Research Journal of Applied Sciences, Engineering and Technology 10(11): 1316-1321, 2015
DOI: 10.19026/rjaset.10.1828
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: February 13, 2015 Accepted: March 7, 2015 Published: August 15, 2015

Research Article

A Framework for Automatic Video Surveillance Indexing and Retrieval

Fereshteh Falah Chamasemani, Lilly Suriani Affendey, Norwati Mustapha and Fatimah Khalid
Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia

Abstract: The manual search through the surveillance video archives for a specific object or event is very time-consuming and tedious task due to the large volume of video data captured by many installed surveillance cameras. Therefore, the solution to accelerate and facilitate this process is to design an automatic video surveillance with the efficient and effective video indexing, video data model, query formulation and language, as well as visualization interface. There are many challenges, for developing a powerful query processing module, formulating complex queries and selecting suitable similarity matching strategy to detect any abnormality based on semantic content of the video using various query types. This study presents a novel video surveillance indexing and retrieval framework to cope with the above challenges. The proposed framework consists of three main modules i.e., pre-processing, query processing and retrieval processing. Moreover, it supports an efficient search and actively refines the retrieval result by formulating various query types including: query-by-text, query-by-example and query-by-region.

Keywords: Data modeling, query formulation, query processing, video indexing, video surveillance, video surveillance retrieval

INTRODUCTION

Nowadays, many surveillance cameras have been installed in public places like banks, airports, parking lots, offices, hospitals and shops to increase the security by real time monitoring of human activities as well as capturing and recording this information for future analysis. Hence, video surveillances mainly supports two applications domain:

- Real time monitoring of environment and generating alarm to prevent dangerous situations or threats by predicting recognized abnormal activities and events.
- Investigating and retrieving specific events (action of object such as abandoned luggage or person entering the forbidden zone) or object (e.g., vehicle, person and luggage) of interest for the after-the-fact activities as evidence forensics (Le *et al*., 2010; Şaykol *et al*., 2010).

Although, many research have been carried out in automatic event recognition (Benabbas *et al*., 2011; Hampapur *et al*., 2005), crowd analysis (Conte *et al*., 2010; Xu and Song, 2010), object detection and tracking on video surveillance (Kim *et al*., 2011), only few works have been dedicated to access the relevant video segment based on the user's intentions for the after-the-fact activities (Chamasemani and Affendey, 2013). In addition, these huge volumes of surveillance video contents bring us many challenges in managing and retrieving useful information efficiently and effectively. Moreover, manually searching the surveillance videos for specific events or objects by security staffs is a tedious and time consuming task which is almost becoming infeasible. Therefore, a practical solution to this problem is to quickly retrieve relevant segment of video based on user query by utilizing semi-automated or automated video retrieval and browsing application (Calderara *et al*., 2006; Hampapur *et al*., 2007; Hu *et al*., 2007). However, a robust video surveillance system should be equipped with powerful data modeling (to extract appropriate features, organizing and storing them in video archive) and retrieval techniques to provide sufficient facilities for detecting specific events or objects in video archives. In addition, developing an efficient query processing algorithm is also essential for accessing the video surveillance archives.

This study focuses on the problem of indexing as well as retrieval of objects-of-interest or events within the stored content of the video surveillance archives. Therefore the main contribution of this study is a novel video surveillance indexing and retrieval framework for forensic investigation (after-the-fact activities retrieval). The three main modules of the proposed framework, namely, pre-processing, query processing and retrieval processing enable the user to formulate various query

Corresponding Author: Fereshteh Falah Chamasemani, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia

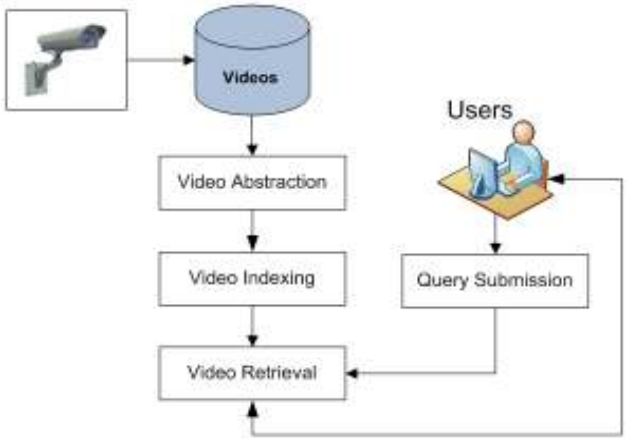

Fig. 1: Architecture of proposed video surveillance indexing and retrieval framework

types (including: query-by-text, query-by-example, query-by-region) and allows active interactions with the retrieval model. Our framework is different from developed framework in (Le *et al*., 2009) since:

- Ours is equipped with its own video analysis module.
- Videos abstraction are used for indexing process so it decreased the processing time and operational cost during retrieval process as shown in Fig. 1.

LITERATURE REVIEW

Stringa and Regazzoni (1998, 2000) proposed a real time video shot detection, indexing and retrieval system. Their system is one of the first real time content-based video surveillance retrieval and indexing systems. It is used to retrieve the detected abandoned/lost luggage in subway station either from its related frame (where the detected lost luggage has been left by a person) or video shot (the last frame among 24 frames of each shot contains the detected lost luggage). Their system stored the frame of the detected lost luggage to use it in the future for retrieving the similar instances of the lost luggage based on features such as: color, shape, texture, 3-D position, movement and compactness. Therefore this system supported only textual query for pre-defined event (lost luggage).

Video Content Analyzer (VCA) was developed by Lyons and his colleagues with five main components including background subtraction, object tracking, event reasoning, graphical user interface plus indexing and retrieval (Lyons *et al*., 2000). VCA extracts and classifies the content of video into people and objects. VCA are also able to recognize event such as person depositing/picking up object, person entering/leaving scene, merging and splitting. Although VCA's graphical user interface provides the facility for retrieving video sequences; nevertheless it is based on only given event queries.

Lee *et al*. (2005) developed an object-based video surveillance retrieval system. Their system equipped with the specific user interface as a search/browse tools from indexed surveillance videos. This interface allows its user to search, browse, filter and retrieve simple events such as presence of persons in a given camera in a specific time or even from other cameras. In fact it can retrieve the presence of suspicious person which appeared in multiple camera viewpoints while this event has already been indexed.

Jung *et al*. (2001) designed an efficient event retrieval system for traffic surveillance based on motion trajectory of the moving objects. Access to this moving object in the semantic level is possible by using a generated motion model as an index key which stored in database (Jung *et al*., 2001). The specific feature of object is used for indexing and searching purpose at different semantic level. Although, their searching module supports different queries (query by sketch, query by example and query by weighting parameter) based on the object trajectory information; it fails to process complex and textual queries.

IBM smart surveillance was developed for video indexing and retrieving by focusing on video data model (Hampapur *et al*., 2005, 2007). Although, this system was successful to detect moving objects, track group of objects, classify objects and event of interest, only predefined events can be queried and processed.

Hu *et al*. (2007) proposed a semantic-based video retrieval framework for video surveillance based on object trajectories. Objects trajectories are extracted from tracked object in the scene then object activity models (at both low level and semantic level retrieval) are hierarchical clustered and learnt from these object trajectories using their spatio-temporal information. Finally several descriptions are added to the activity model to properly index data. Their framework supports query by sketch-based trajectories, query by multiple objects and query by keyword. However, the semantic level retrieval of their framework supports only few activities such as turn left/right/south/north and normal/high/low speed.

SURVIM is a developed data model for online video surveillance which supports different abstraction levels (Durak *et al*., 2007). The framework of SURVIM includes these four main modules: data extractor, video data model, query user interface and query processing. Users of SURVIM are able to retrieve video segmented based on different query types (query by semantic, spatial, size-based, trajectory and temporal). SURVIM did not extract low level feature of objects such as color, shape and velocity; therefore it suffers from inability in successful object classification. Their model also failed to process those queries with incomplete indexing.

Visual Surveillance Querying Language (VSQL) was proposed by Şaykol *et al*. (2005) with the main focus on semantic and low-level features for surveillance video retrieval supports only query-by-text. They developed their query language to support scenario-based query processing system which provides a mechanism for an effective offline inspection (Şaykol *et al*., 2010). The two main drawbacks of their system are: firstly, they performed exact matching since during indexing phase the event are recognized and object are detected and tracked. Therefore, their system failed to process those queries with incomplete indexing. Secondly, the users are restricted to formulate their queries from limited set of predefined events and scenario (in their system scenario is specified as a sequence of events arranged temporally and enriched with object-based low level features).

Le *et al*. (2008, 2009, 2010) developed a general framework for video surveillance indexing and retrieval. Their framework equipped with a Structured Query Language (SQL) to retrieve surveillance video at both event and object level. The retrieval process can be done based on query by text, query by example and query by region. In their system the simple specified events plus interval time of their relations construct the composed events. They developed their framework based on this assumption that the incoming videos are partially indexed. An external video analyzing module was responsible to index the content of video by performing object detection, object tracking and event recognition.

Nam and his colleagues proposed a data model for human activity recognition that support complicated activity by combining a set of basic activities (Nam *et al*., 2013). They used activity labels to defined validity or invalidity of activity combinations and restricted the human activity into symmetric or asymmetric.

Although several works have been dedicated for retrieving object and event on video surveillance archives, still there are many challenges need to be fulfilled. Designing a powerful query processing module and formulating complex query comprising the combination of information and spatial-temporal relations of objects and events is not easy task. Furthermore, developing a similarity matching strategy which enables to match various types of queries and video index is another open problem.

OVERVIEW OF THE PROPOSED METHODOLOGY

Figure 2 illustrates structure and functions of our proposed indexing and retrieval framework. This framework is based on these three main modules: pre-processing, query processing and retrieval processing. Furthermore, it is designed as an effective, efficient and convenient means for automatic object and event indexing as well as retrieval from video surveillance archives which works on both low and semantic levels. However, the proposed framework contains the entire processing steps needs to accomplish retrieval task. The following subsections give a quick review of each framework modules.

Pre-processing module: Surveillance cameras captured raw video, then compressed and stored it into video database based on their location and time information. Then, these stored videos are abstracted in order to decrease processing time and computational cost for performing further tasks such as video indexing, query processing and retrieval.

Video abstraction: Video abstraction is an automatic way for extracting important information from large-scale video which speeds up video indexing and retrieval by avoiding from performing unnecessary and redundant information. Dynamic video skimming and static video summary are two types of video abstraction. Video skimming is constructed from a collection of image sequences with their related audio; in fact video skimming is a brief representation of the source video. While, static video summary is the simplest ways for providing an abstracted video by extracting and selecting the representative video frames which called keyframes. The common way for extracting these keyframes is based on semantic level of a video. However, determining the proper and informative keyframe is not easy tasks (Jiang and Qin, 2010; Sabbar *et al*., 2012). Moreover, most of moving objects in real video surveillance application are important since their presence can be referred as evidence in case of crime occurrence.

In video surveillance the moving objects are the most informative and representative part which appear in video frames. Therefore, the efficient video abstraction should contain these three main characteristics, first, it should be sufficiently short and compact; second, it should include necessary elements of moving objects which appeared in the original video; third, the temporal order of moving object should be preserved (Chiang and Yang, 2015). We developed an

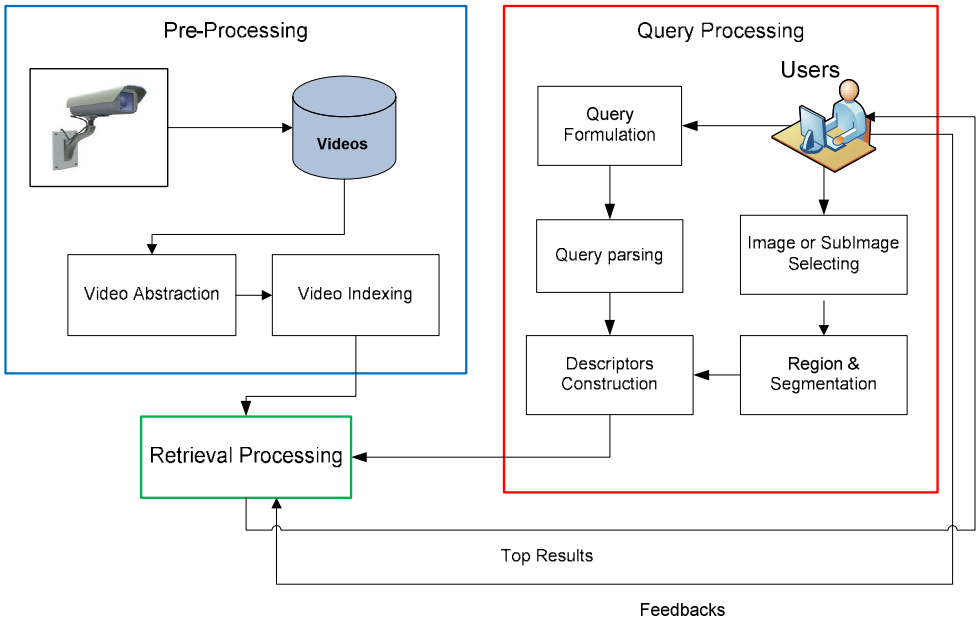

Fig. 2: Three modules of the proposed framework

adapted shot segmentation algorithm that extracted representative keyframes using clustering method for performing video abstraction.

Video indexing: Our video indexing is composed of the following three main sub modules: video analysis, feature extractions and data indexing. Video analysis is responsible for detecting interested stationary or mobile objects as well as tracking them (track these objects in successive frames) and event recognition (by analyzing the behaviors of these detected objects). Therefore, the result of this step is physical objects, events and trajectories.

In this framework detecting object of interest was performed by employing an adaptive background modelling from four common approaches of object detection including: background modelling, segmentation, supervised classifiers and point detectors. These detected moving objects are categorized into person, physical object and grouped object using their feature vectors which extracted during feature extraction. Feature extraction is performed to extract low level feature of object like color, shape, velocity, bag of region and trajectory.

The results of the two previous sub modules are used in the data indexing according to our data model. The Data model is responsible for determining what kind of features needs to be extracted and the way they should be organized and indexed in database. Hence, in the pre-processing module after detecting objects or recognition events their information (including the low level feature as well as spatio-temporal relation among them) are stored in the related frame based on our data model.

Query processing module: Query processing module is the essential part of our proposed framework. Therefore, query formulation, query parsing and query matching are responsible for retrieving the accurate result even the event or object is fully or partially indexed. The query processing module will start with formulating a textual query, selecting whole image as an example query, or part of image as a region query. A visual query-specification interface is devised to facilitate the query formulating or selecting image/sub-image processes. The next step is to parse user's query (query-by-text) or extracted feature after performing region segmentation (query-by-example or query-by-region). Moreover, to enhance the performance of retrieval result, the submitted query can include the low level feature of object such as shape, color, size and also specific occurrence of event such as time interval, spatio-temporal relation of grouped objects. Then, the query parsing checks the vocabulary of the words, analyze the syntax of query and separates the textual queries. On the other side, the selected region is represented using the proposed bag-of-regions segmentation algorithm which is computed with respect to the region's dominant color and using the color distance for quantizing regions.

Retrieval processing module: The ability to efficiently retrieve and browse the accurate results from video archive considering users' satisfaction according to their submitted query is the most critical aspect of any retrieval system. To this end, we employed the matching techniques to retrieve those objects or events which satisfy user's query either they are fully or

partially indexed. Moreover, the proposed relevance feedback algorithm will get the user feedback to refine the retrieval result by interaction among system and user. The user's feedbacks are learnt by developing a learning algorithm to further improve the retrieval performance as well as indexing these retrieved object or recognized event based in these feedbacks on video archive.

CONCLUSION

This research presented a novel and efficient framework for automatic video surveillance indexing and retrieval based on the problem of existing discussed works. However, successful design and development of the proposed framework will lead to accurate video retrieval with low processing time and operational cost. The three main components of this framework were pre-processing, query processing and retrieval processing modules. We briefly described the important functionalities of each module. We are currently implementing the video abstracting and indexing module to show the feasibility of the proposed framework. The creditability of it will be shown by performing various experiments using benchmark datasets.

ACKNOWLEDGMENT

This research is supported by the Fundamental Research Grant Scheme (FRGS/1/2014/ICT07/UPM/02/2) from the Malaysian Ministry of Education.

REFERENCES

Benabbas, Y., N. Ihaddadene and C. Djeraba. 2011. Motion pattern extraction and event detection for automatic visual surveillance. J. Image Video Process., Vol. 2011, Article No. 7.

Calderara, S., R. Cucchiara and A. Prati, 2006. Multimedia surveillance: Content-based retrieval with multicamera people tracking. Proceeding of the 4th ACM International Workshop on Video Surveillance and Sensor Networks (VSSN, 2006). Santa Barbara, California, USA, pp: 95-100.

Chamasemani, F.F. and L.S. Affendey, 2013. Systematic review and classification on video surveillance systems. Int. J. Inform. Technol. Comput. Sci., 5: 87.

Chiang, C.C. and H.F. Yang, 2015. Quick browsing and retrieval for surveillance videos. Multimed. Tools Appl., 74(9): 2861-2877.

Conte, D., P. Foggia, G. Percannella, F. Tufano and M. Vento, 2010. A method for counting moving people in video surveillance videos. EURASIP J. Adv. Sig. Pr., 2010: 231240.

Durak, N., A. Yazici and R. George, 2007. Online surveillance video archive system. In: Cham, T.J. *et al*. (Eds.), MMM 2007. LNCS 4351, Springer-Verlag, Berlin, Heidelberg, pp: 376-385.

Hampapur, A., L. Brown, J. Connell, A. Ekin, N. Haas *et al*., 2005. Smart video surveillance: Exploring the concept of multiscale spatiotemporal tracking. IEEE Signal Proc. Mag., 22(2): 38-51.

Hampapur, A., L. Brown, R. Feris, A. Senior, Chiao-Fe Shu *et al*., 2007. Searching surveillance video. Proceeding of IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS, 2007), pp: 75-80.

Hu, W., D. Xie, Z. Fu, W. Zeng and S. Maybank, 2007. Semantic-based surveillance video retrieval. IEEE T. Image Process., 16: 1168-1181.

Jiang, P. and X. Qin, 2010. Keyframe-based video summary using visual attention clues. IEEE MultiMedia, 17: 64-73.

Jung, Y.K., K.W. Lee and Y.S. Ho, 2001. Content-based event retrieval using semantic scene interpretation for automated traffic surveillance. IEEE T. Intell. Transp., 2: 151-163.

Kim, J.S., D.H. Yeom and Y.H. Joo, 2011. Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems. IEEE T. Consum. Electr., 57: 1165-1170.

Le, T.L., M. Thonnat, A. Boucher and F. Brémond, 2008. A query language combining object features and semantic events for surveillance video retrieval. In: Satoh, S., F. Nack and M. Etoh (Eds.), MMM 2008. LNCS 4903, Springer-Verlag, Berlin, Heidelberg, pp: 307-317.

Le, T.L., M. Thonnat, A. Boucher and F. Bremond, 2009. Surveillance video indexing and retrieval using object features and semantic events. Int. J. Pattern Recogn., 23: 1439-1476.

Le, T.L., A. Boucher, M. Thonnat and F. Brémond, 2010. Surveillance video retrieval: What we have already done? Proceeding of 3rd International Conference on Communications and Electronics (ICCE, 2010).

Lee, H., A. Smeaton, N. O'Connor and N. Murphy. 2005. User-interface to a CCTV video search system. Proceeding of the IEE International Symposium on Imaging for Crime Detection and Prevention (ICDP, 2005), pp: 39-43.

Lyons, D., T. Brodsky, E. Cohen-Solal and A. Elgammal, 2000. Video content analysis for surveillance applications. Proceeding of Philips Digital Video Technologies Workshop.

Nam, Y., S. Hong and S. Rho, 2013. Data modeling and query processing for distributed surveillance systems. New Rev. Hypermedia Multimedia, 19: 299-327.

Sabbar, W., A. Chergui and A. Bekkhoucha, 2012. Video summarization using shot segmentation and local motion estimation. Proceeding of the 2nd International Conference on Innovative Computing Technology (INTECH), pp: 190-193.

Şaykol, E., U. Güdükbay and Ö. Ulusoy, 2005. A database model for querying visual surveillance videos by integrating semantic and low-level features. In: Candan, K.S. and A. Celentano (Eds.), MIS 2005. LNCS 3665, Springer-Verlag, Berlin, Heidelberg, pp: 163-176.

Şaykol, E., U. Güdükbay and Ö. Ulusoy, 2010. Scenario-based query processing for video-surveillance archives. Eng. Appl. Artif. Intel., 23: 331-345.

Stringa, E. and C.S. Regazzoni, 1998. Content-based retrieval and real time detection from video sequences acquired by surveillance systems. Proceeding of International Conference on Image Processing (ICIP, 1998), 133: 138-142.

Stringa, E. and C.S. Regazzoni, 2000. Real-time video-shot detection for scene surveillance applications. IEEE T. Image Process., 9: 69-79.

Xu, Y. and D. Song, 2010. Systems and algorithms for autonomous and scalable crowd surveillance using robotic PTZ cameras assisted by a wide-angle camera. Auton. Robot., 29: 53-66.