A Survey on Sports Video Annotation Frameworks

N.Chidambaram[a,1] and G.Nallavan[b]

[a,1]*Research Scholar, Department of AST & T, TNPESU, Chennai, India*
[b]*Asst. Prof., Department of AST & T, TNPESU, Chennai, India*

Abstract. Video annotation technique delivers many additional video processing capabilities for several applications. Sports broadcast video content is unique in regard to wealth of information as compared to any other video. Sports video annotation is becoming popular among researchers in recent times because of wide range of applications and challenges it pose. The demand for optimized design of framework for sports video annotation is at peak. This paper surveys state-of-the-art in annotation framework design, particularly for sports applications and provides insight into future aspects. This survey may help researchers to further conceive and develop advanced universal frameworks applied to all sports.

Keywords. Sports video annotation, annotation framework design, annotation architecture, broadcasting video, Machine Learning,Neural Networks

1. Introduction

Several technological developments in recent past of video broadcasting and presentation have improved sports video broadcasting quality with significant increase in the number of subscribers. Apart from the conventional TV broadcasting, the boom in OTT platform, web-based (internet) telecasting, reduction in data cost and higher data rate led to a tremendous growth of sports entertainment industry in the recent times. Sports video annotation is regarded as an assisting mechanism in many sports video applications like analysis, retrieval, indexing, summarization[2], browsing/surfing, content mining, video skimming [11], providing supplementary information [10], generate metadata/metainfo for advanced techniques such as artificial intelligence and machine learning [11], video management and many more boundless areas [3,13].

The people in front of various devices watching the broadcast/telecast sports video outnumber the people watching it live on the stadium. So broadcasters have the responsibility to convince the needs of these viewers/consumers to commercially succeed and continuously entertain and retain them as long term subscribers. Annotation in sports video is crucial for broadcasters or even end-users to satisfy their commercial or personal needs respectively [1]. The upcoming sections are categorized as follows: Section 2 outlines the uses of video annotation in various sports video as a specific case.

---

[1] N.Chidambaram, Research Scholar, Department of Advanced Sports Training and Technology, TamilNadu Physical Education and Sports University, Melakottaiyur, Chennai, Tamil Nadu, India.
E-mail: researchofchidambaram@gmail.com

Section 3 analyzes various sports video annotation attributes found in literature. Section 4 deals with existing challenges and future aspects on this area. In section 5, we summarize this work and its importance in current research context.

2. Sports Video Annotation

Sports video annotation is still considered as hot research area among researchers due to its knowledge generating capabilities [14] and commercial value [11]. It conveys additional useful information for all kinds of videos [12]. However, the richness of content in sports video is completely different from other general videos. Furthermore, each sport video is diverse in nature, thus demanding different ways of approach in video annotation. For example, ball-only type of sport video annotation is inappropriate for racquet based sport [16].

Framework design of sports video semantic annotation is challenging, complex and very demanding [1,3]. Design of conventional or small scale framework for large scale dataset is becoming impractical [6]. The motive of video annotation and its framework is to relate the video features (low-level) and semantic labels (high-level) [3,13,14]. Moreover the framework should define semantic meaning of objects, events and context [8]. It implies scene interpretation at higher level and data acquisition. It is considered as the toughest task even for computer vision techniques [11,16].

2.1. Framework design/structure

The framework design falls into any one of two major categories. They are specific or generic. Specific frameworks are suitable for many sports videos and they concentrate more on sports/games related annotation tasks and generic frameworks are applied for several genres of videos with sports video as a subset. The Survey related to AVA (Automatic Video Annotation) is proposed by [13] and applicable to all general videos. The literature survey related to Sport video as specialization is carried out in this work.

2.1.1. Specific Framework

Many frameworks proposed in literature are available as sports specific and some of them can be extended to general video applications. This section lists all sport- specific video annotation frameworks and briefly discusses them. Changsheng Xu et al [1] Contributed framework for video summarization and retrieval of sports video. Here, two-level annotation scheme is used. First level gives the overall summary taken from webcasting text. And second level annotates every event of video employing semantics of text as well as video boundaries grabbed from alignment of text or video.

Mentzelopoulos et al [4] has provided a system for extracting shot boundaries using low-level feature video processing algorithms. Campos et al [5] proposed an automatic sports video annotation framework based on Bayesian reasoning framework aiming to annotate court sport videos at all cognitive levels with adaptability and event classification at any time with user request.

Assfalg et al [9] contributed semantic annotation system that utilizes visual cum graphical features on the video frames. Additionally, color histogram is used for object tagging. Xue et al [10] designed an AVA system for archival sports video. This work delivers rich metainfo of the videos archived.

Kolekar et al [11] labeled video clips by automatic segmentation of broadcast videos and framework provides answers of difficult queries related to video clips. Deng et al [16] proposed data analysis annotation framework. It is dedicated to racquet sports videos and has provision for tools to carry out interactive annotation. In addition, they used supporting computer vision algorithms.

2.1.2. Generic Framework

The generic frameworks in literature have potential to satisfy the requirements of sports video annotation. These frameworks are considered in this survey of sports applications, because of having provision to modify them for sport needs. Zhang et al [2] proposed a semi-supervised learning framework with six types of sports events for analysis. It uses labeled, unlabeled, small scale and large scale videos to train the model. Aote and Potnurwar [3] had undergone a novel approach to define a two- level keyframe extraction method for AVA.

Hwang et.al [6] attempted to provide deep insight from enormous video datasets available in internet to train. Getahun and Birara [8] used audio element of scenes to assist identification of object and event using high level architecture. Islam et al [14] highlighted the importance of distributed framework for AVA. The concept is spatio and spatio-temporal oriented that provides application based solution for users. Human action in sports is taken for their analysis. Huskey and Hill [15] facilitated dedicated video pane in video annotation interface with many functionalities.

3. Literature Survey

Various attributes of sports video annotation framework are listed in Table 1 with both specific and generic framework taken into consideration.

Table 1. Attributes of Sports Video Annotation Framework

Reference	Framework Type	Approach	Applications	Sports Use Cases	User/Personal Preference
[1]	Specific	Web Casting Text	Semantic annotation Indexing & Retrieval	Soccer, Basket ball and other sports/ games	Summary Creation
[2]	Generic	Semi-supervised Learning	Training for event detection and annotation	Basket ball	Search and browse Videos

[3]	Generic	Machine Learning	Shot detection, keyframe and feature extraction	Generally sports	None
[4]	Specific	Active region detection & extraction	Automatic Video segmentation for annotation	Football, Squash & Basketball	None
[5]	Specific	Anomaly Detection & Transfer Learning	Annotate court sports video	Tennis	Event Classification
[6]	Generic	Mapreduce training	annotation for large datasets	Basket ball (As framework input)	None
[7]	Generic	Convolutional Neural Network	Analysis and Management	General	None
[8]	Generic	Video Scenes & associated audio	Event and Object Identification	Basket ball (Shot & Scene Identification)	Video addition, threshold setting, result visualizing and XML or SRT file generation for annotation
[9]	Specific	Visual & graphical features using neural networks	annotate videos at different layers of semantic significance	Several sport videos & Studio/Interview shots	Retrieval of specific shots on demand
[10]	Specific	Computer Vision	archival sports video	Baseball	None
[11]	Specific	Event & Concept level	Semantic labeling	Soccer	Answers difficult queries
[14]	Generic	Spatial & Spatio-temporal	1.End-user distributed VA services 2.Developer algorithm services	Baseball (Pitch), Skate boarding & running	create new VA algorithms through VA & APIs
[16]	Specific	Data Analysis with CV Techniques	Multiple level video data annotations	Racquet sports	User event acquiring from videos & offers interactive tools

4. Future aspects

With the advent of advanced learning methods such as machine learning, computer vision and neural networks [3,7,16], the annotation task becomes easier. But, the existing data acquisition abilities suffer from limitations [16] and faces challenges [3]. The need for universal annotation framework for all sports is essential. Preference must be given to scalability, adaptability and applicability features of a framework.

5. Conclusion

This paper summarizes domain-specific and generic frameworks with application to sports video. The characteristics/attributes of frameworks related to sports video annotation are presented. Investigators, particularly in field of sports video annotation may find this work useful for their optimized framework design.

References

[1] Changsheng Xu, Jinjun Wang, Hanqing Lu, and Yifan Zhang, A Novel Framework for Semantic Annotation and Personalized Retrieval of Sports Video, *IEEE Trans. Multimedia*. (2008) 421–436.

[2] T. Zhang, C. Xu, G. Zhu, S. Liu, and H. Lu, A Generic Framework for Video Annotation via Semi-Supervised Learning, *IEEE Trans. Multimedia*. (2012) 1206–1219.

[3] S.S. Aote, and A. Potnurwar, An automatic video annotation framework based on two level keyframe extraction mechanism, *Multimedia Tools Appl*. (2019) 14465–14484.

[4] M. Mentzelopoulos, A. Psarrou, A. Angelopoulou, and J. García-Rodríguez, Active Foreground Region Extraction and Tracking for Sports Video Annotation, *Neural Process Lett*. (2013) 33–46.

[5] T. de Campos, A. Khan, F. Yan, N. Davar, D. Windridge, J. Kittler, and W. Christmas, A framework for automatic sports video annotation with anomaly detection and transfer learning, ACASVA, 2012.

[6] S.J. Hwang, J. Lee, B. Varadarajan, A. Gordon, Z. Xu, and A. (Paul) Natsev, Large-Scale Training Framework for Video Annotation, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, Anchorage AK USA, 2019: pp. 2394–2402.

[7] W. Tong, L. Song, X. Yang, H. Qu, and R. Xie, CNN-based shot boundary detection and video annotation, in: 2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, IEEE, Ghent, Belgium, 2015: pp. 1–5.

[8] F. Getahun, and M. Birara, Towards a Scene-Based Video Annotation Framework, in: 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), IEEE, Bangkok, Thailand, 2015: pp. 306–313.

[9] J. Assfalg, M. Bertini, C. Colombo, and A.D. Bimbo, Semantic annotation of sports videos, *IEEE Multimedia*. (2002) 52–60.

[10] Y. Xue, Y. Song, C. Li, A.-T. Chiang, and X. Ning, Automatic Video Annotation System for Archival Sports Video, in: 2017 IEEE Winter Applications of Computer Vision Workshops (WACVW), IEEE, Santa Rosa, CA, USA, 2017: pp. 23–28.

[11] M.H. Kolekar, K. Palaniappan, and S. Sengupta, A novel framework for semantic annotation of soccer sports video sequences, in: IET 5th Conference on Visual Media Production (CVMP 2008), IEE, London, UK, 2008: pp. 8–8.

[12] V. Mittal,Annotation Framework for Video, Google LLC, Mountain View, CA (US Patent), 2020.

[13] K. Randive, and R. Mohan, A State-of-Art Review on Automatic Video Annotation Techniques, in: A. Abraham, A.K. Cherukuri, P. Melin, and N. Gandhi (Eds.), Intelligent Systems Design and Applications, Springer International Publishing, Cham, 2020: pp. 1060–1069.

[14] M.A. Islam, M.A. Uddin, and Y.-K. Lee, A Distributed Automatic Video Annotation Platform, *Applied Sciences*. (2020) 5319.

[15] S. Huskey, and R. Hill, Multi-Modal Collaborative Web-based Video Annotation System, University of North Carolina, (US Patent), 2018.

[16] D. Deng, J. Wu, J. Wang, Y. Wu, X. Xie, Z. Zhou, H. Zhang, X. Zhang, and Y. Wu, Event Anchor: Reducing Human Interactions in Event Annotation of Racket Sports Videos, 2021.