HAL
open science

Enhanced 2D/3D Approaches Based on Relevance Index for 3D-Shape Retrieval

Mohamed Chaouch, Anne Verroust-Blondet

To cite this version:

Mohamed Chaouch, Anne Verroust-Blondet. Enhanced 2D/3D Approaches Based on Relevance Index for 3D-Shape Retrieval. SMI 2006 - International Conference on Shape Modeling and Applications, Jun 2006, Matsushima, Japan. 10.1109/SMI.2006.11 . hal-00802602

HAL Id: hal-00802602

https://inria.hal.science/hal-00802602v1

Submitted on 20 Mar 2013

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.

Enhanced 2D/3D Approaches Based on Relevance Index for 3D-Shape Retrieval

Mohamed Chaouch, Anne Verroust-Blondet
INRIA Rocquencourt
Domaine de Voluceau, B.P. 105
78153 Le Chesnay Cedex, FRANCE
{Mohamed.Chaouch,Anne.Verroust}@inria.fr

Abstract

We present a new approach for 3D model indexing and retrieval using 2D/3D shape descriptors based on silhouettes or depth-buffer images. To take into account the dispersion of information in the views, we associate to each view a relevance index which will be afterward used in the dissimilarity computation. The performance of this new approach is evaluated on the Princeton 3D Shape Benchmark database.

1 Introduction

The problem of developing robust and efficient tools for searching a shape in a database of 3D models has become a real challenge, considering the number of 3D models available on the Internet. We have chosen to study here 2D/3D retrieval methods based on the computation of the 2D views of the 3D models. When examining the different views associated to a 3D object, one can see that some of them are more relevant than other ones. In the following, we propose to take into account the relevance associated to the 2D views in the shape matching algorithm. We introduce several relevance index models to measure the amount of significant information associated to a 2D view and show their efficiency on two 2D/3D retrieval approaches based respectively on silhouettes and on depth-buffer images. In section 2 we present an overview of previous work on 3D similarity and shape retrieval, focusing on 2D/3D approaches. Our method and its evaluation on the Princeton 3D Shape Benchmark Database are presented respectively in section 3 and 4. Finally, we conclude the paper and present ideas for future work in section 5.

2 Previous work

Existing shape benchmarks [16, 8, 20, 3] propose databases of polyhedral models. We will here focus on 2D/3D approaches and let the reader consult [19, 18, 26, 2] for a more complete overview of 3D retrieval methods.

2.1 2D/3D approaches

The main idea consists in associating to the 3D model a set of projections from multiple viewpoints. The 3D shape is indirectly represented by various 2D-shape descriptors so that the 3D-shape matching is transformed into similarity measuring between 2D-shapes. This principle is common to all 2D/3D approaches. The difference between these approaches lies in the choice of: the pose normalization, the box of projection, the size of the images, the nature of the 2D-shape descriptors and the size of the signatures.

Silhouettes [5, 12, 3, 13, 1] and depth-buffer images [7, 14, 21] are the most widespread projections. A silhouette indicates the region of a 2D-image which contains the projection of the visible points of the 3D object. Although the silhouette images are easy to store and very useful in some cases, they contain only little information about the 3D shape. Then, to collect more topological and structural properties of the 3D model, it proves to be interesting to use depth-buffer images.

In [3], the authors present a representation of 3D model by ten silhouettes rendered from a set of views covering the extended sphere. Next, each silhouette is encoded by its Zernike moments and Fourier descriptors. The dissimilarity of two 3D models is defined as the minimum of the sum of the distances between features extracted over all rotations and all pairs of vertices on the corresponding dodecahedrons. Filali Ansary et al. [1] propose an approach based on characteristic views selection and introduce a novel probabilistic Bayesian approach for 3D model retrieval from these views. In [12], Mahmoudi and Daoudi consider 7 silhouettes to represent 3D model. To describe these silhouettes, they use curvature scale space (CSS) 2D descriptors which provide compact and highly discriminating representations. In [5], each object is associated to 72 views which are structured in a shock graph. The shape retrieval process is based on a shock graph matching algorithm. Min [13] proposes a model adapted to a 2D user sketch query interface.

Ohbuchi et al. [14] employ a set of 42 depth-buffer images to represent each 3D model. To ensure that their method is invariant under rotation, they introduce a new algorithm for measuring similarity between two 3D models. In [21], a global optimization algorithm is employed to estimate the best relative pose.

2.2 Contributions of the views in the 3D-shape description

By examining the 2D images associated to a 3D model, we notice that some views are more relevant than the others in the 3D-shape description. For example, if one looks at the 2D-shapes of Figure 2, he will quickly recognize a human on the frontal view, a little less on the profile view. The upward view taken in an isolated way will not be more significant. Thus if we measure the similarity between two models by visual comparison, we have to take it into consideration. In the following, we present a process which estimates differently the contribution of each image in the 3D-shape description according to its relevance.

3 Similarity search of 3D models

To respect the diversity of information among various views, we will associate to each silhouette or depth-buffer image a relevance value which will be used in the estimation of the degree of similarity between two 3D objects. This approach can be applied to any type of 2D/3D descriptors. Here, we consider two 2D/3D shape descriptors based on the Fourier transform of rendered silhouettes and depth-buffer images of the 3D models, presented in [7]. These descriptors require pose normalization to provide invariance under similarity transformations. Thus, there are four major steps for measuring the similarity between 3D models:

1. Normalize pose of 3D model. The Principal Component Analysis (PCA) generates an alignment of a 3D-mesh model into a canonical coordinate system frame by translating, rotating, reflecting, and scaling the set of vertices. We have retained the "Continuous" PCA [24, 25] because it appears to be more complete and the most stable of all the approaches we have studied.

2. Extract feature vectors (2D/3D descriptors).

3. Determine relevance indices of projection images.

4. Compute the dissimilarity between 3D models.

3.1 2D/3D Descriptors

Two 2D/3D shape descriptors of [7] have been retained in our 3D model retrieval process. They both need a CPA normalization process.

The first one uses silhouettes. Each 3D model is projected perpendicularly on the planes of his own canonical bounding cube, in order to generate three silhouette images. Each silhouette $\mathbf{s_i} = \{s_i(a,b);\ s_i(a,b) \in \{0,1\}, a = 0,...,N-1, b = 0,...,N-1\}, i \in \{1,2,3\}$, is defined as an outer contour approximated by K equally-spaced sequential points $\mathbf{c_i} = \{c_i(t); t = 0,...,K\}, i \in \{1,2,3\}$. The c_i are selected, computing the intersection of the contour with a set of rays emanating from the image center O and having a uniform angular distribution. They form the input to the one-dimensional discrete Fourier transform (1D-DFT). The shape feature vector of a 3D model is formed by the low frequency part coefficients of the Fourier spectrums, $\mathbf{F}^i$, of the three contour images.

The second 2D/3D shape descriptor is based on depth-buffer images. Six depth-buffer images are associated to the faces of the extended bounding box. The $6\ N$xN pixels images $\mathbf{v_i} = \{v_i(a,b);\ v_i(a,b) \in [0,1], a = 0,...,N-1, b = 0,...,N-1\}, i \in \{1,...,6\}$, with $N = 2^l$, are transformed using the two-dimensional discrete Fourier transform (2D-DFT) to represent the feature in the spectral domain. Finally, each depth-buffer image $\mathbf{v_i}$, $i \in \{1,...,6\}$, is represented by the absolute values of k low-frequency Fourier coefficients, $\mathbf{F}^i$, and constitute the feature vector of dimension $6k$.

3.2 Relevance index models

The relevance index indicates the density of information contained in the 2D image and is calculated for each view characterizing the 3D model. Since the optimal way of measuring the relevant information about 2D-shape is not prescribed, we can consider a variety of different methods to define relevance index. It depends on the nature of information extracted from the 2D-shape such as area, contour, curvature, depth, structure, connectivity, compactness... In this section, for only one information type, we define one relevance index. However, to capture another relevance over different aspects and characteristics of a 2D-shape, we can combine different relevance indices described in what follows. To compute the relevance indices, we can use several statistical measures such as average or variance.

3.2.1 Silhouette relevance indices

We propose two methods to compute the relevance indices of silhouette images:
The first is classic: we compute the number of non-null pixels on the image, i.e., the area of the projected surface of the 3D model on the corresponding face of the bounding box.

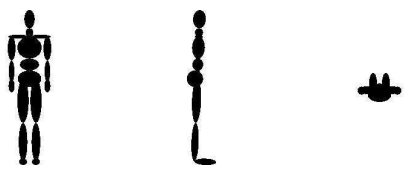
$$R_a = card\{s_{ab} | s_{ab} = 1, 0 \leq a, b \leq N-1\}, \qquad (1)$$

where s_{ab} is the pixel value of the image at position (a,b). We can also retain this relevance value for depth-buffer im-

ages, keeping only depth pixels with positive values.

Figure 1. Limitation of the relevance interpretation (using R_a).

Using the area of the projected surface, the results will give the smallest relevance values for the images of the second row of Figure 1, but it is obvious that the corresponding views are more relevant than on the first row. Then the R_a is not suitable for measuring the relevance index in some cases. To moderate the influence of the area, we can take the square root of the relevance defined in Equation (1) as relevance index, $R_{sa} = \sqrt{R_a}$ (see the values of R_{sa} for the human biped model in Figure 2).



46.81 31.40 21.79

Figure 2. Three silhouettes images of Human-biped model and their R_{sa} (%) relevance values.

The second relevance index model introduces the outer contour of the silhouette in order to keep more 2D shape information. Hence, the contour length is not suitable to represent the relevance index, because it is sensible to local perturbation of the surface. In general, similar shapes can have different contour lengths (see the examples in Figure 3). Therefore, to define a more robust relevance index which depends on the outer contour, we choose the average cord of a 2D mesh, i.e. the average of the lengths of all possible cords connecting two contour points.

$$R_c = \sum_{a=0}^{N-1}\sum_{b=0}^{N-1}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} \frac{\delta_{c_{ab}.c_{pq}}}{L(L-1)}\sqrt{|a-p|^2+|b-q|^2} \tag{2}$$

$$\delta_{x.y} = 1 \text{ if } x = y = 1 \text{ and } \delta_{x.y} = 0 \text{ otherwise.}$$

where c_{ab} is the pixel value at position (a, b) and L is the contour length. The pixels where $c_{ab} = 1$ define the outer contour of silhouette.

To overcome the limitation presented in Figure 3, we can also define the relevance index using a set of particular points detected from the 2D-shape. The considered points can be the higher-curvature points or a subset of interest points as in [11]. However, this information type can well represent the relevance for some particular cases but can be much less efficient for many 3D models due to the unstable behavior of these key points.

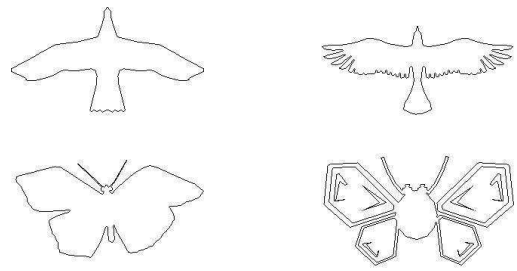


Figure 3. Similar 2D shapes with different contour lengths.

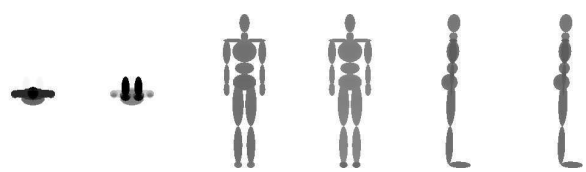
3.2.2 Depth-buffer relevance indices

We propose two methods to compute the relevance indices of depth-buffer images:

The first one introduces the depth. We take the sum of all values of the non-null pixels on the depth-buffer image (we compute the volume enclosed between the visible parts of the 3D object and the opposite plane of the bounding box).

$$R_d = \sum_{a=0}^{N-1}\sum_{b=0}^{N-1} v_{ab}, \tag{3}$$

where v_{ab} is the pixel value of the depth-buffer image at position (a, b).



8.99 8.34 27.61 26.60 14.23 14.23

Figure 4. R_g (%) relevance values for the depth-buffer images of the human-biped model.

The last relevance is more global. It is the sum of the distances between the center of mass of the 3D model and all visible points of the 3D model:

$$R_g = \frac{1}{2w}\sum_{a=0}^{N-1}\sum_{b=0}^{N-1} d_{ab}, \tag{4}$$

$$d_{ab} = \sqrt{|a-N/2|^2+|b-N/2|^2+2w|v_{ab}-1/2|^2},$$

where $2w$ is the length of the sides of the extended enclosing bounding box. The values of R_g for the human-biped model are given in Figure 4.

3.2.3 Normalized relevance indices

Let R^i be the relevance indices associated respectively to the views $1 \leq i \leq N_v$. We use the normalized relevance indices $\bar{R}^i$:

$$\bar{R}^i = R^i / \sum_{i=1}^{N_v} R^i. \quad (5)$$

In fact, the proposed relevance indices depend on the nature of retained information from the 2D-shape. Consequently, for one database, an appropriate selection of the relevance, depending on the present models, could yield a better diversity of information among various views.

3.3 Matching two 3D models

To compare two 3D models $O1$ and $O2$, we generate the feature vectors $\mathbf{F}_1^i$ and $\mathbf{F}_2^i$, $1 \leq i \leq N_v$ of the silhouettes or depth-buffer images. Then, we compute the relevance indices of the images associated to each object, $\bar{R}_1^i$ and $\bar{R}_2^i$, $1 \leq i \leq N_v$, where N_v is the number of views. Finally, the dissimilarity between a pair of 3D models $O1$ and $O2$ is:

$$\Delta(O1, O2) = \sum_{i=1}^{N_v} d(\bar{R}_1^i \mathbf{F}_1^i, \bar{R}_2^i \mathbf{F}_2^i). \quad (6)$$

We take $N_v = 3$ (resp. $N_v = 6$) for the silhouette (resp. depth-buffer) based approaches.

4 Experimental results

For our tests, we use the Princeton 3D Shape Benchmark database (see [18] for a detailed description of the database). The experimental results presented here were obtained with the following parameters: 256x256 size for the projection images; 100 (resp. 78) low-frequency Fourier coefficients for one silhouette (resp. depth-buffer) image; l_1 distance for shape similarity computation. In what follows, SA and DBA will denote the methods without relevance index, ESA and EDBA will denote our enhanced approach based on silhouettes and depth buffer.

To study the ability of each method to discriminate between classes of models, we performed a more general search experiment over the entire database. In this case, for each approach, all similarities between two models from the database are calculated. Each object in the database is selected as query object and the models are retrieved according to the rank of computed similarities.

In order to choose the best relevance index for the Princeton 3D Shape Benchmark database, we tested all proposed relevance indices with the studied approaches (cf. table 1). According to our experiment, R_{sa} the square root of the relevance defined in Equation (1) for silhouettes and R_g defined in Equation (4) for depth-buffer images were slightly better than the others. In the rest of the paper, we present the results of enhanced approaches with them.

Method	NN	FT	ST
ESA R_a	52.59	29.15	39.40
ESA R_{sa}	55.46	29.64	40.02
ESA R_c	54.69	29.59	39.23
EDBA R_a	60.12	34.18	43.87
EDBA R_d	61.78	34.69	43.85
EDBA R_g	60.53	34.40	44.37

Table 1. Comparing the enhanced 2D/3D approaches with different relevance indices using the NN, FT and ST global averages (%).

To compare objectively the retrieval effectiveness of the proposed approaches, we computed Precision-Recall diagrams (the query is not counted in the answer as in [23]) and four quantitative measures for evaluating query results (see [18] for a description of this measures): **(1)** The Nearest neighbor (NN), **(2)** The First-tier (FT), **(3)** the Second-tier (ST), **(4)** The Discounted Cumulative Gain (DCG).
The NN, FT, ST and DCG measures and Precision-Recall diagrams of Figure 5 show that the approaches based on relevance index outperformed the classic approaches. Moreover, if we compare these measures with the results listed in table 4 of [18], the DBA is in third position, after the Radialized Spherical Extent Function (REXT) descriptor [22] while the EDBA is in second position, between the Light Field Descriptor (LFD) [17] and the REXT descriptor.

To illustrate how relevance indices improve the 2D/3D approaches, we present four queries of the desk chair class in Figure 6. In fact, our retrieval system returns a complete list of models ranked by similarity to the query model. Only the nine top matches from it are retained in Figure 6. The query model is at the top of Figure 6 and the retrieved models are presented in decreasing similarity order from left to right and top to down. Each returned model is represented by its most significant depth-buffer image. The EDBA retrieves seven relevant models, one more than the DBA results and the ESA gives the same number of desk chair than the SA. Note that the nine first models returned by the EDBA and the ESA methods are chairs, contrarily to the SA and the ESA methods which also return cars and plants. For all queries illustrated, we notice that the first models retrieved by our methods are visually very similar to the query in term of global shape. These results show the contribution of the relevance index in the studied 2D/3D approaches. Moreover, it emerges from queries that our approaches are more effective than the approaches without relevance index to retrieve the 3D models belonging to the query classes, and return the objects that we would intuitively consider as similar.

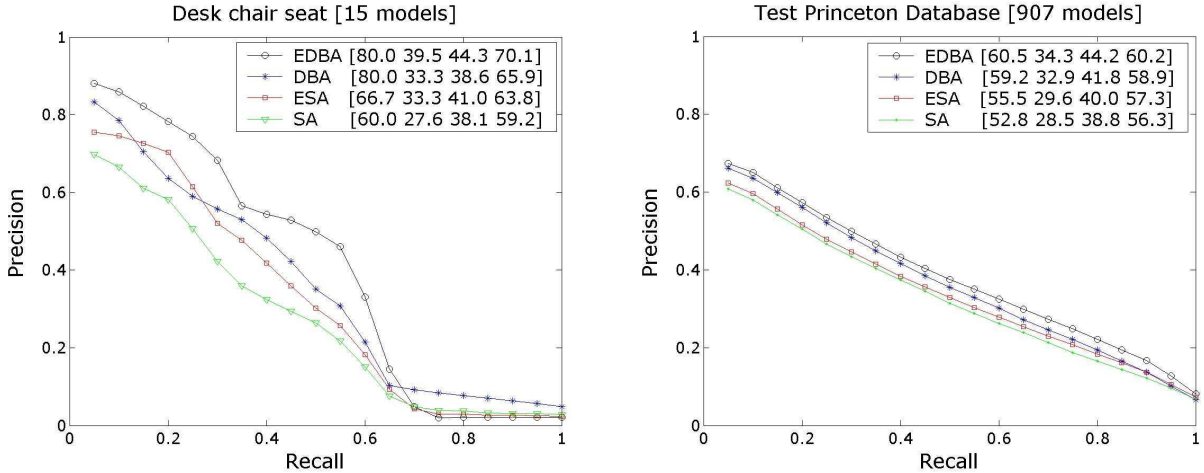


Figure 5. Average Precision-recall curves for the desk chair class and the global database using the two proposed approaches and the two initial approaches without relevance index. The mean NN, FT, ST and DCG values are given in the legends.

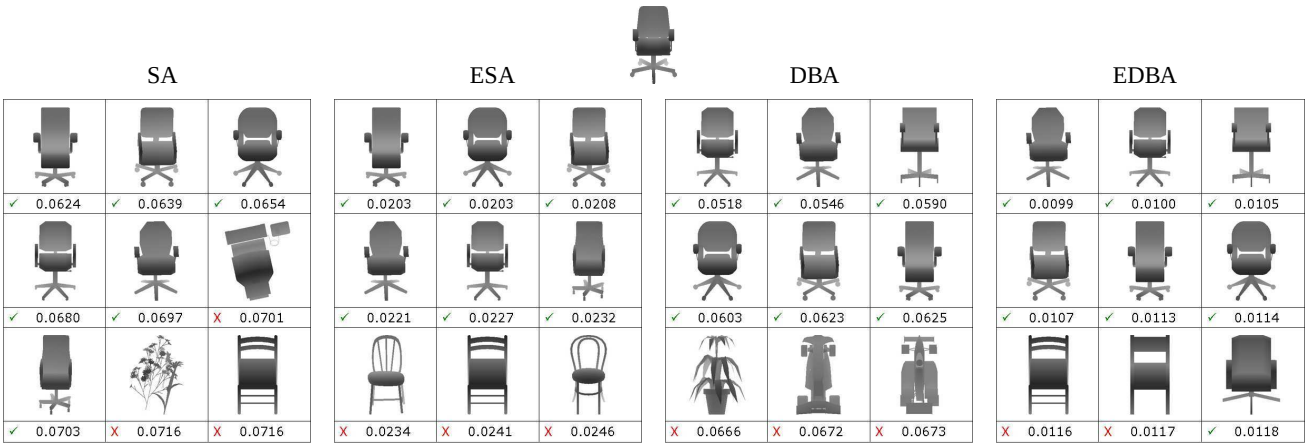


Figure 6. Top 9 matched objects with ESA, ESA, DBA and EDBA methods. The similarities between the query model and the retrieved models are given below corresponding images. ✓ and x indicate that the retrieved models belong or don't belong to the query's class, respectively.

5 Conclusion

In this paper, we have presented a technique to enhance 2D/3D approaches for 3D-shape retrieval. We have associated a relevance index to each 2D image to provide a more accurate description of the 3D models and used them in the dissimilarity measurement process. Several relevance index models have been proposed for silhouette and depth-buffer based descriptors. Their efficiency has been compared with the silhouette and depth-buffer approaches of [23] on the Princeton 3D Shape Benchmark Database. Our NN, FT, ST and DCG global average scores produce better results for both silhouette and depth-buffer based methods and thus our method improves 2D/3D approaches for 3D-shape retrieval without increasing their overall computational cost.

It has been shown by Shilane et al. [18] that the Lightfield Descriptor [3] was the most efficient 3D retrieval method. Then it would be interesting to measure the effect of introducing relevance indices in their approach.

In fact, the relevance index models presented here are empirical: their value should represent the "density" of information contained in the image and should express how much the image characterizes the 3D-shape w.r.t. the other images associated to the 3D model. Other features such as the 2D-shape curvatures should be considered to design new relevance index models for silhouette and depth-buffer based approaches. Improvement in the design of relevance index models could be done also by studying human visual object-recognition system as it has been done in [4, 6, 9] for

3D model recognition purpose. A relevance index model based on saliency could be deduced from [10] for depth buffer images. Nevertheless, it seems difficult to find the perfect relevance index model: the paper of Polonsky et al. [15] which compares various methods computing the best viewpoint to visualize a 3D object indicates that any of them is good for a number of examples and fails for others. Although our purpose here is slightly different, we think that we should obtain similar results for relevance index models.

Acknowledgements

We would like to acknowledge the anonymous reviewers for their helpful comments and suggestions. This work has been supported in part by the DELOS Network of Excellence on Digital Libraries (EU IST Network of Excellence G038-507618).

References

[1] T. F. Ansary, M. Daoudi, and J.-P. Vandeborre. 3D model retrieval based on adaptive views clustering. In *3rd International Conference on Advances in Pattern Recognition (ICAPR 2005)*, Bath, United Kingdom, Aug. 2005.

[2] B. Bustos, D. A. Keim, T. Schreck, and D. Vranic. An experimental comparison of feature-based 3D retrieval methods. In *Second International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'04)*, Thessaloniki, Greece, Sept. 2004.

[3] D. Chen, X. Tian, Y. Shen, and M. Ouhyoung. On visual similarity based 3D model retrieval. *Computer graphics forum*, 22(3):223–232, Sept. 2003 (http://3d.csie.ntu.edu.tw/).

[4] F. Cutzu and M. Tarr. The representation of three-dimensional object similarity in human vision. In *SPIE Proceedings from Electronic Imaging: Human Vision and Electronic Imaging II, 3016,*, pages 460–471, San Jose, CA, 1997. SPIE.

[5] C. Cyr and B. Kimia. 3D object recognition using shape similarity-based aspect graph. In *International Conference on Computer Vision (ICCV)*, pages 254–261, 2001.

[6] W. Hayward and P. Williams. Viewpoint dependence and object discriminability. *Psychological Science*, 11(1):7–12, 2000.

[7] 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.

[8] U. Konstanz 3D models similarity search engine. http://merkur01.inf.uni-konstanz.de/cccc/.

[9] R. Lawson, H. Bülthoff, and S. Dumbell. Interactions between view changes and shape changes in picture - picture matching. *Perception*, 32:1465–1498, 2003.

[10] C. Lee, A. Varshney, and D. Jacobs. Mesh saliency. *ACM Transactions on Graphics*, 24(3):659–666, 2005.

[11] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[12] S. Mahmoudi and M. Daoudi. 3D models retrieval by using characteristic views. In *The International Conference on Pattern Recognition (ICPR 2002)*, pages 11–15, Québec, Canada, Aug. 2002.

[13] P. Min. *A 3D Model Search Engine*. PhD thesis, Princeton University, Jan. 2004.

[14] 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, Californica, USA, Nov. 2003.

[15] O. Polonsky, G. Patanè, S. Biasotti, C. Gotsman, and M. Spagnuolo. What's in an image? *The Visual Computer*, 21(8-10):840–847, 2005.

[16] U. Princeton 3D models search engine. http://shape.cs.princeton.edu/search.html.

[17] Y. Shen, D. Chen, X. Tian, and M. Ouhyoung. 3D model search engine based on lightfi eld descriptors (poster). In *Eurographics 2003*, Granada, Spain, Sept. 2003.

[18] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The Princeton shape benchmark. In *Shape Modeling and Applications Conference, SMI'2004*, pages 167–178, Genova, Italy, June 2004. IEEE.

[19] J. Tangelder and R. Veltkamp. A survey of content based 3D shape retrieval methods. In *Shape Modeling and Applications Conference, SMI 2004*, pages 145–156, Genova, Italy, June 2004. IEEE.

[20] U. Utrecht 3D shape retrieval engine. http://www.cs.uu.nl/centers/give/imaging/3Drecog/-3Dmatching.html.

[21] N. Vajramushti, I. A. Kakadiaris, T. Theoharis, and G. Papaioanno. Effi cient 3D object retrieval using depth images. In *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR' 04)*, pages 189–196, New York, Oct. 2004. ACM Press.

[22] D. Vranic. An improvement of rotation invariant 3D-shape descriptor based on functions on concentric spheres. In *ICIP 2003*, pages 757–760, Barcelona, Spain, Sept. 2003.

[23] D. Vranic. *3D Model Retrieval*. PhD thesis, University of Leipzig, 2004.

[24] D. Vranic, D. Saupe, and J. Richter. Tools for 3D-object retrieval: Karhune-Loeve transform and spherical harmonics. In *2001 Workshop Multimedia Signal Processing*, Cannes, France, Oct. 2001.

[25] D. V. Vranic and D. Saupe. 3D shape descriptor based on 3D Fourier transform. In K. Fazekas, editor, *EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS 2001)*, pages 271–274, Budapest, Sept. 2001.

[26] T. Zaharia and F. Prêteux. 3D versus 2D/3D shape descriptors: A comparative study. In *Proceedings SPIE Conference on Image Processing: Algorithms and Systems III - IS & T / SPIE Symposium on Electronic Imaging, Science and Technology '03, San Jose, CA*, volume 5298, Jan. 2004.