Enhanced 3D Model Retrieval System through Characteristic Views using Orthogonal Visual Hull

Shuen-Huei Guan* Ming-Kei Hsieh* Chia-Chi Yeh* Bing-Yu Chen†

National Taiwan University

1 Introduction

The introduction of the Princeton Shape Benchmark database (PSB) [Shilane et al. 2004] has given rise to extensive research into ways to accurately perform 3D matching. Among the extant methods proposed, the LightField Descriptor (LFD) [Chen et al. 2003] based method currently is the most accurate, but it suffers from a deficiency in on-line matching speed. The contribution of this research is twofold. First, fewer characteristic views are used instead of huge images to boost the on-line retrieval time. Second, depth information is employed to keep the retrieval accuracy.

2 Description

In Chen et al.'s system [Chen et al. 2003], several images, taken from different views around, are used to represent a 3D model, which are then transformed into descriptors. Thereafter, the matching of any two 3D models is achieved by matching two groups of 2D images by those descriptors. 100 images for each model are required, represented by 4 descriptors, which are Fourier descriptor for contour region, Zernike moment descriptor for shape region, circularity and eccentricity.

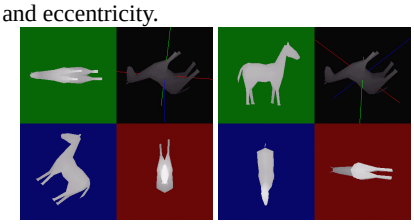

Figure 1: Different pose estimations by PCA (left) and OVH (right).

Our proposal to improve on-line speed is based on reducing the number of images required for each 3D model. Some other research made use of three principal axes by Principal Component Analysis (PCA), applied in pose estimation phase. Although PCA has nice meaning in statistics, it does not fit most cases in 3D model retrieval and its resulting views do not fit user's input of 2D sketches, as illustrated in Figure 1.

Hence, we propose Orthogonal Visual Hull (OVH) to establish pose estimation. Visual hull is an algorithm to reconstruct a 3D model from several views. Given a 3D model and images taken around, since visual hull does not capture concavities, the volume of the reconstructed model is always greater. In other words, the smaller volume we get, the closer approximation it fits the original model. There are two criteria to determine the number and mutual relation of views: 1) The views can be transformed into characteristic axes without obstruction and 2) the visual hull by these views is very close to the original model. Therefore, we chose three orthogonal views because there are three characteristic axes in pose estimation, and relatively orthogonal views carve much space in most cases. Then the next task is to get the views forming the least volume. The global minimum is extracted by the heuristic: (i) find local extrema from widely varying starting orientation of the 3D model by an iterative stochastic method or downhill simplex algorithm as Figure 2, (ii) pick the most extreme of these. The method is slow, but off-line and we are exploring to use faster algorithms, like conjugate gradient descent. Sometimes, the extreme is not unique. In our experiment of randomly picking one extrema to 3D model retrieval system [Chen et al. 2003], the resulting retrieval accuracy is nearly the same.

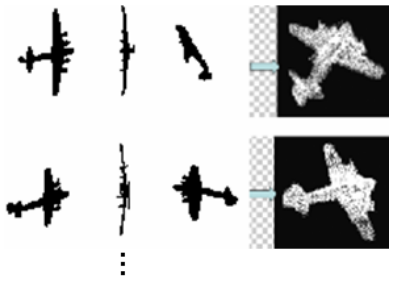

Figure 2: Heuristically pick the characteristic views that form the least volume.

Using fewer images raises the possibility of losing information necessary for 3D matching. By employing the Generic Fourier Descriptor (GFD) [Zhang and Lu 2002], we are able to compensate the information loss because it supplies additional depth information. Therefore, for each image, its contour, shape and depth information are all covered. In order to make use of GFD, the system takes 2 images for each characteristic axes. Consequently, the total number of images taken from each 3D model is reduced from 100 to 6. The number of pair-wise image matching is reduced from 5460 to 24. And the retrieval accuracy is nearly the same to the original system.

In summary, this research enhances the extant best matching method [Chen et al. 2003] by two means: 1) improving the on-line speed 93 times faster at the cost of 5 times deceleration of the off-line process, and 2) keeping the retrieval accuracy. The proposed OVH is also a brand new idea for pose estimation. Additionally, the shapes of the resulting characteristic views generated by OVH more closely fit human visual perception.

	Storage Size (bytes)	Off-line Generate Time (s)	On-line Match Time (ms)
LFD	4,700	3.25	1.300
LFD with OVH	2,592	15.72	0.014

References

CHEN, D.-Y., TIAN, X.-P., SHEN, Y.-T., AND OUHYOUNG, M. 2003. On visual similarity based 3d model retrieval. *Conputer Graphics Forum 22*, 3, 223–233. (Proc. EG 2003).

SHILANE, P., MIN, P., KAZHDAN, M., AND FUNKHOUSER, T. 2004. The princeton shape benchmark. In *Proc. SMI 2004*.

ZHANG, D., AND LU, G. 2002. Generic fourier descriptor for shape-based image retrieval. In *Proc. ICME 2002*, vol. 1, 425–428.

*e-mail:{drake, lionkid, id5}@cmlab.csie.ntu.edu.tw

†e-mail:robin@ntu.edu.tw