HARMONIC SILHOUETTE MATCHING FOR 3D MODELS

Ameesh Makadia, Mirkó Visontai, and Kostas Daniilidis

Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA, 19104, USA
{makadia, mirko, kostas}@cis.upenn.edu

ABSTRACT

The ability to perform fast retrieval from a database of 3D models is becoming a growing necessity as the number of models in circulation is rapidly increasing. Several of the many existing methods dealing with this problem consider similarity measures based on visual appearance. This idea of comparing models using their respective silhouettes performs well on a number of benchmarks, but comes with a few inherent limitations, the biggest of which is that at the time of comparison all possible rotational alignments between two models need to be considered. In this paper we present two retrieval algorithms based on a silhouette representation. The first method shows how model similarity can be computed using fast harmonic matching techniques, and the second method reduces the problem to fast vector differencing using rotation-invariant properties of the representations.

1. INTRODUCTION

Laser-scanned objects, CAD models, and image-based reconstructions are just a few of the sources contributing to the rapidly growing number of publicly available 3D models. Along with these vast resources of 3D collections comes the need for a fast, large-scale model retrieval and matching system. Although the availability of 3D information has sparked a number of methods which take advantage of the complex geometric information for each model, some successful methods are based on visual similarity [1, 2]. Object silhouettes can provide enough information for recovery, eliminating the need to process often complex 3D structural or surface information. The success of silhouette-based algorithms relative to those which rely on descriptive local features can in part be attributed to the fact that local geometric structure may vary widely among objects from the same class.

Another popular class of methods considers global representations of the models ([3, 4, 5, 6]) and in particular cases the representations of the models analyzed using the spherical harmonic transform ([3, 7], among others). We will also use the spherical harmonic representation and its rotational invariants in a similar manner, as will be explained in later sections.

The methods we present in this paper are motivated by the visual similarity method of [1]. The basic idea is based on approximating the light field [8] of a 3D model by capturing silhouettes from a fixed set of positions on the sphere. While their method proves promising, it has a few inherent limitations. The spherical positions of the silhouettes are restricted by the fact that comparisons can only be made for rotations which map the samples onto themselves. There is no natural way to perform approximate comparisons, which is a necessity considering very large databases may be queried, and there is no flexibility in the rotations that can be tested. Finally, for any pair of models, a brute force traversal through all possible rotational alignments must be made (although a hierarchical approach can help speed up the comparisons). We will present in this paper two methods for model retrieval which address these concerns. The first method treats computes similarity as a correlation of tangent bundles on the sphere (where each tangent plane represents a different model silhouette). We show how such a comparison can be computed as a multiplication in the Fourier domain, alleviating the need to perform a comparison for each possible rotation directly in the spatial domain. The second method also utilizes the silhouette-based representation. Instead of performing correlations, harmonic rotational invariants of the silhouette representations are used to encode a small feature vector. In this case the similarity between two models is just the Euclidean distance between two such vectors.

2. LIGHT FIELD REPRESENTATION

The model retrieval method in [1] considers silhouettes taken at 20 (in practice only 10 from one hemisphere are needed) different locations on the sphere, and there are only 60 rotations which map these points onto themselves. In order to create a denser sampling of silhouette positions and to consider more rotations, the only solution is to recreate the configuration of 10 vertices at a different reference orientation. Repeating this configuration 9 times, a total of 100 silhouettes are generated, and the number of rotations which need to be traversed before a distance measure is obtained rises to 5,460 [1]. We would like to develop a method that is more flexible to the number of silhouettes that can be used, a method which has natural approximation capabilities for speed considerations, and one that does not require the brute-force traversal of all possible rotational alignments.

We begin by describing a modified representation of the light field for a 3D model. Instead of capturing silhouettes from pre-determined positions, we can specify the locations given a desired resolution. Given a spherical bandwidth B, both of the spherical angles (θ for co-latitude, and ϕ for longitude) will be sampled uniformly at $2B$ locations, resulting in a total of $2B^2$ samples on the sphere. See figure 1 to visualize this sampling on the sphere. Only the value of B needs to increase until the desired spacing is achieved. A silhouette will then be generated from each of these samples, and this collection of silhouettes will be the model's light field representation. The silhouette at sphere point $p(\theta, \phi)$ is a binary function obtained by orthographically projecting the 3D model onto the tangent plane at p. The orientation of the tangent plane is given by the rotation $R = R_z(\phi)R_y(\theta)$ (i.e. the tangent plane at the north pole maps onto the tangent plane at p via the rotation R).

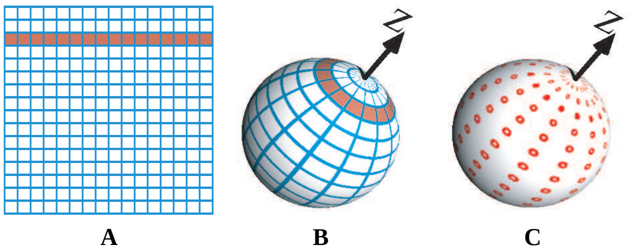

Fig. 1. **A** shows a spherical grid with 256 samples. The sphere is sampled uniformly in the angles, creating a square grid. **B** depicts the corresponding regions mapped onto the sphere. The highlighted samples correspond to the highlighted row in **A**. The sample centers (origins of the tangent planes) are shown in **C**.

3. MODEL SIMILARITY AS CORRELATION

We now present the first method to measure similarity of two models, which are represented with their respective light fields. For the moment, consider the continuous light field function $L(p, v)$ which gives the binary value of the point v on the silhouette taken from spherical location p. If we process the silhouettes to generate smaller features which may encode some translational or rotational invariants, then our light field representation can be stored as a vector-valued function on the sphere, given as $L(p, x)$. In other words, the value at $L([0 \;\; 0 \;\; 1]^T)$ is some feature vector computed for the silhouette obtained from the north pole, and the value of x denotes an element in this vector. For simplicity we will use the centroid-distance functions and Zernike moments used in [9, 1]. If we define the similarity of two feature vectors to be their correlation coefficient, then we can claim the similarity of two 3D models as the maximum correlation coefficient of their light field representations over all possible rotational alignments. In other words, model similarity is given by the maximum of the following rotational function:

$$G(R) = \int_x \int_p L_1(p, x) L_2(R^T p, x)\, dp\, dx \qquad (1)$$

The key here is in recognizing the underlying spherical integration as a correlation of two spherical signals. It has been shown that the spherical correlation $G'(R) = \int L_1(p)L_2(R^T p)\, dp$ can be estimated efficiently as a multiplication in the spherical Fourier domain. We refer readers to [10, 11] for the details, and for other applications of the spherical correlation alignment. We will write $\hat{f}^l_m$ for the spherical Fourier coefficients of degree l and order m, $\hat{f}^l$ for the vector of $(2l + 1)$ coefficients $\hat{f}^l = [\hat{f}^l_l \hat{f}^l_{l-1} \cdots \hat{f}^l_{-l+1} \hat{f}^l_{-l}]^T$, and $\hat{G'}^l_{mp}$ for the coefficients of the Fourier transform defined on the rotation group $SO(3)$. The spherical correlation theorem states that the Fourier transform of the spherical correlation function G can be obtained as

$$\hat{G'}^l_{mp} = \hat{L_1}^l_m \hat{L_2}^l_p$$

and so the samples of the correlation function are recovered with $G'(R) = ISOFT(\hat{G})$ where $ISOFT$ is the inverse SO(3) Fourier transform. Our light field correlation can now be written as

$$G(R) = \int_x ISOFT(\hat{G})(x)\, dx \qquad (2)$$

If we let B represent the bandwidth of the spherical functions (meaning only coefficients up to degree $B - 1$ are computed), then the inverse SO(3) Fourier transform will leave us with with $2B$ samples in each of the three Euler angles, giving us an accuracy up to $\pm \left(\frac{180}{2B}\right)^\circ$ in α and γ and $\pm \left(\frac{90}{2B}\right)^\circ$ in β. Here α, β, γ are the traditional ZYZ Euler angles. Fast spherical Fourier transforms (SFT) can be computed in time $O(B^2 \log^2 B)$, and fast $SO(3)$ Fourier transforms (SOFT) in $O(B^3 \log^2 B)$ [12, 10]. These fast algorithms require uniform sampling in the angles, and this is the primary motivation for our choice of spherical sampling.

4. ROTATIONAL INVARIANTS

For some applications that require searching through large databases, pairwise model comparison using fast correlations may not be sufficiently fast. The correlation function we are considering measures the alignment quality for all orientations, but for retrieval only the overall best alignment is important. We propose to use invariant properties of the spherical harmonic coefficients (see [7, 13] for other uses of these invariants) to encode a feature vector which does not depend on the orientation of the original polygonal model.

As spherical functions are rotated by elements of the rotation group $SO(3)$, the Fourier coefficients are "modulated" by the irreducible representations of $SO(3)$:

$$f(\eta) \mapsto f(R^T\eta) \Longleftrightarrow \hat{f}^l \mapsto U^l(R)^T\hat{f}^l \tag{3}$$

The U^l matrix representations of $SO(3)$ are unitary, and ensure the distribution of energy among frequencies does not vary.

$$||U^l(R)\hat{f}^l|| = ||\hat{f}^l||, \forall R \in SO(3) \tag{4}$$

By considering only the magnitudes of the coefficient vectors, the light field feature size is further reduced. The total size is equal to $\lfloor \frac{B}{2} \rfloor N$, where B is the spherical bandwidth and N is the size of the individual silhouette features (in the next section it will become clear why it is $\frac{B}{2}$ instead of just B). For example, consider a model for which we render a very large number of silhouettes (a bandwidth of $B = 17$ means we must render 578 silhouettes in one hemisphere). Assuming we keep 35 Zernike coefficients and 10 contour distance coefficients for each silhouette (as in [1]), we can represent a 3D model with just $(35 + 10) * 8 = 360$ elements, which are stored in one vector. The distance between models is just the Euclidean distance between these vectors.

5. COMPUTATIONAL CONSIDERATIONS

Knowing that the silhouette generated at any point p on the sphere is just a projection of the model onto the silhouette plane, it is clear that the silhouette generated from the antipodal point $-p$ will just be a reverse image. From the invariance of the Zernike and contour features, $L(p, x)$ is has the even property such that $L(p, x) = L(-p, x)$. For such functions, all spherical Fourier coefficients of an odd degree are zero. Also, since the spherical function is real-valued, the coefficient vectors $\hat{f}^l$ exhibit the hermitian property that opposite orders are related by conjugation. These two facts mean we only need to compute $\hat{f}^l_m$ for l even and $m \geq 0$.

Regarding the rotational matching, it is tempting to perform retrieval hierarchically, since a natural coarse-to-fine similarity can be computed by simply varying the bandwidth at which the correlation is performed. Let B be the bandwidth of a spherical function f, and let $B' < B - 1$. We will denote the inverse SFT of $\hat{f}^l$ as $ISFT\{\hat{f}^l\}_{l=0,\ldots,B-1}$. We use the subscript $(l = 0, \ldots, B - 1)$ to denote the inverse transform is using all degrees less than B, meaning the standard inverse transform. It is straightforward to see the following

$$\begin{aligned} ISFT\{\hat{f}^l\}_{l=0,\ldots,B-1} &= ISFT\{\hat{f}^l\}_{l=0,\ldots,B'} + \\ &\quad ISFT\{\hat{f}^l\}_{l=B'+1,\ldots,B-1} \end{aligned}$$

This same decomposition principle holds for the $SO(3)$ Fourier transform. This shows that reconstructing your signal at a higher resolution or bandwidth includes the computation for a lower resolution transform, so there is no wasted or redundant processing in a coarse-to-fine approach.

Regarding computation times, the time to compare two models using rotational invariants is less than 0.0001 seconds (this is just vector differencing) on a 2GHz machine. The execution time for measuring similarity using correlation is just less than 0.1 seconds (for a bandwidth of $B = 17$, meaning 578 silhouettes). The remaining computational effort is spent on generating the model silhouettes, and figure 2 plots the time required to generate silhouettes at varying bandwidths and numbers of polygons. The current implementation is a preliminary effort and does not take advantage of the many optimization opportunities within OpenGL or modern graphics cards.

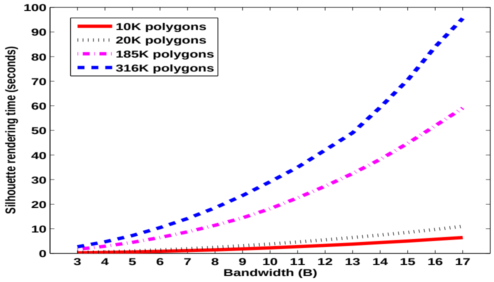

Fig. 2. A plot showing the rendering time to generate silhouettes for various bandwidths and model sizes. For a bandwidth B, the number of silhouettes rendered is $2B^2$. For example, it takes approximately 95 seconds to generate 578 silhouettes of a model with $316K$ polygons.

6. RESULTS

Two versions of the correlation method and two versions of the rotational invariants comparison method were entered in SHREC, the 3D Shape Retrieval Contest [14, 15]. The four entries were entered under the name "Makadia." The first run compared models using the correlation method, but only Zernike features were used. The second run was again the correlation method, but both Zernike and contour-distance features were used. The third run compared models with the faster invariant vector comparison, using only Zernike features, while the fourth run performed the vector comparisons using Zernike and contour-distance features. Of the 17 main categories of analysis, the correlation methods finished first in 11. The faster vector-invariant comparisons, however, finished in the middle of the pack in all categories. The obvious tradeoff is in accuracy versus comparison times.

Figure 3 shows that the accuracy of the correlation method may be achieved without performing pairwise correlations

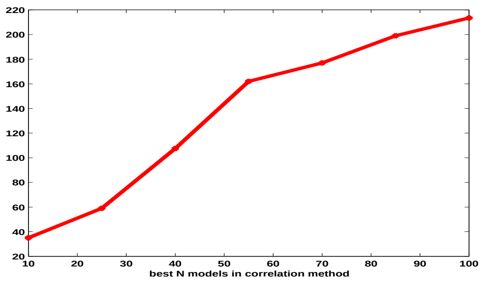

Fig. 3. This plot shows how many models in the ranked list (obtained with vector comparisons) you need to traverse before finding 50% of the best N matches in the ranked list obtained with correlations. The plot shows the median over all queries. For example, 50% of the best 100 matches from the correlation method will appear in the first 213 matches from the ranked list obtained with fast vector comparisons.

against an entire database. The faster invariant vector comparisons can be used to generate a much smaller set of possible matching objects, and the correlations can be used to provide an accurate ranking within this pruned set.

7. CONCLUSION

In this paper we presented two new approaches for comparing 3D models. The first method considers the best possible correlation alignment between the models' light field representations. The benefit of this approach is in the fast correlation estimation using the Spherical Fourier transform, and the flexibility and approximation allowed by varying the number of coefficients used. The second method utilizes the rotational invariants of the spherical transform to encode an entire model light field with just one small feature vector. This allows us to compute model distance with fast Euclidean distance measurements.

The algorithms proposed in this document can be extended in many ways. It may be beneficial to retain more information beyond a binary silhouette image. One could capture surface orientations, or a depth map, where each pixel marks the distance from the model to the silhouette plane, to help disambiguate between very different surfaces which generate similar silhouettes. It is also of importance to investigate

8. REFERENCES

[1] Y.-T. Shen D.-Y. Chen, X.-P. Tian and M. Ouhyoung, "On visual similarity based 3D model retrieval," in *Eurographics*, 2003.

[2] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The princeton shape benchmark," in *Shape Modeling International*, Genova, Italy, June 2004.

[3] D. V. Vranic and D. Saupe, "3d model retrieval with spherical harmonics and moments," in *Proceedings of the 23rd DAGM-Symposium on Pattern Recognition*, London, UK, 2001, pp. 392–397, Springer-Verlag.

[4] M. Ankerst, G. Kastenmüller, H.-P. Kriegel, and T. Seidl, "3D shape histograms for similarity search and classification in spatial databases," in *Advances in Spatial Databases, 6th International Symposium, SSD'99*, R. Güting, D. Papadias, and F. Lochovsky, Eds., Hong Kong, China, 1999, vol. 1651, pp. 207–228, Springer.

[5] B. K. P. Horn, "Extended gaussian images," *IEEE*, vol. 72, pp. 1671–1686, 1984.

[6] M. Kazhdan, *Shape Representations and Algorithms for 3D Model Retrieval*, Ph.D. thesis, Princeton University, 2004.

[7] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Symposium on Geometry Processing*, June 2003.

[8] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. of ACM SIGGRAPH*, 1996, pp. 31–42.

[9] D. S. Zhang and G. Lu, "An integrated approach to shape based image retrieval," in *Proc. of 5th Asian Conference on Computer Vision (ACCV)*, Melbourne, 2002, pp. 652–657.

[10] P. J. Kostelec and D. N. Rockmore, "FFTs on the rotation group," in *Working Paper Series, Santa Fe Institute*, 2003.

[11] A. Makadia and K. Daniilidis, "Rotation recovery from spherical images without correspondences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, 2006.

[12] J.R. Driscoll and D.M. Healy, "Computing fourier transforms and convolutions on the 2-sphere," *Advances in Applied Mathematics*, vol. 15, pp. 202–250, 1994.

[13] A. Makadia and K. Daniilidis, "Direct 3D-rotation estimation from spherical images via a generalized shift theorem," in *IEEE Conf. Computer Vision and Pattern Recognition*, Wisconsin, June 16-22, 2003.

[14] AIM@SHAPE, ," `http://give-lab.cs.uu.nl/shrec/shrec2006/`.

[15] R. Typke, R. C. Veltkamp, and F. Wiering, "Evaluating retrieval techniques based on partially ordered ground truth lists," in *Proceedings International Conference on Multimedia & Expo*, 2006.