

Semantic Description and Information Retrieval Research of Surveillance Video in Smart Transportation System

Boxiong Yang^{1, a}, Jing Huang^{1, b} and Yuqi Yang^{2, c}

¹ Information Technology College, Beijing Normal University, Zhuhai, China

² Zhuhai Branch, Graduate School of Beijing Normal University, Zhuhai, China

^ayangboxiong@bnuz.edu.cn, ^bjhuang@bnuz.edu.cn, ^c516624914@qq.com

Keywords: Smart Transportation System, Semantic Description, Information Retrieval

Abstract. It is a hotspot to retrieve the desired information or find specific target inside the mass traffic video information. To achieve video content, it is necessary to build up a bridge between the low-level visual information and the high-level semantics. In this paper, firstly we propose a framework of intelligent traffic information semantic description. And then the semantic description framework of intelligent traffic information based on the standard of MPEG-7 is brought out. The framework can describe the multi-layer video information from the underlying visual features to the high-level semantic features, it can make up the gap of semantic and ultimately facilitate the users to retrieve mass video information. Finally, the accuracy and rapidity of the retrieval is demonstrated by the example of a traffic video semantic description and retrieval. It greatly reduces the working intensity of video processing workers because this method can handle most cases rapidly.

1. Introduction

With the rise of the wisdom urban construction, urban information which is the main composition of high-definition traffic surveillance video, presents a blowout growth trend. In the face of massive data, the city managers pay more attention to how to efficiently retrieve the information in which they are interested[1]. How to implement Visual Queries Content-Based (CBVQ)? The issue has caused the attention of researchers. Some prototype systems have come out, such as the QBIC IBM system, the WebSeek system of Tsinghua University, Columbia University, TV-FI system, etc.

It is a hotspot to retrieve the desired information or find specific target in mass traffic video information. The query method based on keywords or interesting points becomes extremely difficult for the reason of the limitation of the mainstream video compression format. The traditional traffic video analysis technology is a analysis of the underlying visual information processing, such as color, texture and contour. But the users often understand the entire video content from the perspective of high-level semantics, for instant, they can distinguish the objects and matters from the video. Inevitably, there is a semantic gap [2].

To achieve video content, we need to build up a bridge between the low-level visual information and the high-level semantics. At present, the research works for this problem has been widely carried out in the industry. One of the most popular solutions is the MPEG-7 standard, which is the Multimedia Description Interface (Content) [3]. The traffic monitoring video information based on the standard of MPEG-7 is described, and the information retrieval based on content is realized in this paper..

2. Semantic Description of Smart Transportation Information

The MPEG-7 standard is called "multimedia content description interface", which provides a standardized description of all kinds of multimedia information. In order to facilitate the further information on the content of multimedia semantic retrieval[4], the description which relates to the content itself, can meet the need of the information in which the users are interested in a quick and effective way. The following content is standardized for MPEG-7: Descriptors、Description Schemes、Description Definition Language. The content describes the definition of language based

on the XML language, which allows the description of the definition and description of the program to be flexible and described, and has a very great scalability[5].

2.1 Video_metadata

Video_metadata, a general description of each frame of the video, which is divided into visual metadata (Visual_metadata) and semantic metadata (Semantic_metadata), does not include video content itself. Visual metadata mainly includes format, size, resolution, color depth, compression method, etc. Semantic metadata mainly includes text annotation, time, place and the video frame number[6].

For example, a frame of video Visual_metadata is described as follows:

```
<Visual_metadata>  
<Format>AVI</Format>  
<Resolution>640*480</Resolution>  
<Colordepth>RGB8bit</Colordepth>  
<Compression>M-JPEG</Compression>  
</Visual_metadata>
```

2.2 Traffic_sensor_metadata

Traffic_sensor_metadata mainly describes the layout of a section of intelligent traffic sensor information, which includes the list of sensors layout, the location of the sensor, and the function of the sensor. The information described by this section also does not contain the video itself either. Joining this section and multi-sensor information content together can make it easier for users to query the information of the whole system. For example, the magnetic sensor information could be described as follows:

```
<Traffic_sensor_metadata>  
<Traffic_sensor_set>  
<Traffic_sendor ID="S1">  
<Sensor_type>Magnet</Sensor_type>  
<Function>Speed\Flow\Occupancy</Function>  
<Traffic_sensor >  
<Traffic_sensor_set>  
<Traffic_sendor_metadata >
```

2.3 Video_DS

Video_DS is a semantic description of the video information collected by the camera. This part is the core of the whole semantic description framework, which directly affects the information retrieval of users[7]. The main task of this part is to describe the traffic incident and the traffic incident in the video, which Video_object_set represents the collection of all traffic objects in the video. In each frame of the video, the extraction of traffic objects is divided into four categories: Vehicle (various motor vehicles and non motor vehicles), pedestrians, roads (one-way, two-way road, left turn lane, turn right lane), and a traffic sign (lane, parking line, traffic signal lamp).

Each element in Video_object_set is called Video_object, which is the traffic object. The visual features and semantic features of each class of objects are shown in Table 1.

Table 1. Visual Features and Semantic Characteristics of Intelligent Object

Objects	Visual Features	Semantic Features
Vehicle	Texture Shape Motion	License Plate
Pedestrian	Color Texture Shape Motion	Whether the Pedestrian
Road	Color Texture Texture	Whether to allow pedestrians pass or stay Whether to allow the vehicle pass or stay Allow what kind of vehicle Allow the direction of the vehicle Whether to allow a left turn Whether to allow a right turn
Traffic Sign	Color Texture Shape Position	What kind of traffic signs Whether to allow the pedestrian crossing Whether to allow vehicles crossing

Video_event_set is a set of video events. These events are traffic events which include vehicle straight running, vehicle left turn, vehicle right turn, running red lights, parking and crossing over the pedestrian, each event is given an ID[8]. Object_node references the element Video_object in Video_object_set, thus avoiding the repeated definition of object. Object_relation represents the relationship between object (Video_object), which is typically shown in Table 2.

Table 2. The Relationship between Common Object

Type of Relationship	Relationship
Space	Top of, Bottom of, Right of, Left of, Adjacent to, Neighboring to, Nearby, Within, Contain
Semantic	Relative of, Belongs to, Part of, Related to, Same as, Is a, Consist of, Crosses

Every traffic event describes the relationship among objects. For example, the object is to drive the vehicle into a red light, traffic lights and parking, for running a red light incident. The object is a parking lane and a stationary vehicle for a parking lot. The relationship between them is in the direction of the vehicle.

Sensor_status represents that the relevant intelligent traffic sensor began to detect information while a traffic incident happens. When the users need to read the sensor information, they can read directly from it. Traffic_sensor_node also needs to refer to the Traffic_sensor in the Traffic_sensor_metadata, which avoids duplication of sensors.

2.4 Traffic environment part

This part mainly describes a surveillance camera in the scope of the whole traffic environmental information, including the region's weather climate conditions, light conditions and the overall traffic information of the road.

3. Examples of Traffic Video Semantic Description

The framework of the intelligent traffic information semantic description framework is described. The frame as shown in Figure 1.



Figure 1 A frame of Traffic Video Data

The frame of video events is to identify a silver white car passing through a virtual line. The silver white sedan and the virtual black lines are considered as an object (CT Video_obje), the characteristics of the silver white car is described as follows:

```
<Video_object ID="1"><!--the ID of the object is "1"-->
<Object_visual_features>
<Obj_color>
<color>Silvery White</color>
</Obj_color>
<MediaUri>Video/object.avi</MediaUri>
<MediaTime>2015-06-12T10:20:30</MediaTime>
<MediaResolution>640*480</MediaResolution>
<MovingRegionLocator>
<SingleFrameLocation>
  <TimeLocator>
    <Frame no="81"/>
  </TimeLocator>
</SingleFrameLocation>
<SpacePartition>
  <Start x="188" y="149">
  <End x="232" y="259">
</SpacePartition>
</SingleFrameLocation>
</MovingRegionLocator>
.....
</Descriptor>
```

The ID is the unique number to mark the target global object. MediaUri marks the objects in the video file name. MediaResolution and MediaTime respectively marks the resolution and the starting

time of video files. MovingRegionLocation is used to locate the target, which contains multiple SingleFrameLocator, which is used to locate the position of the object in each frame.

The intelligent traffic information semantic description framework proposed in this paper can describe the video information from the perspective of semantics in Figure 2. The results of these video information are stored in the database, which greatly facilitates the users to query the video information from the perspective of semantics.

4. Traffic Video Retrieval Based on Semantic

The basic principle of traffic video retrieval based on semantics is access to the key frame of the video through the decomposition and analysis of video. The semantic definition of video is reconstructed according to the characteristics of the key frame. Specific principles are shown in Figure 2. After finishing the process of analysis, segmentation, feature extraction, and object recognition of the video stream, the semantic information of the target is established and thus the semantic object and content formed. The content includes information about the temporal and spatial relations, the target category, the direction of action and so on.

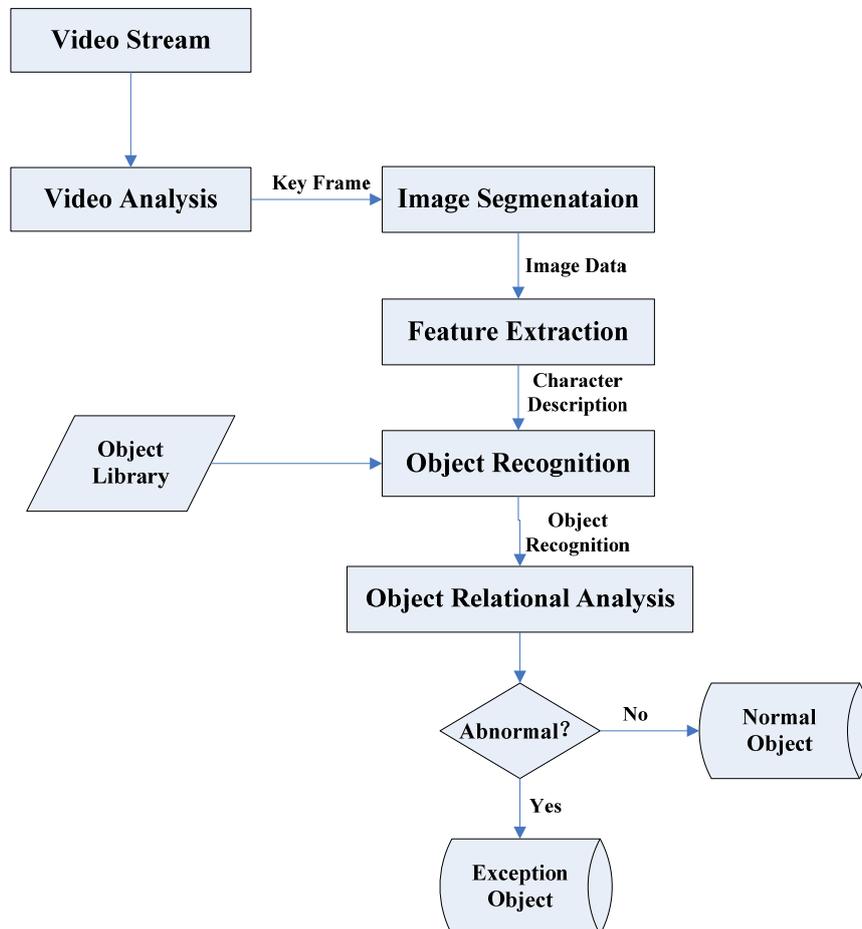


Figure 2. The Relationship between Semantic and Content of Traffic Video Object

Taking the video tracking and retrieval in traffic surveillance, for example, the video information retrieval based on MPEG-7 semantic description is introduced as following. According to the semantic description of MPEG-7, it can support the retrieval of object type, moving object trajectory, object appearance time and so on. Developers can directly use the graphical interface to provide search options, which can also be retrieved to using a predetermined search semantics.

Query language EBNF is defined as follows:

*Query := (TypeRestriction/SpatialRestriction/TemporalRestriction)**

TypeRestriction := object type in (TypeSet)

*TypeSet := (Car/Bicycle/Pedestrian)**

Temporal Restriction :=temporal restriction from NUM to NUM

For example, the following query command string is used to query the vehicle between 20 frames to 200 frames in Figure 3.



Figure 3. Video Tracking and Retrieval Results

Results from Figure 3 shows that the ID1 in video frame can be well retrieved from the ID1, which is a gray and white car with a tracking object tag. In the entire 270 frames of the video, the continuous 5 frame images of the cars can be track the target and locked well.

5.Summary

In this paper, firstly we propose a framework of intelligent traffic information semantic description. And then the semantic description framework of intelligent traffic information based on the standard of MPEG-7 is put forward. The framework can describe the multi-layer video information from the underlying visual features to the high-level semantic features, it can make up the semantic gap, and ultimately facilitate the users to retrieve mass video information. Finally, the accuracy and rapidity of the retrieval is demonstrated by the example of a traffic video semantic description and retrieval. It greatly reduces the working intensity of video processing workers for the method can handle most cases. Our subsequent work is the further study of the information fusion model according to the information structure characteristics of various intelligent traffic sensors. Meanwhile, we further improve the scalability and compatibility of the semantic description framework of intelligent transportation information, and incorporate more traffic information into the whole framework.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (No.61272364).

References

- [1] Simon Jirka, Arne Bröring, Christoph Stasch. Discovery Mechanisms for the Sensor Web. *Sensors*, 2013, 9: 2661-2681.
- [2] FemkeReitsma, JohnLaxton, StuartBallard, WernerKuhn, AliaAbdelmoty. Semantics, ontologies and eScience for the geosciences. *Computers & Geosciences*, 2014, 35: 706-709.
- [3] Michael Compton, Corey Henson, Holger Neuhaus, Laurent Lefort, Amit Sheth. A Survey of the Semantic Specification of Sensors. In *2nd International Workshop on Semantic Sensor Networks, at 8th International Semantic Web Conference*, 2013, 17-32.
- [4] Cory A. Henson, Josh K. Pschorr, Amit P. Sheth, Krishnaprasad Thirunarayan..SemSOS: Semantic sensor Observation Service. *2009 International Symposium on Collaborative Technologies and Systems*:44-53.
- [5] Chao Chen, Sumi Helal. Sifting Through the Jungle of Sensor Standards. *IEEE Pervasive Computing archive*, 2008, 7(4): 84-88.
- [6] Joshua S. Madin, Shawn Bowers, Mark P. Schildhauer, Matthew B. Jones. Advancing ecological research with ontologies. *Trends Ecol Evol*. 2012, 23(3): 159-168.
- [7] M.F. McCabe, E.F. Wood, R. Wójcik, M. Pan, J. Sheffield, H. Gao, H. Su. Hydrological consistency using multi-sensor remote sensing data for water and energy cycle studies. *Remote Sensing of Environment*, 2012, 112: 430-444.
- [8] Remote Sensing of Environment, 2012, 112: 430-444.
- [9] Amit Sheth et al., SA-REST: Semantically Interoperable and Easier-to-Use Services and Mashups, *IEEE Internet Computing*, 2013, 11(6): 91-94