

Automatic Pedestrian Recognition Using Real-Time Motion Analysis

Patrick VANNOORENBERGHE¹, Cina MOTAMED¹, Jean-Marc
BLOSSEVILLE², and Jack-Gérard POSTAIRE³

¹ Laboratoire d'Analyse des Systèmes du Littoral
195, Rue du Pasteur Martin Luther King, BP 649
62228 Calais CEDEX, France.

² Institut National de Recherche sur les Transports et leur Sécurité
2, Avenue du Général Malleret-Joinville,
94114 Arcueil CEDEX, France.

³ Centre d'Automatique de Lille, Equipe Image et Décision
Université des Sciences et Technologies de Lille,
59655 Villeneuve d'Ascq CEDEX, France.

Abstract. This paper presents a real-time vision system for automatic pedestrian recognition using image sequences analysis. We first introduce a robust motion detector based on a specific background structural model. The motion detection process exploits the structural difference between the model and features extracted from the current scene. The pedestrian recognition process analyses the signal provided by the motion detector in order to separate pedestrians from other moving objects (vehicles, buses, ...). The detection signal processing consists in the extraction of several kinematics and pattern features of each classes of objects. The percentage of good pedestrian discrimination, tested with real traffic image sequences under several illumination conditions, is about 85%.

1 Introduction

Pedestrian motion parameters are required by traffic engineers, architects and town-planners in the planning, design and operation of facilities such as shopping centres, business areas, airports, and pedestrians road-crossings. Parameters of interest include flow, density, speed, direction. Advances in video image processing technology suggest that there is a potential for the production of a system which will be suitable for automatic monitoring of pedestrian scenes. In addition, vision-based systems offer several important advantages than other detection technologies. The first one is the complete flexibility to specify a detection zone of variable shape and area. Another one is the incorporation of features which extract trajectory information from pedestrian activity. Despite these advantages, video-detectors are still not widely used in real-world applications. The majority of practical applications have been in indoor environments, often with the help of controlled or structured lighting to simplify the extraction of features and depth information from the image. Since 1982, E. Oscarsson is interested

in pedestrian detection in urban environment [Osc82]. By 1990, Deparis and Duvieubourg analysed the human behaviour on subway quays [Dep94]. Their works attach to evaluate risks of passenger falls on tracks. Recent studies have demonstrated that vision systems can provide a good solution to pedestrians behaviour [Mec94, Gla95, Rou94, Vel94, Rea95]. The major barriers to successful implementation of vision-based pedestrian detectors can be categorised as being due to illumination (several light sources), reflection, occlusion, perspective and pedestrians shapes. We propose, in this paper, an original real-time vision system to recognise pedestrians evolving on the road from other users of the city. Section 2 proposes an overview of the system which is composed of two main processes. We first introduce a motion detection algorithm based on a specific background structural model in the section 3. In order to separate pedestrians from the other objects (vehicles, buses, ...), we introduce a global measurement characterising scene objects (section 4). Finally, we present the results of this real-time pedestrian recognition process. In the section 5, images are proposed to evaluate the performance of the system.

2 System overview

We present in this section, an overview of the proposed vision system. The problems presented by urban scenes include variable illumination conditions, high levels of occlusion, variability in pedestrians shapes with time and of course the discrimination between the traffic users. We are interested in the problem of video information analysis given by a fixed camera aiming the urban scene (on traffic lights for example) and objects evolving on a plan (the road). The vision system is composed of two different processes : the motion detection process and the pedestrian recognition process. The motion detection process makes use of the digital images to provide a binary information concerning each pixel of surveillance areas. For each frame i of the sequence, outputs of the process can take two different forms :

- detection signals $D_{i,j}$ which represent the spatial occupation in the area j of the current frame i ,
- a binary image M_i which corresponds to the mask of the moving objects.

In order to provide a good adaptability of the system, surveillance areas can be easily changed simply by reprogramming the system without any physical realignment. The pedestrian recognition process is based on the analysis of the motion segmented by the above process. Shapes of the moving objects, global occupation of the scene and other kinematics and pattern features are extracted from each frame i of the sequence. This information is important for determining the behaviour of the object in the field of view of the camera. Some indicators are then computed in order to recognise pedestrians from other users of the city.

3 Motion detection process

Several methods have been proposed in order to extract motion information from a sequence of video images according to two main approaches [Nag83]. In fact, motion detection can be undertaken at two different levels: pixel-level or object-level. Object-level methods are based on two successive steps : first, the extraction of high level features (corners, angulous points, edges, ...) describing the objects, and a matching of these features which determines the displacement of each object between two successive frames of the sequence. The difficulty is to find a limited number of features which are relatively difficult to extract from a real image sequence due to the variations in orientation and size of the objects. Pixel-level methods analyse the spatio-temporal variations of gray levels in order to estimate the movement [Hor81]. However, algorithms based on this principle suffer from the sensibility to noise and are very expensive in terms of processing complexity and time. Direct approaches, based on the analysis of differences between successive frames have also been investigated [Jai84] but they need a priori information which is not available in real life situation. We describe, in this section, the principle of the proposed motion detection algorithm based on a reference model and propose some results.

3.1 Principle of the motion detection algorithm

A typical urban scene may be illuminated by several light sources like sun, vehicle headlights or street lighting which can vary rapidly in intensity (sun combined with cloud movements) and direction. For these reasons, it is very important that the motion detector is not influenced by illumination changes. We propose a motion detection process based on the comparison between a specific structural model and features extracted from the current scene. The main step of the motion detection algorithm is the construction of the structural model derived from images of the sequence [Van95]. The result of the detection is directly linked with the quality of the model. Figure 1 presents the structure of the proposed motion detection algorithm.

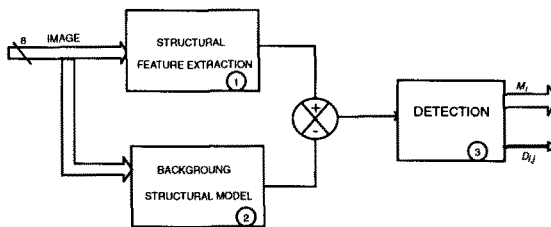


Fig. 1. Structure of the motion detection algorithm

The construction of the model (module 2) consists in researching neighbouring relations in a predefined direction. In order to establish the background of

the scene, we search the grey level **transition** between two pixels of the same frame in a pre-defined direction. The structure of the background of the scene is analysed in eight elementary directions $\Theta_n = n * \frac{\pi}{4}$ for $n = 0, 1, \dots, 7$. Let (X_0, Y_0) be the coordinates of a pixel P_0 with grey level I_0 . We say that a grey level transition T_{Θ_n} in a direction Θ_n occurs between P_0 and its nearest neighbour P_{Θ_n} of coordinates $(X_{\Theta_n}, Y_{\Theta_n})$ and grey level I_{Θ_n} if $|I_0 - I_{\Theta_n}| > T$ where T is the minimum amplitude of grey level transition which can be adjusted by the analyst. The model at the point P_0 can also take the form of the following equations (Equations (1) and (2)):

$$P_0 \xrightarrow{Model} [P_{\Theta_n}, \Delta I^M T_{\Theta_n}] \text{ for } n = 0, 1, \dots, 7 \quad (1)$$

$$\text{where } \Delta I^M T_{\Theta_n} = |I_0^M - I_{\Theta_n}^M| \quad (2)$$

Hence, for each pixel of the surveillance area, a background structural model is then generated and used for motion detection. The motion detection process operates by comparison with features extracted from the current scene for the same pixels P_{Θ_n} . For each image, all differences between the structural model and current features reveal an intrusion in the field of view of the camera. Therefore, for a frame i , we must compute all the transitions T_{Θ_n} in the n directions and obtain the following structural features (Module 1) for each pixel :

$$P_0 \xrightarrow{Frame i} [P_{\Theta_n}, \Delta I^i T_{\Theta_n}] \text{ for } n = 0, 1, \dots, 7 \quad (3)$$

$$\text{where } \Delta I^i T_{\Theta_n} = |I_0^i - I_{\Theta_n}^i| \quad (4)$$

For each pixel of the surveillance area, any difference between the structural model of the background (Equation (1)) and that of the current frame (Equation (3)) reveals a modification of the scene and yields to an intrusion detection at this location. Illumination modifications contribute to change slightly the amplitude of the gradient between two pixels corresponding to a transition in the image. The detection decision is also made by computing the indicator $I = \sum_{n=0}^7 I_n$ where :

$$I_n = \begin{cases} 1 & \text{if } [\Delta I^M T_{\Theta_n} - \Delta I^i T_{\Theta_n}] > V \\ 0 & \text{elsewhere.} \end{cases} \quad (5)$$

The value of the threshold V is adjusted in order to tolerate some dispersion of the amplitudes of the gradient without detection. The decision concerning the detection at a pixel location is taken by comparing the value to a confidence level adjusted to 4 in our case. Finally, for each frame i , the shapes of the intruding objects are convoluted with a filter for noise reduction.

3.2 Results

The figure 2 represents the result of pedestrians and vehicles detection on the crossing. Black areas represent the masks of the moving objects. We can note

that the surveillance area does not correspond to the entire image. Concerning the system performance, the processing rate is around four frames per second. The good detection rate, especially tested for the European Programme PRAXI-TELE, is greater than 99% and experiments show that the system generates less than one false alarm per day and per camera.

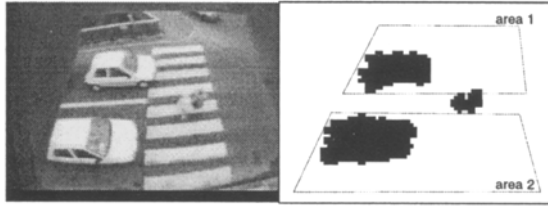


Fig. 2. Results of detection

4 Pedestrian recognition process

The objective of our work is to estimate individual pedestrian time travel for crossing the road. This estimation is made by tracking moving characteristics of the scene. Studied scene is complex because it contains several varieties of objects (cars, pedestrians, buses, ...). For our application we separate the scene into different areas. In order to separate pedestrians from other objects, we introduce a global measurement inside each area. It brings us information related to content of the current scene. This information (the nature of the object) will be used to reduce characteristics to track. In fact, the tracking process is activated when the scene contains only pedestrians. Detection signals $D_{i,j}$ (Occupation signal D_i of the area j) is computed by counting the number of pixels that have been changed at each frame i of the sequence in the area j . The detection signal $D_{i,j}$ is analysed in order to discriminate pedestrians from other moving objects of the area j . We have observed that generally cars have high speed behaviour and a high level detection indicator. On the opposite, pedestrians can be recognised with their slowness and their low surface $D_{i,j}$. It produces some ambiguities when a crowd appears in the scene or when a car moves slowly. These particular cases are treated separately. At the first level we threshold the detection signal $D_{i,j}$ for separating classic isolated pedestrians by computing a new indicator $Dt_{i,j}$. This indicator is calculated with the following expressions :

$$Dt_{i,j} = \begin{cases} D_{i,j} & \text{if } D_{i,j} \geq T_1 \text{ or if } D_{i,j} = 0 \\ 1 & \text{elsewhere.} \end{cases} \quad (6)$$

Isolated pedestrian separation Isolated pedestrians are detected when the detection signal $Dt_{i,j}$ is equal to 1. The threshold level T_1 is chosen in order to

have a detection surface less than $3m^2$. Then, for each sample of the signal $Dt_{i,j}$, a relation $Dr_{i,j}$ is computed. It characterises the link between speed and surface of moving objects in area j of the scene. $Dr_{i,j}$ takes the next expression :

$$Dr_{i,j} = \frac{\overline{Dt_{i,j}}}{Base_{i,j}} \quad (7)$$

where $\overline{Dt_{i,j}}$ is the $Dt_{i,j}$ mean level calculated on three neighbours and $Base_{i,j}$ represents the presence time like you can see on the following figure 3.

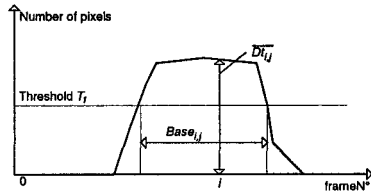


Fig. 3. Measurement of the indicator $Dr_{i,j}$

Car with high speed detection If $Dr_{i,j}$ is higher than T_2 , we consider that the area j contains moving car. The threshold T_2 takes into account minimum speed (20 km/h) and surface ($3m^2$) of moving car. In opposite, a value of $Dr_{i,j}$ lower than the threshold T_2 explains that the area contains crowd or cars with low speed.

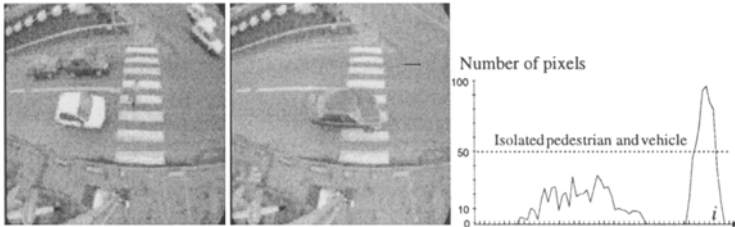


Fig. 4. Example of isolated pedestrian and vehicle detection

Discrimination between crowd and low speed cars Concerning the occupation signal $D_{i,j}$, we can observe more deviation for crowd in opposite with car which has smooth behaviour because they are rigid, so their surfaces do not change when they move. In addition, we can observe that the detection surface of crowd contains more spatial discontinuity (Fig. 5). We propose also two measurements of these variations and discontinuities :

- By counting the number of local maximums of the $Dt_{i,j}$ signal; For this approach, experimental results shows that the percentage of good recognition of crowd from car, is near 70%.
- By measuring spatial discontinuity of detection shape (M_i), experimental results shows 75% of good discrimination. The measurement is made by characterising the texture of binary detection image.

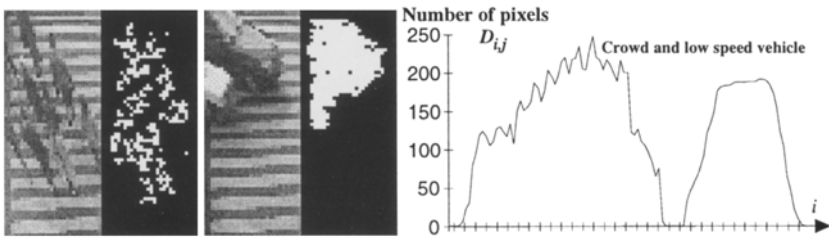


Fig. 5. Crowd and car detection image (M_i)

5 Results

The vision system described in the above sections, was tested on real image sequences under variable traffic and illumination conditions on two different sites. First site corresponds to a local outdoor test-site at I.N.R.E.T.S. (National Institute of Transportation and Security) near Paris. Images arriving directly in the laboratory, enable us to run tests on the two processes and conveniently assess effects of environmental variation (shadows, clouds, sun, rain, and night). The second site is the Place Chatelet in Paris. Tests show that the motion detection process can detect vehicles and pedestrians passing through the surveillance area nearly free of errors. The results of objects detection under variable weather conditions are shown in Table 1 (2 hours experimental sequences). Concerning

Table 1. Object detection under variable weather conditions

	Day	Night	Day+rain	Night+rain
Isolated pedestrian	95%	30%	80%	22%
High speed cars or buses	82%	92%	71%	81%
Crowd	80%	21%	48%	13%
Low speed cars	81%	95%	73%	75%

the recognition process, we can recognise efficiently four varieties of objects behaviour crossing the road (isolated pedestrian, cars with high speed, crowd and

isolated car with slow speed). Experimental results show that the critical case is when we have pedestrian with slow car simultaneously in the same area. The processing rate of the system is near four images per second.

6 Conclusion

A real-time vision system able to recognize automatically pedestrians in uncontrolled environments has been presented. The proposed system is based on the modelization of the structure of the scene. The quality of the detection is improved when the background is high textured. We have presented an automatic pedestrian recognition system based on the analysis of the detection signal. The discrimination process works on dynamic features extracted by global behaviour of the scene. An indicator is deduced from the total occupation of the surveillance area. The signal processing enables us to point out kinematics and pattern characteristics of each class of objects. The proposed algorithm has been tested on real traffic image sequences. The percentage of good pedestrian discrimination is about 85%. The processing rate is near 4 images/s. This project is supported by I.N.R.E.T.S. and the Nord/Pas de Calais Regional Council.

References

- [Nag83] Nagel H.H.: Overview on image sequence analysis. *Proc. Image Sequence Processing and Dynamic Scene Analysis*, pp 2-39, 1983.
- [Hor81] Horn B.K.P. and Schunk: Determining Optical Flow. *Artificial Intelligence*, Vol. 17, 1981.
- [Jai84] Jain R.: Difference and accumulative difference pictures in dynamic scene analysis. *Image and Vision Computer*, 1984.
- [Osc82] Oscarsson E.: TV-Camera detecting pedestrians for traffic light control. *Proceedings of IMEKO 1982*, pp 275-282, 1982.
- [Dep94] Deparis J.P. and Duvieubourg L.: Intrusion Detection by active linear cameras. *Advanced Applications in Transportation Technologies*, Capri, 1994.
- [Mec94] Mecocci A., Bartolini F. and Cappellini V.: Image sequence analysis for counting in real time people getting in and out of a bus. *Signal Processing* 35, pp 105-116, 1994.
- [Gla95] Glachet R., Bouzar S. and Lenoir F.: Estimation des flux de voyageurs dans les couloirs du métro par traitement d'images. *R.T.S. N 46*, Mars 1995.
- [Rou94] Rourke A. and Bell M.G.H.: An image-processing system for pedestrian data-collection. *Proceedings of the 7th International Conference on Road Traffic Monitoring and Control*, London, April 1994, pp 123-126.
- [Vel94] Velastin S.A., Yin J.H., Davies A.C., Vicencio-silva M.A., Allsop R.E. and Penn A.: Automated measurement of crowd density and motion using image processing. *Road Traffic Monitoring and Control*, 26-28 April 1994, Conference publication No. 391, * IEE, 1994.
- [Rea95] Reading I.A.D., Wan C.L. and Dickinson K.W.: Developments in pedestrian detection. *Traffic Engineering + Control*, pp 538-542, 1995.
- [Van95] Vannoorenbergh P., Motamed C., Blossville J.-M. and Postaire J.-G.: Detection of intrusion using a background structural model. *Computers ind. Engng.*, Vol. 29, No. 1-4, pp 675-679, 1995.