iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** II **Month of publication:** February 2018

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue II, February 2018- Available at www.ijraset.com

A Survey on Semi-Supervised Image to Video Adaptation for Video Action Recognition

Dr.M.Preetha[1], K.Monica[2], H.Nivetha[3], S.Priyanka[4]
[1]*Associate Professor, Department of Computer Science and Engineering*
[2, 3, 4]*UG Student, Department of Computer Science and Engineering S.A. Engineering College Chennai-600077*

Abstract: ***The action knowledge can be effectively learned from motion videos or images human action methods has shown. Experiments on public benchmark datasets and real world datasets show that our method outperforms several other state-of-the-art action recognition methods. However, few efforts has been made to improve the performances of action recognition in videos by adapting the action knowledge conveyed from still images to videos In this paper, an adaptation method has been used to enhance action recognition in videos by adapting knowledge from images. The adapted knowledge is used to learn the correlated action semantics by exploring the common components of both labelled videos and images. The existing video action recognition methods suffer from the problem of lacking sufficient labelled training videos. In such cases, over fitting would be a potential problem and the performance of action recognition is restrained. Thus, the over-fitting can be alleviated and the performance of action recognition is improved. Meanwhile, the extend the adaptation method to a semi-supervised framework which can leverage both labelled and unlabeled videos.***

Keywords: Adaption, computer vision, human action recognition, semi-supervised frameworks

I. INTRODUCTION

Videos on the Web are uploaded by users and produced by hand-held cameras or smart phones, which may contains of camera shake, occlusion, and cluttered background. Thus, these videos contain, which large intra class variations within same category it is now challenging task to recognize human action in such videos.

A large number of local features, motion scale invariant feature transform are extracted from videos then and all local features are quantized into a histogram vector using bag-of-words representation. Finally, the vector-based classifiers are used to perform the recognition in testing videos. When the videos are quit simple, recognition methods have are improved in promising results. However, noises and uncorrelated information may be incorporated into the bow during the extraction and quantization of the local features. Therefore, these methods are usually not robots. And could not be generalized well when the videos contain camera shake, occlusion, cluttered background, and so on. In order to improve the recognition accuracy, meaningful components of actions, e.g., related objects, human appearance, and so on, should be utilized to form a clearer semantic interpretation of human actions. Recent efforts have demonstrated the effectiveness of leveraging related object or human poses. These methods may require to training process with large amounts of videos to obtained good performance, especially for real world videos. However, it is quite challenging to collect enough labelled videos that cover a diverse range of action poses. knowledge adaptation from images to videos have shown better performance in application of cross media recognition and retrieval. Knowledge adaptation, also known as transfer learning, aims to propagate the knowledge from auxiliary domains to target domains. Thus, labelled video dataset as the target domain and an elated image dataset as auxiliary domain, the adapted knowledge can help improving performance of video action recognition.

In most recent situation, most of the knowledge adaptation algorithms requires sufficient label data in the target domain. In real world application, However, most videos are unlabeled or weak-labelled. Collecting weak-labelled videos is time consuming and labor intensive.

II. ADAPTATION METHODS

The method has been successfully utilized in area of multimedia analysis and computer vision. Knowledge adaptation method can be classified into two categories. The one requires that the data of target and auxiliary domains are represented using the common feature, which indicates the same type of features with the problem dimension. Thus, these methods cannot deal with the problem that the feature space from different domains are different. In adaptation from images to videos, the data of different domains may be represented by heterogeneous features, which means different features between target and auxiliary domains. In the order to extract the same type of features, the videos are represented as a sequence of key frames. However, in this way the underlying

©IJRASET (UGC Approved Journal): All Rights are Reserved

temporal information and videos may be lost and therefore the recognition performances may be constrained. There have been many adaptation methods which could adapt knowledge between domains in heterogeneous feature space. The most similar work as ours are heterogeneous features based structural adaptive regression, heterogeneous feature augmentation multiple kernel transfer learning and max-margin domain transforms. This method builds upon by leveraging the shared structures between them. It have same features of target and auxiliary domains, which is utilize to adapt the knowledge from auxiliary domain. Experimental results show that two kinds of features are more robust than only one features. Another typical and effective approach is to apply mapping function algorithm. The introducing mapping function, data from two domains can be readily compared in a common subspace and obtain the optimal projection of target and auxiliary domains. However it is difficult to train this model because it involves more parameters, which needs to be well set and tuned. The image features used in could not well characterize the dynamic information in videos. The transform matrix is learned to map the features from the target domain to the auxiliary domains. This method was designed for the supervised learning scenario. It is still unclear how to extend these methods to semi-supervised leaning methods. In summary, as we do not make feature augmentations, the paper is different form HDF and MMDT. Moreover, compared to MKTL and HFSAR, our method is more efficient and can utilize the unlabeled target videos, respectively.

III. COMPUTER VISION

It is used to seeks the intelligent and useful descriptions of visual scenes and sequences, and of the objects that populate them, by performing operations on the signals and it is received from video cameras. This problems are raised in AI (Artificial Intelligence) which used to building general vision machines would entail, to solve the problem they use artificial intelligence. It is an efficient way for, building flexible and robust visual representations of the world. It is also used to maintain and interfacing with attention, goals and plans. They are described based on building a signal-to-symbol converter. The external world present itself only by the signal on sensory surface, such as video camera, retina, microphone etc., which express few of the information required for intelligent understanding of the environment. The signal must be converted ultimately into symbolic representation whose manipulation allows the machine. A video camera contains a dense array which is an independent sensors, which convert incident photons focused by the lens to each point into a charge proportional to the light energy. The charge are coupled to capacity which allow a voltage to be read out in a sequence scanning the array. The photon flux into such small catchment areas is a factor limiting further increases in resolution by simply building denser imaging arrays

IV. HUMAN ACTION RECOGNITION

It is a widely studied computer vision problem. It is used in the application of HAR include video surveillance, health care and human interaction. The video based human activity recognition giving a over view of various approaches as well as their evolution by covering both the representative classical literatures and the state-of-art approaches. Human action as a inherent hierarchical structure that indicates different levels of which can be consider as a three-level categorization. Firs level, there is will be atomic element and these action primitives constitute more complex human activities. After the primitive level, the action activity comes as second. Finally, the complex interactions form the top level, which refers to the human activities that involves more than two persons and objects. Action primitives are those atomic actions at the limb level, such as “stretching the left arm,” and “raising the right leg. “Automatic actions are performed by a specifically, we refer the terminology human activities as all movements of the three layers and the activities as the middle level of human activities. Human activities like walking, running, and waving hands are categorized in the action level. Finally, similar interaction are human activities that involve two or more person and objects. The additional person or object is an important characteristic for interaction.

V. SEMI SUPERVISED

For the system, Semi-supervised is a method used to enable machines to classify both tangible and intangible objects. The objects the machines need to classify or identify could be as varied as inferring the learning patterns of students from classroom videos to drawing inferences from data theft attempts on servers. To learn and infer about various types of data based on which the machines need to learn from large, Structured and unstructured data they receive regularly. The acquisition of labeled data for a learning problem often requires a skilled human agent or a physical experiments. Semi Supervised is a class of supervised learning tasks and techniques that also make use of unlabeled data for training-typically a small amount of labeled data with large amount of data. Semi-supervised learning fails between supervised learning and unsupervised learning. Many machine-learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy. The cost associated with the labelling process thus may render a fully labelled can be of great practical value. Semi-supervised learning is also of theoretical interest in machine and as a model for human learning.

©IJRASET (UGC Approved Journal): All Rights are Reserved

VI. LITERATURE SURVEY

A. *Label Information Guided Graph Construction For Semi-Supervised Learning*

The paper [1] is based on semi-supervised learning methods which use the label information of observe sample in the label propagation stage, while ignoring such valuable information when learning the graph. The enforcing the weight of edges between labeled samples of different classes to the state of the art graph learning methods, such as the low rank representation learning method called semi supervised low rank representation. This is on both synthetic and real datasets demonstrate the label information indeed helps preserve the block diagonal structure of the coefficient matrices and significantly improves the performances of existing graph learning methods.

B. *A Multimedia Retrieval Framework Based on Semi-Supervised Ranking And Relevance Feedback*

In this paper [2] they use two algorithm local regression and global alignment to learn a robust Laplacian matrix for ranking. In a local linear regression model is used to predict the ranking scores of its neighboring points, where semi supervised are used for long term relevance feedback algorithm to reline the multimedia data representation. They are applied to several content based multimedia retrieval application, including cross media retrieval and 3D motion/pose data retrieval. It is used on four data sets have demonstrated its advantage in precision, robustness, scalability, and computational efficiency.

C. *Pose Primitive Based Human Action Recognition in Videos or Still Images*

In this paper [3] the learning mode, parameter representing poses and activities are estimated from videos. In run mode, the method can be used both for videos and still images. For recognizing pose they used Histogram of Oriented Gradient (HOG) based descriptor to better cope with articulated poses ad cluttered background. Action are represented using HOG and it is recognition is done based on a simple histogram comparison. Where they are not rely on background subtraction or dynamic features and thus allows for action recognition in still images.

D. *Learning Discriminative Key Poses for Action Recognition*

In this paper [4] Poses in video frames are described by the proposed extensive pyramidal features which include the Gabon, Gaussian, and wavelet pyramids. These features are able to encode the orientation, intensity and contour information and there by provide an informative representation of human poses. They use Ada boost algorithm to learn a subset of discriminative poses. Bayes nearest neighbor is proposed for the final action classification, which is demonstrated to be more accurate and robust than other classifier.

E. *Multilevel Chinese Takeaway Process and Label based Process For Rule Induction in the context of Automated Sports Video Annotation*

The paper [5] propose four variants of a hierarchical hidden Markov model strategy for indication in the context of automated sport video annotation including a multilevel Chinese takeaway ;process based on the chinses restaurant process and a novel Cartesian product label-based hierarchical bottom-up clustering method that employs prior information contained within label structures. Optimal performance is obtained using a hybrid topological structures with CLHBC generated event labels. That the methods proposed are generalizable to other rule based environments including human driving behavior and human actions.

F. *Semi Supervised Features Selection via Spline Regression for Video Semantic Recognition*

The paper [6] used to improve both the efficiency and accuracy of video semantic recognition, it can perform feature selection on the extracted video features to select subset of features form the high dimensional feature set for a compact and accurate video data representation. This develop semi supervised feature selection algorithms to better identify the relevant video features, which are discriminative target classes by effectively exploiting the information underling the huge amount of unlabeled video data. This is based on framework for video semantic recognition and it is used to discriminative information between labeled training videos and the local geometry structure of all the videos.

G. *Discriminative Tracking Using Tensor Pooling*

The paper [7] is to represent target templates and candidates directly with sparse coding tensor. Local sparse representation has been successfully applied to visual tracking, owing to its discriminative nature and robustness against local noise and partial occlusions. Local sparse codes computed with a template actually although three-order tensor according to their original layout, although most existing pooling operators covert the codes to a vector by concatenating or computing .Where it is used to deliver more informative

©IJRASET (UGC Approved Journal): All Rights are Reserved

and structured information, which potentially enhances the discriminative power of the appearances model and improves the tracking performances.

H. Structure Preserving Binary Representations for RGB-D Action Recognition

The paper [8] focuses on, Local representation for RGB-D video data fusion with a structure preserving projection. To acquire a general feature for the video data, we convert the problem to describing the gradient fields of RGB and depth information of video sequences. With the click fluxes of the gradient fields, which include the orientation and the magnitude of the neighborhood of each point a new kind of continuous local descriptor called local flux feature. This obtain a fused local a binary representation for RGB-D action recognition.

I. Learning Spatio-Temporal Representations For Action Recognition: A Genetic Programming Approach

In this paper [9] , instead of using handcrafted features, we automatically learn spatio-temporal motion features for action recognition. This is used to achieve via an evolutionary method. Genetic Programming which evolves the motion features descriptor on a population of primitive 3D operators. The GP- evolving features extraction methods is evaluated on four popular action datasets, namely KTH, HMDB51,UCF. This develop an adaptive learning methodology using GP (Genetic Programming) to evolve discriminative spatio temporal representation, which simultaneously fuse the color and motion information, for high level action recognition tasks.

J. Transfer Latent SVM for joint Recognition and Localization of Actions in videos

This paper [10] is based on web images and weakly annotated training videos. The model takes trainings videos which are only annotated with action label as input for alleviating the laborious and time-consuming manual annotations of action locations. For the purpose of improving the localization we collect an number of web images which are annotated with both action labels and action location to learn a discriminative model by enforcing the local similarities between videos and web images. A structural transformation based on randomized clustering forest is used to map the web images to video for handling the heterogeneous features of web images and videos.

VII. CONCULSION

The overall performances of video action recognition, we propose an classifier of IVA, which can borrow the knowledge adapted from images based on the common visual features. Meanwhile, it can fully utilize the heterogeneous features of unlabeled videos to enhance the performance of action recognition in videos. To validate, that the knowledge learned from images can influence the recognition accuracy of videos and that different recognition results are obtained by using different visual cues. Experimental results show that the proposed IVA has better performances of video action recognition, compared to the state of the art methods. And the performance of IVA is promising when only few labeled training videos are available.

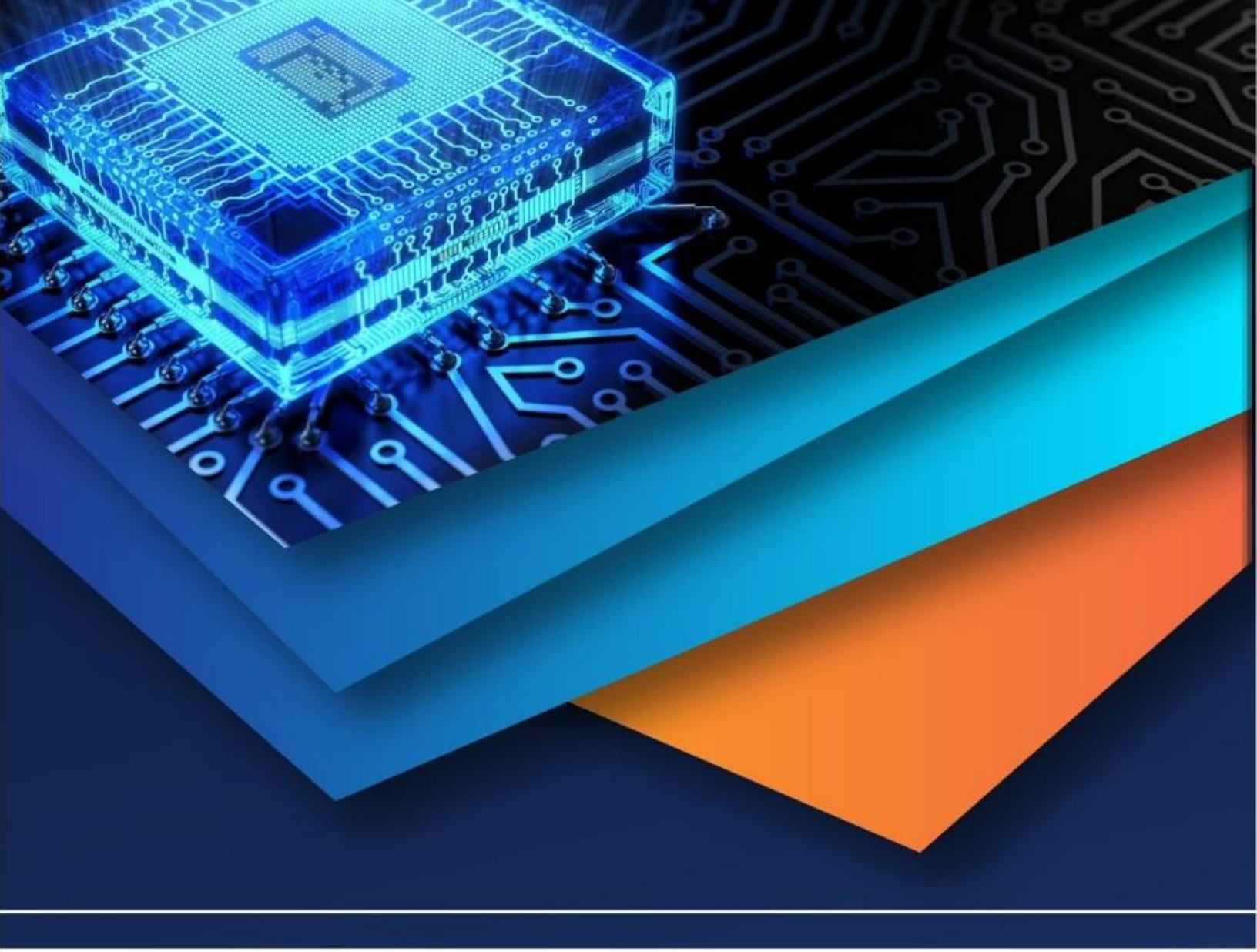
REFERENCES

[1] B. Ma, L. Huang, J. Shen and L. Shao, “Label Information Guided Graph Construction for Semi-Supervised Learning ” IEEE Trans. Cybern., To be published Doi:10.1109/TCYB.2015.2477879.
[2] M. Varma and B. Babu, “A Multimedia Retrieval Framework Based on Semi-Supervised Ranking And Relevance Feedback” in Proc. 26th Annu. Int. Conf. Mach. Learn. NY, 2009, PP. 1065-1072.
[3] Wu, Y. Zhang, W. Lin, “Post Primitive Based Human Action Recognition in Video or Still Images , ” IEEE Trans. Cybren., To be published Doi:10.1109/TCYB.2015.2493538.
[4] J. David, ”Learning Discriminative Key Poses for Action Recognition” In Proc IEEE Workshop Detect. Events Video bc, 2001, PP. 39-46
[5] M. Martin,” Multilevel Chinese Take way Process and Label based process for Rule Induction in the context of Automated Sports Video Annotation” behavior classification of motion. Pattern recognition, 38:1033-1043, 1996.
[6] J. Luo, M. Shah, “Semi Supervised Features Selection via Spline Regression for Video Semantic Recognition ” in Proc. IEEE conf. Pattern recognition., 2009, PP.1996-2003.
[7] C. Jerry, “Discriminative Tracking Using Tensor Pooling”, Proc IEEE Conf Pattern Recognitions, 1996
[8] Xie, P. Xu, H. Sun, ”Structure Preserving Binary Representation for RGB-D action Recognition ”, Pattern Recognition. Lett., vol. 25, No. 7, PP. 677-577,2005.
[9] J. Liu, J. Luo, “Learning Spatio- Temporal Representation for Action Recognition A Genetic programming approach “, Proc. IEE Conf. Pattern Recognition, 2008, PP.1-8.
[10] J. Mike, “Transfer Latent SVM for joint Recognition and Localization of action in video” Proc. ICML Workshop Statistical Regulation System Learning, PP. 132-137, 2004
[11] Y. Yang, J. Luo, “Global Alignment for Cross Media System” Proc. ACM Conf. Multimedia, 2009.
[12] F. Wu et al., “Video Domain Concept Using Adaptive SVMS,” in Proc IEEE Trans. PP. 354-775, 2003.
[13] T. K. Goorge, B. Martin “Mathematical Notes on Three Modes FA” Vol. 11, No. 4, PP. 279-311, 1996.
[14] A. Kevin, “ ECG Signals with Hidden Models”. Med., Vol. 8, No.5, 453-471,1996

©IJRASET (UGC Approved Journal): All Rights are Reserved

[15] K. Thomas, "Semi Supervised in Machine Learning" ,Vol. 10, No.20, 357-103, 2013.
[16] H. Jerry , "Application to Image Database", Proc IEEE conf Pattern 1996
[17] J. Martin, behavior classification of motion Pattern Recognition 38:1033-10433,1995
[18] K. David "Structure Motion Image for Reocgnizing Human Action" in Proc IEEE Conf. Pattern recognition 2009v
[19] Wu, Y. Luo "Better Practices for Learning to Recognize Action Using FV " IEEE Trans. Cybren 2015
[20] J. Shao, "Discriminative Tacking Using Tensor Pooling" IEEE Tran Workshop Detect Events 2001,PP. 39-46

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)