Rapid Pedestrian Detection Based On Movement Trend

Ruohua Li[1, a], Taihong Wang[1, b]

[1]College of Electrical and Information Engineering, Hunan University, Changsha 410082, China;

[2]Key laboratory for Micro-nano Optoelectronic Devices of Ministry of Education, Hunan University, Changsha 410082, China.

[a]ruohua_li@163.com, [b]th_wang@hotmail.com

Keywords: Pedestrian detection; Kalman filter; Prediction; Verification

Abstract. This paper presents a movement trends based approach for pedestrian detection aiming at reducing the consumption of feature calculation caused by sliding windows. A new approach to predict the location of pedestrian is proposed by combining the movement trend of objects, extracted by improved background segmentation algorithm, with Kalman filter. The keypoint descriptor BRISK (Binary Robust Invariant Scalable Keypoints) is presented to verify the predicted location and make it reliable. Experiment results on PETS dataset report that the algorithm is 10.9 times faster than SVM+HOG method and keep a better accuracy at the same time.

Introduction

Pedestrian detection is widely used in surveillance, safety and robotics field. To get better detection quality, research are focusing on better features, additional data, and context information [1]. Sliding-windows is the one of the most popular method in pedestrian field. Through the enumeration approach many algorithms get high accuracy. HOG+SVM [2] is one of the most representative method. It is obvious that the sliding-windows leading to huge computation of feature. Thus to use it in real time calculation seems to be hard. Paper [3] use stixel model to accelerate the process, paper [4] use subwindows to reduce the selected windows.

In this paper, we propose a novel pedestrian detection method to avoiding using sliding windows in video sequences. We use motion trend of pedestrian to accelerate the HOG+SVM. Kalman filter [5] is also been adapted and improved in pedestrian tracking. BRISK [6] descriptor get great performance in our research by verifying the selected regions. Altogether we get speed-ups by 10.9, without suffering a loss in detection quality.

Implementation Details

The whole process of implementation is mainly divide into three parts. Firstly we use HOG+SVM detector to initial the static model. Secondly, the segmentation is executed to update the background model and get the motion trends for Kalman filter to track those models. Thirdly, we use BRISK descriptors to verify the result. Expect for the HOG+SVM is proposed by Piotr Dollar [2] trained by INRIA dataset without change, all the detail is below.

Segmentation of motion trends. In the surveillance systems, objects are generally moving continuously. In order to separate background and moving objects, a background model $B_g(\mathrm{t})$ is used as follows

$$B_g(x,y,t) = \begin{cases} I(x,y,t) & , \quad t = 1 \\ (1-\delta)B_g(x,y,t) + \delta I(x,y,t), & \quad t > 1 \end{cases} \tag{1}$$

Where $I(t)$ is the t-th image of video sequence. δ is a parameter which is used for adjusting up-date rate of background. Foreground moving regions $R(t)$ can be written as

$$R(t) = I(t) - B_g(t-1) \tag{2}$$

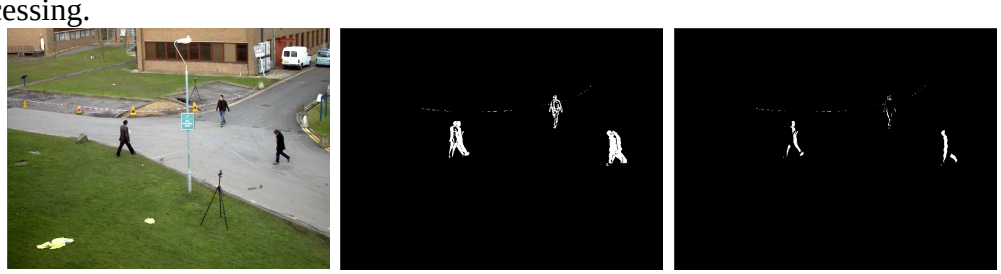
Image binary operation as Eq. 3 is used for avoiding the noise caused by the color differences of movement region.

$$D(x,y,t)=\begin{cases}0, & R(x,y,t)<Th\\ 1, & R(x,y,t)\geq Th\end{cases} \tag{3}$$

Where Th is the threshold to distinguish background and motions. To predict a moving object location, we not only need the position where it exist in the current frame, but also the tendency of its movement. The traditional difference method cannot extract the moving trend of an object, thus we improved it as Eq. 4

$$T(t)=D(t)-D(t)\cap D(t-1) \tag{4}$$

Where $T(t)$ represent the tendency of objects in the t-th frame. Fig. 1 shows every step after image processing.



(a) (b) (c)

Fig.1 Illustration of motion trend area ((a) is the original frame, (b) is the extracted motion area extracted, (c) is the motion trend extracted by our method)

Motion tracking. Kalman filter is a recursive estimator. It is characterized that we only need to know the previous signal at each recursive computation. Thus, making records of all the previous data is unnecessary, in other words, it is little resources consuming and time saving. What's more, the output have the minimum mean square error. In this paper, we establish model $\mathrm{M}_{t,i}(x_{t,i},y_{t,i},vx_{t,i},vy_{t,i})^T$ for each person. $(x_{t,i},y_{t,i})$ indicates the center coordinate of the model while $(vx_{t,i},vy_{t,i})$ represent the speed. Assume at the t-th frame that we have N_t objects detected, especially $i\in\{k|1\leq k\leq N_t,k\in Z^+\}$.

According to Kalman filter theory from [5] for each model we have State Equation Eq.5 and Observation equation as Eq.6.

$$\boldsymbol{M}_{t,i}=\boldsymbol{\varphi}_{t|t-1,i}\boldsymbol{M}_{t-1,i}+\boldsymbol{W}_{t-1,i} \tag{5}$$

$$\boldsymbol{Z}_{t,i}=\boldsymbol{H}_{t,i}\boldsymbol{M}_{t-1,i}+\boldsymbol{V}_{t,i} \tag{6}$$

Where $\boldsymbol{\varphi}_{t|(t-1),i}$ denotes the state transition matrix of i-th object from time $t-1$ to t. $\boldsymbol{H}_{t,i}$ represents the observation matrix. The value of the matrix are as Eq.7. $\boldsymbol{W}_{t-1,i}$ is system noise vector, $\boldsymbol{V}_{t,i}$ is observation vector. $\boldsymbol{Q}_{t,i}$ and $\boldsymbol{R}_{t,i}$ are the covariance matrix of $\boldsymbol{W}_{t-1,i}$ and $\boldsymbol{V}_{t,i}$.

$$\boldsymbol{\varphi}_{t|(t-1),i}=\begin{bmatrix}1 & 0 & \Delta t & 0\\ 0 & 1 & 0 & \Delta t\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}\qquad \boldsymbol{H}_{t,i}=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix} \tag{7}$$

Where Δt is time interval from $t-1$ to t. $\boldsymbol{Q}_{t,i}$ is the covariance matrix of $\boldsymbol{W}_{t-1,i}$ as Eq.8, $\boldsymbol{R}_{t,i}$ is the covariance matrix of $\boldsymbol{V}_{t,i}$ as Eq. 9:

$$\mathbf{Q}_{t,i}=E\{\boldsymbol{W}_{t-1,i}\boldsymbol{W}_{t-1,i}^T\} \tag{8}$$

$$\mathbf{R}_{t,i}=E\{\boldsymbol{V}_{t,i}\boldsymbol{V}_{t,i}^T\} \tag{9}$$

The object i state is predicted by Kalman filter recursive as follows Eq.10 to Eq.14.

$$\boldsymbol{M}_{t|t-1,i}=\boldsymbol{\varphi}_{t|(t-1),i}\boldsymbol{M}_{t-1|t-1,i} \tag{10}$$

$$\boldsymbol{P}_{t|t-1,i}=\boldsymbol{\varphi}_{t|t-1,i}\boldsymbol{P}_{t-1|t-1,i}\boldsymbol{\varphi}_{t|t-1,i}^{\boldsymbol{T}}+\mathbf{Q}_{t,i} \tag{11}$$

$$\boldsymbol{K}_{t,i}=\boldsymbol{P}_{t|t-1,i}\boldsymbol{H}_{t,i}^T\left[\boldsymbol{H}_{t,i}\boldsymbol{P}_{t|t-1,i}\boldsymbol{H}_{t,i}^T+\mathbf{R}_{t,i}\right]^{-1} \tag{12}$$

$$\boldsymbol{M}_{t|t,i}=\boldsymbol{\varphi}_{t|(t-1),i}\boldsymbol{M}_{t-1|t-1,i}+\boldsymbol{K}_{t,i}\left[\boldsymbol{Z}_{t,i}-\boldsymbol{H}_{t,i}^T\boldsymbol{M}_{t|t-1,i}\right] \tag{13}$$

$$\boldsymbol{P}_{t|t,i}=\left[\boldsymbol{I}-\boldsymbol{K}_{t,i}\boldsymbol{H}_{t,i}^T\right]\boldsymbol{P}_{t|t-1,i} \tag{14}$$

Where $\boldsymbol{P}_{t|t-1,i}$ denotes the posteriori error matrix and $\boldsymbol{P}_{t-1|t-1,i}$ priori error matrix in the moment $t-1$.

However, motions of pedestrians in next two frames are taken as uniform motions by Kalman

filter model. Once the pedestrian moves fast, the filter will not be able to accurately calculate the next state. For handing the shortcoming, we use motion trend graphs to approximate the real motion. Through the Kalman filter Eq.10- Eq.14, $\boldsymbol{M}_{t|t-1,i}(x_{t|t-1,i}, y_{t|t-1,i}, v_{x,t|t-1,i}, v_{y,t|t-1,i})^T$ indicate the position of pedestrian in ideal model. For pedestrian are moving continuously, we set a circle $C_i(cx_i, cy_i, r_i)$, which satisfies the Eq.15, as in Fig.2 to limit next rectangle center position.

$$\begin{cases} cx_i = x_{t|t-1,i} \\ cy_i = y_{t|t-1,i} \\ r_i = \mu\Delta t(v_{x,t|t-1,i}^2 + v_{y,t|t-1,i}^2)^{\frac{1}{2}} \end{cases} \tag{15}$$

Where μ is a parameter to control the sensibility of motion-detection. The motion trend graph $T(t)$ in segmentation is used to reaching the real position $\boldsymbol{M}_{t,i}$. Circle C_i is used to limiting the center of predict rectangle while $T(t)$ finding the exact location.

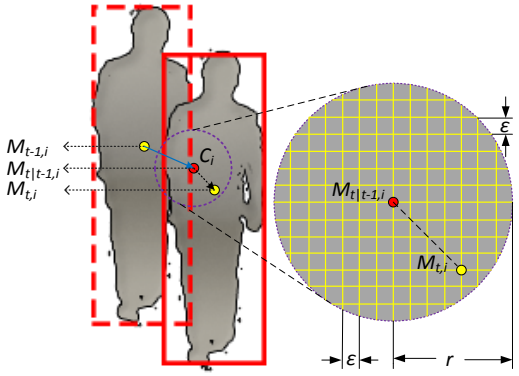


Fig.2 Illustration of prediction by Kalman filter and motion trend

Position $\boldsymbol{M}_{t,i}$ is calculated by Eq.16. The pedestrian is bounded by rectangle $R_{t,i}(xr_i, yr_i, w_i, h_i)$, in which (xr_i, yr_i) is the geometrical center of $R_{t,i}$. The width and height of rectangle $R_{t,i}$ are w_i and h_i respectively.

$$\begin{cases} S\left(R_{t,i}(xr_i, yr_i, w_i, h_i)\right) \geq S\left(R'_{t,i}(xr'_i, yr'_i, w_i, h_i)\right) \quad \forall(xr'_i, yr'_i) \in C_i \\ \boldsymbol{M}_{t,i} = \left[xr_i, yr_i, \frac{x_{t,i} - xr_i}{\Delta t}, \frac{y_{t,i} - yr_i}{\Delta t}\right]^T \end{cases} \tag{16}$$

Where S is a score function designed to get the predicted rectangle weights of pedestrian by this method. The function can be written as Eq.17. Also we find to calculate all the rectangle which center is in C_i is unnecessary. To handle this problem we set up parameter ε as the center moving step in Fig. 3.

$$S(R_{t,i}) = \sum_x \sum_y [(1-\alpha)T(x,y,t) + \alpha D(x,y,t)] \tag{17}$$

Especially in Eq.17 $x\epsilon[xr_i - \frac{w_i}{2}, xr_i + \frac{w_i}{2}]$, $y\epsilon[yr_i - \frac{h_i}{2}, yr_i + \frac{h_i}{2}]$. α is a parameter to balance the weight of motion and motion-trend.

Verification. In previous section we get the probably position $\boldsymbol{M}_{t,i}$ of the pedestrian by motion-trend and Kalman filter. Although it can exclude the cases that is absolutely worry, it cannot avoid the situations that new location really contain the pedestrian. In order to solve this problem we extract BRISK descriptor of each pedestrian when detected at the first time by static detection. We choose HOG+SVM method as the static detection method.

We initial keypoint model K as Eq. 18 for each pedestrian i which is detected by HOG+SVM at the first time.

$$K_i^{init} = \{(r_j, f_j)\}_{j=1}^{N_{f,t}} \tag{18}$$

Where $r \in R^2$ denotes the coordinates of each keypoint in frames and $f \in \{0,1\}^{512}$ denotes the BRISK feature descriptor. While $N_{f,t}$ denotes the total number of keypoints in the pedestrian rectangle in the t-th frame.

When we detected a new pedestrian position $\boldsymbol{M}_{t+n,i}$ by method mentioned in previous section, we should get the keypoints again and save to K_i^{t+n}. In order to judge whether the new location is

legal or not, we compute the Hamming distance between K_i^{t+n} and the initial model K_i^{init} by Eq. 19.

$$d_{i,j}(f^t, f^{t+n}) = \sum_{k=1}^{512} \left(f_i^t, f_j^{t+n}\right) \tag{19}$$

According to Ref. [2] the corresponding keypoint must satisfy the condition that the nearest neighbor is closer than the second-nearest neighbor by 80 percentage and distance should less than 90. We define the number of all the matched keypoints as N_{match}. To verify whether the new position is legal or not, we use Eq. 20.

$$\frac{N_{match}}{N_{f,t}} \geq 50\% \tag{20}$$

Although the verification step cost much time, it largely improves the precision of detection.

Analysis. To give a reliable estimate for the algorithm proposed by this paper, we conduct experiment both on accuracy and detection-time. In order to evaluate the performance of every stage in our method, we divide the algorithm into two approaches. One is only adapted with prediction method. The other with one more step of verification mentioned above. Both are using the HOG+SVM for static pedestrian detection. We name the first method as HKPT and the second as HKPTV. The letter H is for HOG+SVM, K for Kalman filter, PT for prediction by motion trend, V for verification. The original method is HOG+SVM by Piotr Dollar. PETS2013 dataset is selected as the video sequence for the evaluation. All the experiment are conducted on the same PC.

Accuracy analysis. We plot Detection Error Tradeoff (*DET*) curves [1] to assess the accuracy. As it shows in Fig. 4. Our improvement is significant.

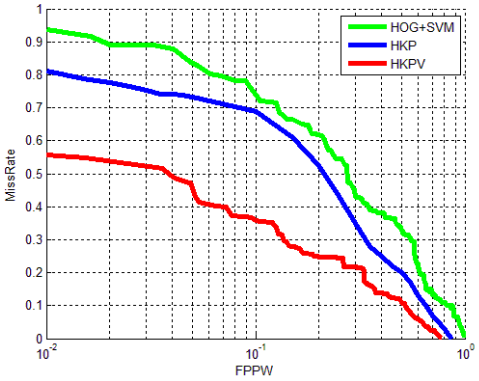


Fig. 4 Accuracy analysis (DET curves) on PETS2013

HKPT reduce the miss rate by 8.53% in average. For using the motion trend as the prior knowledge can largely reduce the detection region and give a guidance for next frames. But this method is not based on pedestrian detection for its only consider the motion trend. Once the environment become complex, in which other object move frequently, the method would not have a good performance. While HKPTV fixed it by the step of verification by BRISK keypoint calculation. It reduced the miss rate by another 19.7% in average. Collecting and verifying the keypoint of every pedestrian can distinguish which box detected by motion trend method really have pedestrian.

Speed analysis. Not like the original method HOG+SVM , which using sliding windows to select the region and compute the HOG descriptor all the time, our improvement only need to calculate the HOG descriptor at the first time the pedestrian appears. Some other step like Kalman filter calculation and new position prediction only need few computations by comparison.

Table. 1 the running time of the three approaches

Method	Frames per second
HOG+SVM	1.2
HKPT	28.6
HKPTV	13.1

From table.1, we can obviously see our method showed impressive result that 10.9 times faster

than the origin. Through HKPTV is slow when compared with HKP, the accuracy is more important. In brief, the algorithm reach high efficiency.

Summary

In this paper, we proposed a novel algorithm base on motion trend segmentation, Kalman filter and BRISK descriptor to solve the low efficiency of pedestrian detection caused by sliding windows method. We firstly using foreground segmentation to get the movement region. In addition, we enhanced it to get the motion trend. Secondly, Kalman filter is combined with the motion trend analysis to predict the pedestrian location by previous frame. At last, to make the result reliable the verification by BRISK descriptor is proposed. The experiment on accuracy and time sides show the algorithm in this paper make significant progress.

Reference

[1] Benenson R, Omran M, Hosang J, et al. Ten years of pedestrian detection, what have we learned?[C]//Computer Vision-ECCV 2014 Workshops. Springer International Publishing, 2014: 613-627.

[2] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005, 1: 886-893.

[3] Benenson R, Timofte R, Van Gool L. Stixels estimation without depth map computation[C]//Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, 2011: 2010-2017.

[4] Lampert C H, Blaschko M B, Hofmann T. Beyond sliding windows: Object localization by efficient subwindow search[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

[5] Paris S, Durand F. A fast approximation of the bilateral filter using a signal processing approach[J]. International journal of computer vision, 2009, 81(1): 24-52.

[6] Leutenegger S, Chli M, Siegwart R Y. BRISK: Binary robust invariant scalable keypoints[C]//Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011: 2548-2555.