	TRECVID MED 11				KTH Actions				UCF Sports			
	PUR	NMI	RI	FM	PUR	NMI	RI	FM	PUR	NMI	RI	FM
LMMC	**39.0**	**28.7**	**89.5**	**22.1**	**92.5**	**87.0**	**95.8**	**87.2**	**76.4**	**71.2**	**92.6**	**60.0**
MMC	36.0	26.6	89.3	20.3	91.2	86.5	95.2	85.5	63.6	62.2	89.2	46.1
SC	28.6	23.6	87.1	20.3	61.0	50.8	73.6	58.2	69.9	70.8	90.6	58.1
KM	27.0	23.8	85.9	20.4	64.8	50.7	84.0	60.6	63.1	66.2	87.9	58.7
NC	12.9	5.7	31.6	12.7	48.0	33.9	72.9	33.1	60.7	53.8	83.4	41.8

Table 1: Clustering results (in %) on the three datasets. The figures boldfaced are the best performance among all the compared methods.

average distance over all the 5-nearest neighbors. Note that these three methods do not use latent variable models. Therefore, for a fair comparison with LMMC, they are directly performed on the data where each video is represented by a vector of tag detection scores. We have also tried KM, NC and SC on the 1,000-dimensional HOG3D features. However, the performance is worse and is not reported here. Furthermore, to mitigate the effect of randomness, KM, NC and SC are run 10 times with different initial seeds and the average results are recorded in the experiments.

In order to show the benefits of incorporating latent variables, we further develop a baseline called MMC by replacing the latent variable model $f_{\mathbf{w}}(\mathbf{x})$ in Eq. 4 with a linear model $\mathbf{w}^\top \mathbf{x}$. This is equivalent to running an ordinary maximum margin clustering algorithm on the video data represented by tag detection scores. For a fair comparison, we use the same solver for learning MMC and LMMC. The trade-off parameter C in Eq. 4 is selected as the best from the range $\{10^1, 10^2, 10^3\}$. The lower bound and upper bounds of the cluster-balance constraint (i.e. L and U in Eq. 4) are set as $0.9\frac{N}{K}$ and $1.1\frac{N}{K}$ respectively to enforce balanced clusters.

Performance measures: Following the convention of maximum margin clustering [32, 33, 29, 37, 38, 16, 6], we set the number of clusters to be the ground-truth number of classes for all the compared methods. The clustering quality is evaluated by four standard measurements including purity (PUR) [32], normalized mutual information (NMI) [15], Rand index (RI) [21] and balanced F-measure (FM). They are employed to assess different aspects of a given clustering: PUR measures the accuracy of the dominating class in each cluster; NMI is from the information-theoretic perspective and calculates the mutual dependence of the predicted clustering and the ground-truth partitions; RI evaluates true positives within clusters and true negatives between clusters; and FM considers both precision and recall. The higher the four measures, the better the performance.

5.1 Results

The clustering results are listed in Table 1. It shows that LMMC consistently outperforms the MMC baseline and conventional clustering methods on all three datasets. Specifically, by incorporating latent variables, LMMC improves the MMC baseline by 3% on TRECVID MED 11, 1% on KTH Actions, and 13% on UCF Sports respectively, in terms of PUR. This demonstrates that learning the latent presence and absence of tags can exploit rich representations of videos, and boost clustering performance. Moreover, LMMC performs better than the three conventional methods, SC, KM and NC, showing the efficacy of the proposed LMMC framework for unsupervised data clustering.

Note that MMC runs on the same non-latent representation as the three conventional methods, SC, KM and NC. However, MMC outperforms them on the two largest datasets, TRECVID MED 11 and KTH Actions, and is comparable with them on UCF Sports. This provides evidence for the effectiveness of maximum margin clustering as well as the proposed alternating descent algorithm for optimizing the non-convex objective.

Visualization: We select four clusters from TRECVID MED 11, and visualize the results in Figure 2. Please refer to the caption for more details.

6 Conclusion

We have presented a latent maximum margin framework for unsupervised clustering. By representing instances with latent variables, our method features the ability to exploit the unobserved information embedded in data. We formulate our framework by large margin learning, and an alter-