CONCENTRIC RING SIGNATURE DESCRIPTOR FOR 3D OBJECTS

Hien Van Nguyen

University of Maryland

Fatih Porikli

Mitsubishi Electric Research Laboratories

ABSTRACT

We present a 3D feature descriptor that represents local topologies within a set of folded concentric rings by distances from local points to a projection plane. This feature, called as Concentric Ring Signature (CORS), possesses similar computational advantages to point signatures yet provides more accurate matches. It produces more compact and discriminative descriptors than shape context. It robust to noise and occlusions. As opposed to spin images, CORS does not require the point normal estimations, therefore it is directly applicable to sparse point clouds where the point densities are insufficiently low. Under the same settings, we demonstrate that the discriminative power of CORS is superior to conventional approaches producing twice as good estimates with the percentage of correct match scores improving from 39% to 88%.

Index Terms— 3D descriptor, matching, retrieval.

1. INTRODUCTION

There has been several attempts in 3D data representation to achieve discriminative power, rotation invariance, robustness to noise, and computational efficiency at the same time. Algorithms developed for this purpose can be classified as global and local descriptors depending on their support regions.

Among popular global descriptors, we can list the *extended Gaussian image* [1] that maps the weighted surface normals onto a sphere, *shape distribution* [2] that forms a histogram of randomly samples pairwise point distances, *super-quadratic* [3], *spherical attribute images*, *COSMO* [4], etc. In general, global shape descriptors are suitable for rigid, cleanly segmented and highly convex objects, however they quickly deteriorate in case of clutter, occlusion, and articulated motion.

Local descriptors are defined on a smaller subset of the object data points. For example, *spin image* [5] considers a cylindrical support region, whose center is at the basis point and north pole is oriented with the surface normal estimate at the basis, and accumulates the points within the support region onto 2D image pixels by keeping a weight proportional to the number of points falling into every subvolume. *3D shape context* [6] is similar to the spin image except that the support region is a sphere. The sphere is segmented into subvolumes by dividing evenly in the azimuth and elevation, and logarithmically in the radial dimensions. A degree of freedom in the azimuth direction remains, which needs be removed before carrying out feature matching. *Spherical harmonic* [7] proposes a wave decomposition to make the shape context descriptor rotation invariant. *Point signature* [8] extracts a 1D vector that represents local volumes by the distance from 3D curves to a plane. Since it does not strictly require normal estimations, the point signature is computationally more advantageous. On the other hand, its oversimplification makes it is less descriptive. Both spin image and shape context inherently generate sparse matrices where most of the coefficients are null. As a result, distance computation becomes inevitably sensitive to the density of the point cloud, particularly for range scan images. Besides, they need error-prone normal vector estimation.

Here, we introduce the concentric ring signatures. CORS is a patch-based descriptor to represent local 3D topologies, thus it is a natural choice for range scan images that capture object surfaces. It describes the local volume by a 2D matrix coefficients that correspond to the deviation of local neighborhood from a fitted reference plane to the local samples. Unlike the volume based descriptors (such as spin image, shape context) that use comparably much higher number of bins, which deteriorates the matching performance, and employ spatial histograms, which require the estimation of point density and the normalization of sampling frequencies, our descriptor is compact and robust against acquisition artifacts. CORS retains rotation invariance and captures information on objects pose without sacrificing any discriminative power.

2. CONCENTRIC RING SIGNATURE

To construct CORS we first find the 3D data points within the local support region, then determine a plane of projection, decide the reference orientation in that plane, and finally compute the patch responses that are arranged into a matrix form as illustrated in Fig. 1.

2.1. Local Support

Let p be a point in a 3D cloud data and r the radius of a spherical volume centered on p. We assign the set of points falling within the spherical volume as the local support $\mathcal{S} = \{p_i : ||p_i - p|| \leq r\}$. The choice of the radius r is data-dependent. For example, larger radius is preferred for smooth

978-1-4577-1303-3/11/$26.00 ©2011 IEEE

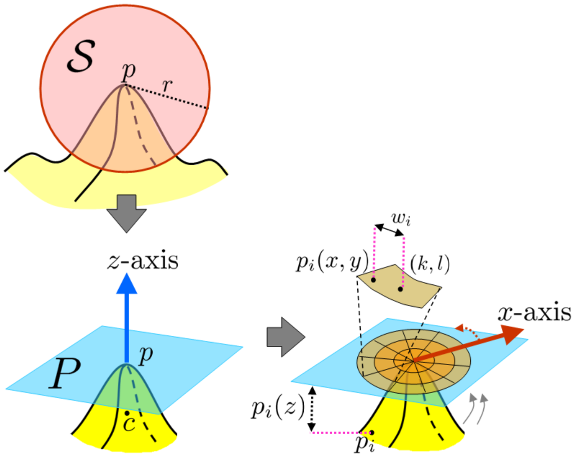

Fig. 1. a) CORS is constructed at p by finding a spherical support region $\mathcal{S}$. b) A plane is fitted to local neighborhood and translated to the point p. The normal direction is taken to be z-axis c) Selecting a reference orientation for x-axis and projecting the distances from the surface to patches.

and rigid shapes while smaller radius is preferred for shapes with articulations or structural variations. As r increases, CORS is more discriminative but more vulnerable against occlusions. In this paper, we perform cross-validation on a subset of databases to choose the value of r that gives best results.

2.2. Plane of Projection and Reference Axes

A plane P is fitted to the local support $\mathcal{S}$. There are two possible choices for plane fitting. One can use all the data points within the local support, fit a plane by least-squares as the system is almost always over-determined, and parallel move the origin of P at p. Alternatively, it is possible to select a subset of points along the perimeter of the local support, e.g. intersecting the sphere support with the object surface.

We define local reference coordinates so that the local descriptor is invariant to camera view. Let c be the (Karcher) mean, that is the coordinate having the minimal overall distance to the other points in the local support, i.e. point c that minimizes $\sum_i \|p_i - c\|$. We set the z-axis to be orthogonal to P and pointing in a direction such as the dot product of the unit vector $\vec{z}$ with vector $\vec{cp}$ is positive. We define a local reference axis (x-axis) so that the local descriptor is invariant to camera view. x-axis points away from p to the projection of the 3D point that has the maximum distance from the fitted plane P within the local support $\mathcal{S}$. y-axis is defined by the cross product $\vec{z} \times \vec{x}$. With such assignments, P corresponds to the xy plane going through point p. These two conditions define a z-axis without any ambiguity.

In case the projection distances from P to the xy plane have more than one peak, multiple reference axes can be generated.

2.3. Populating Patches

After fitting the plane and determining the reference axes, each 3D point p_i in the local neighborhood $\mathcal{S}$ is now represented by a tensor $p_i(x, y, z)$ independent of the camera viewing angle. The z coordinates $p_i(z)$ correspond to the distance from the plane in this tensor, and the xy-plane coordinates $p_i(x, y)$ correspond to the projection on the plane P. We estimate a representative elevation value of the given data points within the patches of this grid as follows:

1. We apply a polar grid along azimuth and radial directions on the xy plane centered on the original point p. The patches of this grid can be considered as the 2D histogram bins. Let $\{(k, l)\}$ be the set of sampled grid locations with $k = 1 \dots K$ and $l = 1 \dots L$, where K and L are the numbers of sampling intervals along the radial and the azimuthal directions, respectively. In other words, we will extract a 2D matrix $F_{K\times L}$ on this grid where each coefficient corresponds to a patch of the grid.

2. At each grid location (k, l), we estimate a representative elevation value $F(k, l)$. Elevation at a location is estimated by interpolating the elevation values of the given 3D points within the four neighboring patches. This significantly improves sparsity related issues, e.g. sudden jumps of the estimated elevation, and boundary issues e.g. being very close to boundary but falling into another bin. The representative elevation score $F(k, l)$ is estimated as follows:

$$F(k,l) = \frac{\sum_i w_i.p_i(z)}{\sum_i w_i} \qquad (1)$$

where p_i are 3D points within the immediate neighboring bins of the bin of (k, l) and the weight is computed as:

$$w_i = \begin{cases} 1/\alpha, & d \leq \alpha \\ 1/d, & \alpha \leq d \leq 2\alpha \\ 0, & \text{otherwise} \end{cases} \qquad (2)$$

and $d = \|(k, l) - p_i(x, y)\|$.

Basically, $F(k, l)$ is the weighted average of elevation of points surrounding (k, l). The contribution of each surrounding point's elevation to the estimation of representative elevation is controlled by a weight w_i negatively proportional to distance to (k, l). Parameter α controls the smoothness of a descriptor. Higher α values yield smoother descriptors while smaller α makes the descriptor sensitive to positional shifts. The choice of α value depends on the sampling interval along azimuth and radial directions. We observed that the average Euclidean distance between bin centers and their adjacent bins is a satisfactory value. Using a fixed value of α makes bins close to the origin in a polar coordinate system look more similar than those further away. α could be

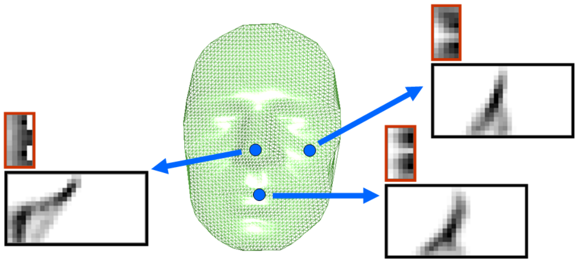

Fig. 2. Illustration of CORS (red border) and spin image (black border) at different points for $r = 15$. For this setting, the dimension of CORS is around 6.5× more compact than that of spin image.

set in an adaptive manner to overcome this issue. Also, imposing a minimum distance constraint enables improving the robustness against small differences in shape very close to the center.

In addition to the mean orthogonal distance from S to the P, the standard deviation of the projection distances and the density of points falling into each bin also possess complementing discriminant power and can be incorporated into similar matrices. An advantage of the mean distance is that it does not require point density estimation and normalization. Figure 2 provides a visual illustration of CORS computed at different locations on a 3D data cloud of human face. Note that the dimension of CORS is 6.5 times smaller than that of spin image. Such dimensional reduction increases the descriptors matching efficiency, yet does not compromise the discriminative power as shown in the experimental analysis section.

In practice, the computation of CORS can be significantly speed up by using the normal vectors whenever available as z-axis of the local reference frame. This eliminates the need of fitting a plane to the neighborhood at every location. For computation of 500 descriptors, CORS takes 1.34 and spin image 2.76 seconds, which indicates CORS is 2× faster than the current state-of-the-art.

3. EXPERIMENTAL ANALYSIS

We conducted several detection, recognition, registration, and retrieval experiments to analyze the performance of CORS descriptors in comparison to existing descriptors including spin images and point signatures. Euclidean distance is used as the dissimilarity measure between two CORS matrices:

$$dist^2(F_1, F_2) = \sum_{k,l} (F_1(k,l) - F_2(k,l))^2. \qquad (3)$$

Matching of CORS descriptors is not limited to Euclidean distance. Since the representation of CORS is in a matrix form, it can be considered to possess a manifold structure where the matching score is defined as the geodesic distance connecting two CORS descriptors on the manifold.

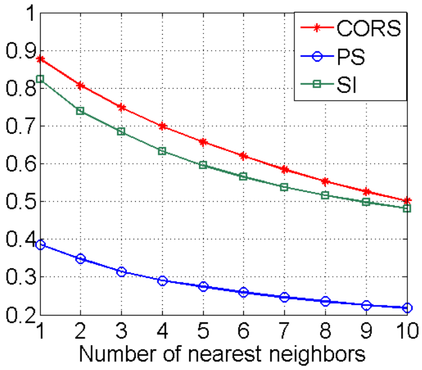

Fig. 3. Ratio of correct matches within k-nearest neighbors.

In the first experiment, we use five synthetic models, each of around 150K points, to compare the saliency, i.e. discriminative power, of point signature, spin image, and CORS. A reference database of 10K signatures is computed at random points on each model. Another 10K set of signatures at randomly sampled points on the same model is used as the query database. We make sure that no point from the reference set is selected for the query set. A query signature at location q_i is said to have a good match to its model at location m_i if their distance $d_i = |m_i - q_i| \leq \epsilon$. We chose ϵ to be 5 times of the scanner resolution. Figure 3 shows the percentage of correct matches within the first k nearest neighbors. CORS out performs both point signature and spin image. The correct matching rate 88% of CORS is approximately 2.3 times higher than that of point signatures 39%. The error rate reduces from 18% for spin image to 12% for CORS, which is more than 33% of improvement. Note that, spin image requires surface normals but not CORS and point signatures.

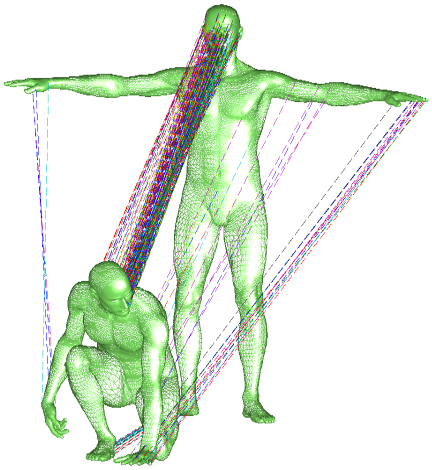

Fig. 4. Point correspondences using CORS for two shapes taken from TOSCA dataset with nonrigid transformations (Note that all correspondences for this pair are correct while the plotted lines are sometimes hidden behind the surface).

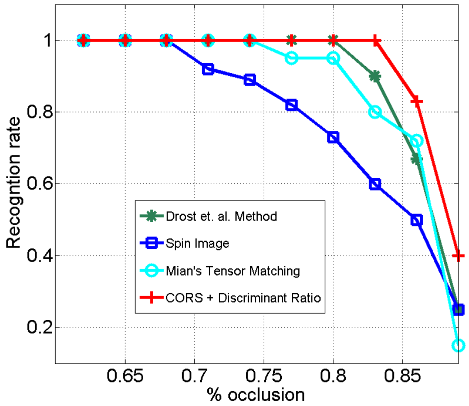

Fig. 5. Recognition rates vs. the percentage of occlusions for spin image, tensor matching, Drost, and CORS.

CORS is robust against articulated motion. Figure 4 gives sample sets of point correspondences found using CORS in two objects. Feature correspondences are accurate even under nonrigid transformations and occlusions.

We run recognition experiments where CORS are computed for a subset of randomly distributed points in both query and database objects, and the alignment is done by RANSAC, and final recognition is determined by CORS distance. We evaluate our method on Mian dataset, which consists of 5 models and 50 scenes taken with a Konica-Minolta range scanner. The scenes in this dataset are highly occluded and cluttered by putting objects very close to each other. We are interested in evaluating the recognition rate that is defined as the number of correct detections over the total number of the scenes. An object is said to be correctly detected if the resulting errors of the translation and pose estimations, compared to the ground truth, are smaller than one-tenth of the object's diameter and 12 °, respectively. These criteria are the same with that of Drost *et al.* [9].

Using CORS our algorithm converges after, on average, only 3 iterations. Figure 5 shows the overall recognition result of our method. As given, it outperforms all other methods in terms of the recognition rate with respect to occlusions. Figure 6 shows sample detection results.

We also run retrieval tests using a bag-of-feature approach is used to retrieve 3D shapes. We compare CORS at a subset of salient points to evaluate the similarity between a model and a query. Sample retrieval results of our method and spin image are presented in Fig. 7.

4. CONCLUSION

We present a compact and discriminative 3D descriptor to represent local topologies of 3D point clouds, which provides superior results for matching and retrieving 3D shapes. In near future, we will explore efficient algorithms for real-time matching of 3D shapes using CORS.

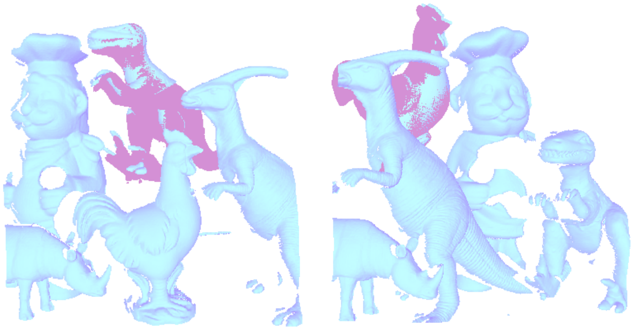

Fig. 6. Recognition and registration of 3D model point clouds into the occluded scenes. The cyan and pink colors are used to render the scenes and models, respectively.

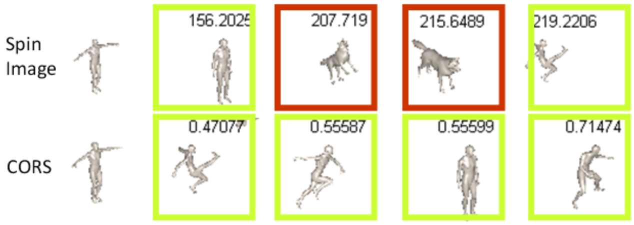

Fig. 7. Top four retrieval results for a sample query of human object. CORS retrieves correct matches.

5. REFERENCES

[1] B. Horn, "Extended Gaussian images," *Proceedings of the IEEE*, vol. 72(12), 1984.

[2] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Transactions on Graphics*, vol. 21, 2002.

[3] F. Solina and R. Bajcsy, "Recovery of parametric models from range images: the case for superquadrics with global deformations," *IEEE PAMI*, vol. 12, 1990.

[4] C. Dorai and Anil Jain, "COSMOS - A representation scheme for 3D free-form objects," *IEEE PAMI*, vol. 19, 1997.

[5] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE PAMI*, vol. 21, 1999.

[6] A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *ECCV*, 2004.

[7] D. Saupe and D. Vranic, "3D model retrieval with spherical harmonics and moments," in *DAGM*, 2001.

[8] C. Chua and R. Jarvis, "Point signatures: a new representation for 3D object recognition," *IJCV*, vol. 25, 1997.

[9] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: efficient and robust 3D object recognition," in *CVPR*, 2010.