ACTIVE 3D RECOGNITION SYSTEM BASED ON FOURIER DESCRIPTORS

E. González, V. Feliú, A. Adán* and Luis Sánchez**
H.T.S. of Industrial Engineering, University of Castilla La Mancha, Ave. Camilo José Cela s/n
Ciudad Real, 13071, Spain

**Higher School of Informatics,University of Castilla La Mancha, Ronda de Calatrava 5*
Ciudad Real, 13071, Spain

***U.E of Technical Industrial Engineering,University of Castilla La Mancha, Ave Carlos III, s/ n*
Toledo, 45071, Spain

Keywords: Active recognition system, next best view, silhouette shape, Fourier descriptors.

Abstract: This paper presents a new 3D object recognition/pose strategy based on reduced sets of Fourier descriptors on silhouettes. The method consists of two parts. First, an off-line process calculates and stores a clustered Fourier descriptors database corresponding to the silhouettes of the synthetic model of the object viewed from multiple viewpoints. Next, an on-line process solves the recognition/pose problem for an object that is sensed by a real camera placed at the end of a robotic arm. The method avoids ambiguity problems (object symmetries or similar projections belonging to different objects) and erroneous results by taking additional views which are selected through an original next best view (NBV) algorithm. The method provides, in very reduced computation time, the object identification and pose of the object. A validation test of this method has been carried out in our lab yielding excellent results.

1 INTRODUCTION

Most computer vision systems used in robotics environments perform 3D object recognition tasks using a single view of the scene (Bustos et al., 2005). Commonly, a set of features is extracted and matched with features belonging to an object database. This is why so many researchers focus their recognition strategy on finding features which are capable of discriminating objects efficiently (Helmer and Lowe, 2004). However, these approaches may fail in many circumstances due to the fact that a single 2D image may be insufficient. For instance, this happens when there are objects that are very similar from certain viewpoints in the database (ambiguous objects); a difficulty that is compounded when we have large object databases (Deinzer et al., 2003).

A well known strategy that solves the ambiguity problem is based on using multiple views of the object. Active recognition systems provide the framework to efficiently collect views until the sufficient level of information for developing the identification and posing estimation tasks is obtained (Niku, 2001).

Previous works on active recognition differ in the way they represent objects, the way they combine information and the way of they plan the next observation (Roy et al., 2004). These systems use 3D representation schemes based on the object geometric model or on the object appearance. Although X recognition based on geometric models might be potentially more effective and allow for the identification of objects in any position, they raise important problems of practical aplicability. Moreover, the methods based on X appearance are currently the most successful approaches for dealing with 3D recognition of arbitrary objects.

Many strategies for solving the 3D object recognition problem using multiple views have been proposed: an aspect graph is used in Hutchinson and Kak (1992) to represent the objects. This criterion handles a set of current hypotheses about the object identity and position. It characterizes the recognition ambiguity by an entropy measure (Dempster-Shafer theory) and evaluates the next best sensing operation by minimizing this ambiguity. Borotsching et al.

González E., Feliú V., Adán A. and Sánchez L. (2007).
ACTIVE 3D RECOGNITION SYSTEM BASED ON FOURIER DESCRIPTORS.
In *Proceedings of the Fourth International Conference on Informatics in Control, Automation and Robotics*, pages 318-325
DOI: 10.5220/0001627003180325

(1999) represent the objects by some appearance-based information, namely the parametric eigenspace. This representation is augmented by adding some probability distributions. These probability distributions are then used to provide a gauge for performing the view planning. Sipe and Casasent (2002) use a probabilistic extension of the feature space trajectory (FST) in a global eigenspace to represent 3D views of an object. View planning is accomplished by determining - for each pair of objects – the most discriminating view point in an off-line training stage. Their approach assumes that the cost of making a mistake is higher than the cost of moving the sensor.

In general, most of these approaches solve the 3D object recognition problem using stochastic or probabilistic models and, consequently, they require a large dataset for training (Deinzer et al., 2006). Here we present a different way to focus on the problem.

The key to our active recognition system consists of using a reduced set of Fourier descriptors to connect and develop the recognition phases: object representation, classification, identification, pose estimation and next best view planning.

We focus the object representation on silhouettes because: they can be robustly extracted from images, they are insensitive to surface feature variations - such as color and texture - and, finally, they easily encode the shape information (Pope et al. 2005). The most popular methods for 2D object recognition from silhouettes are based on invariant moments or Fourier descriptors. Invariant moments exhibit the drawback that two completely different silhouettes may have the same low order invariant moments, which may lead to ambiguities in the recognition process. Fourier descriptors yield much more information about the silhouette, and only similar silhouettes exhibit similar Fourier descriptors. Since we consider the objects to be non-occluded and the background to be uncluttered, we use a representation scheme in which the silhouettes from different viewpoints are represented by their Fourier descriptors.

This paper is organized as follows. Section 2 presents an overview of the method. Section 3 describes our object identification/pose estimation approach. Section 4 details the next best view method. Section 5 shows the performance of our method by carrying out experiments on a real platform, and some conclusions are stated in Section 6.

2 OVERVIEW OF THE METHOD

In this method the scene silhouette (silhouette of the 3D object to be recognized) is recognized among a set of silhouettes (in our case, 80 or 320 per object) of a group of objects through an algorithm based on Fourier descriptors. Therefore, X recognition of the silhouette of the scene involves both object identification and pose. The method consists of off-line and on-line parts.

The off-line process consists of building a structured database of silhouettes belonging to a generic set of objects. Firstly, a high precision three-dimensional model of each object is obtained by means of a laser scanner sensor. Next, this model is viewed from a set of homogeneous viewpoints obtaining the corresponding set of 2D silhouettes.

The viewpoints correspond to the vertexes of a tessellated sphere with origin in the centre of mass of the object. Figure 1 shows an object model inside the tessellated sphere, the projected image of the model and its silhouette from a specific viewpoint.

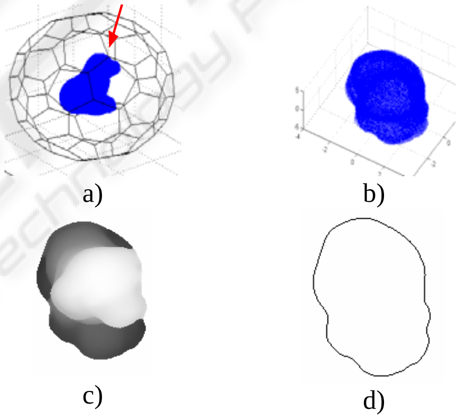
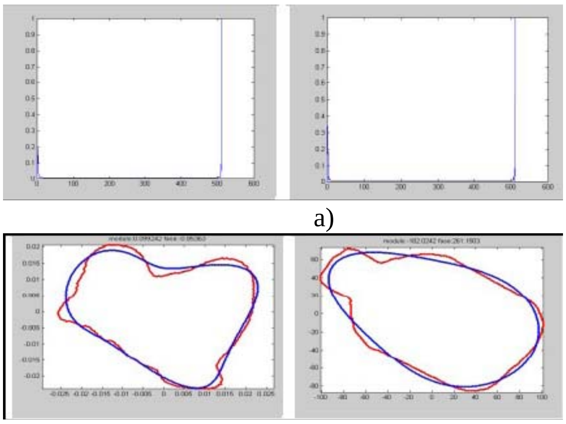


Figure 1: a) Object model put into the tessellated sphere b) View of the object model from a specific viewpoint, c) Depth image d) 2D silhouette.

The database is structured in clusters using only three Fourier descriptors. To build the clustering we used a k-means algorithm (Netanyahu et al., 2002). This strategy allows us to split the silhouette search space in zones where the silhouettes are roughly similar. Consequently, the cost of the recognition process is dramatically reduced. Figure 2 a) shows the most important Fourier descriptors modules for a couple of silhouettes. In our case, we have taken the second, third and next to the last values. Figure 2 b) presents the reconstructed silhouette with the three Fourier descriptors superimposed on the original one. Note that by selecting the most meaningful Fourier components it is possible to work with approximate shapes. Figure 3 shows a spatial representation of the clusters that have been extracted in our database.

The on-line process is designed to solve the recognition/pose problem of an object that is viewed by a camera in a real environment. The essential steps are: Fourier descriptor calculation, classification (discrimination) process, identification/pose calculation and next view algorithm. Next, a brief explanation of these steps is provided.


a)

b)

Figure 2: a) Fourier descriptors modules b) silhouette (red) and silhouette recovered with three Fourier descriptors (blue).

To calculate the Fourier descriptors a suitable image preprocessing is carried out on the original image. Specifically this process consists of filtering, thresholding and contour extraction. Next the points of the contour are taken as a sequence of complex numbers and the Fourier descriptors are finally computed.

The discrimination phase classifies the silhouette of the scene into a single or a set of clusters. The selected clusters constitute the work sub-space in the pose phase. Formally,

Let $O_1, O_2, ...O_N$ a database of N objects, C_k the *kth* cluster, $k \in [1..K]$, S_k^{nm} the *n-th* silhouette of the object m, p_k the *k-th* cluster prototype, D the Euclidean distance, R_k the *k-th* cluster radius where $R_k = \max D(p_k, S_k^{nm})$ and z the silhouette of the scene to be matched. The subspace S_{sub} will be formed by the clusters, which verify one or both of the following conditions:

Criterion 1: If $D(p_k, z) < R_k$ *then* $C_k \in S_{sub}$

Criterion 2:
If $\cdot| D_i - \min D_k | < \varepsilon$ then $C_i \in S_{sub}, i \in [1..K]$

The criterion 1 is satisfied for cases where z is inside a cluster whereas criterion 2 corresponds to cases where the silhouette z is inside an area with very high cluster density or where the scene silhouette falls outside the clusters. Thus, the discrimination process sets a work subspace S_{sub} with a reduced database of silhouettes.

The identification phase, which is carried out in S_{sub}, yields, in general, a reduced set of candidate silhouettes. The reason for taking only a few candidates is as follows. Matching and alignment techniques based on contour representations are usually effective in 2D environments. Nevertheless, in 3D environments these techniques have serious limitations. The main problems with the contour based techniques occur due to the fact that the information on the silhouettes may be insufficient and ambiguous. Thus, similar silhouettes might correspond to different objects from different viewpoints. Consequently, a representation based on the object contour may be ambiguous, especially when occlusion circumstances occur.

In essence, the identification phase compares the silhouette from the scene with the silhouettes in the subspace S_{sub} by means a quadratic error minimization applied on the modulus of the Fourier descriptors. If the identification/pose process proposes more than one candidate silhouette, then the solution is ambiguous and it is necessary to apply the Next Best View planning method (NBV). Figure 4 shows a scheme with the main process.

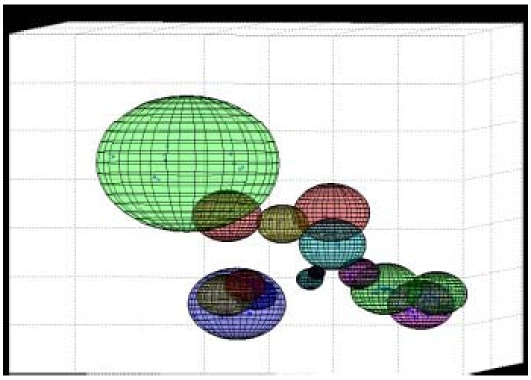


Figure 3: Spatial representation of the silhouette clusters.

In most cases, the recognition and pose estimation phase is finished after several views of the object are taken and only one candidate is found. In this process, the position of the next view is calculated through an algorithm based on the set of candidate silhouettes obtained in the previous view. This will be explained in Section IV. Figure 4 shows a scheme of the main process.

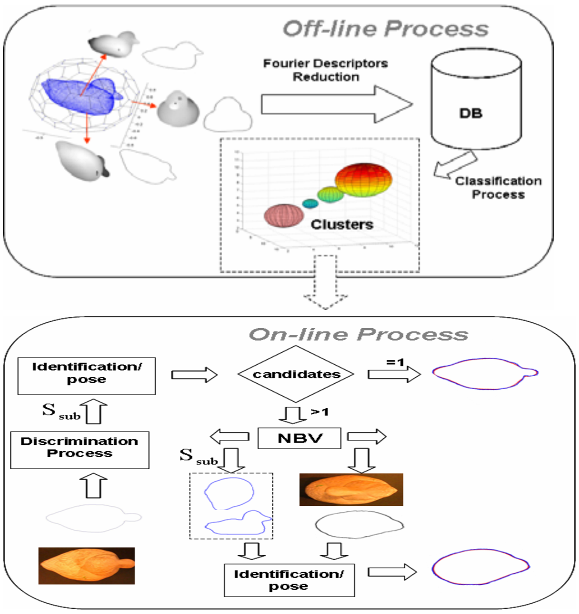


Figure 4: Diagram of the active recognition system.

3 OBJECT RECOGNITION AND POSE ESTIMATION PROCEDURE

Fourier descriptors can be used to represent closed lines (silhouettes). They can be made invariant to translations and rotations, and allow easy filtering of image noise (Deinzer et al., 2003). Assume a contour *l(n)* composed of N points on the XY plane:

$$l(n) = [x(n), y(n)], \quad n = 0..N-1 \tag{1}$$

where the origin of index n is an arbitrary point of the curve, and n and n+1 are consecutive points according to a given direction (for example clockwise direction) over the silhouette. Assume also that points over the curve have been regularized in the sense that two consecutive points are always at the same Euclidean distance. Let us define the complex sequence and its discrete Fourier transforms Z(k)=F(z(n))as:

$$z(n) = x(n) + jy(n) \tag{2}$$

$$Z(k) = \sum_{n=0}^{N-1} z(n)\exp(-j2\pi kn/N) \quad 0 \le k \le N-1 \tag{3}$$

Assume also a data base of *R* silhouettes $s_r(n)$, $0 \le n \le N_r - 1 \text{with} 1 \le r \le R$, whose respective discrete Fourier transforms are $S_r(k)$.

A critical aspect of our method is its computation speed because we want to recognize objects in real time. Then the FFT algorithm is used to obtain the Fourier descriptors. Then N must be power of 2 and both the scene silhouette and the silhouettes of the data base must be regularized to a number of $N = 2^{N'}$ points.

The basic problem to be solved in our method is to match the scene silhouette *z(n)* to some silhouette $s_{r^*}(n)$ of the data base, under the assumptions that *z(n)* may be scaled (λ), translated ($c_x + jc_y$) and rotated (φ) with respect to the best matching silhouette of the data base, and that the reference point on *z(n)* (n = 0) may be different from the reference point of that data base silhouette (we denote that displacement δ). The next section deals with selecting the silhouettes and obtaining X *c*, δ, φ, λ.

3.1 Close Silhouettes Selection

Suppose that *z* is the silhouette of an object captured by the camera and that it corresponds to the silhouette $s_{r^*}(n)$ that belongs to the silhouette database. In general, *z* is matched to $s_{r^*}(n)$ after displacement, rotation, scaling and centre translation parameters are found. In general for $s_r(n)$, in the space domain:

$$z(n) = D(s_r(n), \delta)\lambda \exp(j\varphi) + c \tag{4}$$

where $D(s_r(n))$ displaces δ units the origin of the sequences $s_r(n)$. Taking Fourier transform:

$$Z(k) = \lambda \exp(j\varphi)\exp(j2\pi k\delta / N)S_r(k) + cN \tag{5}$$

- Translation: Since all silhouettes have the coordinate origin at their centre of mass $S_r(0) = 0$ and from expression (5) , $c = Z(0)/N$.
- Close silhouettes identification.: Defining $\widehat{Z}(k) = Z(k) - cN$, the modulus of expression (5) holds:

$$\left|\widehat{Z}(k)\right| = \lambda\left|S_r(k)\right| \tag{6}$$

The matching procedure minimizes the mean squared error between the Fourier descriptors of the scene silhouette and the silhouettes of the database. Given a pair of silhouettes $(z(n), s_r(n))$, the similarity index J_r is defined as:

$$J_r(\lambda) = (\left|\widehat{Z}\right| - \lambda\left|S_r\right|)^t (\left|\widehat{Z}\right| - \lambda\left|S_r\right|) \tag{7}$$

where $\left|\widehat{Z}\right| = (|\hat{Z}(0)|, \ldots |\hat{Z}(N-1|)^t$, $|S_r| = (|S_r(0)| \ldots |S_r(N-1|)^t$,| . |=absolute value.

Minimizing $J_r(\lambda)$ with respect to the scaling factor λ, we obtain:

$$\lambda_r^\circ = \frac{|\widehat{Z}|^t |S_r|}{|S_r|^t |S_r|} \quad , \; J_r^\circ = |\widehat{Z}|^t |\widehat{Z}| - \frac{(|\widehat{Z}|^t |S_r|)^2}{|S_r|^t |S_r|} \qquad (8)$$

After calculating J_r° for all silhouettes of the data base we select the silhouettes which verify $J_r^\circ \le U$, U being a specific threshold. In this case, we have a set of many ambiguous silhouettes $\{\mathrm{S}_{c1}, \mathrm{S}_{c2}, \ldots \mathrm{S}_{cf}\}$ and it is necessary to select another viewpoint to solve the ambiguity problem (NBV section).

3.2 Pose Calculation

Let us denote $L = (r_1, r_2, \ldots, r_m)$ the set of indexes of the candidate silhouettes. In order to select the best candidate among L candidates, a more accurate procedure is carried out. This procedure uses the complete complex Fourier descriptors (not only the modules as in the previous process). As a result of this process, a new similarity index f° is obtained and the pose parameters λ, φ, δ are calculated.

The cost function to be minimized is (see (4)):

$$f\ (\lambda,\theta,\delta) = (\hat{Z} - q\,P\,(\delta))^{t^*}\ (\hat{Z} - q\,P\,(\delta)) \qquad (9)$$

where

$\hat{Z} = (\hat{Z}(0), \ldots, \hat{Z}(N-1))^t$,

$P_r(\delta) = (s_r(0), s_r(1)\exp(j2\pi\delta / N), ,\ldots, s_r(N-1)$,

$\exp(j2\pi\delta(N-1)/N))^t$

$*$ denotes conjugate, t^* denotes transpose conjugate, and r is restricted now to set L

Let us denote $\lambda \exp(j\varphi) = q$, optimizing (9) respect to the complex number q :

$$q^\circ(\delta) = \frac{Z^t\ P^*(\delta)}{S^{t^*}\ S}, \; f^\circ(\delta) = Z^{t^*}\ Z - \frac{\left|Z^{t^*}\ P(\delta)\right|^2}{S^{t^*}\ S} \qquad (10)$$

(notice that $P_r^{t^*}(\delta)\ P_r(\delta) = S_r^{t^*} \cdot S_r$).

Taking into account that $0 \le \delta \le N-1$ is integer, the right expression of (10) is calculated for all possible values of δ and $f_r^\circ =_\delta^{\min} f_r^\circ(\delta)$ is determined.

Then f_r° is the similarity index of the silhouette r in the fine matching process, δ_r° is the corresponding displacement and q_r° is obtained from ¿the? left equation of (10) particularized to δ_r° . Rotation and scaling are estimated from q_r° :

$$\lambda_r^\circ = |q_r^\circ|\ ;\ \varphi_r^\circ = \angle q_r^\circ \qquad (11)$$

4 NEXT BEST VIEW PLANNING

The goal of this phase is to provide a solution to the ambiguity problem by taking a set of optimal viewpoints. When an ambiguous case occurs, we move the camera to another viewpoint from which the silhouettes of the candidate objects are theoretically very dissimilar.

As said before, in our scheme representation we associate each silhouette stored in the database with a viewpoint of the tessellated sphere. Then, the first step in the NBV consists of aligning the candidate spheres (corresponding to the viewpoints in our models) with the scene sphere.

Let *T_R(S)* be the tessellated sphere, *N'_{Rx}* the camera position and *N_{R1}* the viewpoint that corresponds to the candidate silhouette S_{ci} . To align the two spheres a rotation must be applied to make *N'_{Rx}* and *N_{R1}* coincident (Adán et al., 2000). Formally:

Let $\vec{\mathbf{u}}(u_x, u_y, u_z) = \dfrac{\overrightarrow{ON}_{R1} \times \overrightarrow{ON'}_{Rx}}{\left|\overrightarrow{ON}_{R1} \times \overrightarrow{ON'}_{Rx}\right|}$ be the normal vector to the plane defined by $\overrightarrow{ON}_{R1}$ and $\overrightarrow{ON'}_{Rx}$, O being the center of T_I. Let θ be the angle between the last two vectors. Then, a rotation θ around the u axis can first be applied to $T_R(S)$. This spatial transformation is defined by the following rotation matrix $\mathbf{R_u}(\theta)$:

$$\mathbf{R_u}(\theta) = \begin{pmatrix} u_x u_x(1-c\theta)+c\theta & u_x u_y(1-c\theta)-u_z s\theta & u_x u_z(1-c\theta)+u_y s\theta \\ u_x u_y(1-c\theta)+u_z s\theta & u_y u_y(1-c\theta)+c\theta & u_y u_z(1-c\theta)-u_x s\theta \\ u_x u_z(1-c\theta)-u_y s\theta & u_y u_z(1-c\theta)+u_x s\theta & u_z u_z(1-c\theta)+c\theta \end{pmatrix} \qquad (12)$$

where $c\theta = cos\theta$ and $s\theta = sin\theta$.

A second rotation φ around the axis $\vec{\mathbf{v}}(v_x, v_y, v_z) = \dfrac{\overrightarrow{ON'}_{Rx}}{\left|\overrightarrow{ON'}_{Rx}\right|}$ is required to achieve the best fitting of $T_R(S)$ to $T_R(S')$ (see Figure 5). The swing angle φ is determined by (14). This last set of points can be obtained by applying a rotation matrix Rv(φ) that depends on a single parameter φ and can be formally expressed as:

$$\mathbf{R_v}(\varphi)=\begin{pmatrix} v_x v_x(1-c\varphi)+c\varphi & v_x v_y(1-c\varphi)-v_z.s\varphi & v_x v_z(1-c\varphi)+v_y.s\varphi \\ v_x v_y(1-c\varphi)+v_z.s\varphi & v_y v_y(1-c\varphi)+c\varphi & v_y v_z(1-c\varphi)-v_x.s\varphi \\ v_x v_z(1-c\varphi)-v_y.s\varphi & v_y v_z(1-c\varphi)+v_x.s\varphi & v_z v_z(1-c\varphi)+c\varphi \end{pmatrix} \quad (13)$$

$$R_1(\varphi,\theta)=R_v(\varphi)\cdot R_u(\theta) \quad (14)$$

Finally the alignment of the spheres is:

$$\mathrm{T}_{\mathrm{R}}^{'}(S)=\mathrm{R}_1(\varphi,\theta)\cdot \mathrm{T}_{\mathrm{R}}(S) \quad (15)$$

In the next step the Fourier energy is calculated for each viewpoint.

Defining $S_{cp} \in O_i$, $S_{cq} \in O_j$ where S_{cp} , S_{cq} ∈ $[S_{c1}, S_{C2} ... S_{cf}]$, where f is the number of candidate silhouettes, and vp is a viewpoint from $T_R(S)$. The energy is computed for all couples of silhouettes as follows:

$$E_{Oi,Oj}^{vp}=\frac{1}{N}\sum_{k=0}^{N-1}|Z_{Oi}^{vp}(k)- \quad Z_{Oj}^{vp}(k)|^2, \qquad (16)$$
$$\forall i \neq j, i \leq f, j \leq f$$

$$E^{vp} = \min(E_{Oi,Oj}^{vp}) \quad (17)$$

The NBV ν is defined as the viewpoint that verifies:

$$E^{\nu} = \max(E^{vp}) \; \forall \, vp \quad (18)$$

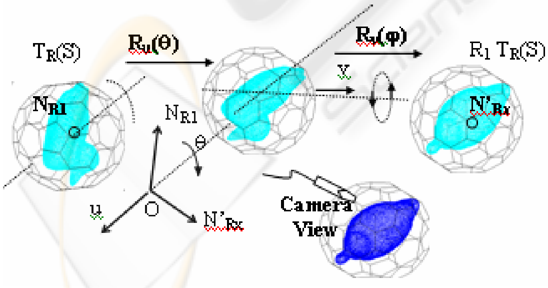


Figure 5: Alignment process between candidate spheres and the scene sphere.

Finally, the camera is moved to the best viewpoint and a new image of the scene is captured and matched with the model silhouette correspondents to the best viewpoint using (9) and (11) equations.

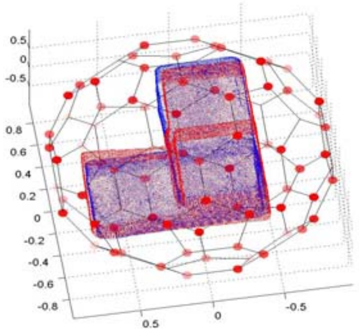


Figure 6: Superimposed models after the alignment process and energy values plotted on the nodes.

5 EXPERIMENTATION

A validation test of this method has been carried out in our lab. The experimental setup is composed of a Stäubli RX 90 Robot with a micro camera Jai-CVM1000 at its end. This system controls the position and vision direction on the camera, the object always being centered in the scene. Figure 7 shows the experimental setup.

In the off-line process, the synthesized models (with 80000 polygons/object) are built through a VI-910 Konica Minolta 3D Laser scanner. At the same time the silhouette database with their respective Fourier descriptors are obtained and stored. Currently we have used databases of 80 and 320 silhouettes/model.

In order to reduce redundant information and to optimize the recognition/pose time, a Fourier descriptors reduction has been carried out on the silhouette models. Figure 2.a shows Fourier descriptors modulus for an example. As X can be seen, the first and the last descriptors are the most meaningful. The reduction procedure consists of considering the intervals [1,X(k)], [X(N-k),X(N)], k=1,…N, until the error/pixel between the original and reduced silhouettes is less than a threshold χ .

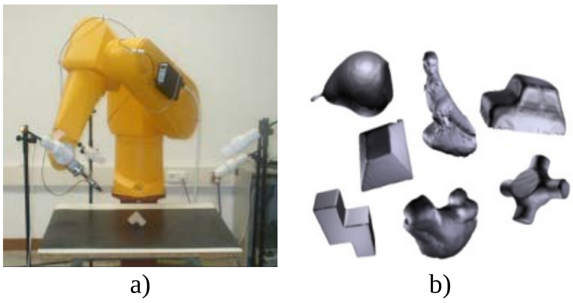


Figure 7: a) Experimental Setup. b) Examples of synthetic models.

The experimentation has been carried out with 19 objects that have been previously modelled with 80 and 320 views, considering descriptor reductions of χ =0.05 and χ =0.5.

In the clustering process we have used 50 clusters. During this phase we use images with resolution 640x480. Each silhouette in the database was stored with a resolution of 512 points and the database size was 12 MB (80 views) and 42 MB (320 views).

Table 1.

# sil.	χ	t_A	ρ_A	t_B	ρ_B
80	0.05	2.652	1.640	2.475	2.382
	0.5	2.047	1.653	1.863	2.473
320	0.05	4.901	1.089	4.739	2.107
	0.5	3.336	1.339	3.096	2.261

The active 3D recognition system worked in all tests achieving 100% X effectiveness. Table 1 shows the results obtained during the recognition process without (A) and with (B) discrimination phase. The results are compared taking into account: the number of silhouettes of the model, threshold for Fourier descriptors reduction (χ), mean square error between the silhouette of the scene and the estimated silhouette (ρ). Variable t is the computation time (seconds) on a Pentium III 800 Hhz processor.

Table II shows in detail the main process rates using a database with 80 silhouettes/model and a reduction factor χ =0.5.

From Tables 1 and 2 the following comments can be made.

- In the whole process, most of the time is devoted to extracting the object's silhouette (88,6% and 94,7%). Note that, this stage includes several image preprocessing tasks like filtering, thresholdin, etc. In part, such a high percentage is also due to the fact that we have used a reduced object database in our experimentation. For large databases (>100-500 objects) this percentage will decrease at the same time that the percentage corresponding to the candidates selection stage will increase.
- Using 320 silhouettes per model increases in a double the execution times with respect the use of 80 silhouettes per model but the ρ decreases by 0,3 percent.

Table 2.

	Algorithm	time (%)
Without clustering	Silhouette extraction	88.6
	Identification	7.9
	Pose estimation	1.4
	NBV	2.1
With clustering	Silhouette extraction	94.7
	Discrimination	0.8
	Identification	2.2
	Pose estimation	0.6
	NBV	1.7

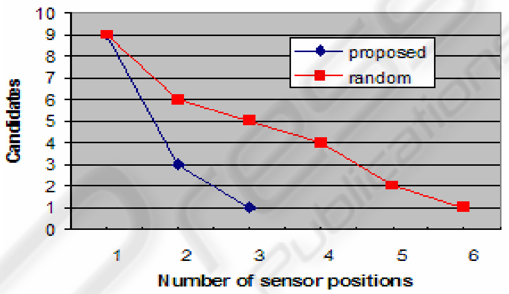


Figure 8: Comparison of discrimination between a random method and our proposed method.

Two experiments were carried out: one running our active recognition system which uses a random selection of the next view, and another computing the next best view from our D-Sphere structure. In Figure 8 we can see the number of candidates in each sensor position for a real case. The test average reported that our method considerably reduced the number of sensor movements: about 62%. The time needed to calculate the next observation position is very short: approximately 1.7% of all the time needed to carry out a complete step of the recognition process. Calculations were performed on a Pentium III 800 Hhz processor. The active 3D recognition system worked in all tests achieving 100% X effectiveness. Figure 9 illustrates a case of ambiguity between two objects and how the system solves the problem. Our NBV method shows much higher discriminative capability than the random method. Thus, the proposed strategy significantly improves the recognition efficiency.

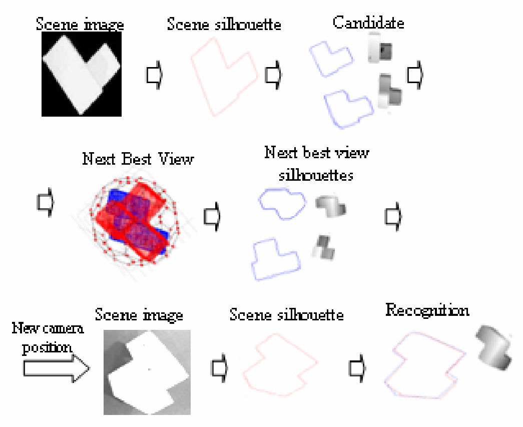


Figure 9: Solving an ambiguous case.

Figure 10 presents some matching and pose estimation results using the proposed algorithm.

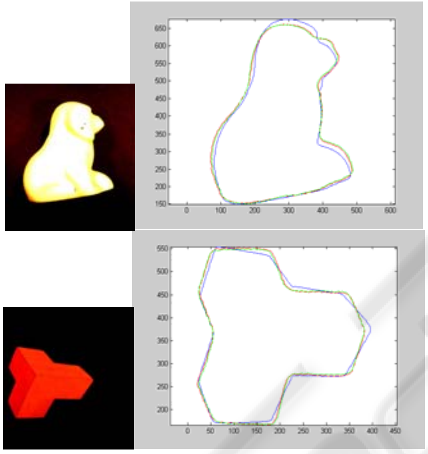


Figure 10: Some examples of 3D recognition without ambiguity.

6 CONCLUSION

This paper has presented a new active recognition system. The system turns a 3D object recognition problem into a multiple silhouette recognition problem where images of the same object from multiple viewpoints are considered. Fourier descriptors properties have been used to carry out the clustering, matching and pose processes.

Our method implies the use of databases with a very large number of stored silhouettes, but an efficient version of the matching process with Fourier descriptors make it possible to solve the object recognition and pose estimation problems in a greatly reduced computation time.

On the other hand, the next best view (NBV) method efficiently solves the frequent ambiguity problem in recognition systems. This method is very robust and fast, and is able to discriminate among very close silhouettes.

REFERENCES

Adán, A., Cerrada, C., Feliu, V., 2000. Modeling Wave Set: Definition and Application of a New Topological Organization For 3D Object Modeling. Computer Vision and Image Understanding 79, pp 281-307.

Bustos, B., Kein, D.A., Saupe, D., Schreck, T., Vranic, D., 2005. Feature-based Similarity Search in 3D Object Databases. ACM Computing Surveys (CSUR) 37(4):345-387, Association For Computing Machinery.

Borotschnig, H. Paletta, L., Pranti, M. and Pinz, A. H. 1999. A comparison of probabilistic, possibilistic and evidence theoretic fusion schemes for active object recognition. Computing, 62:293–319.

Deinzer, F., Denzler, J. and Niemann, H., 2003. Viewpoint Selection. Planning Optimal Sequences of Views for Object Recognition. In Computer Analysis of Images and Patterns, pages 65-73, Groningen, Netherlands, Springer.

Deinzer, F., Denzler, J., Derichs, C., Niemann, H., 2006: Integrated Viewpoint Fusion and Viewpoint Selection for Optimal Object Recognition. In: Chanteler, M.J. ; Trucco, E. ; Fisher, R.B. (Eds.) : British Machine Vision Conference

Helmer S. and Lowe D. G, 2004. Object Class Recognition with Many Local Features. In Workshop on Generative Model Based Vision 2004 (GMBV), July.

Hutchinson S.A. and Kak. A.C., 1992. Multisensor Strategies Using Dempster-Shafer Belief Accumulation. In M.A. Abidi and R.C. Gonzalez, editors, Data Fusion in Robotics and Machine Intelligence, chapter 4, pages 165–209. Academic Press,.

Netanyahu, N. S., Piatko, C., Silverman, R., Kanungo, T., Mount, D. M. and Wu, Y., 2002. An efficient k-means clustering algorithm: Analysis and implementation. vol 24, pages 881–892, july.

Niku S. B., 2001. Introduction to Robotics, Analysis, Systems, Applications. Prentice Hall.

Poppe, R.W. and Poel, M., 2005. Example-based pose estimation in monocular images using compact fourier descriptors. CTIT Technical Report series TR-CTIT-05-49 Centre for Telematics and Information Technology, University of Twente, Enschede. ISSN 1381-3625

Roy, S.D., Chaudhury, S. and Banerjee. S., 2004. Active recognition through next view planning: a survey. Pattern Recognition 37(3):429–446, March.

Sipe M. and Casasent, D., 2002. Feature space trajectory methods for active computer vision. IEEE Trans PAMI, Vol. 24, pp. 1634-1643, December.