Automatic Visual Concept Detection in Videos: Review

Nilam B. Lonkar [1], Dinesh B. Hanchate [2]

[1] Student of Computer Engineering, Pune University
VPKBIET, Baramati, India
[2] Professor of Computer Engineering, Pune University
VPKBIET, Baramati, India

---***---

Abstract - *The video contains various objects, scene and action. Due to the various concept presents in video event detection is becoming difficult. Traditionally, visual concept detection carried out some significant improvement. For instance, the concept is defined by human and it has only one classifier. But this process is manual so it is time consuming. Automatic visual concept detection tries to imitate this process and perform the concept learning automatically. The aim of this study is to develop an automatic system that can learn visual concepts from the video. This detection is divided into three main areas i.e. event analysis, attribute concept and transfer knowledge respectively. In this paper various methods are like domain selection machine (DSM), cross-domain learning, feature and model selection etc. be reviewed*.

***Key Words*: Classification, event analysis, video recognition, visual concept detection.**

1. INTRODUCTION

With the fast development in video innovation, more data is accessible as advanced video information. Video ordering and recovery is rising as an imperative and challenging issue in multimedia applications. Different components, e.g. , shading, shape, surface, movement, shut inscription and discourse are being utilized for recovering recordings. In the majority of the current strategies, the recovery is either in light of some low-level features depending on the cases. The semantic (high-level) ideas are valuable and fundamental for questioning video databases. Along these, automatically extricating concepts or occasions in the video is a huge necessity for recovery.

Concept learning is divided into three parts, event analysis, attribute concepts and knowledge transfer, respectively. For event analysis, there are many methods that are proposed for event detection and recognition which are vision based. There are two methods; Domain Selection Machine (DSM) for event reorganization and another method is learning, natural language description for text description which is used to understand video. For visual attribute detection, various methods such as unseen object class detection and complex video semantic label methods are used. Here, the relation between object class and attributes are found using relatedness. In knowledge transfer, many methods have been proposed to improve classification accuracy. In that object and scene collection, this method is used to find out user intended searched for objects in the video. Another is a domain transfer method which is used to find out the relation between source domain and target domain.

2. RELATED WORK

Domain Selection Machine (DSM) [1] is based on web images which are introduced for event recognition of videos. To choose the most relevant source domains, it introduces something different information based on regularization. Support Vector Regression's (SVR) main aim is to use the insensitive loss, which focus on target classifier which shares common decision result based on the consumer videos which are not labelled with the chosen source classifier. To handle the imbalance between data delivery of two domains, i.e., web video domain and consumer video domain. Here, context information does not deal with videos and images, but only visual features are considered.

Advantages:

1. When the examples of the objective spaces are furthermore represented by another sort of elements i.e., the ST highlights, DSM are chosen for the most applicable source spaces.

2. The objective choice capacity and an area choice vector successfully are explained in the optimization calculation which is the most important source areas.

Disadvantages:

1. The grouping exhibitions in the objective area, irrelevant source spaces might be unsafe.

For textual description, images are gathered from web, H. Zhang and J. Guo [2] proposes a novel cross-domain learning mechanism which is used for delivering the correlation knowledge among different information sources used to divide condition of textual description missing in the image. For generating web multimedia substances, it not handles only a large amount of web multimedia objects, but also deals with most valid correlation knowledge. It proposes unique image representation model Bag-Of-visual-Phrases (BoP) which organizes the continuous and semantic information. It represents images that are both the word and the phrase level. But, all this process of feature retrieval is done offline only.

In concept based representation method [3], to handle multimedia event, recounting approach plans a pilot analysis. It gives tips on, what basis this decision is making up and why this video is categorized in this event. The recounting covers all additional semantic declaration of the event classification. So, this approach is generally suitable for any supplement classifier.

J. Liu, Q. Yu [4] proposes a method for learning natural language descriptions of the training videos. It is an easy way to collect and scale the number of actions and roles. Detailed information about spatial mortal annotation for the atomic action event may be recognized, where the state of the art method is required. The natural language descriptions for event recognition model corporates the role models which are trained previously.

Advantages:

1. Semantic relatedness (SR) have measured relation between a video description and an activity or role label. It additionally contains a posterior regularization objective in videos.

Disadvantages:

1. Natural language descriptions, just gives a coarse high state summary of the occasions happening in the videos.

In heterogeneous features and model selection of event based media classification paper [5], it targets the basic problem handling of social media analysis. They presents a relation between event and media. It finds how the specious and missing metadata is represented in the social site. It collects an event from social site, an gives event dataset and then finds similarity among media data and events. It then spread features with their identity for decreasing the amount of missing values. Then finds whether the values are missing or not.

Advantages:

1. To take the models in view of temporal, area and visual element of each occasion system uses SVM, decision tree and random forest techniques.

Z. Ma, Y. Yang, Z. Xu [6] proposes a method of detecting hard events, through multi source video attributes. Here specifically correlation vector is used for constructing correlation functions. The extra information about video attribute is used for detecting complicated event. The indication about the attribute of video is nothing but semantic labels assigned by the researcher. So this method not only considers map about video data attribute, but also refers to the vector of correlation which relates video attribute to complex event.

Advantages:

1. Video attributes are used as additional information on detecting events. It combines multiple features on both complex event videos and attribute videos for learning the detector.

Disadvantages:

1. When the utilization of number of attribute are limited, it is difficult to learn an intermediate representation.

In knowledge adaptation method [7], the multimedia refers the infer knowledge for detecting event. It introduces various semantic concepts related to the Ad Hoc method of target videos. Firstly, this approach mines shared inconsistency and noise among the various video. This is very important to collect positive examples.

To learn detection of unseen object, Christoph H. Lampert [8] proposes the class attributes transfer method. The object detection is generally depends on human in the targeted object rather than training images. The attributes involved like shape, color and geo information. Class attributes and semantics are categorized in intermediate stage. So this is very fast and easy way to include human in the system.

Advantages:

1. In an attribute layer, it is conceivable to manufacture a learning object recognition framework that are not required any preparation pictures of the objective classes.

Disadvantages:

1. Attribute based classification has solved the object classification problem by transferring information about classes.

J. Sivic and A. Zisserman [9] explains object and scene collection method. It finds user intended searches for those objects are found in the video. The videos are distributed physically, for using the region as track. In this system, inconsistency between information is reduced. The main goal is to effectively identify the object present in the video. Here in some cases, no explicit interest is provided at a time in this system.

Advantages:

1. The worldly coherence of the video inside a shot is utilized to track the areas keeping dismissing insecure districts and reducing the impacts of noise in the description.
2. The objective is indicate as a subpart of a picture and this is sufficient for quasi-planar inflexible objects.

Disadvantages:

1. The textual vocabulary is not static. In text retrieval, systems growing as new documents are added to the accumulation.

3. CONCLUSIONS

Reviewing this literature, an approach towards a new idea of developing an automatic system for concept detection from video is to be proposed. This core idea of implementation with the help of event analysis method, where automatically mine visual concepts from the text, may give efficient and effective system which requires very less user interaction.

ACKNOWLEDGEMENT

This paper would not have been written without the valuable advices and encouragement of Dr. D. B. Hanchate, guide of ME Dissertation work. Author's special thanks go to Prof. S. A. Shinde and Prof. S. S. Nandgaonkar, Head of Computer Department and Hon'ble principal Dr. M. G. Devamane, for their support and for giving an opportunity to work on survey of concept detection tool.

REFERENCES

[1] L. Duan, D. Xu and S.-F. Chang, "Exploiting web images for event recognition in consumer videos: A multiple source domain adaptation approach," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, June 2012.

[2] W. Lu, J. Li, T. Li, W. Guo, H. Zhang and J. Guo, "Web multimedia object classification using cross-domain correlation knowledge," IEEE Transactions on Multimedia, vol. 15, no. 8, pp. 1920-1929, 2013.

[3] Q. Yu, J. Liu, H. Cheng, A. Divakaran and H. S. Sawhney, "Multimedia event recounting with concept based representation," in Proceedings of the ACM Conference on Multimedia, 2012.

[4] J. Liu, Q. Yu, O. Javed, S. Ali, A. Tamrakar, A. Divakaran, H. Cheng and H. S. Sawhney, "Video event recognition using concept attributes," in Proceedings of the IEEE Workshop on Applications of Computer Vision, 2013.

[5] X. Liu and B. Huet, "Heterogeneous features and model selection for event-based media classification," in Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval. ACM, 2013.

[6] Z. Ma, Y. Yang, Y. Cai, N. Sebe and A. G. Hauptmann, "Knowledge adaptation for ad hoc multimedia event detection with few exemplars," in Proceedings of the ACM International Conference on Multimedia, 2012

[7] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[8] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in Proceedings of the IEEE International Conference on Computer Vision, vol. 2, 2003.

[9] M. Zaharieva, M. Zeppelzauer and C. Breiteneder, "Automated social event detection in large photo

collections," in Proceedings of the ACM International Conference on Multimedia Retrieval. ACM, 2013.

[10] Y. Yang, Z. Ma, Z. Xu, S. Yan and A. G. Hauptmann, "How related exemplars help complex event detection in web videos," in Proceedings of the IEEE International Conference on Computer Vision, 2013.

[11] G. Csurka, C. R. Dance, L. Fan, J. Willamowski and C. Bray, "Visual categorization with bags of key points," in Proceedings of the European Conference on Computer Vision, Workshop, 2004.

[12] X. Yang, Q. Song, and Y. Wang, "A weighted support vector machine for data classification," International Journal of Pattern Recognition and Artificial Intelligence, vol. 21, no. 5, pp. 961976, 2007.