ENHANCED CANONICAL CORRELATION ANALYSIS WITH LOCAL DENSITY FOR CROSS-DOMAIN VISUAL CLASSIFICATION

Wei-Jen Ko[1,2], *Jheng-Ying Yu*[1,2], *Wei-Yu Chen*[1,2], *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

ABSTRACT

Real-world visual classification tasks typically need to deal with data observed from different domains. Inspired by canonical correlation analysis (CCA), we propose an enhanced CCA with local density for associating and recognizing cross-domain data. In addition to maximizing the correlation of the projected cross-domain data, our CCA model further exploits the local density information observed from each domain. As a result, our CCA not only exhibits excellent abilities in identifying representative data, noisy data like outliers can be further suppressed during the derivation of our CCA subspace. In our experiments, we successfully apply the proposed methods for solving two cross-domain classification tasks: person re-identification and cross-view action recognition.

***Index Terms*—** Canonical Correlation Analysis, Person Re-identification, Cross-View Action Recognition

1. INTRODUCTION

When dealing with real-world visual classification problems, data observed beforehand and those to be recognized might be collected from different domains and thus exhibit distinct feature distributions. For example, one might need to recognize the context of an image captured by a smartphone, while the training data are obtained from the Internet. As a result, how to adapt the information from the source to target domains of interest is a challenging task.

For solving the above problem of domain adaptation, we typically focus on learning feature or classification models in the target domain (or a common feature space) by leveraging label and data information across domains. For example, Sugiyama et al. [1] proposed an instance re-weighting approach of covariate shift, which derives the target classification model by re-weighting the labeled samples projected from the source domain, with the goal of minimizing the approximated empirical classification error in the target domain. Such domain adaptation methods have been successfully applied to applications like cross-domain object recognition or cross-language text categorization.

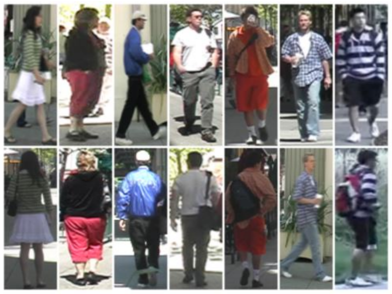

Fig. 1: Person re-identification (PR-ID) as a cross-domain visual classification task. Note that each column shows images of the same person captured by cameras in different views.

To identify the identity of the subjects across different cameras, person re-identification (PRID) can also be viewed as a cross-domain classification problem (see Figure 1), which plays an important role in applications of surveillance and video forensics. Since the cameras are typically distributed at locations with very different angle and lighting conditions (plus possible occlusion), it makes the matching of an image pair captured by different cameras very difficult.

Recently, researchers advanced learning-based approaches for solving PRID problems. For instance, Zheng et al. [2] proposed the PRDC algorithm, which is able to calculate and compare the relative distances across camera pairs. On the other hand, Prosser et al. [3] and Avraham et al. [4] regarded PRID as ranking and domain adaptation problems, respectively. In addition, metric learning has been applied to derive a proper distance metric, aiming at projecting cross-camera images into a feature space for matching purposes. Recently developed approaches include Large Margin Nearest Neighbor (LMNN) [5], Information Theoretic Metric Learning (ITML) [6], and Logistic Discriminant Metric Learning [7].

Canonical correlation analysis (CCA) [8] is a subspace learning algorithm, which aims at learning a common feature space by observing cross-domain data pairs, with the objective to maximize the correlation between the projected cross-domain data pairs. A major advantage of CCA is its ability of relating heterogeneous cross-domain data (i.e., source and target domain data in different feature representation). CCA has

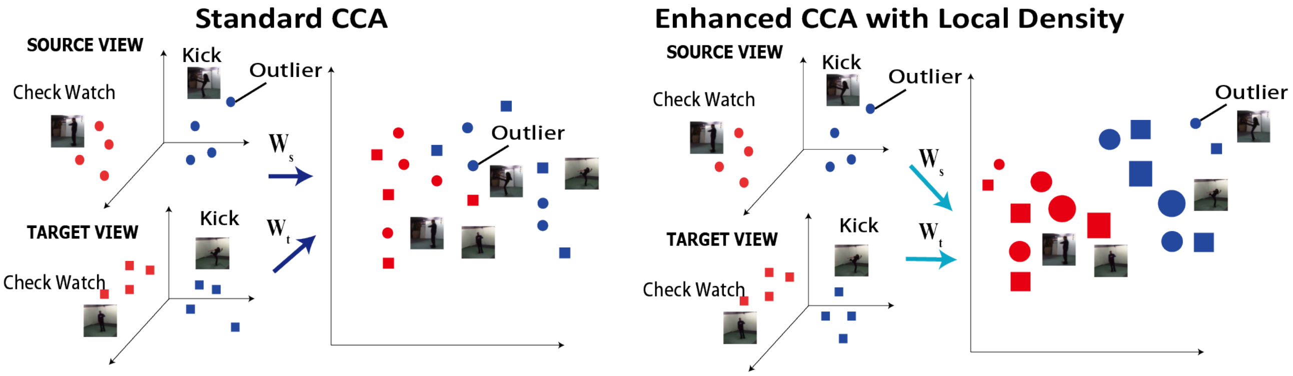

Fig. 2: Illustrations of standard and our enhanced CCA with local density. Note that our CCA is able to identify instances with higher contributions (shown in larger dots), and thus exhibits improved ability in suppressing outlier data.

been successfully applied to solve a variety of cross-domain visual classification tasks including PRID.

Several variants if CCA has been proposed, including the Ranking CCA for learning query and image similarities, which simultaneously learns a bilinear query image similarity function and adjusts the subspace to preserve the preference relations.[9] Luo et el. proppposed the tensor canonical correlation analysis, which maximizes the canonical correlation of more than two views simultaneously.[10]

In this paper, we propose a novel learning-based algorithm for cross-domain visual classification. Based on CCA, our method aims at deriving a common feature space which maximizes the correlation between projected cross-camera data. Moreover, we exploit *local density information* observed from cross-domain data, which identifies the contributions of each instance during the above association. As verified in our experiments, our method is able to better associate cross-camera data while suppressing noisy or outlier data observed from either domain.

We now summarize our contributions as follows:

- We propose an enhanced CCA with local density observed from cross-domain data for solving cross-domain visual classification problems.
- Our proposed model is able to identify representative instances while suppressing contributions from noisy or outlier data when relating cross-domain data.
- We apply our enhanced CCA for solving the tasks of person re-identification and cross-view action recognition. Our method is shown to perform favorably against CCA-based approaches.

2. OUR PROPOSED METHOD

2.1. A Brief Review of Canonical Correlation Analysis

For the sake of completeness and the ease of discussion for the remaining of this paper, we now provide the formulation of standard CCA. Given n data pairs across two different domains $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n] \in \mathbb{R}^{d_x \times n}, \mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ...\mathbf{y}_n] \in \mathbb{R}^{d_y \times n}$, CCA derives a pair of projection vectors $\mathbf{w}_x$ and $\mathbf{w}_y$, which maximizes the correlation coefficient ρ whic is calculated as follow:

$$\max_{\mathbf{w}_x,\mathbf{w}_y} \rho = \frac{\mathbf{w}_x^T \mathbf{X}\mathbf{Y}^T \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{X}\mathbf{X}^T \mathbf{w}_x}\sqrt{\mathbf{w}_y^T \mathbf{Y}\mathbf{Y}^T \mathbf{w}_y}}. \tag{1}$$

We note that, solving the optimization problem for CCA can be equivalently formulated as follows:

$$\begin{aligned} &\max_{\mathbf{w}_x,\mathbf{w}_y} \mathbf{w}_x^T \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{w}_y \\ &s.t. \mathbf{w}_x^T \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{w}_x = 1 \\ &\mathbf{w}_y^T \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{y}_i - \mathbf{y}_j)(\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{w}_y = 1. \end{aligned} \tag{2}$$

Later in Section 2.2, we will apply and extend the above formulation for our proposed CCA algorithm.

2.2. Enhanced CCA with Local Density

We now introduce our proposed method for solving PRID problems. In standard CCA, all training cross-domain data pairs are viewed as equally important. In other words, they would be applied to derive the CCA projections with the same contributions.

However, when dealing with real-world image data, we should be able to identify those with more representative or discriminating information, while suppressing the ones which are corrupted due to noise or occlusion. Based on the above observation and motivation, we present an enhanced CCA by exploiting cross-domain local density information.

In our work, we consider training instances with more neighbors are more representative (i.e., exemplars), and thus such data are less likely to be the outliers for affecting the resulting CCA model. To be more specific, we aim at identifying cross-domain training instances with different corresponding weights a_i, b_i. These weights can be easily calculated based on the distance between each instance of interest to its neighbors. As a result, our enhanced CCA with local density can be formulated as:

$$\max_{\mathbf{w}_x,\mathbf{w}_y} \mathbf{w}_x^T \sum_{i=1}^{n}\sum_{j=1}^{n}(a_i\mathbf{x}_i - a_j\mathbf{x}_j)(b_i\mathbf{y}_i - b_j\mathbf{y}_j)^T\mathbf{w}_y$$

$$s.t. \mathbf{w}_x^T \sum_{i=1}^{n}\sum_{j=1}^{n}(a_i\mathbf{x}_i - a_j\mathbf{x}_j)(a_i\mathbf{x}_i - a_j\mathbf{x}_j)^T\mathbf{w}_x = 1 \quad (3)$$

$$\mathbf{w}_y^T \sum_{i=1}^{n}\sum_{j=1}^{n}(b_i\mathbf{y}_i - b_j\mathbf{y}_j)(b_i\mathbf{y}_i - b_j\mathbf{y}_j)^T\mathbf{w}_y = 1.$$

In the proposed CCA formulation in (3), the instance weights a_i, b_i are calculated by the sum of the observed local density $\mathbf{S}$, i.e.,

$$a_i = \sum_{k=1}^{n}\mathbf{S}_x(i,k) \quad (4)$$

$$b_i = \sum_{k=1}^{n}\mathbf{S}_y(i,k), \quad (5)$$

where the local density is defined as an exponentially decaying function of distance with a threshold t. For each instance pair $(\mathbf{m}_i, \mathbf{m}_j)$ in either domain $\mathbf{m} \in \{\mathbf{x}, \mathbf{y}\}$, we have

$$\mathbf{S}_{\mathbf{m}}(i,j) = \begin{cases} e^{-d(\mathbf{m}_i,\mathbf{m}_j)/\bar{d}}, & \text{if } e^{-d(\mathbf{m}_i,\mathbf{m}_j)/\bar{d}} > t \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

$$\text{and } \bar{d} = 2\sum_{p=1}^{n}\sum_{q=1}^{n} d(\mathbf{m}_p, \mathbf{m}_q)/n(n-1). \quad (7)$$

In the above formulation, $d(\mathbf{m}_i, \mathbf{m}_j)$ denotes the distance between $\mathbf{m}_i$ and $\mathbf{m}_j$, and $\bar{d}$ is the mean distance of all data pairs within the same domain. We note that, in order to suppress the contribution of corrupted image data due to occlusion, we separate such images from the remaining ones by taking the Chebyshev distance as the distance metric, which is the maximal distance between variables of vectors.

With the above derivations, our CCA projection vectors $\mathbf{w}_x$ and $\mathbf{w}_y$ can be solved by a generalized eigenvalue decomposition problem as follow:

$$\mathbf{C}_{xy}(\mathbf{C}_{yy})^{-1}\mathbf{C}_{xy}^T\mathbf{w}_x = \eta\mathbf{C}_{xx}\mathbf{w}_x, \quad (8)$$

where $\mathbf{C}_{xy} = \boldsymbol{\alpha}_x{\boldsymbol{\alpha}_y}^T, \mathbf{C}_{xx} = \boldsymbol{\alpha}_x\boldsymbol{\alpha}_x^T, \mathbf{C}_{yy} = \boldsymbol{\alpha}_y\boldsymbol{\alpha}_y^T$, $\boldsymbol{\alpha}_x$ and $\boldsymbol{\alpha}_y$ denote the derived weighted matrices from the associated domain with data mean removed. To avoid singularity problems and overfitting, we add regularization terms λ_x, λ_y and solve the following problem instead:

$$\mathbf{C}_{xy}(\mathbf{C}_{yy} + \lambda_y\mathbf{I})^{-1}\mathbf{C}_{xy}^T\mathbf{w}_x = \eta(\mathbf{C}_{xx} + \lambda_x\mathbf{I})\mathbf{w}_x. \quad (9)$$

Finally, we derive $\mathbf{w}_y$ by $\mathbf{C}_{yy}^{-1}\mathbf{C}_{xy}\mathbf{w}_x/\eta$.

2.3. Locality-Preserving CCA vs. Our CCA

It is worth noting that, locality-preserving CCA (LPCCA) [11] and ALPCCA [12] are also extensions of CCA, which focus on preserving the local structure observed from either domain data when deriving the common feature space. We now explain how our CCA is different from such locality preserving versions, and why ours is expected to achieve improved performance when associating cross-domain data.

The formulation of LPCCA is derived as follow:

$$\max_{\mathbf{w}_x,\mathbf{w}_y} \mathbf{w}_x^T \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{S}_x(i,j)(\mathbf{x}_i - \mathbf{x}_j)\mathbf{S}_y(i,j)(\mathbf{y}_i - \mathbf{y}_j)^T\mathbf{w}_y$$

$$s.t. \mathbf{w}_x^T \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{S}_x(i,j)(\mathbf{x}_i - \mathbf{x}_j)\mathbf{S}_x(i,j)(\mathbf{x}_i - \mathbf{x}_j)^T\mathbf{w}_x = 1$$

$$\mathbf{w}_y^T \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{S}_y(i,j)(\mathbf{y}_i - \mathbf{y}_j)\mathbf{S}_y(i,j)(\mathbf{y}_i - \mathbf{y}_j)^T\mathbf{w}_y = 1. \quad (10)$$

And, that for the ALPCCA is written as:

$$\max_{\mathbf{w}_x,\mathbf{w}_y} \rho = \frac{\mathbf{w}_x^T\mathbf{X}(\mathbf{I} + \mathbf{S}_x + \mathbf{S}_y)\mathbf{Y}^T\mathbf{w}_y}{\sqrt{\mathbf{w}_x^T\mathbf{X}\mathbf{X}^T\mathbf{w}_x}\sqrt{\mathbf{w}_y^T\mathbf{Y}\mathbf{Y}^T\mathbf{w}_y}}. \quad (11)$$

From the above equations, it can be seen that LPCCA aims at preserving locally relative (or structural) information observed from data in each domain, while ALPCCA adds additional terms for integrating the above observed information into the data covariance matrix.

As verified later by our experiments, both LPCCA and ALPCCA do not exhibit sufficient abilities in identifying and distinguishing between representative and noisy data (as ours does). From (3), we see that our CCA derives sample-wise weighting and applies such local density information for learning the cross-domain transformation. This is very different from the above existing locality-preserving versions, which observe pair-wise data information even the instances are noisy or corrupted.

3. EXPERIMENTS

3.1. Person Re-Identification

We first address the cross-domain visual classification task of person re-identification. To conduct the experiments, we take the VIPeR [13] dataset, which contains 632 persons with 2 images per person captured by 2 cameras. Images in this dataset has extensive variations in viewpoint, pose, and illumination. We randomly pick 316 persons for training and the remaining 316 persons for testing. Example images of VIPer dataset are shown in Figure 1.

We resize each image into 128*64 pixels, and divide it into horizontal stripes. For each stripe, we extract weighted

Table 1: Top-N recognition rate on VIPER.

Rank	1	10	20	50	100
LMNN [5]	17	54	69	88	96
ITML [6]	13	53	71	90	97
PRDC [2]	16	54	70	87	97
EIML [14]	22	63	78	93	98
ICT [4]	14	60	78	-	-
RPLM [15]	27	69	83	95	99
eSDC [17]	27	62	76	-	-
SalMatch [16]	30	65	-	-	-
CCA [8]	18	55	68	83	95
LPCCA [11]	19	56	71	85	95
ALPCCA [12]	17	53	67	84	93
Our CCA	29	77	88	97	99

Table 2: Recognition results on the IXMAS dataset.

CAM	BoBW [22]	CCA+[21]	Ours+ [21]
0/1	67.52	75.95	79.69
0/2	63.91	75.76	78.33
0/3	59.28	77.84	79.54
0/4	57.48	75.38	77.57
1/0	65.72	75.57	75.75
1/2	62.51	75.01	73.78
1/3	57.39	75.47	76.06
1/4	55.02	74.05	75.60
2/0	61.08	78.22	78.03
2/1	57.95	76.04	78.93
2/3	57.29	77.94	81.66
2/4	57.29	77.65	78.03
3/0	60.61	78.31	78.33
3/1	56.91	76.61	77.12
3/2	61.27	79.55	78.93
3/4	53.13	74.24	76.36
4/0	48.01	70.08	75.15
4/1	47.44	71.31	71.96
4/2	55.11	74.15	76.51
4/3	45.51	74.05	73.03
Average	57.47	75.66	77.02

color histogram from 8 channels in the RGB, Lab, and HSV color spaces (discarding the V channel). To emphasize the center of the image and eliminate background influence, the weights are decided by an exponential Gaussian kernel. The color histograms are concatenated with a 4-bin histogram of oriented gradients (HOG) descriptors and Local Binary Patterns (LBP) extracted on a part of the image centered at the torso and legs. In our work, we fix parameters $\lambda_x = \lambda_y = \eta = 1, t = 0.65$.

After deriving the CCA subspace using the training image pairs, we project the probe images and the gallery ones onto this space for matching purposes. To measure the similarity between projected cross-domain data, we apply the cosine similarity and search for the nearest neighbor of each probe image in the gallery.

We compare our results with those produced by baseline and popular PR-ID methods such as PRDC [2], DDC, EIML [14], ICT [4], RPLM [15], SalMatch [16], and eSDC [17]. We also compare with the original CCA,LPCCA [11],and ALPCCA [12]. We conduct the experiments with 5 random trials, and lists the performance in Table 1. It can be observed that our proposed CCA improved the rank 1 recognition rate by 12% compared to the original CCA, and achieved comparable performance as SalMatch did. At rank 10 and above, our method outperforms all other compared methods. Both LPCCA and ALPCCA were not able to produce satisfactory performance, as explained in Section 2.3.

We note that, while a deep-learning version of CCA [18] is available, which replaces the two projection vectors in the standard CCA (one for each domain) by two neural networks. However, after implementing this CCA, we did not observe satisfactory results on solving this PR-ID problem (possibly due to the lack of a large amount of data for training). Therefore, we do not include its results in Table 1.

3.2. Cross-View Action Recognition

In addition, we apply our proposed method for solving cross-view action recognition problems. We consider the IXMAS [19] dataset, which contains videos of 11 different action categories. Each action is performed 3 times by 12 actors, and a total of 5 camera views are available.

To describe the action images, we extract feature descriptors as defined in [20], and contract a group of spatiotemporal cuboids (at most 200). For each video, the cuboids are quantized into 1000 visual words. In our experiments, we randomly choose two thirds of the cross-view action images in each action category as cross-view data pairs for learning our CCA. Then, the remaining one third of the images in the source domain for training the SVM classifier as proposed in [21]. Finally, the rest of the images in the target view are for testing. We repeat the above procedure ten times and report the average recognition performance.

We compare the performance of our method with the bag-of-bilingual-words (BoBW) model proposed in [22], and the use of standard CCA with the SVM proposed in [21]. The recognition results of different approaches are listed in Table 2. From this table, we see that our proposed method performed favorably against the other two state-of-the-art methods. Based on our experiments in Sections 3.1 and 3.2, the effectiveness of our proposed CCA for solving cross-domain visual classification can be successfully verified.

4. CONCLUSIONS

In this paper, we presented an enhanced CCA with local density for solving cross-domain visual classification problems. Our proposed CCA not only identifies representative cross-domain data when relating different domains, it can further suppress noisy or corrupted data during the learning process. This cannot be easily achieved by existing locality-preserving CCA. Our experiments on person re-identification and cross-view action recognition supported the use of our method, which outperformed CCA-based approaches.

5. REFERENCES

[1] M. Sugiyama and M. Kawanabe, "Machine learning in non-stationary environments: Introduction to covariate shift adaptation," in *MIT Press*, 2012.

[2] W. Zheng, S. Gong, and T. Xiang., "Re-identification by relative distance comparison.," in *IEEE T-PAMI*, 2012.

[3] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking," in *BMVC*, 2010.

[4] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, "Learning implicit transfer for person re-identification," in *ECCV Workshop*, 2012.

[5] K. Weinberger and L. Saul., "Fast solvers and efficient implementations for distance metric learning," in *ICML*, 2008.

[6] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon., "Information-theoretic metric learning," in *ICML*, 2007.

[7] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you?metric learning approaches for face identification," in *ICCV*, 2009.

[8] H. Hotelling, "Relations between two sets of variates," in *Biometrika*, 1936, vol. 28, pp. 321–377.

[9] C. W. Ngo T. Yao, T. Mei, "Learning query and image similarities with ranking canonical correlation analysis," in *Proc. ICCV*, 2015.

[10] Y. Wen K. Ramamohanarao C. Xu Y. Luo, D. Tao, "Tensor canonical correlation analysis for multi-view dimension reduction," in *stat.ML*, 2015.

[11] S. Chen T. Sun, "Locality preserving CCA with applications to data visualization and pose estimation," in *Image and Vision Computing*, 2007.

[12] D. Zhang F. Wang, "A new locality-preserving canonical correlation analysis algorithm for multi-view dimensionality reduction," in *Neural Process Letters*, 2013.

[13] D. Gray, S. Brennan, and H . Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. PETS*, 2007.

[14] M. Hirzer, P. Roth, and H. Bischof, "Person re-identification by efficient impostor-based metric learning," in *IEEE AVSS*, 2012.

[15] M. Hirzer, P. M. Roth, M. Kostinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification.," in *ECCV*, 2012.

[16] R. Zhao, W. Ouyang, and X. Wang., "Person re-identification by salience matching," in *IEEE ICCV*, 2013.

[17] X. W. R. Zhao and W. Ouyang, "Unsupervised salience learning for person re-identification," in *IEEE CVPR*, 2013.

[18] J. Bilmes G. Andrew, R. Arora and K. Livescu, "Deep canonical correlation analysis," in *ICML*, 2013.

[19] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," in *Computer Vision and Image Understanding*, 2006, vol. 104, pp. 249–257.

[20] P. Dollr, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *IEEE Int. Workshop VS-PETS*, 2005.

[21] Y. R. Yeh, C.-H. Huang, and Y. C. F. Wang, "Heterogeneous domain adaptation and classification by exploiting the correlation subspace," in *IEEE Transactions on Image Processing*, 2014.

[22] J. Liu, M. Shah, B. Kuipers, and S. Savarese, "Cross-view action recognition via view knowledge transfer," in *IEEE CVPR*, 2011.