Blob Motion Statistics for Pedestrian Detection

Paulo Vinicius Koerich Borges
Autonomous Systems Laboratory, ICT Centre, CSIRO
1, Technology Court, Pullenvale, QLD, 4069, Australia
Email: vini@ieee.org

Abstract—Video analysis aiming at efficient pedestrian detection is an important research area in computer vision and robotics. Although this is a well studied topic, successful detection still remains a challenge in outdoor, low resolution images. We present efficient detection metrics which consider the fact that human movement presents some characteristic patterns. Unlike many methods which perform an intra-blob analysis based on motion masks, we approach the problem without necessarily considering the pixel distribution inside the blob. Therefore, we apply periodicity analysis not to the pixel luminances inside the blob, but by analyzing the motion statistics of the tracked blob as a whole. We propose the use of three cues: (i) a cyclic behavior in the blob trajectory, (ii) an in-phase relationship between the change in blob size and position, and (iii) a correlation between blob size and vertical position, assuming that the camera is set up sufficiently high. These features are combined according to the Bayes classifier for improved performance. Experiments present numerical error rates and comparisons with other methods, illustrating the applicability of the proposed method.

I. INTRODUCTION

Pedestrian detection is a significant and often major task in robotics. On vehicles, for example, it can increase safety, using on-board cameras to avoid collisions with people. Considering static cameras mounted on infra-structure, automated pedestrian detection is often used as a part of an intelligent video surveillance system. Although pedestrian detection from static cameras has been the subject of research for many years, several issues are still challenging and can influence the detection reliability. Examples include different styles of clothing, various possible articulations, variable poses, occlusions and shadows, which all affect the segmentation and classification.

For systems in which the detection is focused on close range pedestrians, silhouette information can be reliably extracted, allowing the use of shape-based classification. However, for distant targets or targets in low quality images, shape information is no longer dependable and motion analysis has to be integrated. Still, even when motion is considered, several issues can severely affect the performance of the algorithm, such as noisy sensors and strong shadows created by the pedestrians themselves.

In this work we propose novel features which are useful to detect human motion. The periodicity analysis performed exploits the following characteristics: (i) apart from the evident periodicity observed in the pixel intensities in a tracked pedestrian, the human walking pattern also presents a subtle yet distinctive periodic behavior motion in its trajectory. For this reason, we perform a trajectory analysis to discriminate pedestrians from other objects based on the trajectory of the center of a tracked blob. (ii) There also exists a distinct “in-phase” and time dependent relationship between the height and the vertical position of a tracked pedestrian, due the shortening in height that humans present when feet are apart when walking. Hence, we extend the trajectory analysis by correlating the the 2-D trajectory with the estimated height of the person. (iii) We also consider the dependence between the blob area and vertical position of the blob. This is motivated by the fact that the height of the tracked blobs decrease as the blob moves away from the camera. This feature can be used not only for pedestrian detection, but for any ground object including vehicles. For this to be valid, the only constraint is that the camera is set up sufficiently high with respect to the ground such that any object that moves towards the top of the image (in 2-D) is consequently moving away from the camera. This relationship can be better predicted if the the homography from the ground to the camera plane is known.

To quantify the above cues, we use a spectral centroid metric to detect periodicity in the 2-D trajectory. In addition, we detect a relationship between the 2-D trajectory and the height of the blob via covariance analysis. The periodicity in the proposed features can be noisy, depending on the video quality. Therefore, we employ pre-processing operations such as rotation and filtering to increase the detection reliability. For improved performance, we combine the features according to the Bayes classifier.

Due to their simplicity, these features can complement existing methods based motion. They can also be integrated to the cases in which shape based methods exhibit high error rates. We perform several tests, using videos from public databases as well as from various surveillance cameras in an industrial site. The results show a satisfactory detection rate, especially considering the non-evident motion at large distances. The proposed classification technique is successfully tested in varying illumination conditions and in the presence of soft and hard shadows.

This paper is organized as follows. In Section II we provide a literature review, contrasting existing methods with the proposed algorithm. In Section III we propose alternatives to detect periodic motion and the corresponding metrics. In Section IV we describe data pre-processing operations for a more efficient detection and in Section V we discuss the Bayes classification used. In Section VI we show results from different experiments, followed by relevant conclusion in Section VII.
II. EXISTING METHODS

Pedestrian detection methods can be divided into two broad classes: shape-based and motion-based methods. In this section we present an overview of current pedestrian detection algorithms, with particular focus on motion-based methods.

A. Shape Based Methods

Shape-based algorithms are trained detectors that search for pedestrians by scanning the whole frame, looking for the pattern that matches well with a pre-defined model. Perhaps the best known examples in this class are the use of histogram of oriented gradients [1] and the combination of local and global cues via a probabilistic top-down segmentation [2]. The main advantage of such approaches is that they can be applied to non-static cameras, as no background segmentation is performed. Techniques extending the above mentioned approaches include using machine learning for combining multiple features (silhouette, appearance, holistic, and parts-based) as input to a support vector machine (SVM) classifier [3] [4].

In addition to the references presented here, other comprehensive surveys broadly review shape based methods [5][6].

B. Motion Analysis Methods

Motion analysis methods consider the cyclic motion in the way humans move as opposed to vehicles, bicycles, and trees, for example.

These methods use a static camera in order to generate a background model and subsequently segmentation [7]. This is followed by a tracking stage, in which the position of the object of interest is determined across frames. The human gait creates pixel or region-wise oscillations, such that the general statistical periodic behavior yields classification. Early works exploiting this characteristic performed a Discrete Fourier Transform (DFT) to quantify pixel oscillations [8] [9]. Extensions of the technique analyze the amount of change in a motion history image [10] or the power spectral similarity in the walking pattern [11]. Cutler and Davis [12] determine the gait period by computing a similarity matrix for every image pair in subsequent frames. He et al. [13] determine the angle formed by the centroid of the object of interest from the video sequence. Details on the tracking approach can be found in noise, such that it can be represented by

\[ M = \begin{bmatrix} x_1, x_2, \ldots, x_N \\ y_1, y_2, \ldots, y_N \end{bmatrix} \]  

and \( N \) is the number of frames considered in the trajectory.

Similarly, we represent the “trajectories” of \( h \) and \( w \) by \( h_i \) and \( w_i \), respectively. Figure 2 illustrates the \((x, y)\) trajectory (in green) for a tracked pedestrian. In the following we consider detection approaches for each of the features proposed.

A. Vertical Periodicity

Following periodicity assumptions similar to Ran et al. [15], assume that \( y \) can be modeled as a periodic signal embedded in noise, such that it can be represented by

\[ y_i = \mu_i + \sum_{k=1}^{M} \alpha_k \cos(\omega_k i) + n_i, \quad \omega_k \neq 0 \]  

instead of analyzing the motion statistics of the tracked blob as a whole, we apply the periodicity analysis not to the pixel luminances inside the segmented area. This novel and simple feature yields efficient human detection, as discussed in the remainder of this paper.
where $n$ represents zero-mean additive Gaussian noise. In order to detect the periodicity, we perform a Fourier-based periodogram analysis [15]. After the periodogram is generated, we calculate its center of mass - or spectral centroid [18] - defined as

$$Q_i = \frac{\sum_{m=0}^{N-1} m|Y_i(m)|}{\sum_{m=0}^{N-1} |Y_i(m)|}$$

where $m$ is frequency and $Y_i$ is the Fourier transform of $y_i$, considering the samples from 1 to $i$. We consider a pedestrian is detected if

$$T_0 < Q < T_1$$

where $T_0$ and $T_1$ are decision thresholds. These thresholds are chosen based on a trade-off between false-positive and false-negative rates. In accordance to [15], our experiments indicate that typical values are $T_0 = 1$ Hz for slow walking and $T_1 = 3$ Hz for running.

**B. Blob Height Periodicity**

We also identify that the blob height $h$ and the blob center $y$ present an “in phase” periodicity. This occurs because as humans walk, the height changes depending on the position of the legs. Our experiments indicate that this change can range from 3 to 5 percent, according to the gait and walking speed. The variation of the blob height can be subtle, and it is reduced with longer distances from the camera and lower image resolution. Although subtle, however, the underlying statistics are well defined and given sufficient samples, periodicity can be detected. This periodic change in blob height is in phase with the vertical projection of the motion of the center of the blob.

In order to quantify the dependency, we model the problem as a simple communication problem in which we want to detected a known frequency signal embedded in noise. Let $h$ represent the periodic signal corresponding to the blob height and let $y$ represent the signal corresponding to the vertical position in the image. In order to detect the “presence” of the periodicity observed in $h$ in the signal $y$, we employ the linear correlation, defined as

$$C(y, h) = \frac{1}{N} \sum_{i=0}^{N} y_i h_i$$

where $h_i$ corresponds to the blob height at frame $i$. This metric yields an efficient performance in the application considered. In addition, it is the theoretical optimal linear filter for maximizing the signal to noise ratio in the presence of additive stochastic Gaussian noise. [19]

**C. Relationship Between Blob Area and Position**

As a semantic cue, we consider the correlation between the blob height and vertical position of the blob. This can be a powerful feature when blobs that are not pedestrians present some sort of periodic motion. A typical example is that of tree branches in windy conditions. The use of this feature is motivated by the fact that the height of a tracked blob decreases as the blob moves away from the camera. Note that this feature is useful in eliminating false-positives, but not false-negatives, as cars or other ground objects show the same characteristic. For this to be valid, the only constraint is that the camera is set up sufficiently high with respect to the ground such that any object that moves towards the top of the image (in 2-D) is consequently moving far away from the camera plane. In practice, this means that the camera is mounted at least as high as the height of a human subject. Favorably, this applies to most cameras in surveillance systems (see Figure 5 and most PETS databases [20]). As a metric, we use the correlation $R$ between $y$ and $(h \ast w)$, where $\ast$ represents element-wise multiplication. Assuming that the origin of the coordinates is on the lower-left corner of the image, $y$ increases as $h \ast w$ decreases, yielding a negative correlation between them. Therefore, any object is labeled as non-pedestrian if $R > 0$.

The dependency is generally non-linear. When the ground can be assumed to be approximately planar, this dependency can be given by the homography transformation between the ground and camera planes.
IV. PRE-PROCESSING

In this section we describe some pre-processing operations, which can reduce the effect of noise in the detection.

In the first step, we filter the data with a low-pass filter \( f_L \) whose cut-off frequency is based on an estimated of the noise \( n \) and on the fundamental frequency of the sinusoidal components in \( y \). After filtering, the signal variability with respect to the vertical axis by rotating the data. This increases the signal-to-noise ratio in \( y \). This operation is motivated by the fact that periodic motion is predominantly vertical, due the the nature of walking (up and down movement) and human posture. Therefore, when finding the periodicity of the rotated path, the underlying motion is more efficiently extracted.

In order to rotate the data, we find the best linear fit to the collected data. The data is rotated by the angle of this fit \( \theta \) such that
\[
[x' \ y']^T = R [x \ y]^T
\]

where
\[
R = \begin{bmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{bmatrix}
\]

This angle is determined via standard linear regression [21] with respect to the horizontal axis. A rotation example is given in Figure 3. Figure 3a shows a scaled version of the path in Figure 2. The rotated version of Figure 3a is shown in Figure 3b, illustrating that the motion with respect to the vertical axis is increased.

![Fig. 3: Illustration of how the rotation increases the signal-to-noise ratio with respect to the vertical axis, by comparing the original (a) and the rotated (b) versions of the pedestrian path from Figure 2.](image)

V. CLASSIFICATION

In this work we the Bayes classifier to combine the features discussed in Section III, although different statistical classifiers could also be tested. A more detailed description of how to combine the features using this technique can be found in [18]. Despite some features having better classification power than others, because all the discussed features are useful to discriminate pedestrians from non-pedestrians, combining them increases the distance between the two classes, and consequently reduces the detection error rate [18], at the expense of increasing computational complexity.

Using this classification framework, the naive thresholds discussed in Section III are not employed, but give the reader a practical reference for the typical values for these parameters.

In a complete commercial system, the features used in the algorithms discussed in Section II can be easily added to the classification framework.

VI. EXPERIMENTS

In this section we show results of experiments in two different industrial environments.

A. Practical Parameters and Considerations

1) Video Details: For the tests, we record footage from several different cameras located across two industrial sites. The resolution of the images is 768 x 576. For one of the sites, the camera locations are indicated in Figure 4, represented by the red dots. Screenshot of some of the views are shown in Figure 5. Due to network oscillations, the average frame rate varies slightly from video to video, ranging from 10 to 12 frames per second. This type of variation is common for networked cameras, which are affected by network traffic, so it makes for a more challenging yet real test scenario. In addition to our local video samples, we also test the algorithm with videos from the PETS database [20], with some of the views shown in Figure 6.

2) Overview of the Tracking Algorithm: The tracking algorithm used assumes a relatively static background for the foreground/background segmentation, which is performed with the technique proposed by Li et al. [22]. This is followed by a blob entrance detection module to threshold moving objects...
B. Results

Similarly to most algorithm based on periodicity analysis, the results depend on the number of frames \( N \) considered. Increasing \( N \) improves the reliability (as more data is used for the decision) at the expense of increased detection delay. In the experiments presented, \( N = 50 \), corresponding to approximately 5 s of video. This yields a good compromise between reliability and delay.

The results in Table I show the detection rates for the algorithm using the three different features proposed: (i) periodogram of \( P_y \), (ii) correlation between vertical blob size \( h \) and vertical position \( y \), \( C \) (iii) correlation \( R \) between change in blob size and blob position. This table also shows the error rates when all the features are combined according to the Bayes classifier, as discussed in Section V. The plots in Figure 7 correspond to the receiver operating characteristics (ROC) curve, which relate the false-positive rate with the true-positive rate for different thresholds. Table II shows the most typical types of errors and their corresponding rates. For overlapping pedestrians, as with any classification algorithm based on motion, the performance depends on the ability of video tracker to overcome occlusions.

C. Comparison to Other Motion Analysis Methods

Compared to other motion analysis methods for pedestrian detection, these rates are on average similar to those reported in [15] and [25], although the rates are heavily dependent on the test set. An identical and fair comparison is hard, as video instances and conditions (frame rate, video quality and pedestrian movement and walking direction) vary. Still, we extract results from [15] (Figure 16 in that article) as accurately as possible and plot them in Figure 7, as indicated by the squares marks. The reported results correspond to \( N = 64 \) and the PETS database is also used in the tests. The results reported in [25] present an extremely low false-positive rate (in the order of \( 10^{-5} \)) with a detection rate of 90%.

As evidenced by the numerical results, the features proposed cannot be considered as better, but complementary metrics. Integrating these methods (in addition to other static image detection techniques such as HOG [1]) with the features proposed here can further reduce the error rates, at the expense of increased computational complexity.

VII. Conclusions

We have proposed a novel pedestrian detection algorithm based on motion-based features, efficient in large areas where a walking pattern is observed. The method analyzes the characteristic gait motion in human walking to discriminate between pedestrians and other tracked objects. In contrast to
several techniques which perform an intra-blob analysis of leg movement, the proposed method analyzes the motion of a tracked blob without considering the motion inside the blob, observing the motion statistics of the full tracked blob. The three features used were: (i) a cyclic behavior in the blob trajectory, (ii) an in-phase relationship between the change in blob size and position, and (iii) a correlation between blob size and vertical position, assuming that the camera is sufficiently high with respect to the ground. We present several experiments where these three features are combined according to the Bayes classifier.

The results show that, particularly in large open areas, the proposed algorithm is efficient, successfully detecting pedestrian in 85% of the cases. In videos in which the pedestrians are very close to the camera and reliable shape information can be extracted, however, the proposed features are often outperformed by traditional algorithms.

Given the method’s simplicity and low computational complexity, it can be easily combined with other techniques. Future work includes the further study of how knowledge of the homography between the ground and the camera plane can be used to increase the reliability and the expected behavior of the suggested metrics.

REFERENCES