4th International Workshop on Video Event Categorization, Tagging and Retrieval (VECTaR), in conjunction with ECCV 2012

Recognizing Actions Across Cameras by Exploring the Correlation Subspace

Chun-Hao Huang, Yi-Ren Yeh, and Yu-Chiang Frank Wang
Research Center for IT Innovation, Academia Sinica, Taiwan

Oct 12th, 2012

Outline

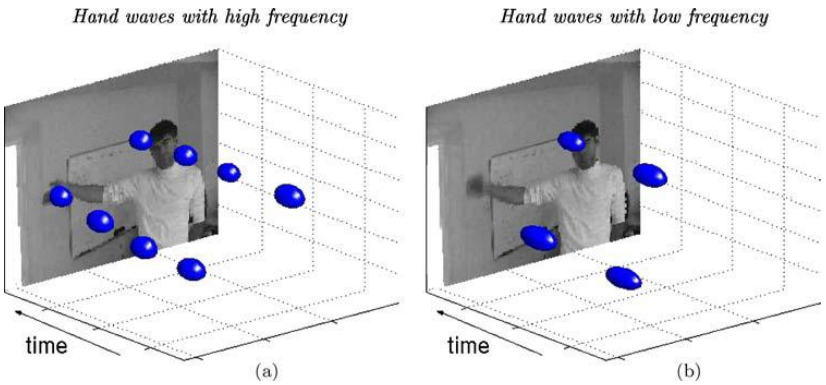
- Introduction
- Our Proposed Framework
  - Learning Correlation Subspaces via CCA
  - Domain Transfer Ability of CCA
  - SVM with A Novel Correlation Regularizer
- Experiments
- Conclusion

Outline

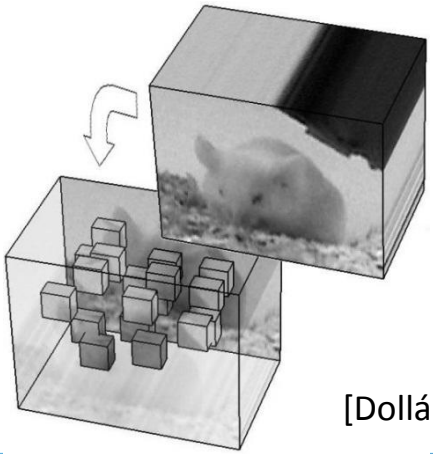
- Introduction
- Our Proposed Framework
  - Learning Correlation Subspaces via CCA
  - Domain Transfer Ability of CCA
  - SVM with A Novel Correlation Regularizer
- Experiments
- Conclusion

Representing an Action

- Spatio-temporal interest points



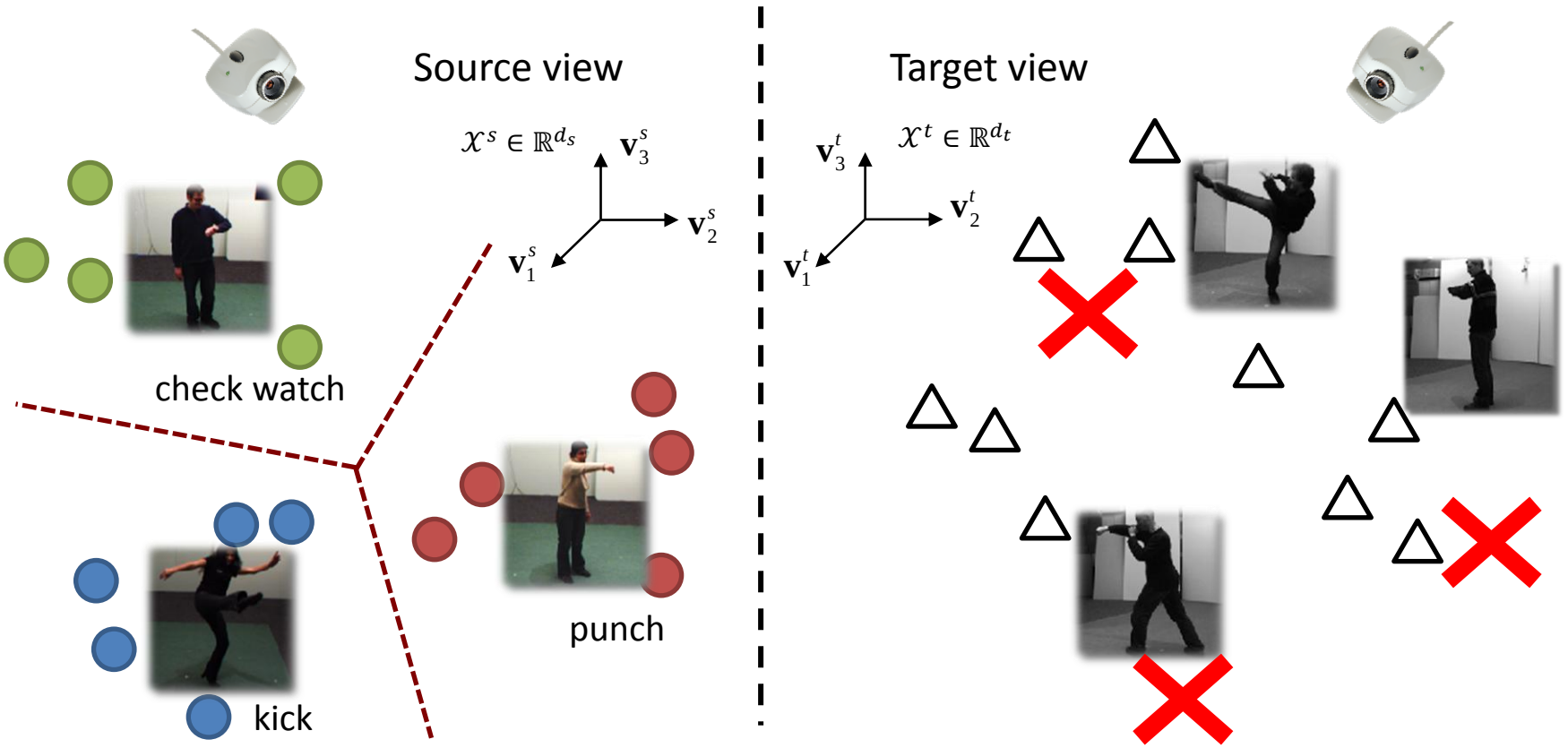
[Laptev, IJCV, 2005]



- Actions are represented as high-dim vectors.
- Bag of spatio-temporal visual word model.
- State-of-the-art classifiers (e.g., SVM) are applied to address the recognition task.

[Dollár *et al.*, ICCV WS on VS-PETS, 2005]

Cross-Camera Action Recognition

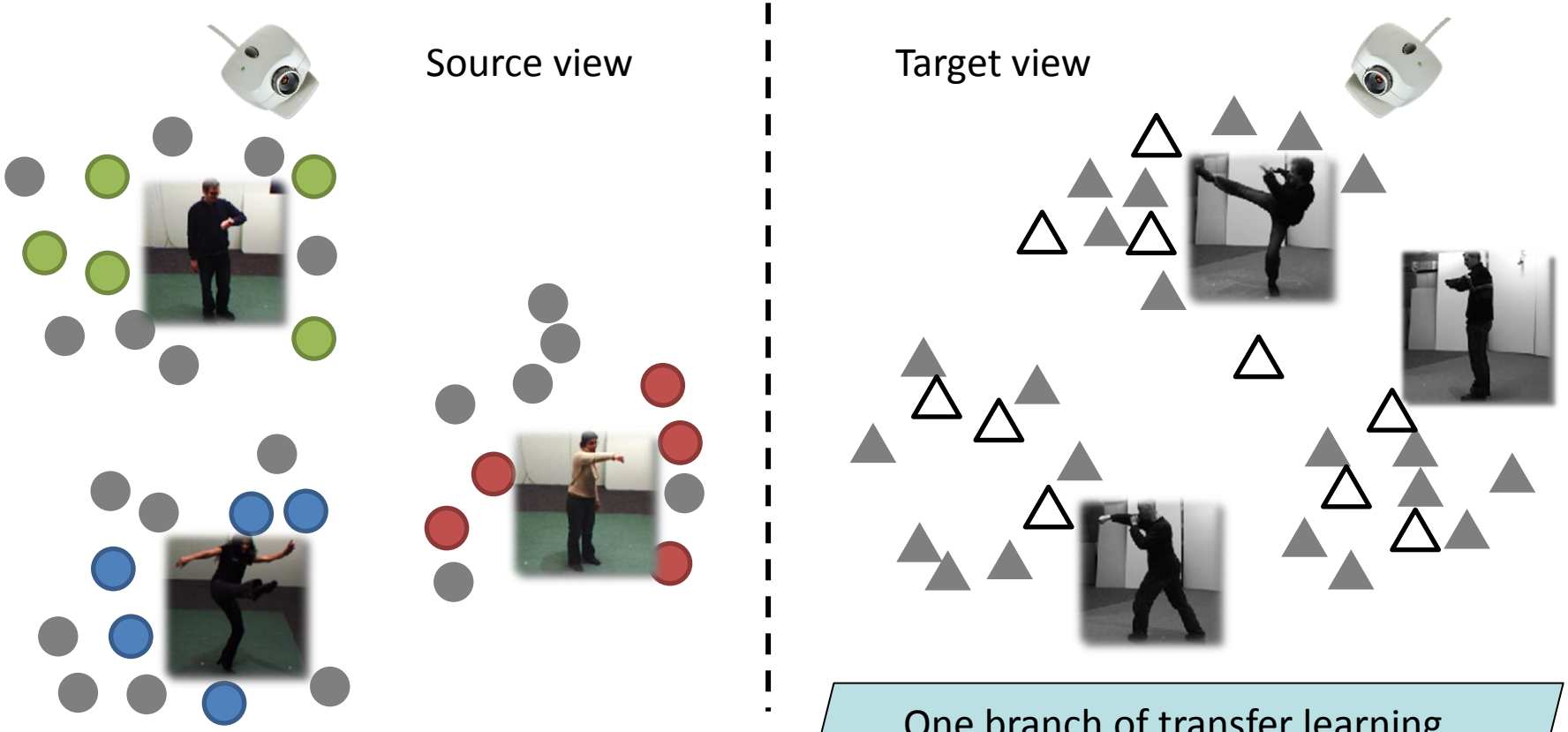


Colored: labeled data

Hollowed: test data

- Models learned at source views typically do not generalize well at target views.

Cross-Camera Action Recognition (cont'd)



Colored: labeled data
Hollowed: test data
Gray: unlabeled data

One branch of transfer learning

- **An unsupervised strategy**:
  - ✓ Only unlabeled data available at target views.
  - ✓ They are exploited to learn the relationship between data at source and target views.

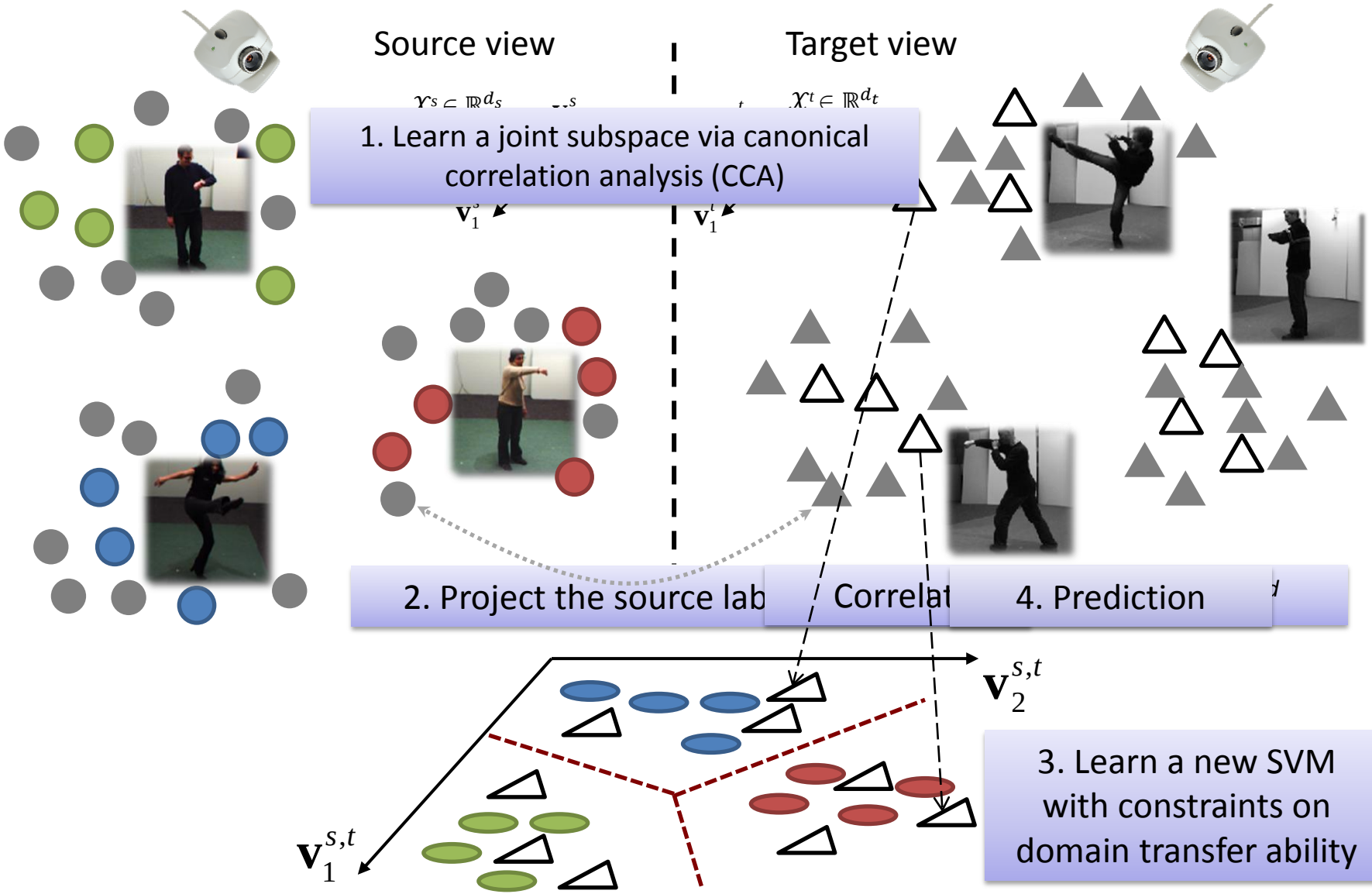
Approaches based on Transfer Learning

- To learn a common feature representation (e.g., a joint subspace) for both source and target view data.
- Training/testing can be performed in terms of such representations.
- How to exploit unlabeled data from *both* views for determining this joint subspace is the key issue.
- Previous approaches:
  1. Splits-based feature transfer [Farhadi and Tabrizi, ECCV '08]
     - Requires frame-wise correspondence
  2. Bag of bilingual words model (BoBW) [Liu *et al*., CVPR '11]
     - Considers each dimension of the derived representation to be equally important.

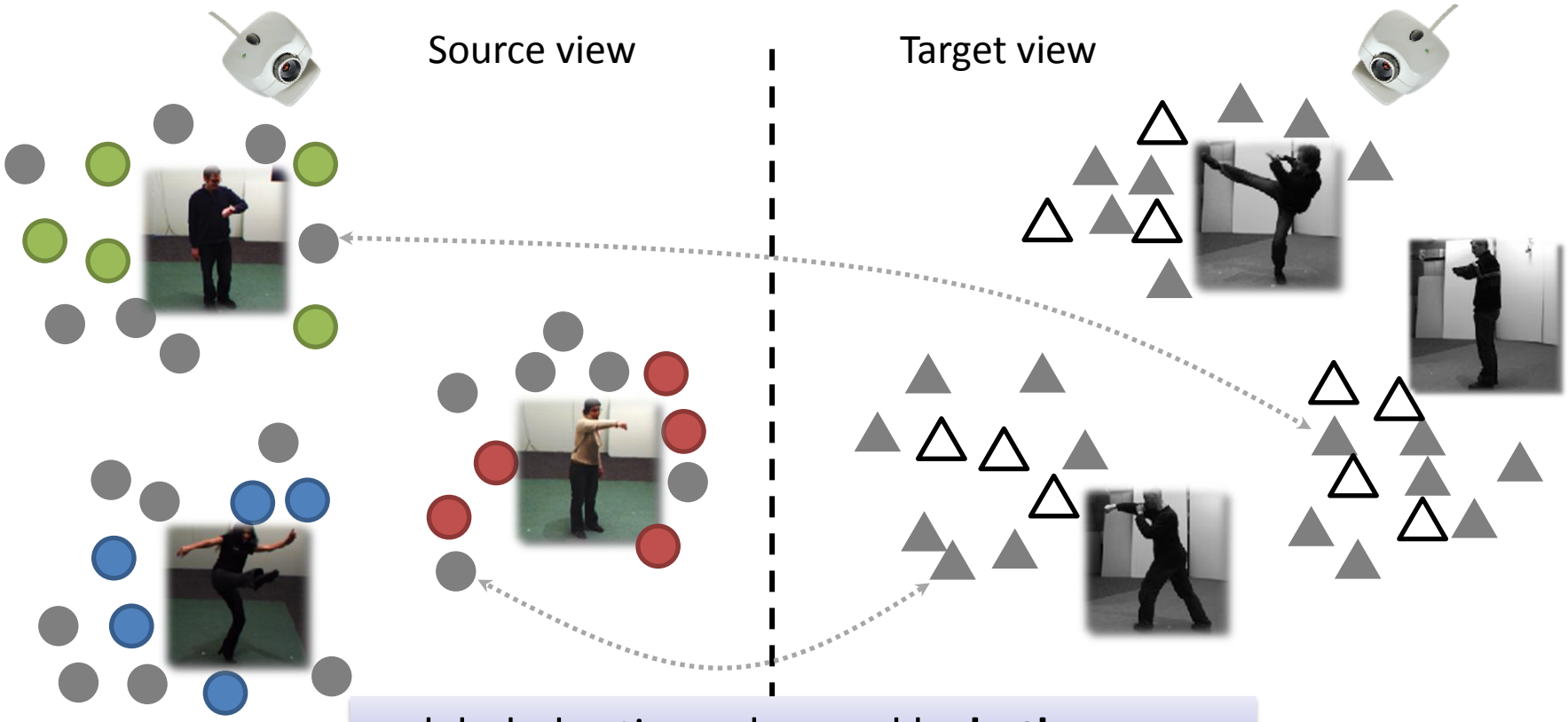
Outline

- Introduction
- Our Proposed Framework
  - Learning Correlation Subspaces via CCA
  - Domain Transfer Ability of CCA
  - SVM with A Novel Correlation Regularizer
- Experiments
- Conclusion

Overview of Our Proposed Method



Requirements of CCA



unlabeled actions observed by **both** cameras

Colored: labeled data
Hollowed: test data
Gray: unlabeled data

: unlabeled data pairs (observed at both views)

Learning the Correlation Subspace via CCA

- CCA aims at maximizing the correlation between two variable sets.
- Given two sets of *n* centered unlabeled observations :

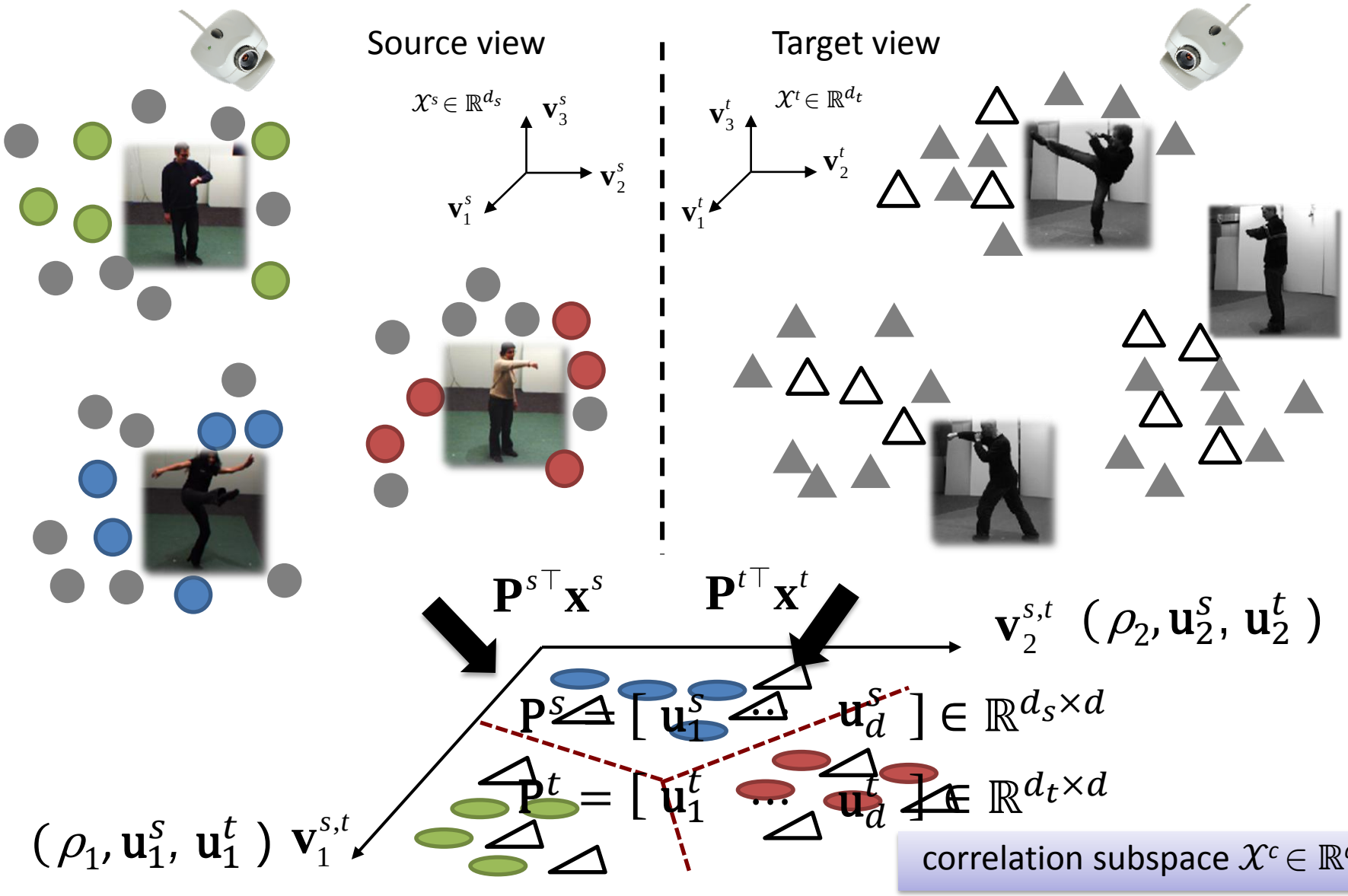
$$\mathbf{X}^s = \left[\mathbf{x}_1^s, \ldots, \mathbf{x}_n^s\right] \in \mathbb{R}^{d_s \times n} \quad \text{and} \quad \mathbf{X}^t = \left[\mathbf{x}_1^t, \ldots, \mathbf{x}_n^t\right] \in \mathbb{R}^{d_t \times n}$$

- CCA learns two projection vectors $\mathbf{u}^s$ and $\mathbf{u}^t$, maximizing the correlation coefficient ρ between projected data, i.e.,

$$\max_{\mathbf{u}^s, \mathbf{u}^t} \rho = \frac{\mathbf{u}^{s\top}\mathbf{X}^s\mathbf{X}^{t\top}\mathbf{u}^t}{\sqrt{\mathbf{u}^{s\top}\mathbf{X}^s\mathbf{X}^{s\top}\mathbf{u}^s}\sqrt{\mathbf{u}^{t\top}\mathbf{X}^t\mathbf{X}^{t\top}\mathbf{u}^t}} = \frac{\mathbf{u}^{s\top}\boldsymbol{\Sigma}_{st}\mathbf{u}^t}{\sqrt{\mathbf{u}^{s\top}\boldsymbol{\Sigma}_{ss}\mathbf{u}^s}\sqrt{\mathbf{u}^{t\top}\boldsymbol{\Sigma}_{tt}\mathbf{u}^t}}$$

where $\boldsymbol{\Sigma}_{tt} = \mathbf{X}^t\mathbf{X}^{t\top}$, $\boldsymbol{\Sigma}_{st} = \mathbf{X}^s\mathbf{X}^{t\top}$, $\boldsymbol{\Sigma}_{ss} = \mathbf{X}^s\mathbf{X}^{s\top}$ are covariance matrices.

CCA Subspace as Common Feature Representation



Outline

- Introduction
- The Proposed Framework
  - Learning Correlation Subspaces via CCA
  - Domain Transfer Ability of CCA
  - SVM with A Novel Correlation Regularizer
- Experiments
- Conclusion

Domain Transfer Ability of CCA

- Learn SVMs in the derived CCA subspace…Problem solved?

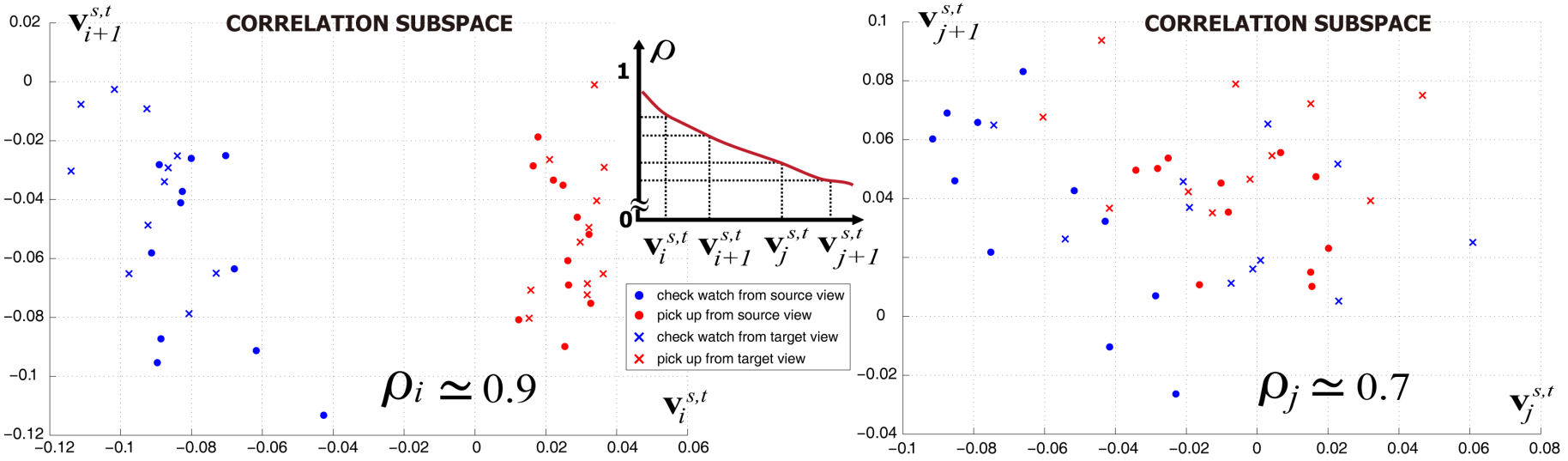
  - Yes and No!

- Domain Transfer Ability:

  - In CCA subspace, each dimension $\mathbf{V}_i^{s,t}$ is associated with a different ρ_i

  - How well can the classifiers learned (in this subspace) from the projected *source view data* generalize to those from the *target view*?

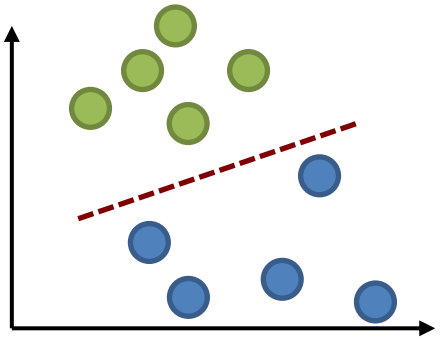
- See the example below…



Outline

- Introduction
- The Proposed Framework
  - Learning Correlation Subspaces via CCA
  - Domain Transfer Ability of CCA
  - SVM with a Novel Correlation Regularizer
- Experiments
- Conclusion

Our Proposed SVM with Domain Transfer Ability



- Proposed SVM formulation:

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{N}\xi_i - \frac{1}{2}\mathbf{r}^\top \mathrm{Abs}(\mathbf{w})$$

$$\text{s.t.} \quad y_i\left(\left\langle \mathbf{w}, \mathbf{P}^{s\top}\mathbf{x}_i^s \right\rangle + b\right) + \xi_i \geq 1, \qquad \xi_i \geq 0, \qquad \forall\left(\mathbf{x}_i^s, y_i\right) \in D_l^s$$

- The introduced correlation regularizer $\mathbf{r}^\top \mathrm{Abs}(\mathbf{w})$:

$$\mathrm{Abs}(\mathbf{w}) \equiv \left[|w_1|,\ |w_2|,\ \ldots,|w_d|\right] \quad \text{and} \quad \mathbf{r} \equiv \left[\rho_1,\ \rho_2,\ \ldots,\rho_d\right]$$

- Larger/Smaller ρ_i

  → Stronger/smaller correlation between source & target view data

  → SVM model w_i is more/less reliable at that dimension in the CCA space.

- Our regularizer favors SVM solution to be dominant in reliable CCA dimensions (i.e., larger correlation coefficents ρ_i imply larger $|w_i|$ values).
- Classification of (projected) target view test data:

$$f(\mathbf{x}) = \mathrm{sgn}\left(\left\langle \mathbf{w}, \mathbf{P}^{t\top}\mathbf{x}^t \right\rangle + b\right)$$

An Approximation for the Proposed SVM

- It is not straightforward to solve the previous formulation with $\mathrm{Abs}(\mathbf{w})$.
- An approximated solution can be derived by relaxing $\mathrm{Abs}(\mathbf{w})$:

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{N}\xi_i - \frac{1}{2}(\mathbf{r}\odot\mathbf{r})^{\top}(\mathbf{w}\odot\mathbf{w})$$

$$\text{s.t.} \quad y_i\left(\left\langle \mathbf{w}, \mathbf{P}^{s\top}\mathbf{x}_i^s \right\rangle + b\right) + \xi_i \geq 1, \qquad \xi_i \geq 0, \qquad \forall\left(\mathbf{x}_i^s, y_i\right) \in D_l^s$$

where $\odot$ indicates the element-wise multiplication.

- We can further simplify the approximated problem as:

$$\min_{\mathbf{w}} \frac{1}{2}\sum_{i=1}^{d}\left(1-\rho_i^2\right)w_i^2 + C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \quad y_i\left(\left\langle \mathbf{w}, \mathbf{P}^{s\top}\mathbf{x}_i^s \right\rangle + b\right) + \xi_i \geq 1, \qquad \xi_i \geq 0, \qquad \forall\left(\mathbf{x}_i^s, y_i\right) \in D_l^s$$

- We apply SSVM* to solve the above optimization problem.

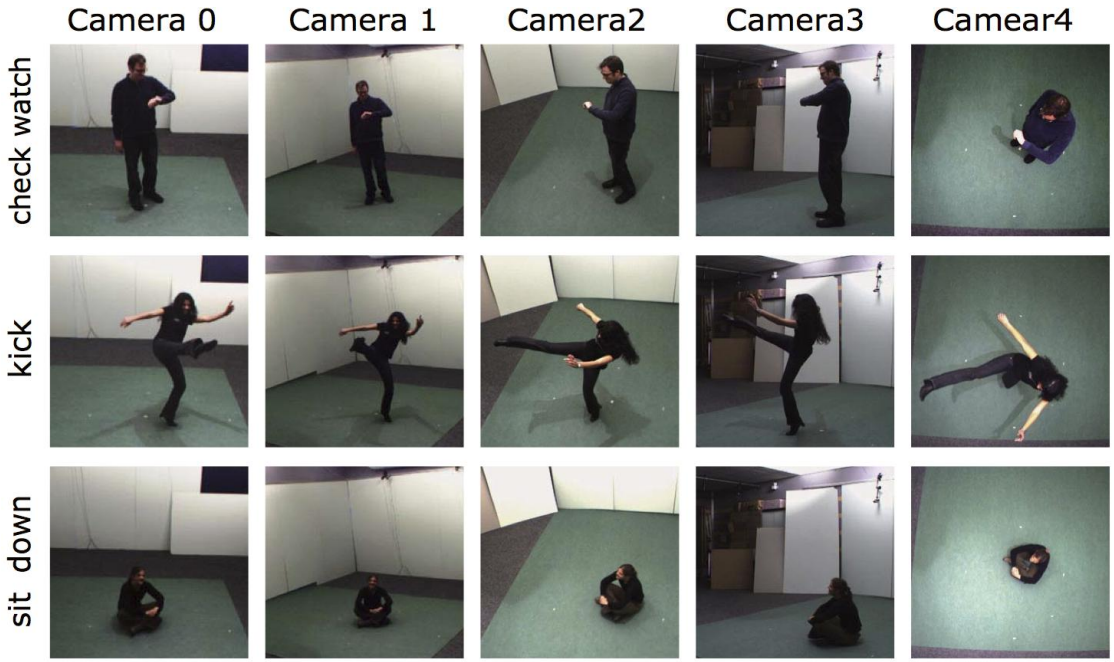
*: Lee *et al*., Computational Optimization and Applications, 2001

Outline

- Introduction
- The Proposed Framework
  - Learning Correlation Subspaces via CCA
  - Domain Transfer Ability of CCA
  - SVM with a Novel Correlation Regularizer
- Experiments
- Conclusion

Dataset

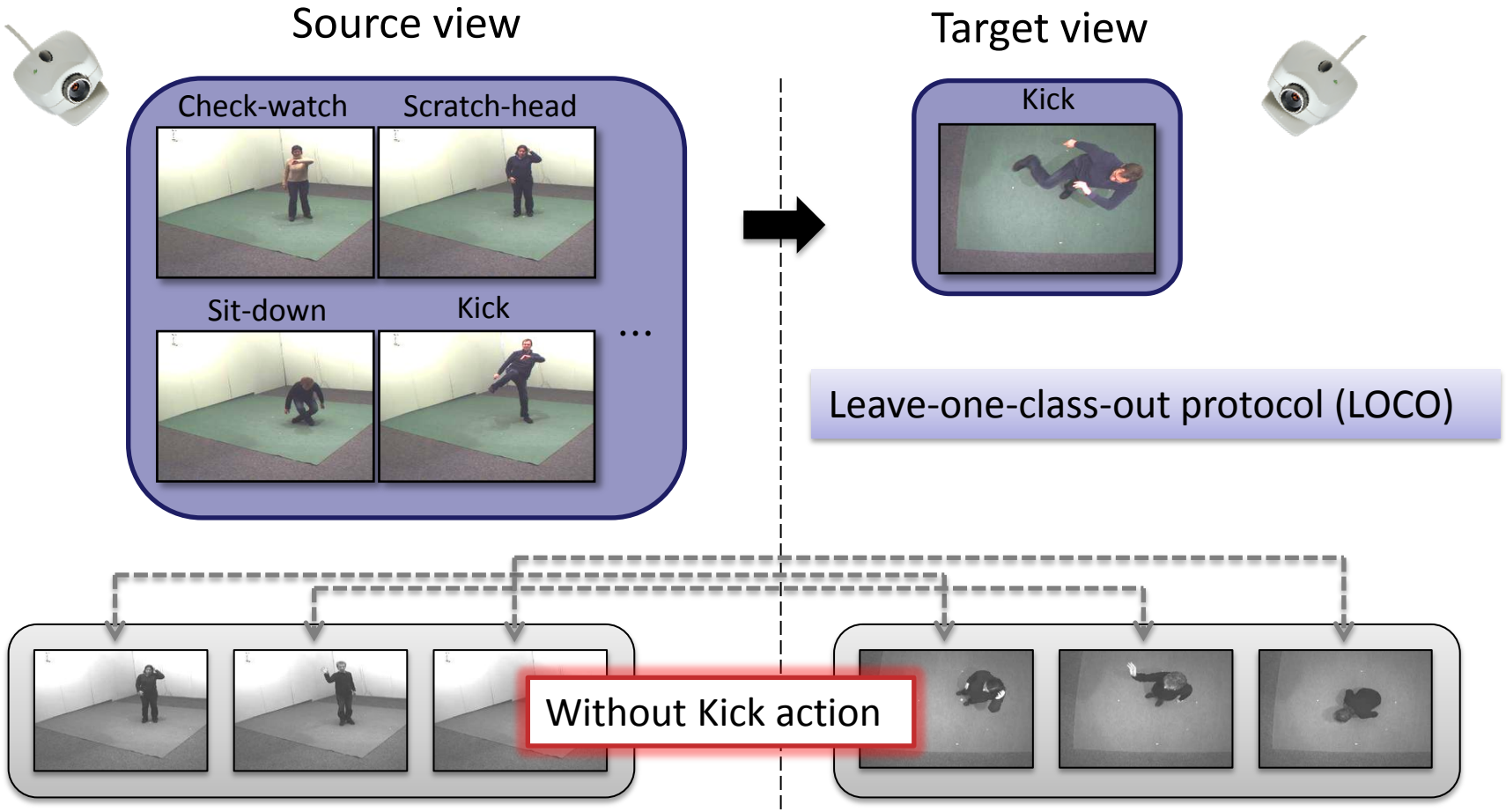
- IXMAS multi-view action dataset
  - Action videos of eleven action classes
  - Each action video is performed three times by twelve actors
  - The actions are captured simultaneously by five cameras



Experiment Setting

1/3 as labeled data: Training and testing

2/3 as unlabeled data: Learning correlation subspaces via CCA



Experimental Results

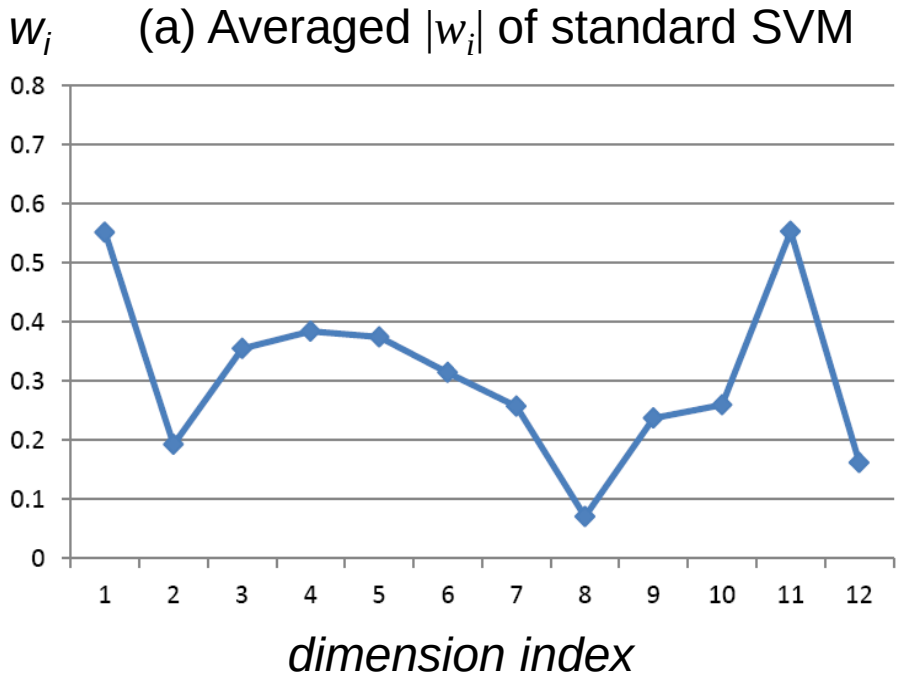
- A: BoW from source view directly
- B: BoBW + SVM [Liu *et al*. CVPR'11]
- C: BoBW + our SVM
- D: CCA + SVM
- E: our proposed framework (CCA + our SVM).

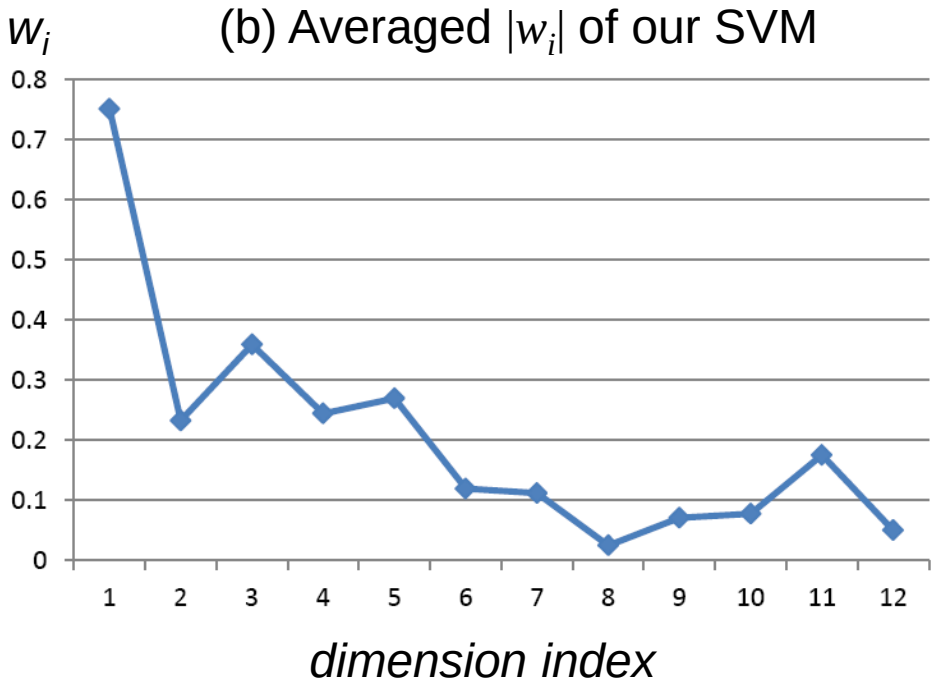
(%)	camera0					camera1					camera2				
	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
c0			-			9.29	60.96	63.03	63.18	**64.90**	11.62	41.21	50.76	56.97	**60.61**
c1	10.71	58.08	59.70	66.72	**70.25**			-			7.12	33.54	38.03	57.83	**59.34**
c2	8.79	52.63	49.34	57.37	**62.47**	6.67	50.86	45.79	59.19	**61.87**			-		
c3	6.31	40.35	44.44	65.30	**66.01**	9.75	33.59	33.27	46.77	**52.68**	5.96	41.26	43.99	**61.36**	**61.36**
c4	5.35	38.59	40.91	54.39	**55.76**	9.44	37.53	37.00	53.59	**55.00**	9.19	34.80	38.28	57.88	**60.15**
avg.	7.79	47.41	48.60	60.95	**63.62**	8.79	45.73	44.77	55.68	**58.61**	8.47	37.70	42.77	58.51	**60.37**

	camera3					camera4				
	A	B	C	D	E	A	B	C	D	E
c0	7.78	39.65	41.36	**63.64**	62.17	7.12	24.60	37.02	43.69	**48.23**
c1	12.02	35.91	39.14	48.59	**54.85**	8.89	26.87	22.22	44.24	**49.29**
c2	6.46	41.46	42.78	60.00	**61.46**	10.35	28.03	33.43	45.05	**51.82**
c3			-			8.89	27.53	28.28	40.66	**41.06**
c4	9.60	27.68	34.60	48.03	**48.89**			-		
avg.	8.96	36.17	39.47	55.06	**56.84**	8.81	26.76	30.24	43.41	**47.60**

Effects on The Correlation Coefficient ρ

- We successfully suppress the SVM model $|w_i|$ when lower ρ is resulted.
- Ex: source: *camera 3*, target: *camera 2*, left-out action: *get-up*





- Recognition rates for the two models were 47.22% and 77.78%, respectively.

Outline

- Introduction
- The Proposed Framework
  - Learning Correlation Subspaces via CCA
  - Domain Transfer Ability of CCA
  - SVM with A Novel Correlation Regularizer
- Experiments
- Conclusion

Conclusions

- We presented a transfer-learning based approach to cross-camera action recognition.
- We considered the domain transfer ability of CCA, and proposed a novel SVM formulation with a correlation regularizer.
- Experimental results on the IXMAS dataset confirmed performance improvements using our proposed method.

Thank You!