Spherical Embedding of Inlier Silhouette Dissimilarities

Etai Littwin, Hadar Averbuch-Elor, Daniel Cohen-Or Tel-Aviv University

In this paper, we introduce a spherical embedding technique to position a given set of silhouettes of an object as observed from a set of cameras arbitrarily positioned around the object. Similar to previous works (e.g., [2, 3, 4]), we assume that the object silhouettes are the only visual cues provided, and thus traditional structure from motion (SfM) techniques based on common feature correspondence cannot be applied successfully. Our technique estimates dissimilarities among the silhouettes and embeds them directly in the rotation space SO(3). The embedding is obtained by an optimization scheme applied over the rotations represented with exponential maps. Since the measure for inter-silhouette dissimilarities contains many outliers, our key idea is to perform the embedding by only using a subset of the estimated dissimilarities. We present a technique that carefully screens for inlier-distances, and the pairwise scaled dissimilarities are embedded in a spherical space, diffeomorphic to SO(3). We show that our method outperforms spherical multi-dimensional scaling (MDS) embedding, demonstrate its performance on various multi-view sets, and highlight its robustness to outliers.

For each view and its associated silhouette, we would like to find the rotations R_i relative to some neutral position for each viewpoint *i*. Let us denote by $D(R_i, R_j)$ the distance between viewpoints *i* and *j*. Note that the camera can produce different views while maintaining the same position in 3D space relative to the object due to rotation around its own principal axis. We assume that a significant portion of dissimilarity measures correlate well with the actual $D(R_i, R_j)$, but our method can tolerate a non-trivial amount of outlier measures.

Given a set of pairwise distances d_{ij} between each pair of viewpoints *i* and *j*, we would like to minimize the following expression in the space of rotations:

$$\sum_{ij} (D(R_i, R_j) - d_{ij})^2$$

The minimization requires the computation of its derivatives with respect to R_i and R_j . Since the first derivatives are not trivial to express with the exponential maps representation, we develop an explicit expression of the first derivatives using the Baker Campbell Hausdorff formula [1].

The problem we address is particularly challenging since the similarity estimates are generally unreliable, and directly applying an MDS embedding introduces a significant distortion. Hence, a robust technique is sought, that may ignore portions of the input data. Our technique finds and considers inlier dissimilarities and ignores outlier ones, and then embeds only a subset of the views (see the illustration in Figure 1). The measure that we employ to estimate the dissimilarities between silhouettes, just like most similarity measures, tends to be more reliable for more similar shapes, and completely unreliable for more dissimilar ones. This may suggest that by simply ignoring large dissimilarity measures, a robust embedding can possibly be obtained. However, as we show in the paper, such a simple approach is not robust enough as some short distance estimates are also erroneous and distort the embedding on the sphere.

The technique that we present is more involved. It carefully defines a graph, that may not necessarily contain all the input silhouettes, nor all their pairwise dissimilarities. The graph is defined by a union of small subsampled matrices, each of which is verified to have a plausible embedding in SO(3). The key idea is to search and sample small sub-sampled matrices that embed well onto a hypersphere. We create an aggregate of such subsampled matrices that have significant overlap and define a graph where the nodes are a subset of the input points and an edge is defined only for a pair that appears in one of the matrices.

The graph of inlier dissimilarities, as a whole, is then embedded into the space of rotations by an optimization that associates relative rotations with

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.



Figure 1: Given a set of silhouettes of an object as observed from a set of cameras arbitrarily positioned around an object, we first compute the full allpairs dissimilarities. We discover the inlier silhouette dissimilarities using our inlier screening technique and obtain a sparse graph. We then perform an optimization to embed the sparse dissimilarities in SO(3). Assuming the contours are associated with photos, then one can place them on a sphere.

the views so that they agree with the dissimilarities defined by the edges of the graph. The direct optimization of our objective function in SO(3) allows solving a rather sparse set of views, without completing large distances as needed in MDS-based techniques. Our embedding technique employs exponential maps and solves the embedding directly in the rotation space SO(3). Our contribution is twofold: First, we present a spherical embedding technique based on exponential maps, and show that it outperforms spherical MDS. Second, we develop an inlier screening technique, and show its robustness to erroneous silhouette dissimilarities.

We show that if the dissimilarities d_{ij} are in full correlation with the ground-truth distances $D(R_i, R_j)$ then our method recovers the rotations accurately. We further show the robustness of our method to erroneous dissimilarities by adding noise to the ground truth data. Moreover, we demonstrate the performance of our method on real data and compare our inlier screening technique with one that uses the k-nearest neighbors (KNN) distances. Our evaluation confirms that under noise or given only a partial distance matrix, our method outperforms the alternatives.

- Sergio Blanes and Fernando Casas. On the convergence and optimization of the baker–campbell–hausdorff formula. *Linear algebra and its applications*, 378:135–158, 2004.
- [2] Yasutaka Furukawa, Amit Sethi, Jean Ponce, and David Kriegman. Structure and motion from images of smooth textureless objects. In *Computer Vision-ECCV 2004.* Springer, 2004.
- [3] Paul McIlroy and Tom Drummond. Reconstruction from uncalibrated affine silhouettes. In *BMVC*. Citeseer, 2009.
- [4] Sudipta N Sinha, Marc Pollefeys, and Leonard McMillan. Camera network calibration from dynamic silhouettes. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–195. IEEE, 2004.