A rapid abnormal event detection method for surveillance video based on a novel feature in compressed domain of HEVC

CONFERENCE PAPER · JULY 2014

DOI: 10.1109/ICME.2014.6890212

Hongyang Chao

Sun Yat-Sen University

A RAPID ABNORMAL EVENT DETECTION METHOD FOR SURVEILLANCE VIDEO BASED ON A NOVEL FEATURE IN COMPRESSED DOMAIN OF HEVC

[1]Huang LI, [2]Yihao ZHANG, [2]Ming YANG, [1]Yangyang MEN, [1]*Hongyang CHAO

[1] School of Software & [2] School of Info. Sci. and Tech., Sun Yat-sen University, China 510275

ABSTRACT

Abstract: Event detection plays an essential role in video content analysis. On the other hand, according to our analysis, the coding structures in new video coding standard High Efficient Video Coding (HEVC) have a high correlation with video contents. Hence there is large potential to identify events by reusing coding structures in HEVC, which can save a huge amount of computational resources. In this paper, we proposed a new compressed-domain feature for abnormal event detection, namely Motion Intensity Count (MIC), which makes use of motion vectors, coding unit and prediction unit modes in HEVC with little computational cost. MIC can well predict the normal paths of moving objects, which enables us to identify motions in unexpected locations where abnormal events are likely to happen. Our experiments show that MIC can correctly detect abnormal events at about 1250 fps.

Index Terms— Abnormal Event Detection, Traffic Accident, Motion Intensity Count, Compressed Domain, HEVC

1. INTRODUCTION

Abnormal event detection plays an essential role in video content analysis. For example, advances in video compression, networking and storage hardware led to the widespread use of traffic video surveillance systems. These systems are characterized by generating and managing a large volumes of video sequences. There arises huge needs for automatic video analysis in order to detect and alarm on abnormal events in real time.

Currently most abnormal event detection methods [7-19] are in pixel domain. However, these methods can be computational intensive because they need to extract video features from raw pixels or estimate the flow dynamics by employing block matching, phase correlation, or gradient based approaches. It makes things more complicated that video data are often compressed by hardware in camera because they save lots of bandwidth and disk space for transmitting and storing purpose. In this situation, pixel-domain event detection could be even more time-consuming because it requires additional resources for fully decoding videos bitstreams into raw pixels.

Challenges arise when video surveillance systems with limited resources demands real-time event detection in multiple video streams, which is a common case in traffic video surveillance systems.

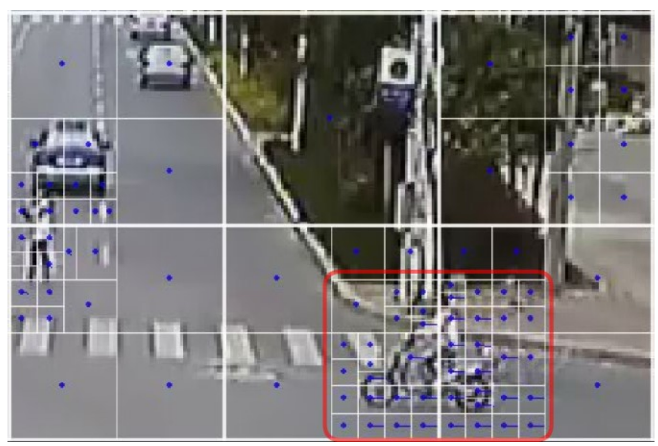


Fig. 1 The compress-domain information in HEVC reflects object segmentations and motion information.

On the other hand, one favorable aspect of compressed videos is that it already contains extra information that may be useful for event detection: additional information such as partition modes and motion vectors can be viewed as coarse analysis of object boundary and optical flow. Moreover, comparing to previous video coding standard H.264 [20], the latest standard High Efficiency Video Coding (HEVC) [21], which came out in early 2013, have even greater potential for compressed-domain event detection. The coding structures of HEVC provide with more information in its compressed domain, because it employs more sophisticated coding technologies such as quad-tree coding structure and more flexible partition mode, which can be well adapted to various video content, resulting in more accurate optical

This work was partially supported by NSF of China under Grant 61173081 and Guangdong Natural Science Foundation, P.R.China, under Grant S2011020001215.

*Corresponding author: Hongyang CHAO (isschhy@mail.sysu.edu.cn)

flow and object boundary approximations. An example is shown in Fig. 1. The moving motorbike is outlined by small blocks and its motion is roughly lined out by motion vectors (MVs). These compressed-domain information can be achieve with little computational cost from a compressed video stream.

How to manipulate the information in compressed-domain information of HEVC to detect abnormal event is exactly the main problem to solve in this paper. As the definitions of abnormal event varies on different occasion, we will use traffic accident detection as an example of the abnormal event detection in this paper.

Existing traffic accident detection algorithms mainly fall into two categories: compressed-domain and pixel-domain approaches. As HEVC just came out in early 2013, currently there is no published work on event detection in the compressed domain of HEVC. For H.264, there are a few related work, which involves object segmentation, tracking, and motion detection [1-6]. As HEVC have different coding structures with H.264, these method cannot be directly applied to HEVC. In pixel domain, there have been a number of traffic accident detection algorithms [7-19]. [7-9] implemented traffic accident detection systems by using Hidden Markov Model. Hu et al predict traffic accidents using 3-D model-based vehicle tracking [10, 11]. Lin et al presented a hardware implementation of real-time accident detection by image tracking [12]. Veeraraghavan et al. predicted traffic accidents by multilevel motion tracking using Kalman filter [13]. Beside these, vehicle behavior and trajectory learning are extensively discussed in [14-19]. All these pixel-domain algorithms need to transform compressed-domain information into pixel domain. In this process, some compress-domain information is loss, such as MVs and partition modes. If we can make use of these information to detect traffic accidents in compressed domain directly, huge amounts of computational resources can be saved.

Aiming to reduce computational cost of abnormal event detection, we develop a fast abnormal event detection algorithm for traffic video surveillance by manipulating the information in compressed domain of HEVC in this paper. The proposed method can accurately detect abnormal events such as traffic accidents with little computational cost by identifying unexpected changes of motion intensity which is measure by our new compressed-domain feature, namely motion intensity count (MIC).

There are mainly two contributions in this paper: the first is that we have proposed a compressed-domain video feature MIC to measure the motion intensity within a video region and a motion intensity transmission method to predict MICs for subsequent frames; the second is that, by identifying unexpected changes of MICs, we develop an abnormal event detection algorithm to detect abnormal events in compressed domain.

The rest of this paper is organized as follows. In section 2, we will introduce our first contribution: the compressed-domain feature MIC and the MIC prediction method. Based on the first contribution, the proposed abnormal event detection algorithm is discussed in section 3, which is our second contribution. The experimental results are given in section 4.

2. MOTION INTENSITY COUNT AND PREDICTION

Our proposed abnormal event detection algorithm will relay on identifying the difference between predicted and actual motion intensity. In this section: a compressed-domain feature for motion intensity measurement and a motion intensity perdition method will be introduced in section 2.1 and section 2.2 respectively.

2.1 Motion Intensity Count (MIC)

For better understanding of the proposed feature, we will firstly briefly explain the related compressed-domain information in HEVC.

Fig. 2 LCU, PU and MV in HEVC

In HEVC, each frame is firstly divided into squares of equal size, i.e. largest coding units (LCUs). Each LCU can be recursively divided into smaller unit until a proper partition reached. Each smallest unit is referred to as prediction unit (PU). For each inter-coded PU, there will be a motion vector (MV) indicating the best match position in the reference frame. As shown in Fig. 2, the red block is one of the LCUs, the green block is one of the PUs and blue lines indicate MVs.

According to our observations and analysis, LCU with intense motion have longer MVs and finer partitions. Thus we propose a new feature, namely motion intensity count (MIC), to represent the motion intensity within a LCU. MIC for each LCU is defined as follows:

$$\mathrm{MIC} = \sum_{i \in \sigma} (\|MV_i\| * area(PU_i)) \ . \tag{1}$$

Here, $\{MV_i | i \in \sigma\}$ is the set of the MVs within the same LCU, $area(PU_i)$ is the area of the PU containing this MV. With this formula, LCU with larger MVs will have higher MIC. An example of MIC calculation for one frame containing 4 LCUs is shown in Fig. 3. The numbers in green indicate the MV. For the top-left LCU with 4 MVs, its MIC is calculate by the formula in purple.

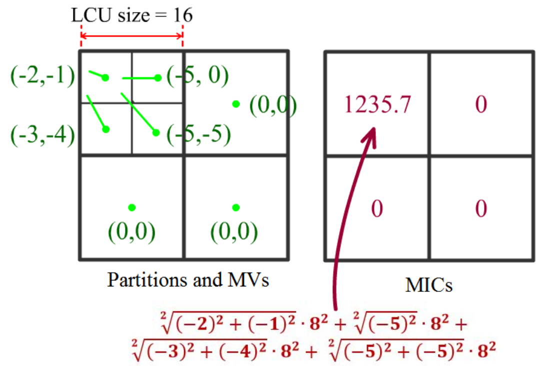
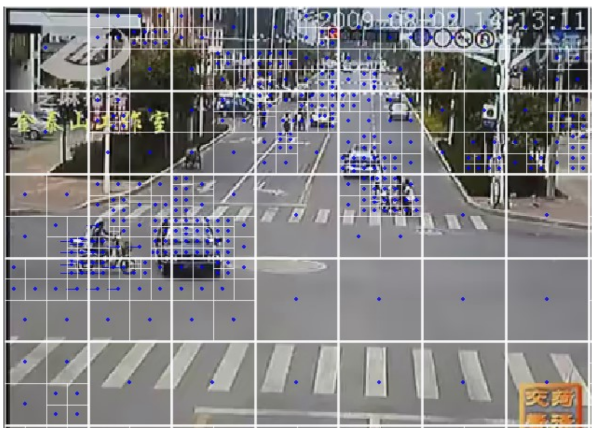
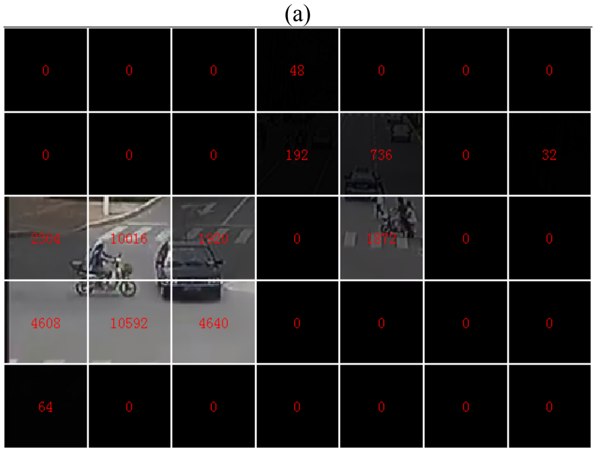


Fig. 3 An example of MIC calculation for one frame containing 4 LCUs



(a)



(b)

Fig. 4 An example of MIC. (a) One original frame with partition and MV information. (b) The frame with MIC for each LCU. Brighter LCU indicates larger MIC.

In order to show the effectiveness of MIC, another real-world example of MIC calculation is shown in Fig. 4. Fig. 4(a) is the original frame with partition and MV information. Fig. 4(b) is the same frame with MIC calculated by above formula for each LCU. A brighter LCU indicates a larger MIC. We can see that the LCUs containing the fast moving motorbike on the left have the larger MICs because there are more intensive motion there than other LCUs in this frame.

2.2 Motion Intensity Count Prediction

In this subsection, we develop a motion intensity transmission method to predict motion intensity changes between different frames caused by uniform linear motion.

In general, normal movement of pedestrians and vehicles can roughly be regarded as uniform linear motions and small deformation within several adjacent frames, especially when frame rate is high. Therefore, for normal motions, we may predict MIC changes linearly for subsequent frames. In contrast, abnormal traffic events such as traffic accidents often come along with sudden and huge changes in both direction and speed. That is to say, object movements involving in a traffic accidents are far from uniform linear motion, which is not linear predictable.

According to the analysis above, in order to distinguish normal events from abnormal ones, we hereby develop a motion intensity transmission model which can predict changes caused by normal movements, i.e. uniform linear motion. The model is illustrated in Fig. 5. For each PU containing a MV in frame n, we predict its new center position in frame m with following formula:

$$P_{pred} = P_{cur} - (m - n) * MV \qquad (2)$$

Here P_{cur} is the center position of the PU in frame n, P_{pred} is the center position of the predicted position in frame m. Then the predicted MIC for each LCU in frame m can be calculated according to the predicted PU locations and MVs. This motion intensity transmission model can well predict the MIC changes caused by linear motion. In other words, if this model fail to predict the MICs in one frame, there is a high probability that abnormal events appear, i.e. traffic accidents happens in this case.

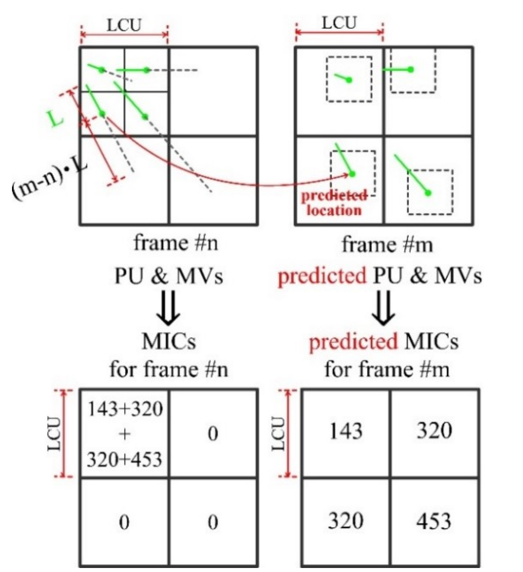


Fig. 5 The proposed motion intensity transmission method

3. PROPOSED ABNORMAL EVENT DETECTION ALGORITHM

Based on the proposed feature MIC in section 2.1 and the MIC prediction method section 2.2, we introduce our new algorithm to detect abnormal traffic event by identifying the difference between the predicted and actual MICs for each frame.

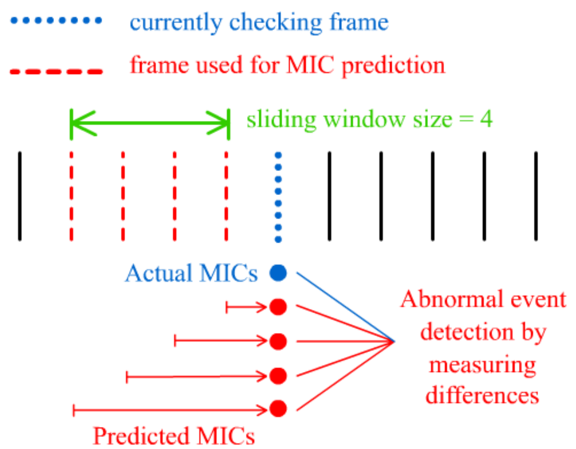


Fig. 6 The proposed abnormal event detection algorithm

The proposed method is depicted in Fig. 6. For each currently checking frame, we predict MICs for current frame by previous *s* frames, which are inside a sliding window of size *s*. For each LCU in current frame, with the predicted MICs and actual MIC we can measure the differences by the following formula:

$$d = \frac{1}{s}\sum_{i=1}^{s}(MIC\_Pred_i - MIC\_Actual)^2 \qquad (3)$$

Here, MIC_Pred_i is the predicted MIC for current LCU by previous i th frame. MIC_Actual is the actual MIC of current LCU. If d exceeds a predefined threshold, then this LCU is identified as containing objects with abnormal events, which are typically traffic accidents for traffic video surveillances.

4. EXPERIMENTAL RESULT

In this section we would evaluate the accuracy and speed of the proposed abnormal event detection algorithm in compressed-domain. All the experiments were run on Windows 7 with Intel Xeon E5420 at 2.50GHz. All videos is compressed with IPPP coding structure. Quantization parameter (QP) is 28. The sliding window size is 5, and the threshold on *d* is 120,000. The proposed abnormal event detection algorithm is evaluated on 16 traffic accidents surveillance videos and 20 normal traffic surveillance videos in MIT traffic dataset[1], which are about 120 minutes long in total. The algorithm is implemented with C++.

[1] http://www.ee.cuhk.edu.hk/~xgwang/MITtraffic.html

Results show that the algorithm can accurately identify all the traffic accidents in 16 videos. On the other hand, only 0.10% was misidentified as traffic accidents in the 20 normal video in MIT traffic dataset when we take "second" as the unit of measurement. Average time consumption for one 480x360 frame is 0.0008s. That is to say, it can process as much as 1250 frames per second.

Fig. 7(a) is a typical LCU with normally moving object. The predicted MICs (red bars) are very close to the real MIC (blue bar). On the contrary, in Fig. 7(b), the predicted MICs are much larger than real MIC because the motorbike come to a stop after hitting a car, which cause an unexpected drop on MIC. The number in red indicates the difference *d* between predicted and actual MICs, which is calculated by formula (3). More results are shown in Fig. 8, LCUs with red boundary indicate the video region with abnormal events.

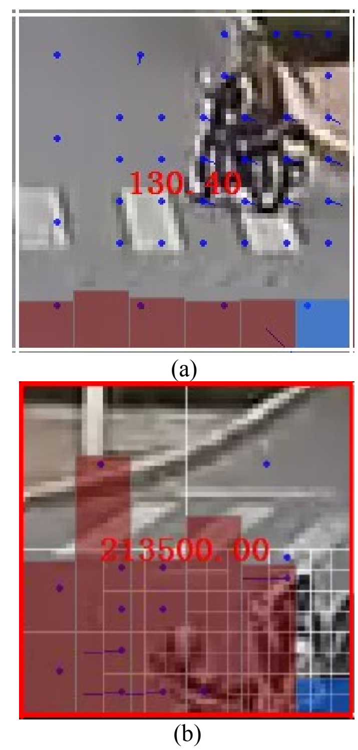


Fig. 7 (a) A typical LCU with normal actions.
(b) A typical LCU with accidents.
(Red bar: predicted MIC. Blue bar: actual MIC)

5. CONCLUSION

In this paper we have proposed an abnormal event detection algorithm based on a novel video feature MIC, which is extracted from compressed-domain information of HEVC. We use traffic accident detection as an example to demonstrate the effectiveness of our algorithm. Experiments show that it can accurately detect abnormal events at 1250 fps for 360p surveillance video.

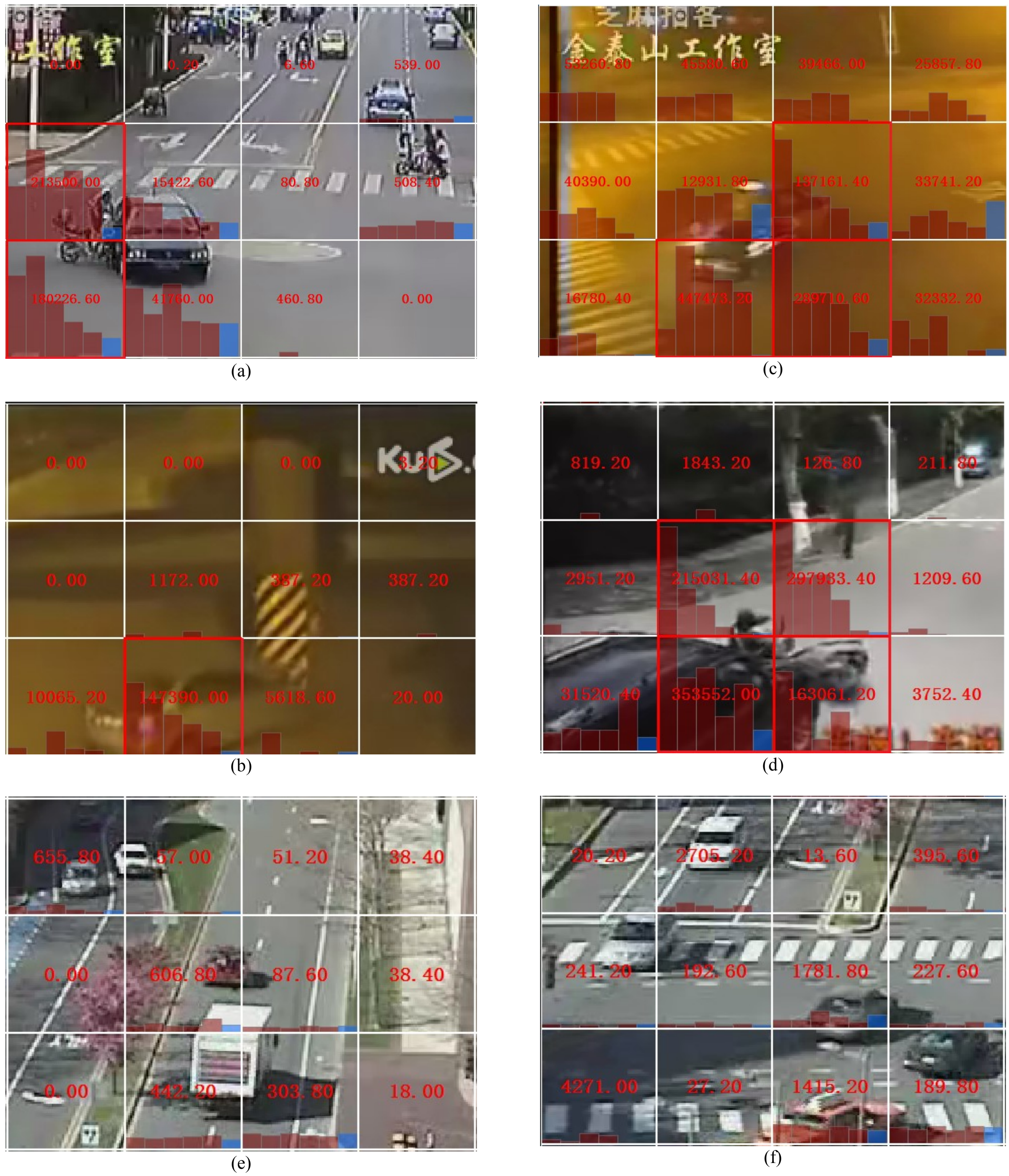


Fig. 8 Some results of abnormal event detection for traffic surveillance videos.
(a)(b)(c)(d) videos with accidents. (e)(f) videos without accidents.

6. REFERENCES

[1] O Akoz, M.E Karsligil, “Video-based traffic accident analysis at intersections using partial vehicle trajectories”, 17th IEEE International Conference on Image Processing (ICIP), pp. 4693-4696, Sept. 2010.

[2] Kun Tao, Shouxun Lin, Yongdong Zhang, “Compressed Domain Motion Analysis for Video Semantic Events Detection”, International Conference on Information Engineering, vol.1, pp. 201-204, July 2009.

[3] Haowei Liu, Ming-Ting Sun, Ruei-Cheng Wu, Shiaw-Shian Yu, “Video activity detection using compressed domain motion trajectories for H.264 videos” , IEEE International Symposium on Circuits and Systems (ISCAS), pp. 3693-3696, June 2010.

[4] Lin, Chia-Wen; Zhi-Hong Ling, “Automatic Fall Incident Detection in Compressed Video for Intelligent Homecare”, 16th International Conference on Computer Communications and Networks, pp. 1172-1177, Aug. 2007.

[5] M. Laumer, P. Amon, A. Hutter, A. Kaup, “A compressed domain change detection algorithm for RTP streams in video surveillance applications” , 13th International Workshop on Multimedia Signal Processing (MMSP), pp. 1-6, Oct. 2011.

[6] K. Yoon, D. DeMenthon, D. Doermann, “Event detection from MPEG video in the compressed domain”, 5th International Conference on Pattern Recognition, vol.1, pp. 819-822, 2000.

[7] Y. Zou, G. Shi, H. Shi, Y. Wang, “Image Sequences Based Traffic Incident Detection for Signaled Intersections Using HMM”, 9th International Conference on Hybrid Intelligent Systems, 2009.

[8] N. Saunier and T. Sayed, “Clustering Vehicle Trajectories with Hidden Markov Models, Application to Automated Traffic Safety Analysis”, International Joint Conference on Neural Networks, pp. 4132-4138, Vancouver, July 2006.

[9] S. Kamijo, Y. Matsushita, K. Ikeuchi, M. Sakuchi, “Traffic Monitoring and Accident Detection at Intersections”, IEEE Transactions on Intelligent Transportation Systems, vol. 1, no. 2, pp. 108-118, June 2000.

[10] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, S. Maybank, “A System for Learning Statistical Motion Patterns”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 9, pp. 1450-1464 ,Sept. 2006.

[11] W. Hu, X. Xiao, D. Xie, T. Tan, S. Maybank, “Traffic Accident Prediction Using 3-D Model-Based Vehicle Tracking”, IEEE Transactions on Vehicular Technology, vol. 53, no. 3, pp. 677-694, May 2004.

[12] C. Lin, J. Tai, K. Song, “Traffic Monitoring Based on Real-Time Image Tracking”, in Proc. of the IEEE International Conference on Robotics & Automation, vol. 2, pp. 2091-2096, September 2003.

[13] H. Veeraraghavan, O. Masoud, N. Papanikolopoulos,“Computer Vision for Intersection Monitoring”, IEEE Transactions on Intelligent Transportation Systems, vol. 4, no. 2, pp. 78-89,June 2003.

[14] B. Morris and M. Trivedi, “A Survey of Vision-BasedTrajectory Learning and Analysis for Surveillance”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 8, pp. 1114-1127, August 2008.

[15] C. Piciarelli, G.L. Foresti, “Online Trajectory Clustering for Anomalous Events Detection”, Pattern Recognition Letters27, 1835-1842, 2006.

[16] F. Porikli, “Learning Object Trajectory Patterns by Spectral Clustering”, IEEE International Conference on Multimedia and Expo, 2004

[17] T. Liao, “Clustering of Time Series Data”, Pattern Recognition 38, pp. 1857-1874, 2005.

[18] X. Wang, K. Tieu, E. Grimson, “Learning Semantic Scene Models by Trajectory Analysis”, 9th European Conference on Computer Vision, pp 110-123, May 2006.

[19] W. Hu, X Xiao, D. Xie, T. Tan, “Traffic Accident Prediction Using Vehicle Tracking and Trajectory Analysis”, in Proc. of the IEEE Intelligent Transportation Systems, vol. 1, pp. 220-225, Oct. 2003.

[20] T. Wiegand, G.J. Sullivan, G. Bjontegaard, A. Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Trans. Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003.

[21] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, ”Overview of the High Efficiency Video Coding (HEVC) Standard”, IEEE Trans. Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.