An experimental evaluation of view-based 2D/3D indexing methods

Raluca Diana Petre, Zaharia Titus, Françoise Prêteux

▶ **To cite this version:**

Raluca Diana Petre, Zaharia Titus, Françoise Prêteux. An experimental evaluation of view-based 2D/3D indexing methods. 2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel, Nov 2010, Israel. pp.924-928, 10.1109/EEEI.2010.5661944 . hal-00738212

HAL Id: hal-00738212
https://hal.science/hal-00738212

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.

An Experimental Evaluation of View-Based 2D/3D Indexing Methods

Raluca-Diana Petre
Institut TELECOM;
TELECOM SudParis;9, rue Charles Fourier, 91011 Evry Cedex, France,
Alcatel-Lucent Bell-Labs
UMR CNRS 8145 – MAP5
Raluca-Diana.Petre@it-sudparis.eu

Titus Zaharia
Institut TELECOM;
TELECOM SudParis;
9, rue Charles Fourier, 91011 Evry Cedex, France
UMR CNRS 8145 – MAP5
Titus.Zaharia@it-sudparis.eu

Françoise Prêteux
Mines ParisTech
75272 Paris, France
Mines ParisTech; 60 Boulevard Saint-Michel, 75272 Paris, France
Francoise.Preteux@mines-paristech.fr

Abstract—This paper proposes an experimental evaluation of state of the art 2D/3D, view-based indexing methods. The principle of 2D/3D indexing methods consists of describing 3D models by means of a set of 2D shape descriptors, associated with a set of corresponding 2D views (under the assumption of a given projection model). Several experiments were conduced in order to examine the influence of the number of views and of the associated viewing angle selection strategies on the retrieval results. Experiments concern both 3D model retrieval and image recognition from a single view. Three 2D shape descriptors were tested in order to determine which of them is the most suited for such approaches. Results obtained show promising performances, with recognition rates from a single view higher than 80%, which opens interesting perspectives in terms of semantic metadata extraction from still images/videos.

Keywords: indexing and retrieval, 2D and 3D shape descriptors, multiview matching, similarity measures, MPEG-7 standard, 3D meshes.

I. INTRODUCTION

Recent innovation in computer imaging has lead to an increasing number of 3D models available for the general public. The tendency to personally model 3D shapes has been replaced nowadays by looking up the object in existing databases. As a result, the focus of the research community has shifted since to 3D model retrieval systems. The main goal of a retrieval system is to extract key features associated to a 3D object (*i.e.* shape, colour, texture, motion) that can be exploited to recover the desired information. Let us note that if not all models are textured, coloured or animated, they always define a 3D shape. Consequently, the great majority of retrieval methods are based on shape recognition techniques. For some comprehensive overviews on this topic, the reader is invited to refer to [1], [2], [3], [4].

A particular family of 3D retrieval methods use the so-called 2D/3D indexing algorithms, which employ a set of 2D representations associated to each model. More precisely, instead of directly describing the 3D content, such approaches exploit a set of 2D images, obtained by projection of the 3D model from several viewing angles. Various 2D shape descriptors are then used to describe each projection image. An outcome of this indexing approach concerns the possibility to compare 3D models not only with 3D models but also with objects detected in 2D images. In particular, this might be useful to derive semantic information from still images/videos.

The rest of the paper is organized as follows. Section 2 proposes a state of the art of 2D/3D shape retrieval approaches. Generic principle and main families of methods are here described and discussed. In Section 3, we propose an experimental protocol and evaluate, under the same experimental conditions (*i.e.* test set, 2D shape descriptor and similarity metrics), the impact of different factors (viewing angles selection, 2D descriptor and matching algorithm) on the performances of both 3D to 3D model retrieval and image recongnition from a single view applications. Finally, Section 4 concludes the paper and opens perspectives of future work.

II. STATE OF THE ART

Before presenting the state of the art approaches, let us recall the 2D/3D indexing methods principle.

A. Principle of 2D/3D indexing methods

The basic principle of the 2D/3D indexing methods relies on the following assumption: two 3D models that are observed from the same perspective present similar views. So, instead of describing the entire 3D objects, a set of images representing the model is selected and used to describe it. Thus, each 3D model M is projected and rendered in 2D using a set of N cameras. Each projection $\pi_i(M)$ is a binary image (also known as silhouette image) that is described using a 2D shape descriptor $d_i(M)$. The set $\{d_i(M)\}$ is further used to encode the model.

When using such an approach, different aspects need to be analyzed: the number of projections to be used, the camera positioning that ensures an optimal description of the object, the choice of the 2D descriptors, and the algorithm to be employed in the matching stage.

Also, invariance aspects with respect to the similarity transforms (*i.e.* rotation, translation, isotropic scaling) should be taken into account. Notably, the 3D rotation invariance is a critical issue since the projections can drastically change from a view to another. In consequence, the choice of the viewing angles is crucial for achieving a 3D pose invariance behavior. Due to this fundamental issue, we have chosen to categorize the state of

978-1-4244-8682-3/10/$26.00 ©2010 IEEE

the art 2D/3D indexing methods with respect to the viewing angle selection procedure involved. Two main families of approaches appear in literature. One very popular category of approaches involves a principal component analysis (PCA) of the 3D geometry, which is exploited in order to achieve a canonical representation invariant under 3D rotation. The second one uses an even distribution of the viewing angles around the object.

B. Methods using PCA-based projection

For the case of PCA-based methods the three principal axes and optionally the four secondary axes of inertia are used as viewing directions. As these axes are object-dependent, all the similar models will be represented by sets of images obtained from the same perspective, ensuring 3D rotation invariance.

Within the ISO/MPEG-7 framework [20], two different methods using seven PCA-based viewing angles were proposed. The first method employs 2D-ART descriptor [5]. When comparing two images, the L_1 distance between 34 ART coefficients is used. In order to compute the global distance between two models, all the matches between the two sets of images are tested and the one giving the minimal distance is selected. The second descriptor proposed by the MPEG-7 group is Contour Scale Space (CSS) [6]. The contour of the 2D shape is extracted and convolved with a set of Gaussian filters. The curvature peaks are thus robustly determined in a multi-scale analysis process. The associated similarity measure between two contours in CSS representation is based on a matching procedure, which takes into account the cost of fitted and unfitted curvature peaks. For the computation of the object-to-object distance, the same approach is used as in the case of 2D-ART-based method.

In [7] two methods are proposed. The first one is based on the Multi-scale curve convexity/concavity (MCC) representation [8]. The silhouette's contour is 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. Two points are similar if their change in position is similar. In order to achieve 2D rotation invariance, all the possible permutations between the two sets of points are tested. The second method proposed in [7] (so-called *Silhouette Intersection - SI*), defines the distance between two images as the area of the symmetric difference [9] of the two silhouettes. The approach is very intuitive and fast to compute, but poorly robust with respect to small variations of the shape or similarity transformations.

In [10] the authors proposed a method using the three projections corresponding to the principal directions determined by PCA. Each image is decomposed into *L=60* concentric circles. The percentage of the object's surface located in each circle is computed and used to form the feature vector. The circular decomposition of the object ensures 2D rotation invariance of the description. The so-called *Principal Plane Descriptor* (PPD) proposed is defined as the concatenation of three feature vectors.

PCA-based methods present the advantage of providing a canonical representation by using an object-dependent coordinate system obtained by applying the PCA. However, as shown in [11], [2], [22], the principal axes may present strong variations in the case of similar objects, resulting in miss-alignment problems.

A second family of approaches proposes to evenly distribute the viewing angles around the 3D object.

C. Methods using evenly distributed viewing angles

Instead of using preferential views depending on the object's geometry, the same set of camera is used for the acquisition of all the sets of images.

A method based on the SI algorithm [7], so-called *Enhanced Silhouette Intersection* (ESI) was proposed in [12]. The vertices of a regular dodecahedron are used as camera angles, resulting in a set of 10 silhouette images. Each of them is described by the SI descriptor, but when computing the distance between two sets of images the authors propose to weight the contribution of each projection by the use of a relevance index (proportional to the root square of the silhouette's area). The algorithm presents an increased robustness as compared to SI, but still remains sensitive to small shape variations.

In [13], authors introduce the *LightField Descriptor* (LFD), which also employs the vertices of a dodecahedron for the viewing angle selection. Each 2D shape is encoded by using Zernike moments [21] and Fourier Descriptors [23] coefficients. The LFD is obtained by used the coefficients resulted from all ten views. As the number of images is not enough to properly describe the 3D shape, 10 LFD are computed by placing the dodecahedron in different positions. A total of 100 images are acquired for each object, offering a very complete description. In the 3D object matching stage, all possible permutations between the two sets of 100 images have to be taken into account, leading to 5460 comparisons to be performed. An intelligent multi-step fitting approach is adopted in order to early reject non-relevant models. The LFD algorithm is reported as one of the most competitive 3D shape retrieval algorithms. However, the cost of such performance is an increased computational complexity.

Starting from the observation that two different objects can have similar projections, the Modified LFD (MLFD) technique is proposed in [14]. The authors skip the 2D size normalization step of the original LFD. The reported results have proved an increase of 4.1% and 3.6% for the First Tier, and Second Tier scores respectively.

In [15], the authors proposed a Compact Multi-View Descriptor (CMVD). Contrary to the other methods based on the even distribution of the camera, the CMVD algorithm starts with the PCA stage. Then, the vertices of a 32-hedron are used to place the camera, resulting in a set of 18 images for each model. Furthermore, each projection is described by using the coefficients form the 2D Polar Fourier Transform, 2D Krawtchouk moments [24] and 2D Zernike moments up to the order of 12, obtaining a total of 212 coefficients. Because in the preprocessing stage a canonical representation of the object was achieved, the authors propose not to test all the possible combinations between the two sets of images but only to take into account the 24 different alignments of the principal axes. Experiments performed on several 3D model databases show that CMVD performs similar to LFD while using a less time consuming algorithm.

All the above-presented methods consider an evenly distribution of the viewing angles over the unit sphere. However, for some regions of the unit sphere the projection images change very slowly in terms of visual aspect. Hence, a small number of images is sufficient to represent such regions. On the contrary, for other regions of the unit sphere the projections might change drastically even for small variations of the viewing angle. In order to eliminate redundancies of the representation, some methods propose to intelligently select a subset of images from all the

projections obtained by the evenly distribution of the camera viewing angles.

In [16], authors present a method based on a similarity aspect graph [17]. The camera is placed over one single plane and 36 projections are obtained by rotating the viewing direction by 5°. A similarity metric is computed between each two successive projections. A subset of representative images, so-called prototypes, is chosen such that each of them optimally represents a cluster of similar views, while maximizing the similarity with the other prototypes. The prototypes are further described by using the shock graph representation [18]. In the matching stage, the query is represented by a single projection, which is compared with all the prototypes of each model in the database. This algorithm is time consuming in both extraction and matching stages. In addition, its main drawback is related to the strong dependency of the associated performances on the viewing plane selected. A similar method is proposed in [19].

Compared with the PCA-based methods, the above-mentioned algorithms generate a higher number of projections (100 for LFD, 18 for CMVD). As a consequence, the computation complexity increases significantly, especially in the matching stage where all possible combinations between images from two sets need to be tested. Aspect graphs and stable view algorithms aim at reducing the number of views in an intelligent way by eliminating the redundant information. Thus, increasing the descriptor extraction complexity, the computation cost of the matching is reduced.

Table 1 synthesizes the various descriptors presented in this section. For each method, the extraction and the matching complexities, respectively denoted by C_E and C_M, are qualitatively estimated (*i.e.* + for low complexity and +++ for high). The numbers of views per model (N_V) as well as the viewing angle selection procedure are also indicated. The last column recalls the 2D descriptor used to describe the projection images.

The state of the art shows an impressive number of techniques. However, it is difficult to establish a fair comparison, since several aspects *(i.e.* viewing angle selection and number, 2D descriptor, matching strategy) can jointly have a relatively important impact on the retrieval performances.

Table 1. **Review of the 2D/3D indexing methods.**

Method	C_E	C_M	N_V	Viewing angle selection	2D descriptor
PPD [10]	+	+	3	PCA	Sums of pixels into concentric circles
MPEG-7 CSS [20]	+	++	7	PCA	Curvature scale space Descriptor
MPEG-7 ART[20]	++	++	7	PCA	Angular Radial Transform
MCC [7]	++	+++	3/9	PCA	Curve evolution at Gaussian filtering
SI [7]	+	+	3	PCA	Binary images
Aspect graph [16]	+++	+++	5 – 10	Aspect graph prototypes	Shock graph
ESI [12]	+	+	10	Even distribution	Binary images
LFD [13]	+++	+++	100	Even distribution	Zernike moments & Fourier descriptor
CMVD [15]	++	++	18	Even distribution	Zernike&Krawtchouk Moment and Fourier descriptor

Therefore, in the next section we propose an experimental evaluation protocol that aims to individually determine the influence of the viewing angle selection mechanism, of the 2D descriptors and of the matching algorithms involved on the retrieval results.

III. Experimental Evaluation

The experiments have been carried out on the MPEG-7 database, which includes 363 objects belonging to 23 categories (*e.g.* humanoids, airplanes, cars, trees, synthetic letters, rifles, missiles, pistols, helicopters etc.). The database presents important intra-class variety in terms of 3D shape (Figure 1).

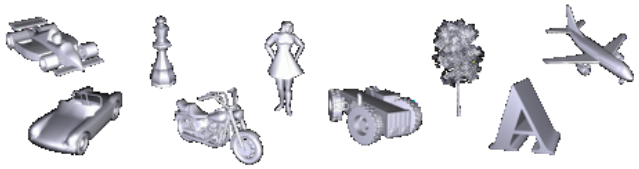

Figure 1: Sample models from the MPEG-7 3D dataset

Before the silhouette acquisition, each object was centered in the origin of the 3D coordinate system and resized w.r.t. its bounding box. For each image, six sets of projections were tested. Two of them are derived from the PCA-based alignment and contain 3 (PCA3) and 7 (PCA7) images per object corresponding to the three principal and four secondary axes of inertia. For the evenly distribution of the viewing angles, the LFD's positioning of the camera was tested. A random 3D rotation has been applied to each object, in order to ensure the objects are placed in an arbitrary position. In addition, we have proposed to test the same distribution for objects previously aligned with the use of the PCA (so-called LFDPCA distribution). Finally, the vertices of an octahedron (Figure 2) presenting different levels of subdivision were used to place the camera. At level zero, the projections correspond to the PCA3 ones, as the object was previously aligned with the coordinate system. From level one and level two of subdivision result 9, respectively 33 projections (OCT9 and OCT33).

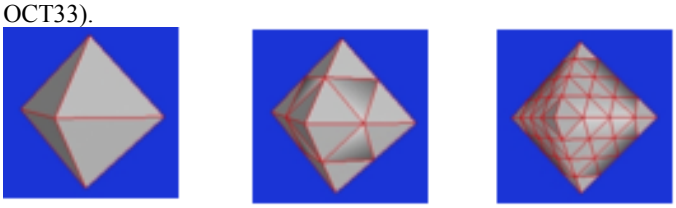

Figure 2: Succesive subdivion of the octahedron

Three different 2D shape descriptors were tested. Two of them are region-based and employ the decomposition of the image onto a set of functions; MPEG-7 2D-ART [5] with 34 coefficients and Zernike moments descriptors [21] with 42 coefficients. The third descriptor retained is the MPEG-7 Contour Scale Space (CSS) [20] which describes the 2D shape's contour.

A. 3D to 3D model retrieval

In order to evaluate the de performance of the retained 3D model retrieval systems two scores, the First Tier (FT) and the Second Tier (ST) have been computed. FT and ST respectively represent the percentage of correct answers among the first Q and 2Q first retrieved results (where Q is the size of the query model's category). Two matching strategies have been tested. The first one, so-called *minimum*, exploits a greedy strategy for fitting the various 2D views. When comparing two 3D models, the best match corresponding to the minimal distance between all combinations of views is first determined. The corresponding

views are considered as aligned and the process is successively applied upon the remaining sets of views until all the views are matched. The second strategy, so-called *diagonal*, assumes the PCA alignment is correct. Thus, each projection from the first set of images is associated to its corresponding view from the second set. For both methods, the global distance between two objects is equal to the sum of distances between the matched images.

Tables 2a-2c present the scores obtained using the MPEG-7 Contour Shape descriptor, the Region Shape descriptor and respectively the Zernike Moments descriptor. We observe that the MPEG-7 CSS representation provides the best results with a gain of 4% and 8% when respectively compared with the MPEG-7 2D-ART and Zernike moments representations.

When considering the viewing angle selection strategy, the same tendency appears no matter what descriptor is employed. Namely, the octahedron-based repartition provides the best results, with quasi-equivalent scores for 9 and 33 views. However, in all cases, the gain in the retrieval rates obtained by using a larger number of views remains marginal (about 3%). This shows that increasing the number of views does not necessarily significantly improve the performances. This fact can be explained by the fact that the number of false positives responses yielded by the 2D descriptors also increases with the number of views. In addition, in the case of objects presenting symmetries, several of the considered views can be similar, which can further influence the results.

Table 2: 3D model retrieval MPEG-7

Table 2a: Contour Shape Descriptor

Matching strategy	Score	PCA3	PCA7	LFD PCA	LFD	OCT9	OCT33
Minimum	**FT**	66.48	66.44	67.18	67.91	68.11	**70.24**
Diagonal		68.83	68.57	69.36	54.72	71.57	**71.79**
Minimal	**BE**	74.54	74.49	76.46	76.03	76.70	**77.06**
Diagonal		77.25	76.77	77.80	70.75	79.21	**79.61**

Table 2b: MPEG-7 Region Shape Descriptor

Matching strategy	Score	PCA3	PCA7	LFD PCA	LFD	OCT9	OCT33
Minimum	**FT**	63.65	64.10	63.72	59.87	**65.44**	65.21
Diagonal		65.57	66.11	65.37	45.19	66.57	**67.47**
Minimal	**BE**	71.86	73.15	71.61	70.66	73.73	**73.84**
Diagonal		73.76	73.54	73.50	58.77	**75.56**	74.71

Table2c: Zernike moments Descriptor

Matching strategy	Score	PCA3	PCA7	LFD PCA	LFD	OCT9	OCT33
Minimum	**FT**	61.47	60.56	58.90	55.86	**61.81**	58.60
Diagonal		61.50	62.22	59.69	41.21	**63.00**	61.11
Minimal	**BE**	69.93	68.61	67.69	65.15	**69.70**	67.02
Diagonal		70.36	69.73	68.33	54.31	**71.29**	69.69

In terms of matching strategy, we can observe that the *Diagonal* provides scores slightly superior to those obtained with the *Minimum* strategy, with an average gain of about 2%. Obviously, for the LFD approach, where no alignment was performed, lower scores are obtained with the *Diagonal* strategy which matches only corresponding views. The fact that the *Minimum* strategy provides lower scores can also be explained by the existence of false positives, which match similar views of non-similar objects.

The second part of our experiments concerns the 2D object recognition.

B. 2D object recognition from a single view

In this case, the query is represented as a single 2D silhouette image. In order to establish a query test set, we have retained 46 models, randomly selected from the MPEG-7 dataset, such that each category is represented by two objects. For each model, three query images, corresponding to different camera positions, have been generated. A first image is obtained as the object's projection onto the principal plane (denoted as PP). A second image (denoted as RAND) is obtained by using a randomly generated angle of view. Finally, a third projection image is obtained. The viewing angle used is also randomly generated, but it is restricted in a ±45° interval around the normal to the principal plane (denoted as RAND45).

Two scores were computed in order to evaluate the performance of the proposed 2D object recognition system. The first one, so-called *First Answer* (FA), provides the percentage of queries where the first retrieved result belongs to the correct category. The second score used, so-called *Recognition rate* (RR), is the percentage of queries to which the correct category is associated with (so-called, recognized category). The recognized is defined as the most represented category within the first 10 retrieved results.

Table 3: 3D model recognition from a single view.

Table 3a: Contour Shape Descriptor

Image type	Score	PCA3	PCA7	LFD PCA	LFD	OCT9	OCT33
PP	**FA**	73.91	73.91	**78.26**	73.91	76.09	76.09
RAND45		73.91	76.09	76.09	**80.43**	78.26	76.09
RAND		30.43	52.17	60.87	**71.74**	54.35	63.04
PP	**RR**	69.57	69.57	67.39	**71.74**	**71.74**	69.57
RAND45		63.04	**80.43**	78.26	71.74	73.91	67.39
RAND		21.74	52.17	**65.22**	56.52	50.00	56.52

Table 3b: Region Shape Descriptor

Image type	Score	PCA3	PCA7	LFD PCA	LFD	OCT9	OCT33
PP	**FA**	71.73	76.08	54.34	78.26	80.43	**82.60**
RAND45		45.65	52.17	60.86	69.56	56.52	**73.90**
RAND		21.73	47.82	56.52	65.21	47.82	**67.39**
PP	**RR**	63.04	63.04	47.82	63.04	**69.56**	**69.56**
RAND45		39.32	52.56	58.10	**67.78**	58.89	66.79
RAND		23.91	45.65	52.17	**56.52**	43.47	50.00

Table 3c: Zernike Moments Descriptor

Image type	Score	PCA3	PCA7	LFD PCA	LFD	OCT9	OCT33
PP	**FA**	78.26	76.09	69.57	76.09	76.09	**78.26**
RAND45		56.52	**71.74**	**71.74**	67.39	58.70	69.57
RAND		32.61	47.83	52.17	60.87	54.35	**67.39**
PP	**RR**	**65.22**	**65.22**	54.35	**65.22**	60.87	58.70
RAND45		41.30	54.35	56.52	**63.04**	54.35	52.17
RAND		28.26	43.48	52.17	**56.52**	39.13	52.17

Tables 3a-3c present the obtained results. In terms of the descriptor used, the same global tendency as in the case of 3D retrieval can be observed: the MPEG-7 CSS performs better than 2D-ART and 2D-ART is superior to the Zernike moments descriptor.

In most of the cases, the best FA and RR scores were obtained when using the PP image (about 80% for the FA and between 65.22% and 71.74 for the RR,). However, in the case of real objects detected in images or videos, the object is observed from

a random or quasi-random direction. For this kind of applications, the scores obtained using the RAND and RAND45 images are more pertinent. In this case, we observe that the LFD-based positioning of the camera offers a good solution, since the associated viewing angle distribution generates a significant gain (of about 10% in average when compared to the PCA7). The increase in views from 9 to 33 for the case of OCT9 and OCT33 is justified only when using the RAND query images, where a gain of 6% to 13% is obtained in terms of RR (7% to 20% in terms of FA).

The results obtained, with recognition rates of up to 80%, show that 2D/3D indexing approaches offer interesting perspectives to be exploited for object recognition and semantic enrichment of visual content.

IV. Conclusions and Future work

This paper presents an overview of the state of the art 2D/3D indexing methods as well as an experimental evaluation of methods. Two issues were considered: the 3D to 3D model retrieval and 3D model recognition from a single view.

The experimental protocol proposed aims at determining the influence of different factors (the viewing angle selection strategy, the number of views used to represent an object, the 2D descriptor used and the matching strategy) on the retrieval results.

Experiments, carried out on the MPEG-7 database, show that using a large set of views does not necessarily increase the retrieval/recognition performances. In particular, for model recognition from a single view, the even distribution of the viewing angles offered by the LFD approach is more appropriate. Concerning the choice of the 2D descriptor, the MPEG-7 CSS representation led to results superior than those offered by the two region-based descriptors considered (Zernike Moments and MPEG-7 2D-ART). In our future work we intend to test combinations of 2D descriptors. At a longer term we will focus our work on object recognition from real-life images/videos.

Acknowledgment

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

References

[1] Zheng Qin, Ji Jia, Jun Qin, "Content Based 3D Model Retrieval: A survey", *Content Based Multimedia Indexing 2008*, June 2008

[2] Johan W.H. Tangelder and Remco C. Veltkamp, "A Survey of Content Based 3D Shape Retrieval Methods", *Proceedings of the Shape Modeling International 2004 (SMI'04)*, pp. 145-156, Genova, Italy, 2004

[3] Y. Yang, H. Lin, Y. Zhang, "Content-based 3D Model Retrieval: A Survey", *IEEE Trans/ on Systems, Man, and Cybernetics – Part C: Applications and Reviews*, Vol. 37, No6, p. 1081-1098, November 2007.

[4] 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.

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

[6] 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

[7] 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.

[8] T. Adamek and N. E. O'Connor, "A multiscale representation method for nonrigid shapes with a single closed contour", *IEEE Trans. Circuits Syst. Video Techn*, Volume 14, Issue 5, pp. 742–753, May 2004.

[9] Helmut Alt, Ulrich Fuchs, Günter Rote, Gerald Weber, "Matching Convex Shapes with Respect to the Symmetric Difference", *Lecture*

[10] J.L. Shih, W.C. Wang, "A 3D Model Retrieval Approach based on The Principal Plane Descriptor", *Proceedings of The Second Internat. Conf. on Innovative Computing, Information and Control (ICICIC)*, pp. 59-62, 2007.

[11] T. Zaharia, F. Prêteux, "3D shape-based retrieval within the MPEG-7 framework", *Proc. SPIE Conf. on Nonlinear Image Processing and Pattern Analysis XII*, Vol. 4304, pp.133-145, San Jose, CA, USA, January 2001

[12] Liuying Wen, Guoxin Tan, "Enhanced 3D Shape Retrieval Using View-Based Silhouette Representation", International Conference on Audio, Language and Image Processing, August 2008.

[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] Tao Yang, Bo Liu, Hongbin Zhang, "3D model retrieval based on exact visual similarity", *9th International Conference on Signal Processing*, December 2008.

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

[16] C. 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.

[17] J. J. Koenderink, A. J. van Doorn, "The singularilarities of the visual mapping". *Biol. Cyber.*, Volume 24, pp. 51–59, 1976.

[18] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Alignment based recognition of shape outlines", *Proceedings of the 4th International Workshop on Visual Form*, May 2001.

[19] 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.

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

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

[22] T.Zaharia, F. Preteux, "Shape-based retrieval of 3D mesh models", *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on Volume 1*, pp. 437-440, August 2002.

[23] D. 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.