Second Asia Pacific Conference on Advanced Research (APCAR, Melbourne, February, 2016)
ISBN:978 0 9943656 37
www.apiar.org.au

REAL-TIME PERFORMANCE EVALUATION METRICS FOR OBJECT DETECTION AND TRACKING OF INTELLIGENT VIDEO SURVEILLANCE SYSTEMS

Kyoung Yeon Kim [a], Yonghwa Kim [b], Jaehyun Park [c], Yoo-Sung Kim [d]
[abcd] INHA University, Incheon 402-751, Korea
Corresponding email: ruddus95@naver.com

Abstract

This paper proposes real-time performance evaluation metrics for the object detection and tracking algorithms used in intelligent video surveillance systems. Since an intelligent video surveillance system should support real-time performance, as well as the spatial accuracy of the object detection and tracking, the performance evaluation tool should evaluate not only the spatial accuracy, but also the real-time property. In this paper, a total of eight performance evaluation metrics for the real-time object detection and tracking system have been proposed. The proposed metrics have been designed by combining the degree of precision of the detected objects, the ratio of the incorrectly detected objects, and real-time performance of the algorithm.

Keywords: Intelligent Video Surveillance, Object Detection, Performance Evaluation Metric, Real-time Performance Evaluation

1.Introduction

Nowadays, the intelligent video surveillance systems have been widely used to realize a safe society. These systems can automatically recognize criminal events from CCTV videos using real-time detection techniques and then notify the securities of the event occurrences (Davies & Velastin, 1990; Han & Seo, 2009; Yun et al., 2009). To recognize criminal events in time, the intelligent video surveillance systems must have not only the accurate object detection capabilities from the CCTV videos, but also the real-time property of the detection algorithms.

To measure the performance of the intelligent surveillance systems, several tools and techniques such as USF-DATE(USF-Detection And Tracking Evaluation), Object Video VEW System, and ViPER-PE(Video Performance Evaluation Resource-Performance Evaluation) have been proposed (Doermann & Mihalcik, 2000; Haering, Venetianer & Lipton, 2008; Kasturi et al., 2009). Among them, ViPER-PE has been widely used. This system was developed to measure the accuracy of the object detections by comparing the candidate records from the detection results with the target records from the ground truth data in terms of the spatial locations of the objects (Doermann & Mihalcik, 2000; Bashir & Porikli, 2006).

For the real-time crime report and protection feature that are widely adopted in recent novel intelligent surveillance systems, the timing performance of the intelligent surveillance system plays an important role in addition to the spatially object detection accuracy. However, most of the existing performance evaluation tools used for the intelligent video surveillance systems focus on the performance of the location-based object detections (Coifman et al., 1998). For example, since ViPER-PE uses the XML schema to represent the ground truth data and the detection results, it does not inherently support to evaluate the timing performance in object recognitions (Kim et al.,

ISBN:978 0 9943656 37
www.apiar.org.au

2013).

In this paper, we propose the metrics that can be used to evaluate the real-time performance of object detection and tracking. This paper is prepared as follows. Section 2 discusses the previous related studies on the performance evaluation for intelligent video surveillance systems. In section 3, a core set of metrics for performance evaluation of the real-time object detection and tracking is proposed. We conclude this paper in section 4.

2. Related Works

To our knowledge, there is no metric for the performance evaluation of the real-time object detection and tracking and no tool that can be useful for evaluating the performance of the real-time object detection and tracking of intelligent video surveillance systems (Bashir & Porikli, 2006; Kim et al., 2013). On the other hand, some metrics to measure the performance of object detections (using simple spatial accuracy) and several performance evaluation tools have been proposed (Doermann & Mihalcik, 2000; Bashir & Porikli, 2006).

One of the major evaluation metrics for the object detection and tracking accuracy is CLEAR Metrics. CLEAR Metrics consist of *MODP (Multiple Object Detection Precision)*, *MODA (Multiple Object Detection Accuracy)*, *MOTP(Multiple Object Tracking Precision)*, and *MOTA (Multiple Object Tracking Accuracy)*. *MODP* uses the mean value of the spatial overlap ratio of the successfully detected objects by calculating the spatial overlap ratio between the ground truth data and the test data recognized by the recognition program for one frame. *MODA* uses the numbers of missed objects and false alarms to evaluate for one frame. *MOTP* calculates the spatial overlap ratio between the ground truth data and test data for all frames and uses the mean of the overlap ratio of successfully detected objects for evaluation. *MOTA* uses the numbers of missed objects, false alarms, and ID switches to evaluate one frame. They are calculated as following equations:

$$MODP^k = \frac{\sum_{i=1}^{M^k} \frac{|G_i^k \cap D_i^k|}{|G_i^k \cup D_i^k|}}{M^k} \tag{1}$$

where G_i^k denotes the i^{th}ground truth object in k^{th} frame. D_i^k denotes the detected object for G_i^k. M^k is the number of mapped object pairs in the k^{th} frame.

$$MODA^k = 1 - \frac{c_m m_k + c_f p_k}{N_G^k} \tag{2}$$

where c_m and c_f are the cost functions for the missed detects and false alarm penalties. N_G^k is the number of ground truth objects in the k^{th} frame. m_k and p_k are the number of missed object in frame k and the number of false positive, respectively.

$$MOTP = \frac{\sum_{i=1}^{M} \sum_{k=1}^{N} \frac{|G_i^k \cap D_i^k|}{|G_i^k \cup D_i^k|}}{\sum_{j=1}^{N} M^j} \tag{3}$$

where N means total number of frames and Mrefers to the mapped objects over the entire track as opposed to just the frame.

$$MOTA = 1 - \frac{\sum_{k=1}^{N}(c_m m_k + c_f p_k + c_s\, ID\ SWITCHES_k)}{\sum_{j=1}^{N} N_G^j} \tag{4}$$

where $ID\ SWITCHES_k$ is the total number of ID switches made by the detected objects for any

ISBN:978 0 9943656 37
www.apiar.org.au

given reference ID.c_s is the cost functions for the ID switches.

As indicated, since CLEAR Metrics only uses the spatial overlap accuracy data without the real-time data for evaluation, in the existing object detection and tracking algorithm evaluation method, the real-time evaluations were simply conducted by evaluating the algorithm execution time and the number of frame loss (Haritaoglu, Harwood & Davis, 2000; Kandhalu et al., 2009). However, the value of recognition is deduced if the execution time is too long and the timing constraints are missed even if the recognition accuracy is very high. For such evaluation purpose, several real-time evaluation methods for evaluation of the accuracy values according to the execution time have been proposed in the other application areas (Abbott & Garcia-Molina, 1988; Kopetz, 2011; Lee, Shin & Easwaran, 2012).

To define the real-time performance, real-time evaluation methods can be considered as either the hard real-time evaluation or soft real-time evaluation. The hard real-time evaluation refers to an evaluation method where the timing constraint is crucial for the recognition performance. Further, the soft real-time evaluation refers to an evaluation method where the scheduled amount of penalty is given according to the value functions if the timing deadline is exceeded at certain amount. Fig. 1 shows the examples of the value functions.

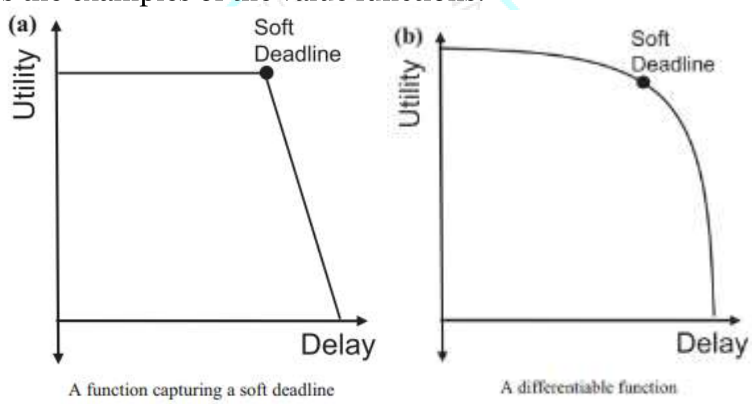

Figure 1: Value functions (Lee, Shin & Easwaran, 2012)

3. Real-time Performance Evaluation Metrics

The real-time object detection and tracking evaluation metrics proposed in this paper is defined by applying penalty to the CLEAR Metrics if the recognition algorithm is not completed within the time deadline. In the hard real-time systems, such as the bomb or hazardous detection system, if the time deadline in the object detection and tracking is missed, the whole system becomes useless. Hence, we defined the hard real-time object detection and tracking evaluation metrics as the extensions of the above CLEAR Metrics by putting the score zero if the deadline is missed. Based on such information, following hard real-time evaluation metrics for the object detection and tracking of the intelligent surveillance systems can be defined. As the object detection evaluation metrics of CLEAR Metrics,$MODP_H$*(Hard Real-time MODP)*and$MODA_H$*(Hard Real-*

ISBN:978 0 9943656 37
www.apiar.org.au

time MODA) are defined to evaluate the precision and accuracy of object detection, respectively, in each frame of input test video.

$$MODP_H^k = \begin{cases} MODP^k & t \le \tau \\ 0 & t > \tau \end{cases} \quad (5)$$

$$MODA_H^k = \begin{cases} MODA^k & t \le \tau \\ 0 & t > \tau \end{cases} \quad (6)$$

$$MOTP_H^k = \begin{cases} \frac{\sum_{i=1}^{M^{1..k}} \frac{|G_i^k \cap D_i^k|}{|G_i^k \cup D_i^k|}}{M^k} & t \le \tau \\ 0 & t > \tau \end{cases} \quad (7)$$

$$MOTA_H^k = \begin{cases} 1 - \frac{c_m m_k + c_f p_k + c_s ID\ SWITCHES_k'}{N_G^k} & t \le \tau \\ 0 & t > \tau \end{cases} \quad (8)$$

$$MOTP_H = \frac{\sum_{k=1}^{N} MOTP_H^k}{N} \quad (9)$$

$$MOTA_H = \frac{\sum_{k=1}^{N} MOTA_H^k}{N} \quad (10)$$

However, since existing MOTP and MOTA of the CLEAR Metrics evaluate the precision and accuracy of the object tracking, respectively, for the entire frame of the input test video, an evaluation equation should be added to evaluate the object tracking for each frame and the real-time object tracking function. For such purpose, $MOTP_H^k$ and $MOTA_H^k$, which indicate the precision and accuracy of hard real-time track in one frame respectively, have been defined. $MOTP_H^k$ of equation (7) is defined by the overlap ratio of the objects successfully tracked up to the k^{th} frame with satisfying time deadline for the precision of object tracking evaluation at the k^{th} frame of the input video. $M^{1..k}$ represents the number of successfully tracked objects from the first frame to the k^{th} frame and M^k represents the number of successful mapped objects in the k^{th} frame. As in equation (8), $MOTA_H^k$ evaluates the object tracking accuracy at the k^{th} frame of the input video with error recognition, detection failure, and ID switch frequency information upon the deadline fulfillment. $ID\ SWITCHES_k'$ refers to the number of objects tracked at the k^{th} frame with mismatched IDs from a pool of mapped IDs of successfully tracked up to the k^{th} frame. $MOTP_H^k$ and $MOTA_H^k$ are considered as tracking failure and evaluated the results as zero, respectively when the recognition time exceeds the deadline. $MOTP_H$ and $MOTA_H$ of the

ISBN:978 0 9943656 37

equations (9) and (10) are defined as the mean values of $MOTP_H^k$ and $MOTA_H^k$ for all frames as the hard real-time object tracking performance evaluation Metrics. Hence, these hard real-time evaluation metrics, $MODP_H$ *(Hard Real-time MODP),* $MODA_H$ *(Hard Real-time MODA),* $MOTP_H$*(Hard Real-time MOTP),* and $MOTA_H$*(Hard Real-time MOTA)* can be used to detect and track important object algorithms such as bomb or hazardous materials.

On the contrary, for the soft real-time system where the value of the recognition decreases significantly during a certain period before it goes meaningless, a value function can be used to extending the existing CLEAR Metrics. Again, the soft real-time evaluation metrics for the object detection and tracking of the intelligent surveillance systems can be defined as follows; hereafter, t is the runtime, τ is the dead line, and $vf(x,t)$ is the value function after the timing deadline.

$$MODP_S^k = \begin{cases} MODP^k t \leq \tau \\ vf\left(MODP^k, \mathrm{t}\right) t > \tau \end{cases} \tag{11}$$

$$MODA_S^k = \begin{cases} MODA^k t \leq \tau \\ vf\left(MODP^k, t\right) t > \tau \end{cases} \tag{12}$$

$$MOTP_S^k = \begin{cases} \dfrac{\sum_{i=1}^{M'} \frac{|G_i^k \cap D_i^k|}{|G_i^k \cup D_i^k|}}{M^k} & t \leq \tau \\ vf\left(\dfrac{\sum_{i=1}^{M'} \frac{|G_i^k \cap D_i^k|}{|G_i^k \cup D_i^k|}}{M^k}\right) & t > \tau \end{cases} \tag{13}$$

$$MOTA_S^k = \begin{cases} 1 - \dfrac{c_m m_k + c_f p_k + c_s ID\ SWITCHES_k'}{N_G^k} & t \leq \tau \\ vf\left(1 - \dfrac{c_m m_k + c_f p_k + c_s ID\ SWITCHES_k'}{N_G^k}\right) & t > \tau \end{cases} \tag{14}$$

$$MOTP_S = \frac{\sum_{k=1}^{N} MOTP_H^k}{N} \tag{15}$$

$$MOTA_S = \frac{\sum_{k=1}^{N} MOTA_S^k}{N} \tag{16}$$

$MOTP_S^k$ and $MOTA_S^k$ have been additionally defined to evaluate the performance of the soft real-time object tracking system frame-by-frame as in the hard real-time performance evaluation

Metrics. $MOTP_S^k$ and $MOTA_S^k$ of the equations (13) and (14), which evaluate the results of the object tracking precision and accuracy in one frame, respectively, equal to $MOTP_H^k$ and $MOTA_H^k$ if the recognition time satisfies the deadline. If the recognition time exceeds the deadline, the accuracy is evaluated with the value function or penalty value added to the successful result. $MOTP_S$ and $MOTA_S$ of the equations (15) and (16) are the soft real-time object tracking performance evaluation metrics and further defined as the average of $MOTP_H$ and $MOTA_H$ for all frames.

The soft real-time evaluation metrics, $MODP_S$ *(Soft Real-time MODP),* $MODA_S$ *(Soft Real-time MODA),* $MOTP_S$ *(Soft Real-time MOTP),* and $MOTA_S$ *(Soft Real-time MOTA)* can be used to evaluate the object detection and tracking which is not seriously dangerous regardless of the timing deadline. The value function for the soft real-time evaluation metrics can be various functions according to the characteristic of the system and the purpose of the recognition program.

Conclusions

In this paper, we have proposed eight performance evaluation metrics for the real-time object detection and tracking of the intelligent video surveillance systems. Such metrics overcome the limitations of the conventional evaluation metrics by evaluating the real-time performance as well as the object detection and tracking accuracy that can be evaluated through the spatial accuracy and real-time evaluation.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (B0101-15-1282-00010002, Suspicious pedestrian tracking using multiple fixed cameras).

ISBN:978 0 9943656 37
www.apiar.org.au

References

i. Abbott, R. & Garcia-Molina, H., 1988. Scheduling Real-Time Transactions. *ACM SIGMOD Record - Special Issue on Real-Time Database Systems*, 17(1), pp. 71-81.

ii. Bashir, F. & Porikli, F., 2006. *Performance Evaluation of Object Detection and Tracking Systems*. Proc. IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 7-14.

iii. Coifman, B., Beymer, D., McLauchlan, P. & Malik, J., 1998. A Real-Time Computer Vision System for Vehicle Tracking and Traffic Surveillance. *Transportation Research Part C: Emerging Technologies*, 6(4), pp. 271-288.

iv. Davies, A. C. & Velastin, S. A., 1990. *A Progress Review of Intelligent CCTV Surveillance Systems*. Proc. 8th IEEE International Conf. Intelligent Data Acquisition and Advanced Computing Systems, Sofia, pp. 417-423.

v. Doermann, D. & Mihalcik, D., 2000. *Tools and Techniques for Video Performance Evaluation*. Proc. International Conf. on Pattern Recognition, pp. 167-170.

vi. Haering, N., Venetianer, P. L. & Lipton, A., 2008. The Evolution of Video Surveillance: An Overview. *Machine Vision and Applications*, 19(5), pp. 279-290.

vii. Han, T. W. & Seo, Y. H., 2009. Emergency Situation Detection using Images from Surveillance Camera and Mobile Robot Tracking System. *Journal of the Institute of Webcasting, Internet and Telecommunication*, 9(5), pp. 101-107.

viii. Haritaoglu, I., Harwood, D. & Davis, L. S. W., 2000. 4: Real-Time Surveillance of People and Their Activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), pp. 809-830.

ix. Kandhalu, A., et al., 2009. *Real-Time Video Surveillance over IEEE 802.11 Mesh Networks*. Proc. 15th IEEE Conf. on Real-Time and Embedded Technology and Applications Symposium, pp. 205-214.

x. Kasturi, R., et al., 2009. Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2), pp. 319-336.

xi. Kim, J. S., et al., 2013. *A Semi-Automatic Video Annotation Tool to Generate Ground Truth for Intelligent Video Surveillance Systems*. Proc. International Conf. on Advances in Mobile Computing & Multimedia, pp. 509-513.

xii. Kopetz, H., 2011. *Real-Time Systems: Design Principles for Distributed Embedded Applications*. Springer New York Dordrecht Heidelberg London: Springer Science & Business Media.

xiii. Lee, J., Shin, I. & Easwaran, A., 2012. Convex Optimization Framework for Intermediate Deadline Assignment in Soft and Hard Real-Time Distributed Systems. *International Journal of Systems and Software*, 85(10), pp. 2331-2339.

xiv. Yun, B. J., et al., 2009. Image Surveillance System using Intelligence. *The Journal of Advanced Smart Convergence*, 9(5), pp. 115-121.

ISBN:978 0 9943656 37
www.apiar.org.au