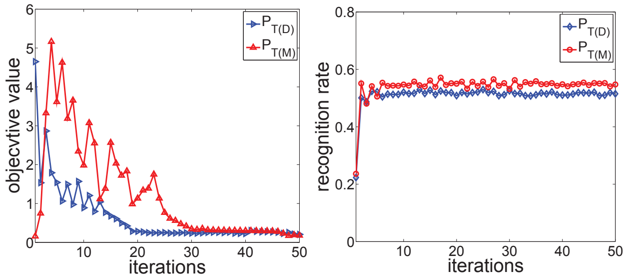

Figure 3: Results of convergence (left) and recognition rate (right) with different inner iterations in two directions individually. The dimensions of $P_{\mathrm{T(M)}}$ and $P_{\mathrm{T(D)}}$ are 50. Here we only show the results of 50 inner iterations.

In total, we have two groups of databases: BUAA&Oulu, CMU-PIE&Yale B, and each has four datasets (two modalities from two databases). For each group of databases, we can select one dataset out of four as the test data (missing modality) and other three as the training data. In both groups, we randomly select one sample per subject from the testing data as the reference data. Note, there is no overlap between the reference and testing data. We repeat this five times using the nearest-neighbor as the classifier, and the average results are reported. There are two groups of experiments: (1) evaluation on convergence and property of two directions; (2) comparisons with other transfer learning algorithms. Our methods can work in two modes: **Our-I** (without dictionary learning), by setting $\varphi = 0$; **Our-II** (with dictionary learning), by setting $\varphi \neq 0$.

Convergence and Property of Two Directions

In this experiment, we first test the convergence and recognition results of Our-I in two directions. Here we define one outer iteration as first to learn $P_{\mathrm{T(D)}}$, then to learn $P_{\mathrm{T(M)}}$. This is different from the inner iteration of the model in Eq. (10) that iteratively updates the subspace independently in one direction. In outer iterations, the proposed transfer learning updates both subspaces in two directions at each round.

In the convergence and recognition experiments, we first show results of different inner iterations in two directions using PCA as the subspace method. We conduct experiments on CMU-PIE and Yale B face databases, and take HR images of CMU-PIE as the testing data. LR of CMU-PIE and LR and HR of Yale B are used for training. The results of convergence and recognition rate in different inner iterations are shown in Figure 3 where both of them converge in two directions. It is observed that the recognition rate converges very fast in the first several iterations, so in practice, we choose small inner iterations (less than 20) for the following experiments. A small number of iterations also facilitates the outer iterations as the number of independent basis of the learned subspace in both directions is gradually decreasing due to the orthogonalization process in updating.

Next, we show how the results were affected by the number of dimensions of the projection, and compare with our model that transfers knowledge only in one direction. We use the same data setting and apply two subspace learning methods: PCA and LDA. Two directions with one outer iteration is first tested, as shown in Figure 4. In PCA, two directions with one outer iteration performs better than model in one direction. In LDA case, however, $P_{\mathrm{T(M)}}$ direction between modalities achieves better results than two directions model. This is because, in PCA case, $P_{\mathrm{T(D)}}$ helps to mitigate the divergence between two databases in terms of data distribution and transfer more modality information to the objective database, while in LDA case, the label information in the auxiliary database may not be applicable to the objective database. This becomes significant when the number of classes in two databases are different. Another reason is the dimensionality of the learned subspace. Since the dimensionality is decreasing due to the orthogonality process in each iteration, we should keep a relative larger number of dimensionality at the beginning; otherwise, the performance will degrade as well. However, as to LDA case, the dimensionality is restricted by the number of the class. This explains why the dimensions of subspace in two directions can only be 30 in Figure 4 (right).

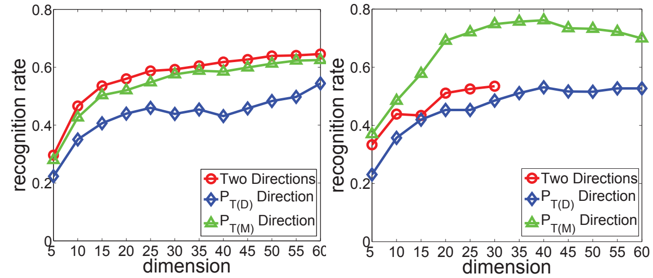

Figure 4: Results of $P_{\mathrm{T(M)}}$, $P_{\mathrm{T(D)}}$, and two directions with one outer iteration. Left is PCA case and right is LDA case.

In addition, we evaluate with more than one outer iterations. Interestingly, we find one outer iteration is adequate for better results if we tune the dimensions of $P_{\mathrm{T(M)}}$ and $P_{\mathrm{T(D)}}$ during the inner iterations appropriately, as shown in Figure 3. One reason might be that since our method is still in the line of traditional transfer learning, one outer iteration is equal to the whole process of traditional transfer learning methods. Another reason is that more outer iterations yields even lower dimensionality of the learned subspace, leading to degenerated results. Therefore, we propose to use one outer iteration in the following comparison experiments.

Comparison with Other Algorithms

In the second group of experiments, we compare Our-I and Our-II with TSL (Si, Tao, and Geng 2010), LTSL(Shao et al. 2012), RDALR(Jhuo et al. 2012), and GFK (Gong et al. 2012) in different subspace settings: PCA(Turk and Pentland 1991), LDA (Belhumeur, Hespanha, and Kriegman 1997), Unsupervised LPP (ULPP) and Supervised LPP (SLPP) (He and Niyogi 2004). Tables 1, 2 show the best results with optimal dimensions for 4 cases by changing training and testing data settings. Figure 5, 6 show the results in different dimensions for one case.

It can be seen that Our-I and Our-II perform much better than compared algorithms. Both LTSL and RDALR perform better than TSL which demonstrates that low-rank constraint is helpful in alignment of different domains. Compared to LTSL and LRDAP that consider one direction knowledge transfer, Our-I and Our-II work better. One thing is our

Table 1: Best results (%) and optimal subspace dimensions of BUAA and Oulu NIR-VIS face databases, where the test data, respectively, are NIR of BUAA (**Case 1**), VIS of BUAA (**Case 2**), NIR of Oulu (**Case 3**) and VIS of Oulu (**Case 4**).

Methods	Case 1				Case 2				Case 3				Case 4			
	PCA	LDA	ULPP	SLPP	PCA	LDA	ULPP	SLPP	PCA	LDA	ULPP	SLPP	PCA	LDA	ULPP	SLPP
TSL	35.8(70)	31.3(90)	29.2(80)	36.8(55)	37.0(60)	28.3(90)	38.2(90)	46.8(85)	39.2(60)	42.2(55)	47.3(70)	45.7(50)	31.5(90)	40.3(40)	39.2(60)	36.2(50)
LRDAP	40.2(90)	38.5(70)	42.8(85)	47.2(75)	33.7(60)	34.5(70)	39.8(85)	50.2(80)	41.3(75)	36.5(80)	42.3(90)	48.2(80)	39.7(70)	42.3(80)	47.5(75)	49.3(90)
GFK	38.3(90)	12.7(39)	40.2(45)	39.5(60)	42.3(95)	15.8(35)	39.2(90)	48.3(90)	39.5(90)	26.8(70)	28.3(80)	45.3(80)	39.2(90)	38.3(60)	42.8(90)	29.3(85)
LTSL	47.2(90)	42.3(80)	50.8(90)	53.5(70)	38.3(90)	41.3(80)	41.2(80)	56.7(90)	41.8(90)	50.7(60)	48.2(40)	54.7(80)	43.3(50)	48.2(85)	52.3(80)	58.8(90)
Our-I	52.3(80)	48.7(80)	59.7(70)	63.7(75)	49.8(80)	43.2(80)	49.3(70)	60.7(90)	48.3(80)	56.8(50)	50.8(90)	55.7(95)	46.3(90)	67.5(50)	58.2(90)	68.5(80)
Our-II	57.2(60)	51.3(65)	56.8(85)	64.5(70)	50.2(80)	42.8(70)	50.7(80)	62.7(90)	49.8(90)	55.7(60)	52.2(60)	56.5(80)	47.3(50)	66.2(85)	59.7(80)	68.8(90)

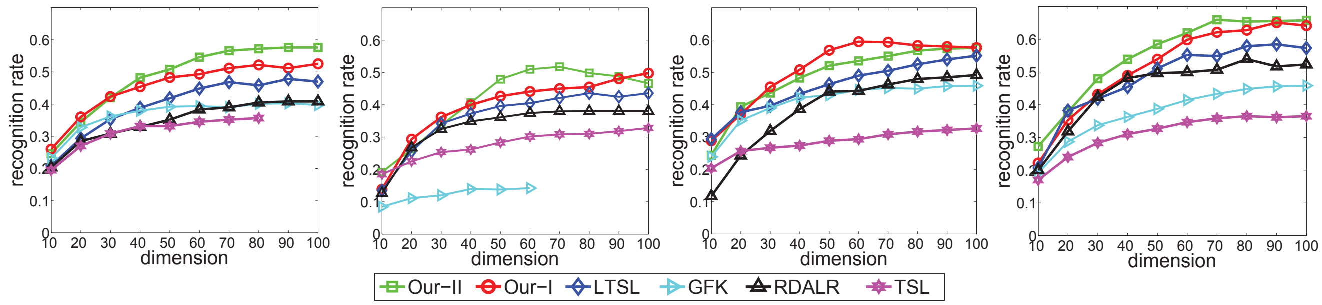

Figure 5: Results of six algorithms on Oulu *vs.* BUAA NIR-VIS face databases (**Case 1**) in four different subspaces. Subspace methods from left to right are PCA, LDA, ULPP and SLPP.

Table 2: Best results (%) and optimal subspace dimensions of CMU-PIE and Yale B face databases, where the test data, respectively, are HR of CMU-PIE (**Case 1**), LR of CMU-PIE (**Case 2**), HR of Yale B (**Case 3**) and LR of Yale B (**Case 4**).

Methods	Case 1				Case 2				Case 3				Case 4			
	PCA	LDA	ULPP	SLPP	PCA	LDA	ULPP	SLPP	PCA	LDA	ULPP	SLPP	PCA	LDA	ULPP	SLPP
TSL	22.0(60)	9.1(N/A)	22.2(60)	22.8(55)	20.3(60)	50.8(50)	27.4(60)	48.7(50)	25.4(40)	8.2(N/A)	35.3(55)	35.5(45)	20.0(55)	21.3(55)	15.2(20)	20.3(50)
LRDAP	42.1(40)	42.8(50)	44.5(60)	48.3(55)	42.8(30)	47.8(45)	50.1(55)	47.3(50)	38.3(45)	38.9(40)	38.5(60)	37.4(35)	42.3(40)	42.9(50)	45.1(60)	47.8(55)
GFK	17.3(20)	12.3(30)	40.2(35)	52.8(55)	17.3(30)	24.1(30)	23.4(40)	49.8(40)	8.3(N/A)	11.2(30)	40.7(60)	37.8(50)	8.3(30)	27.8(15)	33.3(60)	32.2(50)
LTSL	56.3(65)	60.1(50)	49.2(35)	49.7(50)	47.8(35)	54.5(40)	56.7(35)	53.2(45)	40.4(45)	43.2(35)	39.3(55)	38.4(60)	32.1(30)	35.6(50)	37.8(35)	36.7(45)
Our-I	60.8(65)	74.5(30)	59.5(55)	62.6(55)	53.2(35)	60.3(60)	58.4(45)	54.5(50)	41.3(50)	45.1(40)	42.2(50)	43.4(50)	38.4(55)	37.8(30)	41.6(55)	41.3(35)
Our-II	61.3(55)	73.1(40)	60.2(45)	63.5(55)	54.8(35)	62.5(45)	59.7(40)	54.2(45)	42.4(40)	47.2(50)	43.3(50)	45.4(55)	37.1(35)	37.6(50)	42.2(45)	43.2(50)

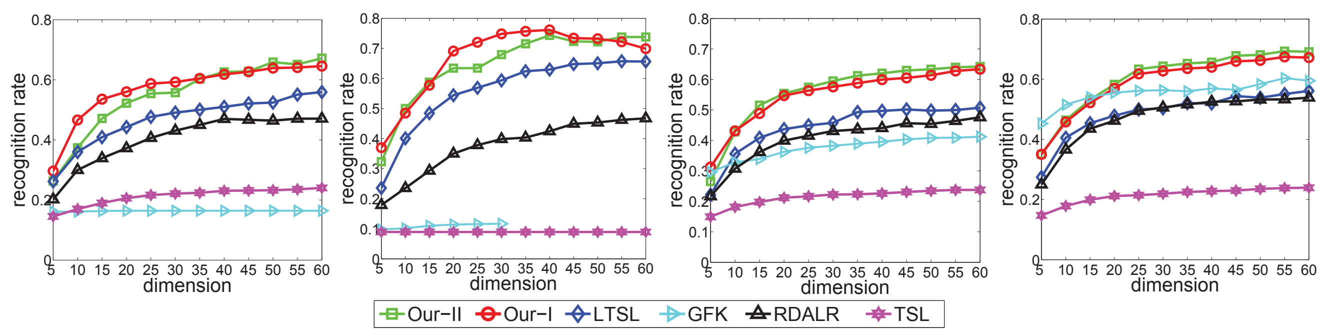

Figure 6: Results of six algorithms on CMU-PIE *vs.* Yale B face databases (**Case 1**) in four different subspaces. Subspace methods from left to right are: PCA, LDA, ULPP and SLPP.

method can compensate missing modality from one database to another, which is also helpful in knowledge transfer between modalities in the same database. In LDA case, our method only learns the subspace in one direction between modalities, but still achieves good performance. We attribute this to the latent factor from the source data which uncovers the missing information of the testing data. Furthermore, we can see that Our-II performs better than Our-I in most cases. This concludes that the dictionary constraint is also useful in finding common subspace suitable for the missing modality, as it can accurately align two domains.

Conclusion

In this paper, we proposed a novel Latent Low-rank Transfer Subspace Learning (L^{2}TSL) algorithm for the *Missing Modality Problem*. With the auxiliary database, our algorithm is capable of transferring knowledge in two directions, between modalities within one database and between two databases. By introducing a dictionary and latent low-rank constraints, our algorithm can learn appropriate subspaces to better recover the missing information of the testing modality. Experiments on two groups of multimodal databases have shown that our method can better tackle the missing modality problem in knowledge transfer, compared to several existing transfer learning methods.

References

Bartels, R. H., and Stewart, G. 1972. Solution of the matrix equation ax+ xb= c [f4]. *Communications of the ACM* 15(9):820–826.

Belhumeur, P. N.; Hespanha, J. P.; and Kriegman, D. J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7):711–720.

Cai, J.-F.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4):1956–1982.

Di, H.; Jia, S.; and Yunhong, W. 2012. The buaa-visnir face database instructions. In *IRIP-TR-12-FR-001*.

Duda, R. O.; Hart, P. E.; and Stork, D. G. 2012. *Pattern classification*. John Wiley & Sons.

Fernando, B.; Habrard, A.; Sebban, M.; Tuytelaars, T.; et al. 2013. Unsupervised visual domain adaptation using subspace alignment. In *IEEE International Conference on Computer Vision*, 2960–2967.

Gong, B.; Shi, Y.; Sha, F.; and Grauman, K. 2012. Geodesic flow kernel for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2066–2073.

Gong, B.; Grauman, K.; and Sha, F. 2013. Reshaping visual datasets for domain adaptation. In *Advances in Neural Information Processing Systems*, 1286–1294.

Guo, Y. 2013. Robust transfer principal component analysis with rank constraints. In *Advances in Neural Information Processing Systems*, 1151–1159.

He, X., and Niyogi, P. 2004. Locality preserving projections. In *Advances in Neural Information Processing Systems*, 234–241.

Huang, D.-A., and Wang, Y.-C. F. 2013. Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition. In *IEEE International Conference on Computer Vision*, 2496–2503.

Jhuo, I.-H.; Liu, D.; Lee, D.; and Chang, S.-F. 2012. Robust visual domain adaptation with low-rank reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2168–2175.

Li, S., and Fu, Y. 2013. Low-rank coding with b-matching constraint for semi-supervised classification. In *International Joint Conferences on Artificial Intelligence*, 1472–1478.

Lin, Z.; Chen, M.; and Ma, Y. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *Technique Report, UIUC*.

Liu, G., and Yan, S. 2011. Latent low-rank representation for subspace segmentation and feature extraction. In *IEEE International Conference on Computer Vision*, 1615–1622.

Liu, R.; Lin, Z.; De la Torre, F.; and Su, Z. 2012. Fixed-rank representation for unsupervised visual learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 598–605.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(1):171–184.

Liu, G.; Lin, Z.; and Yu, Y. 2010. Robust subspace segmentation by low-rank representation. In *27th International Conference on Machine Learning*, 663–670.

Long, M.; Wang, J.; Ding, G.; Shen, D.; and Yang, Q. 2012. Transfer learning with graph co-regularization. In *25th AAAI Conference on Artificial Intelligence*, 1033–1039.

Long, M.; Wang, J.; Ding, G.; Sun, J.; and Yu, P. S. 2013. Transfer feature learning with joint distribution adaptation. In *IEEE International Conference on Computer Vision*, 2200–2207.

Ni, J.; Qiu, Q.; and Chellappa, R. 2013. Subspace interpolation via dictionary learning for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 692–699.

Pan, S. J., and Yang, Q. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10):1345–1359.

Pan, Y.; Lai, H.; Liu, C.; Tang, Y.; and Yan, S. 2013. Rank aggregation via low-rank and structured-sparse decomposition. In *26th AAAI Conference on Artificial Intelligence*, 760–766.

Pan, W.; Xiang, E. W.; and Yang, Q. 2012. Transfer learning in collaborative filtering with uncertain ratings. In *25th AAAI Conference on Artificial Intelligence*, 662–668.

Shao, M.; Castillo, C.; Gu, Z.; and Fu, Y. 2012. Low-rank transfer subspace learning. In *IEEE 12th International Conference on Data Mining*, 1104–1109.

Shekhar, S.; Patel, V. M.; Nguyen, H. V.; and Chellappa, R. 2013. Generalized domain-adaptive dictionaries. In *IEEE Conference on Computer Vision and Pattern Recognition*, 361–368.

Si, S.; Tao, D.; and Geng, B. 2010. Bregman divergence -based regularization for transfer subspace learning. *IEEE Transactions on Knowledge and Data Engineering* 22(7):929–942.

Tan, B.; Xiang, E. W.; Zhong, E.; and Yang, Q. 2013. Multi-transfer: Transfer learning with multiple views and multiple sources. In *SIAM International Conference on Data Mining*.

Turk, M., and Pentland, A. 1991. Eigenfaces for recognition. *Journal of cognitive neuroscience* 3(1):71–86.

Wang, S.; Zhang, L.; Liang, Y.; and Pan, Q. 2012. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2216–2223.

Yang, J.; Yin, W.; Zhang, Y.; and Wang, Y. 2009. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM Journal on Imaging Sciences* 2(2):569–592.

Zhuang, Y.; Wang, Y.; Wu, F.; Zhang, Y.; and Lu, W. 2013. Supervised coupled dictionary learning with group structures for multi-modal retrieval. In *27th AAAI Conference on Artificial Intelligence*, 1070–1076.