ISSN:2320-0790

An International Journal of Advanced Computer Technology

Efficient Video Annotations by an Image Groups

K .Mahi balan[1], J. Sathya kumar[2], K .Rajakumari.[3]

[1,2]UG Student, Department of CSE, Bharath University, Tamilnadu, India.

[3]Asst.Professor, Department of CSE, Bharath University, Tamilnadu, India.

Abstract: Searching desirable events in uncontrolled videos is a challenging task. So, researches mainly focus on obtaining concepts from numerous labelled videos. But it is time consuming and labour expensive to collect a large amount of required labelled videos for training event models under various condition. To avoid this problem, we propose to leverage abundant Web images for videos since Web images contain a rich source of information with many events roughly annotated and taken under various conditions. However, information from the Web is difficult .so,brute force knowledge transfer of images may hurt the video annotation performance. so, we propose a novel Group-based Domain Adaptation learning framework to leverage different groups of knowledge (source target) queried from the Web image search engine to consumer videos (domain target). Different from old methods using multiple source domains of images, our method makes the Web images according to their intrinsic semantic relationships instead of source. Specifically, two different types of groups (event-specific groups and concept-specific groups) are exploited to respectively describe the event-level and concept-level semantic meanings of target-domain videos.

Keywords: Video annotation, Domain adaptation, A-KML, ASVM.

I. INTRODUCTION

Digital cameras and mobile phone cameras have become popular in our daily life. The ever expanding video collections have motivated a real necessity to provide effective tools to support video annotation and retrieval.

However, video annotation still remains a challenging problem due to the highly cluttered background, large intra-class variations and significant camera motions In this paper, we focus on the event annotation of real-world unconstraint consumer videos, which have long-term spatially and temporally dynamic object interactions that happen under certain scene settings Recently, a number of previous methods have been proposed to effectively analyze events in videos These works require labeled training videos to learn robust classifiers and can achieve promising results with sufficient labeled training data. However, the labeling process is time consuming and labor expensive that users are generally reluctant to annotate abundant videos. Since it is difficult to acquire enough knowledge from labeled videos, many researchers have tried to seek another source of labeled data and transfer the related knowledge from these data to videos. Fortunately, Web image searching engines have become increasingly mature and can offer abundant and easily accessible knowledge. Moreover, the image datasets from the Web are more diverse and less biased than homegrown datasets, which makes them more realistic for real world tasks. Recently, several methods are proposed to address the problem of knowledge transformation across the image domain and the video domain. In Web images are incrementally collected to learn classifiers for action recognition in videos. Wang et al. proposed to obtain knowledge for consumer videos from both labeled Web images and a small amount of labeled videos. Duan et al. developed a multi-domain adaptation scheme by leveraging Web images from different sources .The main motivation behind their methods is that the keyword based search can be readily used to collect a large number of relevant Web images without human annotation.

Though it is beneficial to learn from Web knowledge, noisy images of little relevance with consumer videos still exist due to random noting and subjective understanding. Under this circumstance, brute force transferring may degrade the performance of classifiers for videos, which is known as negative transfer. Therefore, it is necessary to effectively summarize Web knowledge and transfer the most

relevant pieces. One strategy to decrease the risk of negative transfer is assigning different weights to different source domains based on their relevance's to the target domain. Recently, several domain adaptation methods were proposed to learn robust classifiers with diverse training data from multiple source domains. Luo et al proposed to maximize the consensus of predictions from multiple sources. Duan et al. developed a multi domain adaptation scheme by leveraging web images from different source domains. In their work, weights are assigned to the images according to their sources, ignoring the intrinsic semantic meaning among the source-domain data. We observe that it is more beneficial to measure the relevance's between Web images and consumer videos according to their semantic meanings instead of their sources. In this paper, we propose to leverage Web images organized by groups, and each group of images stands for one event-related concept. Specifically, we manually define several concept-level query keywords to construct multiple groups in which the images of the same group have similar concepts. We refer this kind of group as concept-specific group. In addition, we propose another kind of groups called event-specific groups to represent events with more descriptive and discriminative

II. RELATED WORK

A. Video Annotation

In recent decades, event annotation in consumer videos has become a challenging problems due to multiple concepts and their complex interactions underlying videos. Several approaches have been proposed to deal with the problem of detecting multiple concepts and modeling the relations between concepts, such as human-object interaction visual context for object and scene recognition scene and action combination, and object, person, and activity relations.

These methods followed the conventional learning framework by assuming that the training and testing samples have the same feature distribution from the same domain. In contrast, our work focuses on annotating consumer videos by leveraging a large amount of loosely labeled Web images, in which the training and testing data come from different Domains having different data distributions.

B. Domain Adaptation for Video Annotation

Domain adaptation (cross-domain learning or transfer learning) methods have been employed over a wide variety of applications, such as sign language recognition text classification and WiFi localization Roughly speaking, there are two settings of domain adaptation: unsupervised domain adaptation where the target domain is completely unlabeled, and semi-supervised domain adaptation where the target domain contains a small amount of labeled data Since the labeled data alone is insufficient to construct well generalized target classifier, a very fruitful line of work has been focusing on effectively using unlabeled target-domain data. In Bruzzone proposed a Domain Adaptation Support Vector Machine (DASVM) to iteratively learn the target classifier by labeling the unlabeled samples in the source domain. Gopalan and Gong used both labeled source domain data and unlabeled target-domain data to infer new subspaces for domain adaptation. Saenko proposed a metric learning method to make the intra-class samples from two domains become closer to each other.Our method belongs to the unsupervised domain adaptation methods , in which the training data consists of a large number of labeled Web images and a few of unlabeled consumer videos..Recently, applying domain adaptation to multimedia content analysis has attracted more and more attentions of researchers Yang et al. proposed an Adaptive Support Vector Machine (A-SVM) method to learn a new SVM classifier for the target domain, which is adapted from a pre-trained classifier from a source domain. Duan et al. proposed to simultaneously learn the optimal linear combination of base kernels and the target classifier by minimizing a regularized structural risk function. And then, they proposed A-MKL to add the pre-learned classifiers as the prior.

Their methods mainly focus on the single source domain setting. To utilize numerous labeled image data in the Web, multiple source domains adaptation methods are proposed to leverage different pre-computed classifiers learned from multiple source domains. In these methods, different weights are assigned to different source domains without taking account of intrinsic semantic relations between source domains. In this paper, we leverage different groups of images queried by different associational keywords to the Web. We insure that the samples in each group are of the same concept, and also ensure that different groups within the same event class are correlated to each other.

III. EXISTING SYSTEM

In existing, a visual event recognition framework for consumer videos by leveraging a large amount of loosely labeled web videos. Observing that consumer videos generally contain large intra-class variations within the same type of events, to measure the distance between any two video clips using a Aligned

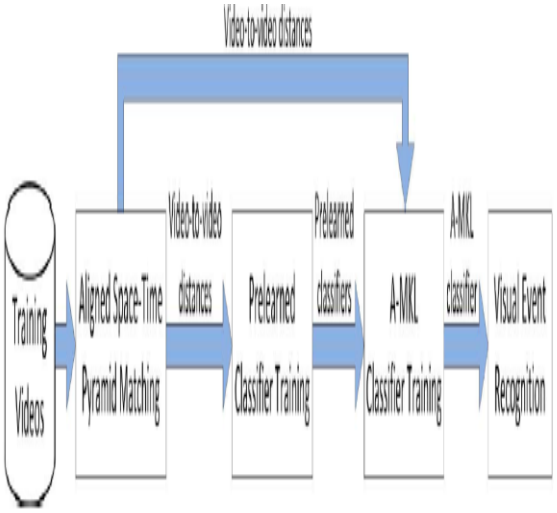
Space-Time Pyramid Matching (ASTPM) and transfer learning method, called as Adaptive Multiple Kernel Learning (A-MKL). A set of SVM classifiers based on the combined training set from two domains by using multiple base kernels from different kernel types and parameters, which are then combined with equal weights to obtain a relearned average classifier. In Adaptive Multiple Kernel Learning, for each event we learn an adapted target classifier based on multiple base kernels and the pre-learned average classifiers from this event class .

Brute Force Knowledge Transfer

Though it is beneficial to learn from Web knowledge, noisy images of little relevance with consumer videos still exist due to random noting and subjective understanding. Under this circumstance, brute force transferring may degrade the performance of classifiers for videos, which is known as negative transfer. Therefore, it is necessary to effectively summarize Web knowledge and transfer the most relevant pieces. One strategy to decrease the risk of negative transfer is assigning different weights to different source domains based on their relevance's to the target domain.

Web images contain a rich source of information with many events roughly annotated and taken under various conditions. In this brute force transferring may degrade the performance of classifiers for videos, which is known as negative transfer.

Figure: 1 AMKL- Adaptive Multiple Kernel Learning



IV. PROPOSED SYSTEM

In this system require labeled training videos to learn robust classifiers and can achieve promising results with sufficient labeled training data. Web image searching engines have become increasingly mature and can offer abundant and easily accessible knowledge. Here the knowledge transformation across the image domain and the video domain. To obtain knowledge for consumer videos from both labeled Web images and a small amount of labeled videos. The main motivation behind their methods is that the keyword based search can be readily used to collect a large number of relevant web images without human annotation. In several domain adaptation methods were proposed to learn robust classifiers with diverse training data from multiple source domains. To leverage web images organized by groups, and each group of images stands for one event-related concept. Specifically, we manually define several concept level query keywords to construct multiple groups in which the images of the same group have similar concepts.

Group-based Domain Adaptation (GDA)

This technique is used for consumer videos by leveraging a large amount of loosely labeled Web images. A large amount of loosely labeled Web images can be readily obtained by

V. ALGORITHM

Step 1: Initialize G concept-specific group classifiers {w 0 } Ss=1 using standard SVM;
Step 2: Initialize E event-specific group classifiers {w (0)S} Ss=s+1 randomly;
Step 3: Compute S (0)S
Step 4: Set m=0;
Step 5: repeat.
Step 6: Return Ws.

VI. CONCLUSION

We have presented a new framework, referred to GDA ,for annotating consumer videos by leveraging a large amount of loosely labeled Web images. Specifically, we exploited concept-level and event-level images to learn concept-specific and event-specific group representation of source-domain Web images. The group classifiers and weights are jointly learned in a unified optimization algorithm to build the target-domain classifiers. In addition, we introduced two new data-dependent regularizers based on the unlabeled target-domain consumer videos for enhancing the generalization of the target classifier. Experimental results clearly demonstrate the effectiveness of our framework. This is the first attempt in transfer learning to weight data according to their semantic meaning instead of their sources. A

possible future research direction is to develop a discriminative common feature space between Web images and consumer videos as well as investigate several criteria to deal with the data distribution mismatch between source and target domains. We are also going to apply our proposed method to other cross domain applications like text-video domain adaptation.

VII. REFERENCES

[1] Y. Jiang, G. Ye, S. Chang, D. Ellis, and A. Louie, "Consumer video understanding: A benchmark database and an evaluation of human and machine performance," in ICMR, 2011, p. 29.

[2] M. R. Nap hade and J. R. Smith, "On the detection of semantic concepts at trecvid," in Proceedings of the 12th annual ACM international conference on Multimedia, 2004, pp. 660–667.

[3] A. Louie, J. Lou, S. Chang, D. Ellis, W. Jiang, L. Kennedy, K. Lee, and A. Yanagawa, "Kodak's consumer video benchmark data set: concept definition and annotation," in Workshop on Multimedia Information Retrieval, 2007, pp. 245–254.

[4] X. Wu, X. Dong, D. Lixin, L. Jiebo, and J. Yunde, "Action recognition using multi-level features and latent structural svm," vol. 23, no. 8, pp.1422–1431, 2013.

[5] Y.-G. Jiang, S. Bhattacharya, S.-F. Chang, and M. Shah, "High-level event recognition in unconstrained videos," International Journal of Multimedia Information Retrieval, pp. 1–29, 2012.

[6] L. Duan, D. Xu, I. Tsang, and J. Lou, "Visual event recognition in videos by learning from web data," in CVPR, 2010, pp. 1959–1966.

[7] Z. Ma, A. G. Hauptmann, Y. Yang, and N. Sebe, "Classifier-specific intermediate representation for multimedia tasks," in Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, 2012,p. 50.

[8] K. Rajakumari, Dr. C. Nalini, National Conference on Advance Trends in Information Computing (NCATIC' 14) "Face Recognition From Sequence Videos in Pose, Illumination Invarient" on 20th August 2014.

[9] N. Ikizler-Cinbis and S. Sclaroff, "Object, scene and actions: Combining multiple features for human action recognition," in ECCV, 2010, pp.494–507.

[10] N. Ikizler-Cinbis, R. Cinbis, and S. Sclaroff, "Learning actions from the web," in CVPR, 2009, pp. 995–1002.

[11] X. W. Han Wang and Y. Jia, "Annotating video events from the web images," in ICPR, 2012.

[12] K. Rajakumari,C. Nalini, "Improvement of Image Quality Based on Fractal Image Compression" in Middle-East Journal of Scientific Research 20 (10): 1213-1217, 2014, ISSN 1990-9233, © IDOSI Publications, 2014.