XML & Fuzzy Based Approach to Video Annotation

Sampada S. Wazalwar, Dr. Latesh G. Malik

Abstract - Video annotation technique is proved to be very important technique for analysis and retrieval of video contents. Video annotation process is complicated because it requires a large database, memory & time for processing. In this paper we have discussed different video annotation techniques & proposed a new technique for video annotation. Proposed system uses XML structures to save the annotations & fuzzy logic for labeling. Partial results are mentioned in the paper.

Index Term - *Frames, segmentation, video annotation, visual features, XML structures.*

————————— ◆ —————————

1. INTRODUCTION

An efficient technique is needed to handle & manage large scale of multimedia archive. It is required urgently to make this unstructured data retrievable & accessible with great ease. Video annotation is nothing but adding the descriptors to the contents of video. It is adding background information about the video [13]. The need for multimedia content analysis techniques in smart environment has dramatically increased because of the invention of smart phones & smart TV [12]. Researchers are continuously working on video annotation techniques & try to find an efficient technique which will resolve the problem of storage & time. Question arise is how to save these annotations for each video as it creates a large database. We have used XML file structures to save the annotations as it becomes easy to retrieve, modify & compare the annotations in XML structures.

Further paper is organized as, Section 1 gives the brief information about different techniques previously proposed, Section 2 gives new proposed method and Section 3 specifies results.

2. LITERATURE SURVEY

Video annotation is gaining an importance because of continuously increasing multimedia archive. In old days annotation was done manually by navigating through various frames of videos. Different software's which are used for manual annotation includes ANVIL, VATIC etc. As manual annotation requires intensive labor cost, need for automatic video annotation technique arises. Different techniques proposed by different authors for automatic video annotation are highlighted in following section.

2.1 Seamless annotation & enrichment of mobile captured video streams

In this, mobile captured video stream annotation technique is proposed. Real time automatically generated tags of currently captured videos are provided by the system. It works as a Client Server system in real-time. [1].

2.2 Semi automatic video content annotation

In this paper [2] video content description ontology is proposed, and a video scene detection algorithm which joints visual and semantics is proposed to visualize and refine the annotation results. Accordingly, various video index strategies are proposed to describe video content.

2.3 Automatic video annotation via hierarchical topic trajectory model

In this paper [3] Hierarchical topic trajectory model (HTTM) is proposed for acquiring a dynamically changing topic model which represents relationship between video frames & associated text labels. Method incorporates 1) Co-occurrences among visual & text information and 2) Temporal dynamics of videos.

2.4 Approach to detect text & caption in videos

In this method, a corner based approach is given to detect text and caption from videos. Algorithm is given to detect moving captions, the motion features are extracted by optical flow, are combined with text features to detect the moving caption patterns [4].

2.5 Automatic annotation of Web videos

In this paper [5], a robust moving foreground object detection method is proposed for videos captured in uncontrolled environment followed by the integration of features collected from heterogeneous domains. Two major steps are given, 1) Consensus Foreground Object Template (CFOT). In this Foreground region estimation is done using Scale-invarient feature transform (SIFT) [16] and 2) Object detection.

2.6 Ontology-based automatic video annotation technique

In this paper [6], an automatic video annotation technique is presented for smart-TV environment which uses LSCOM ontologies [8], MPEG-7 visual descriptors [7], domain ontology & high level concept extraction.

2.7 Generic Framework for Video Annotation via Semi-Supervised Learning

In this paper [9], a novel approach based on semi-supervised learning is proposed. System is having three features: 1) Event recognition 2) Event Localization and 3) Semantic search & navigation. A Fast Graph-based Semi-Supervised Multiple Instance Learning (FGSSMIL) algorithm & de_noising technique is used to detect positive frames.

2.8 Fast Semantic Diffusion for Large-Scale Context-Based Image & Video Annotation

In this paper [10] concept of semantic diffusion is given, graph diffusion formulation is used to improve concept annotation scores. Input to the system is the prediction scores of classifiers or manual labels tagged by web users. Semantic graph is applied in Semantic Diffusion to refine concept annotation results.

3. PROPOSED METHOD FOR VIDEO ANNOTATION

Video annotation is a very complicated process & it gets more complex with the size of video. Considering all the methods previously proposed by different authors, their respective advantages & disadvantages, a new method for video annotation is proposed here. Proposed video annotation method

basically uses XML structures to store the annotation & finally fuzzy logic is used do the annotation of testing video.

System basically starts with manual annotation for training. A video is taken as input, then for each frame YCbCr value of each pixel is taken. As one video consist of thousands of frames so we will go for key frame detection. Key frame detection is done on the basis of threshold which is set to calculate difference between adjacent frames. After key frame detection, feature values of each frame are calculated using CSD & EHD, system asks us to annotate each key frame & these annotations are saved in XML structures. After all the annotation are saved & training is over, a video for automatic annotation is given & system will annotate the video automatically by comparing it with the feature value & respective annotation in XML file structure.

4. EXPERIMENTAL RESULTS

It takes time to process the video, thus while running the system a short video of tornado (natural disaster) 29 seconds is processed which consist of 703 frames. From these 703 frames key frames are detected based on some threshold value of comparison. If the threshold value is taken as 500 then after applying key frame detection algorith, 6 key frames from 50 frames are detected approximately. By observing the value of key frames length of shot can be detected.

In this way XML file is updated for each video. Whenever a new video is given for annotation, its CSD & EHD values will be calculated & will be compared with the contents of XML file, considering some difference limited with threshold similar annotations will automaticaly be given to the frames of new video. For more accuracy fuzzy logic is used to label the frames based on some conditions.

After key frame detection, CSD i.e. color structure descriptor is applied on each frame & feature values are calculated. 192 values are calculated for each frame.
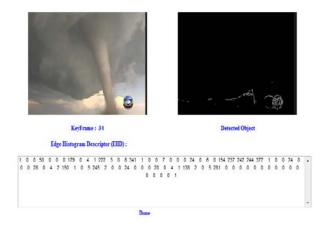
Figure 1. Key frame detection process

Figure 3: 80 bin vector EHD

After applying CSD & EHD on each key frame, these values are stored in XML files. Respective annotations of key frame is also saved in same XML file. Figure below shows the annotation i.e. 'Tornado' applied to key frame 1 of video T5.mp4 based on the values of CSD & EHD. In this way annotations are saved during training, and then are used for automatic annotation of new videos while testing.

Figure 2. Detected shots

To represent the texture of objects in frames a descriptor for edge distribution in the image is used which is called as Edge Histogram Descriptor (EHD) [11]. It expresses only the local edge distribution in the image. The edges in the image are categorized into 5 types: vertical, horizontal, 45-degree diagonal, 135-degree diagonal and non-directional edges result of the texture extraction by EHD is an 80-bin feature vector and each bin has its own semantics in terms of location and edge type.

After applying EHD on each key frame we get 80-bin feature vector for each key frame. Figure shows the values obtained after applying EHD on single frame.

Figure 4: XML file created

5. CONCLUSION

Looking towards different techniques that are highlighted, each technique is having some advantages and disadvantages. Different technique works in different environment with different level of accuracy and timing constraints. There is need to improve the existing methods in terms of accuracy & time, further it needs to handle a large amount of database. In proposed method we are dealing with the problem of time consumption efficiently with the use of XML structures & fuzzy logic is used to improve

the accuracy. Combination of XML structures & fuzzy logic contributes towards new efficient video annotation technique.

REFERENCES

[1] Motaz El-Saban, Xin-Jing Wang, Noran Hasan., "Seamless annotation and enrichment of mobile captured video streams in real time", IEEE 2011.

[2] Xingquan Zhu, Jianping Fan, Xiangyang Xue, Lide Wu and Ahmed K. Elmagarmid, " Semi-automatic video content annotation", Proceeding of Third IEEE Pacific Rim Conference on Multimedia, pp. 37-52, 2008.

[3] Takuho Nakano, Akisato Kimura, "Automatic video annotation via hierarchical topic trajectory model considering cross-modal correlations" IEEE 2011.

[4] Xu Zhao, Kai-Hsiang Lin, Yun Fu, "Text From Corners: A novel approach to detect text and caption in videos", IEEE transaction on image processing, vol..20, No. 3, MARCH 2011.

[5] Shih-Wei Sun, 2011.Yu-Chiang Frank Wang, "Automatic annotation of web videos", IEEE 2011.

[6] Jin-Woo Jeong, Hyun-Ki Hong, and Dong-Ho Lee, "Ontology-based Automatic Video Annotation Technique In Smart TV Environment", IEEE Transactions on Consumer Electronics, Vol. 57, No. 4, November 2011.

[7] ISO/IEC 15938-5 FDIS Information Technology: MPEG-7 Multimedia Content Description Interface - Part 5: Multimedia Description Schemes. 2001.

[8] L. Kennedy and A. Hauptmann "LSCOM lexicon definitions and annotations version 1.0", DTO Challenge Workshop on Large Scale Concept Ontology for Multimedia. ADVENT Technical Report #217- 2006-3, Columbia University, March 2006.

[9] Tianzhu Zhang, Changsheng Xu, Guangyu Zhu, "A Generic Framework for Video Annotation via Semi-Supervised Learning", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 14, NO. 4, AUGUST 2012.

[10] Yu-Gang Jiang, Qi Dai, Jun Wang, Chong-Wah Ngo, "Fast Semantic Diffusion for Large-Scale Context-Based Image & Video Annotation", 3080 IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 21, NO. 6, JUNE 2012.

[11] Chee Sun Won, Dong Kwon Park, and Soo-Jun Park., "Efficient use of Mpeg- edge histogram descriptor", ETRI Journal 2002; vol. 24, no. , pp.23-30.

[12] MoonKoo Kim, JongHyun Park, "Demand forecasting and strategies for the successfully development of the smart TV in Korea" *Proceeding of* International Conference on Advanced Communication Technology, pp.1475-1478, 2011.

[13] http://www.youtube.com/t/annotations_about