UNSUPERVISED DOMAIN ADAPTATION USING MANIFOLD ALIGNMENT FOR OBJECT AND EVENT CATEGORIZATION

Suranjana Samanta and Sukhendu Das

V.P. Lab, Dept. of CS&E, Indian Institute of Technology Madras, India.
Email: ssamanta@cse.iitm.ac.in, sdas@iitm.ac.in

ABSTRACT

This paper describes a method of cross-domain object and event categorization, using the concept of domain adaptation. Here, a classifier is trained using samples from the source/ auxiliary domain and performance is observed on a set of test samples taken from a different domain, termed as the target domain. To overcome the difference between the two domains, we aim to find an optimal sub-space such that the instances from both the domains follow similar distributions when projected onto the sub-space. Along with the distributions, the underlying manifolds of the two domains are aligned in the sub-space to reduce the difference in structure of the data from the two domains. The local spatial arrangement of the instances in both the domains are also preserved in the optimal sub-space. Results show that the proposed method of unsupervised domain adaptation provides better classification accuracy than a few state of the art methods.

***Index Terms*—** Domain adaptation, transfer learning, manifold alignment, trace minimization, classification.

1. INTRODUCTION

The volume of images and videos to be a analyzed are increasing at an enormous rate due to availability of cheap hardwares. However, it is difficult to annotate and create a sufficient number of labeled training samples from various datasets, to perform application tasks like visual categorization, detection, recognition, retrieval etc. in images and videos. Domain adaptation (DA) is the process which uses labeled training samples available from one domain to improve the performance of statistical tasks to be done on test samples drawn from a different domain. The domain from which the training samples are obtained is called the source domain, and the domain from which the test samples are used is the target domain. To estimate the distribution of target domain, few training samples are also necessary from the target domain. Using the training samples from both the domains, the performance of a classifier on the test samples from target domain improves using the proposed method of DA.

The work has been partially funded by Tata Consultancy Services (TCS).

Our aim is to determine an optimal sub-space, such that after projection the distribution of the transformed source and target domains in the sub-space are similar. The local spatial arrangements of instances in both source and target domains are preserved in the sub-space. In addition, inter-class distance (between means) in the source domain is maximized, while the disparity between the structure of source and target domains is minimized, by representing the structure with a set of landmark points in the data. To deal with non-linear transformation of data, we use the concept of Reproducing Kernel Hilbert Space (RKHS) [1, 2] and estimate a suitable sub-space using the kernel function. Performance of the proposed method of unsupervised DA have been observed for object and event categorization in images and videos respectively. Comparative study with state of the art works suggests that the proposed method gives better result for both the tasks.

The rest of the paper is as follows. Section 2 briefly describes the state of the art of the related works. Section 3 explains the proposed method of DA. Section 4 shows the experimental results and section 5 concludes the paper.

2. BRIEF LITERATURE REVIEW

DA has gained enormous importance in the recent past. Depending on the kind (nature) of training samples in the target domain, the two specific categories of DA are: (i) supervised - very few number of labeled training samples are available; and (ii) unsupervised - large number of unlabeled (no class labels) training samples are available.

The proposed method uses the concept of local linear embedding or manifold alignment [3–5] and projects a higher dimensional data from the kernel space to a lower dimensional feature space. Pan et al. [1] proposed transfer component analysis (TCA), which minimizes the disparity of distribution by considering the difference of means between two domains and it also preserves local geometry of underlying manifold. We use similar ideas from TCA with variations. Instead of considering the covariance matrix to capture the scatter of the data, we consider the structure of the data represented by a few chosen landmark points, which is an advantage when the underlying distribution of the data is non-Gaussian. In addition, an optimal sub-space has been also calculated using the

concept of geodesic path in the Grassmannian manifold between source and target domains [6, 7]. Fernando et. al. [8] has calculated a subspace using eigen-vectors of two domains such that the basis vectors of transformed source and target domains are aligned. Manifold alignment has also been used for domain adaptation earlier. Wang et al. [9] has considered the manifold of each domain and estimated a latent space, where the manifolds of both the domains are similar to each other. However, the structures or the distributions of the domains have not been considered in this case. Application of DA for improved results of object categorization and video classification have been discussed in [6–8, 10–12].

3. PROPOSED METHOD OF DOMAIN ADAPTATION

Let X and Y be the source and target domains having n_X and n_Y number of instances respectively and d be the number of features (i.e. dimension). Let $\Phi(X)$ and $\Phi(Y)$ be the corresponding source and target domain data in kernel space of dimension d_k, where $\Phi(.)$ is the kernel function. If A is a matrix representing a dataset, then a_i represents the i^{th} instance of A and $tr(A)$ represents the trace of the matrix. The main principle of DA is to minimize the disparity in distribution between two domains. Distributions of two domains are same if their means are equal in the RKHS (see [2] for constraints). We estimate a transformation/projection matrix $W_{d_k \times d}$ onto an optimal sub-space, such that transformed source and target domain data have similar distributions. Let $\Phi(\tilde{X}) = \Phi(X)W$ and $\Phi(\tilde{Y}) = \Phi(Y)W$ be the transformed source and target domains. Let, $D \in \mathbb{R}^{(n_X+n_Y)\times d}$ be the unified dataset combining X and Y, and K be the corresponding Gram matrix defined as, $K = \Phi(D)\Phi(D)^T = \begin{bmatrix} K_{XX} & K_{XY} \\ K_{XY}^T & K_{YY} \end{bmatrix}$. Then, the transformed data is: $\Phi(\tilde{D}) = KZ$ where, $W = \Phi(D)^T Z$ ($Z \in \mathbb{R}^{(n_X+n_Y)\times d}$, estimated as unknown). K_{XX}, K_{XY}, K_{YY} are the Gram matrices defined as: $K_{XX} = \Phi(X)\Phi(X)^T$ and $K_{XY} = \Phi(X)\Phi(Y)^T$.

3.1. Difference in means of the two domains

The mean of $\Phi(\tilde{X})$ and $\Phi(\tilde{Y})$ are: $m_{\tilde{X}}^{\Phi} = \frac{1}{n_X}\sum_{i=1}^{n_X}\Phi(\tilde{x}_i)$ and $m_{\tilde{Y}}^{\Phi} = \frac{1}{n_Y}\sum_{i=1}^{n_Y}\Phi(\tilde{y}_i)$ respectively. The square of the distance between means of two domains is given as:

$$\begin{aligned} d_\mu^2 &= \left(m_{\tilde{X}}^{\Phi} - m_{\tilde{Y}}^{\Phi}\right)\left(m_{\tilde{X}}^{\Phi} - m_{\tilde{Y}}^{\Phi}\right)^T \\ &= tr(W^T\Phi(X)^T I_1 \Phi(X)W) \\ &\quad -tr(W^T\Phi(Y)^T 2I_2\Phi(X)W) + tr(W^T\Phi(Y)^T I_3\Phi(Y)W) \\ &= tr\left(W^T\Phi(D)^T\begin{bmatrix} I_1 & -I_2^T \\ I_2 & I_3 \end{bmatrix}\Phi(D)W\right) \\ &= tr\left(Z^T K\begin{bmatrix} I_1 & -I_2^T \\ I_2 & I_3 \end{bmatrix}KZ\right) \end{aligned} \quad (1)$$

where, $[I_1]_{n_X\times n_X}$, $[I_2]_{n_Y\times n_X}$ and $[I_3]_{n_Y\times n_Y}$ are matrices containing all elements as $1/n_X^2$, $1/n_X n_Y$ and $1/n_Y^2$ respectively. Similar notations and pattern of derivation (as in Eqn. 1) are followed in the rest of the paper.

3.2. Estimating inter-class distance in source domain

High inter-class distance is a desired property for a classification task. Hence in the subspace formed using W, the means of different classes should be far apart in $\Phi(\tilde{X})$. Let us consider a vector m_c of length n_X, such that $m_c(i) = 1/n_X^c$, iff $\Phi(x_i)$ belongs to class c (otherwise 0); where n_X^c is the number of instances belonging to the c^{th} class of $\Phi(X)$. Hence, the mean of the c^{th} class of $\Phi(X)$ is given by: $m_c\Phi(X)$ and the corresponding mean in $\tilde{X}$ is given by $m_c\Phi(X)W$. If C is the number of classes, then the sum of the squares of distances between all pairs of means of classes in $\Phi(\tilde{X})$ is:

$$\begin{aligned} d_\delta^2 &= \sum_{i=1}^{C}\sum_{j=1,j\neq i}^{C}(m_i - m_j)\Phi(X)WW^T\Phi(X)^T(m_i - m_j)^T \\ &= tr\left(Z^T K\begin{bmatrix} M & \mathbf{0}_{n_X\times n_Y} \\ \mathbf{0}_{n_Y\times n_X} & \mathbf{0}_{n_Y\times n_Y} \end{bmatrix}KZ\right) \end{aligned} \quad (2)$$

where, $M = \sum_{i=1}^{C}\sum_{j=1,j\neq i}^{C}\left(m_i^T m_i - 2m_j^T m_i + m_j^T m_j\right)$. d_δ^2 is not calculated for $\Phi(Y)$ due to lack of class-labels.

3.3. Preserving local spatial arrangement of data

The local spatial arrangement of the data can be captured using a neighborhood graph build on the data [13]. Let, the Euclidean distance between $\Phi(x_i)$ and $\Phi(x_j)$ be given as, $d_E^{ij} = K_{XX}(i,i) + K_{YY}(j,j) - 2\times K_{XY}(i,j)$. Also, consider a symmetric adjacency matrix $[A_X]_{n_X\times n_X}$ representing a minimal spanning tree (MST) build on the source domain $\Phi(X)$, which is defined as:

$$A_X(i,j) = A_X(j,i) = \begin{cases} 1 & \text{if } \forall k,\ d_E^{ij} < d_E^{ik} \text{ and } j \neq k \\ 0 & \text{otherwise} \end{cases}$$

To preserve the MST in the transformed space, the distance of $\Phi(\tilde{x}_i)$ from $\Phi(\tilde{x}_j)$ should be less than that from $\Phi(\tilde{x}_k)$, if $A_X(i,j) = 1$ and $A_X(i,k) = 0$, $\forall k, k \neq j$. Hence if $A_X(i,j) = 1$, for any $\{i,j\}$, we minimize the square of distance between $\Phi(\tilde{x}_i)$ and $\Phi(\tilde{x}_j)$. The required sum of square of distances to be minimized is thus:

$$\begin{aligned} d_{S-MST}^2 &= \sum_{i=1}^{n_X}\sum_{j=1}^{n_X}A_X(i,j)d_E^{ij} \\ &= tr\left(Z^T K\begin{bmatrix} 2(B_X - A_X) & \mathbf{0}_{n_X\times n_Y} \\ \mathbf{0}_{n_Y\times n_X} & \mathbf{0}_{n_Y\times n_Y} \end{bmatrix}KZ\right) \end{aligned} \quad (3)$$

where, $[B_X]_{n_X\times n_X}$ is a diagonal matrix defined as: $B_X(i,i) = \sum_{j=1}^{n_X}A_X(i,j)$. Similarly, we define another measure on a MST build on the target domain Y, which can be derived as:

$$d_{T-MST}^2 = tr\left(Z^T K\begin{bmatrix} \mathbf{0}_{n_Y\times n_Y} & \mathbf{0}_{n_X\times n_Y} \\ \mathbf{0}_{n_Y\times n_X} & 2(B_Y - A_Y) \end{bmatrix}KZ\right) \quad (4)$$

where, $[A_Y]_{n_Y \times n_Y}$ is the adjacency matrix of the MST build on Y and $[B_Y]_{n_Y \times n_Y}$ is the corresponding diagonal matrix.

3.4. Finding landmark points and estimating difference in shapes of two domains

We consider two types of distance measures for detecting the landmark points of data in kernel space - Euclidean (d_E) and manifold (d_M) distances. The manifold distance between two instances is the cost of the shortest path between them in the MST build on the given data [14], as obtained in Sec. 3.3.

To calculate the landmark points, we first find two 'extreme points' in the data. The two nodes in MST which are most far apart from each other are considered as the pair of extreme points. To detect these extreme points, a breadth first search is performed on the MST and the node with maximum distance from the root (any one of the instances can be taken as the root) is considered to be the first 'extreme point'. Next, we perform breadth first search with the first extreme point as the root. The node with the maximum distance from the root is considered to be the second 'extreme point'.

Using this pair of extreme points as the first two landmark points, other landmark points are detected in an iterative manner. Landmark points are detected using an approach similar to that given in [14]. At each step, two adjacent landmark points are taken and the instances lying in between them are considered. The instance which has the maximum difference between the d_E and d_M from its adjacent landmark points, is considered as a new landmark point. A pre-defined odd number of instances, lp, are selected as landmark points.

To make the structure of $\Phi(\tilde{X})$ and $\Phi(\tilde{Y})$ similar, the square of Euclidean distances between the corresponding landmark points in $\Phi(X)$ and $\Phi(Y)$ are minimized, using an approach similar to [15]. The required criterion of Euclidean distance is defined as:

$$\begin{aligned} d_L^2 &= \sum_{i=1}^{lp} (\Phi(\tilde{x}_i) - \Phi(\tilde{y}_i))(\Phi(\tilde{x}_i) - \Phi(\tilde{y}_i))^T \\ &= tr\left(Z^T K \begin{bmatrix} I_4 & -I_6^T I_5 \\ -I_5^T I_6 & I_7 \end{bmatrix} KZ \right) \end{aligned} \quad (5)$$

where, $[I_4]_{n_X \times n_X}$, $[I_5]_{lp \times n_Y}$, $[I_6]_{lp \times n_X}$ and $[I_7]_{n_Y \times n_Y}$ are matrices, defined as:

$$I_4(i,i) = \begin{cases} 1 & \text{if } x_i \text{ is a landmark point of } \Phi(X) \\ 0 & \text{otherwise} \end{cases}$$

$$I_5(i,j) = \begin{cases} 1 & \text{if } y_j \text{ is the } i^{th} \text{ landmark point of } \Phi(Y) \\ 0 & \text{otherwise} \end{cases}$$

$$I_6(i,j) = \begin{cases} 1 & \text{if } x_j \text{ is the } i^{th} \text{ landmark point of } \Phi(X) \\ 0 & \text{otherwise} \end{cases}$$

$$I_7(j,j) = \begin{cases} 1 & \text{if } y_j \text{ is a landmark point of } \Phi(Y) \\ 0 & \text{otherwise} \end{cases}$$

3.5. Estimating optimal sub-space and projected data

We estimate the transformation matrix Z which projects the source and target domains onto a new space satisfying the conditions as described in Secs. 3.1 - 3.4. We estimate Z by minimizing the combined cost function (Eqns. 1 - 5), as:

$$f(Z) = d_\mu^2 + d_{S-MST}^2 + d_{T-MST}^2 + d_L^2 - d_\delta^2 = tr(Z^T P_K Z) \quad (6)$$

where, $P_K = K\tilde{P}K$, and $\tilde{P}$ is defined as:
$\begin{bmatrix} I_1 + I_4 + 2(B_X - A_X) - M & -I_2^T - I_6^T I_5 \\ -I_2 - I_5^T I_6 & I3 + 2(B_Y - A_Y) + I_7 \end{bmatrix}$.
Hence, the desired optimization function can be written as:

$$\underset{Z}{\text{minimize}} \quad tr(Z^T P_K Z) \quad (7)$$

$$\text{subject to} \quad Z^T K Z = I \quad (8)$$

This is a trace minimization problem. Since P_K is symmetric and K is a positive semi-definite matrix, Z is given as [16]:

$$P_K z_i = \alpha_i K z_i \quad (9)$$

where, z_i is the i^{th} column of Z, for a scalar value α_i. Z can be obtained by eigen-analysis of $P_K^{-1}K$, and considering the d eigen-vectors corresponding to the d least eigen-values. If Z_1 and Z_2 are the matrices containing the first n_X rows and the last n_Y rows of Z, we obtain the final transformed domains projected onto W as:

$$\begin{bmatrix} \Phi(\tilde{X}) \\ \Phi(\tilde{Y}) \end{bmatrix} = \Phi(D)W = KZ = \begin{bmatrix} K_{XX}Z_1 + K_{XY}Z_2 \\ K_{XY}^T Z_1 + K_{YY} Z_2 \end{bmatrix} \quad (10)$$

4. EXPERIMENTAL RESULTS

We evaluate our proposed method of unsupervised DA without using class information on a synthetic data and also on two real-world datasets. We have used Gaussian kernel function to build the Gram matrices. Experiments performed using the different methods of DA, are discussed in the following.

Synthetic dataset - Figure 1 (a) shows the instances of the source and target domains in green and blue points respectively, where I and II show two examples of 2-D (500 instances per domain) and 3-D (1075 instances per domain) data respectively. As we are not using class-labels, δ_μ^2 given in Eq. 2 is not used in Eqns. 6 - 8 for estimating Z. Transformed source and target domain instances are shown in Fig. 1 (b). Here, the landmark points are connected by red lines in both the domains. As seen in Fig. 1 (b), the dissimilarities in the structure of the two domains are reduced after transformation (in I & II). Also, the source and target domain instances overlaps in Fig. 1 (b) showing that the disparity in the distributions of data is reduced after transformation. KL-divergence between the two domains before and after transformation for Fig. 1 (I-a) and (I-b) are 12.715 and 0.046, while that for Fig. 1 (II-a) and (II-b) are 115.429 and 1.189

respectively. This illustrates that the proposed method of DA minimizes the disparity in distribution and structure of both the domains, specially when there exists a well-defined structure of data obtained from both the domains.

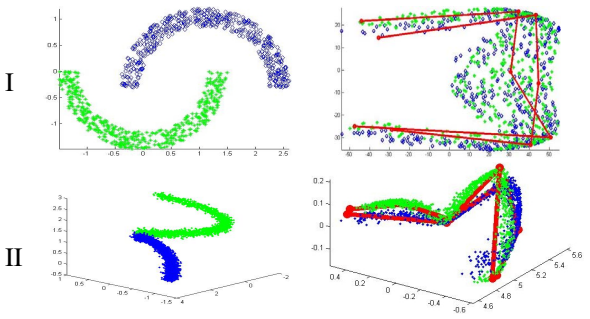

Fig. 1. Source and target domain instances are marked in green and blue points, in (a) as input and (b) after transformation by DA, for two sets of synthetic data: I (2-D) and II (3-D). Red lines in (b) join the detected landmark points.

Object Categorization in images - We evaluate the performance of the proposed method of DA for improving the results of object categorization using Office + Caltech dataset, as in [6]. Here we have four domains: Amazon (A), Caltech (C), Dslr (D) and Webcam (W), and 10 classes of objects in each of the domains. SURF [17] features are extracted from the images and a codebook of size 800 is formed. We follow the same experimental protocols as described in [6], [7] and have considered 7 landmark points in each of the domains. Table 1 shows the classification accuracy for 12 different pairs of source and target domains, using a 10-fold cross validation. We compare our method with TCA [1], GFS [6], GFK [7] and SA [8], while NA denotes the 'No Adaptation', where only the source domain samples are used for training the classifier. From the experimentations, we infer that the proposed method of unsupervised DA gives better result than the state of the art works in majority (9 out of 12) of the cases.

Event categorization in videos - We use 3 video datasets: Kodak [12,18], YouTube [12,18] and CCV dataset [19] as the 3 domains. We consider the YouTube data as the source domain (contains weakly labeled data) collected from Internet, and observed the classification accuracies on Kodak and CCV domains. We consider 6 common classes (events) between YouTube (906 videos) and Kodak (195 videos) as in [12], and 5 classes (events) between YouTube (787 videos) and CCV (2440 videos) as in [20]. For the first case, we use the distance matrices of Kodak and YouTube domains using SIFT and spatio-temporal (ST) features (HOG and HOF) as shared by the authors in [12]; and for the second case, we have considered SIFT and ST features and built bag of words feature using a codebook of size 5000 (as the features given in CCV dataset [19] represents each video by a 5000 dimensional feature). Five landmark points are considered in each of the domains, and five samples per class have been randomly selected from the target domain for training the SVM classifier [21] with Gaussian kernel. We have compared our proposed method of DA with TCA [1]. Figure 2 shows the Mean Average Precision (MAP) for the two cases of event categorization using both SIFT and ST features separately using 10-fold cross validation. Results show that our proposed method of DA gives the best classification accuracy in both cases.

Table 1. Classification accuracy (in %-age) of Office+Caltech dataset [6] using different techniques of DA. Best classification accuracy is highlighted in bold.

Method	C→A	D→A	W→A	A→C	D→C	W→C
NA	21.5	26.9	20.8	22.8	24.8	16.4
TCA [1]	21.96	16.81	13.43	16.18	17.67	11.14
GFS [6]	36.9	32	27.5	35.3	29.4	21.7
GFK [7]	36.9	32.5	31.1	35.6	29.8	27.2
SA [8]	39.0	**38.0**	37.4	35.3	32.4	32.3
Proposed	**54.34**	37.66	**38.47**	**46.17**	**32.60**	**32.60**

Method	A→D	C→D	W→D	A→W	C→W	D→W
NA	22.4	21.7	40.5	23.3	20.0	53.0
TCA [1]	16.69	22.8	32.31	23.60	22.03	44.69
GFS [6]	30.7	32.6	54.3	31.0	30.6	66.0
GFK [7]	35.2	35.2	70.6	34.4	33.7	74.9
SA [8]	37.6	39.6	80.3	38.60	**36.80**	**83.6**
Proposed	**37.93**	**46.49**	**84.62**	**38.82**	34.58	79.67

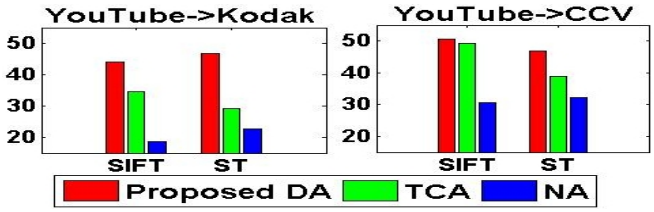

Fig. 2. Mean average precision (MAP) obtained using two sets of features from three real-world datasets. Proposed method of DA (in red) performs better than TCA [1] (in green) and 'No Adaptation' (in blue) techniques.

5. CONCLUSION

We have proposed a new method of unsupervised DA, using the concept of manifold alignment by landmark points. The proposed method reduces the disparity of distribution and the structure of data between source and target domains, when projected in the optimal sub-space. The concept of RKHS enables processing the cases of non-linear transformations in data. The optimal dimension for the subspace can be studied further. The method can improve the performance of cross-domain face recognition and object localization.

6. REFERENCES

[1] Sinno Jialin Pan, I.W. Tsang, J.T. Kwok, and Qiang Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2011.

[2] Arthur Gretton, Alex Smola, Jiayuan Huang, Marcel Schmittfull, Karsten Borgwardt, and Bernhard Schölkopf, "Covariate shift by kernel mean matching," *Dataset shift in machine learning, Chap. 8*, pp. 131–160, 2009.

[3] Chang Wang, *A Geometric Framework For Transfer Learning Using Manifold Alignment*, Ph.D. thesis, University of Massachusetts at Amherst, 2010.

[4] Bogdan Raducanu and Fadi Dornaika, "A supervised non-linear dimensionality reduction approach for manifold learning," *Pattern Recognition*, vol. 45, no. 6, pp. 2432–2444, 2012.

[5] Masashi Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *Journal of Machine Learning Research*, vol. 8, pp. 1027–1061, 2007.

[6] Raghuraman Gopalan, Ruonan Li, and Rama Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *International Conference on Computer Vision*, 2011, pp. 999–1006.

[7] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2066–2073.

[8] Basura Fernando, Amaury Habrard, Marc Sebban, and Tinne Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *International Conference in Computer Vision*, 2013.

[9] Chang Wang and Sridhar Mahadevan, "Heterogeneous domain adaptation using manifold alignment," in *International Joint Conferences on Artificial Intelligence*. 2011, pp. 1541–1546, AAAI Press.

[10] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell, "Adapting visual category models to new domains," in *European Conference on Computer Vision*, 2010, pp. 213–226.

[11] Suranjana Samanta and Sukhendu Das, "Domain adaptation based on eigen-analysis and clustering, for object categorization," in *International Conference on Computer Analysis of Images and Patterns*, 2013, pp. 245–253.

[12] Lixin Duan, Dong Xu, Ivor W. Tsang, and Jiebo Luo, "Visual event recognition in videos by learning from web data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1667–1680, 2012.

[13] Mehrtash Tafazzoli Harandi, Conrad Sanderson, Arnold Wiliem, and Brian C. Lovell, "Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures.," in *IEEE Workshop on Applications of Computer Vision*, 2012, pp. 433–439.

[14] Jun Li and Pengwei Hao, "Finding representative landmarks of data on manifolds," *Pattern Recognition*, vol. 42, no. 11, pp. 2335–2352, 2009.

[15] Deming Zhai, Bo Li, Hong Chang, Shiguang Shan, Xilin Chen, and Wen Gao, "Manifold alignment via corresponding projections," in *British Machine Vision Conference*, 2010, pp. 3.1–3.11.

[16] E. Kokiopoulou, J. Chen, and Y. Saad, "Trace optimization and eigenproblems in dimension reduction methods," *Numerical Linear Algebra with Applications*, vol. 18, pp. 565–602, 2011.

[17] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-up robust features (SURF)," *Computer Vision Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[18] Lixin Duan, Dong Xu, and Shih-Fu Chang, "Exploiting web images for event recognition in consumer videos: A multiple source domain adaptation approach," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1338–1345.

[19] Yu-Gang Jiang, Guangnan Ye, Shih-Fu Chang, Daniel Ellis, and Alexander C. Loui, "Consumer video understanding: A benchmark database and an evaluation of human and machine performance," in *International Conference on Multimedia Retrieval*, 2011, pp. 29:1–29:8.

[20] Lin Chen, Lixin Duan, and Dong Xu, "Event recognition in videos by learning from heterogeneous web sources," in *IEEE conference on computer vision and pattern recognition*, 2013, pp. 2666–2673.

[21] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.