3D model-based still image object categorization

Raluca Diana Petre, Titus Zaharia

▶ **To cite this version:**

Raluca Diana Petre, Titus Zaharia. 3D model-based still image object categorization. SPIE, Aug 2011, United States. pp.81360C, 10.1117/12.904964 . hal-00738214

HAL Id: hal-00738214
https://hal.science/hal-00738214v1

Submitted on 3 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.

3D Model-based Still Image Object Categorization

Raluca-Diana Petre[(1)], Titus Zaharia[(1)],
[(1)]Institut TELECOM; TELECOM SudParis, ARTEMIS Department; UMR CNRS 8145 MAP5
9 rue Charles Fourier, 91011 Evry Cedex, France
{Raluca-Diana.Petre, Titus.Zaharia}@it-sudparis.eu

ABSTRACT

This paper proposes a novel recognition scheme algorithm for semantic labeling of 2D object present in still images. The principle consists of matching unknown 2D objects with categorized 3D models in order to infer the semantics of the 3D object to the image. We tested our new recognition framework by using the MPEG-7 and Princeton 3D model databases in order to label unknown images randomly selected from the web. Results obtained show promising performances, with recognition rate up to 84%, which opens interesting perspectives in terms of semantic metadata extraction from still images/videos.

Keywords: indexing and retrieval, object classification, 2D and 3D shape descriptors.

1. INTRODUCTION

The amount of multimedia content (still image, video, 2D/3D graphics…) available today for the general public is continuously increasing. Within this context, disposing of powerful search and retrieval methods becomes a key issue for efficient indexing and intelligent access to audio-visual material.

This paper addresses the issue of automatic 2D object recognition. The goal is to automatically identify the semantics of the image/video objects present in the content, with the help of computer vision methods. To achieve this goal, the majority of state of the art methods [1], [2] exploit some machine learning techniques. Such approaches need to perform a learning stage on sufficiently large databases. However, the results strongly depend of the learning sets used. In addition, the scalability issues when dealing with a large number of target concepts is still an open issue of research.

The approach proposed in this paper relies on different principle which consists of integrating in the recognition process some *a priori* knowledge, driven from existing 3D models. Today, numerous 3D graphical model databases are available. In addition such databases are semantically categorized, *i.e.* the semantics of each item is known. The objective then is to apply 2D/3D indexing and matching algorithms in order to determine, for a 2D object given as a query, the most similar candidate 3D models from the considered 3D repository. By analyzing the categories to which belong the retrieved objects, we can infer the semantics of the query object (Figure 1).

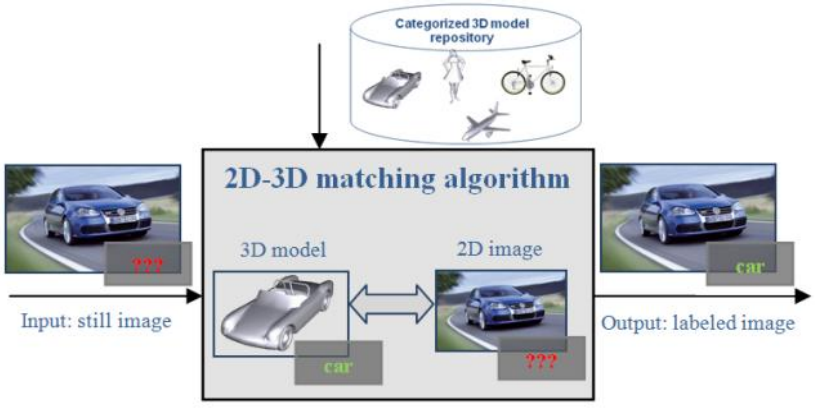

Figure 1. 2D categorization using 3D labeled models.

2. STATE OF THE ART

In the field of automatic object categorization, the great majority of the existing approaches is based on machine learning (ML) techniques [11], [12]. Such algorithms aim to automatically learn to recognize complex structures.

Two main families of approaches, supervised and unsupervised, can be considered. 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 [13], [14], [15]. Even if the supervised approaches may be highly accurate [16], they can often suffer from overfitting [17]. 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 even completely unlabelled data. Some commonly used unsupervised machine learning methods are K-means, mixture methods, K-nearest neighbors… For some examples, the reader is invited to refer [18], [19]. 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 a large set of features. The corresponding computational complexity becomes in such cases intractable [20]. 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 order to overcome the sensitivity of such methods to the object's 3D pose, in this paper we consider a different approach, based on 2D/3D shape indexing methods. The principle consists of introducing in the recognition process *a priori* 3D information, with the help of existing 3D models.

In order to be associated with 2D content, the 3D models are described by a set of 2D images, corresponding to 3D/2D projections obtained with different viewing angles. The obtained projection images are further described by exploiting 2D shape descriptors.

One of the most effective 2D/3D indexing methods is based on the Light Field Descriptor (LFD) [21] and uses a set of 100 evenly distributed views. Each projection is encoded by both Zernike moments [22] and contour-based Fourier coefficients [23]. In [25], the same descriptors (Zernike moments and Fourier coefficients) are combined with the Krawtchouk moments [24] in order to describe each of the 18 silhouettes composing the compact multi-view descriptor (CMVD). Within the ISO/MPEG-7 framework [5], two different methods using three or seven PCA-based viewing angles are proposed. The first one employs 2D-ART descriptor [6] while the second one represents the 2D object in the contour scale space (CSS) [4]. Another multi-scale curve representation was proposed in [26]. The silhouette's contour is here sampled into $N=100$ points and convolved with 10 different Gaussian filters. The variation in position of each point is computed and used to describe the shape. Concerning the selection of the considered viewing angles, several strategies are proposed in [27] and [28], where authors aim at minimizing the number of views by selecting some prototypical silhouette images among a large number of views.

Despite a rich state of the art of 2D/3D indexing algorithms, all of the above-presented methods target mostly 3D model retrieval applications and thus, they are not directly applicable for 2D image classification purposes.

The idea of using categorized 3D synthetic models for real object recognition has been recently exploited by Toshev *et al.* in [29], and by Liebelt *et al.* in [30]. The algorithm proposed in [29] aims at classifying objects automatically segmented from videos. Thus, the query is represented not only by one image but by a set of instances of the object. The 3D models used in the recognition stage are described by a set of 20 projections. These images are chosen by k-means clustering of 500 evenly distributed projections. The algorithm proposed in [30] makes use of textured synthetic 3D models (in contrast with the algorithm proposed in this paper and with the one presented in [29] which use exclusively shape information). Appearance features are selected and used in order to build for each class a visual codebook of $K=2000$ clusters.

In this paper, we propose a new image recognition framework. The recognition performances are evaluated using larger datasets (up to 23 query categories compared to 2 or 3 classes used in [30] and [29] respectively). In addition, we present a comparative experimental evaluation of different 2D/3D indexing methods.

3. THE PROPOSED ALGORITHM

Let us first briefly recall the principle of 2D/3D indexing methods. Each 3D model in a given 3D database is represented as a set of 2D views, which correspond to 3D-to-2D projections of the 3D model from several viewing angles (

Figure 2). Each view represents a 2D shape which is further characterized with the help of a set of 2D shape descriptors. In order to allow matching between 3D models and 2D objects, the same shape descriptor is used for the query objects, which are first segmented from the query image.

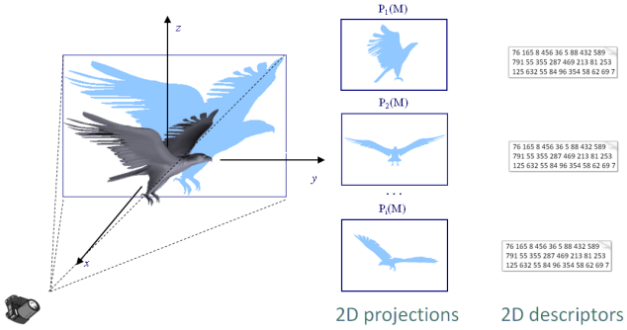

Figure 2. The principle of 2D/3D indexing techniques.

The 3D objects are first normalized in size and 3D pose, by applying alignment with respect to it's the three axes of inertia, computed using a Principal Component Analysis (PCA) [9].

In order to generate the set of projections, we have used several viewing angles selection strategies [3].

The first one, suggested by the MPEG-7 Multiview descriptor uses the projections on the three principal planes and optionally four secondary views (so-called PCA3 and PCA7 strategies) (Figure 3).

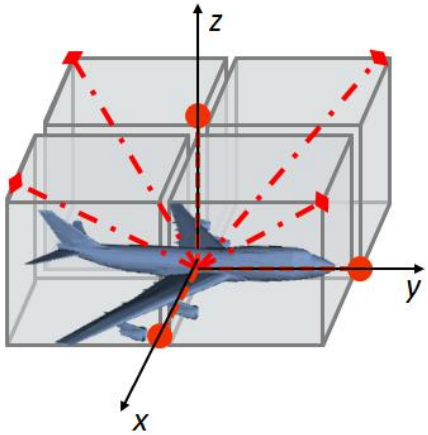

Figure 3. PCA-based positioning of the camera.

A second approach places the camera on the vertexes of a octahedron surrounding the 3D model. In order to obtain additional views, the edges of the octahedron are successively divided, resulting into 3, 9 and 33 vertexes (and implicitly the same number of views) (Figure 4). These strategies are called OCTA3, OCTA9 and OCTA33.

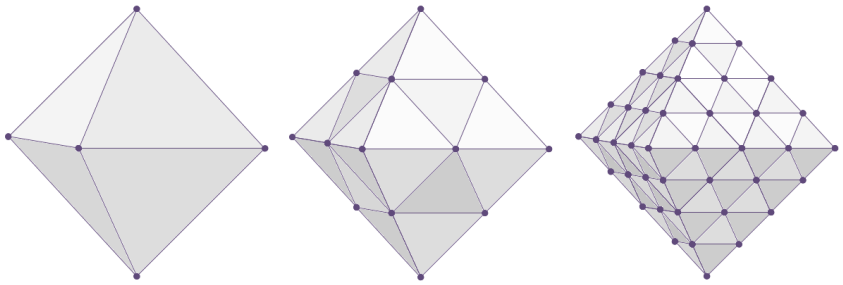

Figure 4. Successively subdividing the octahedron.

Furthermore, by using the vertices of a dodecahedron, we have placed the cameras uniformly around the canonical representation of the object (LFDPCA) (Figure 5.a). Finally, we have used the same repartition of the camera given by the octahedron, but we have applied a random rotation of the 3D model (LFD) (Figure 5.b). This choice is justified by the fact that the objects in real images are represented in a quasi-random pose.

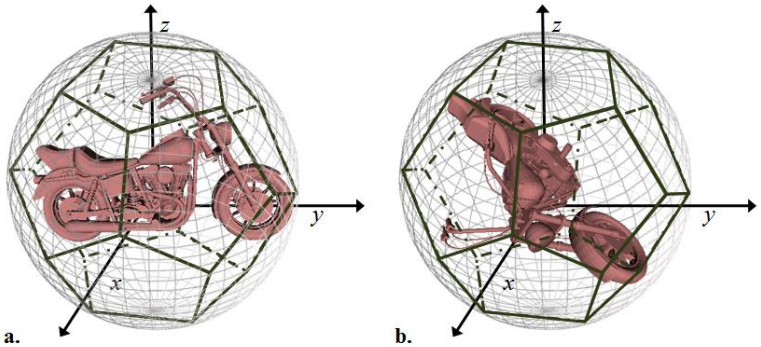

Figure 5. Dodecahedron-based positioning of the camera. a. The LFDPCA projection strategy. b. the LFD projection strategy.

Concerning the 2D shape descriptors, we have tested two contour-based descriptors and two region-based descriptors:

- The first one is the MPEG-7 Contour Shape Descriptor (CS)[4], [5] which stores the predominant curvature peaks (with curvature values and associated curvilinear abscises), robustly detected with the help of a contour scale space analysis [4]. When comparing two objects encoded by their CS descriptor, the distance is given by a matching procedure which takes into account the cost of fitted and unfitted peaks.

- We also propose in this paper a second contour-based shape descriptor, so-called Angular Histogram (AH). The AH descriptor represents, the angular distribution corresponding to three consecutive contour samples. If the sampling step is low, the samples are close one from the other and the resulting histogram depicts the local behavior of the object. On the contrary, when using higher sampling steps, the global properties of the object are obtained. The AH descriptor is composed of 5 angular histograms generated with different sampling steps. The associated similarity measure between two AH descriptors is the L_1 distance.

- The third is the MPEG-7 Region Shape descriptor (RS). Here, we make use of a basis of angular radial transforms (ART) [5], [6]. The image is decomposed on this basis and the first 34 decomposition coefficients are used as descriptor. In order to achieve rotation invariance, only the absolute value of these coefficients is stored. The distance between two RS representations is given by the L_1 distance between the 34-dimensional descriptors.

- Finally, we have tested the 2D Hough Transform (HT) [7] which describes a 2D shape as a cumulative distribution represented in the space of lines from the 2D plane. Here again the L_1 distance is employed in order to compute the distance between two descriptors. Because the HT is not invariant to rotation and mirror reflection, several relative positions between the two objects are tested. As a rotation in *xOy* space corresponds to a cyclic permutation in the HT space, the descriptor is extracted only once but a different distance is computed for each possible permutation. Finally, only the one giving the minimum distance is kept.

In order to evaluate which descriptor is more appropriate for 3D model-based object recognition and which is the best way of projecting a 3D model, we propose further an experimental evaluation of the above presented methods.

4. RESULTS

Experiments have been carried out on the MPEG-7 3D Model database [10] and on the Princeton Shape Benchmark (PSB) [8]. The MPEG-7 repository consists of 362 3D models divided in 23 categories (humanoids, pistols, rifles, helicopters, airplanes, trees with and without leaves, spherical object, fingers, cars, formula 1, tanks, trucks, chess pieces, cylindrical objects, motorcycles with 2 and 3 wheels, screwdrivers and letters A, B, C, D, E) (Figure 6). The PSB database includes 1814 3D models classified into 161 categories. Here, the proposed classification is more precise, having the class airplanes divided in biplanes, fighter jet, glider airplane, F117, commercial airplane... The PSB repository also includes models of animals, humanoids, body parts, furniture, plants, vehicles, sea vessels, tools, musical instruments, buildings… (Figure 7).

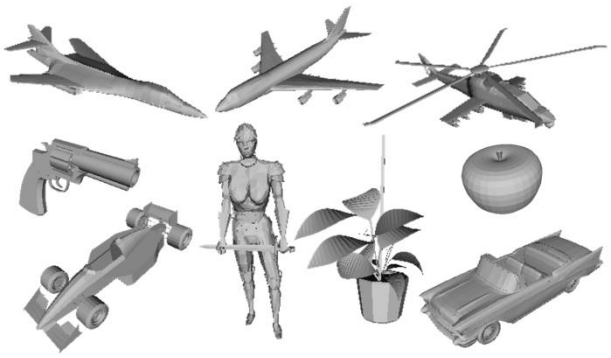

Figure 6. Example of models from the MPEG-7 database

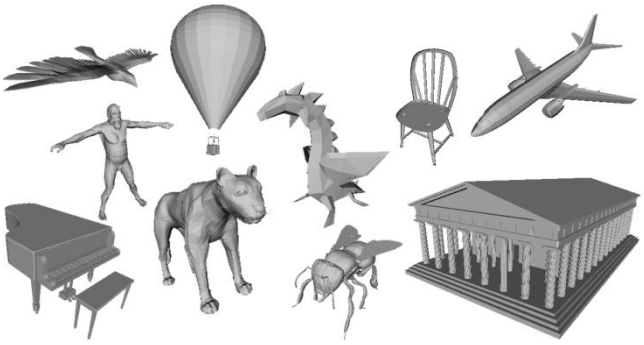

Figure 7. Example of models from the PSB database

We have also used a query database of 2D objects consisting in 115 images randomly chosen from the web (5 images for each MPEG-7 category). When the PSB database was employed, only 65 images were tested. The objects of interest were manually segmented from each image and stored as binary images (Annex A).

Each 2D query object is compared to the set of all the 3D models in the database. Based on distance measures in the space of the considered descriptors, the 3D models M_i are sorted in decreasing order of similarity to the query object O. The distance $d(O, M)$ between a 2D object O and a 3D model M is equal to the minimum distance between the object O and all the projections $P_i(M)$ of the model:

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

Further, we analyze which are the k (with k an integer number) most represented categories among the first top retrieved (up to a number Q) 3D objects. A recognition rate, denoted as $R(k)$ is defined for each number k, as the percentage of cases where the correct category of the query 2D object is determined within the first k most represented categories.

Tables 1 and 2 present the recognition rates obtained, for both considered 3D model databases. The reported recognition rates are $R(1)$, $R(2)$, $R(3)$. In the case of the Princeton database, where the number of existing categories is more important (161 classes), we have also reported the score $R(10)$.

We observe that in most cases the LFD and OCTA33 projection strategies achieve the maximal performances in terms of recognition rates, whatever the considered descriptors.

Thus, for the MPEG-7 database we respectively reach 60% recognition rate for the CS descriptor and 70.4% for AH. We have also tested a combination strategy in order to decide the retrieved categories by taking into account the results returned by multiples descriptors. This can help to exploit possible complementarities between descriptors. Thus, by combining the CS and AH descriptors (which provide the highest recognition rates) the global recognition rate slightly improves (72.2%).

In the case of the PSB database, the same global behaviors can be observed. Thus, the descriptors providing the highest recognition rates are here again CS and AH, with scores *R*(3) of 64.6% and 60%, respectively, when considering the OCTA33 projection strategy. When analyzing the scores *R*(10), we can observe that the recognition rates for the same descriptors increase up to 76,9%. Moreover, when combining AH and CS we obtain a *R*(10) score of 84,6%.

The obtained results show the interest of introducing in the object recognition process a 3D model-related information, exploited with the help of 2D/3D indexing techniques.

We believe that such information can be integrated within existing machine learning techniques dedicated to semantic inference objectives, in order to both speed-up the recognition process (*e.g.*, by reducing the number of candidate categories for a given query object) and to achieve more robust recognition rate b y combining both 2D and 3D information.

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 Figure 8.

Table 1. Recognition Rate for the MPEG-7 database

CS	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	33,9	34,8	37,4	33,9	33,9	37,4	37,4
R(2)	41,7	53,9	52,2	50,4	41,7	51,3	51,3
R(3)	53,9	61,7	**59,1**	60,0	53,9	56,5	**60,0**

RS	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	24,3	22,6	28,7	27,0	24,3	26,1	30,4
R(2)	36,5	37,4	40,9	37,4	36,5	42,6	46,1
R(3)	40,9	45,2	**46,1**	45,2	40,9	50,4	**54,8**

AH	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	30,4	35,7	44,3	42,6	30,4	32,2	38,3
R(2)	47,8	55,7	60,9	56,5	47,8	48,7	60,0
R(3)	56,5	61,7	**67,0**	62,6	56,5	60,0	**70,4**

H	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	18,3	20,9	27,0	24,3	18,3	28,7	34,8
R(2)	27,0	29,6	35,7	30,4	27,0	36,5	41,7
R(3)	37,4	37,4	**46,1**	35,7	37,4	43,5	**49,6**

CS+AH	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	34,8	39,1	41,7	39,1	34,8	37,4	40,9
R(2)	47,8	54,8	61,7	54,8	47,8	53,0	58,3
R(3)	56,5	60,0	**72,2**	60,9	56,5	64,3	71,3

Table 2. Recognition Rate for PSB

CS	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	32,3	41,5	40,0	41,5	32,3	41,5	44,6
R(2)	43,1	53,8	53,8	50,8	43,1	49,2	58,5
R(3)	49,2	58,5	**58,5**	55,4	49,2	56,9	**64,6**
R(10)	63,0	**76,9**	72,3	69,2	63,0	69,2	72,3

RS	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	26,2	20,0	23,1	24,6	26,2	29,2	32,3
R(2)	30,8	27,7	32,3	41,5	30,8	43,1	40,0
R(3)	38,5	35,4	38,5	41,5	38,5	46,2	**46,2**
R(10)	55,4	49,2	55,4	55,4	55,4	63,1	60,0

AH	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	27,7	40,0	40,0	36,9	27,7	35,4	44,6
R(2)	40,0	50,8	49,2	53,8	40,0	50,8	52,3
R(3)	49,2	55,4	52,3	58,5	49,2	**60,0**	53,8
R(10)	66,2	70,8	72,3	73,8	66,2	73,8	**76,9**

H	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	10,8	12,3	21,5	18,5	10,8	26,2	26,2
R(2)	12,3	15,4	32,3	23,1	12,3	32,3	33,8
R(3)	15,4	20,0	36,9	24,6	15,4	35,4	**40,0**
R(10)	30,8	35,4	41,5	29,2	30,8	49,2	52,3

CS+AH	PCA3	PCA7	LFD	LFDPCA	OCTA3	OCTA9	OCTA33
R(1)	35,4	49,2	41,5	47,7	35,4	43,1	49,2
R(2)	44,6	58,5	56,9	56,9	44,6	55,4	61,5
R(3)	55,4	67,7	58,5	58,5	55,4	63,1	**69,2**
R(10)	64,6	76,9	80,0	75,4	64,6	76,9	**84,6**

Figure 8. 2D/3D retrieval and categorization with the proposed system.

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

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.

5. CONCLUSION AND FUTURE WORK

In this paper we have presented a novel approach for 2D object semantic categorization, based on 2D/3D indexing methods. Different descriptors and projection strategies have been evaluated and compared. Experimental results have shown that CS and AH descriptors provide best recognition rate. In addition, their combination can further improve the corresponding recognition rates.

Our future work firstly concerns the extension of the proposed approach to 2D video objects. The multiple views of the query object available in this case can greatly help to disambiguate the recognition process. A second axis of research concerns the integration of complementary cues such as internal edge/contour information, since for the moment solely silhouettes are considered.

6. ACKNOWLEDGMENT

This work has been performed within the framework of the UBIMEDIA Research Lab, between Institut TELECOM and Alcatel-Lucent Bell-Labs.

REFERENCES

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

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

[3] 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.

[4] 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

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

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

[7] R. O. Duda and P. E. Hart. Use of the hough transformation to detect lines and curves in pictures. Commun. ACM, 15(1):11–15, 1972.

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

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

[10] 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.

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

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

[13] 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.

[14] 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.

[15] 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.

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

[17] 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.

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

[19] 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.

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

[21] 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.

[22] R. Mukundan and K. R. Ramakrishnan, "Moment Functions in Image Analysis: Theory and Applications", World Scientific Publishing Co Pte Ltd., September 1998.

[23] S. Zhang and G. Lu. "An Integrated Approach to Shape Based Image Retrieval", Proc. of 5th Asian Conferenceon Computer Vision (ACCV), pp. 652-657, Melbourne, Australia, January 2002.

[24] P.T.Yap, R.Paramesran and S.H.Ong, "Image Analysis by Krawtchouk Moments", IEEE Transactions on Image Processing, Vol. 12, No. 11, pp. 1367-1377, November 2003.

[25] Petros Daras, Apostolos Axenopoulos, "A Compact Multi-View descriptor for 3D Object Retrieval", International Workshop on Content-Based Multimedia Indexing, June 2009.

[26] T. Napoléon, T. Adamek, F. Schmitt,N.E. O'Connor, "Multi-view 3D retrieval using silhouette intersection and multi-scale contour representation", SHREC 2007 - Shape Retrieval Contest, Lyon, France, June 2007.
[27] Cyr and B. Kimia, "3D object recognition using shape similarity-based aspect graph", Proc. 8th IEEE Int. Conf. Comput. Vision, Vancouver, BC, Canada, pp. 254–261, 2001.
[28] H. Yamauchi, W. Saleem, S. Yoshizawa, Z. Karni, A. Belyaev, H.-P. Seidel, "Towards Stable and Salient Multi-View Representation of 3D Shapes", IEEE Int. Conf. on Shape Modeling and Applications, 2006, pp.40-40, 14-16, June 2006.
[29] Toshev, A. Makadia, and K. Daniilidis: Shape-based Object Recognition in Videos Using 3D Synthetic Object Models, IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, 2009
[30] J. Liebelt, C. Schmid, and K. Schertler. Viewpoint-independent object class detection using 3D Feature Maps. In IEEE CVPR, 2008.

ANNEX A

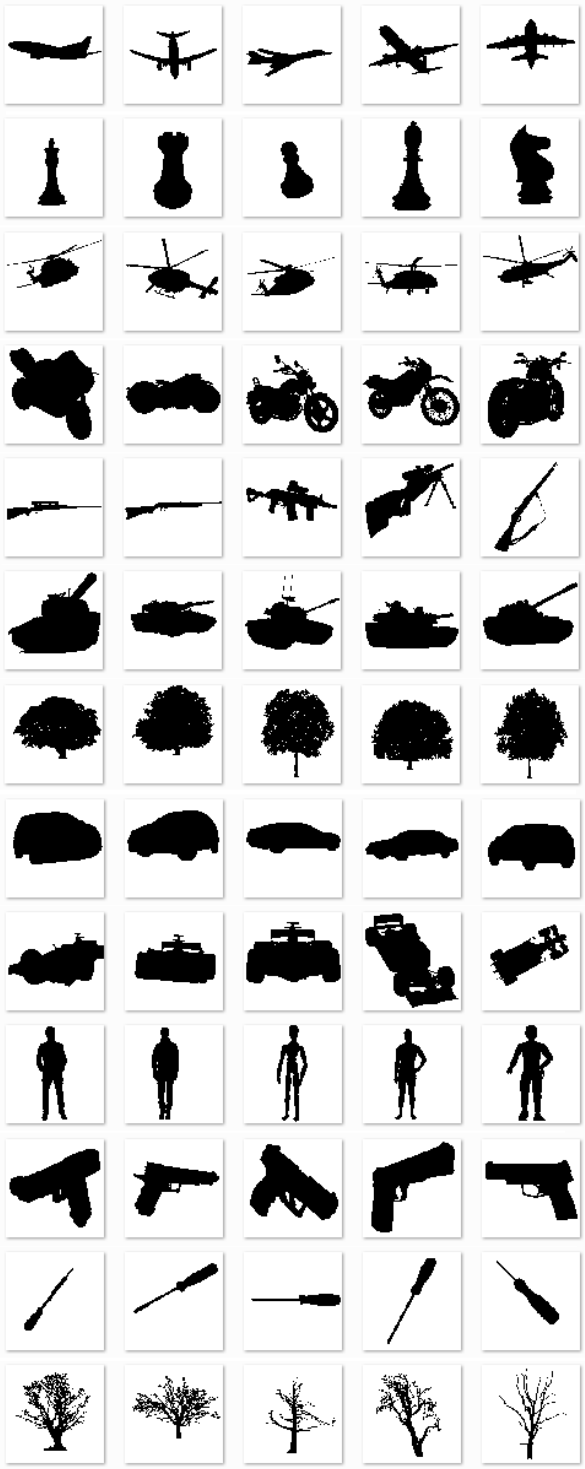

Figure A1. The 65 images tested with both PSB and MPEG-7 database.

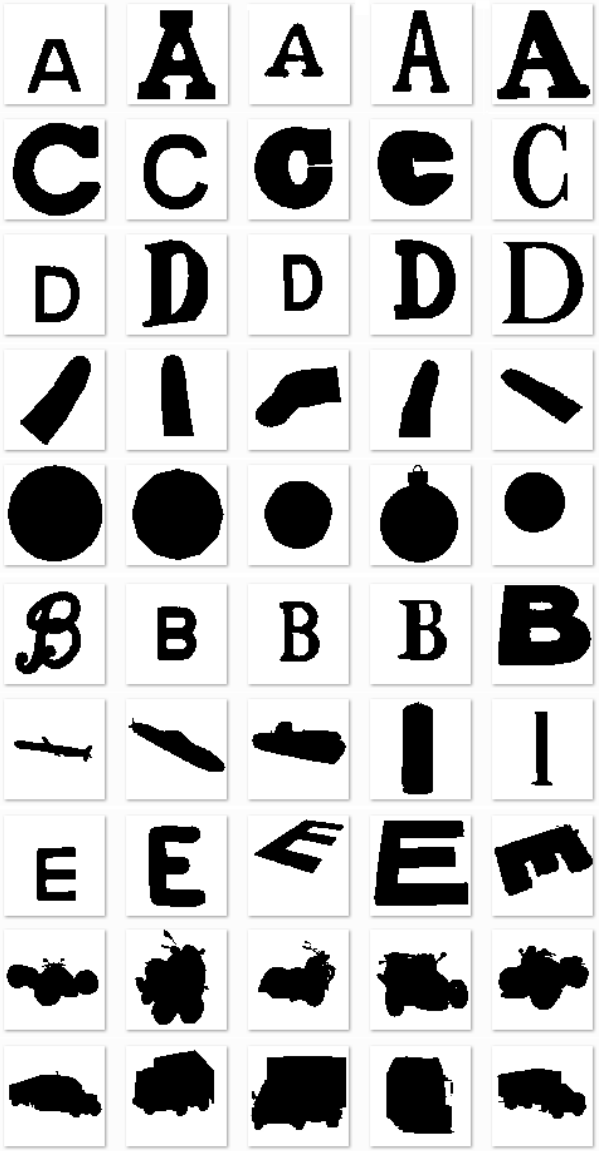

Figure A2. The 50 images tested only with the MPEG-7 database