Exploiting Sparsity, Geometry and Statistics for Video Processing and Understanding Using a Camera Network

Rama Chellappa
Dept. of Electrical and Computer engineering and the Center for Automation Research, University of Maryland, College park, MD 20742

Abstract

Camera networks typically operate in an expansive mode (collect video data at will and without conscience!), leaving to processing and understanding algorithms to cope up with data volume, communication and power demands. With recent advances in compressive sensing (CS) and manifolds, we can start developing novel paradigms for video processing and understanding applications in a camera network. In collaboration with researchers at Rice University, we have recently developed a CS framework for background subtraction, where we directly recover background subtracted images using CS and discuss its applications in some communication constrained multi-camera computer vision problems. Specifically, we have shown how to apply the CS theory to recover object silhouettes (binary background subtracted images) when the objects of interest are sparse in the spatial domain. We cast the background subtraction as a sparse approximation problem and provide different solutions based on convex optimization and total variation. Such methods will be very useful for processing the voluminous video data collected by a camera network for tracking and object and activity recognition problems.

Features derived from video sequences are often found to lie on special manifolds such as Stiefel and Grassmanian. By exploiting the geometry of these special manifolds, one can derive more efficient particle-filter and other trackers on manifolds. For example, the camera geometry can be leveraged to derive importance sampling schemes in the implementation of particle filters. Another issue that is relevant is incorporating the correlation among the sensed data acquired as is being done by researchers in Syracuse. If sufficient samples are available, one can fit kernel-based probability density functions as we have shown for Bayesian recognition of activities.

Recently, in collaboration with Prof. Aunj Srivastava, we have developed a model-based approach to account for the effect of temporal variations (time warps) on activity recognition problems, while simultaneously allowing for the spatial variations in the descriptors. The model is composed of a *nominal activity trajectory* and a *function space* capturing the probability distribution of activity-specific time warping transformations. We use the square-root parameterization of time warps to derive geodesics, distance measures, and probability distributions on the space of time warping functions. One can then design Bayesian algorithms which treat the execution rate function as a nuisance variable and integrate it out using Monte Carlo sampling, to generate estimates of class posteriors. This approach allows us to learn the space of time warps for each activity while simultaneously capturing other intra- and interclass variations.

Simultaneous exploitation of CS theory and statistical inference on manifolds appears to be a promising approach for many video processing and understanding problems in a camera network.