DESIRE: A COMPOSITE 3D-SHAPE DESCRIPTOR

Dejan V. Vranić

Dejan.Vranic@gmx.de

ABSTRACT

The topic of this communication is shape-similarity search for 3D-mesh models. We present and evaluate a composite 3D-shape feature vector (DESIRE), which is formed using **de**pth buffer images, **si**lhouettes, and **r**ay-**e**xtents of a polygonal mesh. We contrast our method with the approach that is declared the best in the recent study. Our experiments suggest that the composite feature vector, which is extracted in a canonical coordinate frame, generally outperforms the competing method, which relies upon pairwise alignment of models. We also provide a Web-based retrieval system as well as publicly available executables for verifying the results.

1. MOTIVATION

The area of 3D-model retrieval attracts more and more researchers. A variety of methods for characterizing 3D-shape have been proposed in recent years. Several surveys of 3D-shape description techniques (e.g., [7]) summarize used ideas, without comparing competing methods quantitatively. In [1], 12 different methods for describing shape are compared on the PSB (Princeton Shape Benchmark) set of 3D-objects, presented by the Princeton Shape Retrieval and Analysis Group. The LightField descriptor (LFD) [2] is declared as superior method in [1]. In [5], 19 different 3D-shape feature vectors are compared on four different test sets (including the PSB), and a composite feature vector outperforms all competitors. Unfortunately, neither the LFD is tested in [5] nor the composite descriptor is tested in [1]. Therefore, the main objective of this paper is to try to find out which of the two approaches is better. In order to enable verification of our results, we provide executables for extracting feature vectors and source code for performance analysis [8].

2. RELATED WORK

In this section, we describe the LightField descriptor [2], which is used in the experiments and discussion (sections 4 and 5). For descriptions of several dozens of 3D-shape retrieval techniques, we refer to [1,5,7].

The extraction of the LFD begins with normalization step in which translation (the center of gravity becomes the origin) and scale (the maximum distance of a point on the surface of the model to the origin becomes 1) are fixed. Then, 100 silhouette images from predefined viewpoints are generated. The 100 cameras are positioned at vertices of 10 dodecahedrons. The distribution of the cameras is nearly uniform. Each of 100 silhouette images is described by 35 coefficients for Zernike moments (also used in MPEG-7) and 10 for Fourier coefficients as proposed in [6]. Thus, the LFD descriptor consists of 4500 components. Rotation invariance is achieved by aligning two models (a query and a candidate) by finding the best correspondence of 10 sets of silhouette images taking into account all possible rotations of dodecahedrons. There are 5460 different rotations, which are necessary to examine in order to align a pair of 3D-objects. When the models are aligned, the dissimilarities between 10 most-corresponding silhouette descriptors are summed-up and regarded as dissimilarity between the models.

Since the maximal distance is used for fixing the scale, we expect potential problems with outliers. In order to demonstrate the problem, we used a model of car with a long outlying antenna as the query and the original executables provided by the authors of the LFD to retrieve (extract features and compute distances) similar models from the PSB set [1]. As displayed in figure 1, the best match is the original model of car (before adding the antenna), while the second and third matches are absolutely irrelevant. Hence, the LFD is not sufficiently robust with respect to outliers. The shown example is an extreme case. Generally, the LFD is one of the most effective 3D-shape descriptors [1].

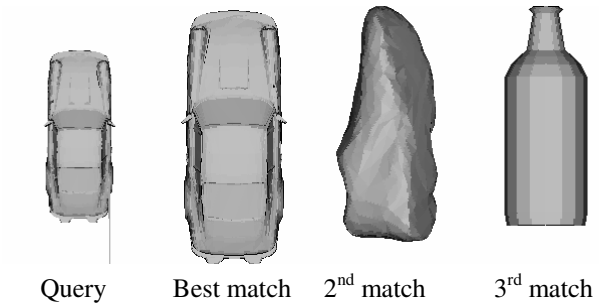

Figure 1. Retrieval using the LightField descriptor.

0-7803-9332-5/05/$20.00 ©2005 IEEE

3. "DESIRE" DESCRIPTOR

As stated in [5], more powerful 3D-shape descriptors can be obtained by combining fundamentally different features aimed at characterizing 3D-shape. The combined features should be effective and "orthogonal" (complementary) to each other. In this section, we present a composite shape descriptor, obtained from the **de**pth buffer-based (DE) feature vector presented in [4,5], the **si**lhouette-based (SI) descriptor [4,5], and the **r**ay-**e**xtent (RE) feature vector [3,5]. We call the composite descriptor "DESIRE".

The DE feature vector describes how distant is the object from a face of a canonical cube by measuring the distance along directions that are *perpendicular* to the face of the cube. The SI descriptor characterizes *contour points* of orthogonal projections of the model on a bounding cube. The RE feature vector gives information about the extent of an object from the center of gravity along *radial* directions. As demonstrated in [5], the retrieval effectiveness of an appropriate composition of descriptors significantly outperforms the original approaches. The feature extraction of the DESIRE descriptor proceeds in five steps: model normalization, extraction of the DE, SI, and RE descriptors, and composition.

In order to secure translation, rotation, scale, and flipping invariance of descriptors, 3D-mesh models are normalized [3,5]. Each triangle mesh model is transformed into a canonical coordinate frame by translating (the center of gravity becomes the origin), rotating (using the Continuous Principal Analysis - CPCA), scaling (the average distance of a point on the surface of the model to the origin becomes 1), and flipping (using a test based on moments) the set of vertices. Complete analytical expressions for transforming a mesh model into canonical coordinates are given in [5]. The CPCA is rather efficient and effective for most categories of 3D-objects. However, certain categories of 3D-models (e.g., cups) are sub-classified by the normalization step. Nevertheless, in spite of known drawbacks, the most effective shape descriptors extracted in the canonical frame outperform competing descriptors obtained by avoiding the PCA [5].

The idea to use depth-buffers for characterizing 3D-shape is introduced in [4], while detailed analysis and exploration of variances of the method are presented in [5]. Briefly, six depth-buffer images of the underlying 3D-object are formed using the canonical cube (CC), i.e., a cube in the canonical frame whose faces are parallel to the coordinate planes, with the center at the origin, and the length of the edge w. Each depth-buffer images serves as the input for the 2D-FFT. Appropriate magnitudes of the obtained Fourier coefficients are used for forming the feature vector $\mathbf{d}=(d_1,\ldots,d_D)$, where $d_1+\ldots+d_D=D$. In [5], the value of w is set to 4. Our additional statistical analysis showed that slightly better results could be obtained by setting w to 3.6. Note that certain parts of 3D-models lay outside the CC and are ignored as outliers.

The SI descriptor is extracted using the canonical bounding cube (CBC) of a 3D-object. The contours of three silhouette images obtained by projecting a 3D-model on the CBC are processed further. Each contour is sampled so that furthest points of intersection between the contour and rays emanated from the origin of the silhouette image and traveling in equiangular radial directions. The distances of the sample points to the origin of the silhouette image serve as the input for FFT, and magnitudes of the obtained coefficients generate the feature vector $\mathbf{s}=(s_1,\ldots,s_S)$, where $s_1+\ldots+s_S=S$.

The RE descriptor is extracted by forming a function on a sphere, by applying the FFT on the sphere, and by using the magnitudes of obtained coefficients as components of the feature vector $\mathbf{r}=(r_1,\ldots,r_R)$, where $r_1+\ldots+r_R=R$. The value of the function on the sphere at the point **u** is equal to the extent of the model in the direction **u**, i.e., the distance between the origin and the furthest point of intersections of the ray traveling in the direction **u** and the polygonal mesh.

The composite feature vector $\boldsymbol{c}$ is formed by concatenating the basic feature vectors, $\mathbf{c}=(\mathbf{d} \mid \mathbf{s} \mid \mathbf{r})$, whence the dimension of the DESIRE descriptor is $C=D+S+R$. Note that the l_1 norm of **c** is equal C. Based on analysis presented in [5], the DE is more effective than the SI descriptor, while the RE is less effective than the SI descriptor. Therefore, the dimensions are chosen so that $D>S>R$, i.e., the "importance" of combined descriptors is fixed according to their performance. We use D=186, S=150, and R=136, whence C=472. For detailed explanation how the DE, SI, and RE descriptors are extracted, discussions about existing constraints, desirable properties of feature vectors that are secured, and tested variances (alternatives) of the methods, we refer to [5].

To demonstrate robustness with respect to outliers, we use the same query as in figure 1. All top three matches (figure 2) are reasonably relevant to the query, when the DESIRE descriptor and the l_1 norm as dissimilarity measure are used. We stress that the depicted models are visualized in scaled canonical coordinates so that all parts of models are captured.

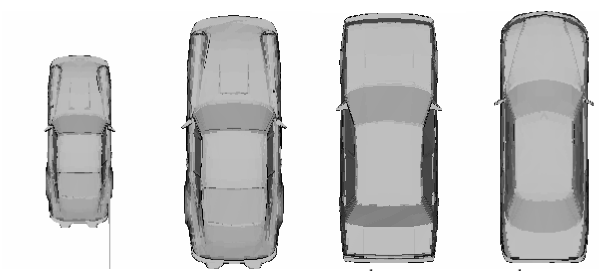

Figure 2. Retrieval using the DESIRE descriptor.

4. EXPERIMENTS

We tested the LFD and DESIRE descriptor using the PSB [1] models (1814 meshes), which are classified into 161 categories. We use standard tools for measuring retrieval effectiveness, precision-recall (PR) diagrams, relevant nearest neighbors (NN), R-precision (RP) (first tier), and Bull-Eye performance (BEP) (second tier) (see [1,5] for details). Firstly, we present PR curves averaged over all models. In figures 3-5, dimensions of the feature vectors are given in the square brackets, while the average precision for recall great or equal 50%, the average precision, BEP, RP, and NN are given in parentheses.

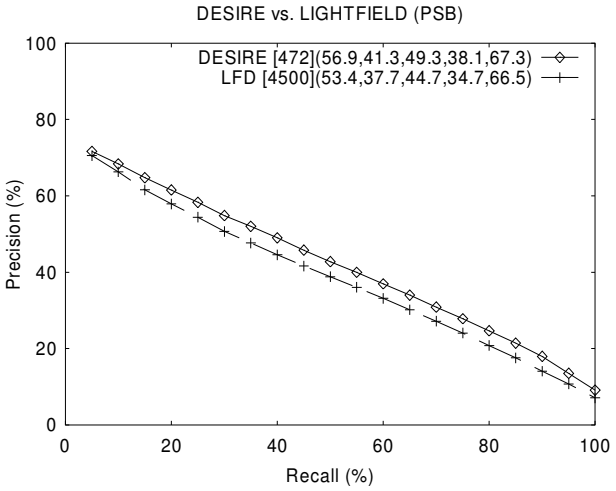

Figure 3. Average precision vs. recall.

Since categories consist of different number of models, we averaged PR diagrams for each category and found the average over categories (figure 4). We also considered a coarser categorization of the PSB set obtained by merging categories with more than two class-keywords and averaged PR curves for all models (figure 5).

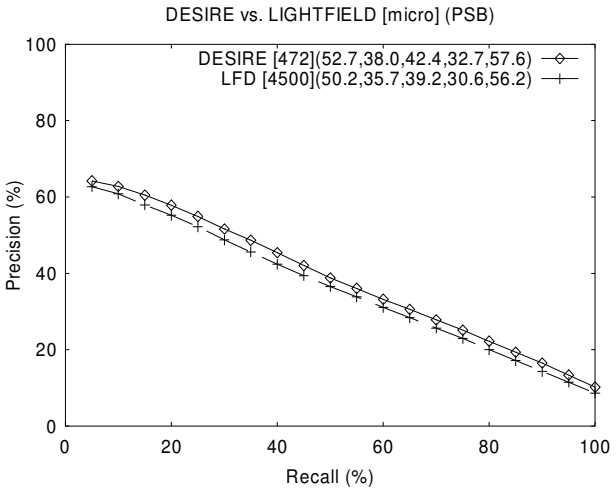

Figure 4. Precision vs. recall averaged over categories.

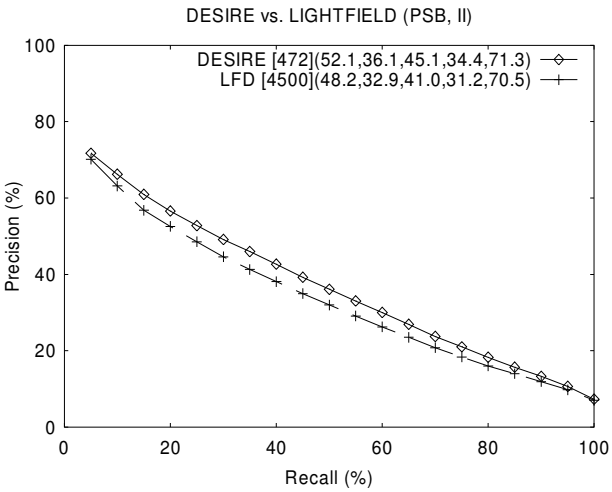

Figure 5. Avg. prec. vs. recall for coarser categorization.

Results shown in figures 3-5 demonstrate that the DESIRE descriptor generally outperforms the LFD. This does not mean that the DESIRE is more suitable descriptor for each category of models. In order to summarize category-wise comparison of the competing descriptors, we found differences between NN, RP, and BEP for DESIRE and LFD (for each of 161 categories), sorted all three sequences in the non-increasing order and displayed them in figure 6. We read that the BEP score is for 90 categories better when DESIRE is engaged, for 57 categories LFD leads to better BEP values, while for 14 (161-90-57) categories both descriptors have the same BEP, and BEP of DESIRE is 4.65% higher on average. As expected from the global results, the DESIRE descriptor is more suitable for most categories. All LFD descriptors are extracted and distances between two LFD descriptors are computed using the original executables provided by authors. The l_1 norm is used as metric for the DESIRE.

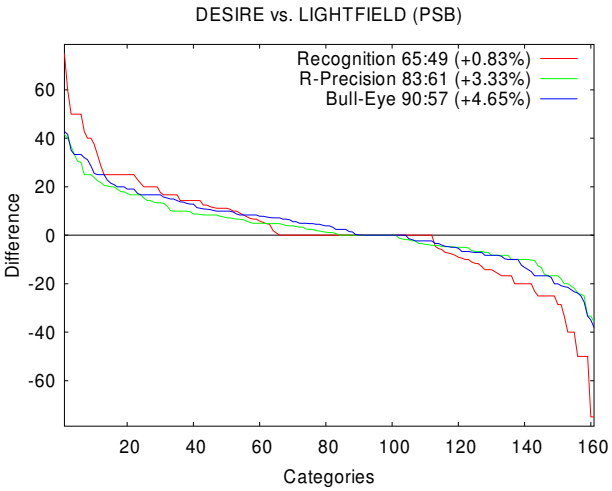

Figure 6. Category-wise comparison for nearest neighbors (recognition), R-precision, and Bull-Eye performance.

We also confronted the two descriptors on our own collection of 3D-models, and the obtained results [8] are even more in favor of the DESIRE descriptor.

Note that the precision–recall curves for the LFD descriptor from [1] and figure 3 are significantly different although the same test set (PSB) is used. The extreme high precision for low recall values in [1] is caused by treating the query model as the best match. Besides, since the original LFD extractor uses OpenGL, the obtained descriptors differ slightly on different PC systems causing additional minor differences in precision-recall curves.

Average feature extraction of the DESIRE descriptor from a normalized model is 135ms. If model loading, checking, and normalization are included, the extraction takes 211ms on a PC with 1 GB RAM and a 3 GHz Pentium 4 processor running Windows XP SP1. The average extraction time for the LFD is 2.3 seconds.

5. DISCUSSION

Results from section 4 suggest that the DESIRE is more effective than the LFD. However, the effectiveness is one of numerous advantages of the DESIRE. If we consider the vector dimensions (DESIRE 472, LFD 4500), the number of necessary rotations of models (DESIRE 1, LFD 5460), the average extraction time (DESIRE 0.2s, LFD 2.3s), and the complexity of computing the dissimilarity measure (matching procedure), then the DESIRE descriptor is a significantly better technique.

The only common aspect of the DESIRE and LFD is the usage of silhouettes. The LFD relies upon results presented in [6], where the centroid distance is compared to several inferior features of contour points. However, the sample contour points, whose centroid distances serve as the input for the FFT, are selected using equal arc length distance. The analysis in [5] states that selecting sample contour points so that their vectors lay on equiangular radial directions (see SI descriptor in section 3) is a more robust approach. Thus, in the trade-off between lost details introduced by using equal angle (only the furthest point of intersection is considered) and the lost of correspondence between similar contours introduced by using equal arc length (an outlier causes a shifting), the preservation of correspondence is more important. Another advantage of using the approach presented in [5] is its applicability in (rare) cases when a 3D-object consists of disjoint parts, whence the silhouette image consists of more than one contour. The method that uses arc length parameterization [6] is restricted to single contour.

Potentially, the performance of the LFD might be improved by changing the method for securing the scale invariance, by characterizing silhouette images in a more suitable manner, by using sets of 10 images for aligning a pair of 3D-models but using fewer images for computing distances (we expect that some of 10 views may represent a "noise" to the overall dissimilarity), or by considering depth buffer images instead of silhouettes (for alignment or for distance computation).

We consider that the most significant weakness of our approach lies in the non-optimal canonical positioning (normalization) of a model. In a forthcoming paper, we will compare the effect of the CPCA [3,5] and the orientation normalization using spherical harmonics [9].

6. CONCLUSION

In summary we have compared two 3D-shape descriptors that are declared the best in recent studies. The comparison is based on the PSB set of 3D-models. We have found that the retrieval effectiveness and the complexity in time and space suggest that the composite descriptor DESIRE outperforms the competing LightField descriptor. In order to enable verification of our results, we provide executables for extracting feature vectors and source code used for presented analysis. A Web-based retrieval system for testing both methods is also available.

6. REFERENCES

[1] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton Shape Benchmark," *SMI 2004,* Genova, Italy, 2004.

[2] D. Y. Chen, X. P. Tian, Y. T. Shen and M. Ouhyoung, "On Visual Similarity Based 3D Model Retrieval", in *Proc. of Eurographics Workshop*, Vol. 22, pp. 223-232, 2003.

[3] D.V. Vranić, D. Saupe, and J. Richter, "Tools for 3D-object retrieval: Karhunen-Loeve Transform and spherical harmonics," *Proc. of IEEE MMSP 2001*, Cannes, France, pp. 293-298, 2001.

[4] M. Heczko, D. Keim, D. Saupe, and D.V. Vranić, A method for similarity search of 3D objects (in German), *Proc. of BTW 2001*, Oldenburg, Germany, pp. 384-401, 2001.

[5] D.V. Vranić, *3D Model Retrieval*, Ph. D. Thesis, University of Leipzig, Germany, 2004.

[6] D. S. Zhang and G. Lu, "A Comparative Study of Fourier Descriptors for Shape Representation and Retrieva,l" *Proc. of Shape Modeling International*, Genova, Italy, 2004.

[7] J. Tangelder and R. Veltkamp, "A Survey of Content Based 3D Shape Retrieval Methods", *SMI 2004*, Genova, pp. 145-156.

[8] DESIRE: Web-Based Demonstration, Results, and Tools, http://merkur01.inf.uni-konstanz.de/ICME2005.

[9] G. Burel and H. Henocq, "Determination of the Orientation of 3D Objects Using Spherical Harmonics", *Graphical Models and Image Processing*, Vol. 57(5), pp. 400-408, 1995.