

Semantic labeling of 2D objects with 3D models

Raluca Diana Petre, Zaharia Titus

▶ **To cite this version:**

Raluca Diana Petre, Zaharia Titus. Semantic labeling of 2D objects with 3D models. 5th IEEE International Conference on Semantic Computing (ICSC 2011), Sep 2011, United States. pp.419-423. hal-00738236

HAL Id: hal-00738236
https://hal.science/hal-00738236

Submitted on 4 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semantic labeling of 2D objects with 3D models

Raluca-Diana Petre
ARTEMIS Department
Institut TELECOM; TELECOM SudParis
Evry, France
Raluca-Diana.Petre@it-sudparis.eu

Titus Zaharia
ARTEMIS Department
Institut TELECOM; TELECOM SudParis
Evry, France
Titus.Zaharia@it-sudparis.eu

***Abstract*—This paper tackles the issue of still image object categorization. The objective is to infer the semantics of 2D objects present in natural images. The principle of the proposed approach consists of exploiting categorized 3D models in order to identify unknown 2D objects, based on 2D/3D matching techniques. Notably, we use 2D/3D shape indexing methods, where 3D models are described through a set of 2D views. Experimental results, carried out on both MPEG-7 and Princeton 3D databases show recognition rates of up to 89.2%.**

Keywords-indexing and retrieval; object classification; 3D mesh; 2D/3D shape descriptors.

I. INTRODUCTION

The last decade has been greatly influenced by the spectacular development of the digital technologies. Nowadays, the availability of low costs audio-visual (AV) acquisition and storage devices leads to very large collections of images and videos that can shared by multiple users. Such databases are used in domains like media, commerce, academia, security or for personal purposes. In this context, a fundamental issue that needs to be addressed concerns the efficient handling of digital content. Retrieving specific resources (*i.e.*, elements of interest for the user) is not possible when a large database is involved. The use of keywords is also restricted by the linguistic variety and needs a prior annotation of the digital content (involving human contribution). Manually labeling the AV material is not a solution that can be taken into account because of human subjectivity as well as time constraints.

The need for automatic object categorization methods appears as a crucial challenge. The objective is to determine automatically the semantic meaning of an object present in an image or video document. The great majority of the existing approaches is based on machine learning (ML) techniques [1], [2]. Such algorithms aim to automatically learn to recognize complex structures and involve two main steps: the learning and the classification stages. First, the system needs a set of examples (*i.e.* the training database). Second, based on the learned examples, the method has to be able to generalize in order to recognize new cases.

The machine learning techniques include two main families: supervised and unsupervised techniques. In the case of supervised methods, the system disposes of several sets of labeled data and aims at finding the function which better discriminates between these sets. Once this function is defined, it can be used to classify new data. Some approaches based on supervised machine learning methods are proposed in [3], [4], [5]. Even if the supervised approaches may be very accurate [6], they can often suffer from overfitting [7]. Another limitation comes from the need of sufficiently large training sets with already classified objects. Results are strongly dependent on the considered training set.

On the contrary, the unsupervised machine learning methods allow training from partially or completely unlabelled data. Some commonly used unsupervised machine learning methods are K-means, Mixture methods, K-Nearest Neighbor… For some examples, the reader is invited to refer [8], [9]. However, in terms of performances, the unsupervised methods are less accurate than the supervised machine learning methods.

In a general manner, when dealing with large databases involving an important number of classes, machine learning approaches need to exploit more features. The corresponding computational complexity becomes in such cases intractable [10]. Another important aspect is that objects may have very different appearances in images because of variation in pose. Thus, the training set should include not merely different examples of objects from each class, but also different instances of objects, corresponding to different poses.

In this paper we present a new method that avoids applying machine learning techniques when dealing with a large variety of objects. Instead of using ML, we propose to use categorized 3D models from existing 3D repositories in the classification process.

The paper is structured as follows. In Section 2 we present the 2D/3D shape-based indexing approach adopted. The object recognition framework is described in Section 3, while the experimental results are presented and analyzed in Section 4. Finally, Section 5 concludes the paper and opens perspectives of future work.

II. SHAPE-BASED 2D/3D INDEXATION

Let us first recall the general principle of 2D/3D indexing methods.

A. The Principle of 2D/3D Indexation

The underlying principle of 2D/3D indexing approaches is based on the following observation: two similar 3D models should present similar views when projected in 2D images. Thus, instead of describing a 3D object in the original 3D

space, the model is represented as a set of 2D views associated to different projection angles (and under the assumption of a given projection model).

Such a strategy makes it possible to compare two different 3D models, but also to compare a 3D model with a 2D object present in 2D images.

In order to obtain a unique set of views, whatever the object's size, position or orientation, each model M is first centered in the origin of the Cartesian system and resized to fit the unit sphere. Furthermore, the model is projected and rendered in 2D from N different viewing angles, resulting in a set of N projections, denoted by $P_i(M)$. In our case, we have exclusively used binary images corresponding to the projected silhouettes (Figure 1. Finally, each projection $P_i(M)$ is described by a 2D shape descriptor $d_i(M)$. The set of all descriptors $\{d_i(M)\}$ yields the 2D.3D representation of the considered 3D object.

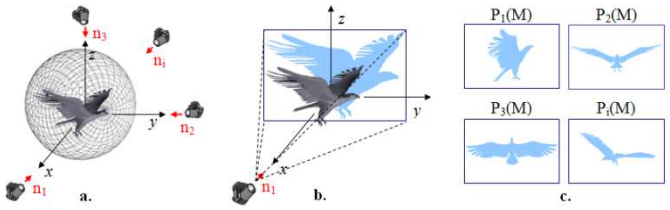


Figure 1. Projecting a model. a. Viewing directions n_i; b. Model projection according to the n_1 direction; c. the resulting silhouette images.

In order to fully implement a 2D/3D indexing approach, several elements have to be specified. The first one is the set of viewing directions $\{n_i\}$ used to perform the projections. Also, the number of images per model has to be carefully considered, since a large number of silhouettes provide more complete descriptions but also increases the computational cost of the subsequent matching algorithms. Finally, the choice of appropriate 2D shape descriptors is fundamental for ensuring a discriminant description.

In the next section we will present different strategies of projection and descriptors considered in our work.

B. The proposed 2D/3D indexing methods

1) The viewing angles selection:

Several strategies for selecting a set of viewing angles can be considered. A first approach, also proposed by the MPEG-7 standard [11], is based on the assumption that the most discriminant views are those corresponding to the principal planes (obtained with the Principal Component Analysis (PCA) [12]) (Figure 2.).

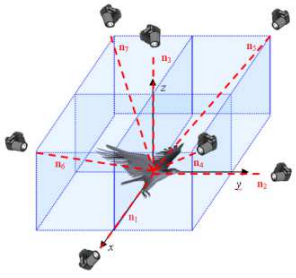


Figure 2. PCA-based positioning of the camera

Moreover, the three principal planes define eight octants. If we consider their bisectors, four additional views can be added to the first three images in order to obtain a more complete representation. From now on, we will refer to these PCA-based strategies as *PCA3* and *PCA7* (with 3 and respectively 7 views).

A second approach for the viewing angles selection aims at evenly distributing the cameras around the model. A first method, introduced in [13], consists of considering the vertices of a regular dodecahedron which results in 10 views. Two sub-cases can be further considered. The first case corresponds to the situation where object's axes of inertia are aligned with the coordinate system (for example, by using PCA). In the second case, the object has an arbitrary, random orientation (Figure 3.). As the dodecahedron-based positioning of the camera is used for the Light Field Descriptor (LFD) [13], we will use the acronyms *LFDPCA* respectively *LFD* for denoting these two projection strategies.

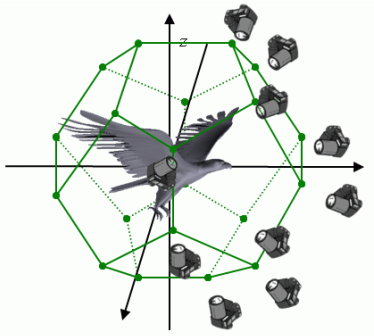


Figure 3. Dodecahedron-based positioning of the camera.

Finally, we have also considered a third projection strategy which combines the previous two, using at the same time the views on the principal planes (given by PCA) and a uniform distribution of the cameras around the object. In this case, the vertices of an octahedron are used as support for the camera [14]. In order to obtain additional views, each face of the octahedron is successively subdivided. At the first level there are 3 viewing directions (which are the same of the PCA3 strategy). At the second level 9 views are obtained while at the third level 33 (Figure 4.). From now on, we will refer to these techniques as *OCTA3*, *OCTA9* and *OCTA33*.

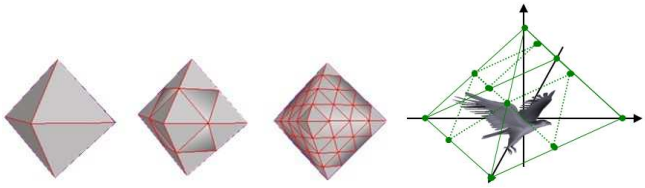


Figure 4. Octahedron-based camera positioning.

2) The 2D shape description:

The second part of the 2D/3D indexing process concerns the extraction of the 2D shape descriptors. As the views are binary images representing only the shape of the object (without inside contours, shadows…), the only features that can be exploited are the exterior contour and the corresponding region of support. In our work, we have considered two

contour-based and two region-based descriptors, briefly recalled here below.

Let us start with the Contour Scale Space (*CSS* or *CS*) descriptor [15] proposed by the MPEG-7 standard [16], [17], [18]. The CSS is obtained by successively convoluting the curve representing the contour with a Gaussian kernel. Using a multi-scale analysis process, the curvature peaks are determined, with curvature value and corresponding position in curvilinear abscise. The associated similarity measure between two CSS representations is based on a matching procedure which takes into account the cost of fitted and unfitted curvatures peaks [18].

The second approach adopted is the MPEG-7 Region Shape (*RS*) descriptor, based on the 2D Angular Radial Transform (ART) [19]. In this case, the object's support function is represented as a weighted sum of 34 ART basis functions. The decomposition coefficients constitute the descriptor. The distance between two shapes is simply defined as the L_1 distance between the absolute values of the ART coefficients.

Another region descriptor is based on the 2D Hough Transform (HT) [20]. Each point p corresponding to the silhouette of the object is represented in the (s, θ) space. If we consider a line l passing through the point p, then θ is the angle between the line l and the Ox axe, and s represents the distance from the coordinate system origin to the line. Therefore, the image can be represented in the (s, θ) cumulative space. The associated similarity measure between two HT representations is the L_1 distance computed for the (s, θ) coefficients.

Finally, we propose a new descriptor, so-called Angle Histogram (AH). The shape contour is first sub-sampled in a number of successive 2D points. The angular histogram is created by computing the angles defined by each three consecutive samples. In our experiments we have used a 18 bins histogram for an 180° interval. Different histograms are obtained, depending on the sampling steps. When the sampling step is small, the histogram will encode the details of the contour, while for big values the global features are extracted. The 2D AH results by concatenating five angular histograms obtained with different sampling steps. A simple L_1 distance is used as similarity measure.

Let us now describe the 2D shape recognition framework proposed.

III. 2D Shape Recognition Framework

Figure 5. presents an overview of the 2D shape recognition framework.

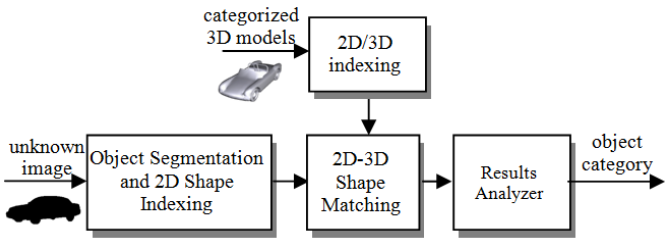


Figure 5. The 2D shape recognition framework.

A 3D categorized database is supposed to be available. Each model in the database is described by the four descriptors presented in Section II.B.2) for all the 6 projection strategies retained (*i.e.* PCA3, PCA7, LFD, LFDPCA, OCTA9, OCTA33).

As input to the system, we consider a binary image corresponding to an unknown object whose semantics needs to be determined. Such an object can be obtained with the help of some semi-automatic segmentation methods [21]. The 2D binary object is also indexed with all the 4 shape descriptors retained.

The distance $d(O,M)$ between the 2D object O and the 3D model M is given by the minimum distance between the 2D object and all the projections $P_i(M)$ of the 3D model.

$$d(O,M) = \min_i d\big(O, P_i(M)\big). \quad (1)$$

In order to retrieve the category of the input image, the system analyzes which are the most similar objects. Next, the categories that are the most represented among the first top retrieved 3D models are presented as potential classes of the 2D object.

In order to measure the performance of this 2D shape recognition system, we have established an experimental evaluation protocol, described in the next section.

IV. Experimental Evaluation

The experiments have been carried out on two different 3D model databases. First, the MPEG-7 dataset [11], which is composed of 362 models divided into 23 semantic classes. Categories include humanoids, airplanes, helicopters, cars, race cars, trees (with and without leafs), rifles, missiles, pistols etc. These models present intra-class variability as well as inter-class similarity. The second 3D database we have used is the Princeton Shape Benchmark (PSB) [22], which includes 1814 models semantically categorized in 161 classes. Compared to the MPEG-7 database, this classification is more precise and presents a hierarchical tree structure which supports sub-classes, For example, a distinction between commercial airplanes, (*e.g.*, biplanes, fighter jet, glider airplane...) is done for the "airplane" category. The PSB database includes various models representing aircrafts, animals, furniture, plants, sea vessels, musical instruments, tools, vehicles …

We have also created a 2D object database consisting of 115 images randomly chosen from the web (corresponding to 5 images for each MPEG-7 category). When using the PSB, only 65 objects have been considered (corresponding to the 13 categories that are common for the MPEG-7 and the PSB databases). For each image, the objects of interest have been manually segmented from the available images.

The performance measure adopted is the recognition rate (RR), defined as the percentage of cases where the correct category is assigned to the input image. In order to associate a class $C(O)$ to a 2D object O, the N most similar 3D models $(M_1 \ldots M_N)$ from the database are considered. Each model belongs to a category. We can thus identify a number of N_C most represented categories among the N top retrieved results

(C_1 ... C_{Nc}). If one of these classes coincides with the category to which belongs the image, then we can state that the recognition has succeeded.

The RR is then defined as a function of the number N_C of possible categories accepted, as described by the following equation:

$$RR(N_C) = \frac{\sum_{i=0}^{N} Recognition\ (O_i, N_C)}{N}., \quad (2)$$

where

$$Recognition(O_i, N_C) = \begin{cases} 1; & if\ C(O) \in \{C_1 \dots C_{Nc}\} \\ 0; & otherwise \end{cases}. \quad (3)$$

In our experiments, we have taken into account one, two or three most represented categories (*Nc = 1,2,3*). In the case of Princeton database, where the number of existing categories is more important (161 classes), we have also computed the score $RR(N_C=10)$.

The parameter N which gives the length of the analysis window has been set to 20, which represents approximately the average size of the categories.

Tables 1 and 2 respectively present the scores obtained using the MPEG-7 and Princeton Shape Benchmark databases.

TABLE I. RECOGNITION RATE FOR THE MPEG-7 DATABASE

a.

CS	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	33.9	34.8	37.4	33.9	37.4	37.4
R(2)	41.7	53.9	52.2	50.4	51.3	51.3
R(3)	53.9	61.7	**59.1**	60.0	56.5	**60.0**

b.

RS	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	24.3	22.6	28.7	27.0	26.1	30.4
R(2)	36.5	37.4	40.9	37.4	42.6	46.1
R(3)	40.9	45.2	**46.1**	45.2	50.4	**54.8**

c.

AH	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	30.4	35.7	44.3	42.6	32.2	38.3
R(2)	47.8	55.7	60.9	56.5	48.7	60.0
R(3)	56.5	61.7	**67.0**	62.6	60.0	**70.4**

d.

H	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	18.3	20.9	27.0	24.3	28.7	34.8
R(2)	27.0	29.6	35.7	30.4	36.5	41.7
R(3)	37.4	37.4	**46.1**	35.7	43.5	**49.6**

e.

CS + AH	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	37.4	40.0	41.7	41.7	38.3	39.1
R(2)	47.0	53.0	60.0	55.7	53.9	60.0
R(3)	58.3	62.6	**71.3**	67.8	61.7	70.4

f.

CS + AH LFD	PCA3	PCA7	LFD	LFD PCA	OCT A9	OCT A33
R(1)	41.7	40.9	41.7	41.7	40.9	39.1
R(2)	55.7	54.8	60.0	56.5	58.3	57.4
R(3)	66.1	68.7	**71.3**	68.7	65.2	68.7

TABLE II. RECOGNITION RATE FOR THE PSB

a.

CS	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	32.3	41.5	40.0	41.5	41.5	44.6
R(2)	43.1	53.8	53.8	50.8	49.2	58.5
R(3)	49.2	58.5	**58.5**	55.4	56.9	**64.6**
R(10)	63.0	**76.9**	72.3	69.2	69.2	72.3

b.

RS	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	26.2	20.0	23.1	24.6	29.2	32.3
R(2)	30.8	27.7	32.3	41.5	43.1	40.0
R(3)	38.5	35.4	38.5	41.5	46.2	**46.2**
R(10)	55.4	49.2	55.4	55.4	63.1	60.0

c.

AH	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	27.7	40.0	40.0	36.9	35.4	44.6
R(2)	40.0	50.8	49.2	53.8	50.8	52.3
R(3)	49.2	55.4	52.3	58.5	**60.0**	53.8
R(10)	66.2	70.8	72.3	73.8	73.8	**76.9**

d.

H	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	10.8	12.3	21.5	18.5	26.2	26.2
R(2)	12.3	15.4	32.3	23.1	32.3	33.8
R(3)	15.4	20.0	36.9	24.6	35.4	**40.0**
R(10)	30.8	35.4	41.5	29.2	49.2	52.3

e.

CS + AH	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	36.9	49.2	46.2	46.2	43.1	44.6
R(2)	47.7	60.0	56.9	60.0	49.2	61.5
R(3)	53.8	66.2	61.5	63.1	56.9	**67.7**
R(10)	67.7	81.5	81.5	83.1	80.0	**84.6**

f.

CS + AH LFD	PCA3	PCA7	LFD	LFDPCA	OCTA9	OCTA33
R(1)	44.6	50.8	46.2	50.8	47.7	44.6
R(2)	56.9	60.0	56.9	56.9	53.8	60.0
R(3)	64.6	66.2	61.5	63.1	61.5	**66.2**
R(10)	83.1	83.1	81.5	**89.2**	84.6	86.2

For both databases, we observe a global behavior regarding the viewing angle selection; in most cases LFD and OCTA33 strategies led to the maximal performances in terms of recognition rate, whatever the considered descriptor. We achieved 60% recognition rate for the CS descriptor and 70.4% for AH when employing the MPEG-7 database. In the case of PSB database, the same global behaviors were observed. CS and AH are the descriptors providing the highest recognition rates, with *RR(3)* scores of 64.6% and respectively 60%. When considering the *RR(10)* scores, the recognition rates increase up to 76.9% for both CS and AH descriptors.

We have also tested our system when the two descriptors with the best performance (*i.e.* CS and AH) were combined (tables I.e. and II.e.). The idea here is to attempt to exploit the possible complementarities between the two descriptors. Therefore, instead of computing the scores based on the most similar models given by the CS, we have also considered those provided when using AH descriptor. Thus, the *RR(3)* scores have increased up to 71.3% when using the MPEG-7 database and up to 67.7% when PSB was used. Also, the recognition rate of the combined descriptor has improved to 84.6% when analyzing the *RR(10)*.

As the LFD strategy provides promising results, while using a small number of views per model, we have also tested the combination between the LFD and all the other projection strategies, reaching a recognition rate of 89.2% (tables I.f. and II.f.), when LFD is combined with LFDPCA.

The results presented above show the interest of integrating some *a priori* knowledge in the recognition process, driven from existing 3D models and exploited with the help of 2D/3D indexing techniques.

Despite the fact that the highest scores are obtained when several candidate categories are taken into consideration ($N_C \neq 1$), we believe that such a multiple response is still very useful. Our framework can be used in order to reduce the number of candidate categories from 161 (in the case of PSB database) to N_C. Therefore, if we integrate this approach within existing machine learning techniques, we can significantly speed-up the recognition process. Such a mixed system would also allow achieving superior recognition rates.

Finally, when considering the issue of 2D/3D object retrieval it is useful to develop appropriate user-interfaces that can help to both evaluate the approaches and perform semi-automatic data annotation. The proposed system is illustrated in Figures 6, 7 and 8.

The user has the possibility to select different descriptors and projection strategies, to perform queries, compare/validate results and finally annotate images.

Figures 6, 7 and 8 contain examples of queries representing a humanoid, an airplane and respectively a tree. It can be observed that the category of the given query image was retrieved within the first three returned positions.

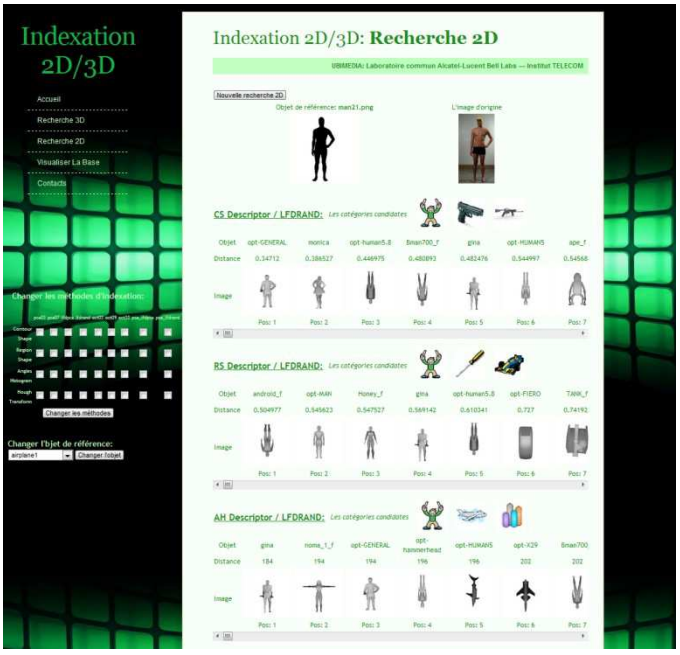


Figure 6. 2D/3D retrieval and categorization with the proposed system, with a query representing a humanoid.

The 2D/3D retrieval and categorization system has been developed with the help of web technologies/services and thus can be remotely accessed by multiple users.



Figure 7. 2D/3D retrieval and categorization with the proposed system, with a query representing an airplane.

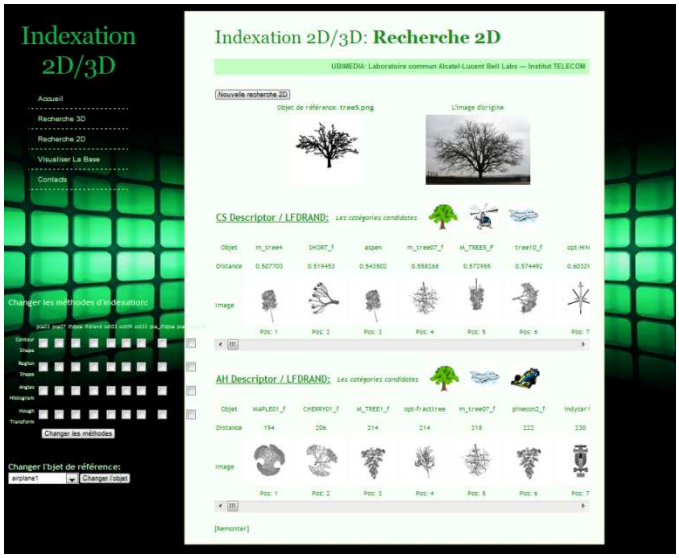


Figure 8. 2D/3D retrieval and categorization with the proposed system, with a query representing a barren tree.

V. Conclusions and Future Work

In this paper we have presented a novel recognition algorithm for semantic labeling of 2D objects extracted from still images. As the projection strategy and the 2D shape descriptors are key issues for the 2D/3D indexing methods, we have analyzed the performance of different such approaches. Thus, we observed that LFD and OCTA33 strategies provide better scores in most cases. When comparing the four descriptors that we have tested, we have observed that the two contour-based descriptors (*i.e.* CS and AH) provided highest

recognition rates. Moreover, we have exploited their complementarity by combining them and thus improving the scores obtained with only one descriptor. We have also computed the recognition rates obtained when combining two indexing methods based on different projection strategies. As a result, the scores were further improved (up to 89.2%) when we considered two viewing angle selection approaches.

In our future work we intend to extend the proposed approaches to 2D video objects. By using a tracking tool, the system will dispose of several views per query model. This additional information could greatly help the recognition process.

Furthermore, we plan to integrate an approach exploiting the internal edges of the shape in order to obtain more discriminant descriptions.

ACKNOWLEDGMENT

This work has been performed within the framework of the UBIMEDIA Research Lab, between Institut TELECOM and Alcatel-Lucent Bell-Labs. We also wish to thank Princeton University for supplying their 3D shape database.

REFERENCES

[1] Mitchell, T. M., 1997. Machine Learning. New York: McGraw-Hill.

[2] Xue, M., Zhu, C., A Study and Application on Machine Learning of Artificial Intelligence, International Joint Conference on Artificial Intelligence, pp. 272, July 2009.

[3] A. Bosch, A. Zisserman, and X. Mu~noz. Image classication using random forests and ferns. In International Conference on Computer Vision, Rio de Janeiro, Brazil, Oct. 2007.

[4] F. Moosmann, B. Triggs, and F. Jurie. Fast discriminative visual codebooks using randomized clustering forests. In Neural Information Processing Systems Conference, Vancouver, BC, Canada, Dec. 2006.

[5] J. Shotton, J. Winn, C. Rother, and A. Criminisi. TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In European Conference on Computer Vision, volume 3951 of Lecture Notes in Computer Science, pages 1{15, Graz, Austria, May 2006.

[6] Deselaers, T., Heigold, G., Ney, H., Object classification by fusing SVMs and Gaussian mixtures, Vol. 43, Issue 7, pp. 2476-2484, July 2010.

[7] Pados, G.A., Papantoni-Kazakos, P., A note on the estimation of the generalization error and prevention of overfitting [machine learning], IEEE Conference on Neural Networks, volume 1, pp 321, July 1994.

[8] M. Weber, M. Welling, and P. Perona. Unsupervised learning of models for recognition. In Proc. ECCV, pages 18–32, 2000.

[9] R. Fergus, P. Perona, A. Zisserman. Object class recognition by unsupervised scale-invariant learning, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2 (2003), pp. 264-271, June 2003.

[10] Li, Ling, Data complexity in machine learning and novel classification algorithms. Dissertation (Ph.D.), California Institute of Technology, 2006.

[11] T. Zaharia, F. Prêteux, 3D versus 2D/3D Shape Descriptors: A Comparative study, In SPIE Conf. on Image Processing: Algorithms and Systems, Vol. 2004 , Toulouse, France, January 2004.

[12] R.A. Schwengerdt, Remote Sensing: Models and Methods for Image Processing, 2nd. Ed., Academic Press, 1997.

[13] Ding-Yun Chen, Xiao-Pei Tian, Yu-Te Shen and Ming Ouhyoung, On visual similarity based 3D model retrieval, Computer Graphics Forum, vol. 22, no. 3, pp. 223-232, 2003.

[14] Petre, R., Zaharia, T., Preteux, F., "An overview of view-based 2D/3D indexing methods", Proceedings of Mathematics of Data/Image Coding, Compression, and Encryption with Applications XII, volume 7799, August 2010.

[15] F. Mokhtarian, A.K. Mackworth, A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves, IEEE Transaction on Pattern Analysis and Machine Intelligence, pp. 789-805, August 1992.

[16] M. Bober, "MPEG-7 Visual Shape Descriptors", IEEE Transaction on Circuits and Systems for Video Technology, Volume 11, Issue 6, pp. 716-719, August 2002 .

[17] B.S. Manjunath, Phillipe Salembier, Thomas Sikora, "Introduction to MPEG-7: Multimedia Content Description Interface", John Wiley & Sons, Inc., New York, NY, 2002.

[18] ISO/IEC 15938-3: 2002, MPEG-7-Visual, Information Technology – Multimedia content description interface – Part 3: Visual, 2002.

[19] W.-Y. Kim, Y.-S. Kim, New Region-Based Shape Descriptor, ISO/IEC MPEG99/M5472, Maui, Hawaii, December 1999.

[20] Hart, P.E.: How the Hough transform was invented" IEEE Signal Processing Magazine, November 2009.

[21] Sapna Varshey, S., Rajpal, R, "Comparative study of image segmentation techniques and object matching using segmentation", Proceeding of International Conference on Methods and Models in Computer Science, January 2010.

[22] Philip Shilane, Patrick Min, Michael Kazhdan, and Thomas Funkhouser, "The Princeton Shape Benchmark", Shape Modeling International, Genova, Italy, June 2004.