Pattern Recognition 47 (2014) 56–71

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Detecting pedestrians on a Movement Feature Space

Pablo Negri [a,b,*], Norberto Goussies [a,c], Pablo Lotito [a,d]

[a] CONICET, Av. Rivadavia 1917, Buenos Aires, Argentina
[b] Instituto de Tecnologia, UADE, Lima 717, Buenos Aires, Argentina
[c] DC-UBA, Buenos Aires, Argentina
[d] PLADEMA-UNCPBA, Campus Universitario, Tandil, Argentina

A R T I C L E I N F O

Available online 6 June 2013

Keywords:
Pedestrian detection
Movement Feature Space
Histograms of oriented level lines
Adaboost cascade
Linear SVM

A B S T R A C T

This work aims at detecting pedestrians in surveillance video sequences. A pre-processing step detects motion regions on the image using a scene background model based on level lines, which generates a Movement Feature Space, and a family of oriented histogram descriptors. A cascade of boosted classifiers generates pedestrian hypotheses using this feature space. Then, a linear Support Vector Machine validates the hypotheses that are likeliest to contain a person. The combination of the three detection phases reduces false positives, preserving the majority of pedestrians. The system tests conducted in our dataset, which contain low-resolution pedestrians, achieved a maximum performance of 25.5% miss rate with a rate of 10^{-1} false positives per image. This value is comparable to the best detection values for this kind of images. In addition, the processing time is between 2 and 6 fps on 640×480 pixel captures. This is therefore a fast and reliable pedestrian detector.

1. Introduction

Pedestrian detection, as well as character and face recognition, vehicle detection, etc., is considered a key chapter of pattern recognition in computer vision. The various applications associated with this subject, such as surveillance, safety systems, robotics, and content-based indexing, attract the interest of researchers and manufacturers, as well as the military and security agents.

This variety results in innumerable algorithms and methodologies that have been recently developed [1–3]. However, a recent and thorough analysis of the state of the art by Dollar et al. [4] concludes that, for certain applications or architectures, pedestrian detection is still an open problem, e.g. cases of low-resolution pedestrian images. Such is the case of a person's 3D motion model [5], which may have between 14 and 22 degrees of freedom, showing the countless combinations of potential positions, let alone the changes in appearance due to clothes, points of views, lighting conditions, etc. Owing to these factors, pedestrian detection through computer vision remains a constant challenge in the field of pattern recognition.

The various pedestrian detection applications can be classified by the type of architecture used in the image acquisition, namely, static cameras or views (still detection), fixed-camera video sequences, or mounted camera sequences (e.g. robots, vehicles). Even though the nature of the images used may be different, the methods are all consistent in finding the best descriptors adapted to the image dataset and a classifier making it possible to find pedestrians based on those descriptors.

In still detection, people are identified in photographic images [6–11]. In general, a person's position or size scale on the image is unknown. The system thus needs to evaluate the entire image in different scales. There are multi scale descriptors, such as the Haar-like filters [6], or the Histograms of Orientations (HOG) [8,9], which are then used by SVM-type classifiers in the first case, or by cascades of boosted classifiers in the second. It is also possible to obtain a pyramid of scales from the original image, calculating the HOG in a fixed scale, and using a linear SVM classifier [7] or a latent SVM classifier [10].

In the case of a sequence of images taken from a fixed camera position, temporal information is generally used. This information may be used creating descriptors based on the video flow [12–14], which require a high frame rate of the sequence. A background model or reference image is also used to identify the pixels that do not belong to this model, and it is inferred that they are part of moving objects. The primitives used may be the level lines [15–17], Mixture of Gaussians [18,19], or a combination of color, texture and HOG channels [20]. Identifying the motion in the scene, assuming that pedestrians are actually moving, makes it possible to reduce the search space for the system, to create new descriptors, or to calculate confidence filters or maps.

* Corresponding author at: Instituto de Tecnologia, UADE, Lima 717, Buenos Aires, Argentina. Tel.: +54 1140007307.
E-mail address: pnegri@uade.edu.ar (P. Negri).

0031-3203/$ - see front matter
http://dx.doi.org/10.1016/j.patcog.2013.05.020

Finally, the application involving on-board cameras is used in robotics for human–computer interaction, or in intelligent vehicles with pedestrian avoidance systems to minimize accidents [21–23]. In this case, the complexity lies in the fact that the camera is in motion, processing the dynamic flow in the sequence which must be solved to identify pedestrians. Gravila [21] uses a disparity map from a stereo system to get a silhouette that is subsequently evaluated on a hierarchical tree of template shapes, using the Chamfer distance. Dollar [11] prefers to perform a frame-by-frame analysis, without using the motion information, and to meet the real-time requirements using a fast cascade detector and a multi-scale calculation of gradient histograms.

The system discussed in this paper seeks to detect pedestrians in outdoor sequences captured by a fixed remote camera. To get images with an adequate resolution for detection, the sequence that is generally obtained is reduced to a frame rate of 1 or 2 fps, given the limited bandwidth available. The applications involved mainly consist of remote surveillance systems, where it is important to detect intruders in restricted or dangerous areas, such as railroads or highways. The detection of people allows a surveillance system to report an improper presence in these dangerous areas so as to carry out appropriate actions. The system should be robust to rapid and significant changes in the scene (e.g. shadows, weather conditions), and to the presence of many other moving objects (e.g. vehicles, trains, etc.), as shown in Fig. 1. These images, which make up the dataset of this paper, correspond to different captures of a camera mounted in a traffic light. Moreover, the camera view is not exactly fixed because of the movement caused by the blowing wind or traffic vibrations.

The proposed detection system consists of four stages: motion detection, hypothesis generation, hypothesis validation and final filtering, as shown in Fig. 2(a).

In the first stage, the motion detection uses a level line based approach, illustrated in Fig. 2(c), generating a Movement Feature Space (MFS). The MFS information is grouped in order to obtain a descriptor family that we call Histograms of Oriented Level Lines (HO2L). Then, they are used by a cascade of boosted classifiers to generate hypotheses of the presence of a person, see Fig. 2(d), restricting the search space to some positions within the image. The hypotheses or regions of interest (RoIs) are then validated by a linear SVM classifier using the HO2L descriptors grouped in a configuration similar to the R-HOG [7]. Validated RoIs, as shown in Fig. 2(e), are then grouped using a Mean Shift Clustering method [24,25] or the Non-Maxima Suppression algorithm. Returned bounding boxes (RoI positions) are considered the system output (see Fig. 2(f)).

The main advantages of the system proposed in this paper include an increased robustness, minimized loss of information, and computation time efficiency.

We have developed MFS based on level lines to obtain an adaptive background model, preserving the orientation of the level lines, and a measure similar to the gradient module. The MFS adapts well to slow changes in the scene while it is robust to rapid variations (e.g. illumination changes or weather conditions). In these situations, people's appearance on the MFS does not change significantly compared to normal conditions.

The calculation of level lines in a transformed HSV color space, called Texton Color Space (TCS) [26], which makes it possible to retrieve color transitions that are lost when working on gray levels, is proposed.

The orientation histograms obtained from the MFS allow for a multi-scale pedestrian detection, avoiding the construction of a dense pyramid of sub-sampled versions of the input image that is very costly in terms of calculation time. They can also be computed quickly using the integral histogram, which offers the possibility of accelerating the detection stage. The combination of generative and discriminative classifiers speeds up the search for pedestrian hypotheses on the image, as proven in [27].

In summary, the system calculates the MFS of the sequence and continues to work on this space without using the information

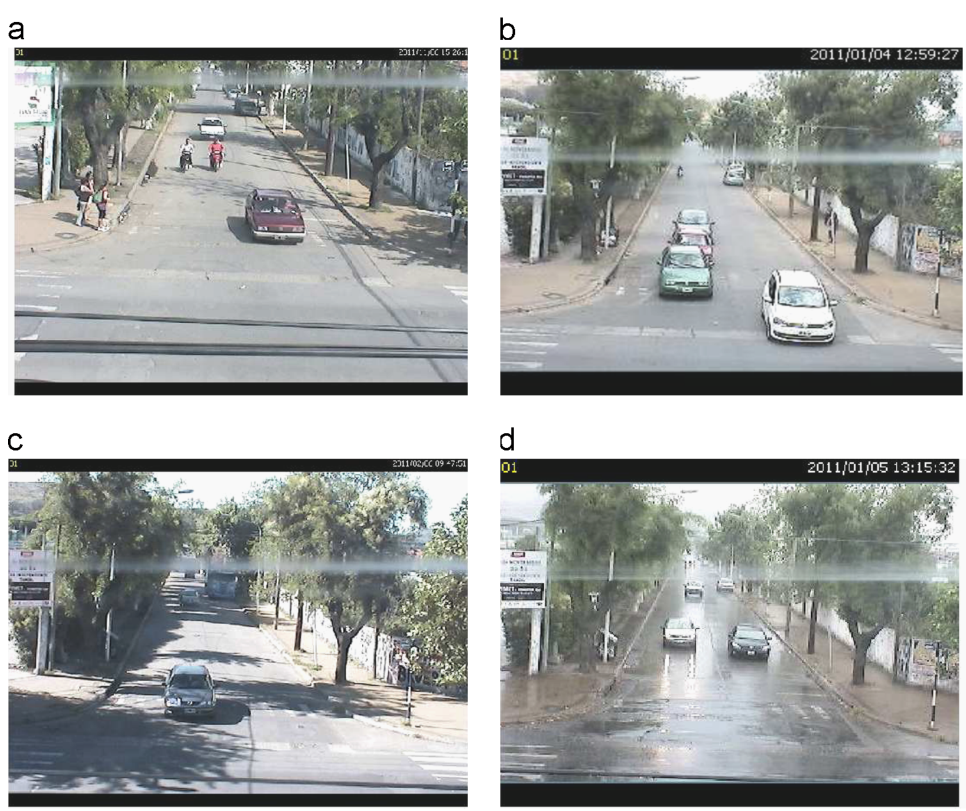

Fig. 1. Capture samples of the recorded sequences. (a) Afternoon view, (b) cloudy view, (c) lateral shadows and (d) rainy view.