A Pseudo Relevance Feedback Based Cross Domain Video Concept Detection

Xu Shaoxi[1], Yang Jing[2], Tang Sheng[1,*], Zhang Yong-Dong[1]
[1]Institute of Computing Technology, Chinese Academy of Sciences, China
[2]School of Computer and Communication, Hunan University of Technology, China
[1]{xushaoxi, ts, zhyd}@ict.ac.cn, [2]yangjingwp@gmail.com

ABSTRACT
Due to the mismatch of data distribution between training and testing data set, the issue of semantic gap in the field of video concept detection becomes more and more serious. To solve this problem, an effective pseudo relevance feedback (PRF) based method is proposed in this paper to build domain adaptive classifiers. Firstly, the mechanism of PRF tries to select some pseudo samples according to the fused estimation for test data given by existing source models. Then, these pseudo samples are integrated into the process of Tradboost based cross domain transfer learning to make the best use of semantic information generalized by existing source models. Extensive experiments demonstrate that the proposed method can effectively enhance the performance of cross domain learning.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *relevance feedback, retrieval models, search process.*

General Terms
Algorithms, Design, Performance, Experimentation.

Keywords
Pseudo Relevance Feedback, Cross Domain Learning, Video Concept Detection.

1. INTRODUCTION
With the development of processor, network and digital storage, an explosive amount of image/video data have emerged. To effectively manage and utilize these resources, video concept detection becomes a very important research topic in the field of information retrieval [1]. Traditional methods in video concept detection are based on the assumption that the distribution of the training and the test data are identical, such as [2][3]. However, due to the increased multimedia information the existing model built on previous video data can't accommodate to the diversity of new emerging data. The collection of new labeled data for new model-building requires expensive and time-consuming human labor. To this end, cross domain learning emerges as an effective technique to adapt models built in source domains to target domains. In general, a domain refers to data of a certain type, from a certain source, or generated in a certain period of time, etc [4]. In source domain, there are large amount of outdate labeled data and models. Only a limit number of new labeled data exist in target domain.

The issue of semantic gap is always a big obstacle in CBIR [5]. The mismatch between training and testing distribution aggravates this situation. It is widely accepted by the public that relevance feedback can effectively bridge the semantic gap in CBIR [6][7][8]. Nevertheless relevance feedback badly violates the rule of video concept detection that completely no manual intervention should be allowed in test dataset. Pseudo relevance feedback provides an automatic local analysis method to improve retrieval performance without an extended interaction, which automates the manual part of relevance feedback [9]. In most situation, the automatic techniques can obtain good performance [10][11][12]. In this paper, a PRF based method is proposed to build domain adaptive detectors, which aims to alleviate the problem of semantic gap when the distributions of training and test data are different. Firstly, the mechanism of PRF tries to select some pseudo samples according to the fused estimation for test data given by existing source models. The semantic information of source domains is well generalized by these source models. The process of pseudo sample selection can extract part relevant with target domain. Then, these pseudo samples are integrated into the process of Tradboost based cross domain transfer learning to make the best use of semantic information generalized by existing source models.

The rest of this paper is organized as follows. In section 2, we provide a short review on the related work. In section 3, the proposed method is presented. Section 4 demonstrates the experimental setting and result analysis.

2. RELATED WORK
2.1 Cross Domain Video Concept Detection
Cross domain learning is a very promising orientation in the machine learning community. A lot of cross domain methods have been applied in the field of video concept detection. To utilize the existing classifiers in the source domain and the limited labeled samples in the target domain, Jun Yang et al. in [13] try to learn adaptive SVM classifiers which are not "far from" the existing auxiliary classifiers and separate the labeled samples in the target domain well. Following this work, a more general formulation of adaptive SVM in [14] was proposed for function-level classifier adaptation. This framework is based on regularized loss minimization principle which simultaneously measures the classification error of the target classifiers and controls the complexity of the hypothesis space. L. X. Duan [15] et al. proposed Domain Transfer SVM to simultaneously learn a kernel function and a robust SVM classifier by minimizing both the

* Corresponding Author. **Address:** 617H No.6 Kexueyuan South Road Zhongguancun, Haidian District Beijing, China; **Postcode:**100190; **Tel:** +8610-62600616; **Fax:** +8610-62611846.

ICIMCS'11, August 5-7, 2011, Chengdu, Sichuan, China.

structural risk function of SVM and the distribution mismatch of samples between source and target domains. In [16] data-dependent regularizer inspired from Manifold Regularization was used for domain adaptation, which assumes that the target classifier should have similar decision values with the pre-computed auxiliary classifiers when the source and target domain are closely relevant. The cross domain SVM developed in [17] tries to learn a new decision boundary by taking into consideration the classification impact of support vectors derived from source classifiers. There are other adaptation techniques that not using SVM as basic classifiers and also obtaining effective and efficient performance. For example, P. Luo et al. in [18] combined Shannon entropy based consensus measure with logistic regression for sample posteriori distribution prediction to transfer information from multiple source domains to a target domain.

2.2 Pseudo Relevance Feedback

Pseudo relevance feedback is originally applied in the area of text document retrieval. As its effectiveness, this technique has been intensively applied in CBIR. Yan et al suggest in [19, 20] that using lowest ranked image examples for negative pseudo relevance feedback (NPRF) because of their high reliability. In [19], they propose to decompose the bottom negative samples into several partitions and combine all the positive examples in query with each partition as training data to build several classifiers, and finally use a logistic regression to combine the outputs of all the classifiers. Meanwhile, the research proposed in [20] is a more general one and gives an in-depth study to the relationship between retrieval score and their performance criterion. Second, the denotation of bag-of-visual-words (BOW) makes the representation of images similar with that of documents thus some PRF techniques in text retrieval can be reused. J. H. Hsiao et al [12] introduce BOW representation into image retrieval task in which the KL-divergence language modeling-based retrieval with dirichlet smoothing become feasible as it is used in [21]. The PRF in [12] is an unsupervised learning process using a linear interpolation between query and feedback model. The probabilistic model of feedback is a summation over language models of the top-ranked retrieval images.

3. THE PROPOSED ALGORITHM

3.1 The Framework of Algorithm

The framework of the proposed algorithm is illustrated in Figure 1. The flow of this framework is summarized as follows:

(1) Using existing model in source domain to predict test samples in target domain and obtaining the initial result list. The aim of this step is to make the semantic information of source model be assigned on the samples in target domain.

(2) Utilizing PRF to extract relevant semantic information embedded in the initial result list and obtaining pseudo samples relevant with the designated concept. Due to the mismatch of distribution between source and target domain and the obstacle of semantic gap, the top-retrieved samples in initial result list contain noise. In other word, the semantic information assigned by the source model is not correct. Digging out relevant semantic information from initial result list as much as possible is a critical issue for PRF in video concept detection. The mechanism of PRF could distill the semantic information and extract relevant part.

(3) Combining the pseudo samples with labeled samples both in source and target domain as new training dataset and based on Tradaboost [22] method training domain adaptive model for target domain.

(4) Using the domain adaptive model obtained in step(3) to predict test samples in target domain and repeating step(2)(3)(4) until the process converges or the performance reduces sharply. Due to computational issues, this process iterates only once. The feedback process can make the best of semantic information generalized by the source model.

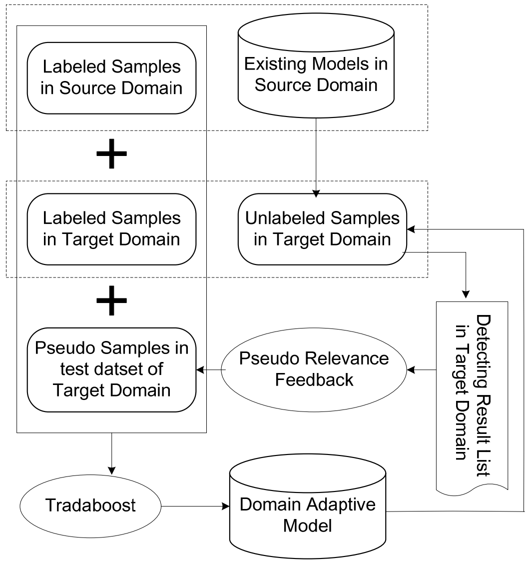

Figure 1. The framework of the proposed algorithm

3.2 Notations

The detection unit is based on shot-level. Each shot is represented by one or two key-frames. Suppose that for each key-frame, p categories of image features are extracted. In source domain, for each concept, the samples are represented as $D_S^j = \{(x_1^{Sj}, y_1^S), \ldots, (x_{ns}^{Sj}, y_{ns}^S)\}$ $(j = 1, \ldots, p, ns$ is the number of samples in source domain). $x_i^{Sj} \in R^n$ is the image feature j of sample $x_i^S (j = 1, \ldots, p)$; $y_i^S \in \{0, 1\}$ is the corresponding label of sample x_i^S $(i = 1, \ldots, ns)$. The existing source models corresponding to p categories of image features are denoted by $M^{S1}, \ldots, M^{Sp}$. In target domain, there are a limited number of labeled samples and plenty of unlabeled test samples. For each concept, the labeled samples in target domain are denoted by $D_T^{lj} = \{(x_1^{lj}, y_1^l), \ldots, (x_{nl}^{lj}, y_{nl}^l)\}$ $(j = 1, \ldots, p, nl$ is the number of labeled sample in target domain). $x_i^{lj} \in R^n$ is the image feature j of sample $x_i^l (j = 1, \ldots, p)$; $y_i^l \in \{0, 1\}$ is the corresponding label of sample x_i^l $(i = 1, \ldots, nl)$. The unlabeled samples in target domain are denoted by $D_T^{uj} = \{x_1^{uj}, \ldots, x_{nu}^{uj}\}$ $(j = 1, \ldots, p,$ nu is the number of unlabeled samples in target domain, $nl \ll nu$). $x_i^{uj} \in R^n$ represents the image feature j of sample $x_i^u (i = 1, \ldots, nu, j = 1, \ldots, p)$.

3.3 Initial Detection by Source Models

Firstly we use the existing models $M^{S1}, \ldots, M^{Sp}$ in source domain to respectively predict test samples $D_T^{uj} = \{x_1^{uj}, \ldots, x_{nu}^{uj}\}$ $(j = 1, \ldots, p)$ in target domain. The prediction results are as follows:

$$\left\{\left(x_1^{uj}, f^{sj}\left(x_1^{uj}\right)\right), \ldots, \left(x_{nu}^{uj}, f^{sj}\left(x_{nu}^{uj}\right)\right)\right\} (j = 1, \ldots, p) \quad (1)$$

$f^{sj}\left(x_i^{uj}\right)$ $(i = 1, \ldots, nu)$ is the probability score of key-frame containing the designated concept. To a certain extent, $f^{sj}\left(x_i^{uj}\right)$

$(i = 1, \ldots, nu)$ includes some semantic information generalized by the source models. Re-ordering the test samples according to the probability scores, the following formulation can be obtained:

$$\left\{\left(x_{(1)}^{uj}, f^{sj}\left(x_{(1)}^{uj}\right)\right), \ldots, \left(x_{(nu)}^{uj}, f^{sj}\left(x_{(nu)}^{uj}\right)\right)\right\} (j = 1, \ldots, p) \quad (2)$$

where $f^{sj}\left(x_{(1)}^{uj}\right) \geq f^{sj}\left(x_{(2)}^{uj}\right), \ldots, \geq f^{sj}\left(x_{(nu)}^{uj}\right)$.

3.4 Semantic Information Extraction by PRF

As the initial detecting results contain noise and are not as ideal as we expect, the mechanism of PRF is utilized to filter out semantic information irrelevant with target domain. The detailed strategy is that for each unlabeled sample in target domain, we calculate:

$$Q(x_i^u) = \sum_{j=1}^{p} w_j * f^{sj}(x_i^{uj}) \quad (3)$$

where w_j is the detection precision of source model $M^{Sj} (j = 1, \ldots, p)$ on the test set in target domain. As the probability scores $f^{s1}(x_i^{u1}), \ldots, f^{sp}(x_i^{up})$ are predicted by different source models, the values might converge in different intervals in [0, 1], which results in unequal processing for the *p* source models. Min-max normalization is utilized to normalize result lists corresponding to different image features and then the normalized probability scores for sample x_i^u are denoted by $f^{s1}(x_i^{u1})', f^{s2}(x_i^{u2})', \ldots, f^{sp}(x_i^{up})'$. Based on these normalized scores, we calculate:

$$Q'(x_i^u) = \sum_{j=1}^{p} w_j * f^{sj}(x_i^{uj})' \quad (4)$$

Re-ordering the test samples according to $Q'(x_i^u)$, the following re-ranking list can be obtained:

$$\left\{\left(x_{(1)}^u, Q'(x_{(1)}^u)\right), \ldots, \left(x_{(nu)}^u, Q'(x_{(nu)}^u)\right)\right\} \quad (5)$$

where $Q(x_{(1)}^u) \geq Q(x_{(2)}^u), \ldots, \geq Q(x_{(nu)}^u)$. The top-ranked *m* samples of the re-ranking list are recognized as positive pseudo samples for feedback; the *m* samples in the bottom of re-ranking list are recognized as negative pseudo samples for feedback.

3.5 Tradaboost based Cross Domain Learning

Adaboost is a popular boosting algorithm which aims to boost the accuracy of weak learners by carefully adjusting the weights of training samples and finally learn a strong classifier. Tradaboost is an extension of Adaboost, which deals with the situation when the distributions of training and test data are different and only related to each other. The weak learners are built based on samples both in source and target domains. Different weighting strategies are adopted for samples respectively in source and target domain. For training samples in target domain, Adaboost is still applied to build the base of the model. That is, the misclassified samples will be focused more and their weights are increased. However, for training samples in source domain, they are wrongly predicted due to distribution changes by the learned model, these samples could be those that are the most dissimilar to the samples in target domain [20]. Therefore, the weights of misclassified samples in source domain need to be decreased to weaken their impacts in the learning process.

Although Tradaboost extends boosting-based learning algorithm to effectively solve the problem of distribution mismatch between training and test dataset, it didn't make full use of the relevant resource in source domain for building adaptive model in target domain, that is, it typically ignored to utilize the existing models in source domain. Actually the existing source models are a kind of very important resource because these models have generalized the characteristic of samples in source domain, which can play a crucial role in the learning process. The proposed method utilizes both the existing labeled samples and models to extract relevant semantic information to assist the building of domain adaptive model, which is an extension of Tradaboost. The detailed strategy is that when the pseudo samples including semantic information assigned by the source models are extracted by utilizing PRF, they are combined with labeled samples in source domain for training Tradaboost based model. Since these pseudo samples come from target domain, their weighting strategy is the same with training samples in target domain.

The algorithm description of the proposed PRF based cross domain learning method is presented in Figure 2:

The PRF based cross domain learning method

Input:

1. labeled samples in source domain D_S^j; labeled samples in target domain D_T^{lj}; pseudo samples $D_T^{(prf)j}$ $(j = 1, \ldots, p)$. The combined training set is defined as follows:

$$D^j = \{(x_1^j, y_1), \ldots, (x_{ns+nl+2m}^j, y_{ns+nl+2m})\}$$

$$x_i^j \in R^n, y_i \in \{0,1\}, \; i = 1, \ldots, ns + nl + 2m,$$

$$x_i^j = \begin{cases} x_i^{Sj} & i = 1, \ldots, ns \\ x_i^{lj} & i = ns + 1, \ldots, ns + nl \\ x_i^{(prf)j} & i = ns + nl + 1, \ldots, ns + nl + 2m \end{cases}$$

x_i^j is the image feature *j* of sample x_i; y_i is the corresponding label of sample x_i; $x_i^{(prf)j}$ is the image feature *j* of pseudo sample $x_i^{(prf)}$.

2. unlabeled sample set in target domain D_T^{uj}.

3. a base learning classifier *C*.

4. the maximum number of iterations *N*.

Initialization: initial weight vector $\mathbf{w}^{1j} = (w_1^{1j}, \ldots, w_{ns+nl+2m}^{1j})$, in general, the weight value for each $w_i^{1j} (i = 1, \ldots, ns + nl + 2m)$ is the same.

For $t = 1, \ldots, N$

1. Set the distribution of training samples as:

$$\mathbf{p}^{tj} = \mathbf{w}^{tj} \Big/ \sum_{i=1}^{ns+nl+2m} w_i^{tj}$$

2. On the training dataset D^j, build classifier C^{tj} with distribution $\mathbf{p}^{tj}$; for each sample $x \in D^j$, classifier C^{tj} returns a probability value $f^{tj}(x) \in [0,1]$ indicating how likely the sample x to be positive (containing the designated concept).

3. On the labeled dataset in target domain, calculating the error of classifier C^{tj}:

$$\epsilon_t^j = \sum_{i=ns+1}^{ns+nl} \frac{w_i^{tj} \cdot |f^{tj}(x_i) - y_i|}{\sum_{i=ns+1}^{ns+nl} w_i^{tj}}$$

4. **if** $\epsilon_t^j \geq 1/2$ **then** adjusting weight vector $\mathbf{w}^{tj}$, reture to step 1.

5. Set $\beta_t^j = \epsilon_t^j/(1-\epsilon_t^j)$ and $\beta^j = 1/(1+\sqrt{2\ln(ns)/N})$.

6. Update the new weight vector：

$$w_i^{(t+1)j} = \begin{cases} w_i^{tj} \cdot (\beta^j)^{|f^{tj}(x_i)-y_i|} & i = 1,\dots,ns \\ w_i^{tj} \cdot (\beta_t^j)^{-|f^{tj}(x_i)-y_i|} & i = ns+1,\dots,ns+nl \\ w_i^{tj} \cdot (\beta_t^j)^{-|f^{tj}(x_i)-y_i|} & i = ns+nl+1,\dots,ns+nl+2m \end{cases}$$

Output： $f(x) = \begin{cases} 1, & \sum_{j=1}^{p}(1-\epsilon_N^j) * f^{Nj}(x) \geq 1/2 \\ 0, & \text{otherwise} \end{cases}$

Figure 2. Description of our proposed algorithm

4. EXPERIMENTS AND OBSERVATIONS

4.1 Experimental Setup

To validate the effectiveness of our proposed assumption, the video data collection provided by TRECVID is detected. We use the development set of TRECIVD2005 as the source data set. The development set and the GroundTruth test set for TRECVID2007 is utilized as the target data set. The GroundTruth test set is recognized as unlabeled dataset in target domain in our experiments. The strategy to extract labeled samples in target domain is that 5% samples in development set of TRECVID2007 and 2% labeled samples in GroundTruth set are randomly selected. The concept collection is provided by TRECVID2007 including 20 concepts. For each key-frame, three categories of image features are extracted, that is: (1) grid-based (5 by 5) color moment in Lab space (CM, 225D); (2) SIFT Descriptors-based Visual Keywords(SIFT, 500D); (3) grid-based (3 by 3) wavelet texture (WT, 81D). The features (1)(2)(3) and their corresponding pre-trained classifiers from source domain are provided by VIREO374[23]. The base classifier *C* is weighting SVM provided by LibSVM[24]. The detecting performance is evaluated with the inferred average precision (Inf AP) [25].

We setup two experimental settings to respectively compare the proposed method (denoted by "CDL") with Tradaboost (denoted by "TRA") and two baselines. The iteration number for both Tradaboost and our method is set as $N = 3$. The two baselines are: (1) Using existing source models to predict test samples in target domain(that is using models of TRECVID2005 to predict GroundTruth set of TRECVID2007, which is denoted by "05to07"); (2) Using models built on the labeled data in target domain to predict test samples in target domain(denoted by "label").

4.2 Experimental Results and Conclusions

The performance of Tradaboost and the proposed method is presented in Table 1. The performance is evaluated based on the average of Inf AP values (Mean Inf AP) for 20 concepts. The percentages in parentheses denote the percent by which the proposed method outperforms Tradaboost. We can see from Table 1 that the settings of CDL have achieved good performance for all the three image features compared with Tradaboost. We can conclude that the introduction of PRF for building domain adaptive classifiers can effectively promote the performance of cross domain learning.

Table 1: the Mean Inf AP of 20 concepts for TRA and CDL

	CM	SIFT	WT
TRA	0.0577	0.0808	0.0390
CDL	0.0901(56.15%)	0.0914(13.12%)	0.0599(53.59%)

The Mean Inf APs of two baselines and the proposed method are presented in Table 2. The percentages in parentheses denote the percent by which the proposed method outperforms the second baseline "label". We can see from Table 2 that compared with the baseline "label", the settings of CDL have achieved good performance for image feature CM and WT, while for image feature SIFT the performance of CDL has reduced by 14.82%. Through serious analysis, we found that the reason for this performance reduction is that the detecting performance by model built on the labeled samples in target domain for image feature SIFT has already been high (that is 0.1073). Based on this precondition, cross domain learning will lead to the effect of "Negative Transfer", which means the source domain data contribute to the reduced performance of learning in target domain[26].

Table 2: Mean Inf AP of 20 concepts for baselines and CDL

	CM	SIFT	WT
05 to 07	0.0462	0.0855	0.0387
label	0.0785	0.1073	0.0531
CDL	0.0901 (14.78%)	0.0914 (-14.82%)	0.0599 (12.81%)

In conclusion, the introduction of PRF into the process of cross domain video concept detection can effectively improve the detecting performance, which proves that the proposed method can alleviate the problem of semantic gap when the distributions of training and test data are different but only related. In our future work, suitable measures to judge transferability between source and target domains need to be proposed to avoid negative transfer. We also will focus on the research of more intelligent PRF strategies [27] for some specific target domain.

5. ACKNOWLEDGMENTS

This work was supported by National Basic Research Program of China (973 Program, 2007CB311105); National Nature Science Foundation of China (60873165); Co-building Program of Beijing Municipal Education Commission.

6. REFERENCES

[1] Cees, G. M. S. and Marcel, W. 2008. Concept-based video retrieval. Journal of *Foundations and Trends in Information Retrieval* 2(Apr. 2008), 215-322.

[2] Meng, W., Xian-Sheng, H., Jinhui, T. and Richang, H. 2009. Beyond distance measurement: constructing neighborhood similarity for video annotation. Journal of *Transactions on Multimedia-Special section on communities and media computing* 11(Apr. 2009), 465-476.

[3] Meng, W., Xian-Sheng, H., Richang, H., Jinhui, T., and Guo-Jun, Q. 2009. Unified video annotation via multigraph learning. Journal of *IEEE Transactions on Circuits and Systems for Video Tec*hnology 19 (May, 2009), 733-746.

[4] Jun, Y. and Alexander, G. H. 2008. A framework for classification adaptation and its application in concept detection. In *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval* (Vancouver, Canada, October 30-31, 2008).MIR'08. ACM, New York, NY, 467-474. DOI = http://doi.acm.org/10.1145/1460096.1460171.

[5] Arnold, W.M. S., Marecl, W., Simone, S., Amarnath, G. and Ramesh J. 2000. Content-based image retrieval at the end of the early years. Journal of *IEEE Transaction on Pattern Analysis and Machine Intelligence* 22(Dec. 2000),1349-1380.

[6] Nastar, C., Mitschke, M. and Meilhac, C. 1998. Efficient query refinement for image retrieval. In *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Santa Barbara, USA, June 23-25, 1998). CVPR'98. IEEE Computer Society Press, Washington DC, 547-552.DOI=http://dx.doi.org/10.1109/CVPR.1998. 698659.

[7] Kinh, T. and Paul, V. 2004. Boosting image retrieval. *International Journal of Computer Vision: Special Issue on Content-Based Image Retrieval* 56(Jan.-Feb. 2004), 17-36.

[8] Yong, R., Thomas, S. H., Michael, O. and Sharad, M.1998. Relevance feedback: a power tool in interactive content based image retrieval. Journal of *IEEE Transaction on Circuits Systems Video Technol: Special Issue on Interactive Multimedia Systems for the Internet* 8(May 1998), 644-655.

[9] Christopher, D. M., Prabhakar, R. and Hinrich, S. 2008. Relevance feedback and query expansion. In *Introduction to Information Retrieval*. New York: Cambridge University Press, 177-194.

[10] Rong, Y., Alexander, G. H. and Rong, J. 2003. Multimedia search with pseudo-relevance feedback. In *Proceedings of the 2nd International Conference on Image and Video Retrieval* (Urbana-Champain,USA,July 24-25, 2003). CIVR'03. Heidelberg: Springer-Verlag Berlin ©2003, 238-247.

[11] Rong, Y., Alexander, G. H., and Rong, J. 2003. Negative pseudo-relevance feedback in content-based video retrieval. In *Proceedings of the 11th ACM International Conference on Multimedia* (Berkeley, USA, November 2-8, 2003). MM'03. ACM, New York, NY, 343-346.

DOI = http://dx.doi.org/10.1145/957013.957087.

[12] Jen, H. H., Chu, S. C., and Ming, S. C. 2008. Visual-word-based duplicate image search with pseudo-relevance feedback. In *Proceedings of IEEE International Conference on Multi-media & Expo* (Piscataway, USA, June 23-26, 2008). ICME'08. IEEE Press Piscataway, NJ, 699-672.

DOI = http://dx.doi.org/10.1109/ICME.2008.4607523.

[13] Jun, Y., Rong Y., and Alexander, G. H. 2007. Cross-domain video concept detection using adaptive SVMs. In *Proceedings of the 15th ACM International Conference on Multimedia* (Augsburg, Germany, September 23-28, 2007). MM '07. ACM, New York, NY, 188-197.

DOI = http://dx.doi.org/10.1145/1291233.1291276.

[14] Jun, Y., and Alexander, G. H. 2008. A framework for classification adaptation and its application in concept detection. In *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval* (Vancouver, Canada, October 30-31, 2008). MIR'08. ACM, New York, NY, 467-474.

DOI = http://dx.doi.org/10.1145/1460096.1460171.

[15] Duan, L., Tsang, I. W., and Xu, D. 2009. Domain transfer SVM for video concept detection. In *Proceedings of the 21st IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Miami Beach, USA, June 20-25, 2009). CVPR'09. IEEE Computer Society Press, Washington DC, 1375-1381. DOI = http://doi.ieeecomputersociety.org/ 10.1109/CVPRW.2009.5206747.

[16] Duan, L., Tsang, I. W., Xu, D, and Chua, T. S. Domain adaptation from multiple sources via auxiliary classifiers. In *Proceedings of the 26th International Conference on Machine Learning* (Montreal, Canada, June 14-18, 2009). ICML'09. ACM, New York, NY, 289-29.

DOI = http://dx.doi.org/10.1145/1553374.1553411.

[17] Jiang, W., Zavesky, E., Chuang, S. F., and Loui, A. 2008. Cross-domain learning methods for high level visual concept classification. In *Proceedings of the 15th IEEE International Conferece on Image Processing* (San Diego, USA, October 12-15, 2008). ICIP'08. IEEE Computer Society Press, Washington DC, 161-164.

DOI = http://dx.doi.org/10.1109/ICIP.2008.4711716.

[18] Luo, P., Zhuang, F. Z. Xiong, H., Xiong, Y. H., and He, Q. 2008. Transfer learning from multiple source domains via consensus regularization. In *Proceedings of the 17th ACM Conference on Information and Knoledge Management* (Napa Valley, USA, October 26-30, 2008). CIKM'08. pp.103-112.

DOI = http://dx.doi.org/10.1145/1458082.1458099.

[19] R. Yan, A. G.. Hauptmann, and R. Jin, "Multimedia Search with Pseudo-relevance Feedback", *in Proc. of Intl. Conf. on Image and Video Retrieval*, pp. 238-247, 2003.

[20] R. Yan, A. G. Hauptamn, and R. Jin, "Negative Pseudo-Relevance Feedback in Content-based Video Retrieval", *in Proc. of ACM intl. conf. on Multimedia*, pp. 343-346, 2003.

[21] Cao, G., Nie, J. Y. Gao, J. and Robertson, S. 2008. Selecting good expansion terms for pseudo-relevance feedback. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, July 20-24, 2008). SIGIR'08. ACM, New York, NY, 243-250.

DOI = http://dx.doi.org/10.1145/1390334.1390377.

[22] Dai, W., Yang, Q., Xue, G., Yu Y. 2007. Boosting for transfer learning. In *Proceedings of the 24th International Conference on Machine Learning* (Corvallis, USA, June 20-24, 2007). ICML'07, ACM, New York, NY, 193–200.

DOI = http://dx.doi.org/10.1145/1273496.1273521.

[23] Jiang, Y. G., Yang, J., Ngo, C. W., and Alexander, G. H. 2010. Representations of keypoint-based semantic concept detection: a comprehensive study. Journal of *IEEE Transactions on Multimedia* 12 (Jan. 2010), 42-53. Vireo374 Available on: http://www.cs.cityu.edu.hk/~yjiang/vireo374/.

[24] Chang, C. C., and Lin, C. J. 2001. LIBSVM: a library for support vector machines. Software available at :

http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[25] Aslam, J. A., Pavlu, V., and Yilmaz, E.2006. Statistical method for system evaluation using incomplete judgments. In *Proceedings of the 29st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Seattle, USA, August 6-11, 2006). SIGIR'06. ACM, New York, NY, 541-548.

DOI = http://dx.doi.org/10.1145/1148170.1148263.

[26] Pan, S. J., and Yang, Q. 2010. 2010. A survey on transfer learning. Journal of *IEEE Transactions on Knowledge and Data Engineering* 22(Oct., 2010), 1345-1359.

[27] Meng, W. and Xian-Sheng, H. 2011. Active learning in multimedia annotation and retrieval: A survey. Journal of ACM Transactions on Intelligent Systems and Technology 2(Feb.,2011).