Estimating the Pose of an Object in a Distributed Camera Network

A. Jorstad*, P. Burlina*, I-J. Wang*, D. Lucarelli*, D. DeMenthon*, R. Tron†, A. Terzis†, and R. Vidal†

*JHU Applied Physics Laboratory, Laurel, MD; and
†JHU Department of Computer Science, Baltimore, MD

Our work addresses pose estimation in a distributed camera framework. Computer vision research has traditionally assumed that visual information acquired by multiple cameras is aggregated and processed at a single collector node. An alternative distributed paradigm consists of having multiple cameras locally process partial information while communicating with only a few of the neighboring camera nodes to compute and disseminate a common estimate of the pose to all nodes without the need for a centralized processor (Fig. 1).

Recent work in sensor networks has examined the local aggregation of information through consensus algorithms. We examine how such processing cameras can best reach a consensus about the pose of an object when they each know a model of the object, defined by a set of point coordinates in the object frame of reference.

The cameras are assumed to be internally and externally calibrated, but they can potentially see only a subset of the object feature points in the midst of clutter points from the background, not knowing at first which image points match which object points or which points are object points or background points.

The cameras individually recover the object's pose by using their knowledge of the model, where pose is defined to be the rotation matrix and translation vector from a known origin that places and orients the object in space (Fig. 2). These pose estimates are then exchanged between neighbors, performing all updates locally to obtain a single estimate that is consistent across all cameras, without requiring a centralized processor.

The traditional consensus algorithm calculates the mean of a set of node values x_i by locally updating each node's value on the basis of how it differs from its immediate neighbors:

$$x_{\text{consensus}} = \frac{1}{J}\sum_{j=1}^{J} x_j \longleftarrow x_i^{(k+1)} = x_i^{(k)} + \epsilon \sum_{j \in N_i} \left(x_j^{(k)} - x_i^{(k)} \right).$$

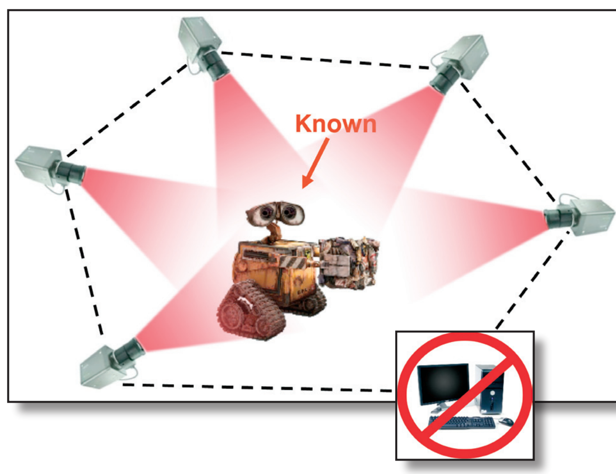

Figure 1. Object pose information is estimated and disseminated throughout the network using local computations between neighboring cameras to arrive at a common estimate of the pose. (WALL-E image is the property of Pixar.)

Figure 2. Information from the object model (left) and from the image with potential feature points (right) is combined to estimate the pose of the object.

We study four variations on the traditional consensus algorithm to exchange pose information. The first simply averages the 3-D coordinates of each object point in space. The second method adds a term to the objective function used in the first method to penalize geometric discrepancies with the known model. The remaining methods perform consensus over the rotations and translations themselves. A 3-D translation vector can be handled in the same way as 3-D point coordinates, but a rotation matrix cannot be meaningfully averaged in Euclidean space. The first of our rotation-based methods represents a rotation as an angle θ about an axis $\vec{u}$, comparing scaled axes at each iteration. The second uses the Karcher mean, performing calculations on the manifold of rotations.

Our main contribution is a novel algorithm that obtains consensus in 3-D world coordinates penalized by a 3-D model, resulting in fewer errors than with other methods. (Using the Karcher mean is also reliable and very fast.) Our algorithm is able to recover a single estimate of the pose that is statistically consistent with the average pose and to effectively disseminate this information through the network irrespective of the network topology (Fig. 3). Only six values need to be exchanged in our distributed camera network, making our method very efficient in a low-bandwidth setting. Applications of decentralized data fusion requiring small robots to collaborate include space exploration, search and rescue, and video surveillance. We plan to incorporate our algorithm into a full object-recognition system.

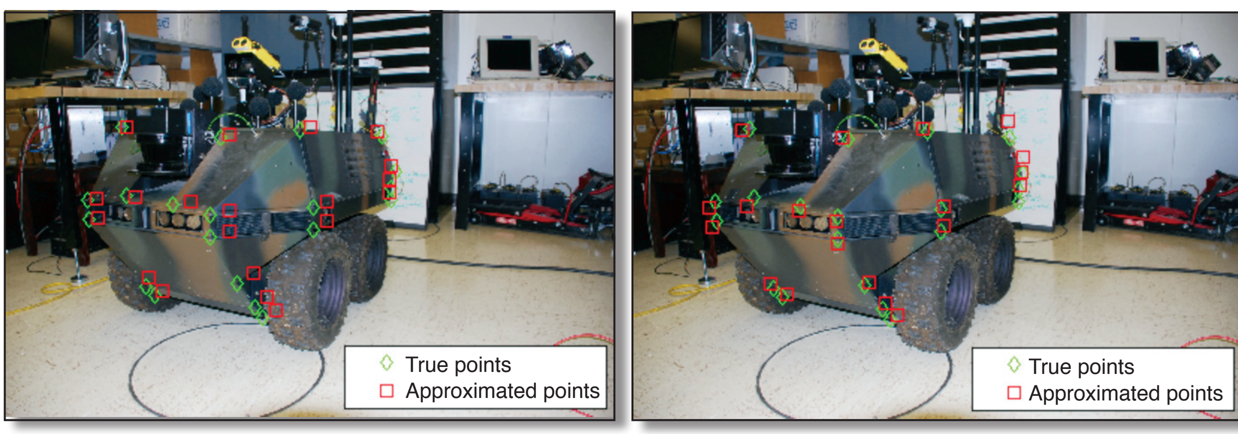

Figure 3. Real imagery. (Left) Initial estimate of pose before consensus iterations. (Right) Pose estimate after consensus iterations are performed, using penalized world coordinates.

For further information on the work reported here, see the references below or contact anne.jorstad@jhuapl.edu.

[1]Jorstad, A., Burlina, P., Wang, I-J., DeMenthon, D., and Lucarelli, D., "Model-Based Pose Estimation by Consensus," *Proc. Int. Conf. on Intelligent Sensors, Sensor Networks, and Information Processing*, pp. 569–574 (2008).

[2]Olfati-Saber, R., Fax, J. A., and Murray, R., "Consensus and Cooperation in Networked Multi-Agent Systems," *Proc. IEEE*, 95(1), 215–233 (Jan 2007).

[3]Tron, R., Vidal, R., and Terzis, A., "Distributed Pose Averaging in Camera Sensor Networks Using Consensus on Manifolds," *Proc. Second ACM/IEEE Int. Conf. on Distributed Smart Cameras* (ICDSC 08) (2008).