

# Event Detection in User Generated Video Content: A Comprehensive study

Ms. Susmitha.A<sup>1</sup> and Dr. Sanjay Jain<sup>2</sup>

<sup>1</sup>Research Scholar, ECE, CMR Institute of Technology,  
Bengaluru, India

<sup>2</sup>Professor, CMR Institute of Technology  
Bengaluru, India

**Abstract** — with the rise of social media there is an impulsive increase in user generated videos on the web day by day. Each minute, 300 hours of user generated video are uploaded in YouTube due to which people find it difficult to look for the appropriate content amongst huge number of videos. So it has become increasingly important to find the solution to automatically figure out the events captured in this large collection of videos according to user's interest. As it is highly challenging, it has grabbed the attention of the researchers. One of the key tasks in this direction is event detection.

Joint analysis of various features in a single modality is a verified and useful approach for a variety of tasks in the computer vision field. From past few years there has been increased interest to the fusion of multiple features from different modalities in the user generated video content as it improves the accuracy of the detection of events. This paper provides some of the key challenges for unconstrained user generated videos. It also presents a few of the research work on multi modal event detection in user generated video. Finally we review its role in present research directions which contain browsing, indexing, retrieval, summarization, surveillance and consumer content management applications.

**Keywords** — Multimodal Event Detection, User Generated Video Content, Modalities, Video Indexing, Video Retrieval, Video Summarization, Computer Vision

## I. INTRODUCTION

The enormous development of digital technology along with the use of internet brought so much popularity for social networking sites such as YouTube, Instagram, and Linked in, Flickr and Pinterest. Growth of social media over the last one decade attracted users to create and to immediately post their own content with no restrictions on the content. As a result the user generated content [3, 5-7, 33, 40, 41, 44, 47, 51, 60, 61, 79] has been increasing rapidly on the web.

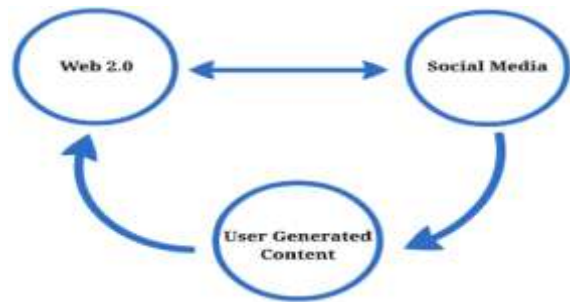


Fig.1. Creation of user generated content over the web with the help of social media (source from Google)

The most widely used user generated contents are text messages (tweets), Audio---Speech and Music, Images or pictures and short sequences of moving images also known as video clips. User generated contents are valuable resources of information capturing people's interests, thoughts and actions. Automatic video understanding [58] is the crucial among them.



Fig. 2 over view of types of videos

### A. Video

Video is short sequences of moving images and audio. A video is only a synchronous arrangement of several frames, each frame being a 2-D representation. So, the important unit in a video is a frame. The video can be considered as a gathering of numerous scenes as shown in fig 3, [12] where scene is as accumulation of shots that have a similar setting. Therefore video consists of enormous amount of content in terms of scenes, shots and frames.

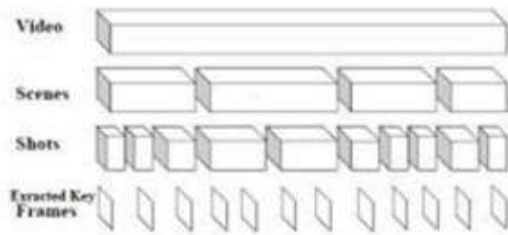


Fig. 3 Internal complex architecture of video

### B. User generated video opportunities and challenges

Most of the user generated videos are captured in uncontrolled conditions. They are unstructured, unfiltered, unverified and badly annotated in comparison with the edited news and broadcast videos.

In general user generated video is

- Of poor quality: As there are restricted capturing situations they may be of poor quality than professionally edited videos.
- Less organized: Both sports and news videos are made after proper editing. However, most UGV is usually captured using their own smartphones by individual users, and without any editing they will be up loaded on the web. So, UGV is unstructured[60].

According to twitter statistics, each minute twitter dynamic users create about 500 million tweets everyday and YouTube users post 300 hours of videos due to which people find it difficult to look for the appropriate content amongst a huge number of videos. Therefore the need of automatic event detection in user generated video data is obvious in many computer vision applications. Finding a solution to automatically figure out the events captured in this large collection of videos is not an easy task. Therefore user generated video data provides both opportunities as well as challenges. The primary challenge is how to handle such huge data in a proficient manner. In addition to that it is complex to search videos based on user interested specific events.

### II. Event Detection

The aim of event detection in user generated videos is to automatically detect the video clips based on user interested specific events in a given video.[1,4,8,9,11,15-17,19,20-22,27-29,36,40,41].User specific interested events detection in UGV is quite challenging task because the videos are captured in restricted environment by non-professional users. They are unstructured and unconstrained. They suffer from poor quality,

deprived lighting, blurring, complex camera motion, chaotic background clutter, and obstructions. An event is a significant occurrence specified by the user. It happens at a certain place at a certain time. For example, human actions like jumping, running and object related events like kicking ball, riding horse etc. Events are characterized by its type, time, location and description. The process of identifying the occurrence of an event in a natural or manmade content is called as event detection. Detecting substantial change in sea-level is a natural event detection and detecting events like smiling, frowning etc. in images or videos is called manmade event detection. The social media platforms allow millions of people use them daily to communicate and share information ranging from world level information for example World Cup, to personal information like wedding, graduation. Massive amount of data is created by individual users in the form of texts, videos, and photos. The research done so far could give efficient solutions for large data storage but retrieving, handling and processing of such large amount of data particularly in videos is still a challenge. So monitoring and evaluating the user-generated video content can produce undoubtedly valuable information.

### A. General event detection methodology

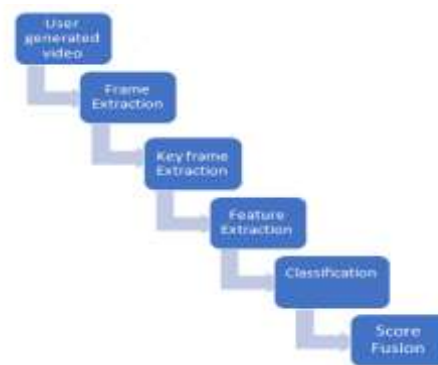


Fig.4. General Event detection methodology

The general event detection methodology is explained as

- Frame Extraction- Depending upon the size of video it consists of number of frames. Frame extraction is to represent the video in terms of image frames.
- Key Frame Extraction- extraction of key frames is the basic step in video related tasks to get rid of the duplicate frames with unnecessary data. The extracted key frames represent the characteristics of the video.
- Feature extraction- It extracts important features from video data to enable semantic understanding. The feature extraction process can be based on visual or audio

features. The visual features are such as spatial, transform, color, texture, shape, edge and boundary, structure, layout, and motion. The most common audio classes in videos are speech, silence, music and the combination of later three. The audio features can be embedded with low level visual features for key frame extraction.


- Classification- Classification is done using classifiers. After extracting features classifiers generate scores based on different model formulations and set of features. These are the most significant concepts in image processing, these are computer based mathematical algorithms developed to encounter the required performance level, at trained data set with a given amount of time. The classifier is trained in such way that the system must easily differentiate the datasets.




The following classifiers are the most widely used classifiers

- Naïve –Bayes Classifier
- Support Vector Machine (SVM)
- Multi-Layer Perceptron
- Clustering Classifier
- ANN classifier
- Bayesian Inference
- Dynamic Bayesian Networks
- Decision Tree (DT)
- Multiple Correspondence Analysis (MCA)
- Maximum Entropy model
- Latent Dirichlet Allocation(LDA)
- Hidden Markov Model(HMM)

- Score fusion –It combines scores computed from different features from different modalities. Based on this decision can be made.

Table.1. Over view of different modalities

S.no	Modality	Example	Description and Related work
	Text		Most of the research work done so far on the single text modality conveyed that the textual features from a video were obtained using either automatic speech recognizer (ASR) or optical character recognition

			(OCR)[51]
2	Audio (Speech and Music)		MFCC (Mel-frequency cepstral coefficients) is a popular and standard feature of an audio. [51]
3.	Visual (Image)		The most widely used low level visual features are spatial-temporal interest points (STIP)[34], scale invariant feature transform (SIFT)[31], histogram of gradients (HOG)[60], histogram of optic flow (HOF)[60], color, GIST, independent subspace analysis (ISA), geometry texton histogram (GTH), transformed color histogram (TCH), local binary patterns (LBP), speeded up robust features (SURF)[79].  The above said features are helpful in extracting different types of visual information from the video. For example ISA feature captures the temporal dynamics in video clips.
4	Motion		It is the representation of kinetic energy. It is used to measure the variation of pixels within a shot, direction of the motion and histogram magnitude.[51]

Event detection in a user generated video using single modality is not up to the mark. Hence multi modal fusion techniques have popped up. So, in

recent years multimodal event detection has been attracting extensive research attention.

### III. Multi modal event detection

Multi modal event detection [23,26,32,34,37,42-45,48,49,56,70,72-73,77] is based on multi modal fusion techniques. (i.e.) fusion of multiple features from different modalities which is referred as multimodal fusion. The fusion of various modalities may give related important information and therefore it is better to know which modalities will contribute major role for accomplishing an analysis task. Fusion of both visual and audio features along with the video textual [54] data in a user generated video will improve the accuracy of event detection. Hence, extraction of helpful features from video one by one to get better recognition of events is a further important task due to the discrete features of the concerned modalities.

As different modalities possess different characteristics it is better to consider the confidence levels of the modalities in completing the required multimedia applications.

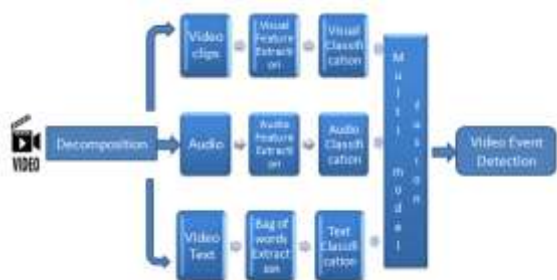


Fig.5. Frame work of multi modal event detection

Table.2. Over view of related work

S . n o	Pap er	Modalities used	Classifier	Multimed ia task
1	[1]	Low level visual features alone	Hidden Markov Model (HMM)	Sport video event detection
2	[2]	Visual. Audio and text modalities	Hidden Markov Model (HMM)	Sports video summarisation
3	[34]	Low level video(spatio-temporal descriptors) and audio((Mel-frequency cepstral coefficients) features	CRF based discriminative learning	Event detection in sports video[

4	[27]	Visual features (HOG, CEDD, color histogram, texture, and wavelet)	Decision Tree (DT), Multiple Correspondence Analysis (MCA)  Support vector machine(SVM)	Automatic Video Event Detection in disaster data set.
5	[60]	Visual features (static Opponent Scale-Invariant-Feature Transform(SIFT) 3D spatial-temporal interest points (STIPs))	Support vector machine(SVM)	Complex event detection in user generated video
6	[53]	Video(color, structure, shape)  Audio(MFCC), textual cues	Support vector machine(SVM)	Semantic concept detection
7	[78]	Audio(ZCR, LPC, LFCC)  video(blob location and area)	Bayesian Inference	Event detection for surveillance
8	[76]	Visual (color SIFT)  Audio (MFCCs), (acoustic segment model)	Latent Support vector machine(LSVM)	Multi media event detection
9	[79]	Sensor data modalities(auxiliary sensors)	Support vector machine(SVM)	Interesting event detection in UGV and extracting appropriate information about the recording activity.
10	[58]	Video and audio	Modern convolution neural network (CNN)	Audio-visual salient event detection
11	[26]	Visual, audio and motion features	SVM with Gaussian kernel	Robust event recognition in videos

#### IV. Various Research Applications

Video is a rich source of information and topics on video data offers a wide range of research applications such as multi modal event detection which further helps in video browsing, video indexing, video summarization, content-based video retrieval, video structuring, semantics modeling and surveillance applications.

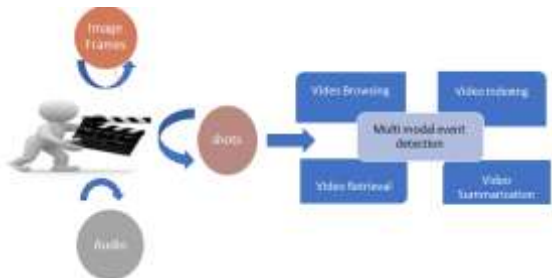


Fig.6. various research applications of multimodal event detection in user generated video content

By using either video browsing or video retrieval the video content can be accessed. Video browsing is useful when we need to get the essence of content. When we know what exactly we are looking for, video retrieval is useful.

##### A. Video indexing

In this busy life, instead of seeing the entire video, most of the users would like to watch the most exciting and interested scenes of the entire video. So, to fulfill the user’s requirements, specific video sequence which is more striking, exciting is grouped to form the indexed video. The goal of video indexing is to deliver more data of the entire video within short time with good accuracy. Color, texture and motion features are very helpful for efficient video indexing.

##### a) Various video indexing techniques

In topical years, video indexing has turned into a developing field of research. It has a vital impact. The different indexing strategies as per points of interest are represented as shown below.

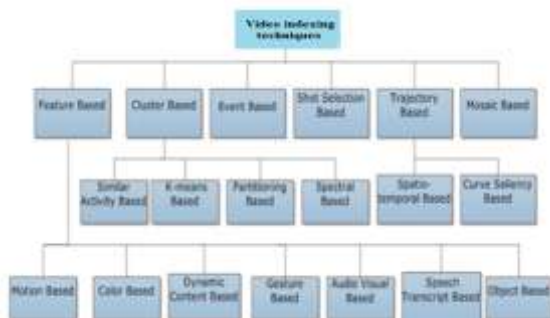


Fig.7. Different video indexing techniques

##### b) Video indexing framework.

The framework of Video indexing is given in detail in the below mentioned diagram

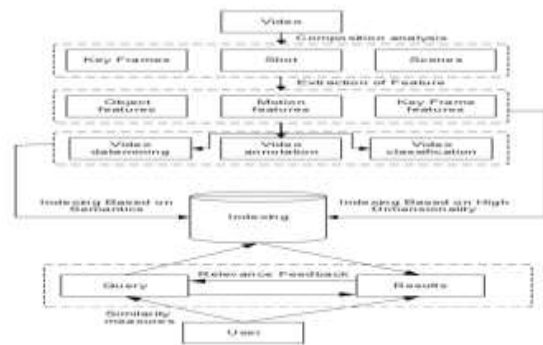


Fig.8. Video indexing frame work

##### B. Video summarization

Summarization is another exciting area because of the necessity to explore the content’s structure. Due to enormous growth of UGV on the web, it is essential to have automatic techniques of producing the brief representation of the video content i.e. summary of the video.

##### a) Video summarization Frame work

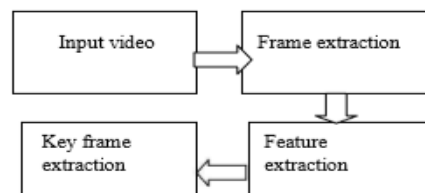


Fig.9. Video summarization frame work

##### C. Video retrieval

For a given video query, how to get a like clip of the video is called video retrieval. The temporal information present in a video content is the important factor that needs to be considered. It may complicate the querying system, but helps in describing useful information. The temporal information brings the concept of motion for the objects present in the video content. Other domains where Video retrieval can be applied are

- News Broadcasting
- Advertising.
- Music video clips. Etc....

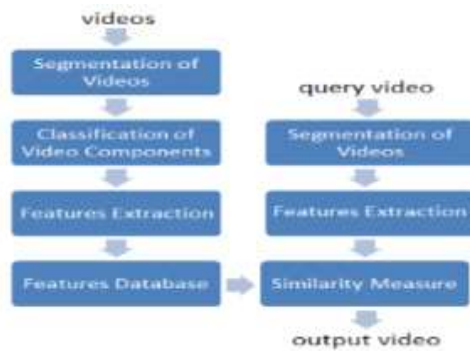


Fig.10. Video retrieval frame work

## V. Conclusion

In recent years, multimedia event detection has been attracting extensive research attention because of the exponential increase in volume of web video data. This paper focused on multi modal event detection in user generated video content where the objective is video clips detection by the key event happening in the clip by the fusion of different features from different modalities. We discussed several challenges that need to be addressed due to the exponential growth of unstructured web user generated video content. We have given an outline of the event detection and the general multi modal event detection frame work. Key points about modalities, features, classifiers, and fusion techniques were presented. We also discussed applications at the end with a few promising directions for future research. We think that this paper can give important insights for researchers who are just starting to investigate this area.

## REFERENCES

- [1] B.Liand ,M.Ibrahim Sezan: "Event detection and summarization in sports video". In IEEE Workshop on Content-Based Access of Image and Video Libraries, 2001
- [2] D. Tjondronegoro, Y.-P. P. Chen and B. Pham: "Sports video summarization using high lights and play-breaks". In Proceedings of the 5<sup>th</sup> ACM SIGMM international workshop on Multimedia information retrieval, (2003)
- [3] Rajiv Ratan Shah: "Multimodal Analysis of User-Generated Content in Support of Social Media Applications". In ICMR'16
- [4] Dong-Jun Park: "Video event detection framework on large-scale video data"
- [5] Manos Schinas ,Yiannis Kompatsiaris,Georgios Petkos, Symeon Papadopoulos : "Multimodal Event Detection and Summarization in Large Scale Image Collections". In ICMR'16
- [6] Yu-Gang Jiang, Xiaohong Zeng, Guangnan Ye, Subhabrata Bhattacharya, Dan Ellis, Mubarak Shah, Shih-Fu Chang Columbia-UCF TRECVID2010 Multimedia Event Detection: Combining Multiple Modalities, Contextual Concepts, and Temporal Matching. In NIST TRECVID Workshop (2010)
- [7] Francesco Cricri1, Kostadin Dabov, Igor D.D. Curcio, Sujeet Mate, Moncef Gabbouj: "Multimodal event detection in user generated videos". In IEEE International Symposium on Multimedia (2011)
- [8] Ge AradMedioni, Isaac Cohen, Franc Eois Bre Âmond, Somboon Hongeng and Ramakant Nevatia: IEEE "Transactions on Pattern Analysis and Machine Intelligence", Vol. 23, No. 8, August 2001: Event Detection and Analysis from Video Streams'
- [9] Cheng Lu & Yun Jin: "Real-time event detection by twitter"
- [10] Anastasios Doulamis & Nikolaos Doulamis& Luc van Gool & Mark Nixon -Guest editorial, Multimedia Tools Appl 'Event-based video analysis/retrieval.In springer(2013)
- [11] Farzindar, A., and Wael Khreich : "A Survey of Techniques for Event Detection in Twitter". In computational intelligence (2015)
- [12] Sachan Priyamvada Rajendra,Dr. Keshaveni N "A Survey of automatic video summarization techniques 'IJEECS,Volume3, Issue1, April2014.
- [13] Shruthi N, Ms. Priyamvada S, "Dominant Frame Extraction for Video Indexing". In IEEE International Conference on Recent Trends in Electronics Information Communication Technology (2017)
- [14] Weiqiyang & jonathan Weir 'Visual event computing'
- [15] Julio Borges, Peter Bozsoky, Michael Beigl "Advances in Event Detection"17 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)
- [16] ArifNurwidyantoro, EdiWinarko "Event Detection in Social Media: a Survey"
- [17] Hila Becker, Mor Naaman ,Luis Gravano "Event Identification in Social Media" Twelfth International Workshop on the Web and Data bases(WebDB2009), June 28, 2009, Providence, Rhode Island, USA.
- [18] Baoxin Li and M. Ibrahim Sezan: "Event Detection and Summarization in Sports Video" 2001 IEEE
- [19] Min Xu, Liang-Tien Chia, Jesse Jin: "Affective Content Analysis in Comedy and Horror Videos by Audio Emotional Event Detection"2005 IEEE
- [20] Smitashree Choudhury ,Harith Alani, , "Personal Life Event Detection from Social Media" Article · January 2014
- [21] Chung-Hong Lee ,Hsin-Chang Yang ,Tzan-Feng Chien and Wei-Shiang Wen "A Novel Approach for Event Detection by Mining Spatio-temporal Information on Microblogs" 2011 IEEE
- [22] Nikolaos D. Doulamis, Anastasios D. Doulamis, Panagiotis Kokkinos, and Emmanouel (Manos) Varvarigos "Event Detection in Twitter Microblogging" iee Transactions On Cybernetics, Vol. 46, No. 12, December 2016
- [23] Viktor Slavkovikj, "Multimodal Analysis for Object Classification and Event Detection"
- [24] Ziyou Xiong, Regunathan Radhakrishnan, Ajay Divakaran, Thomass. Haung "A unified frame work for Video Summarization, Browsing and Retrieval".
- [25] Stéphane Marchand-Maillet "Content-based video Retrieval: An overview"
- [26] Mennan Güder, Nihan Kesim Çiçekli: "Multi modal video event recognition based on association rules and decision fusion" .In Multimedia Systems (2018)
- [27] Samira Pouyanfar and Shu-Ching Chen: "Automatic Video Event Detection for Imbalance Data Using Enhanced Ensemble Deep Learning"
- [28] M. Xu, L.-Y. Duan, C. Xu, M. Kankanhalli, and Q. Tian. "Event detection in basketball video using multiple modalities". In Proceedings of the Joint Conference of the Fourth International Conference on Information, Communications and Fourth Pacific Rim Conference on multimedia., volume 3, dec. 2003
- [29] ZhicongTan, Peng Zhang,Jianlong Tan and Li Guo: "A multi-layer event detection algorithm for detecting global and local hot events in social networks".In ICCS(2014)
- [30] Klaser, A., Marszalek, M., Schmid, C: A "spatio - temporal descriptor based on 3d-gradients". In BMVC (2008)
- [31] van de Sande,K.E.A,Gevers,T., Snoek, C.G.M: "Evaluating color descriptors for object and scene recognition". PAMI32(9)(2010)

- [32] Lan,Z.Z.,Bao,L., Yu, S.I.,Liu, W., Hauptmann, A.G.: “Double fusion for multimedia event detection”. In: ICME (2012)
- [33] Tamrakar, A., Ali, S., Yu, Q., Liu,J., Javed, O.,Divakaran, A.,Cheng, H.,Sawhney,H.s.: “Evaluation of low level features and their combinations for complex event detection in open source videos”. In:CVPR(2012)
- [34] Ehtesham Hassan, M Gopal, Santanu chaudhury, Vikram Garg: “A Hybrid Framework for Event Detection Using Multi-modal Features.” In IEEE International Conference on Computer Vision Workshops(2011)
- [35] Nevenka Dimitrova, Hong-Jiang Zhang, Behzad Shahraray, Ibrahim Sezan,Thomas Huang Champaign: “Applications of Video-Content Analysis and Retrieval”
- [36] J. Nisha, C. Santanu, S. Roy, P. Mukherjee, K. Seal, and K. Talluri : “A novel learning-based frame work for detecting interesting events in soccer videos”. In Proceedings of the Sixth Indian Conference on Computer Vision, Graphics Image Processing, Dec. 2008
- [37] M.Pijl, S.vandePar, and C.Shan : “An event-based approach to multi-modal activity modeling and recognition”. In Proceedings of the International Conference on Pervasive Computing and Communications, 2010.
- [38] N. Dalal and B. Triggs: “Histograms of oriented gradients for human detection”, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.
- [39] S.-C. Chen, S. Sista, M.-L. Shyu and R. L. Kashyap: “Augmented transition networks as video browsing models for multimedia databases and multimedia information systems,” in 11th IEEE International Conference on Tools with Artificial Intelligence, 1999.
- [40] M.L. Shyu, Z. Xie,c M. Chen and S.C. Chen: “Video semantic event/concept detection using a subspace-based multimedia data mining framework”, IEEE Transactions on Multimedia 10(2) (2008).
- [41] Y. Yan, Q. Zhu, M.L. Shyu and S.C. Chen: “A classifier ensemble framework for multimedia big data classification”, in 17th IEEE International Conference on Information Reuse and Integration, 2016.
- [42] Yasin Yilmaz, Alfred O. Hero: “Multi modal event detection in Twitter Hash tag networks”. In Springer (2016)
- [43] Mohammad Soleymani, David Garcia, Brendan Jou, Bjorn Schuller, Shih-Fu Chang, Maja Pantic: “A Survey of Multimodal Sentiment Analysis”. In Image and Vision Computing (2017)
- [44] Natarajan,P.,Wu,S.,Vitaladevuni,S.N.P.,Zhuang,X.,Tsakali dis,S.,Park,U.,Prasad,R : “Multi modal feature fusion for robust event detection in web videos”. In CVPR (2012)
- [45] Jiang, L., Hauptmann, A.g., Xiang,G: “Leveraging high level and low level features for multi media event detection”. In ACM-MM (2012)
- [46] Du tran, Junsong Yuan, David Forsyth: “Video event detection: from sub volume localization to spatio temporal path search”. In IEEE Transactions on pattern analysis and machine intelligence (2014)
- [47] Joannis Kompatsiaris, Sotiris Diplaris, Symeon Papadopoulos: “Extracting emergent semantics from large scale user generated content”. (2012)
- [48] Virginia Fernandez Arguedas,Qianni Zhang and Ebroul Izquierdo: “Multi modal Fusion in Surveillance applications”
- [49] S.-F.Chang, D.Ellis, W.Jiang, K.Lee, A.Yanagawa, A.C.Loui,J.Luo : “Large scale multi modal semantic concept detection for consumer video”. In ACM MIR (2007)
- [50] Gregory K. Myers , Ramesh Nallapati , Julien van Hout , Stephanie Pancoast , Ramakant Nevatia ,Chen Sun Amirhossein Habibiian ,Dennis C. Koelma , Koen E. A. van de Sande ,Arnold W. M. Smeulders , Cees G. M. Snoek: “Evaluating multimedia features and fusion for example-based event detection.” In Machine Vision and Applications (2014)
- [51] Pradeep K. Atrey , M. Anwar Hossain , Abdul motaleb El Saddik , Mohan S. : “Multimodal fusion for multimedia analysis: a survey”. In Multimedia Systems (2010)
- [52] Thomas B. Moeslund , Omar Javed, Yu-Gang Jiang, R. Manmatha: Special issue on Multimedia Event Detection
- [53] Adams, W., Iyengar, G., Lin, C., Naphade, M., Neti, C., Nock, H., Smith, and J.: “Semantic indexing of multimedia content using visual, audio, and text cues”. EURASIP J. Appl. Signal Process. (2003)
- [54] Chang, S.F., Manmatha, R., Chua, and T.S.: “Combining text and audio-visual features in video indexing”. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, IEEE Computer Society, Philadelphia (2005)
- [55] Chetty, G., Wagner, M.: “Audio-visual multimodal fusion for biometric person authentication and liveness verification”. In: NICTA-HCSNet Multimodal User Interaction Workshop, Sydney (2006)
- [56] Hsu, W., Kennedy, L., Huang, C.W., Chang, S.F., Lin, C.Y: “News video story segmentation using fusion of multi-level multi-modal features in TRECVID 2003”. In: International Conference on Acoustics Speech and Signal Processing. Montreal, QC (2004)
- [57] Xu, H., Chua, and T.S.: “Fusion of AV features and external information sources for event detection in team sports video.” ACM Trans. Multimed. Comput. Commun. Appl. (2006)
- [58] Petros Koutras, Athanasia Zlatinsi and Petros Maragos: “Exploring CNN-based architectures for multimodal salient event detection in videos”. In Proc. 13th IEEE Image, Video, and Multidimensional signal processing (IVMSP) Work shop (2018)
- [59] Thorsten K’uhnappel , Tele Tan1, Svetha Venkatesh , Eric Lehmann: “Calibration of audio-video sensors for multimodal event indexing”. In IEEE ICASSP (2007)
- [60] Jinlin Guo, David Scott, Frank Hopfgartner, Cathal Gurrin: “Detecting complex events in user generated video using concept classifiers”.In CBMI (2012)
- [61] Lamberto Ballan, Marco Bertini, Alberto Del Bimbo, Lorenzo Seidenari , Giuseppe Serra: “Event detection and recognition for semantic annotation of video”. In Multimedia Tools Appl(2011)
- [62] Snoek, C., Worring, and M.: “Multimodal video indexing: A review of the state-of-the-art. Multimedia Tools and Applications 2005”
- [63] Brezeale, D., Cook, D.: “Automatic video classification : A survey of the literature.” IEEE Transactions on Systems, Man, and Cybernetics 38(3), 416–430 (2008)
- [64] Yining Deng, B.S. Manjunath, "Content-based Search of Video Using Color, Texture, and Motion", IEEE, pg 534-537, 1997
- [65] Lavee, G., Rivlin, E., Rudzsky, and M.: “Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video”. IEEE Transactions on Systems, Man, and Cybernetics 39(5), 489–504 (2009)
- [66] Yang, J., Jiang, Y.G., Hauptmann, A.G., Ngo, and C.W.: “Evaluating bag-of-visual-words representations in scene classification”. In: Proc. of Int’l Workshop on Multimedia Information Retrieval (MIR) (2007)
- [67] Ballan, L., Bertini, M., Del Bimbo, A., Serra, and G: “Video annotation and retrieval using ontologies and rule learning”. IEEE Multimedia (in press) (2010)
- [68] San Miguel, J., Martinez, J., Garcia. A: “ontology for event detection and its application in surveillance video”. In: Proc. of Int’l Conference on Advanced Video and Signal-Based Surveillance (AVSS) (2009)
- [69] Yan Ke, Martial Hebert: “Event detection in crowded videos”. In IEEE international conference on computer vision (2007)
- [70] Youngjune Gwon, William M.Camp bell, Kevin Brady and Douglas Sturim: “Multimodal Sparse Coding for Event Detection”. In Multimodal Machine Learning Workshop at NIPS (2015)

- [71] YuchiLiu', Lingda Wu, Zhen Lei, Yuxiang Xie: "Hierarchically Multi-modal Indexing of Soccer Video". In IEEE (2006)
- [72] Yu-Gang Jiang, Subhabrata Bhattacharya, Shih-Fu Chang, Mubarak Shah: "High-level event recognition in unconstrained videos". In Multimedia Info Retr (2013)
- [73] Sergio Escalera, Eloi Puertas, Petia Radeva, and Oriol Pujol: "Multi-modal laughter recognition in video conversations". In IEEE (2009)
- [74] Mohamed Elhoseiny, JingenLiu HuiCheng, Harpreet Sawhney, Ahmed Elgammal: "Zero-Shot event detection by multimodal distributional semantic embedding of videos". (2016)
- [75] Aasif Ansari, Muzammil H Mohammed: "Content based video retrieval systems - methods, techniques, trends and challenges." In International Journal of Computer Applications (2015)
- [76] Sangmin Oh, Scott Mc Closkey, Ilseo Kim, Arash Vadat, Kevin J. Cannons, Hossein Hajimirsadeghi, Greg Mori, A. G. Amitha Perera, Megha Pandey, Jason J. Corso: "Multi media event detection with multimodal feature fusion and temporal concept localization". In Machine Vision and Applications (2014)
- [77] Francesco Cricri1, Kostadin Dabov, Igor D.D. Curcio, Sujeet Mate, Moncef Gabbouj: "Multimodal extraction of events and of information about the recording activity in user generated videos". In Multimedia tools and applications. In (2014)
- [78] Atrey, P.K., Kankanhalli, M.S., Jain, R: "Information assimilation framework for event detection in multimedia surveillance systems".
- [79] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: "SURF: Speeded up robust features". Computer Vision and Image Understanding 110(3), 346–359 (2008)