Towards the recognition of 3D free-form objects

Christian Schutz, Heinz Hugli

University of Neuchatel, Institute of Microtechnology
Tivoli 28, CH-2003 Neuchatel, Switzerland
schutz@imt.unine.ch, hugli@imt.unine.ch

ABSTRACT

This paper investigates a new approach for the recognition of 3D objects of arbitrary shape. The proposed solution follows the principle of model-based recognition using geometric 3D models and geometric matching. It is an alternative to the classical segmentation and primitive extraction approach and provides a perspective to escape some of its difficulties to deal with free-form shapes. The heart of this new approach is a recently published iterative closest point matching algorithm, which is applied variously to a number of initial configurations. We examine methods to obtain successful matching. Our investigations refer to a recognition system used for the pose estimation of 3D industrial objects in automatic assembly, with objects obtained from range data. The recognition algorithm works directly on the 3D coordinates of the objects surface as measured by a range finder. This makes our system independent of assumptions on the objects geometry. Test and model objects are sets of 3D points to be compared with the iterative closest point matching algorithm. Substantially, we propose a set of rules to choose promising initial configurations for the iterative closest point matching; an appropriate quality measure which permits reliable decision; a method to represent the object surface in a way that improves computing time and matching quality. Examples demonstrate the feasibility of this approach to free-form recognition.

KEYWORDS: 3D vision, closest point matching, 3D free-form object recognition, pose estimation, range imaging

1. INTRODUCTION

The work described in this paper is carried out in the context of knowledge based 3D object recognition. We developed a hybrid recognition system, which combines range and intensity images to generate and verify object hypothesises[7]. One application of the vision system is to recognize known objects in a robot environment, which permits to update a virtual representation of a robot workspace[10]. The recognition of free-form 3D objects is one of the major problems in computer vision. We discuss a new approach for this task avoiding the classical segmentation and primitive extraction procedures.

The range images are acquired with a range finder working on the principle of space coding with projected stripe pattern and triangulation. Unlike intensity images, range images provide directly the intrinsic geometric information of the object. The separation of objects from the background, assumed to be plane, is performed easily in range images, since they do not contain texture. In this paper, we consider isolated objects, represented by the measured 3D points of the visible part of their surface.

In a previous work[4], we proceed according to the classical segmentation approach. The range image of the test object is first segmented into planar faces. These are then grouped and compared to a polyhedral model. However, segmentation of objects becomes much more difficult when dealing with free-form objects. As pointed out by other authors[6, 14], it is not clear, how object parts should be defined and how reliable segmentation may work.

To deal with free-from objects, we propose to use the measured 3D points themselves as a representation for objects of complex shape. No segmentation of the acquired data is needed and the surface represented by its 3D points is directly matched to the model. The model itself can be represented by any geometric primitive that allows an Euclidean distance measure. For example, models can be a set of points or polygonal models. They can be constructed experimentally by merging several point sets obtained from different range finder views or derived from CAD models. This allows us to treat objects of any form, since we are not limited to predefined geometric subparts and there is no effort needed to build a high level description of the model.

Christian Schutz, Heinz Hugli, "Towards the recognition of 3D free-form objects", Proc. SPIE, Int. Symposium on Intelligent Robots and Advanced Manufacturing, Intelligent Robots and Computer Vision XIV: Algorithms, Techniques Active Vision and Material Handling, Vol 2588, pp 476-484, Philadelphia, Oct. 1995

An interesting technique to registrate sets of 3D points, called iterative closest point algorithm (ICP), has been proposed by Besl[1]. As a typical geometric-based approach, the ICP algorithm needs a priori knowledge to give a rough estimate of the transformation between the test and model object. A good initial estimation of the transformation is crucial to convergence of the iterative algorithm, since its zone of convergence is limited.

Other researchers[2, 3, 5, 11, 12, 14] have used similar algorithms to track objects and register surfaces. In most of these systems the initial transformation, from where the iterative algorithm is launched, is entered by an operator or estimated, for example when tracking objects. As far as we know, the ICP algorithm has not yet been used to perform object recognition.

In a previous paper, we reported experiments to investigate the usefulness of the ICP algorithm to recognize very simple objects[13]. The presented results show that the translation between test and model is of minor influence to successful matching. On the contrary, rotation around one axis limits the convergence to a zone of about 80 degrees. Asymmetric objects give a slightly better performance, whereas subparts can only be matched for some specific initializations of ICP. Furthermore the algorithm converges quickly and allows subsampling of the point set representing the objects, which significantly reduces the computing time.

In this paper we present further experiments which now refer to more complex 3D objects. They show how the ICP algorithm may be successfully applied to object recognition. Section 2 presents rules to build a set of starting configurations for the ICP algorithm which helps to overcome its limited convergence and make even subpart matching feasible. In section 3, we propose a measure of matching quality best suited for the decision process. Finally, a method to subsample the cloud of points representing the objects, is introduced in section 4. The presented methods improve the recognition process in speed and quality. Successful application to real objects is shown in section 5.

2. RECOGNITION ALGORITHM

The heart of most recognition systems is an algorithm used to compare two objects and to deliver a measure of similarity. We use the mentioned ICP algorithm for this task. The algorithm searches first for every point of a test set, the point of a model set with the smallest Euclidean distance. These pairs of closest points between two surfaces to be matched are then used to calculate the translation and the rotation, which minimize the mean square distance or error. The test object is then translated and rotated by the resulting transformation. This procedure is applied several times until the error falls below a threshold or the number of iteration exceeds a predefined constant. Fig. 1 gives an overview of the basic working principle of ICP. The reader may find a complete description of the algorithm in the original publication[1] or in a previous paper of our research group[13].

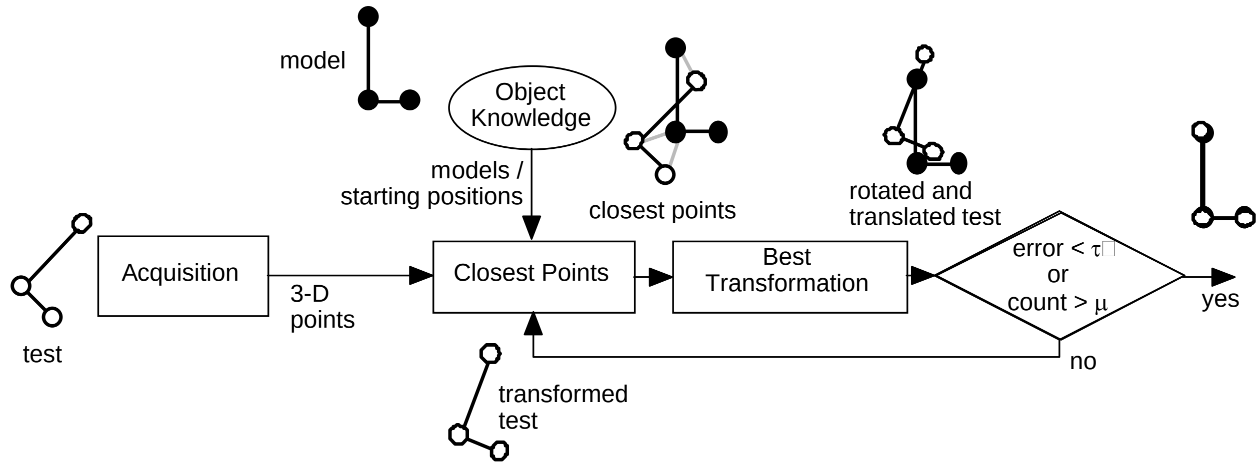


Fig. 1. Working principle of the iterative closest point algorithm (ICP)

The simplicity of this algorithm allows a fast implementation. Besides, it does not require any data pre-processing or local feature extraction, which makes it easily applicable to free-form objects and permits to handle reasonably well noise effects that occur in real range data. The algorithm is proved to converge to a minimum[1]; practically, this requires usually few iterations. There is no guarantee that this minimum is the global one. Most of the computing time of the algorithm is spent in the closest point search. We have implemented a fast closest point search algorithm using k-D tree[14], which helps to reduce the computing cost significantly[8]. The ICP algorithm has a great flexibility to use models of different kinds, since only the closest point search routine has to be adapted. Our closest point search implementation can deal with models represented by clouds of points or polygons, where the appropriate function is selected automatically.

In a 3D object recognition system, the matching algorithm used to compare a test with the models in the database should allow the test to be any view of the corresponding model placed in any pose. Since the ICP algorithm converges only for a limited set of transformations between a test and its model, it has to be adapted to recognize reliably subparts of 3D objects. We propose a set of appropriatly chosen starting configurations, so that at least one of them leads towards the global minimum.

First, given a view axis defined by the camera position and the center of mass of the test, we place the model behind the test, as shown in Fig. 2. It is important that the two objects are not too far away from each other to avoid unstable point coupling in the first iteration of ICP. This happens when all points of the test are coupled with only one point in the model and may result in a bad rotation during distance minimization. Practically, we choose the maximal model radius as distance between both centers of mass. This placement ensures that the test surface not visible from the camera always faces the model and also excludes that test surfaces are compared with invisible model surfaces.

Second, we select a number of view points distributed uniformly on the sphere, circumscribing the model as drawn in Fig. 2. The model is now oriented so that every view point is lying once on the view axis. Furthermore, the model is rotated stepwise around the view axis for each of these configurations. The so defined starting configurations for the ICP algorithm are selected in such a way, that their convergence zones cover the whole space around the model. Starting configurations with promising matching are used in further tests using smaller rotation steps when turning the model. This use of different starting configurations will ensure that the matching converges at least once towards a successful matching.

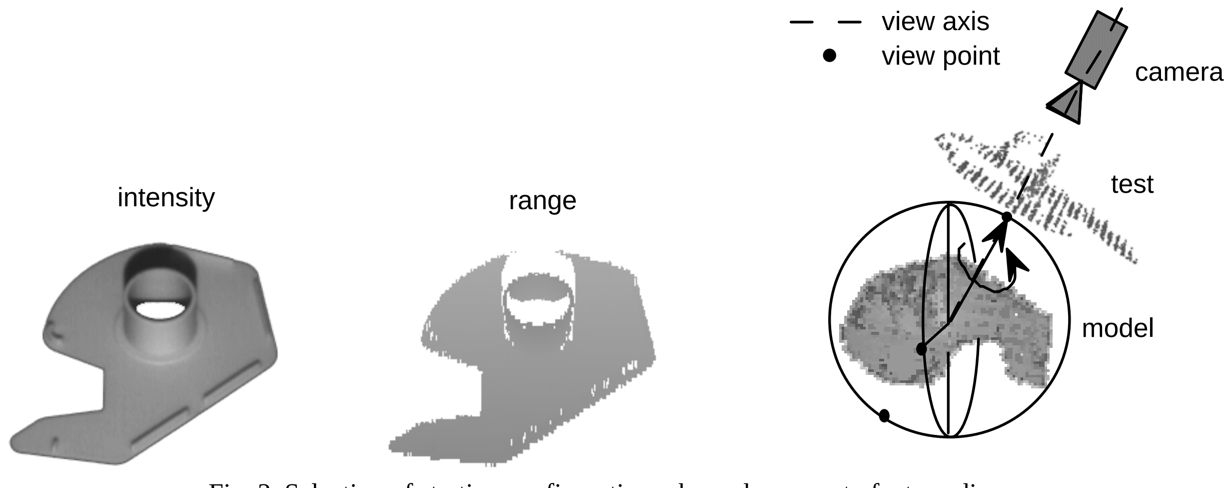


Fig. 2. Selection of starting configurations showed on a part of a tape dispenser

At a first glance the necessity of multiple starting configurations seems to introduce much overhead. But since the convergence zone is relatively large, the number of starting configurations can be kept low. Furthermore, the algorithm converges quickly and allows an estimation of the quality of a starting configuration after few iterations of ICP. This allows to prune the search tree. Starting configurations of high quality are then used to do more iterations in order to find a final matching. The definition of matching quality will be presented in the next section.

If the database consists of several models, the procedure described above has to be applied to every model. Since ICP converges fast, there is no need to trace the whole search tree for every model. Wrong models may be excluded early.

Other researchers[3, 5] proposed to use the curvature during the closest point search to match similar points and therefore enlarge the convergence zone. But even so, multiple starting configurations are needed. We prefer to use a relatively large amount of well chosen starting configurations without the overhead of calculating curvatures, which needs sophisticated filtering of the range image.

3. MATCHING QUALITY

A successful recognition system needs a good error measurement which reflects the quality of the matching. In our recognition approach, we start the ICP algorithm from different configurations doing only few iterations to gain speed. Then only best configurations will be selected and used for further iterations to obtain final matching. This decision is based on the matching quality. Since ICP minimizes the mean square distance between the points of two objects, it seems obvious to use the minimized mean square distance as decision criteria. But experiments showed that this criteria is insufficient to discriminate the quality of starting configurations. For example the mean square distance for the two cases shown in Fig. 3 differs by only 20%, which does not reflect very much the large difference between the two cases where the gray model object is in two completely different configurations.

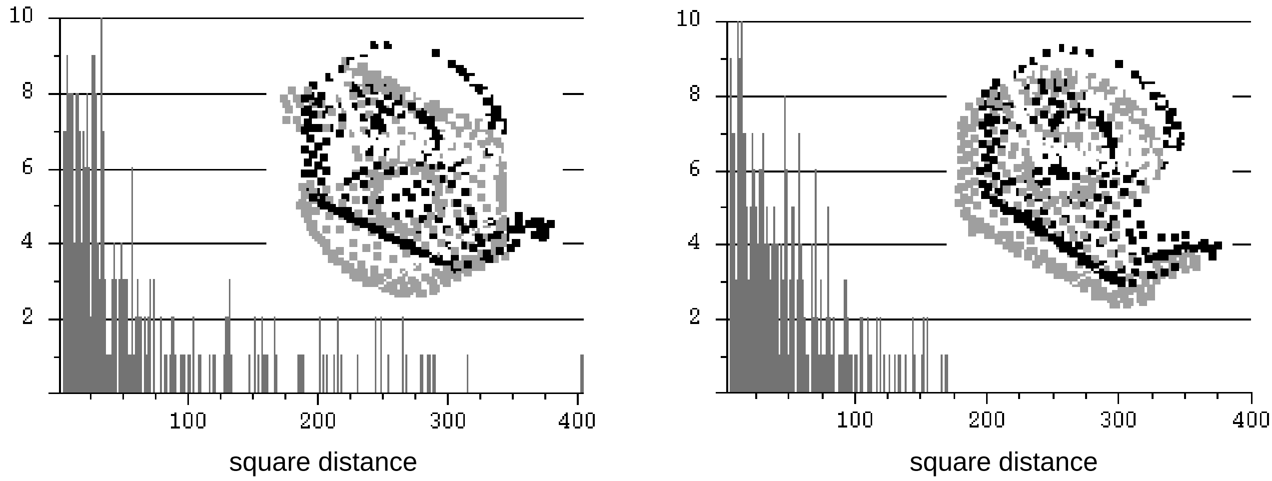


Fig. 3. Histogram of square distance between closest points for a bad and a good case

The square distance histograms corresponding to the two matches show that the distributions of the square distances differ even for similar means of square distance. Zhang proposed to include the deviation of the distances to qualify a matching[14]. We calculate statistics for the square distances to gain speed by omitting the root calculation. In fact the deviation of the square distances is twice as large for the bad case compared to the good one shown in Fig. 3. So, we finally define the matching error as the sum of the mean and the deviation of the square distances as calculated in Eq. 1. Cases with low error will indicate a promising matching and may be used for further iterations.

$$
\begin{aligned}
& d_i = \left\| x_i - p_i \right\|, \quad x_i \in test, \quad p_i \in model \\
& \mu = \frac{1}{N} \sum_{i=1}^{N} d_i^{\,2} \\
& \sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} \left(d_i^{\,2} - \mu \right)^2} \\
& error = \mu + \sigma
\end{aligned}
\tag{1}
$$

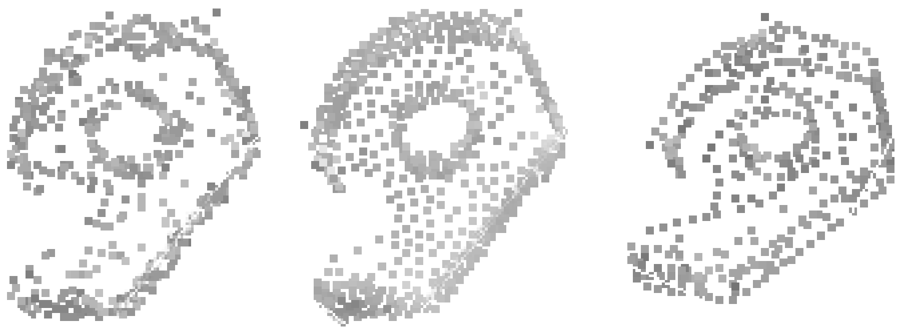
4. OBJECT REPRESENTATION

The execution time of the ICP algorithm depends largely on the number of points representing the test and model objects and there is a real interest to keep this number as low as possible. We investigate therefore three methods for point number reduction.

The first approach is to select linearly every i-th point in the set. This method is straightforward and fast, but since points are not distributed uniformly over the objects surface obtained from the range finder, zones with more acquisition points have a larger weight when calculating the best transformation in ICP. A possible solution to this problem would be the introduction of weighted couplings for the closest points when calculating the best transformation. A smaller weight could be associated to points with less neighbors. However, this neighborhood inspection is quite complex and outliers may obtain too much influence during minimization.

A better method should reduce the number of points while preserving a more or less uniform point distribution over the object surface. Therefore we propose an iterative point grouping algorithm. We start with a random point out of the object set and mark all its neighbors within a radius R. The marked points are then replaced by their center of mass in the object set. Again, we select randomly an yet untreated point and do the same grouping. This procedure is repeated until no more points are left. If an object surface has been sampled densely enough by the range finder, the points resulting from the grouping have a distance of about 2R, where R is the neighborhood radius. Object details may be lost, if R is not small enough. Note that the test and model objects have been grouped using the same neighborhood radius R. Their similar point distribution results in better matching. The neighborhood search may be implemented very efficiently using kD-tree.

A third method to reduce the number of points, is the selection of the points lying on borders. We use a range image segmentation algorithm[9] to extract the points with high curvature. Fig. 4 shows the resulting representations for an object, a tape dispenser part, when applying the three approaches, called linear reduction, 3D grouping and border selection.



linear reduction 3D grouping border selection

Fig. 4. Three methods to reduce the number of points

Linear reduction and 3D grouping reduced the amount of points to about 600, whereas the selection of points with high curvature yields a subset of 400 points. Considering the two objects with the same amount of points, the point distribution is much more uniform for the one treated by the 3D grouping. This has significant impacts on the decision quality during recognition. Next figure Fig. 5 shows the final error for the matching of objects, reduced by the three methods described above.

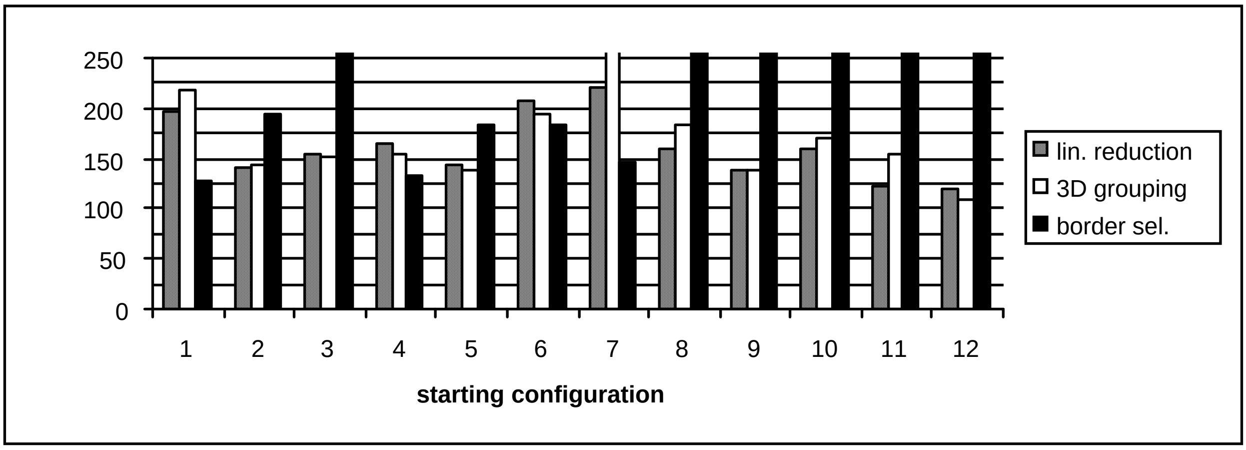


Fig. 5. Matching error for different point reduction algorithms

As explained in section 2, the ICP algorithm is launched from different starting configurations. Search paths starting at configurations with bad quality are pruned and more iterations are done at the best configurations. This implies that the best configuration is reliably detectable. We observe in Fig. 5 that the best discrimination is obtained with objects reduced in number of points by 3D grouping. Indeed, with this method the difference between the most promising configuration (12) and the second best (5) is much greater, than the difference corresponding to configurations (12, 11) obtained with linear reduction. The selection of border points is not stable enough and the ICP algorithm does not converge towards the correct solution (12).

So far, we considered model and test represented by set of points. An alternative approach uses polygonal object models. Polygonal models as shown in Fig. 6 may reduce the complexity of the closest point algorithm even more, since the same object may be described with less than hundred polygons. Although there are less parts representing a model, the calculation of the closest point becomes more complicated.

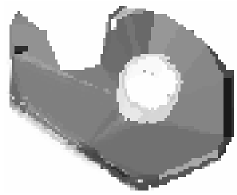


Fig. 6. Polygonal model for a tape dispenser part

5. RESULTS

As stated in the introduction, our recognition system is used in a robot environment. In an experimental configuration, the robot has to assemble a tape dispenser by plugging the tape into the base part and then closing it with the cover part. Therefore, our model database consists of the three parts to be assembled. A typical robot workspace is shown in Fig. 7, where the background has been eliminated for visualization convenience. Derived from the range image, the z-image gives the height above the working table for every image point. It is used to separate the objects, assumed not to touch each another. We simply set a threshold at the workplace height and obtain a binary image with the object zones (see Fig. 8). With this method we can extract objects even if the background has a complex texture, since texture does not appear in the z-image any more.

Every extracted test object is matched with all models. We start therefore the ICP algorithm for every model applying the rules defined in section 2. We select six view points uniformly distributed on the model sphere and do four rotations of 90 degrees for every view point aligned to the view axis. The ICP algorithm is expected to converge for at least one of these 24 configurations since its convergence zone is about 80 degrees. Recognition reliability and speed can be changed using more or less view points.

The best model together with its two best configurations are selected. Several new configurations near to the two selected ones are used to start the ICP algorithm again calculating more iterations. Finally, we decide for the best configuration and do a final matching performing more than 20 iterations. The recognized model is then placed in the scene and projected on the intensity image to verify the result, as presented in Fig. 8 and Fig. 9.

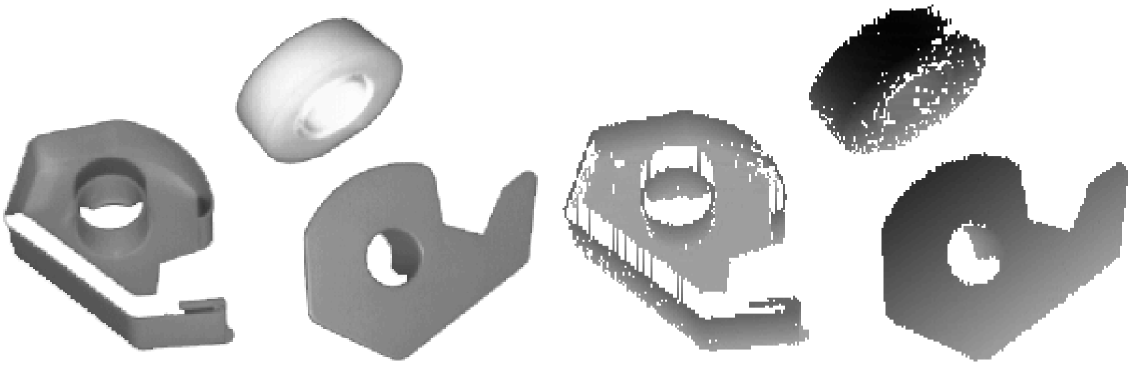


Fig. 7. Intensity and z-image of parts to be assembled

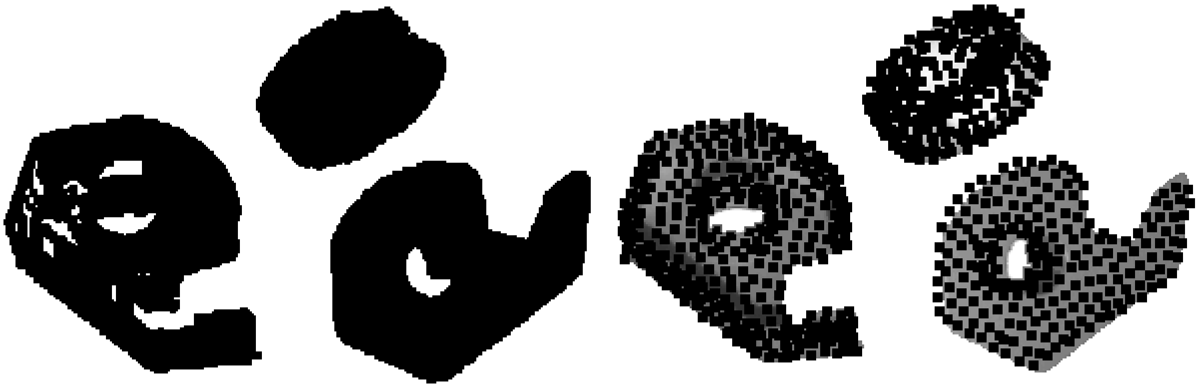


Fig. 8. Thresholded z-image and recognized models superposed to the intensity image

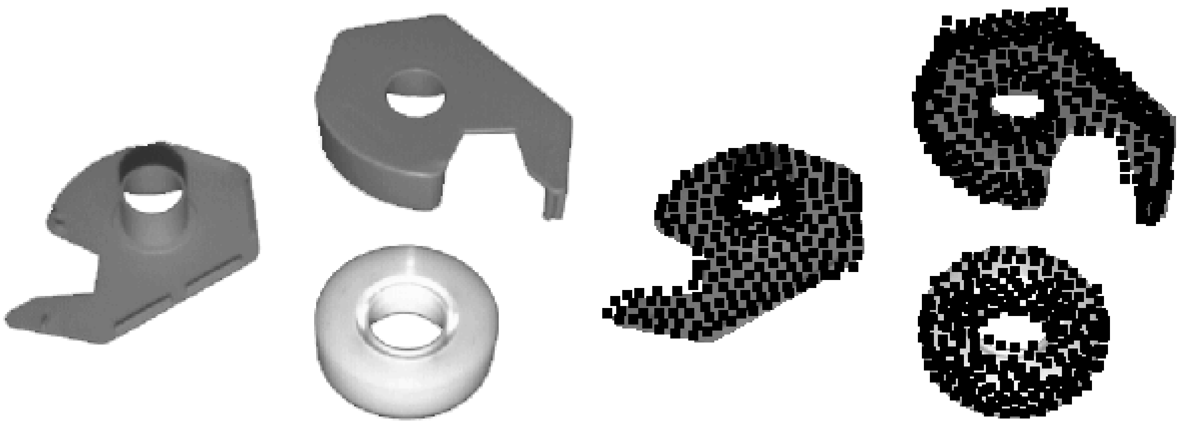


Fig 9. Recognition results for an other scene

6. CONCLUSIONS

The presented work is a contribution to a 3D object recognition approach which is easily applicable to free-from objects. The approach is based on geometric matching and applies to objects represented by sets of points or polygonal models. It differs from the classical approach which requires object segmentation and model construction in terms of geometric primitives.

Using the ICP algorithm at the heart of the recognition, we proposed a number of methods and techniques to extend its use to the recognition of 3D objects obtained from range images.

Since the ICP algorithm is not directly applicable to object recognition because of a limited convergence zone, we proposed a set of starting configurations using the knowledge of the camera position to overcome this handicap and to allow object or subpart matching.

We showed that the integration of the square distance deviation in the quality measure helps to extract reliably promising starting configurations for further observation. This pruning of the search tree at an early stage of the recognition procedure allows a concentration of the matching effort to the interesting configurations.

Aiming at lowering the computation cost, we considered and compared three methods for reducing the size of the point set representing the objects. Best results are obtained with the 3D point grouping algorithm.

Presented results show the successful recognition of the three parts of a tape dispenser and demonstrate the feasibility of the approach. Its intrinsic flexibility makes the approach applicable to any object form.

In the future, we will address possible limitations of the proposed approach which may arise when dealing with large number of models and having occlusion in the scene.

7. ACKNOWLEDGMENTS

This research has been funded by the Swiss national Priority Program in Informatics, Knowledge-based systems, under project number 5003-344336.

8. REFERENCES

1. P. Besl and N. McKay. "A method for registration of 3-D shapes," *IEEE Transactions On Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239-256, 1992.
2. Y. Chen and G. Medioni. "Object modelling by registration of multiple range images," *Image and Vision Computing,* Vol. 10, No. 3, pp. 145-155, 1992.
3. J. Feldmar and N. Ayache. "Rigid and affine registration of smooth surfaces using differential properties," *Lecture Notes in Computer Science*, Vol. 801, No. 2, pp. 397-406, 1994.
4. P. Gingins and H. Hügli. "Model-based 3D object recognition by a hybrid hypothesis generation and verification approach," *Intelligent Robots and Computer Vision XIII: Algorithms and Computer Vision*, Boston, SPIE 2353, pp. 67-77, 1994.
5. G. Godin and P. Boulanger. "Range image registration through viewpoint invariant computation of curvature," *From Pixels to Sequences, Proc. conf. ISPRS,* Zurich, Vol. 30, pp. 170-175, 1995.
6. M. Hebert, J. Ponce, T. E. Boult, A. Gross and D. Forsyth. "Representation for computer vision," *Report of the NSF/ARPA workshop on 3D object*, CUNY Graduate School and University Center, 1994.
7. H. Hügli, Ch. Schütz, Ch. Baur, E. Natonek. "Knowledge-based 3-D vision system," *Proceedings of Information Conference Module 2,* Knowledge based systems, Swiss National Science Fundation, pp. 46-53, 1994b.
8. H. Hügli, Ch. Schütz, D. Semitekos. "Geometric matching for free-form 3D object recognition," to be published in: *ACCV'95*, Singapore, 1995.
9. G. Maitre. *Segmentation et traitement preliminaire des images de profondeur*. These de doctorat a l'Universite de Neuchatel, 1994.

10. E. Natonek, Th. Zimmermann, L. Flückiger. "Model based vision as feedback for virtual reality robotics environments," *IEEE Virtual Reality Annual Symposium*, North Carolina, 1995.

11. G. Turk and M. Levoy. "Zippered polygon meshes from range images," *Computer Graphics*, ACM Siggraph 94 Proc, Orlando, Fla., pp. 311-318, 1994.

12. D. A. Simon, M. Hebert and T. Kanade. "Techniques for Fast and Accurate Intra-Surgical Registration," *Journal of Image Guided Surgery*, Vol. 1, No. 1, pp.17-29, 1995.

13. Ch. Schütz and H. Hügli. "Recognition of 3-D objects with a closest point matching algorithm," *From Pixels to Sequences*, Proc. conf. ISPRS, Zurich, Vol. 30, pp. 128-133, 1995.

14. Z. Zhang. "Iterative point matching for registration of free-form curves and surfaces," *International Journal of Computer Vision*, Vol. 13, No.2, pp. 119-152, 1994.