INTERNATIONAL JOURNAL OF RESEARCHES IN BIOSCIENCES, AGRICULTURE AND TECHNOLOGY

www.vmsindia.org

A REVIEW: SMART VIDEO FORENSICS SYSTEM USING SEMI-AUTOMATED ACTIVITY FILTER FOR SECURE, EFFICIENT FOOTAGE GENERATION

Ajay B Gadicha[1] and M. V.Sarode[2]
[1] P. R. Patil College of Engineering & Technology Amravati.
[2] Jagadambha College of Engineering & Technology,Yavatmal

Abstract:
State of the art computer vision algorithms detect objects in real-time video and determine basic activity information such as object type (human, vehicle, etc.), object trajectory, and interactions with other objects. This information is encoded as a light-weight stream of activity-based meta-data that can be distributed over a network, analyzed in real-time to detect security threats, and archived for forensic analyses. To enable flexible forensic analysis, an activity query schema is described that is designed specifically for physical security applications. It is based on a traditional Boolean binary tree approach, but with specific modifications for activity queries. This schema is implemented through an intuitive user interface that enables even unsophisticated users to create complex activity queries. The described mechanism of separating the complex computer vision process from activity queries means that forensic analysis can be achieved at extremely high-speeds - database query time instead of video processing time. Finally a case study is presented concerning the use of this mechanism to optimise the performance of a real-time security system through iterative forensic analysis.
Keywords: Video Security, video sequence, visual cryptography

Introduction:

These days Digital Video Technology is growing at a rapid pace and due to advancement in technology it becomes very easy to record many videos. A bulk of digital contents such as news, movies sports, and documentaries are used on the internet. Moreover the need of surveillance has increased significantly due to increase in the demand of security gazettes especially after 9/11(Sept 11, 2001).Many video Cameras are found at public place, public transport, bank, airport, etc resulting in the large amount of information which is difficult to be processed in the real time. Furthermore, storage of huge amount of video data is not that easy. It's very important to quickly retrieve and browse huge amount of data effectively because end user wants to get important aspect of data. To solve this problem proposed system works very significantly to generate the secure, efficient footage from captured or original sequence of videos.. While most existing video summarization approaches aim to identify important frames of a video from either a global or local perspective, I propose a top-down approach consisting of scene identification and scene summarization. For scene identification, it represents each frame with global features and utilizes a scalable clustering method. Proposed system then formulate scene summarization as choosing those frames that best cover a set of local descriptors with minimalredundancy.

PROCESSING AND ANALYSIS OF VIDEO:

The goals of the processing and analyzing is to identify the more important segments of video and to assign dominance to each segment, and to Segment video into larger and meaningful units beyond the shots, such as scenes or events, so that video poster can be constructed for each individual unit.

a) Segmentation of Video into Shots :

To analyze the content of a given video sequence, we need to first segment the sequence into individual shots. Such analysis is computationally expensive when full resolution video frames are used because of the huge data size. In addition, most video sequences are captured and stored in various standard compression formats. For compressed video in motion JPEG or MPEG-1 standard compression formats, we use the algorithms in [14] based on dc images extracted directly from compressed video to segment a video into shots.

b) Visual and Temporal Classification of Shots :

A first step toward video processing is the semantic labeling of video shots. We want to associate with each shot , a label , which provides description of the content of the shot. For example, a shot of a news anchor person could be labeled as "news anchor," "news room," or "man behind table." Such semantic description, however, is difficult to derive at by automatic means. Instead, we make use of a direct consequence of the montage presentation of film—that parallel or simultaneous events are depicted one after another, with substantial repetitions of shots with similar contents. To capture such repetitions, clustering of video shots can be performed.

c) Finding Story Units in Video:

A shot is the fundamental unit of a video production, but a shot by itself does not have much meaning in a story; instead, a basic unit of the story can be a scene, which is comprised of a number of interrelated shots unified by location or dramatic incident. In a given scene, we often find multiple parties coexist, and the shots of these parties are juxtaposed and linked together with multiple shots of the same party. Shots from different scenes are not juxtaposed, except at the transition from one scene to the next. Because of the intense interactions between shots in a scene, the use of a label sequence can be used to segment a video into large logical units, called the story units, each of which closely approximates a scene.

d) Detecting Meaningful Temporal Events:

In addition to segmenting video into story units or scenes, it is also advantageous to recognize common temporal events within a scene. The use of label sequences can be used to recognize dialogue and action events. Using the degree of repetition or the lack of repetition in a subsequence of labels, we propose formulations in [18] that translate common filmmaking rules into their analytical equivalents. Based on these formulations, algorithmic and systematic methods are then derived to analyze video for the extraction of video segments that resemble the characteristics of the temporal event models. Consequently, we are able to classify a video subsequence into one of three categories: dialogue, action, and others. A dialogue refers to actual conversation or a conversational like montage presentation of two or more concurrent processes. In motion pictures, to present two or more processes as simultaneous and parallel, they have to be shown one after the other in a sequential temporal order. Dialogue events are characterized by intense interactions between two dominant parties, possibly interspersed by establishing a shot or shots of other parties. Models are constructed to capture such repetitive nature of two dominant shots and to also accommodate "noise" labels.

An action is characterized by a progressive presentation of shots with contrasting visual contents to express a sense of fast movement and achieve strong emotional impact. Such a sequence of shots is most likely found in a scene where there is a rapid unfolding of the story, where the camera is not fixated at a location or event, and where there is a significant amount of object movement. In such a sequence, there is typically little or no recurrence of shots taken from the same camera, of the same person, or background locale. A model is then constructed to capture the lack of label repetitions.

OBJECTIVES:

This research work is focusing on various methodologies like object detection, shot boundary detection with key frame extraction, silhouette detection.

This research work will look to meet some of following objectives:

- To improve the accuracy of video shot detection so that it generates efficient footage for investigation.
- To create summary of video to enable a quick browsing of data from large video database.
- To incorporate the real time conditions for video summarization
- To develop a high speed cryptographic algorithm for video encryption.

To design the prototype level implementation of the proposed model, so that it can be commercialized.

LITERATUER REVIEW

Extensive research work has been carried out by academic and industries to summarize the content of videos so that to generate the cluster scene from captured or real time videos.

Ravi Mishra et al [2014] proposed a Comparative study of block matching algorithm and dual tree complex wavelet transform for shot detection in videos. Mishra presents a comparison between the two detection methods in terms of various parameters like false rate, hit rate, miss rate tested on a set of different video sequence [1].

Zhe Ming Lu et al [2013] proposed a Fast Video Shot Boundary Detection Based on SVD and Pattern Matching. It is based on segment selection and singular value decomposition (SVD) [2]

Donate et al [2010] presented Shot Boundary Detection in Videos Using Robust Three-Dimensional Tracking. The proposal is to extract salient features from a video sequence and track them over time in order to estimate shot boundaries within the video [3].

To perform multifractal analysis on video **Goran J. Zajić et al [2011]** proposed video shot boundary detection based. Low-level features (color and texture features) are extracted from each frame in video sequence then are concatenated in feature vectors (FVs) and stored in feature matrix [4].

Sandip T et al [2012] proposed Key frame Based Video Summarization Using Automatic Threshold & Edge Matching Rate. Firstly, the Histogram difference of every frame is calculated, and then the edges of the candidate key frames are extracted by Prewitt operator [9].

Zhang, L et al. [2014] proposed a novel algorithm to summarize one-shot landmark videos. The algorithm can optimally combine multiple unedited consumer video skims into an aesthetically pleasing summary. In particular, to effectively select the representative key frames from multiple videos, an active learning algorithm is derived by taking advantage of the locality of the frames within each video. Toward a smooth video summary, we define skimlet, a video clip with adjustable length, starting frame, and positioned by each skim. Thereby, a probabilistic framework is developed to transfer the visual cues from a collection of aesthetically pleasing photos into the video summary. The length and the starting frame of each skimlet are calculated to maximally smoothen the video summary. At the same time, the unstable frames are removed from each skimlet. Experiments on multiple videos taken from different sceneries demonstrated the aesthetics, the smoothness, and the stability of the generated summary [13].

By using triangle model of Perceived Motion Energy (PME) motion patterns T. Liu, H. J. Zhang, and F. Qi [2003] proposed the frames at the turning point of the motion acceleration and motion deceleration are selected as key frames. The key-frame selection process is threshold free and fast and the extracted key frames are representative measures are computed for each descriptor and combined to form a frame difference measure. Fidelity, Shot Reconstruction Degree, Compression Ratio qualities are used to evaluate the video summarization [10].

Xingquan Zhu, XindongWu, et. al. proposed [2004] a hierarchical video summarization strategy that explores video content structure to provide the users with a scalable, multilevel video summary. First, video-shot- segmentation and keyframe-extraction algorithms are applied to parse video sequences into physical shots and discrete keyframes. Next, an affinity (self-correlation) matrix is constructed to merge visually similar shots into clusters (supergroups). Since video shots with high similarities do not necessarily imply that they belong to the same story unit, temporal information is adopted by merging temporally adjacent shots (within a specified distance) from the supergroup into each video group. A video-scene-detection algorithm is thus proposed to merge temporally or spatially correlated video groups into scenario units. This is followed by a scene-clustering algorithm that eliminates visual redundancy among the units [5].

Fernando Barreiro-Megino, Jos´e M. et al.[2009] presents an image to video adaptation system that transmodes images in order to be viewed on small displays without a significant loss of information. The followed approach is to automate the process of manual browsing and zooming through an image by simulating the movement of a virtual camera [11].

1. PROPOSED WORK:

By considering the following scenario the simplest method of parsing video data for efficient browsing, retrieval, and navigation is segmenting the continuous video sequence into physical shots and then selecting a constant number of keyframes for each shot to depict its content. Unfortunately, since the video shot is a physical unit, it is incapable of conveying independent scenario information. In proposed system cluster scene then encrypted using visual cryptography algorithm then sends to the base station.

3.1 Proposed Methodology

Video Summarization: Initial work would be based on application of video forensics to video summarization. In this process, proposed system has algorithms like object detection, shot boundary detection with key frame extraction, silhouette detection, background modeling under various illumination conditions so that the video can be summarized so as to provide the highlights of the video.

Key frame extraction: would be used to extract only the frames where movements are present while Object detection would be used to inform the user about the type of object present in the video. Visual Cryptography: The summarized video would be divided into multiple parts using the concept of visual cryptography, so that any unwanted intruder must not be able to get the contents of the video. This is the first of two steps of securing the video. Proposed system mainly focuses on to send the video securely to the base station Video Encryption Each of the shares generated by the Visual Cryptosystem would be then processed to perform video encryption. Fast encryption techniques like RC6, MES Version 2 and others can be used to perform real time video encryption. The shares would be encrypted by different algorithm and different keys. Meta data about each key and each algorithm would be stored inside a MASTER FRAME which would be separately encrypted using a pre-decided custom encryption technique. TTLS for Secure Communication: Once the entire video is summarized, visually and analytically encrypted. Proposed System uses a tunnel based security protocol for communicating the video from the transmission end to the reception end with minimum possibility of hacking into the system.

Result Analysis :

Once the system is developed, the analysis would be performed based on security, accuracy and quality parameter.

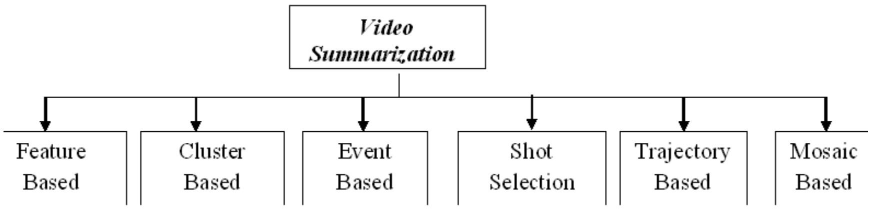

Figure : 1 Hierarchical structure of Video Summarization

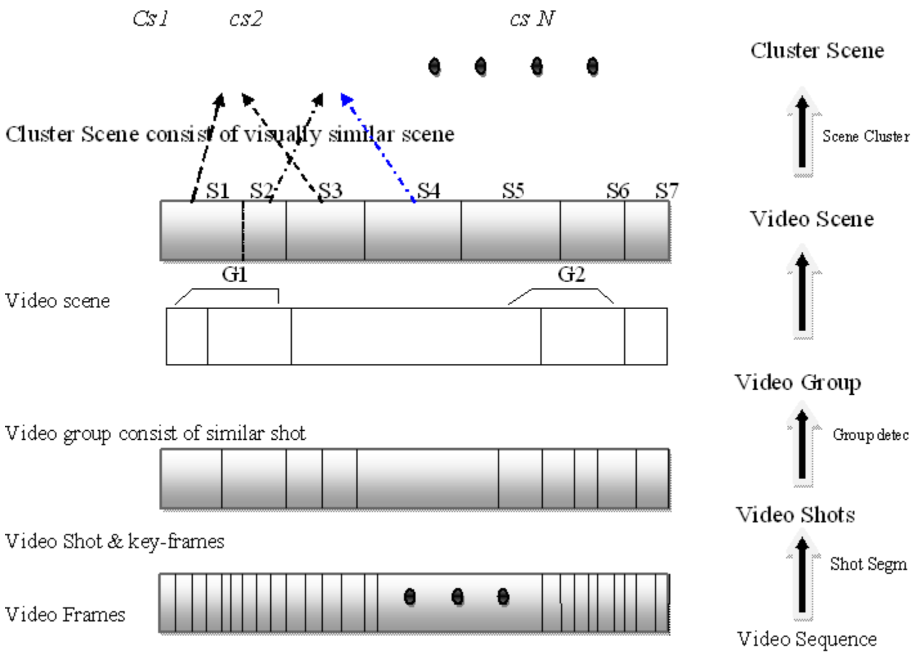

Figure 2. Pictorial video content structure

Conclusion:

Video Forensics is the scientific examination, evaluation and comparison of video in legal matters, whether criminal or civil in nature. In some circumstances, it may not be completely clear what is actually taking place with respect to an event captured on video. Many of researchers have worked on easily convert and enhance any type of surveillance video into an easily viewable format that can be used back as DVD or video clip. This work will focus on how the key frame extraction from the video summarizations and provide the security so that the required video reaches to base stations or destination as a secured content. For this proposed system contains semi-automated filter i.e.(taking input from the user) for the detection of suspect or object in any circumstances.

References:

1. Ravi Mishra ,S.K.Singhai,M. Sharma "Comparative study of block matching algorithm and dual tree complex wavelet transform for shot detection in videos" International Conference On Electronic system, signal processing and computing technologies(ICESC), pp 450-455 Jan 2014.
2. Zhe Ming Lu and Yong Shi "Fast Video Shot Boundary Detection Based on SVD and

Pattern Matching" IEEE TRANSACTIONS ON IMAGE PROCESSING Volume: 22, Issue: 12, pp 5136 – 5145, Dec. 2013.

3. Arturo Donate and Xiuwen Liu Member IEEE, "Shot Boundary Detection in Videos Using Robust Three-Dimensional Tracking" IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 64 – 69 , 13-18 June 2010.

4. Goran J. Zajić, Irini S. Reljin, Senior Member, IEEE, and Branimir D. Reljin, Senior Member, IEEE, "Video Shot Boundary Detection based on Multiracial Analysis" IEEE JOURNAL ON IMAGE PROCESSING, VOL. 3, NO. 2, pp 1403-1407 , 2011.

5. Xingquan Zhu, XindongWu, Jianping Fan, Ahmed K. Elmagarmid ,Walid G. Aref "Exploring video content structure for hierarchical summarization" SPRINGER MULTIMEDIA SYSTEMS,10: Digital Object Identifier Volume:10 pp 98–115, 2004

6. Sergio Benini, PierangeloMigliorati, and Riccardo Leonardi , "Statistical Skimming of Feature Films" Hindawi Publishing Corporation INTERNATIONAL JOURNAL OF DIGITAL MULTIMEDIA BROADCASTING Volume 2010 , Article ID 709161.

7. Genliang Guan And Zhiyong Wang,Shaohui Mei,Max Ott, Mingyi He,David Dagan Feng "A Top-Down Approach for Video Summarization" ACM TRANSACTIONS ON MULTIMEDIA COMPUTING, COMMUNICATIONS, AND APPLICATIONS, Vol. 11, Issue 1, August 2014.

8. Minerva M. Yeung, Member IEEE, and Boon-Lock Yeo, Member IEEE "Video Visualization for Compact Presentation and Fast Browsing of Pictorial Content" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 7, Issue. 5, pp 771 - 785, OCTOBER 1997.

9. Mr. Sandip T. Dhagdi, Dr. P.R. Deshmukh "Key frame Based Video Summarization Using Automatic Threshold & Edge Matching Rate" International Journal of Scientific and Research Publications, Volume 2, Issue 7, pp 1-12 ,July 2012.

10. T. Liu, H. J. Zhang, and F. Qi, "A novel video key frame extraction algorithm based on perceived motion energy model," IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, Vol. 13, Issue. 10, pp 1006-1013, Oct 2003.

11. Fernando Barreiro-Megino, Jos´e M. Mart´ınez "Visual Tools for ROI Montage in an Image2Video Application" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 19, NO. 12, pp 1927-1932, DECEMBER 2009.

12. Vijay Badrinarayanany, Fabio Galassoy and Roberto Cipolla "Label Propagation in Video Sequences" IEEE TRANSACTIONS ON SYSTEM AND CIRCUITS V0LUME:10 ISSUE:4 , pp 11-18 , 2010.

13. Zhang, L, Gao, Y Hong, R. Hu, Y "Probabilistic Skimlets Fusion for Summarizing Multiple Consumer Landmark Videos" IEEE Transactions on Multimedia, Volume:17, Issue: 1,pp 40 – 49, 23 Jan 2015.

14. Costeira J, Kanade T (1994)A multi-body factorization method for motion analysis. Technical Report, CMU-CS-TR-94-220, Department of Computer Science, Carnegie Mellon University, Pittsburgh

15. Shi J, Malik J "Normalized cuts and image segmentation" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 8, pp 888-905 , AUGUST 2000