Recognizing Actions Across Cameras by Exploring the Correlated Subspace

Chun-Hao Huang, Yi-Ren Yeh, and Yu-Chiang Frank Wang

Research Center for IT Innovation, Academia Sinica, Taipei, Taiwan
{paulchhuang,yryeh,ycwang}@citi.sinica.edu.tw

Abstract. We present a novel transfer learning approach to cross-camera action recognition. Inspired by canonical correlation analysis (CCA), we first extract the spatio-temporal visual words from videos captured at different views, and derive a correlation subspace as a joint representation for different bag-of-words models at different views. Different from prior CCA-based approaches which simply train standard classifiers such as SVM in the resulting subspace, we explore the *domain transfer ability* of CCA in the correlation subspace, in which each dimension has a different capability in correlating source and target data. In our work, we propose a novel SVM with a correlation regularizer which incorporates such ability into the design of the SVM. Experiments on the IXMAS dataset verify the effectiveness of our method, which is shown to outperform state-of-the-art transfer learning approaches without taking such domain transfer ability into consideration.

1 Introduction

Action recognition has been an active research topic for researchers in the areas of computer vision and image processing. However, in practical scenarios, one typically needs to deal with multiple cameras with different lighting, depression angle, etc. conditions. Moreover, actions of interest might not be seen by a particular camera in advance, and thus no training data for that action is available. Therefore, it is expected that most existing single-view action recognition approaches cannot be easily extended for cross-view action recognition due to poor generalization [1].

While some researchers proposed to extract view-invariant representations for cross camera action recognition (e.g., [2, 3]), *transfer learning* [4] has recently been applied to address this problem [5, 6]. The purpose of transfer learning is to transfer the knowledge observed from one or few source domains to the target domain, so that the task in the target domain (e.g., predicting the action of interest captured by a new camera) can be solved accordingly.

Based on canonical correlation analysis (CCA) [7], we present a transfer learning based approach (via CCA) for cross camera action recognition. Our method aims at determining a correlation subspace as a shared representation of action models captured by different cameras. However, the correlation between the projected source and target view data will be different in each dimension of

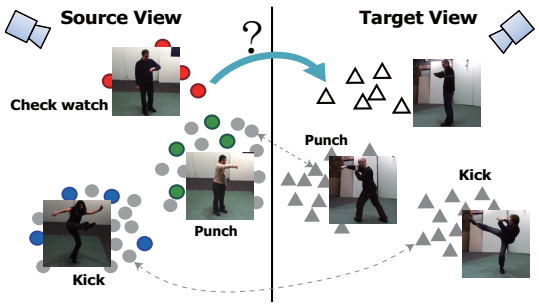

Fig. 1. The scenario of cross-camera action recognition. Note that instances in circles and triangles are actions captured by the source and target view camera, respectively. Our approach aims at utilizing labeled training data (colored circles) at the source view and unlabeled data pairs (in gray) from both views for recognizing unseen actions (in white) at the target view.

this subspace, depending on the corresponding correlation coefficient. Therefore, we need to take such *domain transfer ability* into consideration when designing the classifier in this joint subspace. We propose a novel SVM formulation, which incorporates such ability into classification in the joint subspace, so that the unseen actions at the target view can be projected and recognized accordingly. As shown in Figure 1, we focus on the scenario of using labeled data captured by the source camera for training (i.e., colored instances in Figure 1), and *no* training data is available at the target view. The unlabeled instance pairs (shown in gray in Figure 1) are collected from both views for transfer learning purposes ([5, 6] also have this requirement). Later in our experiments, the effectiveness of our proposed method will be verified.

2 Related Work

2.1 Action Recognition

One can divide existing works on action recognition into two categories: human body modeling and action representation [8]. The former aims at tracking joints of human body model and recognizing actions by predicting poses [9], while the latter utilizes spatial and temporal information for recognizing the associated action (e.g., spatiotemporal curvatures of 2D trajectories [2] or space-time volumes [10]). Inspired by the use of bag-of-words models for image classification, researchers also advocate the extraction of spatio-temporal descriptors [11, 12] for constructing the corresponding bag-of-words model for recognition. In such cases, actions are thus described by histograms of *visual words*.

2.2 Cross-View Action Recognition

For cross camera/view action recognition, only labeled instances collected by one or multiple source view cameras are available for training. Since *both* training

and test data at the target view cannot be seen in advance, this scenario makes cross-view action recognition very challenging. Some researchers aim at designing view-invariant representation [2, 3]. Alternatively, one can approach this problem as solving a matching task according to the quality of recovered geometry [13].

Recently, transfer learning has attracted the attention from researchers, and it has been successfully applied to cross-camera action recognition. The goal of transfer learning is to first learn a model to distinguish between different actions using training (labeled) data $\mathcal{D}_l^s$ from the source view domain $\mathcal{X}^s \in \mathbb{R}^{d_s}$. Once this model is observed, transfer learning aims at mapping this model into the target view domain $\mathcal{X}^t \in \mathbb{R}^{d_t}$ by utilizing unlabeled instance pairs $(\mathcal{D}_u^s, \mathcal{D}_u^t)$ collected by cameras at both source and target views. Generally, these approaches focus on determining a *shared representation* for both views when representing a data instance. For example, Farhadi and Tabrizi [5] propose to learn split-based features for source-view frames based on local data structure. They convert such features to the corresponding frames at the target view, so that actions at the target view can be encoded and recognized accordingly. However, their method requires the assumption that the local data structures at two domains are consistent, which might not be practical. Li and Zickler [14] characterize the source and target domains as two points on a Grassmann manifold, and they take the sampled points between them (along the geodesic) as the shared feature representation. Their approach considers data in different feature spaces lie on a low-dimensional manifold, and thus implicitly assumes their local structures are similar. Besides the implicit assumption of similar local structures for both domains, another concerns for the above methods is the requirement of $d_s = d_t$, i.e., the feature dimensions of source and target domains must be the same, which also limits their practical uses. Recently, Liu *et al.* [6] advocate to construct a bilingual codebook as a shared feature representation for both domains. With unlabeled data collected from both domains, their approach learns a shared codebook for two views in terms of a bipartite graph, and the bilingual words are obtained by spectral clustering. Although this approach does not require similarities of local data structure and allows features dimensions of the two views to be different, the shared feature attributes are considered to be *equally important*, which may not be preferable if the (shared) features extracted from each domain have uncoordinated contributions.

3 Our Proposed Method

3.1 Learning Correlation Subspace via CCA

The idea of applying transfer learning for cross-view action recognition is to determine a common representation (e.g., a joint subspace) for features extracted from source and target views, so that the model trained from the source-view data can be applied to recognize test data observed at the target view. Among existing methods [15, 5, 16, 6], canonical correlation analysis (CCA) is a very effective technique. It aims at maximizing the correlation between two variable sets [15, 16] and thus fits the goal of this work.

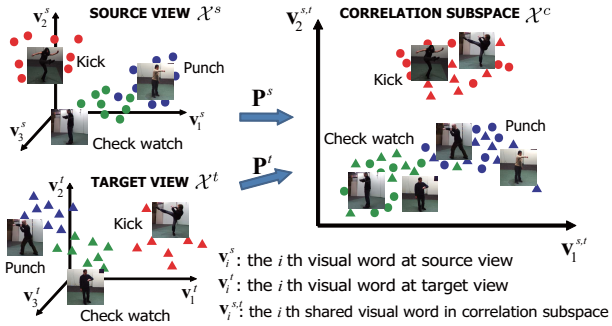

Fig. 2. Transfer learning via CCA [15]. Note that $\mathbf{P}^s$ and $\mathbf{P}^t$ are the projection matrices derived by CCA.

For the sake of completeness, we briefly review CCA as follows. Given two sets of n centered unlabeled observations $\mathbf{X}^s = [\mathbf{x}_1^s, \ldots, \mathbf{x}_n^s] \in \mathbb{R}^{d_s \times n}$ and $\mathbf{X}^t = [\mathbf{x}_1^t, \ldots, \mathbf{x}_n^t] \in \mathbb{R}^{d_t \times n}$ ($\mathbf{x}_i^s \in \mathcal{D}_u^s$ and $\mathbf{x}_i^t \in \mathcal{D}_u^s$) in source and target views respectively, CCA learns the projection vectors $\mathbf{u}^s \in \mathbb{R}^{d_s}$ and $\mathbf{u}^t \in \mathbb{R}^{d_t}$, which maximizes the correlation coefficient ρ:

$$\max_{\mathbf{u}^s, \mathbf{u}^t} \rho = \frac{\mathbf{u}^{s\top} \mathbf{\Sigma}_{st} \mathbf{u}^t}{\sqrt{\mathbf{u}^{s\top} \mathbf{\Sigma}_{ss} \mathbf{u}^s} \sqrt{\mathbf{u}^{t\top} \mathbf{\Sigma}_{tt} \mathbf{u}^t}}, \tag{1}$$

where $\mathbf{\Sigma}_{st} = \mathbf{X}^s \mathbf{X}^{t\top}$, $\mathbf{\Sigma}_{ss} = \mathbf{X}^s \mathbf{X}^{s\top}$, $\mathbf{\Sigma}_{tt} = \mathbf{X}^t \mathbf{X}^{t\top}$, and $\rho \in [0, 1]$. As suggested by [16], $\mathbf{u}^s$ in (1) can be solved by a generalized eigenvalue decomposition problem:

$$\mathbf{\Sigma}_{st} (\mathbf{\Sigma}_{tt})^{-1} \mathbf{\Sigma}_{st}^\top \mathbf{u}^s = \eta \mathbf{\Sigma}_{ss} \mathbf{u}^s. \tag{2}$$

Once $\mathbf{u}^s$ is obtained, $\mathbf{u}^t$ can be calculated by $\mathbf{\Sigma}_{tt}^{-1} \mathbf{\Sigma}_{st} \mathbf{u}^s / \eta$. In practice, regularization terms $\lambda_s \mathbf{I}$ and $\lambda_t \mathbf{I}$ need to be added into $\mathbf{\Sigma}_{ss}$ and $\mathbf{\Sigma}_{tt}$ to avoid overfitting and singularity problems. As a result, one solves the following problem instead:

$$\mathbf{\Sigma}_{st} (\mathbf{\Sigma}_{tt} + \lambda_t \mathbf{I})^{-1} \mathbf{\Sigma}_{st}^\top \mathbf{u}^s = \eta (\mathbf{\Sigma}_{ss} + \lambda_s \mathbf{I}) \mathbf{u}^s. \tag{3}$$

Generally, one can derive more than one pair of projection vectors $\{\mathbf{u}_i^s\}_{i=1}^d$ and $\{\mathbf{u}_i^t\}_{i=1}^d$ with corresponding ρ_i in a descending order (i.e., $\rho_i > \rho_{i+1}$). Thus, the source (target) view data $\mathbf{X}^s$ ($\mathbf{X}^t$) projected onto $\mathbf{u}^s$ ($\mathbf{u}^t$) will lie in the *correlation subspace* $\mathcal{X}^c \in \mathbb{R}^d$, which is spanned by $\{\mathbf{v}_i^{s,t}\}_{i=1}^d$.

Figure 2 shows a CCA example for cross-view action recognition. Given data of three action classes in source and target views ($\mathcal{X}^s$ and $\mathcal{X}^t$), CCA determines projection matrices $\mathbf{P}^s = [\mathbf{u}_1^s, \ldots, \mathbf{u}_d^s] \in \mathbb{R}^{d_s \times d}$ and $\mathbf{P}^t = [\mathbf{u}_1^t, \ldots, \mathbf{u}_d^t] \in \mathbb{R}^{d_t \times d}$. Once the correlation subspace $\mathcal{X}^c \in \mathbb{R}^d$ is derived, unseen test data at the target view can be directly recognized by the model trained from the source view data projected onto $\mathcal{X}^c$.

3.2 Domain Transfer Ability of CCA

As discussed in Section 3.1, unseen test at the target view can be first projected onto the CCA correlation subspace $\mathcal{X}^c$, and thus the model learned from the

source view data at this subspace can be applied for recognition. It is worth repeating that each dimension $\mathbf{v}_i^{s,t}$ in this subspace is associated with a different correlation coefficient ρ_i; the higher ρ_i is, the closer the projected data from different domains are. It is obvious that, a better *domain transfer ability* is resulted for the dominant dimensions $\mathbf{v}_i^{s,t}$ with larger ρ_i, and thus one should take such ability into consideration when designing a classification model in this correlation subspace.

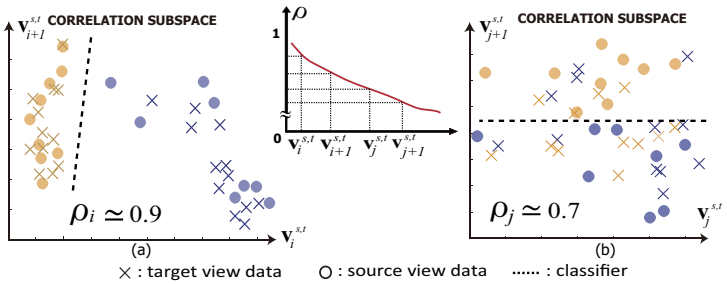

Fig. 3. Projecting source and target view instances from the IXMAS dataset into different correlation subspaces using projection vectors with different ρ.

Figure 3 illustrates this issue by projecting source and target view data onto different 2D correlation subspaces, in which one subspace is associated with ($\mathbf{v}_i^{s,t}$ and $\mathbf{v}_{i+1}^{s,t}$) with higher ρ, and the other one is constructed by ($\mathbf{v}_j^{s,t}$ and $\mathbf{v}_{j+1}^{s,t}$) with smaller ρ values. The dash lines represent the classifier learned from projected source view data (since no labeled data in the target domain is available). From Figure 3(a), we see that the location of projected source and target data with the same label are close to each other, since the two basis vectors correspond to larger ρ values. On the other hand, as shown in Figure 3(b), the distributions of projected source and target view data are different due to a lower ρ. As a result, the classifier learned from projected source view data (i.e., the dash lines) cannot generalize well to the projected target view ones. In other words, poorer domain transfer ability will result in increased recognition error, even the classifier is well designed using the projected source view data.

To overcome such limitations for CCA in transfer learning, we advocate the *adaptation* of the learning model based on the domain transfer ability. Based on the formulation of support vector machine (SVM), we propose a new SVM formulation which takes such ability into account, and it can be applied to address cross-view recognition.

3.3 The Proposed SVM Formulation

Generally, if the ith feature attribute exhibits better discrimination ability, the standard SVM would produce a larger magnitude for the corresponding model (i.e., a larger $|w_i|$). As discussed earlier, transfer leaning via CCA does not take the domain transfer ability into account when learning the classifiers in the

correlation subspace and thus degrades the recognition performance. To address this problem, we introduce a correlation regularizer and propose a novel SVM formulation which integrates the domain transfer ability and class discrimination in a unified framework. Due to the introduction of such ability, the generalization of our SVM for transfer leaning will be significantly improved.

The proposed SVM solves the following problem:

$$\begin{aligned} &\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{N}\xi_i - \frac{1}{2}\mathbf{r}^\top \mathrm{Abs}(\mathbf{w}) \\ &\text{s.t. } \; y_i(\langle \mathbf{w}, \mathbf{P}^{s\top}\mathbf{x}_i^s\rangle + b) + \xi_i \geq 1,\ \xi_i \geq 0,\ \forall(\mathbf{x}_i^s, y_i) \in \mathcal{D}_l^s, \end{aligned} \tag{4}$$

where $\mathrm{Abs}(\mathbf{w}) \equiv [|w_1|, |w_2|, \ldots, |w_d|]$ and $\mathbf{r} \equiv [\rho_1, \ldots, \rho_d]$ is the correlation vector in which each element indicates the correlation coefficient of CCA for each projection dimension. Note that only labeled source domain data $\mathbf{x}_i^s \in \mathcal{D}_l^s$ is available for training (not target domain data), and y_i is the associated class label. Parameters C and ξ are penalty term and slack variables as in the standard SVM. We have $\mathbf{P}^{s\top}\mathbf{x}_i^s$ as the projection of source domain data $\mathbf{x}_i^s$ onto the correlation subspace $\mathcal{X}_c$. The proposed term $\mathbf{r}^\top \mathrm{Abs}(\mathbf{w})$, which is introduced for model adaptation based on CCA, can be regarded as a similarity measure for $\mathbf{r}$ and $\mathbf{w}$. More precisely, a smaller correlation coefficient ρ_i would enforce the shrinkage of the corresponding $|w_i|$, and thus suppresses the learned model along the ith CCA projection vector; on the other hand, a larger ρ_i favors the contribution of the associated $|w_i|$ when minimizing (4).

Since it is not straightforward to solve the minimization problem in (4) with $\mathrm{Abs}(\mathbf{w})$, we seek the approximated solution by relaxing the original problem into the following form:

$$\begin{aligned} &\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{N}\xi_i - \frac{1}{2}(\mathbf{r}\odot\mathbf{r})^\top(\mathbf{w}\odot\mathbf{w}) \\ &\text{s.t. } \; y_i(\langle \mathbf{w}, \mathbf{P}^{s\top}\mathbf{x}_i^s\rangle + b) + \xi_i \geq 1,\ \xi_i \geq 0,\ \forall(\mathbf{x}_i^s, y_i) \in \mathcal{D}_l^s, \end{aligned} \tag{5}$$

where $\odot$ indicates the element-wise multiplication. We can further simplify (5) as:

$$\begin{aligned} &\min_{\mathbf{w}} \frac{1}{2}\sum_{i=1}^{d}(1-\rho_i^2)w_i^2 + C\sum_{i=1}^{N}\xi_i \\ &\text{s.t. } \; y_i(\langle \mathbf{w}, \mathbf{P}^{s\top}\mathbf{x}_i^s\rangle + b) + \xi_i \geq 1,\ \xi_i \geq 0,\ \forall(\mathbf{x}_i^s, y_i) \in \mathcal{D}_l^s. \end{aligned} \tag{6}$$

We refer to (6) as our proposed SVM formulation. Recall that $0 < \rho_i < 1$ in CCA, so that the convexity of the proposed objective function is guaranteed. It can be seen that, depending on the derived correlation coefficients, the formulation in (6) is effectively weighting each component of the regularization term accordingly. As a result, this modified SVM automatically adapt the derived classification model $\mathbf{w}$ based on the domain transfer ability of CCA, and thus it exhibits better generalization in recognizing projected unseen test data in the

correlation subspace (as confirmed by our experiments). The decision function for classifying unseen test data at target domain is shown as follows:

$$f(\mathbf{x}) = \text{sgn}\ (\langle \mathbf{w}, \mathbf{P}^{t\top}\mathbf{x}^t \rangle + b), \tag{7}$$

where $\mathbf{P}^t$ projects the input test data $\mathbf{x}^t$ from the target domain onto the correlation subspace $\mathcal{X}_c$.

4 EXPERIMENTS

4.1 Dataset and Experiment Settings

We consider the IXMAS multiview action dataset [3] which contains action videos of eleven action classes. Each action video is performed three times by twelve actors, and the actions are synchronically captured by five cameras, as shown in Figure 4. For a fair comparison with recent works such as [6], we extract descriptors defined by [11] and describe each action video as a group of spatio-temporal cuboids (at most 200). For each view these cuboids are quantized into $N = 1000$ visual words. As for data partition, we randomly choose two thirds of the video instances in each class as unlabeled data, and the rest are labeled data for training purposes. We follow the leave-one-action-out strategy as [6] did, which means we consider only one unseen action class at the target view to be recognized, and we exclude all instances of that class at both views when selecting the unlabeled data. The regularization terms λ_s and λ_t in (3) are both empirically set as 50. Instead of using a predetermined dimension number d (as [16] did), we select $\mathbf{v}_i^{s,t}$ with the corresponding correlation coefficient ρ_i above 0.5 for spanning the correlation subspace, and only the labeled data projected from the source view to this subspace are used for training. We repeat the above setting for each action class of interest, and report the average recognition performance in Figure 5.

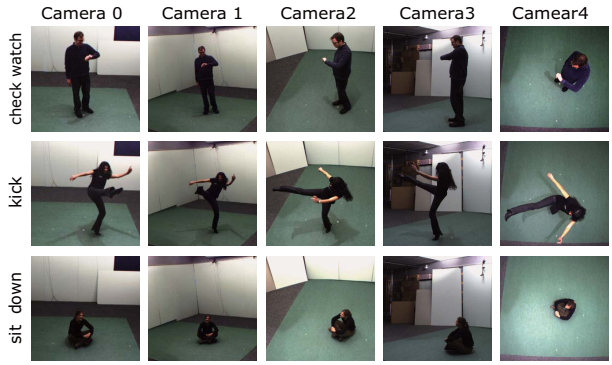

Fig. 4. Example actions of the IXMAS dataset. Each row represents an action at five different views.

4.2 Discussions

To compare our performance with other approaches, we consider the methods of direct prediction using classifiers learned at the source view (i.e., standard BoW without transfer learning), and the bag-of-bilingual-words (BoBW) model proposed in [6]. We note that, the above two approaches apply the standard linear SVM after deriving the feature representation for training/testing. Besides CCA [15], to argue that our SVM can be extended to other methods based on joint feature representations, we also consider a variant of BoBW [6]. We first compute the correlation between the source and target view data in terms of the derived BoBW, and apply our SVM with the correlation regularizer using the associated correlation coefficients (i.e., BoBW + our SVM in Figure 5).

	camera0						camera1						camera2					
	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
cam0			-				9.29	60.96	63.03	63.18	63.23	**64.90**	11.62	41.21	50.76	56.97	56.67	**60.61**
cam1	10.71	58.08	59.70	66.72	65.40	**70.25**				-			7.12	33.54	38.03	57.83	**61.97**	59.34
cam2	8.79	52.63	49.34	57.37	58.33	**62.47**	6.67	50.86	45.79	59.19	59.60	**61.87**				-		
cam3	6.31	40.35	44.44	65.30	61.87	**66.01**	9.75	33.59	33.27	46.77	48.43	**52.68**	5.96	41.26	43.99	61.36	**63.74**	61.36
cam4	5.35	38.59	40.91	54.39	51.52	**55.76**	9.44	37.53	37.00	53.59	49.24	**55.00**	9.19	34.80	38.28	57.88	57.88	**60.15**
avg.	7.79	47.41	48.60	60.95	59.28	**63.62**	8.79	45.73	44.77	55.68	55.13	**58.61**	8.47	37.70	42.77	58.51	60.06	**60.37**

	camera3						camera4					
	A	B	C	D	E	F	A	B	C	D	E	F
cam0	7.78	39.65	41.36	**63.64**	57.37	62.17	7.12	24.60	37.02	43.69	42.22	**48.23**
cam1	12.02	35.91	39.14	48.59	46.92	**54.85**	8.89	26.87	22.22	44.24	41.36	**49.29**
cam2	6.46	41.46	42.78	60.00	61.31	**61.46**	10.35	28.03	33.43	45.05	46.11	**51.82**
cam3				-			8.89	27.53	28.28	40.66	41.01	**41.06**
cam4	9.60	27.68	34.60	48.03	45.51	**48.89**				-		
avg.	8.96	36.17	39.47	55.06	52.78	**56.84**	8.81	26.76	30.24	43.41	42.68	**47.60**

Fig. 5. Performance comparisons on the IXMAS dataset. Note that each row indicates the source view camera (for training), and each column is the target view camera for recognizing the unseen action class. We consider the methods of A: BoW without transfer learning [11], B: BoBW [6], C: BoBW + our SVM, D: CCA + linear SVM [15], E: CCA + nonlinear SVM, and F: our proposed framework (CCA + our SVM).

From Figure 5, we see that the method without transfer learning (i.e., columns A) achieved the poorest results as expected. While the BoBW model (columns B) and the approach of CCA (columns D) remarkably improved the performance by determining a shared representation for training/test, the use of our SVM for BoBW (columns C) produced comparable or better results than the simple use of BoBW did, and the integration of CCA with our proposed SVM (columns F) achieved the best performance. Comparing the results shown in columns C and F, although our SVM taking the correlation of the source and target view data was able to improve the recognition performance, it would still be desirable to derive such correlation from a correlation-based transfer learning approach such as CCA. This explains why our approach combining CCA and imposing the resulting correlation coefficient into the proposed SVM formulation achieved the best recognition performance.

We further investigate the effectiveness of the proposed SVM over the standard one in terms of domain transfer ability. Figure 6(a) and (b) show the averaged value $|w_i|$ of each attribute in the standard and our SVM models using the

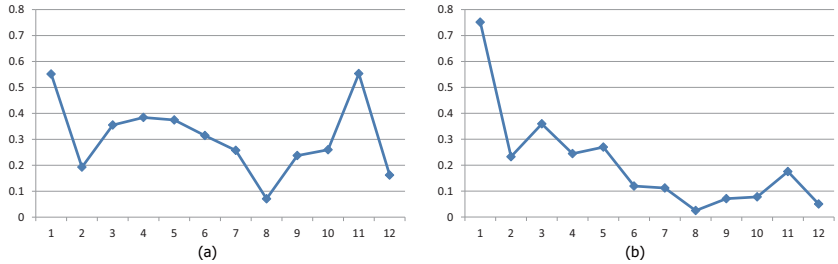

Fig. 6. Comparisons of the averaged $|w_i|$ values: (a) standard SVM and (b) our proposed SVM. The horizontal axis indicates the index of the dimension in the correlated subspace (arranged according to the associated correlation coefficients in a descending order). The vertical axis shows the associated $|w_i|$ values. The recognition rates for the two SVMs on recognizing the action "get-up" are 47.22% and 77.78%, respectively.

IXMAS dataset, respectively. From Figure 6(a), we see that the standard SVM aims at separating data the in the correlated subspace without considering the domain transfer ability (i.e., the correlation between projected data), and thus we still observe prominent $|w_i|$ values at non-dominant feature dimensions (i.e., the 11th dimension). On the other hand, in Figure 6(b), our proposed SVM suppresses the contributions of non-dominant feature dimensions in the correlated subspace, and thus only results in large $|w_i|$ values for dominant feature dimensions. The actual recognition rates for the two models were 47.22% and 77.78% for the action "get-up." Such a significant recognition improvement verifies that the leaning and enforcement of domain transfer ability of our proposed SVM model are preferable for transfer learning based cross-view action recognition.

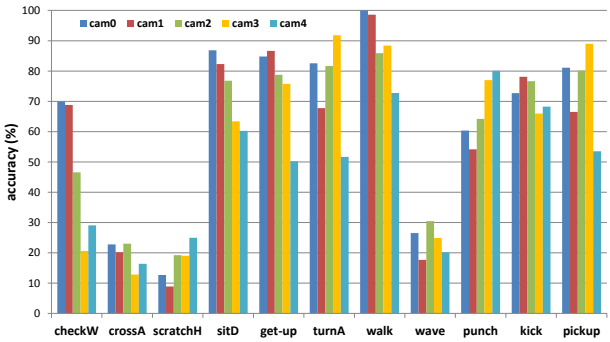

Fig. 7. Average recognition rates at different target views for each action category.

Figure 7 compares the recognition performance of each action for different target views. As expected, we observe that the transfer of recognition models is more challenging for certain actions/views (e.g., cross-arms, wave, etc. actions only with movements of arms). In general, camera 4 (top view) obtains the lowest recognition rate, and it is mainly due to the ambiguity between different torso-associated actions observed at this view.

5 Conclusions

We proposed a transfer learning based approach to cross-camera action recognition. By exploring the correlation subspace derived by CCA using unlabeled data pairs of source and target view data, we presented a novel SVM formulation with a correlation regularizer. The proposed SVM takes the domain transfer ability into consideration when designing the classifier at the correlation subspace. As a result, only projected and labeled training data from the source view are required when designing the classifier in the resulting subspace (i.e., no training data at the target view is needed). Experimental results on the IXMAS dataset confirmed the use of our proposed framework for improved recognition, and we verified that our approach outperformed state-of-the-art transfer learning algorithms which did not take such domain transfer ability into consideration.

References

1. Holte, M., Tran, C., Trivedi, M., Moeslund, T.: Human action recognition using multiple views: a comparative perspective on recent developments. In: ACM MM joint workshop on HGBU. (2011)
2. Rao, C., Yilmaz, A., Shah, M.: View-invariant representation and recognition of actions. IJCV **50** (2002) 203–226
3. Weinland, D., Ronfard, R., Boyer, E.: Free viewpoint action recognition using motion history volumes. CVIU **104** (2006) 249–257
4. Pan, S., Yang, Q.: A survey on transfer learning. IEEE TKDE **22** (2010) 1345–1359
5. Farhadi, A., Tabrizi, M.: Learning to recognize activities from the wrong view point. In: ECCV. (2008)
6. Liu, J., Shah, M., Kuipers, B., Savarese, S.: Cross-view action recognition via view knowledge transfer. In: CVPR. (2011)
7. Hotelling, H.: Relations between two sets of variates. Biometrika **28** (1936) 321–377
8. Poppe, R.: A survey on vision-based human action recognition. IVC **28** (2010) 976–990
9. Tran, C., Trivedi, M.: Human body modelling and tracking using volumetric representation: Selected recent studies and possibilities for extensions. In: ICDSC. (2008)
10. Blank, M., Gorelick, L., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. In: ICCV. (2005)
11. Dollár, P., Rabaud, V., Cottrell, G., Belongie, S.: Behavior recognition via sparse spatio-temporal features. In: ICCV joint workshop on VS-PETS. (2005)
12. Laptev, I.: On space-time interest points. IJCV **64** (2005) 107–123
13. ul Haq, A., Gondal, I., Murshed, M.: On dynamic scene geometry for view-invariant action matching. In: CVPR. (2011)
14. Li, R., Zickler, T.: Discriminative virtual views for cross-view action recognition. In: CVPR. (2012)
15. Blitzer, J., Foster, D., Kakade, S.: Domain adaptation with coupled subspaces. In: AISTATS. (2011)
16. Hardoon, D., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: An overview with application to learning methods. Neural Computation **16** (2004) 2639–2664