible (since the process may fail) object position hypothesis, expressed in Euler coordinates with respect to the camera reference frame.

Now the filtering process starts with the initialization of the Kalman filter as explained in 2.5.1; it should be noticed that Λ_y also refers to coordinates of intersection between edges and hence simple geometric transformations of the plain (ρ,ϑ) uncertainties are done ([11], [17]).

After the initialization, the process of improving the pose estimate goes on by back-projecting the object onto the image plane, looking for additional matches (Section 2.4) and applying the Extended Kalman filter equations, as shown in 2.5.2.

3.2 Experimental Results: Simulation

A simulation experiment consists of taking an object model, rototranslating it with known Euler coordinates and projecting it onto the image plane with given camera parameters; all edges are given (ρ,ϑ) uncertainty. For brevity, we expose here only two different sets simulations (Figure 2).

Both sets have in common the kind of object, its position in the space and the (camera) parameters of the perspective projection. Six points of the object base (see Frame 4 and 8) have been given an equally distributed Gaussian uncertainty along their reference axis. Frames 1,5 show the two images along with the initial pose estimates; frames 2,6 give the final estimate *not considering* object uncertainty and frames 3,7 show the final pose estimate *including* object uncertainty, whose final uncertainty ellipses are displayed in frames 4 and 8. The focal length uncertainty was taken into account in both cases. The position is given as $(\phi, \theta, \psi, p_x, p_y, p_z)$, with angle measured in radiants and lengths in millimeters.

Set A (1-4) **Edge 1** has a 4% shift from its nominal position in the object reference frame. It can be easily seen that the final estimate with deterministic model (2) is actually worse than the initial one (1) because of the shift of **Edge 1** in the object that is not accounted for in the model. With object uncertainty we get an almost perfect estimate (3) and also the instance of the object has been properly updated (4).

Set B (5-8) Here **Point 6** has a 5% shift-up with respect to the nominal prototype and the simulated focal length is significantly different (10%) from the true one; this causes an enlargement of the *expected* image with respect to the actual one. Due to the combination of two disturbing factors, the final estimate without object uncertainty is rather bad (6), even poorer than the inital one (5). Conversely, including the object uncertainty we get a very good final estimate of the pose (7); the only error is (as expected) in the distance since, in order to "fit" the object into the image taken with a smaller focal length, the system "put it slightly further". In the model instance (10) **Point 6** has been correctly updated according to the image data and its uncertainty has considerably diminished.

Even so few examples show that the inclusion of the object uncertainty helps

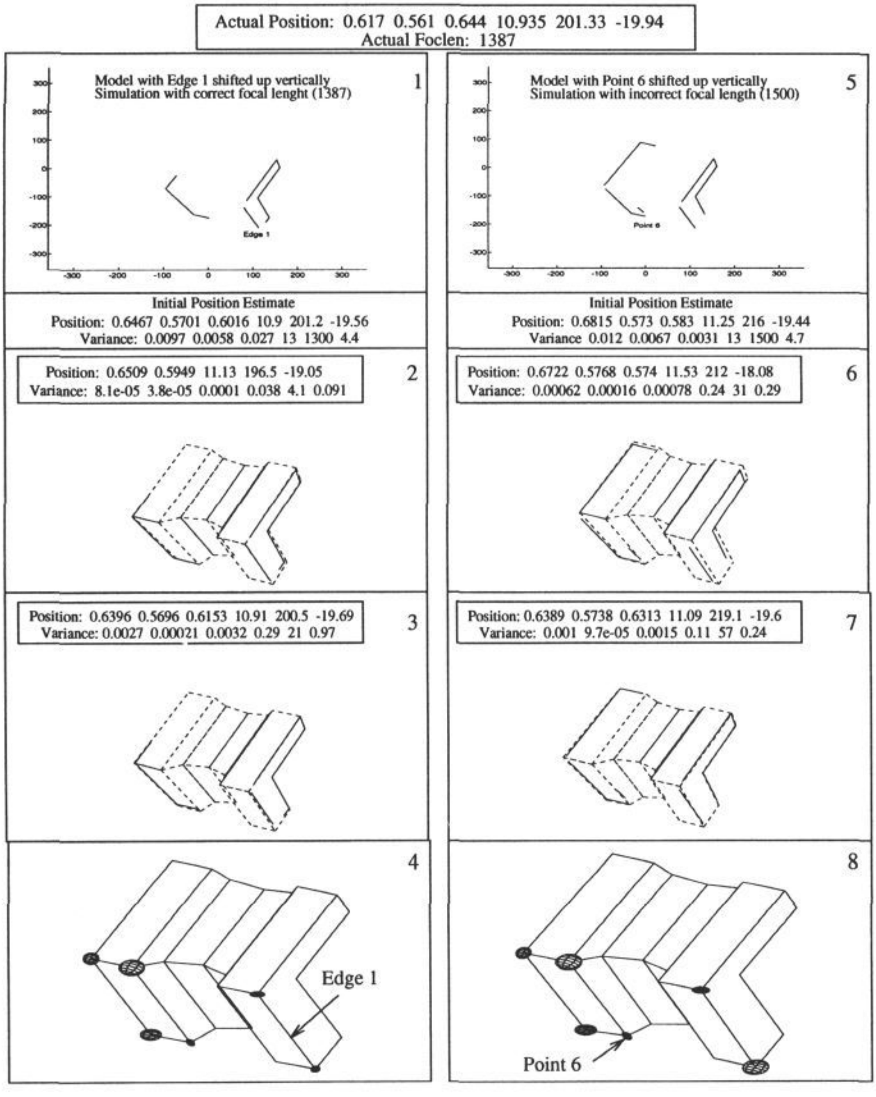

Figure 2: Two sets of simulation experiments.

improve the pose estimation in presence of external disturbance (like a wrong knowledge of the camera parameters and slightly different geometry of instances of objects in the scene) and reduces the sensibility of it under different conditions (e.g. compare 3/7 and 2/6 in **Set A** and **Set B**).

3.3 Experimental Results: Real Images

The uncertainty propagation method illustrated in Sec. 2 has also been tested on real images, such as the bottle with smoothed side edges shown in Figure 3(left) with superimposed edges.

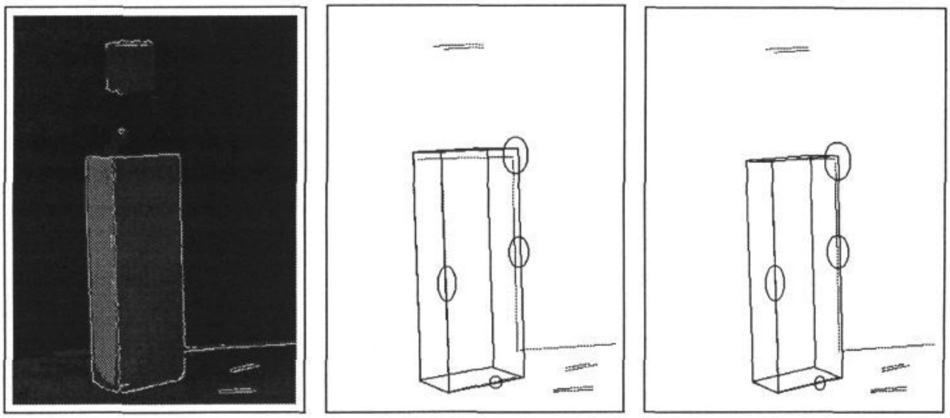

Figure 3: Experimental results with a real image of a smoothed-edge bottle

Figure 3(middle) shows the final estimate of the bottle body superimposed to the segmented image *without* taking account of camera and object uncertainty. It can be easily seen how the top and the right side have a poor fit with the image. The line uncertainty alone, which is well below 1 pixel, cannot account for this displacement. However, this displacement can be accounted for if the uncertainty both in the calibration parameters and in the object model is considered, as shown by the final estimate of the object pose in Fig. 3(right). Small circles indicates zones in which the improvement is relevant.

	ϕ		θ		ψ		p_x		p_y		p_z	
True	90.0		0.0		-60.0		0.0		116.0		0.15	
Det.	92.47	0.65	1.73	1.08	-61.36	0.56	0.27	1.63	113.00	3.01	1.00	0.87
Unc.	92.59	1.81	1.22	2.92	-60.63	2.26	0.25	2.42	115.25	3.04	0.34	0.58

The table shown above gives the true values and the estimates of the position (pairs of value and standard deviation) with deterministic model and camera parameters (DET) and with uncertainty (UNC); angles are in degrees and lengths in centimeters. The overall improvement in the estimate is evident and, in particular, the standard deviation is more in accordance with the real error, while it is much smaller in the case in which only line uncertainty is considered. This shows the inadequacy of an uncertainty model that neglects the uncertainty in either the projection parameters or the model geometry.

4 Conclusions

This paper analyses the effect of different uncertainty sources in object localization. Of the considered uncertainty sources – namely image features, camera projection parameters and object geometry – only the first one was considered in previous approaches to uncertainty propagation. Simulation results had shown how the position estimate benefits from taking all the three above uncertainty sources into account. Experimental results have demonstrated the application of the methodology to the localization of real objects.

5 Acknowledgements

The system described in this work will be used by a robot designed to help disabled peoples in various activities. V.Caglioti and D.Sorrenti acknowledge the support of the Italian National Research Council. M.Pilu is partially sponsored by INMOS ltd. The authors want to thank David Wren and Andrew Fitzgibbon for proof-reading the final document.

References

[1] N. Ayache and O. D. Faugeras. Mantaining representations of the environment of a mobile robot. *Robotic Research*, (4):337–350, 1988. MIT Press, USA.

[2] G. BIerman. *Factorization Methods for Discrete Sequential Estimation.* Academic Press, 1977.

[3] R. Bolles, P. Horaud, and M. Hannah. 3DPO: A three dimensional part orientation system. In M. Brady and R. Paul, editors, *Robotic Research: The First Symposium*, Cambridge, 1984. MIT Press.

[4] V. Caglioti. On the uncertainty of straight-lines in digital images. *CVGIP-Graphical Models and Image Processing*, 55(4):225–270, 1993.

[5] V. Caglioti. The planar three-line junction perspective problem with application to the recognition of polygonal patterns. *Pattern Recognition*, 26(11), 1993.

[6] T. A. Cass. Polynomial-time object recognition in the presence of clutter, occlusion, and uncertainty. In G. Sandini, editor, *Computer Vision - ECCV '92*, Berlin, 1992. Lecture Notes in Computer Science, Springer-Verlag.

[7] H. F. Durrant-Whyte. Consistent integration and propagation of disparate sensor observations. In *IEEE International Conference of Robotics and Automation*, pages 1464–69, 1986.

[8] H. F. Durrant-Whyte. *Integration, Coordination and Control of Multisensor Robot System.* Kluwer Academic Publisher, 1988.

[9] H. F. Durrant-Whyte. Uncertain geometry in robotics. *IEEE Journal of Robotics and Automation*, 4(1), Feb. 1988.

[10] S. Gordon and W. Seering. Real-time part position sensing. *IEEE PAMI*, 10(3), 1988.

[11] F. Mainardi and M. Pilu. Analisi della incertezza nella localizzazione di oggetti a partire da immagini singole. Tesi di laurea, Politecnico di Milano, Dec. 1990. In Italian.

[12] D. W. Marquardt. *Journal of the Society for the Industrial and Applied Mathematics*, 11:431–441, 1963.

[13] J. Porrill. Optimal combination and constraints for geometrical sensor data. *International Journal of Robotics Research*, 7:66–77, 1988.

[14] F. Shweppe. *Uncertain Dynamic Systems.* Prentice-Hall, 1973.

[15] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, (4), 1987.

[16] M. Waite, M. Orr, R. Fisher., and J. Hallam. Statistical partial constraints for 3D model matching and pose estimation problems. *Proceeding of the British Machine Vision Conference*, pages 105–114, Sept. 1993.

[17] S. Yi, R. M. Haralick, and L. Shapiro. Error propagation in machine vision. *Machine Vision and Applications*, 7:93–114, 1994.