heterogeneous sources of information.

The problem of combining multiple segmentations can be posed as a cluster ensemble problem. While classifier ensembles have been widely used in the Machine Learning and Data Mining communities, researchers have only recently started exploring cluster ensemble problems [22]. This is primarily because the cluster ensemble problem is inherently more difficult, since we no longer have well defined classes. From a linear algebra point of view clustering has been studied as a matrix factorization problem. Traditionally, SVD based methods have been used for this purpose. However, for image data (which is non-negative), the bases produced by these methods are not easily interpretable since they do not enforce non-negativity constraints. *Non-Negative Matrix Factorizations(NMF)* [10] produces a matrix factorization which respects non negative constraints, thereby producing directly interpretable and more representative bases. On a parallel front, recently Li *et al*. [11] have shown that consensus clustering (an algorithm for solving cluster ensembles) may be posed as a NMF problem. In this paper we propose the use of NMFs for finding the consensus segmentation. We also explore incorporating domain constraints in the consensus process to produce higher quality segmentation maps.

Cho and Meer [2] also proposed a consensus segmentation approach. The system is based on a bottom up Region Adjacency Graph (RAG) pyramid method which merges regions until a threshold on similarity is reached. The base segmentations are generated using a grayscale consistency metric, these are then used to compute a co-occurrence probability field for pixels grouped together in the segmentations. The probability field is in turn used as the metric to compute the final consensus segmentation again via the RAG pyramid. Our approach is related in that we search for an assignment of pixels to segments which best matches the mean co-occurrence $\widetilde{M}_{ij}$ (see Section 2) for each pixel or object pair. We determine the final segmentation using NMF rather than a multiscale approach, which allows us to avoid setting thresholds on region similarity.

Zhang *et al*. [26] propose combining an ensemble of Spectral Clustering results computed using randomly generated scale parameters to construct a consensus segmentation of SAR images. The authors propose several approaches for combining segmentation maps including a majority voting scheme and a hypergraph-based metaclustering algorithm. They conclude that, of the voting and hypergraph techniques, the segmentation which maximizes sum of the normalized mutual information between the base segmentations and the consensus is the best solution. However, it has been shown in [11] that the NMF approach outperforms both the naive voting scheme and the more advanced hypergraph approach.

In fields such as object labeling and image retrieval, researchers prefer segmentation approaches which exploit prior knowledge or models related to the ultimate task goals. Obviously the more context and semantic information that can be included the better the segmentation will be from the task viewpoint. Our goal is to generate useful segmentations for tasks where there is no prior knowledge which can be applied, for example systems using superpixels to simplify images [9, 17], or using coherent patches for Stereo correspondence [27]. There are however several authors who exploit multiple segment maps which are worth mentioning in this context. Russel et al. [19] use multiple base segmentations for object labeling, but rather than combine the maps, they select the best segment among all maps for an object, based on learned object class appearance. Malisiewicz and Efros [13] explore whether arbitrarily shaped segments provide better support for object recognition and demonstrate that sampling many segmentation maps allows them to find tighter segments for objects. Hoiem et al. [8] attempt to label a scene with geometric classes again by sampling and evaluating multiple segmentations to find the one with the best spatial support for the labeling task. Rabinovich et al. [16] use a set of segmentations determined to stable under slight image changes, a signature for each segment is computed and used to classify it based on training image signatures. All of these systems use multiple segmentations as a sampling of segmentation space, to obtain the best support (tightest fit) for labeling objects in the scene.

Our NMF framework provides a flexible general method for combining a set of maps as well as additional constraints such as smoothness or potentially boundary information. The base maps themselves can arise from any combination of segmentation algorithm and feature modality. A significant advantage is the ability to combine information from many sources in a uniform way. Allowing each modality to contribute a map has advantages over combining attributes in a single high dimensional feature vector. We see in Figure 1, that the curse of dimensionality causes poor segmentation performance for a simple k-means based segmentation using the stacked features.

Techniques such as ours, which use connectivity (co-occurrence) matrices between pixels, present a problem due to their size. We describe a method to scale the problem by essentially computing regions which have a preconsensus: spatially linked pixels which belong to one segment in all segmentation maps. These superpixels or *objects* allow us to compute consensus segmentations even for large images. We also present a no bells and whistles alternative which is able to perform fast consensus segmentation.

Finally, Sections 2 and 3 present details of the proposed approach and Section 4 presents the evaluation of our system on the Berkeley image segmentation database and comparison of our results to those for Mean Shift [4] and Effi-