Journal of Communications Technology, Electronics and Computer Science, Issue 9, 2016
ISSN 2457-905X

Pedestrian Crossings Detection by using Driving Assistance Systems

Al Smadi Takialddin
Department of Communications and Electronics Engineering,
College of Engineering, Jerash University, 311, Jerash-Jordan'
Jerash-Jordan
e-mail takialdiina@yahoo.com

***Abstract*— This paper mainly studies Driving Assistance Systems and Detection Pedestrian Crossings of traffic and control, many years around the world and company studies have been conducted on intelligent transport systems (ITS). Intelligent vehicle, (IV) the system is part of a system which is designed to assist drivers in the perception of any dangerous situations before, to avoid accidents after sensing and understanding the environment around it. Methodology: we made an analysis of the peculiarities of the task of surveillance for pedestrian crossings and presented a detection system which these features into account. The system consists of a detector based on histograms of oriented gradients, and activity detector. The proposed Results tested detection precision and performance of the proposed system. The motion of the work is to combine the proposed system and the tracker. The results show that an adequate application of the quality and performance of the developed algorithm of detection of objects of interest in the work.**

Keywords*— *Object Detection, Activity Detector, ITS, Computer Vision.

I. Introduction

The protection of critical transportation assets and infrastructure is an important topic these days. Transportation assets such as bridges, overpasses, dams and tunnels are vulnerable to attacks. In addition, many of these facilities exist in areas of high pedestrian traffic, making them accessible to attack, while making the monitoring of the facilities difficult. Testing the solution on real Intersections would be so difficult, especially when the designer needs to continually make changes to his design and use the output data for analysis and other purposes. Another way of modelling is required that gives the designer the flexibility to test his designs and obtain results much easier and faster. "Smart video" system to understand the behaviour of human beings is a very important one. One of the most popular trends in the development of these systems is traffic surveillance allowing solving such tasks as traffic rules control; traffic congestion monitoring, stolen vehicle search [1].

Detection Pedestrian Crossings Use Driving Assistance Systems, of the system which is designed to assist drivers in the perception of any dangerous situations before, to avoid accidents and understanding the environment around it [2]. To date, there have been numerous studies into the recognition.

II. Traffic Tracking Systems

First, The Compulsory Component of the Traffic Tracking systems is robust computer vision algorithms that are able to track the motion of vehicles and pedestrians on video.

Traffic accidents have become one of the most serious problems. The reason is that most accidents happen due to the negligence of the driver. Rash and negligent driving could push other drivers and passengers in danger on the roads. More and more accidents can be avoided if such dangerous driving condition is detected early and warned other drivers. Most of the roads, cameras and speed sensors are used for monitoring and identifying drivers who exceeded the permissible speed limit on roads and motorways, this simplistic approach and there are no restrictions.

- Intensity resolution is about 8 bits/pixel for each channel (RGB) and Most computer vision applications work with monochrome images
- Temporal resolution is about 40 ms (25 Hz), SNR is about 50 Db (Pulnix camera spec.)
- One camera gives a raw data rate of about 450 MBytes/s (color) 150 MBytes/s (mono)

The presented work is concerned with the task of pedestrian crossings surveillance (Fig. 1a and 1b) to control compliance of the Traffic Rules (TR) at them. The purpose of the work is to develop a task adapted algorithms for detection and tracking of road users. The main difficulties in the development of such systems are high requirements to operation speed and stability of algorithms in continuous shooting. The tasks assigned to video surveillance systems are not new [2], [3]. At the moment, there are dozens of objects detection, recognition and tracking algorithms. However, in practice, even advanced algorithms of computer vision do not give consistently good results in real objects surveillance [4], [5]. Such cases, at least, require algorithms to be adjusted for specific shooting conditions (camera calibration for the 3D layout of the scene, setting a specific background, illumination, hardware properties). Thus, you cannot just use any computer vision algorithms; you must take into account peculiarities of the problem and combine existing approaches to solve it. This paper presents an analysis of the peculiarities of the video surveillance task for pedestrian crossings and proposes a pedestrian detection system optimized for this work.

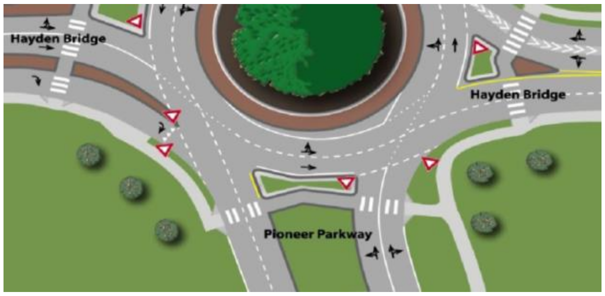


Fig. 1. Example of a frame captured by the surveillance camera at pedestrian crossing.

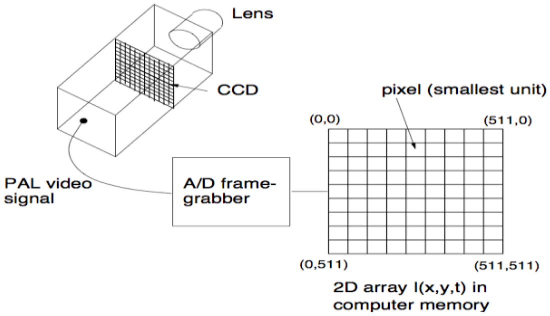


Fig. 2. A typical digital SLR CCD measures about 24×16 mm and contains about 6×106 sampling elements (pixels).

III. Materials and Methods

Detection and tracking of objects of a particular class are associated with a number of difficulties, such as a interclass variability of objects of interest, objects occlusions, noise, etc. When performing the surveillance task for pedestrian crossings we should be mindful of the fact that certain difficulties occur to a lesser extent due to the following features:

- Stability of pedestrians position: Pedestrians cross the roadway in vertical position
- Size of the objects of interest varies within a small range
- Pedestrians and vehicles do not lean when moving
- Relative low velocity of the objects
- New objects appear in certain places of the frame
- Trajectory of pedestrians and vehicles motion fall under a number of simple patterns and Stationary camera
- Stable background excluding changes in weather conditions and time of day

In addition, to control the traffic rules compliance at pedestrian crossings we do not need to know the exact number of pedestrians at the crossing and separate them from each other. What is important is the mere fact of pedestrian's presence the task complexity lies in the following problems solving:

- Frequent occlusion of objects
- Camera noise and changes caused by weather conditions
- Complicated background
- Possible unevenness of illumination (if, for example, one part of the crossing is in shadow while the other is well lit)

Practical implementation imposes a number of requirements:

- Operation in real time mode
- Minimum of settings at each new point of surveillance
- A small number of false positives and missed violations

Let us consider the typical aspects that define the work of the system of detection and tracking of objects of interest (targets 1):

- A detector that detects objects of interest based on particular points, gradients, structural models and other approaches
- Signs of the tracking object colour, shape, singular points, gradients and others
- A model of representation of the object attributes (pattern, histogram, point/rectangle/ellipse/ "skeleton"/deformable model)
- A mechanism of target localization in subsequent frames (neighbourhood scanning, Kalman filter, particle filter, nuclear localization
- A match algorithm of different objects traces in different frames (method of nearest neighbour, minimum matching, two- and multi-frame algorithms)
- A control unit that maintains a list of tracked objects and recognizes the behaviour of objects in their trajectories

This article provides a detailed description of learning and work of the target detector and, also the activity detector which reduces the load on the computationally complex detector of objects. If we add a tracker to the system the activity detector becomes a link between the tracker and the detector revealing new objects for tracking and restoring tracking of objects lost by the tracker (Fig. 2).

This work we used a detector of objects based on histograms of oriented gradients [5] implemented in [10] which is one of the best detectors currently available. Its work includes sequential scanning of test images by sliding window method, extracting the histogram of oriented gradients and comparison with the reference histogram. The histogram of oriented gradients is information about the type of direction gradients typically found in a particular place of the target object (Fig. 3a and 3b). Histograms are built for the positive and negative learning examples; the classifier learns according to these examples, for instance, SVM (Fig. 3c). Usually objects' detectors are launched using the image pyramid, but due to low variability of size of detected objects in this task we can restrict the launch of the detector only by one fixed scale.

The data used for learning and [6] only every fourth frame is marked, which gives a total of about 900 positive examples for learning. This set has been extended to 2200 examples by marking the intermediate frames using linear interpolation of the trajectory. Also, examples of sedentary people and examples of people visible from the front or the rear were removed from the learning set, as these examples are not important for the work; they increase the interclass variability of detected objects rather than interfere with the work of classifier. Quality measurement results (Table 1) show that detection recall has increased by 15% (description of the quality metrics-in section 4).

TABLE I. COMPARISON OF THE EFFECTIVENESS OF THE LEARNING SETS

Learning set of examples Precision	Recall	F-score
Original set (900 positive examples)	1.0000	0.7007
Extended set (2200 positive examples)	0.9954	0.8571
Extended set with hard negative examples	0.9973	0.7551

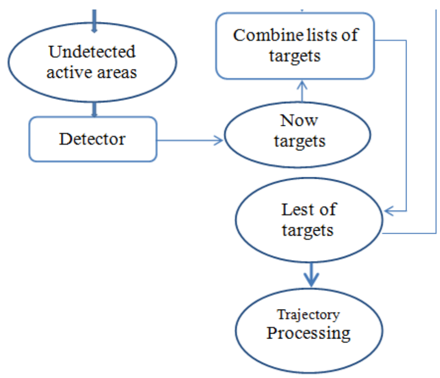
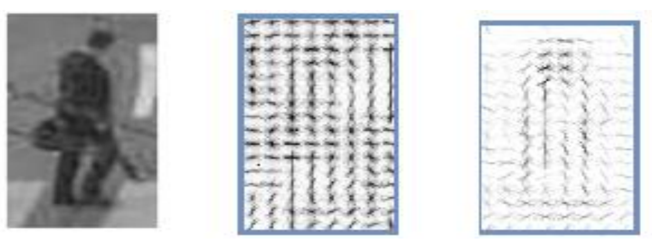
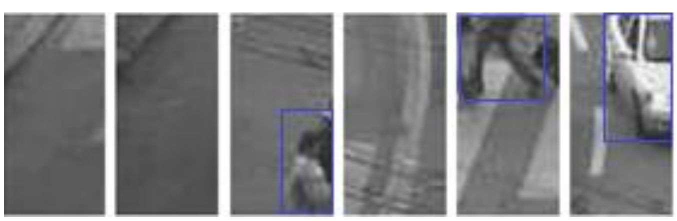


Fig. 3. Tracking pedestrians and capturing pedestrian images.



a) *b)* *c)*

Fig. 4. Positive learning example (a) its histogram of gradients (b). Learning classifier (c).



a) b) c) d) e) f)

Fig. 5. (a) An active pixels mask after morphological operations, rectangles forming around only large blobs; (b) rectangular areas that are processed by the detector; (c) detected pedestrians

We conducted an experiment to reduce the number of false positives using a mining hard negatives technique [11] it is well known that the number of mined negative examples for learning is significantly greater than that of positive ones. Therefore, to reduce the number of false positives of the detector, it is desirable to choose hard for the classifier negative examples, similar to pedestrians (Fig. 4). Thus, the classifier builds a more precise "pedestrian-not pedestrian" border. As we can see from Table 1, the use of this technology has slightly increased the precision, but significantly reduced the recall. As a result, we selected an extended set of hard negative examples.

Detection algorithms are computationally very expensive, so the displacement pitch of the sliding window was increased to 8 pixels, which gave acceleration of more than 10 times with a small loss of quality. The area of detection was also reduced to the area of crossing without roadway intended for vehicles only. This gave further twofold acceleration. Furthermore, to accelerate the detection of objects we used an activity detector as described in the following subsections.

IV. TESTING

The activity detector is designed to highlight areas of the frame where certain changes are found as compared to the background. It allows running computationally complex detector only in these areas, as well as working of the system in the "background" mode while the crossing is empty. From a technical point of view, the background mode also allows the calculation unit to save energy [7].

The designed activity detector uses a background model based on median values of pixels in the last frame (Fig. 5a):

$$B(x,y,t) = median_{i=1_k}\left\{I(x,y,t-i)\right\} \ldots\ldots\ldots\ldots\ldots\ldots Q1$$

where, $I(x,y,t)$-t-th frame, $B(x,y,t)$-background model for t-th frame.

Each frame (Fig. 5b) is compared with the obtained background model and the pixels, wherein the difference exceeds a predetermined threshold are considered active (Fig. 5c), i.e., belonging to the objects, not to the background:

$$B(x,y,t) = \left|\left|I(x,y,t) - B(x,y,t)\right| \geq H\right\| \ldots\ldots\ldots\ldots\ldots\ldots Q2$$

here, a (x, y, t) the active pixel mask for the t-the frame, H finalization threshold.

As shown in Fig. 5c, the results of this approach are far from ideal:

- Some parts of the objects slightly differ from the background, so the shape of the selected connected regions (blobs) differs from the true shapes of objects and several areas can correspond to one object
- Shadow distorts the object's shape, because they are marked as active regions
- Camera noise and slightly changing background objects provoke appearance of small false areas

Morphological operations of erosion and dilation [8] by disc structural elements apply to mask active pixels to remove noise and combine parts of objects into united binary blobs:

$$Q(x,y,t) = \max\left[e(x,+x',y+y',t \middle| x'^2 + y'^{2 \le R^2 dilate} \right] ....Q4$$

Q mask for t-the frame.

It should be noted that on the basis of the experiments the established erosion radius ($R_{er\text{-}rode}$ = 5) is considerably smaller than the radius of following dilatation (R_{dilate} = 15). The blobs the size of which exceeds the predetermined threshold (marked with white rectangles) are selected among obtained blobs. The resulting blobs do not reproduce the exact shape of objects, but this configuration is much more convenient for subsequent processing.

After that, a bounding rectangle is formed around each of the filtered blob, this rectangle is cut out of the frame and forwarded to the detector Due to the fact that the blobs obtained do not reproduce the shape of pedestrians perfectly, it is unknown whether the entire object is covered by blob and if not, what part of it (top, bottom, middle) is. That is why the transformation of coordinates of the rectangle limiting the blob (B_{top}, B_{bottom}, B_{left}, B_{right}) into coordinates of the rectangle directed to the detector (D_{top}, D_{bottom}, D_{left}, D_{right}), is as follows:

$$D_{to\ p} = \min\left(B_{to\ p}, B_{bottom} - H \right) - G$$
$$D_{bottom} = min\left(B_{bottom}, B\ to_p \right) + G$$
$$D_{left} = \min\left(B_{left},\ B_{right} - W \right) - G$$
$$D_{right} = \min\left(B_{right}, B_{lift} + W \right) + G ....$$
$$.................................Q5$$

The W height and the width (respectively) of the scanning window (128×64), G the width of the additional gap correcting errors of blobs outlining (G = 10). The window size corresponds to the size of pedestrians on video (Fig. 3). The size of the gap is chosen experimentally (section 4). Each pair of the resulting rectangles that are considerably overlapped is combined into one rectangle until there are no couples left. Each of the resulting rectangles is separately processed by the detector. An area in which the detector works is reduced by 20-90% (depending on the number of people on the crossing [9]. A general chart of the pedestrian detection algorithm is shown in Fig. 6.

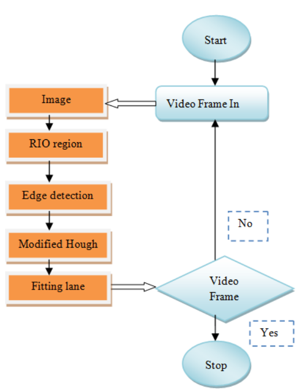


Fig. 6. Chart of pedestrian detection.

V. Detection Quality

The following indicators were selected as detection quality metrics are shown in Table 2.

TABLE II. PARAMETERS DEFINITION

Parameters	Definition
Precision (P)	The ratio of the number of correctly detected pedestrians to the total number of detections from the detector
Recall (R):	The ratio of the number of correctly detected pedestrians to the total number of registered pedestrians in thereference marking;
F-score or Van	Harmonic average of precision and recall. F-score combines theseindicators into a single value
Operation time (T)	The algorithm in seconds

As seen from Table 3, the resulting operation time exceeds the duration of the video, which indicates that this algorithm cannot be used in the real time mode and we need to use the tracker together with the detector or to substantially optimize the algorithm.

TABLE III. OPERATION TIME OF THE PROPOSED ALGORITHM PHASES

Operation	Video 1		Video 2	
	Sec	%	Sec	%
Obtaining an active pixels mask	2.469	0.089	1.223	0.042
Morphological operations	6.953	0.253	3.384	0.115
Rectangles forming	6233	0.227	6.514	0.222
Rectangles merging	0.158	0.006	0.318	0.011
Detector operation	11.637	0.424	17.951	0.611
Total	27.45	1.0	29.39	1.0

We conducted an experimental analysis of various modifications of the algorithm and the proposed configuration, which allowed us to make the following conclusions:

- Complete image processing reduces precision as the number of false positives outside crossing increases; recall varies slightly. Operation time increases twice as much due to a larger area for processing
- Scanning of the entire area of the crossing (not just active regions) slightly increases the recall that indicates that the activity detection misses some significant areas. In the case of a sparse crowd of pedestrians the activity detector reduces the time, but in the case of a dense crowd additional operations on separate parts allocation slow the algorithm
- Reduction of the pitch of the sliding window can greatly increase the recall; however, it is achieved through an

extremely large increase in the operation time of the detector

- Narrowing the gap around the rectangles marked for detector reduces the recall and increase of the gap does not change it, which indicates that the gap of 10 pixels is enough

The operation time of phases of the proposed algorithm was measured to evaluate the performance (Table 2). Most of the time (40-60 %) is spent on the work of the detector itself, especially when the pedestrian flow is dense (Video 2). Also a lot of time is given to morphological operations (Matlabfunctions immediate, imerode) and to the rectangle formation (regionprops function), but it's worth noting that these phases can be accelerated through optimization of these operations for a specific task.

VI. Conclusions

The analysis of the quality and performance of the developed algorithm of detection of objects of interest in the task of video surveillance for pedestrian crossings has shown that it is impossible to track objects in real time using only the detector due to its computational complexity. The next step is to equip the system with a tracker capable of tracking the targets found.

Acknowledgment

This Work Was Supported by the Deanship of Scientific Research, Jerash University.

References

[1] Al Smadi, T., 2014. Real-time lane detection for driver assistance system. Circuits Syst., 5: 201-207. DOI: 10.4236/cs.2014.58022

[2] Al Smadi, T., 2011. Computing simulation for traffic control over two intersections. J. Adv. Comput. Sci. Technol. Res., 1: 10-24. https://www.sign-ific-ance.co.uk/index.php/JACSTR/article/view/379

[3] Al Smadi, T., H.A. Al Issa, E. Trad and K.A. Al Smadi, 2015. Artificial intelligence for speech recognition based on neural networks. J. Signal Inform. Process., 6: 66-72. DOI: 10.4236/jsip.2015.62006

[4] Maggio, E. and A. Cavallaro, 2011. Video Tracking: Theory and Practice. 1st Edn., John Wiley and Sons, Hoboken, ISBN-10: 0470749644, pp: 292.

[5] Al Smadi, T.A. and M. Maitah, 2013. An efficiency and algorithm detection for stenography in digital symbols. Int. J. Comput. Netw. Inform. Security, 6: 34-38. DOI: 10.5815/ijcnis.2014.01.05

[6] Yeo, H.S., B.G. Lee and H. Lim, 2015. Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware. Multimedia Tools Applic., 74: 2687-2715. DOI: 10.1007/s11042-013-1501-1

[7] Liu, Y., W. Zhang and Y. Zhang, 2015. Dynamic neuro-fuzzy-based human intelligence modeling and control in GTAW. IEEE Trans. Automat. Sci. Eng., 12: 324-335. DOI: 10.1109/TASE.2013.2279157

[8] Bian, Z.P., Junhui Hou, Lap-Pui Chau and N. Magnenat-Thalmann, 2015. Fall detection based on body part tracking using a depth camera. IEEE J. Biomed. Health Informat., 19: 430-439. DOI: 10.1109/JBHI.2014.2319372

[9] Zhu, C. and W. Sheng, 2011. Motion- and location-based online human daily activity recognition. Pervasive Mobile Comput., 7: 256-269. DOI: 10.1016/j.pmcj.2010.11.004

[10] Rautaray, S.S. and A. Agrawal, 2015. Vision based hand gesture recognition for human computer interaction: A survey. Artificial Intell. Rev., 43: 1-54. DOI: 10.1007/s10462-012-9356-9

[11] Dariush, B., K. Fujimura and Y. Sakagami, 2011. Vision based human activity recognition and monitoring system for guided virtual rehabilitation. U.S. Patent Application No. 12/873,498.