Vision Based Surveillance System

Leanne Attard
Department of Communications and Computer Engineering
University of Malta
Msida, Malta, MSD 2080
Email: leanneattard@yahoo.com

Reuben A. Farrugia
Department of Communications and Computer Engineering
University of Malta
Msida, Malta, MSD 2080
Email: reuben.farrugia@um.edu.mt

***Abstract*—Due to the numerous amounts of surveillance cameras available, security guards seem to be ubiquitously watching over. However, the number of existing cameras exceeds the number of humans to monitor them and the supervision of all the sensors' output is costly. Thus, video footage from cameras is most often only used as a forensic tool. This suggests the need of an intelligent video surveillance system providing continuous 24–hour monitoring, replacing the traditional ineffective systems.**

This paper presents an automated vision based surveillance system which is capable to detect and track humans and vehicles from a video footage. Simulation results have shown that the Object Classification module manages to achieve an accuracy of 97.31% and 97.14% for the person and vehicle classification respectively. Furthermore, the system manages to successfully track the objects 97% of the time under no occlusion and 94.14% in presence of occlusion.

I. Introduction

The cost of purchase, installation and maintenance of surveillance cameras is getting cheaper, rendering them to be ubiquitous. On the other hand, the manpower for their supervision is expensive. Most of the existing cameras installed in banks, stores and parking lots are usually monitored sparingly if not merely used as a forensic tool.

Surveillance cameras are a far more useful tool if instead of passively recording footages, they detect events requiring attention as they happen, exploiting their benefit as an active and real-time medium. Continuous monitoring of surveillance video captured in real-time at selective indoor and outdoor public places is required for crime and accident prevention. This can be achieved through an intelligent surveillance system which obtains a description of what is happening in an area from a camera and then takes appropriate action based on the video footage, such as alerting security officers and law enforcement organizations, as salient events happen for appropriate coordination and response.

This paper presents, an autonomous video surveillance system which is capable to detect and track both human and vehicle activity within a scene. The proposed method is composed of three modules: (1) *Motion Detection* is used to detect the objects of interest (moving objects within a scene) and post-processing in order to extract interesting regions, (2) *Object Classification* is used to recognize humans and vehicles within the scene and (3) *Object Tracking* is used to track the persons and vehicles to store their trajectories. The novelty of this paper resides in the integration of advanced object recognition and tracking procedures. Simulation results show that the system is able to achieve a classification accuracy of 97.31% and 97.14% for human and vehicle classification respectively. Furthermore, the Object Tracking system achieves an average accuracy of 97% under no–occlusion and 94.14% in presence of occlusion.

The remainder of this paper is organized as follows. Section II presents the related work found in literature. Sections III, IV and V present the modules adopted by the proposed system in some depth. Section VI presents the simulation results while section VII draws the concluding remarks.

II. Related Work

A new generation of video surveillance systems is emerging transforming traditional Closed-Circuit Television (CCTV) cameras into an active intelligent security system, by employing advanced behavior recognition software. The detection of moving objects and their classification allows for further tracking, behavior analysis and event recognition. Until now, intelligent systems with a specific functionality, usually funded by governments, have been employed only in particular places. A survey on intelligent video surveillance systems is presented in [1].

In ADVISOR [2], new algorithms were developed for motion detection, people tracking and behavior recognition. Advanced traffic management systems (ATMS), which interpret the camera data to evaluate and re-program traffic signal lights in real-time were introduced in some countries to improve traffic flow. The latter include work part of the California PATH program [3] of the University of California, generating detailed traffic scene analysis and making inferences about traffic events such as vehicle lane changes and stalls.

Existing video surveillance systems reported in literature explore various aspects in automatic surveillance. Tracking of human body parts as well as humans in groups, pose recognition of people and detection of objects carried by individuals were part of the W4 system proposed in [4]. In [5] a frame–rate Lehigh Omni-directional Tracking System (LOTS) was proposed for the tracking of targets in a perimeter security setting. A real-time system for tracking people and interpreting their behavior was proposed in [6]. Piciarelli et al. [7] suggested a system for anomalous event detection based on trajectories.

III. Motion Detection

The *Motion Detection* module is made up of three sub–modules, as illustrated in Fig. 1. The first sub–module identifies the regions of interest involving moving objects, and therefore suppresses the background through the use of temporal data. Different background subtraction methods were considered in this work. These include Frame Differencing, Approximate Median [8] and Running Average. However, the background subtraction technique which we propose involves the subtraction of an initial static background from the current frame. A predefined threshold was used to check if the change in pixel value is due to motion or not using

$$f_g(t) = \begin{cases} 1, & \text{if } D(t) > T \\ 0, & \text{Otherwise} \end{cases} \quad (1)$$

where

$$D(t) = \mid f(t) - f_b(t) \mid \quad (2)$$

where $f(t)$ is the current frame, $f_b(t)$ represents the background model and $f_g(t)$ represents the foreground image. This background subtraction strategy was used in this work because it is simple and obtains satisfactory results.

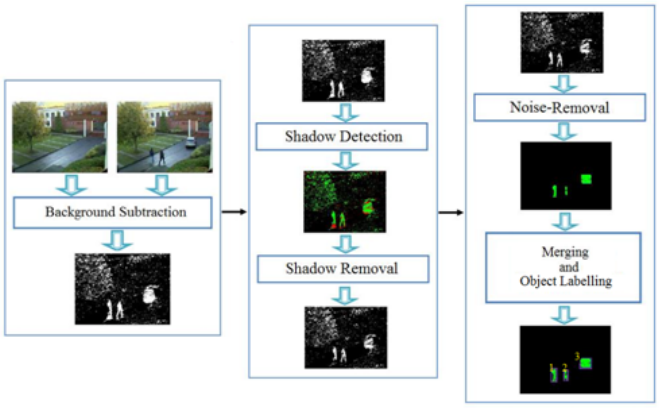

Fig. 1. Motion Detection Process.

As shown in Fig. 1, the previous sub-module leaves a lot of noise which must be suppressed. One source of noise is shadows. This method incorporates a shadow detection and removal strategy using the normalized RGB color space model. The algorithm exploits the fact that when a shadow is cast upon an object, the object intensity values of all three color channels are modified by the same value δ, such that $R_s = R + \delta$, $G_s = G + \delta$ and $B_s = B + \delta$. Therefore, the normalized components, before and after shadowing are assumed to be approximately equal according to

$$\frac{R_s}{R_s + G_s + B_s} \approx \frac{R}{R + G + B} \quad (3)$$

where only the R channel is considered since all three channels are expected to behave exactly the same.

The shadow detection technique also involves a threshold for distinguishing intensity change due to shadow as opposed to that due to motion. The discrepancy between the R and R_S is computed using

$$\left| \frac{R_s}{R_s + G_s + B_s} - \frac{R}{R + G + B} \right| < T_s \quad (4)$$

where if the difference between the normalized components is below a threshold T_s, then the change in intensity is assumed to be due to shadowing. Thus this must not be detected as part of the moving object. Fig. 2 depicts the shadow detection removal process.

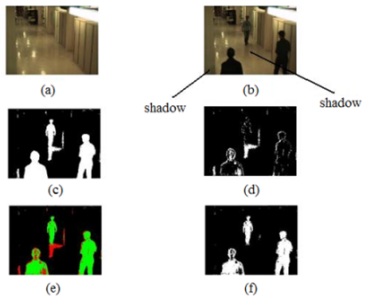

Fig. 2. Shadow detection and removal process (a) Background model, (b) Current Frame (c) Output from the *Motion Detection* module, (d) Shadow Detection, (e) foreground is shown in green while shadow is depicted in red and (f) Resulting frame with suppressed shadows.

Additional residual noise is generally present due to insignificant moving objects in the scene such as swaying tree branches in the background. This noise was suppressed using the dilation and erosion morphological operators. The individual pixels were merged and labeled using an 8–connected component algorithm, which objects will be used by latter algorithms to recognize and track the moving object.

IV. Object Classification

The *Object Classification* module adopted by the proposed system employs the distribution of the gradient orientation of an object in an image through the use of the Histogram of Oriented Gradients (HOG) [9]. Fig. 3 summarizes the feature extraction process. It involves a weighted voting of a pixel in a histogram according to the orientation of its gradient. The image undergoes an intensity transformation to map the darker intensity range onto a larger range such that the darker regions are lightened and the edges caused by shadow are reduced. Gradient computation was made using both Sobel and Simple 1-D ($[-1, 0, 1]$) masks, with the former performing better in the person classifier and the latter more appropriate for the car classifier, as seen in Table I and II. The image is then dissected in cells and the orientation values of the pixels are accumulated in a local histogram through a vote. Overlapping blocks of cells are then contrast normalized to provide for better illumination change.

This feature extraction method was applied to known datasets for Persons [10], [11] and Vehicles [12], [13] to train two separate Support Vector Machines (SVM) using LIBSVM [14]. Both Linear and Radial Basis Function (RBF) were used as kernels.

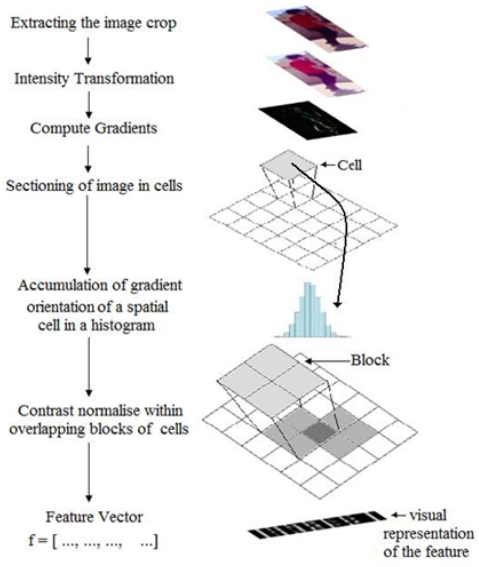

Fig. 3. Histogram of Orientation Gradient Extraction Process.

V. Object Tracking

The main purpose in a vision-based surveillance system is to detect the objects in the scene and follow them, acquiring information on their location and appearance as they move. This system achieved this through a correspondence matching procedure of the objects from one frame to the next using the HSV color histogram and distance cues.

The HSV color space is dissected into twenty-one sections as suggested in [15]. The cross–sectional H–S hexagon is partitioned into seven cells as shown in Fig. 4 (a). The *Value* V component is divided into three partitions $(0, 0.45)$, $(0.45, 0.75)$ and $(0.75, 1)$.

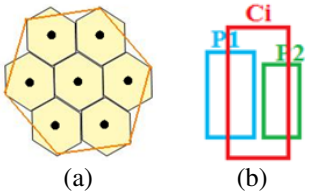

Fig. 4. (a) HSV Model (b) Grouping of occluded objects

Every pixel is quantized to one of the 21–bins and the resulting color histogram is generated. Each previous object in the video sequence has its histogram compared to the histogram of current objects which lie close to the previous object. The insertion distance was computed using

$$D = 1 - \sum_{i=1}^{n} min(h_{1i}, h_{2i}) \tag{5}$$

where h_{1i} and h_{2i} represent the color histogram of the previous and current objects respectively. If $D < T_c$ (T_c is a predefined threshold), the object are considered the same and thus the trajectory file is updated accordingly.

The objects in the scene generally interact amongst themselves and may get across each other such that one is occluding the other from the camera's field of view. If two or more objects occlude each other, the current object may match to a number of previous objects. In this case, an occlusion group C_i is formed as showing in Fig. 4 (b) and its color histogram is computed. The group is tracked until the respective objects split into individual objects. When splitting is detected, the individual objects are compared to all the previous objects pertaining the occlusion group until a match is found.

VI. Simulation Results

The *Motion Detection*, *Object Classification* and *Object Tracking* modules were combined to form the autonomous vision based surveillance system. The *Motion Detection* module adopts the Static Background Subtraction method described in section III. The object classification adopts the SVM classifier to recognize humans and vehicles within the scene. Two different kernel functions were considered in this work, linear and radial basis function (RBF). The RBF kernel optimizes two distinct parameters (C and σ).

The cross–validation technique implemented in LIBSVM was used for parameter selection. The goal of this technique is to identify good parameters so that unknown data can be accurately predicted, thus providing a more generalized classification. It involves, segmenting the training set into v subsets of equal size. Each subset is then sequentially tested using the classifier trained on the remaining $v-1$ subsets. The cross-validation accuracy representing the percentage of data which was correctly classified is calculated. A grid–search is then applied such that the pair (C and σ) which maximizes the cross–validation accuracy are selected. The optimal parameter pairs were (2, 0.0078125) and (8, 0.0078125) for the person and the vehicle classifier respectively.

Tables I and II summarize the classification rates of the human and vehicle classifier respectively. These results show that the RBF kernel function achieves better results for the person classifier whereas the Linear function resulted in higher accuracy, recall and precision for the vehicle classifier. It can be further seen that the *Object Classification* module achieves maximum accuracy of 97.31% and 97.14% for the person and vehicle classifiers respectively and an average precision and recall rates of around 95%.

In order to evaluate the performance of the *Object Tracking* module, the following performance measures were adopted

- *Track Detection Rate (TDR)* – ratio of true positive for an object to total ground truth points for the object
- *Occlusion Success Rate (OSR)* – ratio of number of successful dynamic occlusions to the total number of occlusions

Eight different video sequences were considered to test the *Object Tracking* process, where both indoor and outdoor environments were considered. These experiments show an average TDR of 96.99% for non occluded objects and 94.14% for occluded sequences. Furthermore, the system achieves an

TABLE I
RESULTS FOR PERSON CLASSIFIER (VARYING THE KERNEL FUNCTION)

HOG Parameters		SVM Configuration			
Mask	Voting Gradient Magnitude (M) Unity (1)	Kernel Type	Accuracy (%)	Recall (%)	Precision (%)
1-D centred	M	RBF	96.04	94.92	97.09
1-D centred	M	Linear	95.23	95.00	95.44
1-D centred	1	RBF	94.73	92.23	97.09
1-D centred	1	Linear	91.04	87.00	94.64
(3x3) sobel	M	RBF	95.96	95.92	96.00
(3x3) sobel	M	Linear	94.46	95.69	93.39
(3x3) sobel	1	RBF	94.15	94.31	94.02
(3x3) sobel	1	Linear	94.04	95.15	93.08

TABLE II
RESULTS FOR VEHICLE CLASSIFIER (VARYING THE KERNEL FUNCTION)

HOG Parameters		SVM Configuration			
Mask	Voting Gradient Magnitude (M) Unity (1)	Kernel Type	Accuracy (%)	Recall (%)	Precision (%)
1-D centred	M	RBF	95.60	95.71	95.49
1-D centred	M	Linear	95.95	95.71	96.17
1-D centred	1	RBF	95.60	94.76	96.37
1-D centred	1	Linear	96.31	95.24	97.32
(3x3) sobel	M	RBF	93.33	94.52	92.33
(3x3) sobel	M	Linear	94.64	94.29	94.96
(3x3) sobel	1	RBF	92.86	90.24	95.23
(3x3) sobel	1	Linear	94.05	92.62	95.34

OSR of 88.89%. A typical example of tracking under occlusion is illustrated in Fig. 5. Fig. 6 illustrated the performance of the proposed algorithm when dealing with vehicles.

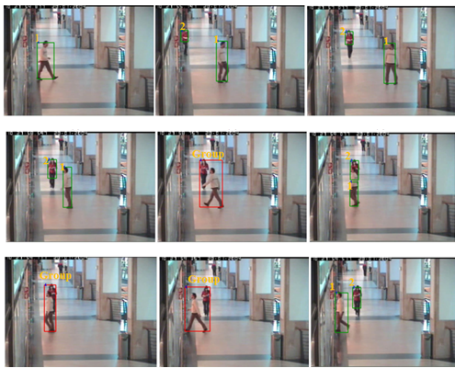

Fig. 5. Human Tracking (2–person occlusion)

Fig. 6. Vehicle Tracking (2–vehicle occlusion)

VII. COMMENTS AND CONCLUSIONS

This paper has presented the implementation of a Vision based Surveillance System which is capable to detect and track both humans and vehicles at a high level of accuracy. The developed system adopts the Static Background Subtraction method to detect foreground pixels which are filtered and grouped using standard morphological operations. The active regions are then classified as person, vehicle or other using the HOG features and SVM classifier which were appropriately trained. Simulation results have shown that the developed system manages to detect 97.31% of the persons and 97.14% of the vehicles within a scene. The system then adopts an object tracking method based on the color segmentation cue. The tracking method manages to achieve a TDR of 94.14% and an OSR of 88.89% for normal surveillance sequences. Therefore, the proposed system can be used even when more than one object is occluded.

These results clearly show that the proposed surveillance system can achieve high detection and tracking of both human and vehicles. The algorithms adopted by the system were successfully tested in both indoor and outdoor environments. Future work involves the real–time implementation of this system proposed in this paper.

REFERENCES

[1] L. P. Roberts. (2005) The history of video surveillance – from vcrs to eyes in the sky. [Online]. Available: http://www.video-surveillance-guide.com/history-of-video-surveillance.htm

[2] N. T. Siebel and S. J. Maybank, "The advisor visual surveillance system," in *in ECCV 2004 workshop Applications of Computer Vision (ACV)*, 2004.

[3] J. Malik, S. Russell, J. Weber, T. Huang, and D. Koller, "A machine vision based surveillance system for california roads," Computer Science Division, University of California, Tech. Rep., 1994.

[4] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 809–830, 2000.

[5] T. E. B. R. Micheals, X. Gao, P. Lewis, C. Power, W. Yin, and A. Erkan, "Frame-rate omnidirectional surveillance & tracking of camouflaged and occluded targets," in *VS '99: Proceedings of the Second IEEE Workshop on Visual Surveillance*. Washington, DC, USA: IEEE Computer Society, 1999, p. 48.

[6] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 780–785, 1997.

[7] C. Piciarelli, C. Micheloni, and G. Foresti, "Trajectory-based anomalous event detection," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 11, pp. 1544 –1554, nov. 2008.

[8] N. J. B. McFarlane and C. P. Schofield, "Segmentation and tracking of piglets in images," *Machine Vision and Applications*, vol. 8, no. 3, pp. 187–193, May 1995.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, 2005.

[10] (2010) Mit pedestrian data. [Online]. Available: http://cbcl.mit.edu/software-datasets/PedestrianData.html

[11] (2010) Inria pedestrian data. [Online]. Available: http://pascal.inrialpes.fr/data/human/

[12] (2010) Caltech vehicle data. [Online]. Available: http://www.vision.caltech.edu/html-files/archive.html

[13] (2010) Pascal vehicle data. [Online]. Available: http://pascallin.ecs.soton.ac.uk/challenges/VOC/databases.html

[14] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[15] K.-M. Chen and S.-Y. Chen, "Color texture segmentation using feature distributions," *Pattern Recognition Letters*, vol. 23, no. 7, pp. 755–771, 2002.