The Journal of Systems and Software 102 (2015) 217–225

Contents lists available at ScienceDirect

The Journal of Systems and Software

journal homepage: www.elsevier.com/locate/jss

Semantic based representing and organizing surveillance big data using video structural description technology

Zheng Xu[b,a,*], Yunhuai Liu[a], Lin Mei[a], Chuanping Hu[a], Lan Chen[a]

[a] *The Third Research Institute of Ministry of Public Security, Shanghai, China*
[b] *Tsinghua University, China*

ARTICLE INFO

Article history:
Received 23 September 2013
Revised 27 May 2014
Accepted 13 July 2014
Available online 24 July 2014

Keywords:
Video structural description
Surveillance big data
Big data representing and organizing

ABSTRACT

Big data is an emerging paradigm applied to datasets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Especially, the data volume of all video surveillance devices in Shanghai, China, is up to 1 TB every day. Thus, it is important to accurately describe the video content and enable the organizing and searching potential videos in order to detect and analyze related surveillance events. Unfortunately, raw data and low level features cannot meet the video based task. In this paper, a semantic based model is proposed for representing and organizing video big data. The proposed surveillance video representation method defines a number of concepts and their relations, which allows users to use them to annotate related surveillance events. The defined concepts include person, vehicles, and traffic sighs, which can be used for annotating and representing video traffic events unambiguous. In addition, the spatial and temporal relation between objects in an event is defined, which can be used for annotating and representing the semantic relation between objects in related surveillance events. Moreover, semantic link network is used for organizing video resources based on their associations. In the application, one case study is presented to analyze the surveillance big data.

1. Introduction

Big data is an emerging paradigm applied to datasets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time (Wigan and Clarke, 2013). Such datasets are often from various sources (Variety) yet unstructured such as social media, sensors, scientific applications, surveillance, video and image archives, Internet texts and documents, Internet search indexing, medical records, business transactions and web logs; and are of large size (Volume) with fast data in/out (Velocity). More importantly, big data has to be of high value (Value). Various technologies are being discussed to support the handling of big data such as massively parallel processing databases (Yuan et al., 2013), scalable storage systems (Zhang et al., 2013a), cloud computing platforms (Liu et al., 2013), and MapReduce (Zhang et al., 2013b). Distributed systems are a classical research discipline investigating various distributed computing technologies and applications such as cloud computing (Yan et al., 2013a, 2013b; Lizhe et al., 2010) and MapReduce (Ze et al., 2014; Dan et al., 2013). With new paradigms and technologies, distributed systems research keeps going with new innovative outcomes from both industry and academy.

Recent research shows that videos “in the wild” are growing at a staggering rate (Cisco Visual Networking Index, 2013; Great Scott, 2013). For example, with the rapid growth of video resources on the world-wide-web, on YouTube[1] alone, 35 h of video are unloaded every minute, and over 700 billion videos were watched in 2010. Vast amount of videos with no metadata have emerged. Thus automatically understanding raw videos solely based on their visual appearance becomes an important yet challenging problem. The rapid increase number of video resources has brought an urgent need to develop intelligent methods to represent and annotate the video events. Typical applications in which representing and annotating video events include criminal investigation systems (Wu and Wang, 2010), video surveillance (Liu et al., 2009), intrusion detection system (Zhang et al., 2008), video resources browsing and indexing system (Yu et al., 2012), sport events detection (Xu et al., 2008), and many others. These urgent needs have posed challenges for video resources management, and have attracted the research

* Corresponding author at: Tsinghua University, China. Tel.: +86 13817917970.
E-mail address: xuzheng@shu.edu.cn (Z. Xu).

[1] www.youtube.com.

http://dx.doi.org/10.1016/j.jss.2014.07.024
0164-1212/

of the multimedia analysis and understanding. Overall, the goal is to enable users to search the related events from the huge number of video resources. The ultimate goal of extracting video events brings the challenge to build an intelligent method to automatically detect and retrieve video events.

In fact, the huge number of new emerging video surveillance data becomes a new application field of big data. The processing and analysing video surveillance data follow the 4 V feature of big data.

(1) **Variety**: The video surveillance data comes from the different devices such as traffic cameras, hotel cameras and so on. Besides the different surveillance devices, these devices also come from the different region. The distributed feature of video surveillance data augments the variety of the resources. For example, in the criminal investigation systems, different video surveillance data from the different surveillance devices are processed and analyzed to detect the related people, car, or things. The variety of video surveillance devices brings the big challenges for storage and management distributed video surveillance data.
(2) **Volume**: With the rapid development of the surveillance devices, for example, the number of surveillance devices in Shanghai is up to 200,000, the volume of video surveillance data becomes the big data. The data volume of all video surveillance data in Shanghai is up to 1 TB every day. The whole volume of all video surveillance data in Shanghai Pudong is up to 25 PB. The huge volume of video surveillance data brings the big challenges for processing and analyzing distributed video surveillance data.
(3) **Velocity**: The video surveillance devices are with fast data in/out. The video surveillance devices usually work in 24 h per day. The video surveillance devices collect real-time videos. The real-time collected videos usually upload to the storage server or data center. The velocity of collecting video surveillance data is faster than that of processing and analyzing them. The high velocity of video surveillance devices brings the big challenges for processing and analyzing video surveillance data. For example, the speed of processing and analyzing video surveillance data is much lower than collecting them.
(4) **Value**: The video surveillance data usually has high value. For example, in the criminal investigation systems, the video surveillance can help the police to find the suspect. In the traffic surveillance system, the video data can detect the illegal vehicles or people. On the other hands, the huge volume brings the challenges for mining the value from the video surveillance data. The phenomenon of “High volume, low value” also exists in the video surveillance big data.

In this paper, a semantic based model for representing and organizing video resources is proposed for bridging the gap between low-level representative features and high-level semantic content in terms of object, event, spatial and temporal relation extraction. The proposed model is named Video Structural Description (VSD). In order to solve the representing and annotating need for objects, events, and spatial–temporal relations during the video understanding process, a wide-domain applicable traffic ontology that uses objects and spatial/temporal relations in an event is developed. In order to organize the video resources based on their association, semantic link network (Zhuge, 2011) based method is used. The major contributions of this paper are summarized as follow.

(1) A whole framework for building domain ontology of VSD is proposed. The basic concepts, events, and relations of a given domain are defined. Moreover, a rule construction standard which is domain independent is given to construct domain ontologies. Domain ontologies are enriched by including additional rule definitions.
(2) The proposed method defines a number of concepts and their relations, which allows users to use them to detect video traffic events. A number of concepts including person, vehicle, and traffic sigh is given, which can be used by users for annotating and representing video traffic events unambiguous. In addition, the spatial and temporal relation in an event is proposed, which can be used for annotating and representing the semantic relation between objects in video traffic events.
(3) In order to organize the video resources, semantic link network based method is used. The semantic link network model can mine and organize video resources based on their associations.
(4) A semantic video annotation tool is implemented for annotating and organizing video resources based on the video annotation ontology. The annotation tool allows annotators to use domain specific vocabularies from traffic field to describe the video resources. These annotated video resources are managed based on the semantic relation between annotations. A semantic-based video organizing platform is provided for searching videos. It supports reasoning operation of the annotations of video resources.

The organization of the paper is as follows. In Section 2, the related work of the proposed work is given. The proposed VSD framework is given in Section 3. In Section 4, the ontology of traffic events domain is built. In Section 5, the semantic link network model is proposed to mine and organize video resources based on their associations. In Sections 6 and 7, the application and case study for mining video surveillance data are given. Finally, the conclusions and future research directions are discussed.

2. Related work

The key issue in semantic content extraction from videos is the representation of the semantic content. Many researchers have studied this from different aspects. A simple representation method may be associated the video events with low level features (texture, shape, color, etc.) using frames or shots from videos. These simple methods do not use any relations between features such as spatial or temporal relations. Obviously, using spatial or temporal relations between objects in videos is important for achieving accurate extraction of events. Researches such as BilVideo (Donderler et al., 2005), extended-AVIS (Sevilmis et al., 2008), multiView (Fan et al., 2001) and classView (Fan et al., 2004) used spatial and temporal relations but do not have ontology-based models for semantic content representation. Bai et al. (2007) presented a semantic based framework using domain ontology. Their work is used to represent video events with temporal description logic. However, the event extraction is manually and event descriptions only use temporal information. Nevatia and Natarajan (2005) gave an ontology model using spatial temporal relations to extract complex events where the extraction process is manual. In Bagdanov et al. (2007), each defined concept is related to a corresponding visual concept with only temporal relations for soccer videos. Nevatia and Natarajan (2005) built event ontology for natural representation of complex spatial temporal events given simpler events. A Video Event Recognition Language (VERL) (Nevatia et al., 2005) that allows users to define the events without interacting with the low level processing is defined. VERL is intended to be a language for representing events for the purpose of designing ontology of the domain, and, Video Event Markup Language (VEML) is used to manually annotate VERL events in videos. The lack of low level processing and using manual annotation are the drawbacks of this study. Akdemir et al. (2008) present a systematic approach to address