Scientific
Research

Detection of Objects in Motion—A Survey of Video Surveillance

Jamal Raiyn

Computer Science Department, Alqasemi College, Baka El Gariah, Israel
Email: raiyn@qsm.ac.il

ABSTRACT

Video surveillance system is the most important issue in homeland security field. It is used as a security system because of its ability to track and to detect a particular person. To overcome the lack of the conventional video surveillance system that is based on human perception, we introduce a novel cognitive video surveillance system (CVS) that is based on mobile agents. CVS offers important attributes such as suspect objects detection and smart camera cooperation for people tracking. According to many studies, an agent-based approach is appropriate for distributed systems, since mobile agents can transfer copies of themselves to other servers in the system.

1. Introduction

Various papers in the literature have been proposed and focused on computer vision problems in the context of multi-camera surveillance systems. The main problems highlighted in these papers are object detection and tracking, and site-wide, multi-target, multi-camera tracking. The importance of accurate detection and tracking is obvious, since the extracted tracking information can be directly used for site activity/event detection. Furthermore, tracking data is needed as a first step toward controlling a set of security cameras to acquire high-quality imageries, and toward, for example, building biometric signatures of the tracked targets automatically. The security camera is controlled to track and capture one target at a time, with the next target chosen as the nearest one to the current target. These heuristics-based algorithms provide a simple and tractable way of computing. Conventional video surveillance systems have many limitations to their capabilities. In one case, conventional video surveillance systems have difficulty in tracking a great number of people located at different positions at the same time and tracking those people automatically. In another case, the number of possible targeted people is limited by the extent of users' involvement in manually switching the view from one video camera to another. With cognitive video surveillance system, mobile agent technologies are more effective and efficient than conventional video surveillance

systems, assuming that a large number of servers with video camera are installed. If one mobile agent can track one person, then multiple mobile agents can track numerous people at the same time, and the server balances the load process of the operating mobile agent on each server with a camera.

We consider the scenario that the smart camera captures two similar objects (e.g. twin), then each object selects a different path. The tracking process will be confusing. Furthermore, the smart camera is limited to cover a certain zone in public place (Indoor). Next section introduces many solutions that have been suggested to the above problem. The suggested solutions to improve the conventional video surveillance system are extended in various ways.

A part of the approaches is to use an active camera to track a person automatically, and thus the security camera moves in a synchronized motion along with the projected movement of the targeted person. These approaches are capable of locating and tracking a small number of people. Another common approach is to position the camera at strategic surveillance locations. This is not possible in some situations due to the number of cameras that would be necessary for full coverage, and in such cases, this approach is not feasible due to limited resources. A third approach is to identify and track numerous targeted people at the same time involving image processing and installation of video cameras at any designated location,

since the image processing increases server load.

The limitation of human perception system in conventional video surveillance system increases the demand to develop cognitive surveillance system. Many of the proposed video surveillance systems are expensive and lack the capability of cognitive monitoring system such as no image analysis. This makes the system lack the ability to send warning signal autonomously in real-time and before the incidents happen. Furthermore, it is difficult and might take a long time for people to locate the suspects in the video after the incidents happen. The problem may get more complete on the larger scale surveillance system. The next generation video surveillance system expected not only to solve the issues of detection and tracking but also to solve the issue of human body analysis. In the literature, it can be found many references in development of sophisticated video surveillance system. In this paper, we introduce the cognitive video surveillance system (CVS). CVS aims to offer meaningful characteristics like automation, autonomy, and real-time surveillance such as face recognition, suspect objects, target detection, and use of cooperative smart cameras. Many face recognition systems have a video sequence as the input. Those systems may require being capable of not only detecting but tracking faces. Face tracking is essentially a motion estimation problem. Face tracking can be performed using many different methods, e.g., head tracking, feature tracking, image-based tracking, and model-based tracking. These are different ways to classify these algorithms.

2. Review of Human Body Analysis

This section introduces various approaches that considered the object detection and object tracking in video surveillance field [1-3]. The analysis of human body movements can be applied in a variety of application domains, such as video surveillance, video retrieval, human-computer interaction systems, and medical diagnoses. In some cases, the results of such analysis can be used to identify people acting suspiciously and other unusual events directly from videos. Many approaches have been proposed for video-based human movement analysis [4-6].

In [7] Oliver *et al.* developed a visual surveillance system that models and recognizes human behavior using hidden Markov models (HMMs) and a trajectory feature. In [8-10] proposed a probabilistic posture classification scheme to identify several types of movement, such as walking, running, squatting, or sitting. In [11] traced the negative minimum curvatures along body contours to segment body parts and then identified body postures using a modified Iterative Closest Point (ICP) algorithm. In addition [12,13] used different morphological operations to extract skeletal features from postures and then identified movements using a HMM framework. Another

approach used to analyze human behavior is the Gaussian probabilistic model. In [14] has been described the real-time finder system for detecting and tracking humans. In [15] proposed a shape-based approach for classification of objects is used following background subtraction based on frame differencing. The goal is to detect the humans for threat assessment.

In [16] presented a method to detect and track a human body in a video. First, background subtraction is performed to detect the foreground object, which involves temporal differencing of the consecutive frames. In [17] presented a novel approach to detect the pedestrians, which is shown to work well in a indoor environment. They make use of a new sensing device, which gives depth information along with image information simultaneously. In [18] proposed method that deals with the direct detection of humans from static images as well as video using a classifier trained on human shape and motion features. The training dataset consists of images and videos of human and non-human examples. In [19] has been suggested to use the mobile agent for multi-node wireless video cooperation in order to reduce redundancy which will result repeated information collection in overlapping regions. In [20] introduced automatic human tracking system based on a video surveillance system enhanced with mobile agent technologies. In [21-23] has been proposed a composite approach for human detection, which uses skin color and motion information to first find the candidate foreground objects for human detection, and then uses a more sophisticated technique to classify the objects. Other approaches extract human postures or body parts (such as the head, hands, torso, or feet) to analyze human behavior.

Motion Detection

This section aims to provide the status of art of the different techniques of motion detection estimation. Various studies have been introduced on the subject and the literature is very plentiful in this province. We are trying to list some methods used methods. The idea is to give an overview of the most commonly used methods and approaches. The most used algorithms for moving objects detection are based on background subtraction. The background subtraction is based on comparing of the current video frame (foreground objects) with one from the previous frames that is called sometimes background.

3. Video Surveillance System

In this section we introduce the system model of the video surveillance system. Video surveillance system has been used for monitoring, real-time image capturing, processing, and surveillance information analyzing.

The infrastructure of the system model is divided in

three main layers: mobile agents that are used to track suspect objects, cognitive video surveillance manage-ment (CVS), and Protocol for communication as shown in **Figure 1**. Each end device, smart camera, covers a certain zone or cell. Smart camera used for collecting parameters of human face.

3.1. Communication Protocol

In the system model has been introduced two communi-cation protocols. The first protocol used for agent-to-agent protocol. Agents used this protocol for communi-cation. The protocol is based on messages exchange as shown in **Figure 2**. The goal of the protocol is to update the agents. The second protocol is used for communica-tion between CVS and mobile agent.

3.2. Mobile Agent Features

Mobile agents are placed in smart camera stations. Mo-bile agent aims to track the suspect object from smart camera station to others. Mobile agent offers various characteristics, e.g. negotiation, making decision, roam-ing, and cloning.

3.3. Cognitive Video Surveillance Management

Cognitive video surveillance (CVS) managed mobile agent handoff in wireless networks. CVS provide the mobile agent with information. Based on received infor-mation mobile agents make decision when and where to move to next smart camera station.

3.4. Tracking Moving Objects

In order to track moving objects, we introduce two strate-gies. The first strategy is based on messaging protocol (msg_protocol). The goal of this msg_protocol is to in-
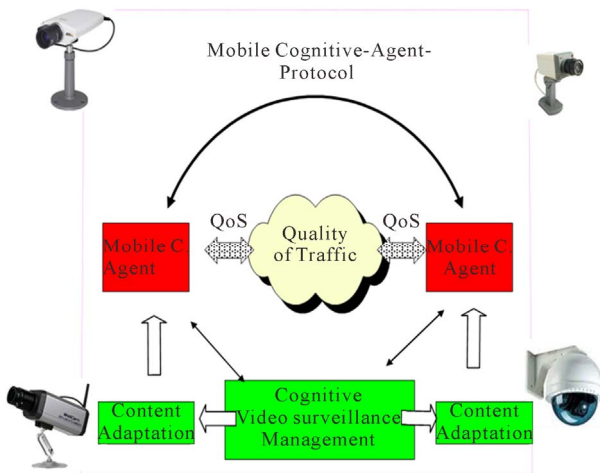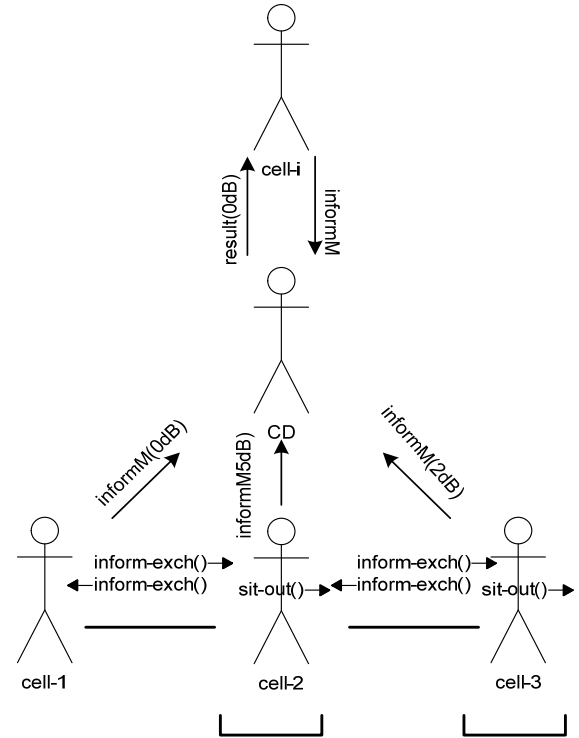

Figure 1. System model.

Figure 2. Agent protocol.

form the mobile agent about the position of the suspect object. The second strategy uses the protocol to help the mobile agent to roaming from point to others.

4. Methodology

Cognitive video surveillance (CVS) uses a data base of images. Pixels are described by a set of binary sequences. Each sequence presents certain properties (color). The database is divided into two separate sets of pixels—the training set and the test set. In both sets there are both pixels, which belong to a certain family of colors (attrib-utes) and sequence, which do not belong.

$$TP = X = \left\{ X_1, X_2, \cdots, X_n \right\}$$
$$TN = Y = \left\{ Y_1, Y_2, \cdots, Y_n \right\}$$

Each image is then divided into frames, a frame being a subset of pixel from the sequence. The number of pixel in each frame is a variable and is dynamically set to ob-tain optimal results.

$$X_1 = \left\{ x_1^1, x_2^1, \cdots, x_n^1 \right\}$$
$$X_2 = \left\{ x_1^2, x_2^2, \cdots, x_n^2 \right\}$$
$$\cdots$$
$$X_n^m = \left\{ x_1^m, x_2^m, \cdots, x_n^m \right\}$$

If for example a certain frame is comprised of 200 segments, the frames might consist of pixels 1 to 10, 2 to

11, 3 to 12, etc. Statistical methods are then applied to find correlation between a certain properties of the frame.

The basic logic of statistical differentiation of pixel is known and widely used in many prediction systems.

$$J = X \oplus Y$$

$$J = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{otherwise} \end{cases}$$

A large number of correlating factors is defined by CVS and grouped in sets. A number is linked with each correlating factor. Each factor is then turned into a single number which represents the strength of the correlation factors for each frame with respect to the probability that this frame belongs to the certain family or not. As a result we have a large number of frames, for each pair of a frame we have a number which is correlated to the probability that this frame belongs to a certain attribute (color similarity) or does not belong.

$$J = \left\{ J_1^1, J_2^2, J_3^3, J_4^4 \cdots \right\}$$

Optimization of J:

$$J_{\text{Prediction}} \left(J_1^* \right) = J_{\text{demand}} + k * \left(\Delta J \right)$$

In addition to the statistical method an innovative method of logical XOR multiplication of matrices is applied to enrich the number of frames, which are potentially contributing to the prediction model.

CVS can be implemented in a dynamic environment – when the training databases are modified the prediction mechanism is modified as well with improved prediction capabilities.

5. Smoothing EMA

In this section we introduce detection model that is based on moving average scheme. There are three types of moving average, that is, simple moving average (SMA), weight moving average (WMA), and exponential moving average (EMA). In this study, an exponential moving average is considered. An exponential moving average uses a weighting or a smoothing factor which decreases exponentially. The weighting for each older data point decreases exponentially, giving much more importance to recent observations while not discarding the older observations entirely. The detection phase focused on the collected data analysis. To increase the accuracy of the forecast model, the abnormal events in the collected data should be considered. The forecast scheme is based on the exponential moving average. The robustness and accuracy of the exponential smoothing forecast is high and impressive. The accuracy of the exponential smoothing technique depends on the weight smoothed factor alpha value of the current demand. To determine the optimal alpha factor value, fitting curve has been considered.

6. Performance Analysis

We have used the object oriented programming language C # to present the image in binary system as shown in **Figure 3**. Hence Binary vectors are implemented in WEKA platform. WEKA is stand for Waikato Environment for Knowledge Analysis. WEKA implements many machine learning and data mining algorithms. As shown
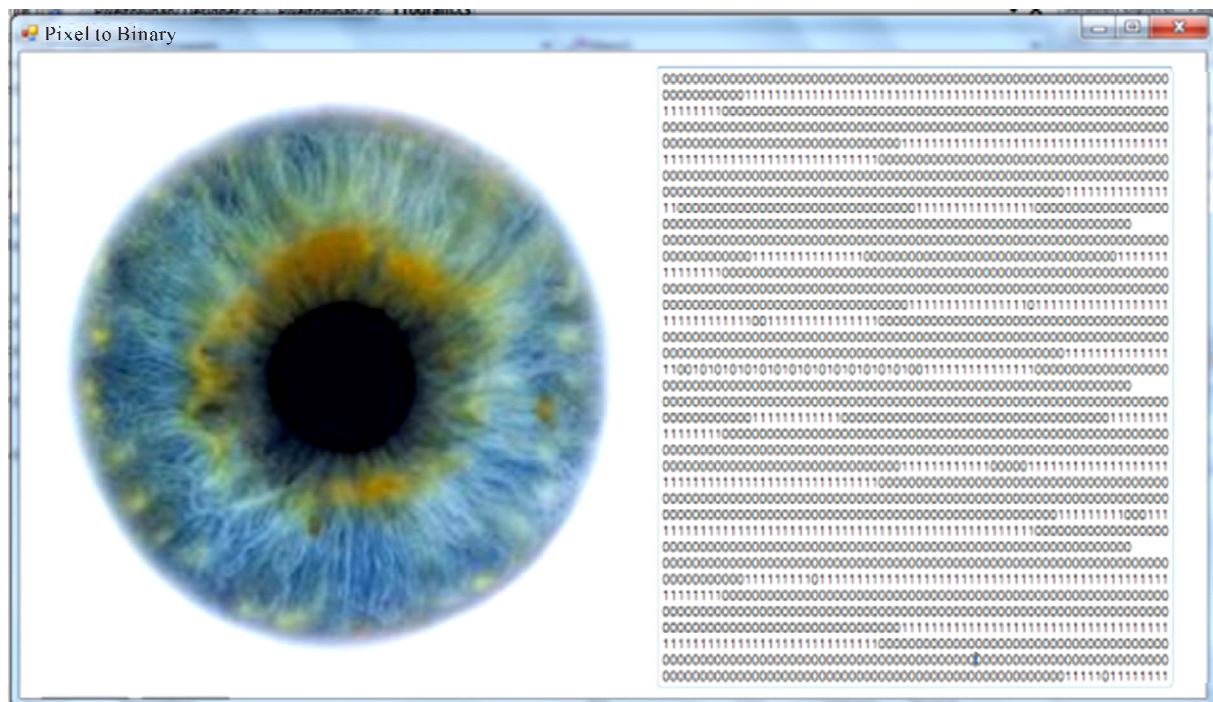

Figure 3. Image representation in binary system.

AIT

in **Figures 4(a)** and **(b)** the image analysis in visual form is based on color classification. WEKA considers the color of the image. The colors are represented in binary system. WEKA clusters the binary vectors. Each cluster represents certain attributes. As shown in **Figure 5** the comparison between simple moving average (SMA), weight moving average (WMA) and exponential moving average (EMA) is based on mean average error (MAE). Furthermore we have compared the actual observations to EMA model as shown in **Figure 6**. Results indicate that all three moving average methods have more or less similar performance in forecasting short-term times. However, as one would expect the method using optimized weights produced slightly better forecasts at a higher computational cost. Quality of forecast is diminished as the time for which forecasts are made is farther in the future. Moving average methods overestimate travel speeds in slow-downs and underestimate them when the congestion is clearing up and speeds are increasing.

7. Conclusion

In this paper, we discussed several methods in the recent literature for human detection from video. We have organized them according to techniques which use back-
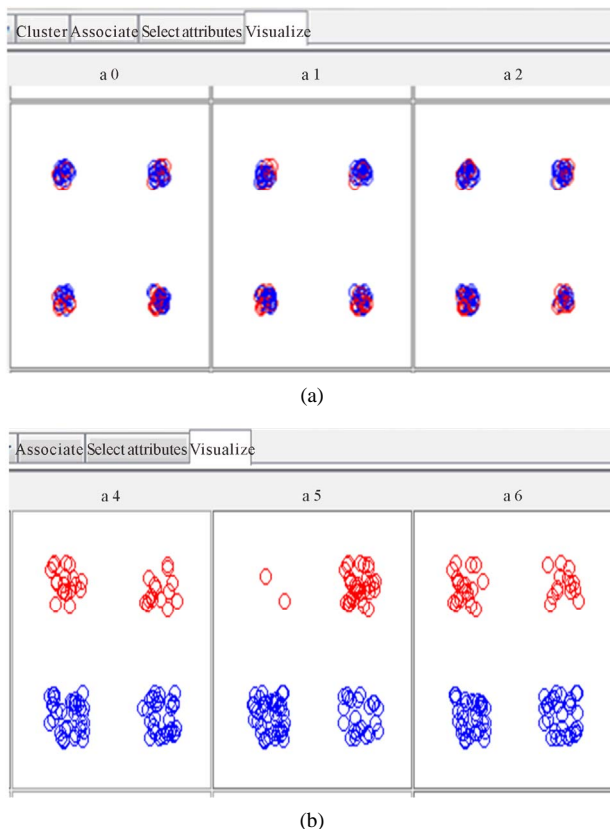

(a)

(b)

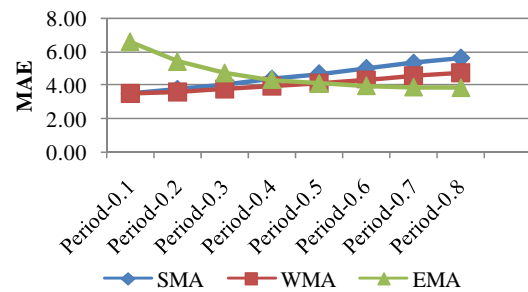**Figure 4. (a) Image analysis; (b) Color classification.**

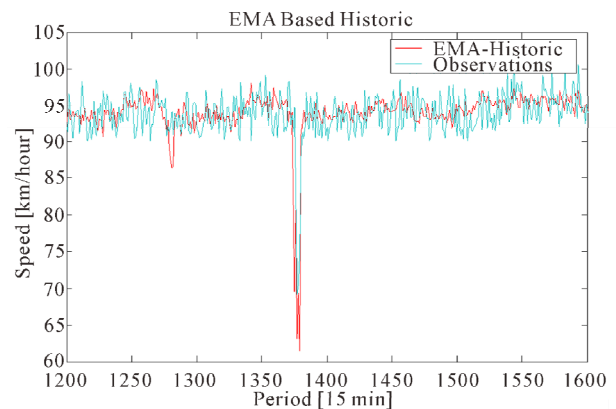**Figure 5. Comparison between MA schemes.**

Figure 6. Actual observation vs. forecasting model.

ground subtraction and which operate directly on the input. In the first category, we have ordered the techniques based on the type of background subtraction used and the model used to represent a human. In the second category, we have ordered the techniques based on the human model and classifier model used. Overall, there seems to be an increasing trend in the recent literature towards robust methods which operate directly on the image rather than those which require background subtraction as a first step. The EMA model can be used for human behaviors prediction.

REFERENCES

[1] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt and L. Wixson, "A System for Video Surveillance and Monitoring," Robotics Institute, Carnegie Mellon University, Pittsburgh, 2000.

[2] I. Haritaoglu, D. Harwood and L. S. Davis, "W[4]: Real-Time Surveillance of People and Their Activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, 2000, pp. 809-830. http://dx.doi.org/10.1109/34.868683

[3] S. Kwak and H. Byun, "Detection of Deominant Flow and Abnormal Events in Surveillance Video," *Optical Engineering*, Vol. 50, No. 2, 2011. pp. 1-8.

[4]   Z. Xu and H. R. Wu, "Smart Video Surveillance System," *Proceedings of the IEEE International Conference on Industrial Technology*, 14-17 March, pp. 285-290.

[5]   S. Aramvith, *et al*., "Video Processing and Analysis for Surveillance Applications," *International Symposium on Intelligent Signal Processing and Communication Systems* (*ISPACS* 2009), 7-9 January 2009, Kanazawa, pp. 607-610.

[6]   P. Bottoni, "A Dynamic Environment for Surveillance," *Proceedings of the* 12*th IFIP TC* 13 *International Conference on Human-Computer Interaction*, Uppsala, 24-28 August, 2009, pp. 892-895.

[7]   N. M. Oliver, B. Rosario and A. P. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, 2000, pp. 831-843. http://dx.doi.org/10.1109/34.868684

[8]   D. Weinland, R. Ronfard and E. Boyer, "A Survey of Vision-Based Methods for Action Representation, Segmentation and Recognition," *Computer Vision and Image Understanding*, Vol. 115, No. 2, 2011. pp. 224-241. http://dx.doi.org/10.1016/j.cviu.2010.10.002

[9]   I. Karaulova, P. Hall and A. Marshall, "A Hierarchical Model of Dynamics for Tracking People with a Single Video Camera," *Proceedings of the British Machine Vision Conference*, 2000, pp. 262-352.

[10] Y. Ren, *et al*., "Detection and Tracking of Multiple Target Based on Video Processing," 2009 *Second International Conference on Intelligent Computation Technology and Automation*, Changsha, 10-11 October 2009, pp. 586-589.

[11] M. B. Augustin, S. Juliet and S. Palanikumar, "Motion and Feature Based Person Tracking in Surveillance Videos," *Proceedings of ICETECT* 2011, Tamil Nadu, 23-24 March 2011, pp. 605-609.

[12] T. J. Broida and R. Chellappa, "Estimation of Object Motion Parameters from Noisy Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 1, 1986, pp. 90-99. http://dx.doi.org/10.1109/TPAMI.1986.4767755

[13] Y. Su, *et al*., "Surveillance Video Sequence Segmentation Based on Moving Object Detection," 2009 *Second International Workshop on Computer Science and Engineering*, Qingdao, 28-30 October 2009, pp. 534-537.

[14] C. Wren, A. Azarbayejani, T. Darrell and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 1997, pp. 780-785. http://dx.doi.org/10.1109/34.598236

[15] M. Ahmad and S.-W. Lee, "HMM-Based Human Action Recognition Using Multi View Image Sequences," *International Conference on Pattern Recognition*, Vol. 1, 2006, pp. 263-266.

[16] Y. Kuno, T. Watanabe, Y. Shimosakoda and S. Nakagawa, "Automated Detection of Human for Visual Surveillance System," *Proceedings of the* 13*th International Conference on Pattern Recognition*, Vienna, 25-29 August 1996, pp. 865-869. http://dx.doi.org/10.1109/ICPR.1996.547291

[17] H. Gou, *et al*., "Implementation and Analysis of Moving Objects Detection in Video Surveillance," *Proceedings of the* 2010 *IEEE International Conference on Information and Automation*, Harbin, 20-23 June 2010, pp. 154-158.

[18] S. Wang *et al*., "A Mobile Agent Based Multi-Node Wireless Video Collaborative Monitoring System," *The* 3*rd International Conference on Advanced Computer Theory and Engineering*, Chengdu, 20-22 August 2010, pp. 35-39.

[19] H. Kakiuch, *et al*., "Detection Methods Improving Reliability of Automatic Human Tracking System," 2010 4*th International Conference on Emerging Security Information*, *Systems and Technologies*, Washington DC, 2010, pp. 240-246.

[20] W. Y. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, Vol. 35, No. 4, 2003, pp. 399-458. http://dx.doi.org/10.1145/954339.954342

[21] T. S. Ling, L. K. Meng, L. M. Kuan, Z. Kadim and A. A. Baha Al-Deen, "Colour Based Object Tracking in Surveillance Application," *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*, Hong Kong, 18-20 March 2009, pp. 459-464.

[22] B. Schiele, "Model-Free Tracking of Cars and People Based on Color Regions," *Image and Vision Computing*, Vol. 24, No. 11, 2006, pp. 1172-1178. http://dx.doi.org/10.1016/j.imavis.2005.06.003

[23] Z. Zhang, "Head Detection for Video Surveillance Based on Categorical Hair and Skin Colour Models," *The* 16*th IEEE International Conference on Image Processing*, Cairo, 7-10 November 2009, pp.1137-1140.