Content-based 3D model retrieval considering the user's relevance feedback

Panagiotis Papadakis[1],[2],[*]

[1] Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos", Greece
[2] Computer Graphics Laboratory, Department of Informatics & Telecommunications, University of Athens, Greece
ppapadak@iit.demokritos.gr

Abstract. This dissertation deals with the problem of enabling precise and real-time retrieval of generic 3D models. We focus on three sub-problems, namely, rotation normalization, discriminative shape representations and relevance feedback-based retrieval. Considering the first problem, we develop a rotation normalization method that computes the object's principal axes using the orientation of the object's surface. We also develop a hybrid rotation normalization scheme than combines both the surface distribution and the surface orientation distribution. Considering the second problem, we develop a set of 3D shape descriptors, namely, the Concrete Radialized Spherical Projection (CRSP), Hybrid and Panoramic Object Representation for Accurate Model Attributing (PANORAMA) shape descriptor. The CRSP descriptor is formed by computing a spherical function-based representation and using the spherical harmonic transform. The Hybrid descriptor is formed by combining the CRSP descriptor with a depth buffer-based representation. The PANORAMA descriptor is formed by projecting the surface of a 3D object and its orientation to the lateral surfaces of a set of cylinders that are centered at the object's centroid and parallel to its principal axes. For each projection we compute the 2D Discrete Fourier Transform and Wavelet Transform. Considering the third problem, we develop a local relevance feedback technique which is based on shifting the feature vectors of 3D objects closer to their cluster centroid in 3D space.

1 Introduction

Information is an interdisciplinary concept used to refer to textual, image, audio and video data and, more recently, 3D data. Common to all these types of information is the need for methods that enable effective search within large collections of information; this holds true for 3D geometrical data, since reusing such data is more efficient than creating it from scratch.

3D geometrical objects constitute an integral part of various applications such as computer-aided design, architecture, computer games development, virtual

*Dissertation Advisor: Theoharis Theoharis, Associate Professor reality, archaeology, computer vision, biometrics and medicine. Regardless of the context of a particular application, as 3D object repositories grow in size and complexity, the need for effective and efficient content-based 3D object retrieval is becomes imperative.

1.1 Problem statement and challenges

3D object retrieval is the process where a query is initiated by a user within a collection of 3D objects and a retrieval system responds to that query by presenting to the user a list of retrieved objects that are sorted in decreasing similarity to the query, according to a predefined measure of object similarity.

This thesis addresses the latter problem of inter-class retrieval, i.e. retrieval within a broad domain of 3D objects where there are no limitations regarding the shape characteristics of the objects and there is no available knowledge regarding their geometrical shape.

In contrast to conventional 3D object retrieval methods, such as keyword-based or sketch-based methods, a more effective approach is content-based retrieval where queries are posed as 3D objects. Using 3D object queries, the retrieval system obtains all necessary information about the object's shape. In detail, a 3D object is transformed into a representation that is used to efficiently and accurately compare it against other 3D objects. This representation is called *shape descriptor* and is commonly used to refer to a feature vector that captures the shape characteristics of a 3D object. An important requirement is that the descriptor of a 3D object should be invariant to the translation, scale and rotation of an object in 3D space, as these characteristics are not considered to affect the measure of similarity between 3D objects. Although translation and scale invariance can be readily attained, rotation invariance is the most complex requirement to be addressed. A shape descriptor should also be efficient in terms of computation and comparison times as well as size requirements; otherwise it is not appropriate for real-time retrieval, neither suitable for large databases.

The above approach to 3D object retrieval is fixed to a predetermined way of measuring similarity between objects without any consideration about the particularities of the query and the collection of 3D objects that are compared against the query. A way to overcome this limitation is to enable the retrieval system to use information about the similarity of the query to a subset of objects, which can then be used to tune the way of measuring similarity and increase performance. The process of providing the relevance information to the system is called *relevance feedback* and it can either be obtained as a result of the user's interaction with the system or it can be inferred from the retrieval system itself. These two types of relevance feedback correspond to supervised and unsupervised learning respectively.

The unsupervised relevance feedback approach is known as local relevance feedback and, since it does not require any user interaction, it facilitates the retrieval task. A local relevance feedback method should be based on non-strict assertions about the relevance of objects in order to minimize its sensitivity to incorrect judgements while while the additional computational overhead should not alter the real-time nature of retrieval.

1.2 Contributions of this dissertation

This dissertation proposes a new framework for 3D object retrieval with contributions in the fields of rotation normalization of 3D objects, 3D shape description and LRF. The proposed framework enables very effective and efficient real-time retrieval within large collections of 3D objects in an unsupervised framework. In detail, the contributions of this dissertation are the following:

- A rotation normalization method that uses the distribution of the orientation of an object's surface in order to determine and align its principal axes with the coordinate axes (Section 2.1). This method enables improved rotation normalization for a variety of 3D object classes where the distribution of the object's surface itself is inappropriate to determine the principal axes of the object.
- A hybrid rotation normalization scheme that exploits both the distribution of the surface itself and the orientation of the object's surface in order to find two alternative but highly complementary ways to determine and align the principal axes of the object with the coordinate axes (Section 2.2). The hybrid scheme provides consistent rotation normalization for a wider range of 3D object classes compared to using either the distribution of the surface itself or the distribution of the surface orientation.
- A spherical function-based shape descriptor that considers not only the surface of a 3D object but also parts that are occluded under particular viewpoints (Section 3.1). The proposed descriptor attains better discrimination ability than the conventional approach where only the surface of a 3D object is considered.
- A hybrid 2D/3D shape descriptor that combines two highly complementary descriptors, namely, the spherical function-based descriptor with a depth buffer-based descriptor and enables improved discrimination ability by integrating the 2D-based with the 3D-based domain of 3D object retrieval methodologies (Section 3.2).
- A panoramic 3D shape descriptor that is based on the usage of cylindrical projections which describe the shape of a 3D object using both the depth and orientation of the object's surface (Section 3.3). The proposed descriptor can sufficiently describe the shape of a 3D object by using a set of three perpendicular cylinders that are uniformly sampled along their lateral surface and enables superior retrieval performance compared to state of the art shape descriptors.
- A LRF scheme within the context of 3D object retrieval using a method that shifts the feature vector of a 3D object closer to its cluster centroid in feature space (Section 4). The proposed scheme improves the retrieval performance without requiring the user's feedback and enables conventional RF methods that can be subsequently employed to benefit since more relevant 3D objects

are presented to the user, thus, enabling him to provide additional training data.

2 Rotation normalization

Ensuring rotation invariance for a wide range of object classes has proven to be a very complicated problem. This is because it is not obvious which shape characteristic should be chosen (e.g. surface distribution, orientation, symmetry, etc) and in which way it should be used when normalizing for rotation, so that it is applicable across object classes.

2.1 Normal PCA

The method that we propose for rotation normalization was introduced in [1]. It uses PCA to determine the principal components of the surface orientation of an object and is insensitive to inconsistent directions of the normal vectors of neighboring polygons.

In general, the purpose of PCA is to uncorrelate a feature space. This is done by computing an orthonormal basis for the feature space where a set of input samples are represented and transforming the initial input samples to the uncorrelated feature space. The basis vectors of the computed orthonormal basis are called principal components and are computed as the eigenvectors of the covariance matrix of the input samples.

We compute the covariance matrix $\mathbf{C}$ of an object's normal vectors as:

$$\mathbf{C} = \frac{1}{A}\sum_{j=1}^{N_p} A_j \mathbf{n}_j \mathbf{n}_j^T \tag{1}$$

where $\mathbf{n_j}$, denotes the normal vector of the j^{th} triangle of the object's surface, A_j the corresponding surface area, A the total surface area of the object, N_p the total number of triangles and T the transpose operation. After the computation of $\mathbf{C}$, we compute its orthogonal eigenvectors and normalize them to the unit L_2 norm which correspond to the principal axes of the object. Next, the eigenvectors are sorted in decreasing (or increasing) order of the respective eigenvalue, i.e. in decreasing (or increasing) order of the variance along each eigenvector. Setting Φ as the matrix having as rows the sorted and normalized eigenvectors, the final step is to apply Φ to the normal vectors of the object's triangles. To achieve this, the transformation is applied to the vertices of the object instead, i.e. to the actual 3D object. In Fig. 1, we give a set of example 3D objects that are consistently aligned using NPCA.

2.2 Hybrid rotation normalization scheme

The hybrid rotation normalization scheme that we propose is based on the interchangeable usage of the distribution of surface points and the distribution of

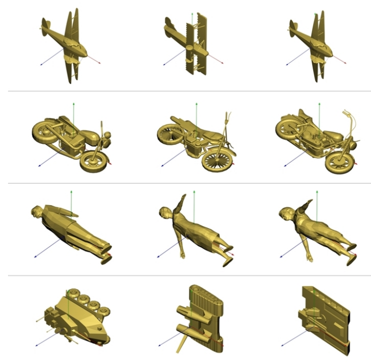

Fig. 1. Examples of consistently aligned 3D objects using the NPCA method.

the surface orientation. Toward this goal, we employ PCA to determine two alternative orientations of the object's principal axes, using the two former shape characteristics. In particular, the two variations of the PCA algorithm that are employed are the Continuous PCA (CPCA) [2] that uses the points of the object's surface and the previously described Normal PCA that uses the orientation of the object's surface. The difference between the two alignment methods is in the input samples that are used for the computation of the covariance matrix. According to the formulation of CPCA the covariance matrix $\mathbf{C}$ of an object's surface is computed as:

$$\mathbf{C} = \frac{1}{12A}\sum_{j=1}^{N_p} A_j \cdot \left[\mathbf{f}(\mathbf{v}_{j1}) + \mathbf{f}(\mathbf{v}_{j2}) + \mathbf{f}(\mathbf{v}_{j3}) + 9 \cdot \mathbf{f}\left((\mathbf{v}_{j1} + \mathbf{v}_{j2} + \mathbf{v}_{j3})/3\right)\right] \quad (2)$$

where $\mathbf{v}_{j1}, \mathbf{v}_{j2}, \mathbf{v}_{j3}$ are the three vertices of the j^{th} triangle, $f(\mathbf{v}) = (\mathbf{v} - \mathbf{m}_s)(\mathbf{v} - \mathbf{m}_s)^T$ and $\mathbf{m_s}$ is the centroid of the object's surface.

After the computation of $\mathbf{C}$, we compute the corresponding orthonormal eigenvectors, sort them in decreasing (or increasing) order of the corresponding eigenvalues, form a transformation matrix Φ having as rows the sorted eigenvectors and apply Φ to the vertices of the 3D object. Thus, we first employ CPCA and then NPCA to obtain two alternative aligned versions of a 3D object that are subsequently used to acquire the corresponding pair of shape descriptors that are used for matching the 3D objects. In order to compare two objects, both aligned versions are used according to the following schema. We compare the CPCA aligned version of the first object to the CPCA aligned version of the second object and similarly for NPCA. This gives two comparison scores for a pair of 3D objects. If a distance measure is employed to compare 3D objects, then the smaller the distance, the more similar the 3D objects are considered to be. Therefore, we set the final score as the minimum of the two distance scores. In figure 2, we give typical examples of cases where 3D objects are better aligned with NPCA than CPCA (Fig. 2 (a)) and vice versa (Fig. 2 (b)).

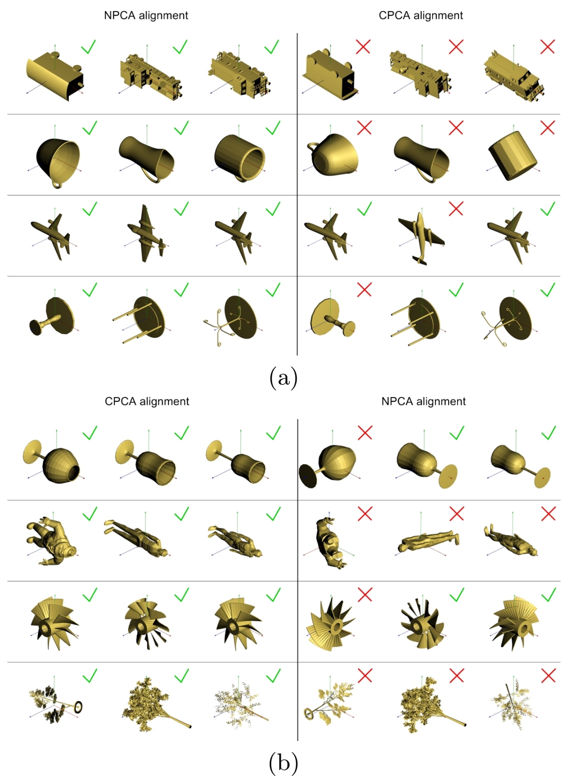

Fig. 2. Comparison of NPCA and CPCA alignments of various 3D objects where: (a) NPCA is more effective than CPCA and (b) CPCA is more effective than NPCA.

3 Content-based 3D Object Retrieval

3D object retrieval based on shape similarity requires an appropriate representation that is used to index and measure the similarity of 3D objects. This representation is called shape descriptor and is commonly used to refer to a feature vector. Ideally, a 3D shape descriptor should be invariant to the translation, rotation, scaling and reflection of 3D objects, compact and at the same time descriptive of the object's shape features and efficient in terms of computation and comparison time.

Extended state of the art surveys for 3D object retrieval methodologies can be found in [3], [4], [5], [6] and [7]. Most approaches can be classified into two main categories according to the spatial dimensionality of the information used, namely 2D, 3D and their combination. 2D-based methods transform the problem of measuring 3D object similarity into the problem of matching a collection of 2D images that characterize the shape of the objects. 3D-based methods process a 3D object as a single geometric entity in 3D space and capture a set of features that are computed over the 3D parameterization of the object.

3.1 Concrete Radialized Spherical Projection descriptor

The Concrete Radialized Spherical Projection (CRSP) descriptor [1] is a spherical function-based descriptor that captures the radial depth of an object's surface and considers the object's interior as well as parts that are occluded from particular viewpoints. It employs the hybrid rotation normalization scheme and uses the spherical harmonic representation of the spherical functions as shape features.

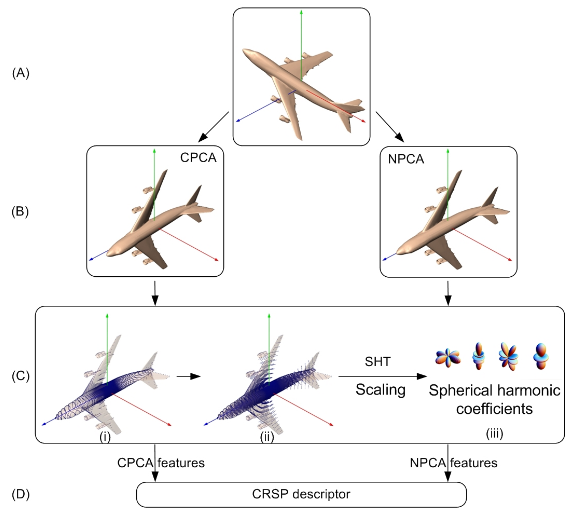

Fig. 3. The overall scheme for the extraction of the CRSP descriptor.

Initially, the model is translated so that its center of mass coincides with the origin (Fig. 3 (A)). This is achieved by computing the object's centroid and translating the object to its centroid. Then we employ the proposed hybrid rotation normalization scheme which gives two aligned versions of the translated model (Fig. 3 (B)) that are hereafter processed in the same way and will finally give two sets of features. In the next stage, each alignment variant is expressed by a set of spherical functions with increasing radius giving the intersections of the model's surface with rays emanating from the origin (Fig. 3 (C)-(i)). In the sequel, the set of spherical functions is processed to find the furthest intersection points from the origin on each ray. If the model is viewed in the direction of a specific ray, then the furthest intersection point leaves invisible the part of the model along the ray which is closer to the origin. We may assume that the invisible part belongs to the 3D model because we perceive a 3D model as a concrete-solid entity. Thereafter, these parts are considered to belong to the model giving a new 3D model representation (Fig. 3 (C)-(ii)). For each spherical function, the spherical harmonics transform is applied producing a set of complex coefficients that are scaled to the unit L_1 norm (Fig. 3 (C)-(iii)).

3.2 Hybrid descriptor

The Hybrid 3D shape descriptor [8] descriptor is formed by linearly combining CRSP with a depth-buffer based shape descriptor that was previously proposed in [9]. By combining a 3D-based descriptor (CRSP) with a 2D-based descriptor [9] and adopting the hybrid rotation normalization scheme, the Hybrid descriptor attains significantly improved performance compared to each individual descriptor component while enabling real-time retrieval at comparatively small size requirements.

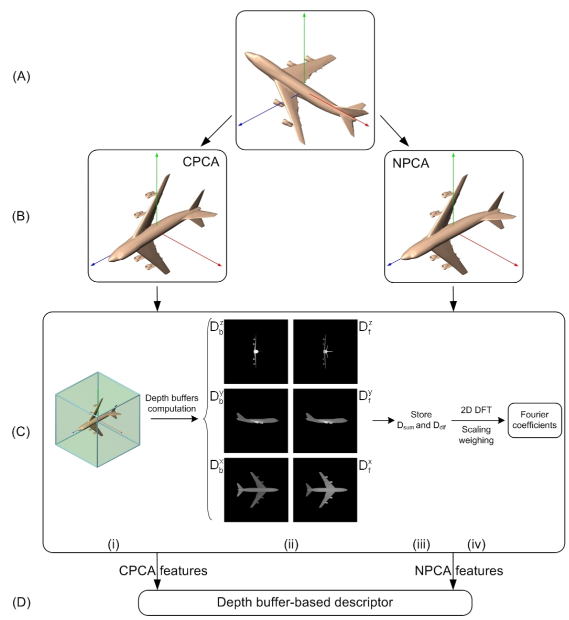

Fig. 4. The overall scheme for the extraction of the depth-buffer based descriptor.

Initially, the model is translated so that its center of mass coincides with the origin (Fig. 4 (A)). This is achieved by computing the object's centroid and translating the object to its centroid. Then we employ the proposed hybrid rotation normalization scheme which gives two aligned versions of the translated model (Fig. 4 (B)) that are hereafter processed in the same way and will finally give two sets of features. In the following, we acquire a set of depth buffers of the 3D object by projecting it to the six faces of a cube and centered at the object's centroid (Fig. 4 (C)-(ii)). Next, we store the difference and the sum of each pair of depth buffers (Fig. 4 (C)-(iii)) and compute the respective 2D Discrete Fourier Transform. Finally, the Fourier coefficients are scaled to their unit L_1 norm and weighed according to the principal directions that they encode (Fig. 4 (C)-(iv)).

3.3 Panoramic Object Representation for Accurate Model Attributing descriptor

The PANoramic Object Representation for Accurate Model Attributing 3D shape descriptor (PANORAMA) [10] employs the hybrid rotation normalization scheme and uses a cylinder-based representation for the description of a 3D object. The cylinders are used to project the surface depth and orientation in a uniform fashion, giving a set of 2D projections that are particularly descriptive and can be readily analyzed using standard 2D features.

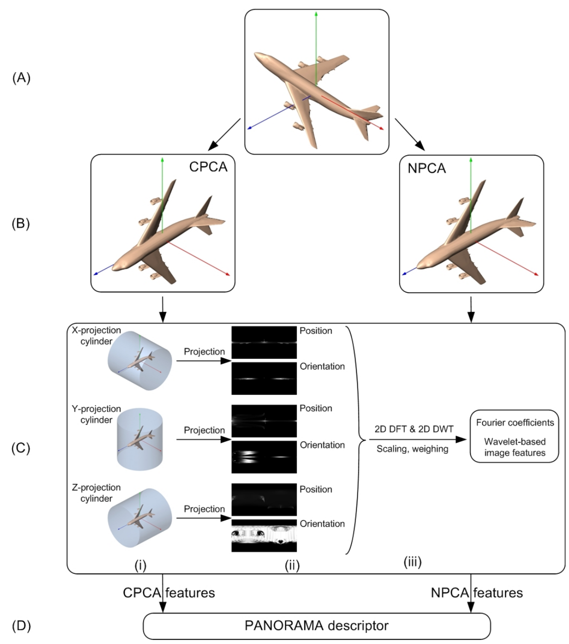

Fig. 5. The overall scheme for the extraction of the panorama descriptor.

Initially, the model is translated so that its center of mass coincides with the origin (Fig. 5 (A)). This is achieved by computing the object's centroid using and translating the object to its centroid. Then we employ the proposed hybrid rotation normalization scheme which gives two aligned versions of the translated model (Fig. 5 (B)) that are hereafter processed in the same way and will finally give two sets of features. In the following, we acquire a set of panoramic views of the 3D object by projecting it to the lateral surface of a cylinder aligned with one of its three principal axes and centered at the object's centroid. The object is projected onto three perpendicular cylinders, each one aligned with one of its principal axes (Fig. 5 (C)-(i)). The cylindrical projections capture the position (depth) of the model's surface as well as its orientation (Fig. 5 (C)-(ii)). For each projection, we compute the 2D Discrete Fourier Transform as well as

2D Discrete Wavelet Transform which are normalized to their unit norm and weighed according to the cylindrical projection that they describe (Fig. 5 (C)-(iii)). The details of the overall procedure are given in the following subsections.

4 Relevance Feedback

Relevance feedback is used as a mean to involve the user in the retrieval process and guide the retrieval system towards the target, in order to enable the machine to retrieve information through adapting to individual categorization criteria and hence bridge the semantic gap. RF was first used to improve text retrieval [11], later on successfully employed in image retrieval systems [12],[**?**],[13],[14] and lately in 3D object retrieval systems [15],[16],[17],[18]. It is the information of relevance with respect to a subset of the retrieved results, acquired from the user's interaction with the retrieval system. Using RF, the retrieval system adjusts its parameters in order to optimally match the user's classification criteria. Then, based on the adjusted parameter setting, a new retrieval session is initiated, new results are presented to the user and the above procedure is repeated until the user is satisfied.

In local relevance feedback the user does not provide any RF at all. Instead, the required information is obtained using only the unsupervised retrieval result. The procedure comprises two steps. First, the user submits a query to the system which uses the low-level features to produce a ranked list of results which is not displayed to the user. Second, the system reconfigures itself by only using the top m closest matches of the list, based on the assumption that they are relevant to the user's query. The relevance assumption for the top m closest matches is not always valid thus irrelevant objects that may lie close to the neighborhood of the query could be erroneously considered as relevant. This case is known as "query drift" phenomenon and implies the scenario where the retrieval system is misled by the irrelevant data and drawn away from the user's target.

4.1 Proposed local relevance feedback technique

The LRF technique that we propose comprises an off-line and an on-line stage. Both stages are based on the idea of moving a feature vector closer to its cluster centroid in feature space, thus improving the performance of a nearest neighbor search, which is similar in spirit to the query modification technique.

During the off-line stage each object of the dataset is used as a query in the remaining set of objects which. For each object, we take the descriptors of the top closest matches which are assumed to be relevant and shift the descriptor of the query object towards a weighted centroid of the top closest matches.

During the on-line stage the system computes the descriptor of the query and compares it to the updated descriptors of the objects of the database. The top m closest matches are used to update the descriptor of the query as in the off-line stage. In the sequel, a new retrieval session is initiated using the updated descriptor of the query, which is closer in feature space to its cluster-class centroid and the results are shown to the user in decreasing order of similarity.

5 Results

We have performed extensive experiments that demonstrate the effectiveness of the proposed rotation normalization schemes, shape descriptors and local relevance feedback method. Our methods clearly advance the state of the art.

In Fig. 6 (a) we compare the performance of NPCA, CPCA and the hybrid rotation normalization scheme using the PANORAMA descriptor in terms of the Discounted Cumulative Gain (DCG [?]) score in the NIST [?] dataset. It is readily observed that the NPCA, CPCA methods exhibit high complementarity which is the reason for the improved performance of the hybrid rotation normalization scheme.

In Fig. 6 (b) we compare the CRSP, Hybrid and PANORAMA descriptors against the best performing state of the art descriptors (DESIRE, LF and SH-GEDT), in terms of precision and recall. Apparently, the proposed 3D shape descriptors are ranked above the compared methods with PANORAMA being the clear winner.

In Fig. 6 (c) we demonstrate the increase in the performance of the CRSP and PANORAMA descriptors that is due to the application of LRF. It is clear that LRF adds a major gain in the retrieval performance, particularly in the case of the PANORAMA descriptor.

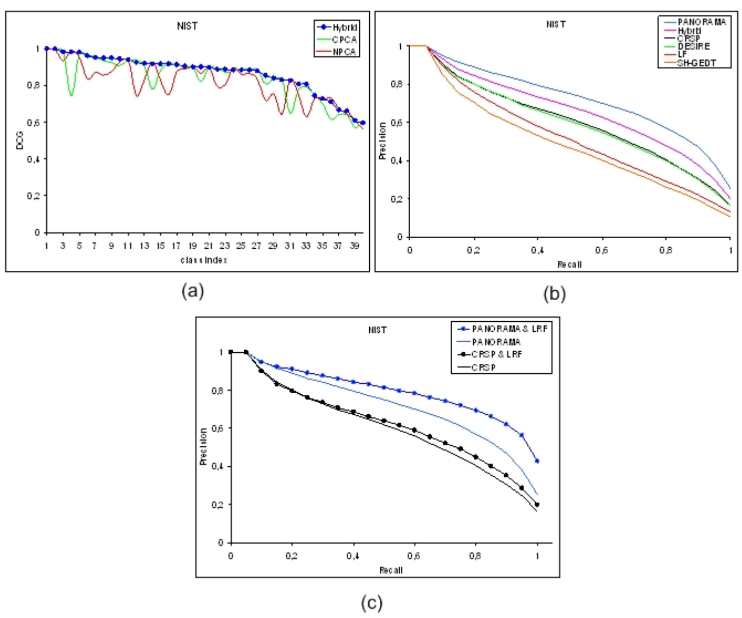

Fig. 6. (a) Performance comparison of the NPCA, CPCA and hybrid rotation normalization scheme; (b) Performance comparison of the PANORAMA, Hybrid and CRSP shape descriptors against the DESIRE, LF and SH-GEDT state of the art descriptors; (c) Performance evaluation of the proposed LRF method.

References

1. Papadakis, P., Pratikakis, I., Perantonis, S., Theoharis, T.: Efficient 3D shape matching and retrieval using a concrete radialized spherical projection representation. Pattern Recognition **40**(9) (2007) 2437–2452
2. Vranic, D.V.: 3D Model Retrieval. PhD thesis, University of Leipzig (2004)
3. Bustos, B., Keim, D., Saupe, D., Schreck, T., Vranić, D.V.: Feature-based similarity search in 3d object databases. ACM Comput. Surv. **37**(4) (2005) 345–387
4. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The princeton shape benchmark. In: Shape Modeling International. (2004) 167–178
5. ans S. Jayanti, N.I., Lou, K., Kalyanaraman, Y., Ramani, K.: Three-dimensional shape searching: state-of-the-art review and future trends. Computer-Aided Design **37** (2005) 509–530
6. Tangelder, J., Veltkamp, R.: A survey of content based 3d shape retrieval methods. In: Proceedings of the Shape Modeling International. (2004)
7. Yang, Y., Lin, H., Zhang, Y.: Content-based 3-d model retrieval: A survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **37** (2007) 1081 – 1098
8. Papadakis, P., Pratikakis, I., Theoharis, T., Passalis, G., Perantonis, S.: 3D object retrieval using an efficient and compact hybrid shape descriptor. In: Eurographics Workshop on 3D Object Retrieval. (2008)
9. Passalis, G., Theoharis, T., Kakadiaris, I.A.: Ptk: A novel depth buffer-based shape descriptor for three-dimensional object retrieval. Visual Computer **23**(1) (2006) 5–14
10. Papadakis, P., Pratikakis, I., Theoharis, T., Perantonis, S.: Panorama: A 3d shape descriptor based on panoramic views for unsupervised 3d object retrieval. International Journal of Computer Vision accepted for publication.
11. Rocchio, J.: 14. In: Relevance Feedback in Information Retrieval The SMART Retrieval System. . (1971) 313–323
12. Cox, I., M.Miller, Minka, T., P.Yianilos: An optimized interaction strategy for bayesian relevance feedback. In: IEEE Conference on Computer Vision and Pattern Recognition. (1998) 553–558
13. C.Nastar, Mitschke, M., Meilhac, C.: Efficient query refinement for image retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition. (1998) 547–552
14. Rui, Y., Huang, T., Mehrotra, S.: Content-based image retrieval with relevance feedback in mars. In: IEEE International Conference on Image Processing. (1997) 815–818
15. Akbar, S., Kung, J., Wagner, R., Prihatmanto, A.S.: Multi-feature integration with relevance feedback on 3D model similarity retrieval. In: Proceedings of the International Conference on Information Integration and Web-based Applications Services. (2006) 77–86
16. Bang, H., Chen, T.: Feature space warping: An approach to relevance feedback. In: IEEE International Conference on Image Processing. (2002)
17. Elad, M., Tal, A., Ar, S.: Content based retrieval of vrml objects: an iterative and interactive approach. In: Proceedings of the 6th Eurographics workshop on Multimedia. (2001)
18. Papadakis, P., Pratikakis, I., Trafalis, T., Theoharis, T., Perantonis, S.: Relevance feedback in content-based 3d object retrieval: A comparative study. Computer-Aided Design and Applications **5**(5) (2008) 753–763