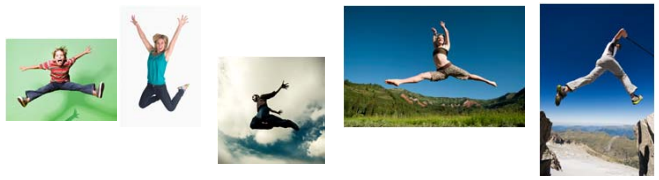
CROSS-VIEW ACTION RECOGNITION VIA LOW-RANK BASED DOMAIN ADAPTATION

Wen-Sheng Tseng[1,2], Lun-Kai Hsu[1,2], Li-Wei Kang[3], Yu-Chiang Frank Wang[2]
[1]Dept. Electrical Engineering, National Taiwan University, Taipei, Taiwan
[2]Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan
[3]Dept. Computer Science & Info. Eng., National Yunlin Univ. of Science & Technology, Yunlin, Taiwan

Introduction

- Cross-view action recognition
- Source view: a sufficient amount of labeled data
- Target view: few labeled data
- Cannot train classifiers at either domain directly



Related Work

Geometry-based approaches

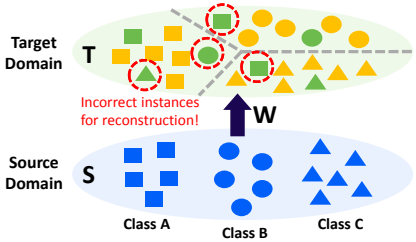
- Rao et al., *IJCV*, 2002.
- Paramesmaran and Chellappa, *IJCV*, 2006.
- Weinland et al., *ECCV*, 2010.

Domain adaption based approaches

- Liu et al., *IEEE CVPR*, 2011
- Huang et al., *ECCV WS*, 2012

Motivation

$$\min_{W,Z,E} \|Z\|_* + \alpha\|E\|_{2,1} \quad \text{s.t.} \quad WS = TZ + E,\ WW^T = I$$



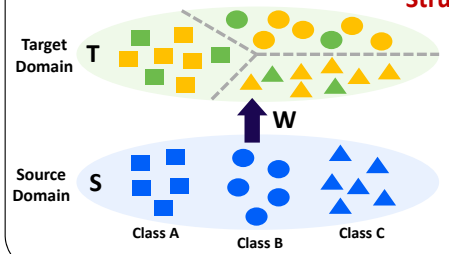
W: transformation matrix
S: source domain images
T: target domain images
Z: reconstruction coefficient matrix
E: error matrix

Jhuo et al., "Robust visual domain adaptation with low-rank reconstruction," *CVPR*, 2012.

Our Proposed Method

$$\min_{Z_i,E_i,W} \sum_{i=1}^{N}\{\|Z_i\|_* + \alpha\|E_i\|_{2,1}\} + \eta\sum_{j\neq i}^{N}\|Z_j^T Z_i\|_F^2 \quad \text{s.t.} \quad WS = TZ + E,\ WW^T = I$$

Structure incoherence (SI)



W: transformation matrix
S: source domain images
T: target domain images
Z: reconstruction coefficient matrix
E: error matrix

Optimization

• Introducing F

$$\min_{F_i,Z_i,E_i,W} \sum_{i=1}^{N}\{\|F_i\|_* + \alpha\|E_i\|_{2,1}\} + \eta\sum_{j\neq i}^{N}\|F_j^T F_i\|_F^2 \quad \text{s.t.} \quad WS_i = TZ_i + E_i,\ Z_i = F_i$$

• Solving Class-Wise Minimization Problems

$$\min_{F_i,Z_i,E_i,W} \|F_i\|_* + \alpha\|E_i\|_{2,1} + \eta\sum_{j\neq i}^{N}\|F_j^T F_i\|_F^2 \quad \text{s.t.} \quad WS_i = TZ_i + E_i,\ Z_i = F_i$$

• Solving The Relaxed Version

$$\min_{F_i,Z_i,E_i,W} \|F_i\|_* + \alpha\|E_i\|_{2,1} + \eta'\|F_i\|_F^2 \quad \text{s.t.} \quad WS_i = TZ_i + E_i,\ Z_i = F_i$$

where $\eta' = \eta\sum_{j\neq i}^{N}\|F_j\|_F^2$

Updating Zi, W, Ei, F_i

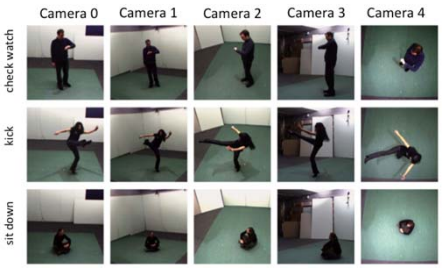
$$F_i^{k+1} = \arg\min_{F_i} \varepsilon\|F_i\|_* + \frac{1}{2}\|X_F - F_i\|_F^2 \quad Z_i^{k+1} = \left(I + T^T T\right)^{-1}\left[T^T\left(W^{k+1}S_i - E_i^k\right) + \frac{1}{\mu^k}\left(T^T Y_i^k - X_i^k\right) + F_i^{k+1}\right]$$

$$W^{k+1} = \left[\left(TZ_i^k + E_i^k\right)S_i^T - \frac{1}{\mu^k}Y_i^k S_i^T\right]\left(S_i S_i^T\right)^{-1} \quad E_i^{k+1} = \arg\min_{E_i} \varepsilon'\|E_i\|_{2,1} + \frac{1}{2}\|E_i - X_e\|_F^2$$

Experiments

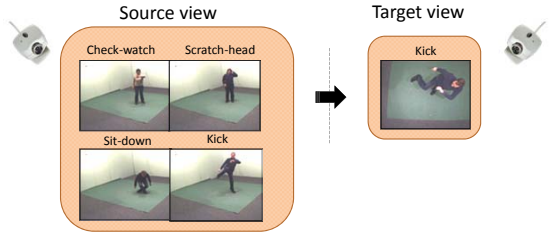
Dataset and Settings

IXMAS dataset
- 11 classes
- 5 camera views
- 36 sequences per action
- Consider BOW features



Training: 36 source-view and 2 target-view instances

Testing: 34 remaining target-view instance



Results

A: SVM (target only), **B:** SVM (source only), **C:** SVM (combine source + target labeled data), **D:** CCA, **E:** BoBW (CVPR'11), **F:** RDALR (CVPR'12), and **G:** Ours

	camera 0							camera 1						
	A	B	C	D	E	F	G	A	B	C	D	E	F	G
cam 0				-				0.48	9.63	16.58	15.90	12.63	**37.17**	**37.17**
cam 1	0.05	11.02	21.82	17.17	22.22	16.04	**25.60**				-			
cam 2	2.41	10.27	18.29	15.40	2.78	6.68	**25.67**	1.18	9.73	12.09	12.88	5.81	24.60	**27.27**
cam 3	0.32	10.16	17.43	10.35	8.84	10.16	**27.54**	0.53	9.73	14.39	13.39	5.05	17.65	**27.27**
cam 4	2.03	11.71	17.70	11.87	18.94	26.47	**26.73**	1.44	8.50	12.46	12.88	4.29	**30.48**	**30.48**
Avg.	1.20	10.79	18.81	13.70	13.19	14.84	**26.13**	0.91	9.40	13.88	13.76	6.94	27.47	**30.55**

	camera 2							camera 3						
	A	B	C	D	E	F	G	A	B	C	D	E	F	G
cam 0	0.64	12.14	33.52	22.22	26.26	**37.97**	30.75	2.30	13.48	**35.13**	6.57	4.29	21.93	33.96
cam 1	0.64	13.26	**34.28**	19.70	4.29	32.36	28.88	0.96	13.53	**33.53**	18.94	16.16	11.23	31.55
cam 2				-				1.55	6.42	**30.59**	16.92	19.19	5.88	26.20
cam 3	0.27	7.27	22.14	14.39	28.28	9.36	**29.95**				-			
cam 4	0.91	10.43	21.39	12.63	17.68	23.53	**26.47**	0.91	9.52	25.83	15.40	15.40	21.93	**25.94**
Avg.	0.61	10.78	27.78	17.23	19.13	25.80	**29.01**	1.43	10.74	**31.27**	14.46	13.76	15.24	29.41

Conclusions

- A low-rank based domain adaptation model for solving cross-view action recognition problems
- Collect more training data at the target domain by transforming labeled data from the source domain
- Our introduced SI adds additional discriminating ability to the resulting training data at the target domain.