ORDERING RANDOM OBJECT POSES

James Massaro and Raghuveer Rao

Electrical Engineering Department
Rochester Institute of Technology
Rochester, NY 14623
massaro.james@gmail.com

ABSTRACT

Complete or partial three-dimensional reconstruction of objects from multiple angle-views, or poses, is important in several applications such as photogrammetry, machine vision, and computer-aided design. Knowledge of the pose angles and their proper ordering are required for accurate reconstruction. When these multiple angle images are acquired in random order and the angle of view information is not available the poses have to be put into proper order. This work presents an approach based on principal component analysis (PCA) for automatic ordering of random object poses. A measure based on local curvature and correlation of the estimated pose trajectory in a multidimensional manifold is also developed to assess confidence in the ordering. In addition to providing a degree of confidence for pose ordering with single cameras, this measure enhances the pose estimation accuracy in double and multiple camera systems by providing a basis for camera selection for different poses. The paper presents theoretical development and experimental results.

Index Terms— principal component analysis, photogrammetry, pose recognition, pose estimation, multi-camera image processing.

1. INTRODUCTION

Most algorithms for 3d object reconstruction from images require knowledge of the object's pose relative to the camera. Examples of reconstruction algorithms can be found in [1, 2, 3]. Unless *a priori* knowledge of the sequence of poses is available, the methods have trouble matching object geometry with motion. Other methods such as photogrammetry, struggle with the process of making measurements from images with unordered data sets [4]. An unordered image data set is one that corresponds to non-monotonically changing view-angles of an object or scene. This usually occurs when images are taken in asynchronous fashion at random instances in time. Some examples of situations where unordered sequences of images occur, can be found in aerial surveillance, and unsupervised learning of object training sets for humanoid robots [4, 5]. Reordering the image sequence to correspond to ascending or descending order of view-angles is necessary to enable good reconstruction. Such object pose ordering has potential applications in image matching, photogrammetry, scene and object recognition, pose recognition, and facial feature analysis [6, 7].

Methods for object pose recognition can be found in [8]-[10]. These methods use large databases of training sets. There has been research addressing object recognition at various poses using image keypoints [6, 11, 12]. However, these do not address the issue of sequencing the pose order of multiple images. Other papers dealing with image matching [4, 7], also do not address these issues.

This paper develops approaches for pose ordering using the idea that different poses of the object can be connected to construct a smooth manifold in eigenspace. Further, a multi-dimensional confidence measure is developed to assess the accuracy of ordering. This confidence measure can also be used in a multi-camera object recognition system to choose the best view for improved pose ordering accuracy.

2. APPROACH

2.1. Principal Component Analysis

The first step is to perform PCA on the image set. Each image $\mathbf{x}_i$ of size MxN in the randomly captured sequence is formed into an MNx1 vector. This is done for the entire set of images and concatenated into a matrix $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_K]$ of random object poses. Normalization is not done across the image set. This way, illumination and shadows are preserved as these are seen as features for ordering poses. The next step is to calculate the eigenspace of the random set. The eigenspace basis matrix $\mathbf{E}$ is composed of the L largest corresponding eigenvalues for some L.

The projection $\mathbf{g}(\mathbf{x}_i)$, of size Lx1, of the image onto the eigenspace is then calculated and used as the feature space.

$$\mathbf{g}(\mathbf{x}_i) = \mathbf{x}_i^T \mathbf{E} \tag{1}$$

2.2. Object Pose Ordering

The proposed approach for ordering images is an iterative process. Let S_j and Θ_j be the set of unordered and ordered images at iteration j, respectively. To begin, S_0 is the entire set of unordered images and Θ_0 is the empty set of ordered images. At iteration $j = 1$, a randomly chosen image is labeled $\mathbf{x}_1$ and moved from S_0 to Θ_0, yielding S_1 and Θ_1. For $j \geq 2$ an image $\mathbf{x}_j$ is moved from S_{j-1} to Θ_{j-1} such that

$$\mathbf{x}_j = \underset{\mathbf{x} \in S_{j-1}}{\operatorname{argmin}} \left(\| \mathbf{g}(\mathbf{x}_{j-1}) - \mathbf{g}(\mathbf{x}) \| \right) \tag{2}$$

Thus, the ordering algorithm picks from the unordered set, the image closest to the last ordered image in eigenspace. Once the images have been ordered using the minimum separation, a confidence measure is computed using local curvature along the trajectory (called the object manifold) of the ordered images in eigenspace.

Let $\mathbf{ds}_j$ be the vector:

$$\mathbf{ds}_j = \mathbf{g}(\mathbf{x}_j) - \mathbf{g}(\mathbf{x}_{j-1}) \tag{3}$$

The cosine of the angle between the vectors is the correlation coefficient:

978-1-4244-2354-5/09/$25.00

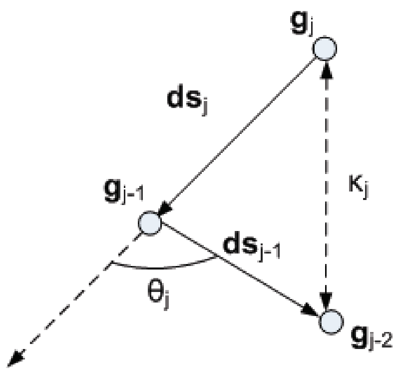

Fig. 1. Curvature explanation in a 2-d eigenspace

$$\cos(\theta_j) = \frac{\mathbf{ds}_j^T \mathbf{ds}_{j-1}}{\|\mathbf{ds}_j\| \, \|\mathbf{ds}_{j-1}\|} \tag{4}$$

The distance between vectors is then computed by subtracting the two vectors. The difference equation is an approximation to the second order derivative for curvature:

$$\kappa_j = \sqrt{(\mathbf{ds}_{j-1} - \mathbf{ds}_j)^T (\mathbf{ds}_{j-1} - \mathbf{ds}_j)} \tag{5}$$

The confidence in ordering metric is given by:

$$c_j = \kappa_j (1 - \cos(\theta_j)) \tag{6}$$

The confidence metric c_j attempts to use a combination of three local image projections to measure the alignment and the curvature. The alignment is equivalent to the congruence coefficient across three images and is equal to zero when they are in a straight line. The curvature acts as a weight across the combination of the three images. A low confidence measure indicates the images are changing slowly and pose classification is more accurate in this region and a high measure of confidence means the images are changing erratically. For this confidence measure, a low value is desirable to indicate correct ordering. The geometric description of these measures can be seen in figure 1.

2.2.1. *Two Camera Ordering*

The method is extended to the case where two synchronized cameras are available. Since the cameras are synchronized, we can compare the times indices of each camera. By independently ordering each camera's images, it can be said that if the time indices do not match, then an error in ordering has occurred. The difficulty lies in determining automatically which camera view is more likely to have incorrectly ordered the images. This is done by comparing the confidence c_j developed in the previous section and time indices of each camera to determined if an error has occurred and which camera has the lower confidence of error. If $c_j^{camera1} > c_j^{camera2}$ then camera 2 is chosen for the next view and vice versa.

2.2.2. *Two Camera Configuration*

The experimental configuration used includes a single camera and an additional second camera 90^o apart in viewing angle to improve performance. The object is placed on a motorized turntable and data is captured. The two cameras are synchronized and the configuration can be seen in figure 2. The image frames are then placed in random order using a random number generator. The desired configuration is chosen in an attempt to untangle the object manifold geometry in eigenspace.

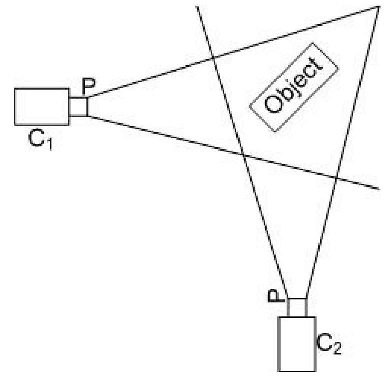

Fig. 2. Two camera configuration

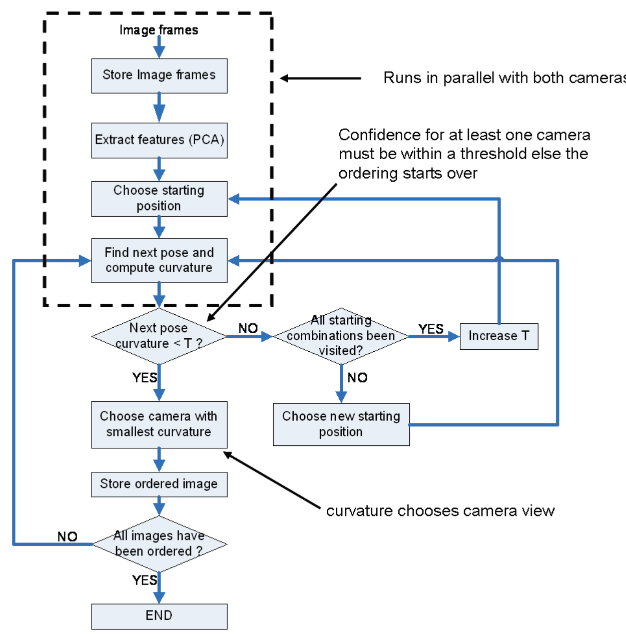

Fig. 3. Flow Chart of two camera algorithm

3. RESULTS

The results were obtained from multiple ordered sequences of object data sets. The objects were scrambled in MATLAB using the uniform random number generator. A known sequence is used to help with assessing the results of ordering. The random number set is used to produce an equal distribution for the full rotation of the object. A set of images that includes a full rotation of the object ensures that the first and last images are connected in eigenspace. It also makes error counting and validation of the ordering simpler. The video frame rate is 16 frames/sec and the number of poses in a set varies from 30-37 depending on the random number seed. The images are downsampled by a factor of 4 to a size of 80x60. This is done to accommodate the large size of the covariance matrix. The formulation of the covariance matrix for PCA is not normalized by illumination. It is thought that illumination is a feature for recognition of objects and scenes used by human vision.

The error rate for object ordering is calculated manually by visual inspection of the images and counting the number of incorrectly sequenced images. An error is defined if the next pose does not correspond to the current pose for the set of images. Error rate is calculated for the entire set by summing the total number of errors and dividing by the number of images in the set.

Objects without distinct features have failed to order properly in the single camera case because of the similarity of views. The objects without saliency in table 1 the tape dispenser and the truck. For these types of objects, the recognition errors generally misclassify the front and back views corresponding to 180^o of rotation. In eigenspace, this is where the manifold wraps around itself and becomes tangled. This can be shown in figure 5, where the circled region represents the problem area when the next angle-view is not the minimum distance. The dotted line shows the errors for ordering with a single camera. The solid line shows the precise manifold structure, corrected by two cameras with the confidence metric. The confidence measure helps to select the camera whose manifold section is the most flat and is the minimum distance across images. Figure 6 shows an example of a section of the manifold where camera 1 has made an error and camera 2 is chosen as the correct order. The solid line in figure 5 shows the correct manifold reconstruction using two cameras. Figure 4 shows an example of the confidence selecting between camera 1 and camera 2 for front and side views of the tape dispenser. The images displayed in figure 4 are seen from the perspective of camera 1. As the front view appears, camera 1 has trouble with recognizing the next pose, therefore, it switches to camera 2.

Table 1. Error Rate of Ordering Various Objects

Object	Single Camera	Two Cameras	Gain
Tape dispenser	17.0%	0.9%	16.1%
Stapler	9.1%	0.3%	8.8%
Truck	6.7%	0.3%	6.4%
House	3.0%	0.9%	2.1%
Airplane	1.5%	0.3%	1.2%
Cylinder	0.6%	0.0%	0.6%
Battery	0.3%	0.0%	0.3%

4. CONCLUSION

A method has been developed for ordering image sets of multiple angle-views using a single camera. This can be done by using PCA and minimum distance classification in eigenspace to sequence the images in the order relative to the angle viewed by the camera. The method has been extended to the case of a second camera being added at a 90^o viewing angle relative to the first camera view. A method to minimize ordering errors is then developed using a superfeature as a confidence meausure to choose between the two cameras. The superfeature is a combination of manifold curvature and correlation coefficient in the manifold eigenspace. The confidence metric is a measure of the flatness and closeness across three images in eigenspace. The results show up to a 16% recognition performance gain from the using the confidence measure with two cameras. The addition of a double camera framework attempts to untangle a complex multidimensional manifold in eigenspace. Along with object ordering, this confidence measure can also be used in a multi-camera pose recognition system where the best camera for recognition can be chosen.

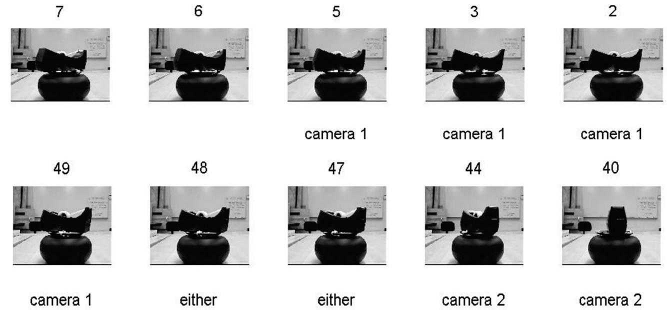

Fig. 4. Ordered images for the tape dispenser as seen from camera 1

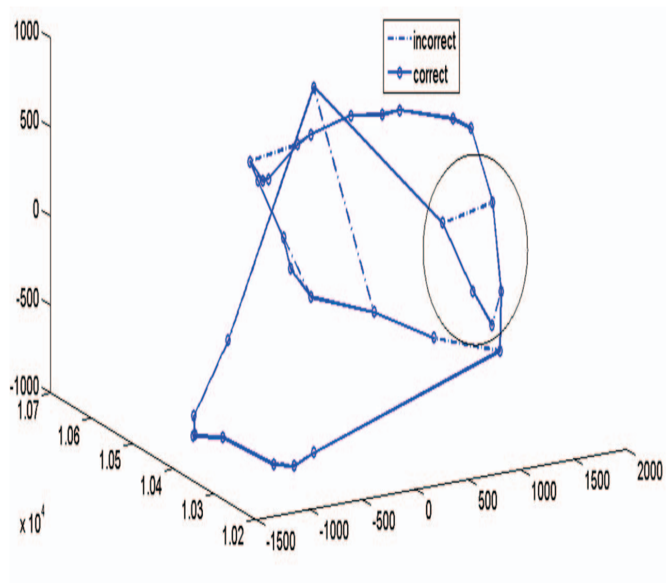

Fig. 5. Reconstruction of the object manifold using a single camera. The circled region displays errors in ordering where the manifold overlaps.

5. REFERENCES

[1] W. Seales and O. Faugeras, "Building three dimensional object models from sequences," *Computer Vision and Image Understanding*, vol. 61, no. 3, pp. 308–324, 1995.

[2] A. Bottino and A. Laurentini, "Shape-from-silhouette when the relative position of the object is unknown," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1484–1493, 2003.

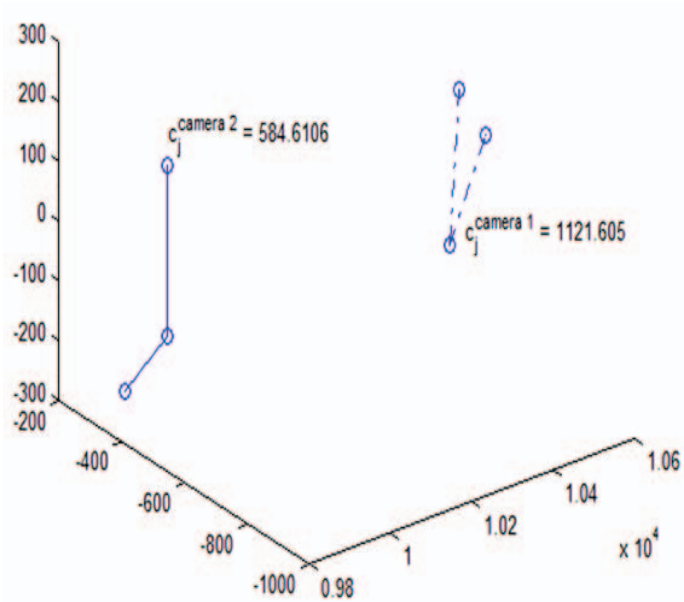

Fig. 6. Selcted region of the manifold where an eratic change has taken place (dotted line) and a correct confidence

[3] S. Baker K. Chueng and T. Kanade, "Visual hull alignment and refinement across time: A 3d reconstruction algorithm combining shape-from-silhouette with stereo," *IEEE conf. on Computer Vision and Pattern Recognition*, 2003.

[4] Marketa Potuckova, "Image matching and it's application to photogrammetry," *Ph.D. Thesis Aalborg University*, 2004.

[5] Lijin Aryananda, "Online and unsupervised face recognition for humanoid robot: Toward relationship with people," *IEEE-RAS International Conference on Humanoid Robots*, 2001.

[6] K. Grauman and Trevor Darrell, "Efficient image matching with distributions of local invariant features," *IEEE conf. on Computer Vision and Pattern Recognition*, vol. 2, no. 20, pp. 627–634, 2005.

[7] I. Ahn J. Ham and Daniel Lee, "Learning a manifold-constrained map between image sets: applications to matching and pose estimation," *IEEE conf. on Computer Vision and Pattern Recognition*, 2006.

[8] D. Beymer, "Face recognition underying pose," *MIT Artificial Intelligence Laboratory*, , no. 89, 1993.

[9] H. Murase and S.K. Nayar, "Learning and recognition of 3d objects from appearance," *Int. J. of Comp. Vis.*, vol. 14, pp. 5–24, 1995.

[10] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–709, 1997.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 2004.

[12] M. Brown and D. G. Lowe, "Unsupervised three dimensional object recognition and reconstruction in unordered data sets," *IEEE 5th Conf. on 3D Imaging and Modeling*, pp. 56–63, 2005.