3D Objects Indexing and Retrieval Based On A New Efficient Optimal 2D Views Selection Method

M.A. Alaoui Mhamdi[1,4] A. Lachkar[2] S. El Alaoui Ouatik[3] D. Ziou[1]

[1] Département d'Informatique, Université de Sherbrooke, Canada

[2] Ecole Nationale des Sciences Appliquées de Fès, Morocco

[3] Faculté de Sciences Dhar El Mahraz, Fès, Morocco

[4] Laboratoire d'Informatique, Statistique et Qualité (LISQ), Morocco

Université sidi Mohamed Ben Abdellah,
Route d'Imouzzer B.P.2626 Fès – Morocco

ayoubalaoui@gmail.com, abdelmonaime_lachkar@yahoo.fr, S_ouatik@yahoo.com, Djemel.Ziou@usherbrooke.ca

Résumé

In this paper, we present a novel view-based approach for efficient 3D objects retrieval. A set of 2D images (multi-views) are automatically generated from the 3D object's views sphere approximated by a polyhedron subdivision loop scheme. We place the 3D object in it. To generate the initial views we place the camera on each of the triangle center of the polyhedron looking at the coordinate origin. For each 2D view associated with a triangle in the view sphere we apply the binarization to the 2D image and we extract the edge of the associated 2D shape. Afterwards, the most two similar adjacent triangles along edges are chosen. The Similarity among the 2D shapes is computed using our early proposed descriptor. Thus, we obtain a partitioned sphere into triangles regions. For each region in the views sphere; we place the camera at its associated center of mass (Local PCA) looking at the coordinate origin to take the most representative 2D view of all 2D views in it. The experimental results illustrate the efficiency of our proposed approach.

Mots clefs

Three-dimensional models, 3D Indexing and retrieval, Optimal 2D Views Selection, 2D shape descriptor,

1 Introduction

In recent years, with the significant advances in 3D acquisition and modeling, three-dimensional objects have become an important multimedia data type with many application possibilities. For example, 3D models can present complex information, and content-based searching problems in large 3D object repositories arise in many practical fields. Example application domains include CAD/CAM, Virtual Reality, medicine, molecular biology, military applications, and entertainment. In this context, content-based retrieval of 3D models has become an important subject of research. Several researchers have investigated the possibility of performing effective retrieval of 3D models from large archives, using shape properties instead of text. For efficient comparison and similarity estimation, 3D models can be represented with a set of meaningful descriptors that encode the salient geometric and topological characteristics of their shapes. The database objects are then ranked according to their distance to the descriptors of the query model. These descriptors can be global [1, 2], local [3, 4], structural [5, 6], transform-based approaches [7, 8] or by 2D/3D approach [9, 10]. This latter, consists to represent and describe a given 3D model by some 2D views. According to our investigation of the most well-known 2D/3D indexing methods proposed in the literature [9, 10, 11, 12, 13], we remark that they can be divided into two categories: in the first one, there is no automatic 2D views selection, in fact each 3D model is presented by a fixed number of 2D views [9, 10]. However, it is a major drawback that makes those methods inefficient. In the second one, the 3D model is presented by an "optimal" number of 2D views which are selected based on different used criteria's and different optimization algorithms [11, 12, 13]. Those methods will be detailed and criticized in the related works section.

The remainder of this paper is organized as follows. In Section II, the related works are presented and discussed. Our new proposed method is detailed and its robustness is discussed in Section III. Section IV presents the interpretation of the obtained results using our proposed method compared to some well-known view-based methods. Finally, conclusion is presented in Section V.

2 Related Works

Within the pattern-recognition and Computer Vision communities, the problem of defining representative 2D views for recognition and representation of 3D objects has recently received significant attention [14, 15].
The 3D shape can indirectly be represented by various 2D shape descriptors associated with projection images. There exist two categories: with fixed associated 2D views and with optimal 2D views selection.

2.1 Fixed associated 2D views

This kind of category consists on attributing a fixed views' number. Mahmoudi et al. [10] introduced a new method based on 2D/3D silhouettes computation. Each 3D object is represented by a set of *seven views*: the *first three* directions are determined by the PCA applied on the 3D object and the other *four* are deducted from the principal views. To index the *seven* silhouettes (Figure 1- (a)) describing the 3D object, they utilized CSS (Curvature Scale Space) organized around an M-Tree index structure. This descriptor characterizes the contour by exploiting the maxima of curvature, identified through a multi-scale analysis.
The original work of Vranic [16] proposed the first descriptor based on depth images, Depth-Based image Descriptor (DBD) was introduced by Heczko et al. [17]. To ensure the geometric invariance behavior, each 3D object is first determined after a PCA and normalized according to a cube with parallel axes to those intrinsic references of the 3D object. By projecting the model on the *six* faces of the cube (Figure 1- (b)), depth images in gray level are calculated and then transformed into Fourier space using the 2D-FFT. The signature of the 3D object is then determined by storing, for each image Fourier processed, the low-frequency coefficients.
Ohbuchi et al. [9] proposed a new descriptor called Multiple Orientation Depth Fourier Descriptor (MODFD). The authors used two stages of the pre-processing with PCA to obtain independent representation of both translation and scale. Invariance against rotation is ensured by calculating the depth images of the 3D model taken from *forty two* different viewpoints equally spaced on the unit sphere. These images cover all possible views of the model by discretizing the space (θ, φ) of the unit ball. Using the method of Zhang [18], each depth image in the Cartesian coordinates system (x, y) is transformed into an image of depth in the system of polar coordinates (ρ, θ) system using a polar map. This latter was then transformed into a Fourier image. They considered only low frequencies to represent the corresponding view. Finally, they got a vector for each view and a set of vectors for each 3D object. The similarity between a pair of MODFD is calculated using the distances between all possible combinations of two sets of vectors. The calculation of similarity is very expensive.
However, presenting a 3D object with a fixed number of 2D views can lead to some major limitations that depend of the 3D shape complexity. The first one when the 3D model is complex and contains more information it leads to the problem of *under views estimation*. In the opposite case, if the 3D model has not a complex structure it can lead to another problem of *over views selection*.

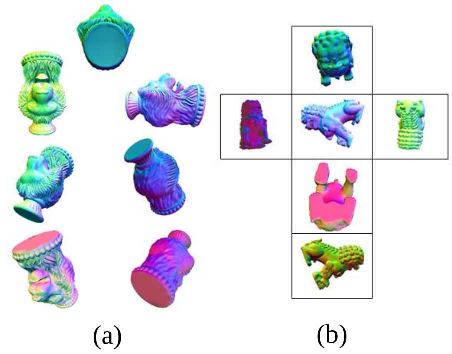

Figure 1 - *Fixed associated 2D views approach: (a) seven silhouettes, (b) six depth images.*

2.2 With optimal 2D views selection

To overcome those limitations of *under and over views estimation* leaded by the fixed 2D views number in the first category; the second one consist on automatic selection of the optimal 2D views.
Mokhtarian et al. addressed in [12] the issue of automatic selection of the best and the optimum number of 2D views for each 3D object. The object boundary of each 2D view is considered as a 2D shape and is represented effectively by less than ten pairs of integer values. These values include the locations of the maxima of its Curvature Scale Space (CSS) image contours. The CSS shape descriptor is expected to be selected for MPEG-7 standardization. They eliminated similar 2D views and selected a relatively small number of 2D views using an optimization algorithm. An unknown object is then recognized by a single image taken from an arbitrary viewpoint. Each object has been modeled using an optimized number of silhouette contours obtained from different viewpoints. This number varies from 5 to 25 depending on the complexity of the object and the measure of expected accuracy. However, it does not always give results aligned with human intuition. The main drawback of this method is its potentially high degree of ambiguity. The positions of zero-crossing point maxima for very deep and sharp concavities and for very long shallow concavities are almost identical. The convex parts of the curve are represented only implicitly by assuming that every concavity must be surrounded by two convexities.

Filali Ansary et al. [13] proposed a method for 3D model indexing based on 2D views, named AVC (Adaptive Views Clustering). The goal of this method is to provide an “optimal” selection of 2D views from a 3D model, and a probabilistic Bayesian method for 3D model retrieval from these 2D views. The characteristic views selection algorithm is based on an adaptive clustering algorithm (K-means) and used statistical model distribution scores to select the optimal number of 2D views. Starting from the fact that all views do not have equal importance, they introduced also a Bayesian approach to improve the retrieval. However, K-means and its variations present some limitations, when clusters have non-globular shapes or widely different sizes or densities. In addition, K-means is not efficient in the case of empty clusters. These factors decrease the accuracy of this 2D views selection method.

While most 3D object representations are complicated and inefficient, conventional multi-2D views representations are based on a large number of 2D views and cannot be used in many applications such as retrieval from large 3D objects databases. Multi-views representations have not yet successfully dealt with the following issues:

- What is the optimal number of 2D views?
- Do the extracted 2D views contain relevant information about the 3D object?

In our work, we consider that the problem of automatic 2D views selection can be divided into two underlying problems: the first one is the suitable used criteria and the second one is the use of an efficient optimization algorithm. Therefore, to enhance the performance of a given optimal 2D views selection; one can enhance both the performance of the used criteria and optimization algorithm or one of them.

For this aim, in this paper, we propose a new method based on two contributions. In the first one, we propose to use our early robust developed criteria [19] and in the second one, we suggest to use our new designed 2D optimal views selection algorithm.

3 The proposed method

The proposed method includes two contributions: the use of our robust developed criteria [19] and our new proposed 2D optimal views selection algorithm.

3.1 3D Pose Normalization

Note that, before applying our proposed 2D optimal views selection algorithm, a 3D PCA normalization must be performed in order to ensure invariance to the different geometric transforms. Indeed, 3D models have arbitrary position, orientation and scaling in 3D space. Since the extracted features are not invariant to position, orientation and scaling; to capture their invariant features, a feasible scheme is to place the model in a canonical coordinates frame to get the pose normalized. Then, a model is scaled, translated or rotated, the placing into the canonical frame. The pose normalization step is done through PCA [20, 21].

Let a 3D object defined by a triangular mesh M represented with a set T of n triangles $T = \{T_i, 1 \le i \le n\}$ and a set P of N vertices $P = \{P_i, 1 \le i \le N\}$. The covariance matrix C of the mesh M is approximated as follow:

$$C = \frac{1}{n}\sum_{i=1}^{n} S_i(g_i - m)(g_i - m)^T$$

Where S_i and g_i are the area of the i^{th} triangle of a shape and its center of gravity, m is the center of mass of the 3D model given by the formula:

$$m = \frac{\sum_{i=1}^{n} S_i g_i}{\sum_{i=1}^{n} S_i}$$

And n is the number of triangles of the 3D object. The process of scaling to a unit sphere (Figure 2 - (a)) is applied before the 3D alignment using the following formulas:

$$D = \max_{i=1,\ldots,N} d(m, P_i)$$

$$P^{'} = \left\{P_i^{'} | P_i^{'} = \frac{1}{D} P_i, P_i \in P, i = 1, \ldots, N\right\}$$

Where N is the number of vertices of the 3D object and P_i its i^{th} vertex.

The 3D alignment (Figure 2 - (c)) step must be performed after centering and scaling the 3D model in the centered unit sphere. Obviously the matrix C is a real symmetric one, therefore its eigenvalues are non-negative real numbers. Then we sort the eigenvalues in non-increasing order and find the corresponding eigenvectors. The eigenvectors are scaled to Euclidean unit length and we form the rotation matrix R which has the scaled eigenvectors as rows. We rotate all points in $P^{''}$ and a new point set is formed:

$$P^{''} = \{P_i^{''} | P_i^{''} = P_i^{'} R, P_i^{'} \in P^{'}, i = 1, \ldots, N\}$$

(a) (b) (c)

Figure 2 - *(a) Centering and Scaling to the Unit Sphere, (b) and (c) are respectively 3D objects before and after PCA Alignment*

3.2 Our Robust 2D Shape Descriptor

Our robust 2D shape descriptor is based on multi-scale analysis. Let $f(u) = \langle (x(u), y(u)) | u \in [0, T] \rangle$ be the parametric representation for a given curve of shape, where T is it arc-length, and u is the curvilinear abscise. And let $\langle g(u,\sigma)|\sigma > 0\rangle$ be set of Gaussians, where for a given σ, $g(u,\sigma)$ is given as follow : $g(u,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-u^2}{2\sigma^2}}$.

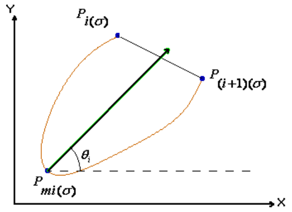

Figure 3 - *Token representation and orientation θ_i.*

The set of the smoothed curves $\langle f(u,\sigma)|\sigma \geq 0\rangle$, which are obtained by the convolution of $f(u)$ with the set of Gaussians $g(u,\sigma)$. For different value of σ is a multiscale representation of the curve of shape $f(u)$. A set of a multiscale curvature $K(u,\sigma)$ that corresponds to the set of curve of shape $\langle f(u,\sigma)|\sigma \geq 0\rangle$ is defined in [23] as follow:

$$K(u,\sigma) = \frac{x_t(u,\sigma)y_{tt}(u,\sigma) - x_{tt}(u,\sigma)y_t(u,\sigma)}{\left(x_t^2(u,\sigma) + y_t^2(u,\sigma)\right)^{3/2}}$$

Where x_t, y_t and x_{tt}, y_{tt} are respectively the first and second derivatives of x and y with respect to t.

Let $P = \{P_i(\sigma)\}_{i=1}^{N}$ be the set of minima that is the set of points such as $K(u,\sigma) = 0$. If we assume that the curvature $K(u,\sigma)$ is continuous, between two consecutive minima $P_i(\sigma)$ and $P_{i+1}(\sigma)$, there is always a maximum of $f(u)$, namely $m_i(\sigma)$, located at the point $P_{m_i}(\sigma)$. For each value of σ, a smoothed curve of shape is obtained, which is decomposed into portions or tokens according to the points P_i. Each token i of the curve $f(u,\sigma)$ is described by the vector $E_{(i\sigma)}(m_i(\sigma), O_i(\sigma))$ (Figure 3), with $m_i(\sigma)$ in [-180, 180] is the curvature at point $P_{mi}(\sigma)$, and $O_i(\sigma)$ in [0, 360] is the orientation defined in polar coordinates of the vector linking the median point of the segment $P_i(\sigma) - P_{i+1}(\sigma)$ with the point $P_{mi}(\sigma)$. Our descriptor is invariant to translation and scale; to assure its rotation invariance, we proposed the use of the principle of force equilibrium. Let $\Gamma_0 = \{\vec{f_i} | 1 \leq i \leq N\}$ be a set of features vectors of a given 2D shape where N is the number of vectors. If $\sum_{i=1}^{N} \vec{f_i} \neq \vec{0}$, then, there exists a vector $\vec{F_0}$ verifying $\sum_{i=1}^{N} \vec{f_i} + \vec{F_0} = \vec{0}$. The equilibrium vector is called the *principal vector* and its direction according the axis OX is called the *principal direction*. The principle consists on computing the principal direction θ and rotating all tokens' orientation vectors by $-\theta$ to let the principal vector of the features vectors of each 2D shape coincides with the OX axis.

3.3 The new proposed optimal 2D views selection algorithm

After performing the 3D normalization using PCA of the 3D targeted object, we construct a unit sphere with a regular mesh using two iterations of the Loop subdivision scheme on an initial tetrahedron to obtain 64 faces. We place the 3D object in it. To generate the initial 2D views, we place the camera on each of the triangle center looking at the coordinate origin. For each 2D view associated with a triangle in the view sphere we apply the binarization to the 2D image and we extract the edge of the associated 2D shape. Each triangle in the views sphere has an associated 2D view and has three adjacent triangles along its edges, thus, the similarity measure is computed between its 2D view and each of the three 2D views associated to its three adjacent triangles, the most two similar adjacent triangles along its edges are chosen. The Similarity among the 2D shapes associated to the 2D views is computed according to the chosen criteria (2D shape descriptor based Contours or Regions). In our case, it is based on our early developed robust shape descriptor [19]. Thus, we obtain a partitioned sphere into triangles regions. For each region in the view sphere; we place the camera at its associated center of mass (Local PCA) looking at the coordinate origin to take the most representative view of all views in it.

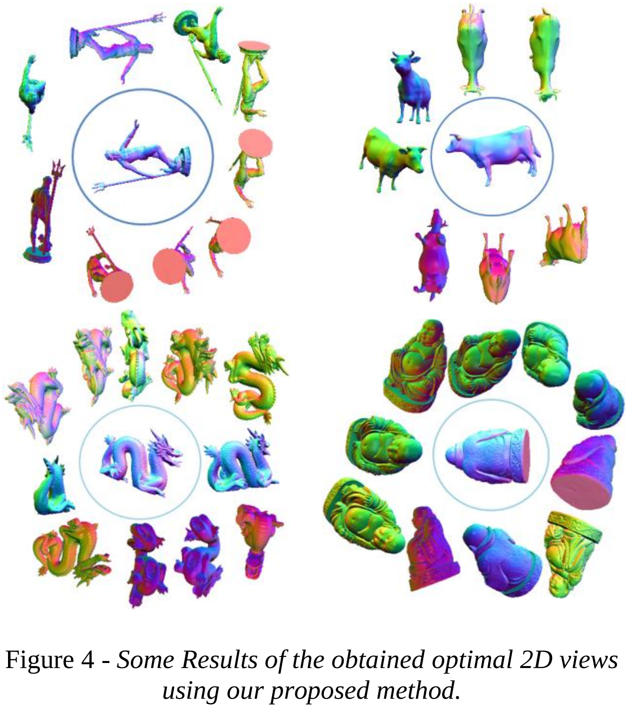

Figure 4 - *Some Results of the obtained optimal 2D views using our proposed method.*

4 3D/3D Matching

Let now A and B be two 3D models, with features vectors F_t^A and F_t^B respectively, and $F_t^A = \bigcup_{i=1}^{N_a} f_i^A$ and $F_t^B = \bigcup_{i=1}^{N_b} f_i^B$ where N_a and N_b are the numbers of the 2D

shapes associated to A and B respectively and f_i^A and f_i^B are the i^{th} shapes' descriptor of A and B respectively.
For an efficient matching procedure among the set of shapes of A and B; in the first step, we compute the distance between an i^{th} shape associated with A and every shape associated with the model B. The smallest of the computed distances is the distance $d(A_i, B)$ given by the following formula:

$$d(A_i, B) = \min_{1 \le j \le N_b} D(f_i^A, f_j^B)$$

$D(f_i^A, f_j^B)$ is the Minkowski distance computed between an i^{th} shape associated with A and a j^{th} shape associated with B according to the our early developed robust 2D shape descriptor [19].
Thus, the distance between the model A and B is given by the following formula:

$$d(A, B) = \frac{1}{N_a} \sum_{i=1}^{N_a} d(A_i, B)$$

5 Experimental Results

The proposed method was experimentally evaluated using the Princeton Benchmark Database [22]. It is one of standard databases of 3D models available on the Web by the team "Princeton Shape Retrieval and Analysis Group" to let researchers evaluating their 3D indexing algorithms on the same 3D database. The Princeton database contains 1814 3D models grouped in high-level semantics classes where the objects of the same class are heterogeneous. For example, the insects' class contains 3D models which represent insects of different shapes but with the same semantic.
To evaluate the proposed method, each 3D model was used as a query object. The retrieval performance was evaluated in terms of "precision" and "recall", where precision is the proportion of the retrieved models that are relevant to the query and recall is the proportion of relevant models in the entire database that are retrieved in the query.

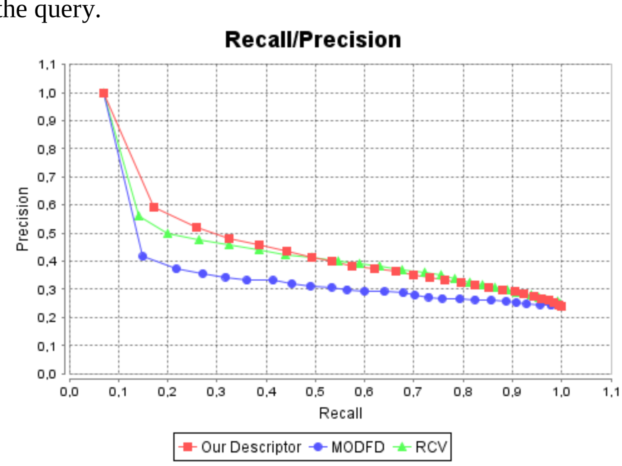

Figure 5 - *Recall/Precision curve evaluating the performance or our new proposed method.*

In order to evaluate our proposed approach, we compared it to the following two methods, which are based on a fixed 2D views number:

- Retrieving 3D shapes based on their appearance [9] that proposed Multiple Orientation Depth Fourier Descriptor (MODFD).
- Retrieval by shape using characteristic views (RCV) [10].

The retrieval experimental results (Figure 5) illustrate the efficiency of our proposed method over similar view-based methods. They showed the importance of local information and the efficiency of our early developed robust 2D shape descriptor.
From the figure 5, when the recall is between 0% and 50%; the most of the retrieved 3D models belong to the same class of the 3D query object in the case of our method compared with the descriptors MODFD and CSS. Which shows that the 2D extracted views using our proposed optimal views selection method, contains more relevant information about the 3D targeted object.
The robustness of our proposed method is from its invariance to the different geometric transforms (translation, scale and rotation). The limitations of our proposed method result in:

- The case of the 3D articulated objects, due to the tokens' orientation change.
- The high computational cost during the 2D views selection procedure, especially where there is a huge amount of local features to be extracted and the multi-scale analysis of each 2D shape worsens it more.
- Our proposed 2D views selection method doesn't take into account the human perception factor.

6 Conclusion and Future works

In this paper, we proposed a new method for efficient 3D model retrieval which contains two contributions: the first one is the use of our robust developed criteria and the second one is our new proposed 2D optimal views selection. Our new algorithm of optimal view selection showed it efficiency of extracting 2D views with more relevant information of 3D objects.
The retrieval experimental results showed efficiency and superiority of our method compared to other well-known methods based on fixed 2D views number. The strength of the proposed method is its robustness in terms of information relevance contained in the 2D extracted optimal views for a given 3D object.
In our future work, we are planning to improve our optimal views selection algorithm in order to take into account human perception factors. We are working on designing a well-suited index structure to improve the

speed response of our 3D developed Search Engine. In addition to using a 3D model (given a priori) as the query, we would like to add 2D sketch, in order to apply our method on various interactive applications, 3D face recognition, occlusion problem, classification of marine species.

Acknowledgements

The authors would like to acknowledge the Stanford Data Archives, AIM@SHAPE Shape Repository and the Princeton Shape Benchmark database [22] for the 3D models.

The authors are also grateful for the constructive and valuable comments from the reviewers of this paper.

References

[1] CHEN, D.-Y., TIAN, X.-P., SHEN, Y.-T., AND OUHYOUNG. On visual similarity based 3D model retrieval. *Computer Graphics Forum* 22, 3, 223–232, M.2003.

[2] KAZHDAN, M., FUNKHOUSER, T., AND RUSINKIEWICZ. Rotation invariant spherical harmonic representation of 3D shape descriptors. In *SGP '03: Proceedings of the 2003 Eurographics/ ACM SIGGRAPH symposium on Geometry processing*, 156–164, S. 2003.

[3] C. Chua and R. Jarvis. Point signatures: A new representation for 3D object recognition. *International Journal of Computer Vision*, 25(1):63–85, 1996.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002.

[5] HILAGA, M., SHINAGAWA, Y., KOHMURA, T., AND KUNII. Topology matching for fully automatic similarity estimation of 3D shapes. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM Press, 203–212, T. L. 2001.

[6] BIASOTTI, S., MARINI, S., SPAGNUOLO, M., AND FALCIDIENO. Sub-part correspondence by structural descriptors of 3D shapes. *Computer-Aided Design* 38, 9, 1002–1019, B. 2006.

[7] Novotni, M. Klein, R. Shape retrieval using 3D Zernike descriptors. *Computer Aided Design*, 36(11):1047–1062. 2004.

[8] Vranic, D. V. Saupe, D. 3D Shape Descriptor Based on 3D Fourier Transform. *EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS '01)*, pages 271–274. Budapest, Hongrie. 2001.

[9] R. Ohbuchi, M. Nakazawa, and T. Takei. Retrieving 3D shapes based on their appearance. In *5th ACM SIGMM Workshop on Multimedia Information Retrieval (MIR 2003)*, Berkeley, California, USA, November 2003.

[10] S. Mahmoudi and M. Daoudi. Retrieval by shape using characteristic views. *The International Conference on Pattern Recognition ICPR02*, pages 457-460, August 11-15, 2002.

[11] J. Lee, B. Moghaddam, H. Pster, and R. Machiraju. Finding optimal views for 3D face shape modeling. In *International Conf. on Automatic Face and Gesture Recognition*, pages 31–36, 2004.

[12] F. Mokhtarian and S. Abbasi. Automatic selection of optimal views in multi-view object recognition. In *The British Machine Vision Conf. (BMVC'00)*, pages 272–281, 2000.

[13] T.Filali Ansary, M. Daoudi, J-P Vandeborre, A Bayesian 3D Search Engine using Adaptive Views Clustering. *IEEE Transactions On Multimedia*,Vol. (9), Issue (1) , Page 78-88, January 2007.

[14] T. Denton, M. Demirci, J. Abrahamson, A. Shokoufandeh, and S. Dickinson. Selecting canonical views for view-based 3-D object recognition. In *17th International Conf. on Pattern Recognition (ICPR'04)*, volume 2, pages 273–276, August 2004.

[15] P. M. Hall and M. J. Owen. Simple canonical views. In *The British Machine Vision Conf. (BMVC'05)*, volume 1, pages 7–16, 2005.

[16] D. Vranic, *3D Model Retrieval*, Ph.D. dissertation, University of Leipzig, 2004.

[17] M. Heczko, D. A. Keim, D. Saupe, and D. V. Vranic. Verfahren zur Ähnlichkeitssuche auf 3d objekten (methods for similarity search on 3D databases). *Datenbank-Spektrum*, 2(2):54-63, 2002.

[18] D. S. Zhang and G. Lu. Shape-based image retrieval using generic Fourier descriptor. *Signal Processing: Image Communication*, 17(10):825-848, November 2002.

[19] H. Silkan, A. Lachkar, S. E. Ouatik, and A. Elkharraz. Rotation Invariant and Robust Shape Descriptor for Content-Based Shape Retrieval. *Information and Communication Technologies International Symposium ICTIS'07 IEEE*. Morocco, Fez, 3-5 April 2007.

[20] Chen, D.Y. and Ouhyoung, M.: A 3D model alignment and retrieval system. *Proc. of International Computer Symposium*, 2002.

[21] Jolliffe, I.T. *Principal Component Analysis*. Springer, 1986.

[22] Shilane, P., Min, P., Kazhdan, M. et Funkhouser, T. The Princeton Shape Benchmark. *Shape Modeling International (SMI '05)*, pages 167–178. Genève, Italie. pages xii, 56, 57, 113. 2004.

[23] Farzin Mokhtarian, Alan K. Mackworth. "A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves". IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 14, Issue 8. Pages: 789 – 805. August 1992.