Non-polyhedral landmark recognition using 3D depth images and partially correct models

Philippe Fillatreau   Robert B. Fisher *

Department of Artificial Intelligence
University of Edinburgh
5, Forrest Hill
Edinburgh EH1 2QL
United Kingdom

In this paper, we address the problem of modeling, recognising and localising non-polyhedral shapes or objects in a partially known scenes, using 3D depth images. The applications this work aims at are mobile robotics for intervention or survey in environments like factories or nuclear plants.

This paper focusses on the problems of the recognition of non-polyhedral landmarks or objects and misplaced feature detection. In our approach, planes and some simplified classes of quadrics play an important part in the different processes. Surface classification takes into account the uncertainties on the measures and relies on probabilistic tests. The surface features belonging to those simplified classes provide a pose estimate which can be refined, if needed, using iconic methods. After registration, we compare the scene description and the a priori model for model validation and change detection.

1   Introduction

This paper deals with the problems of modeling, recognising and localising non-polyhedral shapes or objects in a partially known scene. Applications are mobile robotics for intervention or survey in environments like factories or nuclear plants. For such environments, the *a priori* models provided by engineering drawings may no longer be an accurate description of the actual factory, because of differences in construction or subsequent modifications. Our work investigates environment modeling issues, focussing on the problem of the correction of the *a priori* maps, using a mobile robot, 3D depth images, and partially correct maps.

Whereas much work has been done previously on 1) world modeling using polyhedra or planar surface facets and 2) scene understanding where the scene is known exactly in advance, current applications require the use of non-planar primitives for scene descriptions, and show the need to tackle cases where the *a priori* model of the environment is only partially correct. These are two important features of the investigated project.

This paper focusses on the problems of the recognition of non-polyhedral landmarks or objects (known by a partially correct CAD-type [10] model) and misplaced feature detection. In this paper, we present the global approach undertaken, and describe the different steps of the procedure: scene description (data segmentation and surface classification), matching and pose estimation, feature misplacement detection. Some simplified classes of non-planar primitives play an important part in all these processes. Surface classification takes into account the uncertainties on the measures and relies on probabilistic tests. The surface features belonging to those simplified classes provide a pose estimate which can be refined, if needed, using iconic methods. After registration, we compare the scene description and the a priori model for model validation and change detection. Finally, we present some experimental results and the next steps of this work. While a mobile vehicle, for the applications this work aims at, will usually be equipped with a multisensor perception system, we focus in this paper on the interpretation of range data.

---

* Email: philippe@aifh.ed.ac.uk, rbf@aifh.ed.ac.uk

The work presented here is based in part on the IMAGINE2 integrated recognition system at the Department of Artificial Intelligence [11].

2 The problem. Overview of Approach

Most research dealing with the problem of environment modeling have concentrated on building *purely geometric* representations; in most cases, environment modeling consists of incremental fusion into a common model of the geometric primitives obtained from the segmentation of the data obtained from successive sensory acquisitions[1, 2, 6, 14, 16, 17]. But geometric models have shown serious limitations with respect of the traditional tasks to perform (robot localisation or navigation for example), because the complexity of those tasks increases as new data are acquired, and the consistency of the models is difficult to keep. Complex environments need heterogeneous representations [8, 9], which must be dedicated to the task to be performed, and several levels of representation are needed [7, 8]:

- a *geometric* level, containing the basic descriptions, and essentially the metrical parameters of the different primitives involved: curves, surfaces, volumetric elements.
- a *symbolic* and *semantic* level, characterising primitives of the same type (that may represent different entities), or sets of primitives (defining landmarks – e.g. a door, a room corner, a group of pipes in a factory – or particular objects as corresponding to a given set of primitives). This allows control of the navigation or localization tasks, choice of relevant information to keep in the model (according to a given task), and rejection of spurious geometric primitives arising in the segmentation.
- a *topologic* level, which can be useful in several ways. Topological relationships between several primitives or semantic entities (for exemple, object A lies at the right of object B, in front of object C,...) can be used for various tasks such as perception planning, or propagation in the model of the environment changes that are detected locally. Besides, a global topologic layer in the model of the environment can be used at a mission definition level. This global layer, for a factory-like environment, could give a representation of the factory as a set of *topologic loci* like corridors and rooms, and the positioning of the local reference frames attached to each of those loci.

At the geometric level, we notice that in man-made environments, many relatively simple features are expected to be present in the environment (walls, polyhedral obstacles, pipes, elbows, ...). Such features can be represented with sufficient accuracy by using a few simple surface primitives (like planes or simplified classes of quadrics). In the strategy defined for landmark recognition and misplaced feature detection, such features play a particularly important part in the control of the process, because (a) they are well-suited for the kind of environments we are interested in (b) they are easier to fit and match than more general primitives and (c) they provide reliable and stable 3D invariants for 3D pose estimation.

In the rest of this paper, we focus on the problems of non-polyhedral landmark recognition and feature misplacement detection. Figure 1 presents the strategy we have defined for the entire process (from sensory data acquisition to feature misplacement detection). We will discuss in detail below the differents steps presented in this diagram.

3 Segmentation and surface classification

Segmentation extracts a collection of homogeneous patches from the range data. A patch corresponds to maximally connected sets of surface points in which the signs of the Gaussian and mean curvature do not change [20]. Local surface shapes are divided into four main classes (planar, cylindrical, elliptic and hyperbolic) and finally planes or quadrics are fitted to the patches.

As simplified classes of quadrics are to play an important part in the object recognition strategy, the next step is to classify the quadric surface patches into the following classes: {planes, cylinders, cones, spheres, others}, labelled as PLAN, CYL, CONE, SPH, OTH. We present here the method we have developped.

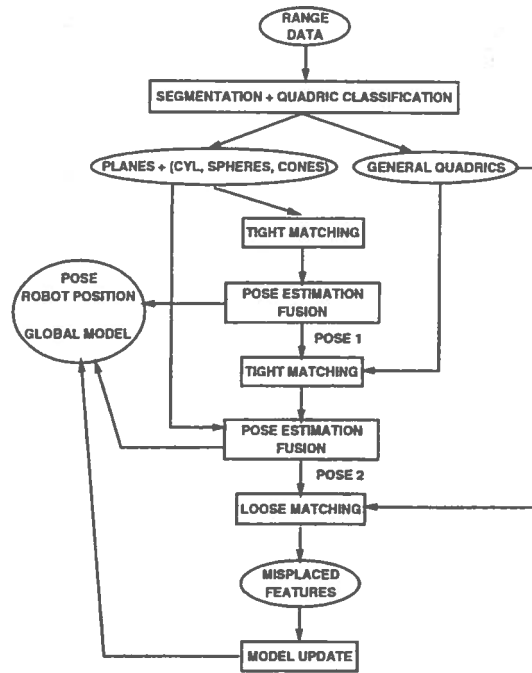Quadrics can be expressed according to the general equation:

Fig. 1.: *Non-polyhedral landmark recognition and change detection: a global approach*

$$X^t.A.X + X^t.B + C = 0$$

Classification can be performed by using the parameters of the canonical forms. Canonicalization finds the (R,T) transform (for central quadrics: ellipsoids and hyperboloids) such that the transformed quadric is expressed by:

$$X'^t.A_{cano}.X' + C_{cano} = 0$$

where $A_{cano} = diag(\lambda_1, \lambda_2, \lambda_3)$ and $X' = R.X + T$.
Then, the simplified classes listed above are characterized by:

- cylinder: $C_{cano} < 0$ and $\lambda_1 = \lambda_2 > 0$ and $\lambda_3 = 0$
- sphere: $C_{cano} < 0$ and $\lambda_1 = \lambda_2 = \lambda_3 > 0$
- cone: $C_{cano} = 0$ and $\lambda_1, \lambda_2 > 0$ and $\lambda_3 < 0$

In the presence of noisy data, it is difficult to perform a reliable classification by using empirical thresholds for testing the relations above.

Thus, we take into account the uncertainties on the measured points, and we propagate them through the canonicalization processes. We know that $C_{cano} = C + T^t.(A.T + B)$ where T is such that $B_{cano} = 2T^t.A.R + B^t.R = 0$. Computing the Jacobians of C_{cano} with respect of the different parameters of the quadric, we can compute the uncertainty on C_{cano} from its 1st order Taylor development. We just recall here that if we have the relation $z = f(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)$, where f is non-linear and $(\hat{x}_i)_{i=1\ldots n}$ are uncertain data having a covariance matrix σ_x, then:

$$\sigma_z = \sum_{i=1}^{n} J_i \sigma_{x_i} J_i^t + \sum_{i \neq j} J_i \sigma_{x_i, x_j} J_j^t$$

where J_i is the Jacobian of f with respect of $\hat{x}_i$, and the second term takes into account the correlations between $\hat{x}_i$ and $\hat{x}_j$.

The uncertainties on the eigenvalues of A are more difficult to estimate; we used an upper bound provided by the Bauer-Fike theorem [13].

Finally, for the patch classification, we compute the *plausibilities* that each quadric patch belongs to each class within the set {cylinder,sphere,cone}, according to the class characterizations given above. The canonical parameters are given by their Gaussian distributions. If we are interested in the signs of the parameters, we compute, from the distributions, the *probabilities* that the parameters have the required sign. To compute the *plausibility* that some parameters have a given value (e.g. $C_{cano} = 0$), we use heuristic functions involving the probabilities that the corresponding parameter is greater or lower than the required value (zero).

The uncertainties on the parameters on the fitting quadrics are evaluated by using the Extended Kalman Filter (the initial state is provided by a least-squares fitting; the associated uncertainties are provided by the non-linear optimization theory [3]).

When this classification has been performed, we fit the classified patches with the relevant dedicated primitive (cylinder, sphere, cone); we verify *a posteriori* the classification by evaluating the quality of that dedicated fitting (the average quadratic distance over the fitted patch, if lower than a given threshold, confirms the results of the classification). The final scene description is a set of surface patches labelled as follows: plane, cylinder, sphere, cone or other (general quadric), as well as their parameters and the uncertainties.

4 Matching and pose estimation

From the scene description obtained, we compute, for the simplified quadrics, the 3D invariants to be used for the matching and pose estimation: length and radius for a cylinder, radius for a sphere, and half angle and length for a cone (lengths, of course, are very sensitive to occlusions). The pose is then computed using the surface normals of planes, the centers of 2 matched spheres, or the axes of cylinders or cones (and possibly other elements like the ends of a cylindric element if entirely seen, or the apex of a cone, if actually perceived).

The matching and object recognition processes in the IMAGINE2 system are described in [12]. The current project aims at improving the procedure. In our approach, the matching and recognition processes are performed in two steps:

1) We first use the planes, cylinders, cones and spheres, because (a) they are well suited for the kind of environments we are interested in, (b) they are easier to fit and match than more general primitives and (c) they provide reliable and stable 3D invariants for 3D pose estimation. This allows us to compute a first estimate of the pose, using those *reliable* matchings, and thus to reduce the uncertainty on the pose estimate.

2) Applying the transform deduced from the first pose estimation to all the surface primitives in the perceived scene (and projecting them onto the model), we can now reduce the complexity on the matching of general quadrics. For the latter, it is more difficult to extract reliable and stable invariants. But if we have a first estimate of the pose, other criteria can be used now for the matching of those primitives (like overlapping, spatial proximity, ...).

In both cases, if needed, an Iterative Closest Point (ICP)-like method can be used for fine position registration [5, 15, 21]. Such a refinement of the pose estimation is mainly needed if:

1) the pose estimation is underconstrained and leaves some degrees of freedom (DOF) (we need at least 3 plane normals, or 2 cylinder or cone axis, or 3 sphere centers to estimate the pose), or

2) the uncertainty on the pose estimate needs to be reduced.

5 Feature misplacement detection

This part of the project deals with the problem of validating the correctness of the model and determining when inconsistencies occur; this is a novel approach with respect of the state of the art, as most research

work so far questions the validity of perceived data rather than the correctness of the model [18], and most research work related to change detection so far has been related to image differencing more than on the detection of important structural changes in the model [4].

We may need to validate the model or detect changes on 2 different levels: (1) on a numerical level, which has already been tackled by using numerical fusion tools (the Kalman filter, notably, has been extensively used); (2) on a structural level, if part of the structure of the model is wrong.

In a first step towards the second issue, the current project is investigating the problem of misplaced features detection along two lines:

- We decompose the matching and recognition strategy into two steps. The first one involves a tight matching procedure, with strong constraints, and will eliminate the primitives or object sub-parts that do not satisfy rigidity constraints; the matchings obtained at that stage allow computation of a pose estimate. In the second step, relaxing the constraints in the matching procedure allows to match those primitives or object sub-parts which do not meet rigidity constraints associated to the former pairings but lie in a small enough (in a statistical sense) neighborhood of the predicted position.
- The final part deals with the analysis of differences. 2 levels of analysis (one geometric level and a semantic one) are needed.

  Indeed, we need to be able to check for the presence, absence or misplacement of *objects* or *semantic entities* in the scene with repect of the information contained in the model (some objects may be in a different position from the one given in the model, some objects may have articulated sub-parts, ...), and to distinguish this case from the case where a few surface features (patches) in a given object description do not match the model. Here, we need at a scene description level a semantic grouping operator.

At a geometric level, we can distinguish:

1. *moved features* for features matched in the loose matching procedure;
2. *missing features* if model features that were predicted to be perceived could not be found in a neighbourhood of the predicted positions;
3. *new features* when perceived features could not be matched with one in the model;
4. *altered features* when features appear "at the same place" (i.e. in a close neighbourhood in a statistical sense) in the new model and the old one, but have different shape parameters.

Both semantic and geometric level should be involved in the model validation and change detection procedures. After registration, the presence of an object could be checked by counting the number of model-data surface pairings obtained from the matching procedure, for the patches describing the considered object in the model (in a similar approach to the one presented in [4], where segments detected in an image are used for the validation of wireframe-type buildings models).

Finally, the model has to be updated according to the former labels and the numerical uncertainties on the features. We do not address this problem in this paper.

6 Experimental results

So far, we have used the segmentation, surface patch classification and registration methods presented above 1) on simulated range images from a synthetic model of a factory-like environment (including simulation of the sensory acquisition with given sensor characteristics, and simulation of noise) and 2) on real range images of small objects and of a factory mock-up acquired with a range sensor [19]. Figures 2 and 3 show respectively a cosine shaded range image of a real scene, and the surface patches obtained after segmentation.

The following table shows details of the results we obtained for the quadric patches classification, giving for each quadric region (using the number given figure 3) the plausibilities P(cyl), P(sphere), P(cone) that it is a cylinder, a sphere or a cone. The last column of the table gives the label finally given to those surfaces; a label among CYL, SPH, CONE is given if the corresponding plausibility is the highest and its difference with the other plausibilities is greater than a given threshold (0.2 in our case). All regions not listed in the table were classified as planes during segmentation.
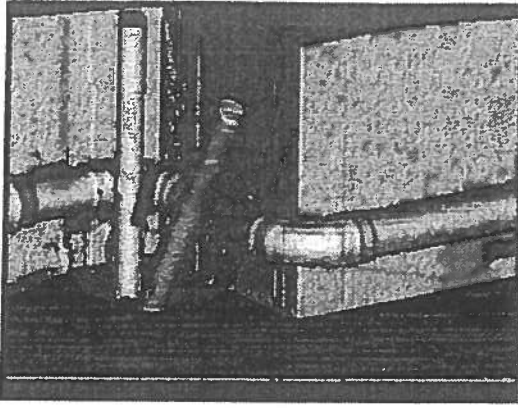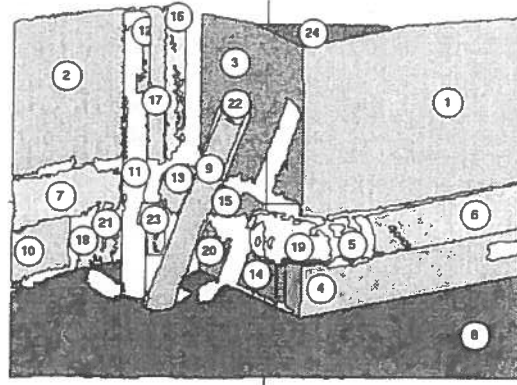
Fig. 2.: *Cosine shaded range data*

Fig. 3.: *Regions image*

The patches corresponding to the description of cylinders in the real scene, and most patches corresponding to planes, have been successfully classified. The quadric patch number 19 corresponds to an elbow between 2 horizontal pipes in the scene, and the quadric patch number 5 corresponds to a junction between a horizontal pipe and an elbow (see figure 2). They have been successfully given the label OTH. Patches number 16, 18, 21, 22, 23 have been classified as quadric patches with the label OTH, whereas they correspond to the description of planar surfaces in the scene. This is due to the fact that they are small patches containing a small number of noisy measure points: a small pipe section, corners, occlusions, lead to noisy data, sensitive to *footprint* (the laser beam intersects several surfaces) or specularity problems. Thus, the corresponding patches could not be labelled as planar patches during segmentation.

Region Num.	P(cyl)	P(sphere)	P(cone)	Class
5	0.05	0	0	OTH
6	0.48	0	0.09	CYL
7	0.54	0	0	CYL
9	0.47	0.03	0.05	CYL
11	0.55	0	0.24	CYL
12	0.33	0	0.11	CYL
13	0.21	0	0	CYL
15	0.34	0.09	0.06	CYL
16	0	0	0.04	OTH
18	0	0	0.01	OTH
19	0	0	0	OTH
21	0.32	0.02	0.15	OTH
22	0	0	0.05	OTH
23	0	0	0.14	OTH

Finally, we present now the results obtained for model-data registration. The factory scene is known by a CAD-type [10] model. We show here two examples. Figures 4 and 5 show an example of a successful hypothesis considered durind the tree-search performed by the matching procedure [11], involving 2 cylinders and one plane pairings. Figure 4 shows the initial pose estimate obtained by using the 2 cylinders and the plane descriptions only (and their associated invariants). The data appear in light colour, the 3D display of the CAD model is displayed in a darker colour. Figure 5 shows the refined pose obtained after calling the ICP procedure. Again, the data appear in light colours, the 3D display of the CAD model is displayed in a darker colour; the matched features are displayed using the darkest colour. Figures 6 and 7 show the results obtained from another hypothesis, involving 3 cylinder pairs this time. The accuracy of the obtained registration is about 1mm; this is illustrated by the overlap between model and data (see

the horizontal pipes, the slanted pipe, and the wall on the right in figure 5, and the horizontal pipes, the slanted pipe and the planes in figure 7).
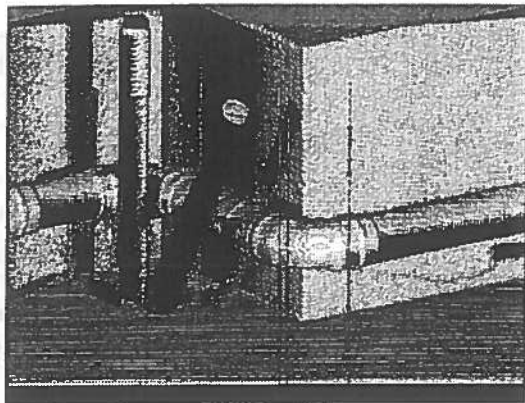

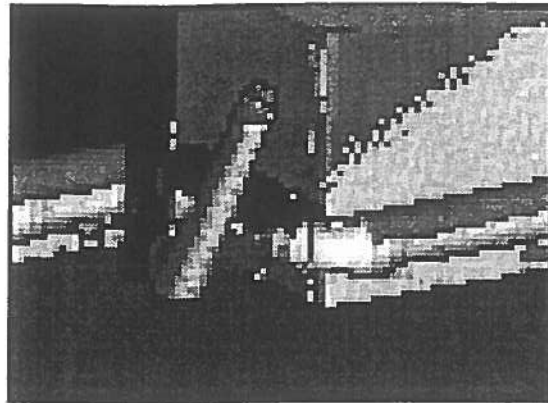
Fig. 4.: *Registration: initial pose estimate*

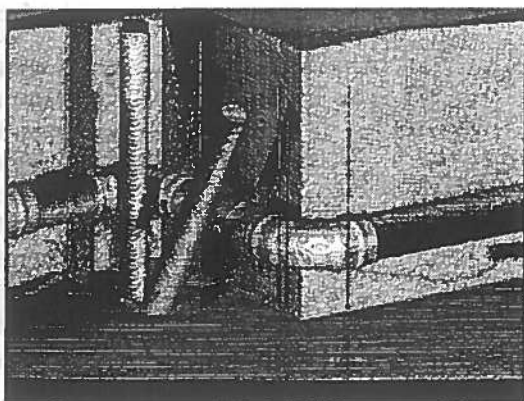Fig. 5.: *Final registration after ICP refinement*
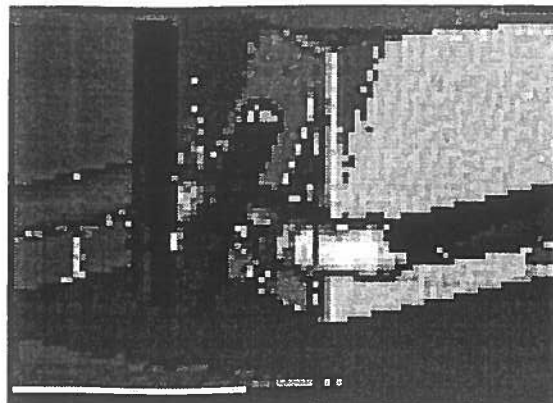

Fig. 6.: *Registration: initial pose estimate*

Fig. 7.: *Final registration after ICP refinement*

7 Current progress and conclusion

The research work presented in this paper deals with non-polyhedral modeling in a factory-like environment. In this paper, we have presented an approach for non-polyhedral landmark recognition and model update. This approach is decomposed in several steps. It first emphasizes the role played by simpler surface primitives (planes, cylinders, cones, spheres) in the scene description and data-model registration. At a scene description level, the surface classification procedure takes into account the uncertainties on the data; they are propagated through the canonicalization processes and used to compute the plausibilities that the patches obtained after segmentation belong to one of the simplified classes above. The corresponding geometric descriptions lead to simpler and stabler 3D invariants for the registration. An ICP like method can be used for pose refinement.

So far, we have used the segmentation, surface patch classification and registration methods presented above 1) on simulated range images from a synthetic model of a factory-like environment (including simulation of the sensory acquisition with given sensor characteristics, and simulation of noise) and 2) on real range images of small objects and of a factory mock-up acquired with a range sensor [19]. Successful results on real images have been provided.

The next steps of this work will further investigate the problem of questioning the model correctness.

References

1. N. Ayache and O. Faugeras. Maintaining representations of the environment of a mobile robot. In *Robotics Research : The Fourth International Symposium, Santa Cruz (USA)*, August 1987.
2. N. Ayache and F. Lustman. Trinocular stereovision. In *10th International Joint Conference on Artificial Intelligence (IJCAI), Milan (Italy)*, August 1987.
3. Y. Bard. Nonlinear parameter estimation. New York; Academic Press, 1974.
4. M. Bejanin, A. Huertas, G. Medioni and R. Nevatia. Model Validation for Change Detection. In *1994 ARPA Image Understanding Workshop, Monterrey, CA, USA*, 1994.
5. P.J. Besl and N.D. McKay A Method for Registration of 3-D Shapes. In *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 14, No 2*, February 1992.
6. J.L. Crowley and P. Stelmaszik. Measurement and integration of 3-D Structures By Tracking Edge Lines. In *1st European Conference on Computer Vision, Antibes (France)*, pages 269–280, 1990.
7. M. Devy and H. Bulata. Landmark-base vs Feature-based Localization of a mobile robot in a structured Environment. In *1st European Conference on Computer Vision, Antibes (France)*, pages 269–280, 1990.
8. P. Fillatreau & M. Devy. *Localization of an Autonomous Mobile Robot from 3D Depth Images using heterogeneous Features.* In IEEE International Workshop on Intelligent Robots and Systems (IROS '93), Yokohama (Japan), Juillet 1993.
9. P. Fillatreau. *Localisation et Modélisation tridimensionnelles pour un robot mobile autonome tout-terrain.* Thèse de l'Université Paul Sabatier, Toulouse (France), Rapport Interne N. 94108, Laboratoire d'Automatique et d'Analyse des Systèmes (C.N.R.S.), Janvier 1994.
10. R. B. Fisher: *SMS: A Suggestive Modeling System For Object Recognition*, Image and Vision Computing, Vol 5, No 2, pp 98 - 104, May 1987.
11. R. Fisher, E. Trucco, A. Fitzgibbon, M. Waite, M. Orr. *IMAGINE - A 3-D Vision System* Sensor Review, Vol 13, No. 2, pp 23–26, April 1993.
12. R. B. Fisher, A. Fitzgibbon, M. Waite, E. Trucco and M. Orr "Recognition of Complex 3-D Objects from Range Data", Progress in Image Analysis and Image Processing (Proc. 7th International Conference on Image Analysis and Processing), S. Impedovo (Ed.), Bari (Italy), September 1993, 509–606.
13. G. Golub and C. F. Van Loan "Matrix Computations", Second Edition. The John Hopkins University Press.
14. X. Lebegue and J. K. Aggarwal. Extraction and Interpretation of Semantically Significant Line Segments for a Mobile Robot. In *IEEE International Conference on Robotics and Automation, Nice, (France)*, May 1992.
15. A. Lorusso, D. W. Eggert and R. B. Fisher. A Comparison of Four Algorithms for Estimating 3-D Rigid Transformations. Research Report N. 737., Department of Artificial Intelligence, University of Edinburgh, 1994.
16. P. Moutarlier and R. G. Chatila. Stochastic Multisensory Data Fusion for Mobile Robot Location and Environment Modelling. In *Proc. International Symposium on Robotics Research, Tokyo*, 1989.
17. F. Nashashibi & M. Devy. *3D Incremental Modeling and Robot Localization in a Structured Environment using a Laser Range Finder.* In IEEE Transactions on Robotics and Automation, Atlanta (USA), Mai 1993.
18. V. Caglioti and F. Mainardi and M. Pilu and D.G. Sorrenti. *Improving Pose Estimation Using Image, Sensor and Model Uncertainty.* In Proc. British Machine Vision Conference BMVC94, York, September 1994.
19. E. Trucco, R. B. Fisher and A.W. Fitzgibbon. *Direct Calibration and Data Consistency in 3-D Laser Scanning*, Proc. British Machine Vision Conference BMVC94, York, September 1994.
20. E. Trucco and R.B. Fisher. *Computing Surface-Based Representations from Range Images*, Proc. IEEE International Symposium on Intelligent Control ISIC-92, Glasgow, August 1992.
21. Z. Zhang. Iterative Point Matching for Registration of Free-Form Curves and Surfaces. In *International Journal of Computer Vision, 13:2*, 1994.

This article was processed using the TEX macro package with SIRS95 style