Feature Representation Scheme for Smart Video Sensor

Imelda Imelda†, Agus Harjoko†† and Pekik Nurwantoro†††

†,††Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
†††Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
†Faculty of Information Technology, Universitas Budi Luhur, Jakarta, Indonesia

Summary
In Jakarta and other cities, the police and the Department of Transportation are still monitoring traffic videos manually. They have difficulties when some of the traffic problems occur simultaneously. Therefore, they need a Smart Video Sensor that can automatically provide the latest information about traffic conditions. The information generated Smart Video Sensor are features. Data inputs from Smart Video Sensor are scenes. The working principle of this video is a streaming video that records scene frame by frame. Then, the Smart Video Sensor will extract the features of the frames. After that, the Smart Video Sensor will provide a feature representation scheme in the form of XML. Furthermore, the Smart Video Sensor will send features in the form of XML representations to TMC. This paper contributes to representing the features of the objects of traffic in XML (eXtensible Markup Language). XML format is selected to represent the features because of its dynamic XML structure. The test results showed that the representation of the features can be read and viewed accordingly.

Key words:
XML, feature representation, traffic objects, features of traffic objects.

1. Introduction

In general, traffic problems consist of traffic congestions, traffic violations and traffic accidents [1], [2]. The causes of traffic congestion, among others are roads that can no longer acccomodate the existing number of vehicles, traffic light settings which are not proportional and not efficient to the number of vehicles on the road, the spilled market and the number of vehicles that stop in random places. Some causes of traffic violations, among others are motorists who disobey the traffic rules, motorists who disobey traffic signals, the motorists who cross the road markings that have been determined, vehicle loads that exceed the capacity of the applicable standard or limit, and vehicles that exceed the speed limit. Many traffic accidents are resulted from driver being too tired, sleepy driver, unfocus driving, overly loaded vehicles that exceed the applicable standards, vehicles that exceeds the safe limits, unsafe distance between vehicles, or disfunctional brakes. Traffic violations can lead to traffic accidents. Traffic accidents can cause death.

The police and the Department of Transportation require traffic information to find the cause of the accidents. The parties need accurate information on the crime scenes. Traffic information needs to be given as soon as possible to enable them to take an immediate action. Traffic information can be a form of report or a public complaint. Traffic information is received by the police through the video, websites, radio, facebook, twitter, call center, SMS, fax, telephone, or people can directly come to report to the nearest police station. The police will then forward the traffic information to the traffic users via website, facebook, twitter, GIS (Geographic Information Systems), call center, vehicle registration online, CCTV (Closed-circuit television), GPS (Global Positioning System), the operator HT, news radio, television news, and information boards on the highway. Twitter Yogyakarta police RTMC police is @RTMC_Jogja. Twitter TMC Polda Metro Jaya is @TMCPoldaMetro. Twitter NTMC Police is @NTMCLantasPolri. A website that provides information about the traffic in Indonesia can be accessed at lewatmana.com. Korlantas official website can be accessed at lantas.polri.go.id. CCTV Streaming Traffic provides information on Road Transport, and Traffic Management Center - Department of Transportation can be accessed at http://www.rttmc-hubdat.com. Live Streaming Video CCTV PT Jasa Marga can be accessed at jasamargalive.com.

However, the traffic information above isn’t sufficient. In Jakarta and other cities, the police and the Department of Transportation have difficulties in finding the cause of accidents because traffic videos are still monitored manually [3]. Monitored manually means that the video recordings are directly sent to TMC (Traffic Management Center). In TMC, the police and the Department of Transportation will observe many video recordings simultaneously. They will have difficulties when some of the traffic problems occur at the same time. For this reason, Smart Video Sensor (SVS) is presented as a solution. SVS is needed because it can automatically provide the latest information about traffic conditions [4]. This statement is reinforced by the survey [4] that the latest technology and market trends is demanding significant needs for a viable solution to video/camera systems and analytic.

Manuscript received January 5, 2018
Manuscript revised January 20, 2018

SVS is an intelligent video camera that utilizes a processor and a software to achieve a certain level of intelligence. Without human intervention, SVS has the ability to analyze automatically. SVS can produce information as needed. SVS can provide a structured and meaningful information needed to process the raw data of video and send the results to where it is needed. SVS can be easily integrated into other devices. SVS can be used efficiently, interactively, precisely and flexibly. It can also be used in some other intelligent applications [5], [6]. Automatically, SVS compresses video data, extracting features of the object traffic, performs XML format and data transmission. SVS is used to process videos, extracting features and generate traffic object features. These features will be delivered and deployed to another application that is suitable and in need. With an SVS, applications in other places will no longer need the stages of preprocessing, segmentation and feature extraction. With an SVS, applications that exist in other places can process and use the features that have been extracted as needed.

There are two alternative places video processing. The first alternative is a system that records and transmits video data to TMC. Then, other systems will receive and process video data in TMC. When another system is receiving video data at TMC, a lot of video data that is received is incomplete because the delivery process often fails. This failure occurs because the network is disconnected, and the bandwidth and the processing power of the CPU (Central Processing Unit) is limited. Due to the incomplete video data, the video data cannot be processed further. When this other system is processing video data at TMC, the system will become very slow in producing an output because much data is processed at the same time. The second alternative, streaming system that records images frame by frame. Systems will extract features from frame. The system will then generate a schematic representation of features in the form of an XML format. Next, the system will send an XML representation of the feature to TMC. Many advantages can be obtained when choosing the second alternative. The police and the Department of Transportation can direct traffic using object features as needed. The police and the Department of Transportation will be able to monitor traffic easier. The police and the Department of Transportation can use video data traffic as evidence of traffic violations or other purposes. Therefore, the second alternative is chosen as the solution.

The contribution of this paper represents the features of the traffic objects in XML format. Feature representation of traffic objects is selected in XML format for a dynamic XML structure. XML format is chosen because the information can be exchanged from one system to another different platform. For example, from Windows to Unix, or from PC to Macintosh and even from the Internet to mobile phones with WAP technology [7]–[9]. In addition, the XML format used for XML can save time tag information. Time tag information is stored to solve the problem of incomplete video data traffic. By storing information, tag means that each frame has a time stamp. With a time stamp, the current frame will be connected to the next frame. In result, it will be easier to track an object.

Several researchers have studied the representation of features. In the study regarding security monitoring indoor (indoor surveillance), a system will extract a frame into a feature, classify objects, track objects, and send features to another place using XML format. The accuracy level for its object classification is above 80% [10], [11]. Another study classification features data in an XML document used the fuzzy c-means clustering algorithm. Accuracy which was obtained on classification reached 85% - 87% [12]. Meanwhile, other researchers examined on saving features in the form of XML because XML can organize the data in a very large number [13]. But no one has discussed in detail the scheme of representation and relation to XML format. Therefore, feature of this representation is presented.

2. Feature Representation

Model representation of features can be seen in Fig.1. The process starts from a video sensor that records scene frame per frame. Then, the noise of the frame is removed. After that, the frame is normalized. Then, the frames are segmented into segment groups. Each segment is labeled. Several features are extracted from each label. The features are shown in the schematic representation. Scheme of representation features generates XML structure. Representation of the features that use XML structure is easier to read. Processing of the data of Fig.1 is discussed in more detail below.

2.1 Sensor Video

Features of the traffic objects come from the video sensor. Data inputs from the video sensor are scenes. The working principle of this sensor is a streaming video that records scene frame by frame. It means that a video camera captures the object in a frame base [14].

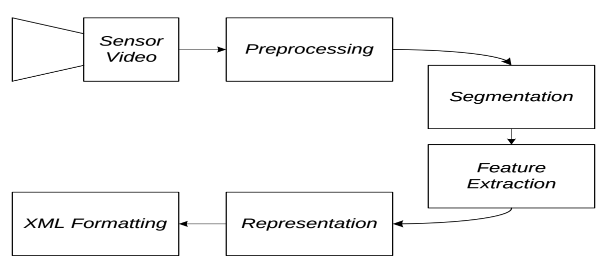

Fig.1 The proposed feature representation model

2.2 Preprocessing

Preprocessing is useful for preparing images to be processed at a later stage. In the preprocessing, noise image data is removed by filtering. There are two ways filtering. First, if an image is too bright, the image will be smoothed (smoothing). Second, if the existing image is less clear, then the image will be sharpened (sharpening). Data input from preprocessing is a frame. Then, each frame will be normalized. The output preprocessing is the data frame that has been normalized.

2.3 Segmentation

Segmentation is useful to separate the background of an object to the foreground the object. Segmentation is also useful to separate an object from another object. Data input from segmentation is a data frame that has been normalized. The outputs of the segmentation are segments. A segment is a group of pixels of the object.
In the early stages, background model will be created to make a background of an object traffic. Background model is created by eliminating the moving object. After that, the system will observe moving objects. Moving objects or speed $\neq 0$ means foreground. While the object is not moving or speed = 0, that's the background model. The method used to determine background and foreground is Background Substraction model [10], [11]. Blob method [15], [16] and connecting labeling [10], [11], [17] is used to separate objects with one another and produce segments.

2.4 Feature Extraction

Features are defined as a function of one or more measurements, each of which defines some quantitative properties of an object, and count like quantify some significant characteristics of the object [18]. Features can be classified as a low-level feature and a high level feature. The low-level feature can be extracted directly from the original image, while the high-level feature must be based on low-level feature [18]. A low-level feature is the feature that can not be unloaded [19]–[21]. Meanwhile, a high-level feature is the feature that can still be lowered [22]–[24]. Nowadays, it is known: there are 53 features that will be extracted. Features consist of 30 low-level features and 23 features a high level. Features low level include: length, width[25], area [26], [27], perimeter [26], [28], height [29], width of the bounding box (WBB), height of the bounding box (HBB), length of major axis, length of the minor axis [28], [30], [31], centroid [27], [32], [33], filled area [28], [30], [31], diameter [25], [34], edge [33], [35], [36], distance [37], gray [38], roughness [27], Region Of Interest (ROI) [33], [38], [39], Histograms of Oriented Gradients (HOG) [40], [41], separation radius of each training example [35], axle, distance, body length, chassis height [42], texture [26], [38], [43], angle [34], [37], elongation [27], convex area [26], [28], wheelbase [44], colour [33], [45], [46], motion [45], [46], moment invariant [26], [32], [47], [48]. Features high level include: size [29], compactness [39], [49], length-width ratio [38], [49], equivdiameter, dispersedness [28], [30], [31], rectangularity [38], [49], solidity [49], [50], the total area of the object [25], the ratio of the distance between the centre of the object and length of the main axis [25], 3D color histograms [50], extent, eccentricity, ellipticity, bounding box, circularity, shape-filling measure [28], area ratio [39], Edge-based Pyramid of Histogram of Orientation Gradients (EPHOG), Intensity-based Pyramid of Histogram of Orientation Gradients (IPHOG) [30], breadth [27], [32], convex hull perimeter [26], [27], the median of the number of edge points [35], shape [38], [45], [46]. Data inputs from the feature extraction are segments. The output of the feature extraction are features of the objects of the traffic. In one frame, there may be a few segments. These segments apart in several places. The segments are labeled to facilitate the determination of features. Once the segments are labeled and then each of its segments will be searched one by one. Features of the objects of traffic are needed elsewhere for different usage. Features can be used for object identification, object detection, object recognition, object classification, object tracking, calculation of the number of objects, an object speed calculation, estimate traffic congestion, and other traffic parameters required by the police and the Department of Transportation.

2.5 Representation

Feature representation scheme of traffic objects in Fig.2 contains information about the frame number, the date and time frame, the number of objects in the frame, the coordinates of each object, and the features of each object. The frame number is needed to determine the sequence of the frame. If the date changes, the frame will start from 1 (one). Date and time frame using UTF (Universal Time Format). The time format is in accordance with international formats are specified in ISO 8601 [51]. The standard format used is complete date plus hours, minutes, seconds and milliseconds. The form: YYYY-MM-DDThh: mm: ss.sTZD. Example: 2015-10-28T12: 05: 45.55 + 07: 00. It means that the frame was recorded on October 28, 2015, at 12: 05: 45: 55 GMT + 7. This time format has been used by the World Wide Web. This format is used to facilitate the use of the time format in a single number. The coordinates x and y are used to distinguish the position of the object with another object, determining the initial position of the object and the calculation of the value of the feature. In the schematic in Fig.2 can be seen every one frame consists of multiple objects. Each object has a few features.

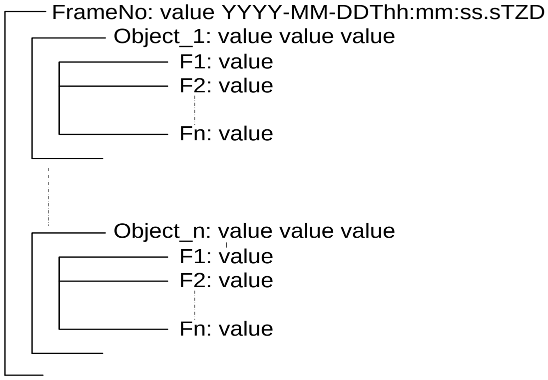

Fig.2 The proposed feature representation scheme

2.6 XML Formatting

XML formatting was formed from feature representation scheme of traffic objects. XML format has a dynamic structure. Dynamic XML structure is necessary because the number of objects in each frame is different, and the number of features in each object is also different. XML structure contains information that does not require a standard format. Text-based XML files can be read XML files without the need for special software. This facilitates the use of XML application developers to debug their programs. In the XML tags and attributes can be used as needed. Data can be stored as an attribute and element content simply by putting it between an opening tag and a closing tag. Characteristics of XML include intelligence, adaptive, simple, portability. XML is said to be intelligence because XML can handle various levels of complexity. Adaptive means that XML can be adapted to create their own language. Simple means that XML is simple and easy to make. Portability means that XML has the ease of exchanging data between diverse systems. With XML format, information can be exchanged easily from one system to another different platform. XML format can be easily processed in a variety of applications. The XML design enable of storing data in a concise and easy to set [52].

Feature representation scheme in Fig.2 is shown in Fig.3 in the XML structure. Tag <?xml version="1.0"?> indicates the XML version used. Label and the contents of the frame number indicated by <FRAMENO>FrameNo </FRAMENO> <FRAME_No> value</FRAME_No>. Time format indicated by <FTIMES>valuetime </FTIMES>. In one frame there can be multiple objects. Name of the object and the object value indicated by <ObjectNm_n>Object_n<ObjectNm_n> <ObjectNo>value </ObjectNo>. The number of objects indicated by n. X and Y coordinates are indicated by <Coordinat_X> value </Coordinat_X><Coordinat_Y> value </Coordinat_Y>. The name and value of the features indicated by <Feature_n>Fn</Feature_n> <Value_n>value</Value_n>. The number of features indicated by n. To mark that all its objects have been extracted then used the symbol <S>-</S>.

```
<?xml version="1.0"?>
<FRAMENO>FrameNo</FRAMENO><FRAME_No>value</FRAME_
No>
<FTIMES>valuetime</FTIMES>
<LIST_O_XNO>
<O_XNO>
        <ObjectNm_1>Object_1<ObjectNm_1><ObjectNo>value</Obje
        ctNo>
        <Coordinat_X> value </Coordinat_X>
        <Coordinat_Y> value </Coordinat_Y>
        <Feature_1>F1</Feature_1><Value_1>value</Value_1>
        <Feature_2>F2</Feature_2><Value_2>value</Value_2>
        :
        <Feature_n>Fn</Feature_n><Value_n>value</Value_n>
</O_XNO>
:
:
<O_XNO>
        <ObjectNm_n>Object_n<ObjectNm_n><ObjectNo>value</Obje
        ctNo>
        <Coordinat_X> value </Coordinat_X>
        <Coordinat_Y> value </Coordinat_Y>
        <Feature_1>F1</Feature_1><Value_1>value</Value_1>
        <Feature_2>F2</Feature_2><Value_2>value</Value_2>
        :
        <Feature_n>Fn</Feature_n><Value_n>value</Value_n>
</O_XNO>
</LIST_O_XNO>
<S>-</S>
```

Fig.3 XML Structure

3. Testing

Testing of feature representation in XML format is conducted to prove that the features can be read and can be seen in conformance with the video frame. Tests are also conducted to ensure that features successfully sent can be read correctly. If the feature is successfully sent and features can be read then the scheme representation is correct .

Testing of feature representation on this paper uses two samples of different data frames. Initially, the smart video sensor automatically determines the observation area / window. Second, smart video sensor counts the number of objects that exist in the area. Third, smart video sensor stores x and y coordinates of each object. Fourth, smart video sensor calculates the features of each object. Fifth, smart video sensors form a representation of features in

XML format (XML representation). Sixth, smart video sensors send the XML representation to TMC. Seventh, another system receives XML representation. Eighth, other systems read XML representation.

In this test, not all features are represented. Only 7 of the 53 features are represented. These features are represented include Height of the Bounding Box (HBB), Width of the Bounding Box (WBB), Area, Bounding Box, Compactness, Equivdiameter, and Dispersedness. Features have been many studies that use it for classification and tracking [28]. To facilitate reading representation scheme then each feature given the symbol. The F1 symbol for the Height of the Bounding Box (HBB). The F2 symbol for the Width of the Bounding Box (WBB). The F3 symbol for the Area. The F4 symbol for Bounding Box. The F5 symbol for Compactness. The F6 symbol for Equivdiameter. The F7 symbol for Dispersedness.

Fig.4 Video Frame 1

In Fig.4, there are traffic objects such as cars, buses, motor1, and motor2. Traffic objects represented are only moving objects. Traffic objects that do not move are not represented.

Fig.5 shows the feature representation scheme of video frame 1. Frame: 2014-01-16T05:58:43.23+07:00 means that the frame was recorded on January 16, 2014 at 5 over 58 minutes, 43 seconds and 23 milliseconds in accordance GMT+7. There are 4 objects which are represented in Figure 5. Each object has several features. The first object is located at the coordinates (135, 201). The first object features include: Height of the Bounding Box (HBB) = 54, Width of the Bounding Box (WBB) = 56, Area = 3024, Bounding Box = 220, Compactness = 0.78, Equivdiameter = 14,839,671, and Dispersedness = 16.01. The nth object located at coordinates (81, 268). The nth object features include Height of the Bounding Box (HBB) = 27, Width of the Bounding Box (WBB) = 13, Area = 351, Bounding Box = 80, Compactness = 0.69, Equivdiameter = 199,928.59 and Dispersedness = 18.23.

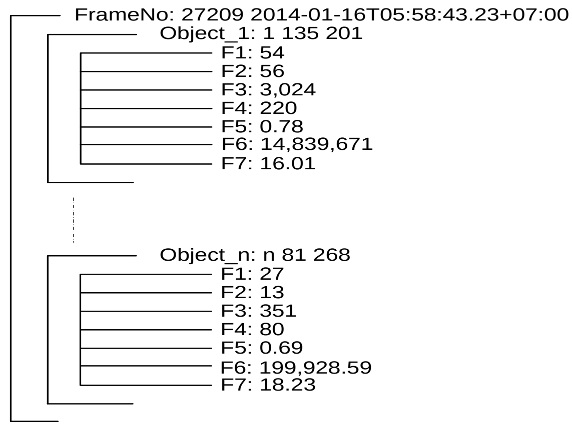

Fig.5 Feature representation scheme of video frame1

Fig.6 shows XML Representation of video frame 1. Tag <?Xml version = "1.0" encoding = "UTF-8"?> Indicates the version of XML being used. The <FRAMENO> FrameNo </FRAMENO> <FRAME_No> 27209 </FRAME_No> means frame to 27209 of the recorded video. The <FTIMES> 2014-01-16T05:58:43.23+07:00 </FTIMES> means frame of this video was recorded on January 16, 2014 at 5 over 58 minutes, 43 seconds and 23 milliseconds in accordance GMT+7. The first object indicated by <ObjectNm_1> Object_1 </ObjectNm_1> <ObjectNo> 1 </ObjectNo>. The first location of the object coordinates on (135, 201) is indicated by <Coordinat_X> 135 </Coordinat_X> <Coordinat_Y> 201 </Coordinat_Y>. The name and value of the first feature of the first object indicated by <Feature_1> HBB </Feature_1> <VALUE_1> 54 </VALUE_1>. The name and value of the second feature of the first object indicated by <Feature_2> WBB </Feature_2> <VALUE_2> 56 </VALUE_2>. The name and value of the third feature of the first object indicated by <Feature_3> Area </Feature_3> <VALUE_3> 3024 </VALUE_3>. The name and value of the fourth feature of the first object indicated by <Feature_4> Bounding Box </Feature_4> <Value_4> 220 </Value_4>. The name and value of the fifth feature of the first object indicated by <Feature_5> compactness </Feature_5> <Value_5> 0.78 </Value_5>. The name and value of the sixth feature of the first object indicated by <Feature_6> Equivdiameter </Feature_6> <Value_6> 14,839,671 </Value_6>. The name and value of the seventh feature of the first object indicated by <Feature_7> Dispersedness </Feature_7> <Value_7> 16.01 </Value_7>. Mark the beginning and end of the reading object is marked with the pair <O_XNO> </O_XNO>. The reading of the same was done for the second and subsequent objects. All objects traffic moving its read to completion. Signs that all objects have been read and the process is completed in one frame indicated by <S> - </S>.

```
<?xml version="1.0" encoding="UTF-8"?>
<FRAMENO>FrameNo</FRAMENO><FRAME_No>27209</FRAME_No
>
<FTIMES>2014-01-16T05:58:43.23+07:00</FTIMES>
<LIST_O_XNO>
<O_XNO>
<ObjectNm_1>Object_1</ObjectNm_1><ObjectNo>1</ObjectNo>
<Coordinat_X>135</Coordinat_X>
<Coordinat_Y>201</Coordinat_Y>
<Feature_1>HBB</Feature_1><Value_1>54</Value_1>
<Feature_2>WBB</Feature_2><Value_2>56</Value_2>
<Feature_3>Area</Feature_3><Value_3>3,024</Value_3>
<Feature_4>Bounding Box </Feature_4> <Value_4>220 </Value_4>
<Feature_5>Compactness</Feature_5><Value_5>0.78</Value_5>
<Feature_6>Equivdiameter</Feature_6><Value_6>14,839,671</Value_6>
<Feature_7>Dispersedness</Feature_7><Value_7>16.01</Value_7>
</O_XNO>
            :
            :
<O_XNO>
<ObjectNm_n>Object_n</ObjectNm_n><ObjectNo>n</ObjectNo>
<Coordinat_X>81</Coordinat_X>
<Coordinat_Y>268</Coordinat_Y>
<Feature_1>HBB</Feature_1><Value_1>27</Value_1>
<Feature_2>WBB</Feature_2><Value_2>13</Value_2>
<Feature_3>Area</Feature_3><Value_3>351</Value_3>
<Feature_4>Bounding Box</Feature_4><Value_4>80</Value_4>
<Feature_5>Compactness</Feature_5><Value_5>0.69</Value_5>
<Feature_6>Equivdiameter</Feature_6><Value_6>199,928.59</Value_6>
<Feature_7>Dispersedness</Feature_7><Value_7>18.23</Value_7>
</O_XNO>
</LIST_O_XNO>
<S>-</S>
```

Fig.6 XML representation of video frame 1

In Fig.7, we can see the video frame features frame 2. Date 2014-01-16T05:58:33.21+07:00 means: the frame was recorded on January 16, 2014 over 58 hours 5 minutes 33 seconds and 21 milliseconds GMT + 7 . From 7 objects that exist in Fig.7, only two objects will be described in detail. The first object is located at the coordinates (107, 190). The first object has features include Height of the Bounding Box (HBB) = 10, Width of the Bounding Box (WBB) = 7, Area 70, Bounding Box = 34, Compactness = 0.76, Equivdiameter = 7951.64 and Dispersedness = 16.51. The nth object is located at the coordinates (328, 288). The nth object has features include Height of the Bounding Box (HBB) = 26, Width of the Bounding Box (WBB) = 18, Area = 468, Bounding Box = 88, Compactness = 0.76, Equivdiameter = 355,428.63 and Dispersedness = 16.55.

Fig.7 Video Frame 2

In Fig.8 is shown XML Representation of video frame 2. Tag <? Xml version = "1.0" encoding = "UTF-8"?> indicates the version of XML being used. The <FRAMENO> FrameNo </FRAMENO> <FRAME_No> 22302 </FRAME_No> means frame to 22302 of the recorded video. The <FTIMES> 2014-01-16T05:58:33.21+07:00 </FTIMES> means frame of this video was recorded on January 16, 2015 at 5 over 58 minutes, 33 seconds and 21 milliseconds in accordance GMT+7. The first object indicated by <ObjectNm_1> Object_1 </ObjectNm_1> <ObjectNo> 1 </ObjectNo>. The first location of the object coordinates on (107.190) is indicated by <Coordinat_X> 107 </Coordinat_X> <Coordinat_Y> 190 </Coordinat_Y>. The name and value of the first feature of the first object indicated by <Feature_1> HBB </Feature_1> <VALUE_1> 10 </VALUE_1>. The name and value of the second feature of the first object indicated by <Feature_2> WBB </Feature_2> <VALUE_2> 7 </VALUE_2>. The name and value of the third feature of the first object indicated by <Feature_3> Area </Feature_3> <VALUE_3> 70 </VALUE_3>. The name and value of the fourth feature of the first object indicated by <Feature_4> Bounding Box </Feature_4> <Value_4> 34 </Value_4>. The name and value of the fifth feature of the first object indicated by <Feature_5> compactness </Feature_5> <Value_5> 0.76 </Value_5>. The name and value of the sixth feature of the first object indicated by <Feature_6> Equivdiameter </Feature_6> <Value_6> 7,951.64 </Value_6>. The name and value of the seventh feature of the first object indicated by <Feature_7> Dispersedness </Feature_7> <Value_7> 16.51 </Value_7>. Mark the beginning and end of the reading object is marked with the pair <O_XNO> </O_XNO>. The reading of the same was done for the second and subsequent objects. All objects traffic moving its read to completion. Signs that all objects have been read and the process is completed in one frame indicated by <S> - </S>.

```
<?xml version="1.0" encoding="UTF-8"?>
<FRAMENO>FrameNo</FRAMENO><FRAME_No>22302</FRAME_No
>
<FTIMES>2014-01-16T05:58:33:21+07:00</FTIMES>
<LIST_O_XNO>
<O_XNO>
<ObjectNm_1>Object_1</ObjectNm_1><ObjectNo>1</ObjectNo>
<Coordinat_X>107</Coordinat_X>
<Coordinat_Y>190</Coordinat_Y>
<Feature_1>HBB</Feature_1><Value_1>10</Value_1>
<Feature_2>WBB</Feature_2><Value_2>7</Value_2>
<Feature_3>Area</Feature_3><Value_3>70</Value_3>
<Feature_4>Bounding Box</Feature_4><Value_4>34</Value_4>
<Feature_5>Compactness</Feature_5><Value_5>0.76</Value_5>
<Feature_6>Equivdiameter</Feature_6><Value_6>7,951.64</Value_6>
<Feature_7>Dispersedness</Feature_7><Value_7>16.51</Value_7>
</O_XNO>
              :
<O_XNO>
<ObjectNm_n>Object_n</ObjectNm_n><ObjectNo>n</ObjectNo>
<Coordinat_X>328</Coordinat_X>
<Coordinat_Y>288</Coordinat_Y>
<Feature_1>HBB</Feature_1><Value_1>26</Value_1>
<Feature_2>WBB</Feature_2><Value_2>18</Value_2>
<Feature_3>Area</Feature_3><Value_3>468</Value_3>
<Feature_4>Bounding Box</Feature_4><Value_4>88</Value_4>
<Feature_5>Compactness</Feature_5><Value_5>0.76</Value_5>
<Feature_6>Equivdiameter</Feature_6><Value_6>355,428.63</Value_6>
<Feature_7>Dispersedness</Feature_7><Value_7>16.55</Value_7>
</O_XNO>
</LIST_O_XNO>
<S>-</S>
```


Fig.8 XML Representation from video frame 2

Experimental results show that the representation of the XML in Fig.8 is in accordance with the state of video in Fig.7. This proves that the feature representation scheme is correct. Thus, the representation scheme can be used to represent the features of video frames.

4. Conclusion

Feature representation of traffic objects is very useful to the police and the Department of Transportation. With this feature, the police and the Department of Transportation can directly use features to more specific needs such as identifying objects, detect objects, object recognition, object classification, object tracking, calculation of the number of objects, the calculation speed of the object, estimation of traffic congestion, and other traffic parameters. Feature representation of traffic objects uses the XML format. XML format is chosen because the information can be exchanged from one system to another. This feature representation need to be tested to prove that the system works well, and the proposed scheme in the XML format can be read properly. The test results showed that the system extracted features successfully, represented features in XML format, and sent the XML representation to TMC. The test results also showed that the XML representation was successfully received by TMC and read well by other systems.

Acknowledgments

Acknowledgments are given to Yayasan Budi Luhur Cakti who has funded this paper. Acknowledgments are also given to the Department of Computer Science and Electronics, Gadjah Mada University which has provided learning facilities in the Lab so that this paper can be completed properly.

References

[1] Kemenhub, “Government Regulation of the Republic of Indonesia no. 32 year 2011 concerning Management and Engineering, Impact Analysis and Requirement of Traffic,” 2011.

[2] Kemenhub, “Government Regulation of the Republic of Indonesia no. 22 year 2009 about Traffic and Road Transportation,” 22, 2009.

[3] A. Munawar, “Sustainable Urban Public Transport Planning in Indonesia , Case Studies in Yogyakarta and Jakarta,” Indones. Students’ Sci. Meet. 2008, no. May, pp. 13–15, 2008.

[4] H. Liu, S. Chen, and N. Kubota, “Intelligent Video Systems and Analytics : A Survey,” IEEE Trans. Ind. Informatics, vol. 9, no. 3, pp. 1222–1233, 2013.

[5] A. N. Belbachir, Smart Cameras. London, UK: Springer Science + Business Media, 2010.

[6] G. Li, J. Zhang, H. Lin, D. Tu, and M. Zhang, “A moving object detection approach using integrated background template for smart video sensor,” Proc. 21st IEEE Instrum. Meas. Technol. Conf. (IEEE Cat. No.04CH37510), pp. 462–466, 2004.

[7] P. Carey, New Perspectives HTML, XHTML and XML, 3rd ed. United States of America: Cengage Learning, Inc., 2010.

[8] P. Carey, New Perspectives on XML, 2nd ed. United States of America: Carey Associates, Inc., 2007.

[9] J. J. Parsons and D. Oja, New Perspective on Computer Concepts 2012. United States of America: Cengage Learning, Inc., 2012.

[10] J. Sánchez, G. Benet, and J. E. Simó, “Video sensor architecture for surveillance applications,” Sensors (Basel)., vol. 12, no. 2, pp. 1509–1528, Jan. 2012.

[11] O. Ripolles, J. E. Simó, G. Benet, and R. Vivó, “Smart video sensors for 3D scene reconstruction of large infrastructures,” Multimed. Tools Appl., Jul. 2012.

[12] Gnana Vardhini. H and A. Abraham, “Classification of XML Document by Extracting Structural and Textual Features,” vol. 2, no. 2, pp. 61–66, 2014.

[13] P. Ajitha and G. Gunasekaran, “Effective Feature Extraction for Document Clustering to Enhance Search Engine using XML,” J. Theor. Appl. Infomation Technol., vol. 68, no. 1, pp. 20–26, 2014.

[14] W. Wang, Y. Shang, J. Guo, and Z. Qian, “Real-time vehicle classification based on eigenface,” 2011 Int. Conf. Consum. Electron. Commun. Networks, pp. 4292–4295, Apr. 2011.

[15] T. C. Chen, Y. Lin, and T. Chen, “Intelligent Vehicle Counting Method Based on Blob Analysis in Traffic Surveillance,” Transportation (Amst)., pp. 1–4, 2007.

[16] J. C. Rubio, J. Serrat, A. M. López, and D. Ponsa, “Multiple-Target Tracking for Intelligent Headlights Control,” vol. 13, no. 2, pp. 594–605, 2012.

[17] B. Wu and C. Kao, “A New Approach to Video-Based Traffic Surveillance,” vol. 14, no. 1, pp. 485–491, 2013.

[18] R. S. Choras, “Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems,” vol. 1, no. 1, 2007.

[19] N. Gupta, “Comparative Study of Different Low Level Feature Extraction Techniques for Content based Image Retrieval,” vol. 1, no. 1, pp. 39–42, 2011.

[20] A. Tamrakar, S. Ali, Q. Yu, J. Liu, O. Javed, A. Divakaran, H. Cheng, H. Sawhney, and S. R. I. I. Sarnoff, “Evaluation of Low-Level Features and their Combinations for Complex Event Detection in Open Source Videos,” pp. 3681–3688, 2012.

[21] A. B. Chan and N. Vasconcelos, “Counting people with low-level features and Bayesian regression.,” IEEE Trans. Image Process., vol. 21, no. 4, pp. 2160–77, Apr. 2012.

[22] C. Ngo, Y. Jiang, X. Wei, W. Zhao, and Y. Liu, “VIREO / DVMM at TRECVID 2009 : High-Level Feature Extraction , Automatic Video Search , and Content-Based Copy Detection High-Level Feature Extraction,” 2009.

[23] S. Chang, J. He, Y. Jiang, E. El Khoury, C. Ngo, A. Yanagawa, and E. Zavesky, “High-Level Feature Extraction and Interactive Video Search,” 2008.

[24] H. Glotin, Z. Zhao, and E. Dumont, “LSIS TREC VIDEO 2009 High Level Feature Retrieval using Compact Profile Entropy Descriptors,” Entropy, vol. 2008, 2009.

[25] S. Fazli, S. Mohammadi, and M. Rahmani, “Neural Network based Vehicle Classification for Intelligent Traffic Control,” Int. J. Softw. Eng. Appl., vol. 3, no. 3, pp. 17–22, 2012.

[26] R. C. Gonzalez, R. E. Woods, and B. R. Masters, Digital Image Processing, Third Edition, vol. 14, no. 2. Prentice-Hall, Inc., 2008.

[27] B. T. Morris and M. M. Trivedi, “Learning , Modeling , and Classification of Vehicle Track Patterns from Live Video,” IEEE Trans. Intell. Transp. Syst., vol. 9, no. 3, pp. 425–437, 2008.

[28] Z. Chen, T. Ellis, and S. A. Velastin, “Vehicle Type Categorization: A comparison of classification schemes,” 2011 14th Int. IEEE Conf. Intell. Transp. Syst., pp. 74–79, 2011.

[29] P. M. Daigavane, P. R. Bajaj, and M. B. Daigavane, “Vehicle Detection and Neural Network Application for Vehicle Classification,” 2011 Int. Conf. Comput. Intell. Commun. Networks, pp. 758–762, Oct. 2011.

[30] Z. Chen and T. Ellis, “Multi-shape Descriptor Vehicle Classification for Urban Traffic,” 2011 Int. Conf. Digit. Image Comput. Tech. Appl., pp. 456–461, Dec. 2011.

[31] Z. Chen, T. Ellis, and S. A. V. Smieee, “Vehicle Detection , Tracking and Classification in Urban Traffic,” in 2012 15th International IEEE Conference on Intelligent Transportation Systems, 2012, pp. 951–956.

[32] X. Li, H. Fu, and J. Xu, “Automatic Vehicle Classification Based on Video with BP Neural Networks,” 2008 4th Int. Conf. Wirel. Commun. Netw. Mob. Comput., no. 2, pp. 1–3, Oct. 2008.

[33] S. Srivastava and P. Singh, “Real-Time Object Tracking Using Colour Feature,” Int. J. Innov. Technol. Explor. Eng., vol. 3, no. 8, pp. 56–59, 2014.

[34] Celil Ozkurt and F. Camci, “Automatic Traffic Density Estimation And Vehicle Classification For Traffic Surveillance Systems Using Neural Networks,” Math. Comput. Appl., vol. 14, no. 3, pp. 187–196, 2009.

[35] F. M. D. S. Matos and R. M. C. R. De Souza, “An image vehicle classification method based on edge and PCA applied to blocks,” 2012 IEEE Int. Conf. Syst. Man, Cybern., pp. 1688–1693, Oct. 2012.

[36] Y. Qiao and Z. Shi, “Traffic Parameters Detection Using Edge and Texture,” 2012 Int. Work. Inf. Electron. Eng., vol. 29, pp. 3858–3862, 2012.

[37] M. Kafai and B. Bhanu, “Dynamic Bayesian Networks for Vehicle Classification in Video,” IEEE Trans. Ind. Informatics, vol. 8, no. 1, pp. 100–109, 2012.

[38] N. C. Mithun, N. U. Rashid, and S. M. M. Rahman, “Detection and Classification of Vehicles From Video Using Multiple Time-Spatial Images,” IEEE Trans. Intell. Transp. Syst., vol. 13, no. 3, pp. 1215–1225, Sep. 2012.

[39] D. Huang, C. Chen, W. Hu, S. Yi, and Y. Lin, “Feature-Based Vehicle Flow Analysis and Measurement for a Real-Time Traffic Surveillance System,” J. Inf. Hiding Multimed. Signal Process. Ubiquitous Int., vol. 3, no. 3, pp. 282–296, 2012.

[40] T. Wang and Z. Zhu, “Multimodal and Multi-task Audio-Visual Vehicle Detection and Classification,” 2012 IEEE Ninth Int. Conf. Adv. Video Signal-Based Surveill., pp. 440–446, Sep. 2012.

[41] S. M. Khan, H. Cheng, D. Matthies, and H. Sawhney, “3D Model Based Vehicle Classification in Aerial Imagery,” IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1681–1687, 2010.

[42] P. Yadav and M. Kaur, “Vehicle Classification Using Novel Edge Detection Technique,” Int. J. Comput. Sci. Commun. Networks, vol. 2, no. 1, pp. 42–45, 2012.

[43] Y. Zou, H. Shi, Y. Wang, G. Shi, and H. Zhao, “A Robust Traffic Parameter Extraction Approach for Surveillance System at Urban Intersection,” 2009. CISP ’09. 2nd Int. Congr. Image Signal Process., pp. 1–5, 2009.

[44] A. Lu, L. Zhong, L. Li, and Q. Wang, “Moving Vehicle Recognition and Feature Extraction From Tunnel Monitoring Videos,” Telkomnika, vol. 11, no. 10, pp. 6060–6067, 2013.

[45] C. Singh and T. Verma, “Tracking of Moving object in Video scene using Neural Network,” Int. J. Adv. Res. Comput. Eng. Technol., vol. 1, no. 10, pp. 127–129, 2012.

[46] A. S. Jalal, “The State-of-the-Art in Visual Object Tracking,” Informatica, vol. 36, pp. 227–248, 2012.

[47] E. Ince, “Measuring traffic flow and classifying vehicle types : A surveillance video based approach,” Turk J Elec Eng Comp Sci, vol. 19, no. 4, pp. 607–620, 2011.

[48] J. Jumi, A. Harjoko, and A. Ashari, “Content-Based Image Retrieval for Asset Management Based On Weighted Feature and K-Means Clustering,” J. Theor. Appl. Inf. Technol., vol. 77, no. 1, pp. 116–124, 2015.

[49] N. U. Rashid, N. C. Mithun, B. R. Joy, and S. M. M. Rahman, "Detection And Classification of Vehicles from A Video Using Time-Spatial Image," 2010 Int. Conf. Electr. Comput. Eng., pp. 502–505, 2010.
[50] Z. Chen, N. Pears, M. Freeman, and J. Austin, "Road Vehicle Classification using Support Vector Machines," 2009. ICIS 2009. IEEE Int. Conf. Intell. Comput. Intell. Syst., vol. 4, pp. 214–218, 2009.
[51] M. Wolf and C. Wicksteed, "Date and time formats," W3C NOTE NOTE-datetime-19980827, no. September 1997, pp. 12–13, 1997.
[52] T. Bray, J. Paoli, C. M. Sperberg-McQueen, and E. Maler, "Extensible Markup Language (XML) 1.0 (Fifth Edition)," Keio, 2013.

Imelda Imelda is a lecturer and researcher at Faculty of Information Techonology, Universitas Budi Luhur, Jakarta, Indonesia. She received the B.S. degrees in Computer Science, STMIK Budi Luhur, Jakarta, Indonesia in 1997 and the master degree in Information Technology, Universitas Indonesia, Jakarta, Indonesia in 2006. Now, she is a doctoral student in Electronics and Instrumentation, Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia. Her research interest include multimedia processing (image, audio, video processing), computer vision, computer security and database management system.

Agus Harjoko is a lecturer and a researcher at Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada. He is also a member of the Indonesian Physics Society and the Indonesian Mathematical Society. He received the B.S. degree in Electronics and Instrumentation from Universitas Gadjah Mada, Yogyakarta, Indonesia and the M.Sc and Ph.D degree in Computer Science both from University of New Brunswick, Canada, in 1986, 1990 and 1996. His research interests include multimedia processing (image, audio, video processing), computer vision and medical instrumentation.

Pekik Nurwantoro is a lecturer and a researcher at Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada. He is also a member of the Indonesian Physics Society. He received the B.S. and M.S. degree in Physics from Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia and the Ph.D degree in Computational Physics, from The University of Birmingham, United Kingdom, in 1987, 1991, and 1998. His research interests include computational physics, oscillator and superconductor.