Eurographics Workshop on 3D Object Retrieval (2010)
I. Pratikakis, M. Spagnuolo, T. Theoharis, and R. Veltkamp (Editors)

3D shape retrieval based on Multi-scale Integral Orientations

Asim Imdad Wagan[1] , Stephane Bres[1] , Hubert Emptoz[1] and Afzal Godil[2]

[1]INSA de Lyon, France
[2]National Institute of Standards and Technology, USA

Abstract

In this paper we describe a novel 3D shape retrieval method based on Multi-scale Integral Orientations. In our approach, a 3D model after normalization is represented by a set of depth images captured uniformly on a unit circle. Then the shape descriptor based on the multiscale version of the localized gradient histogram is calculated for each depth image. Finally, the comparison is performed using a simple Euclidean distance to prove the effectiveness of the shape descriptor for the 3D shape retrieval. We have then applied our algorithm on the Princeton shape benchmark and got results with performance similar to the Light Field Descriptor. In the future we are planning to use this descriptor with the bag of features and machine learning based approaches to get even better results.

Categories and Subject Descriptors (according to ACM CCS): I.2.10 [Computer Graphics]: Vision and Scene Understanding—Shape

1. Introduction

With advances in 3D modeling and scanning technologies, large number of 3D models are created and widely used in computer graphics and visualization, archeology, bio-imaging and medical imaging applications, defense and security, industrial design and games, entertainment and film industry and so on. This has created an impetus to develop effective 3D shape retrieval algorithms for these domains. This has made the field of 3D shape retrieval become an active area of research in the 3D community. One of the main area of research in 3D Shape retrieval fields is the shape descriptor (feature descriptor). The shape descriptors can be classified into four main categories : view-based, graph-based, statistics-based and transformed-based, or as local and global shape descriptors. Recent work shows that view-based approaches perform better than other approaches for 3D shape retrieval. The view-based approaches can also be used for query searching interface, based on binary images or 2D sketches.

One of the main drawback of view-based methods is that there a infinite number of possible view. One very effective solution is to apply pose normalization to the object via Principal Component Analysis (PCA) than uniformly sample the view direction on a unit sphere.

For our method we fist apply pose normalization to the object via Principal Component Analysis (PCA) than uniformly sample the view direction on a unit sphere to create 20 depth images for each model. we then apply our algorithm based on the Multiscale Integral orientations on the depth images to calculate the shape descriptor. Then comparison is performed using a simple Euclidean distance to proving the effectiveness of the shape descriptor for the 3D shape retrieval.

2. Related Works

In this section we will only describe some of the more relevant methods with respect to our work. For more information about different 3D shape retrieval methods we refer the reader to the survey paper [TV08].

Silhouette have frequently been used for view-based techniques for 3D model retrieval. In Light Field Descriptor by Chen et al. and DESIRE, by Vranic et al. [CTSO03, VS00, Vra04], the authors developed a descriptor based on the silhouettes from multiple directions for visual similarity based comparison. Initial methods proposed for 3D object searching were based on the histograms [AKKS99] and [OFCD02]. Although these methods are very simple and fast to compute, the results were not that promising. Laga

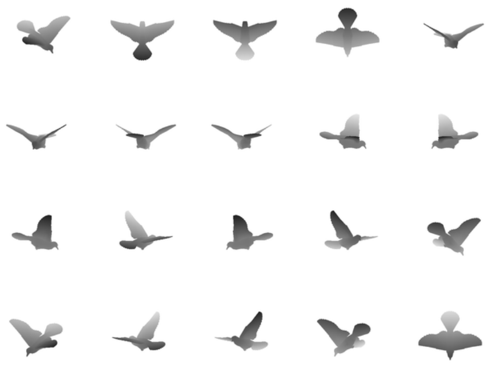

Figure 1: *Model Alignment and projection*

et al. [LTN06] have discussed the generation of a descriptor for 3D shape comparison using spherical wavelet transforms. Recently in [KH90, FHK*04] some work has been done to define shape features using Krawtchouk moments. In [SMKF04] discussed the creation of the Princeton Shape Benchmark, and analyzed some of the contemporary algorithms for 3D shape retrieval. Pu Jiantao has suggested to use the cutting planes in [JJY*04]. Papadakis et al [PTK07] have proposed a method based on the features calculated from the 3D depth buffer. Recently in [PPT*08], the authors have used the hybrid descriptor which is composed of 2D features from depth buffers and 3D features based on spherical harmonics with some interesting results. In [KH90] authors discussed methods to compare images based on the Zernike moments and compared the performance of Zernike moments with other type of image similarity measures. Ohbuchi et al has used the SIFT based invariant features and Bag of words based scheme to generate features for the 3D models for 3D shape retreival [OF, OOFB08]. MSIO descriptor has the same origin in principle to the SIFT descriptor which was proposed in [Low04, Low99].

3. 3D Model Projection

We first apply pose normalization to the 3D object via Principal Component Analysis (PCA) than uniformly sample the view direction on a unit sphere to create 20 depth images. Before calculating the descriptor, we create an image composed of all these different depth images into 4x5 grid, which is shown in the figure 1.

4. Multi-scale Integral Orientations

Multiscale Integral Orientations are the localized orientation gradients which are calculated using the integral images. These orientations are binned inside a set of blocks and layers. The blocks are by definition a rectangular region inside a layer in the Multiscale integral orientations. Multiscale orientations are similar to the feature descriptor used by the Scale Invariant Feature Transforms [Low04] but they also are different in many other ways as the local regions and the calculation is done in a radically different way. Let us consider an image over which, we want to calculate the MSIO features. To calculate the multiscale integral orientations, we first calculate the gradient of an image in x and y direction. As with other similar type of algorithms [DT05] it has been found that the most suitable form of gradients are the simple gradient filters in horizontal and vertical directions. Using these gradients, we calculate the directions for each pixel using the equation $\theta = 180.\pi.\arctan\left(\frac{d_{y(i,j)}}{d_{x(i,j)}}\right)$, and magnitude of each direction is calculated using the equation, $m = \sqrt{d^2_{x(i,j)} + d^2_{y(i,j)}}$. After calculating the orientations for each pixel in the image, angles are quantized and then scaled by the magnitude at that locations. The number of angles usually used are 8 and 16 but they can be be changed according to the utility of the application, but mostly 8 angles are sufficient for most image analysis tasks. In the next step separate matrices are constructed which are composed from the each quantized and weighted angles. Final orientation histogram is shown in equation below, where W is the matrix of quantized and scaled orientations, $(rx1, ry1)$ and $(rx2, ry2)$ are the starting and ending points of the bounding box of the region, and g is the number angles for which we want to calculate the descriptor :

$$\partial = \left\{ \begin{array}{ccc} \sum_{x=rx1}^{rx2} \sum_{y=ry1}^{ry2} W_{\theta 1}(x,y) & \sum_{x=rx1}^{rx2} \sum_{y=ry1}^{ry2} W_{\theta 2}(x,y) & \cdots \\ & \sum_{x=rx1}^{rx2} \sum_{y=ry1}^{ry2} W_{\theta g}(x,y) & \end{array} \right\} \quad (1)$$

Each block in the final descriptor is a histogram composed of the gradient vectors in the block in the each layer a sample layer composed of 16 blocks is shown in 2. Block sizes can be different depending upon the type of object being analyzed and layers will basically overlap each other. Usual layer sizes also depend on the application for calculations performed in normal circumstances layer sizes of 4x4, 6x6, 8x8 are enough, where each layers contains the given number of blocks in the given composition of layers. An example configuration of layers is shown in the image 3.

Speed optimization is a necessary process for this kind of descriptor because without speed optimization a normal computation of this descriptor over a complete image will be very time consuming operation. To optimize the speed of the multiscale integral orientation descriptor calculation, we convert each direction into an integral image which basically provides a constant time calculation of each block in the descriptor in constant time. This is described in the next section. Integral images play an important role in the calculation of this descriptor because they provide a fast way to compute

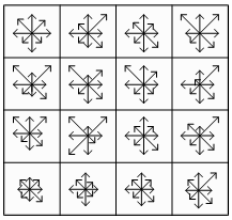

Figure 2: *Single Layers composed of blocks in MSIO*

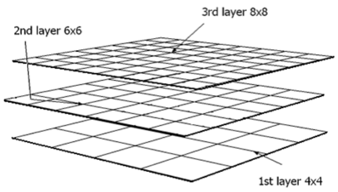

Figure 3: *Multiscale Integral Orientation Layers*

descriptor in constant time for any rectangular region in the image. Integral images were first used in computer vision by the [VJ04, VJ01] in the seminal paper on object detection. Integral images are also known as the cumulative sum of or summed area tables. Integral images by definition are the cumulative sum of the numeric values inside a new table. The integral images can be computed efficiently in a single pass over the image, using the fact that the value in the summed area table at (x,y) is just the combination of the previous computed values in the neighborhood of the current pixel. In calculating the multiscale integral orientations, if we are interested in calculating the descriptor with integral images, we calculate the integral images for the each quantized angle in the image. Let us assume, if we are interested in calculating the descriptor of an image region and we have chosen 8 angles to quantize then, we create 8 integral images. These integral images of the orientation vectors can be used in constant time to calculate the descriptor of any rectangular region in the image. To normalize the descriptor, we apply the block based normalization which although not that accurate is enough for fast computation of the descriptor. The gaussian values are precomputed for each block. These values are then multiplied with each block according to the position. Multiscale orientations can be beneficial in many cases as they provide more stability and accuracy then just normal orientation histograms. The orientations have been proved very efficient for detection of objects in many cases but in some case they have also given poor results.

4.1. 3D Shape matching

To calculate the descriptor of these objects, we combine these images as a single image, as most of the models in our test database are basically aligned so, we have chosen to rely on the linear combination of these objects to form a single image but in the case of non aligned object first an alignment algorithm can be used to align them then the depth images are generated from these aligned models.

Layer 1	4	5
Layer 2	8	10
Layer 3	12	15
Layer 4	16	20
Layer 5	20	25
Layer 6	24	30

Table 1: *Layer configuration for MSIO for 3D matching.*

The descriptor is calculated using the multiscale integral orientations for the complete image but keeping in the mind the separation of each projection. So for a 20 images based projection image, we have used 6 layers and the configuration of layers is shown in the table 1. This is done so that they localize the individual projection separately. This whole process can be done without combining the images altogether but for performance reason, we have chosen a single image approach. In this way descriptors for all the models in the database are generated. The descriptor for the query model is then compared with the descriptors of the objects in the database and the models which have the low distances are the models with the most similar shape.

5. Results

Experiments were performed on Princeton Shape Benchmark(PSB) [SMKF04]. Most of the results were obtained by setting the camera on the 20 different positions around the model. The depth images calculated from these 20 positions are combined to form a single image for which, we calculate the MSIO descriptor. The 3D shape searching performance is characterized by the Precision and Recall curve for more details see [SMKF04]. The precision and recall results are shown in Figure 4 for the Princeton benchmark database using the above described algorithm. The results show the comparison between our method which is based on the MSIO based descriptor and the results of some other algorithms. In the figure 4, D2 [OFCD02] is one of the oldest algorithms but it is also one of the most simplest algorithm. Another algorithm, we have compared is based on the spherical harmonics [KFR03], which is shown in the results with the tag SHD. Final comparison is done with the silhouette based algorithm [CTSO03] and it is marked as LFD in the results.

From discussion above we can deduce that MSIO algorithm has comparative performance to some of the algorithms described above. The results of the LFD based algorithm are very close to our method, but it is interesting to mention that the comparison speed for our method is many

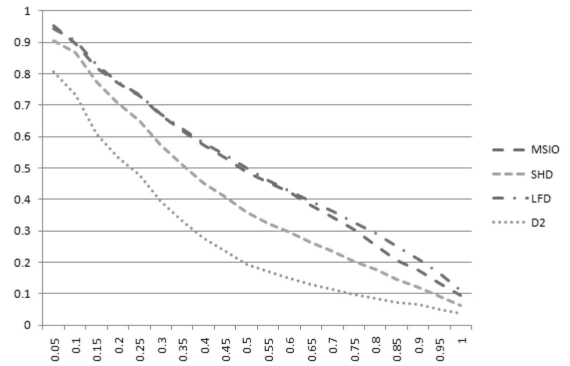

Figure 4: *Results obtained from PSB with test and training dataset*

times faster than LFD, because in LFD method, the permutations of the set of descriptors are used to compute the minimum distance which is a time consuming process.

6. Conclusion

The algorithm presented here is extensible and we have only explored few parameters, but we can still improve the performance by increasing the number of the layers, and the distribution of the blocks to achieve better results. But the main goal of this research is to show the effectiveness of the multi-scale integral orientations for the field of 3D shape retrieval. Some of the previous techniques for 3D shape searching can be reformulated using this algorithm, which will increase their accuracy. Further improvements are possible by extending this algorithm by using hybrid feature vectors, which will improve the accuracy a bit further. In future we plan to use bag of features and machine learning based approaches to improve the results. In future we also plan to target some other kind of 3D Shape matching problems, such as partial matching and models with articulation.

References

[AKKS99] Ankerst M., Kastenmiiller G., Kriegel H., Seidl T.: 3D Shape Histograms for Similarity Search and Classification in Spatial Databases. *Advances in Spatial Databases: 6th International Symposium, SSD'99: Hong Kong, China, July 20-23, 1999: Proceedings* (1999). 1

[CTSO03] Chen D., Tian X., Shen Y., Ouhyoung M.: On Visual Similarity Based 3D Model Retrieval. *Computer Graphics Forum 22*, 3 (2003), 223–232. 1, 3

[DT05] Dalal N., Triggs B.: Histograms of oriented gradients for human detection. In *International Conference on Computer Vision & Pattern Recognition* (INRIA Rhône-Alpes, ZIRST-655, av. de l'Europe, Montbonnot-38334, June 2005), Schmid C., Soatto S., Tomasi C., (Eds.), vol. 2, pp. 886–893. 2

[FHK*04] Frome A., Huber D., Kolluri R., Bulow T., Malik J.: Recognizing Objects in Range Data Using Regional Point Descriptors. *Computer Vision, ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004: Proceedings* (2004). 2

[JJY*04] Jiantao P., Jiantao P., Yi L., Guyu X., Hongbin Z., Weibin L., Uehara Y.: 3d model retrieval based on 2d slice similarity measurements. In *Proc. 2nd International Symposium on 3D Data Processing, Visualization and Transmission 3DPVT 2004* (2004), Yi L., (Ed.), pp. 95–101. 2

[KFR03] Kazhdan M., Funkhouser T., Rusinkiewicz S.: Rotation invariant spherical harmonic representation of 3D shape descriptors. In *Proc. Eurographics Symposium on Geometry Processing (SGP'03)* (2003), vol. 43, pp. 156–164. 3

[KH90] Khotanzad A., Hong Y.: Invariant image recognition by Zernike moments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 12*, 5 (1990), 489–497. 2

[Low99] Lowe D.: Object recognition from local scale-invariant features. vol. 2, Corfu, Greece, pp. 1150–1157. 2

[Low04] Lowe D. G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110. 2

[LTN06] Laga H., Takahashi H., Nakajima M.: Spherical Wavelet Descriptors for Content-based 3D Model Retrieval. *Proceedings of the IEEE International Conference on Shape Modeling and Applications 2006 (SMI'06)-Volume 00* (2006). 2

[OF] Ohbuchi R., Furuya T.: Accelerating Bag-of-Features SIFT algorithm for 3D model retrieval. 2

[OFCD02] Osada R., Funkhouser T., Chazelle B., Dobkin D.: Shape distributions. *ACM Transactions on Graphics (TOG) 21*, 4 (2002), 807–832. 1, 3

[OOFB08] Ohbuchi R., Osada K., Furuya T., Banno T.: Salient local visual features for shape-based 3D model retrieval. In *Shape Modeling International* (2008). 2

[PPT*08] 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), pp. 9–16. 2

[PTK07] Passalis G., Theoharis T., Kakadiaris I.: PTK: A novel depth buffer-based shape descriptor for three-dimensional object retrieval. *The Visual Computer 23*, 1 (2007), 5–14. 2

[SMKF04] Shilane P., Min P., Kazhdan M., Funkhouser T.: The Princeton Shape Benchmark. *Shape Modeling Applications, 2004. Proceedings* (2004), 167–178. 2, 3

[TV08] Tangelder J., Veltkamp R.: A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications 39*, 3 (2008), 441–471. 1

[VJ01] Viola P., Jones M.: Rapid object detection using a boosted cascade of simple. In *Proc. IEEE CVPR 2001* (2001). 3

[VJ04] Viola P., Jones M. J.: Robust real-time face detection. *International Journal of Computer Vision 57*, 2 (2004), 137–154. 3

[Vra04] Vranic D.: *3D model retrieval*. PhD thesis, University of Leipzig, Germany, 2004. 1

[VS00] Vranic D., Saupe D.: 3D Model Retrieval. *Proc. Spring Conference on Computer Graphics and its Applications (SCCG2000), May* (2000), 3–6. 1