

A Study on Relational Semantic Video Content Extraction Using Bayesian Network Classifier

S.Rajarajeshwari¹, G.Kalaimathi Priya², S.Grace Mary³, Dr. V.Sambath kumar⁴

¹IP^r M.E(CSE), ShivaniEnggCollegeTiruchirappalli

²Assistant professor(CSE), ShivaniEnggCollegeTiruchirappalli

³Associate professor(CSE), ShivaniEnggCollegeTiruchirappalli

⁴Department of CSE, ShivaniEnggCollegeTiruchirappalli

Abstract: Multimedia data mining is an emergent field, which consists of image mining, video data mining etc. The content based extraction in videos is an important application due to the rapid growth in the video based application. Semantic content is the high level concepts that is the objects and events, for example in a basketball game, the player, a ball are all consider to be an object and the actions takes place throughout the game is consider to be the events. In concept extraction the events and objects are extracted, and it is not that much efficient for the user because the relationship between the objects are missing. Therefore the Bayesian network classifier is used to extract the relationship between the objects. It allows the user to extract the semantic video content more efficiently. The Bayesian network classifier uses an ontology which includes the construction of domain ontology. The VISCOM is used to construct ontology for a given domain and the rule based model is used to define some complex situation more effectively. Using Bayesian Network Classifier improves the extraction process by providing the semantic content with relation between the object.

Index terms: Multimedia mining, Video data mining, Semantic content, Domain ontology, Bayesian network classifier.

I. Introduction

Data mining is the process of digging through and analysing massive sets of data and then extracting the required data, the extraction of buried information from large databases, is a powerful new technology with great potential to help users to focus on the salient information in their data warehouses. Data mining can be classified into many types such as traditional data mining, visual data mining, multimedia data mining etc. Nowadays a data mining plays a vital role in an entertainment field. It can be satisfied by a Multimedia data mining called as Multimedia Miner which extends the Data mining system by handling multimedia data.

Multimedia data mining is the process of extracting multimedia data such as image, audio, video etc. Multimedia data are stored in a database known as multimedia database. Multimedia database sustain large amount of storage capacity, because the multimedia data such as video, image etc have need of gigabytes of storage. In multimedia data mining, storage and search technique need to be incorporated with standard data mining methods. The video data mining is a developing field in a multimedia data mining. The extraction of video content is quite tricky process. The video content is classified into three types such as raw video data, low level features and high level features. The raw video data is considered to be the video attributes. Generally the video is the collection of frames. The video attributes refers to the frame length, frame rate, format etc. The low level features extend the attributes of the raw data which include its colour, audio, texture, shape, motion etc. The last feature is the high level concept known as semantic content which include the objects and events. The annotation tool is used to extract the video content but it extract only the raw video data and the low level feature and also it is a time consuming process. Extracting the low level and raw video data is not that much efficient to the user, therefore the semantic content is extracted i.e the objects and the events. The objects in the videos can be identified by using three methods. First method is manual method, identifying objects manually. Second method is the fully automated method in which the objects are identified by using certain algorithm which include special features such as face recognition, moving object etc. Third method is the semi automatic method which is the combination of both manual and automatic model. The manual method is the time consuming process. The objects in the video cannot be directly identified and extracted. The video contains lots of frames and those frames are considered to be an image and by using those images the objects are identified and extracted. Moving Picture Expert group is also known as Multimedia Content Description Interface. MPEG contains various types of multimedia information and it is classified into six types. First, Basic structure, which include grid layout, time series, multiple view, spatial 2D coordinates, and temporal relation. Second, Colour Descriptor which include dominant colour, scalable colour, colour layout and colour structure. The Dominant colour defines the region consist of small number of colours; to identify the region which has small number of colours the colour quantization is used. Scalable colour uses the bits in the image. Colour layout is in the case

of sequence to sequence matching and it is achieved by using spatial distribution of colour. Colour structure is used to identify the object in case of disconnected region. Third, Texture Descriptor consists of Homogeneous texture which is used in case of similar retrieval of pattern. Edge histogram represents the spatial distribution of edges. Fourth, Shape descriptor represents the shape i.e. the pixels within the frame and it is known as region space.

II. Literature Survey

The various methods for content extraction provide only the raw video data and the semantic video content, that is the objects and the events but the relation between the object is missing and it can be achieved by using the Bayesian network classifier

A. Event detection using video streams

The author[4] has been proposed Event detection using video streams, Video streams are collected by stationary platforms, mobile ground vehicles, or UAV. Video stream applications are widely used in many areas, one of the important application is the surveillance monitoring. The surveillance video is mainly used for the security purpose, to monitor the human, vehicle, and the actions takes place in a particular place. The moving objects from a video stream are extracted by using detection and tracking module. The behavior inference module uses these trajectories, which has only a rough description to recognize and classify object motion.

The first step is to track the detected object through the help of sequence of images and the detected objects is matched to determine the trajectories of the objects and it is done by using the low level features such as object template, color or texture etc. The second step is the object is recognized and classified by using the spatial and temporal properties. The frames are processed by using the above two steps. The moving region is identified by using the current and the previous frame. It is achieved by using the buffered set of the frames that is the processed frames are buffered and stored. The graph representation is done by using the last 5 to 10 frames in the buffered set. The drawback is only the objects are extracted and the event that is the actions between the objects is not extracted.

B. Multiple Event Detection and Representation in Videos

The author [6] developed a novel method to detect Multiple event, it is designed to recognize event involving multiple agents in a video. The novel method is implemented which describes the structure of events from the video and it is composed of temporally related sub events and that is known as the event correlation graph which is used to signifies the frequency of occurrence of reliant sub events. Second the clustering concept is used to group the interrelated sub events. The events with in the cluster are noted as high weighted and the events outside the cluster are noted as low weighted.

The video event graph is represented by the DAC, the vertex represents the sub event and the edges represent temporal relation between two vertices. The main drawback here is the sub event dependency. The event correlation graph is the EWDH. ECG captures all the variations in temporal order of sub events as well as the frequency of occurrence of the chain of sub events in a video. It was implemented in the field of surveillance, railroad monitoring domain and in some kind of meeting video. It detects the events consists of multiple agents but the disadvantage is the resultant output is not in a sequential manner.

C. Automatic Extraction of Salient Object using Colour Histogram Method

The author [8] has been proposed Colour Histogram technique for automatic detection of salient object using shot boundaries. This technique explains salient objects are detected and tracked automatically and the Spatial temporal features are constructed for the video database system. The color histogram based shot boundary detector is used to find the shot boundaries. The salient object are tracked and detected in each shot. Then the object is labeled manually and stored in a database. Two processes are involved in the detection of salient object such as preprocessing and segmentation the preprocessing is done to remove the noise and to flatten the features such as color, audio for healthier segmentation process.

In preprocessing the 3×3 bilateral filtering technique is used to remove the noisy in the video frames, the texture of the video frame is alone flatten and the edges of the frames are maintained without lost. The human vision is approximated by using the Lab color space technique. The undirected graph is constructed in which each node represents the pixel and the edge connects the pair of the pixel. The constructed graph is segmented by using a greedy approach. The sorting is done according to the weight of the edge. If the edge weight between a pair of nodes is less than the maximum edge weight of both of the components, they are merged and maximum edge weight of the new component is updated. If the resulted segmented video is over segmented the subsequent region merging operation is done and the region adjacency graph is constructed.

Face detection is a process in which the image is given as an input to the classifier, if the output is 1 the object is present in the region, if it is 0 the object is not presented in that region. The salient object is then tracked by labeling the object. The advantage in this approach is the object is extracted accurately along with the face detection. The drawback is only the objects are extracted automatically and the events among the objects are not extracted. managing high dimensional data.

D. Event extraction using Genetic algorithm

The author [15] uses Genetic algorithm for Event extraction. OVDAM is developed to extract the object, event and concept automatically and the low-level feature values for objects and relations between objects are determined. The video key frame segments are determined by using N Cut image segmentation algorithm and the segments is classified by using genetic algorithm classifier. In OVDAM use the ontology to extract components of the model. The ontology is a collection of items and their interrelationship. In video data mining, ontology is used to extract the semantic content. Videos are the collection of the set of images, therefore the images should be segmented by using N cut image segmentation algorithm, which it uses a features such as brightness, color etc to segment the image. The segmentation process can be done more than a time to achieve a better result of segmentation. The result of the segmentation process is used for the construction process by using GA approach and it is improved by using fuzziness.

In GA algorithm the Fitness function is used to calculate the accuracy of the result. Event ontology is a major component of video sequences. Temporal information between events like before, after, meets, during, starts is used. Relations between the objects and events such as spatial relation between the objects and object trajectories are used to define the events. The query process in multimedia data mining is quite different from other query processing.

The OVDAM provide the user the ontology supported querying capability Queries containing objects, events, spatial temporal clauses, trajectory clauses, and low level features of objects are answered and the domain ontology is also used in the query processing. The drawback is the domain ontology model is not that much efficient to retrieve the concept.

E. Semantic Event Extraction from Basketball games using Multi Model analysis

The author [26] has been performed a Multi Model method to perform a Event Extraction from basketball games. The basketball games have more events with faster transition of shots and scenes, which may lead to false positive and bias in event detection. Second, the time stopping and unpredictable game break in basketball games make it more challenging for alignment between time tags in web casting text and event moment in broadcast video.

A novel multimodal framework for semantic event extraction for basketball games has been developed. It is achieved by using the following methods, the first method is the web casting text analysis it exploit unsupervised clustering method that automatically detect event from web casting text, the second method is the broadcast video analysis in which the text event time tag in the video stream is mapped to detect event moment and event boundary in the video, the third method is the text video alignment it employ the HMM to detect the temporal event structure and the event boundaries.

The final step is the semantic annotation and indexing in which the nouns and verbs are taken to annotate the event segments and the log file attached is also used which contains the information of time, event type, player and team. It improves the, unsupervised clustering based method to automatically detect event from web casting text, A statistical approach is used to detect event boundary in the video. The game time recognition algorithm is improved to recognize the break time in basketball games. The drawback is, it lacks on personalized video presentation to provide more flexible view ship to meet viewers various preferences, which the output for the personalized video is not in a sequential manner the flow is not depicted in the output.

III. Conclusions

From the literature survey, we discussed about the existing content extraction methods from videos. The semantic content provides the entire knowledge about the video i.e. the object and the events in the corresponding video. Though it is extracted it does not provide the relation between one object over another. To get rid of this problem a Bayesian network classification is used which improve the accuracy of the output text by providing relation over one object with another.

References

- [1]. L.S. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Haratoglu, and M.J. Black, "Visual Surveillance of Human Activity," Proc. Third Asian Conf. Computer Vision (ACCV), vol. 2, pp. 267-274, 1998.
- [2]. M. Petkovic and W. Jonker, "Content-Based Video Retrieval by Integrating Spatio-Temporal and Stochastic Recognition of Events," Proc. IEEE Int'l Workshop Detection and Recognition of Events in Video, pp. 75-82, 2001. 60 IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 1, JANUARY 2013

- [3]. L.S. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Haratoglu, and M.J. Black, "Visual Surveillance of Human Activity," Proc. Third Asian Conf. Computer Vision (ACCV), vol. 2, pp. 267-274, 1998.
- [4]. G.G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia, "Event Detection and Analysis from Video Streams," IEEE Trans. Pattern Analysis Machine Intelligence, vol. 23, no. 8, pp. 873-889, Aug. 2001.
- [5]. S. Hongeng, R. Nevatia, and F. Bremond, "Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods," Computer Vision and Image Understanding, vol. 96, no. 2, pp. 129-162, 2004.
- [6]. A. Hakeem and M. Shah, "Multiple Agent Event Detection and Representation in Videos," Proc. 20th Nat'l Conf. Artificial Intelligence (AAAI), pp. 89-94, 2005.
- [7]. M.E. Do'nderler, E. Saykol, U. Arslan, O. Ulusoy, and U. Gu'du' kbay, "Bilvideo: Design and Implementation of a Video Database Management System," Multimedia Tools Applications, vol. 27, no. 1, pp. 79-104, 2005.
- [8]. T. Sevilmis, M. Bastan, U. Gu'du' kbay, and O. Ulusoy, "Automatic Detection of Salient Objects and Spatial Relations in Videos for a Video Database System," Image Vision Computing, vol. 26, no. 10, pp. 1384-1396, 2008.
- [9]. M. Ko'pru' lu', N.K. Cicekli, and A. Yazici, "Spatio-Temporal Querying in Video Databases," Information Sciences, vol. 160, nos. 1-4, pp. 131-152, 2004.
- [10]. R. Nevatia, J. Hobbs, and B. Bolles, "An Ontology for Video Event Representation," Proc. Conf. Computer Vision and Pattern Recognition Workshop, p. 119, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1384914, 2004.
- [11]. U. Akdemir, P.K. Turaga, and R. Chellappa, "An Ontology Based Approach for Activity Recognition from Video," Proc. ACM Int'l Conf. Multimedia, A. El-Saddik, S. Vuong, C. Griwodz, A.D. Bimbo, K.S. Candan, and A. Jaimes, eds., pp. 709-712, <http://dblp.uni-trier.de/db/conf/mm/mm2008.html#AkdemirTC08>, 2008.
- [12]. Y. Yildirim, "Automatic Semantic Content Extraction in Video Using a Spatio-Temporal Ontology Model," PhD dissertation, Computer Eng. Dept., METU, Turkey, 2009.
- [13]. T. Yilmaz, "Object Extraction from Images/Videos Using a Genetic Algorithm Based Approach," master's thesis, Computer Eng. Dept., METU, Turkey, 2008.
- [14]. Y. Yildirim and A. Yazici, "Ontology-Supported Video
- [15]. Modeling and Retrieval," Proc. Fourth Int'l Conf. Adaptive Multimedia Retrieval: User, Context, and Feedback (AMR), pp. 28-41, 2006.
- [16]. Y. Yildirim, T. Yilmaz, and A. Yazici, "Ontology-Supported Object and Event Extraction with a Genetic Algorithms Approach for Object Classification," Proc. Sixth ACM Int'l Conf. Image and Video Retrieval (CIVR '07), pp. 202-209, 2007.
- [17]. V. Mezaris, I. Kompatsiaris, N.V. Boulgouris, and M.G. Strintzis, "Real-Time Compressed-Domain Spatiotemporal Segmentation and Ontologies for Video Indexing and Retrieval," IEEE Trans. Circuits Systems Video Technology, vol. 14, no. 5, pp. 606-621, May 2004.
- [18]. D. Song, H.T. Liu, M. Cho, H. Kim, and P. Kim, "Domain Knowledge Ontology Building for Semantic Video Event Description," Proc. Int'l Conf. Image and Video Retrieval (CIVR), pp. 267-275, 2005.
- [19]. W. Chen and D.S. Warren, "C-logic of Complex Objects," PODS '89: Proc. Eighth ACM SIGACT-SIGMOD-SIGART Symp. Principles of Database Systems, pp. 369-378, 1989.
- [20]. J.F. Allen, "Maintaining Knowledge about Temporal Intervals," Comm. ACM, vol. 26, no. 11, pp. 832-843, 1983. [20] M.J. Egenhofer and J.R. Herring, "A Mathematical Framework for the Definition of Topological Relationships," Proc. Fourth Int'l Symp. Spatial Data Handling, pp. 803-813, 1990.
- [21]. M. Vazirgiannis, "Uncertainty Handling in Spatial Relationships," SAC '00: Proc. ACM Symp. Applied Computing, pp. 494- 500, 2000.
- [22]. P.-W. Huang and C.-H. Lee, "Image Database Design Based on 9D-SPA Representation for Spatial Relations," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 12, pp. 1486-1496, Dec. 2004.
- [23]. I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet, B. Groszof, and M. Dean, "Swrl: A Semantic Web Rule Language," technical report, W3C, <http://www.w3.org/Submission/SWRL/>, 2004.
- [24]. "Prote'ge' Ontology Editor," <http://protege.stanford.edu/>, 2012. [30] "Jena: A Semantic Web Framework," <http://www.hpl.hp.com/semweb/>, 2012.
- [25]. C. Xu, J. Wang, K. Wan, Y. Li, and L. Duan, "Live Sports Event Detection Based on Broadcast Video and Web-Casting Text," MULTIMEDIA '06: Proc. 14th Ann. ACM Int'l Conf. Multimedia, pp. 221-230, 2006.
- [26]. Y. Zhang, C. Xu, Y. Rui, J. Wang, and H. Lu, "Semantic Event Extraction from Basketball Games Using Multi-Modal Analysis," Proc. IEEE Int'l Conf. Multimedia and Expo (ICME '07), pp. 2190- 2193, 2007.