Human tracking using the Kalman filter

Ernesto Pérez Costa

INAOE
http://www.inaoep.mx

Abstract. Human tracking systems are commonly used for visual surveillance, but are also a key component in systems with other objectives such as activity recognition and behavior understanding. A variety of methods have been developed for tracking single or multiple pedestrians in static or moving cameras by exploiting different types of image information. The goal of this paper is to examine possible methods and based on these, to develop a good human detection and tracking algorithm.

Keywords: tracking, detection, Kalman filter

1 Introduction

A variety of methods have been developed for tracking single or multiple pedestrians in static or moving cameras by exploiting different types of image information.

Background subtraction

Background separation methods track foreground objects by measuring the differences between video frames and a learned background model. These methods tend to be fast and often have the additional advantage of recovering foreground object silhouettes.

- **University of Cranfield:** Tracking using a combination of background separation and optical flow
- **University of Edinburgh:** Long term tracking from an overhead camera
- **University of Leeds:** Tracking system making use of Active Shape Models to separated foreground regions (EPSRC IMV project)
- **University of Loughborough:** Background separation with foreground silhouette extraction
- **University of Reading:** Further development of the Active Shape Model tracking system (EU ADVISOR project)

Appearance and motion

Appearance based pedestrian tracking methods work by identifying tracks of image regions that look like people. These methods can be advantageous in dense crowds where occlusions can cause other methods to fail, but are usually more computationally expensive than background separation based methods.

- **University of Leeds:** Pedestrian tracking system making use of Condensation to fit template models
- **University of Oxford:** Pedestrian tracking system based on detecting heads

2 System Overview

In a detection method, one of the popular feature detection, Histogram of Oriented Gradients (HOG) was used. Because the HOG performance is slow with video frames, linear Support Vector Machine (linear SVM) classifiers are applied for better performance. Tracking bases on the detection results, to be more precise, Kalman filter uses detected features for tracking and also for predicting the position of the pedestrian [Fig. 1].

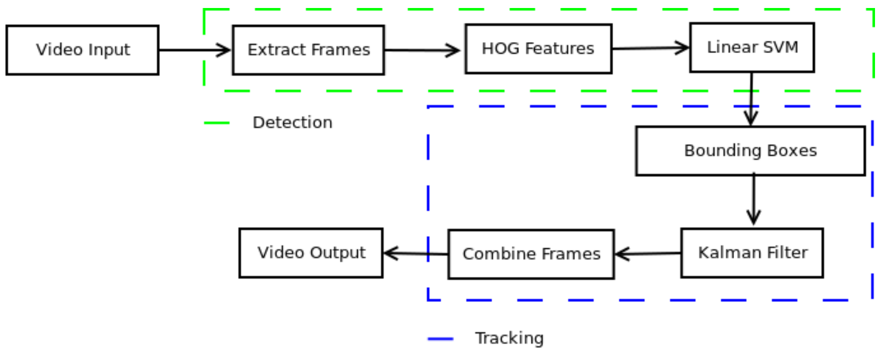

Fig. 1. System Overview

3 Methodology

3.1 Detection

An overview to object detection chain with HOG and SVM, is given in [Fig. 2]. The main idea is that, even without knowing the corresponding gradient or edge positions, by the distribution of local intensity gradients or edge directions, local object appearance and shape can be well characterized. An examlpe of the detector performance is given in [Fig. 3]

3.2 Tracking

Thresholding

Thresholding is one of the simplest methods of image segmentation, which can create binary images from a grayscale image. The purpose is to improve the

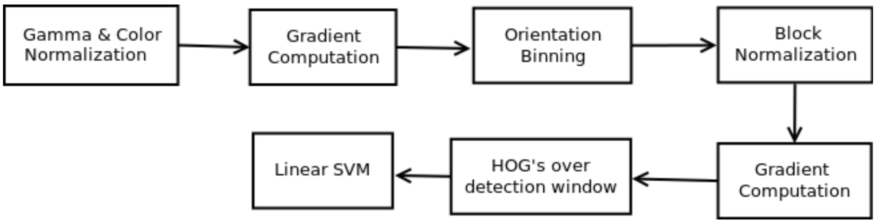

Fig. 2. Detection overview

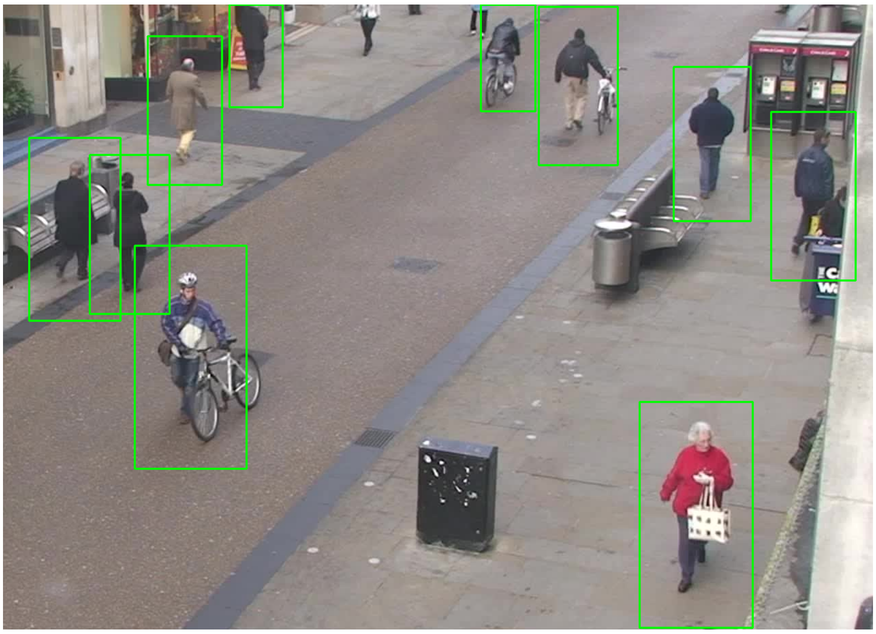

Fig. 3. HOG + Linear SVM detection

quality of the image and extract pixels from image, which represent an object. The working principle of the simplest methods of thresholding is to replace pixels of the image with black or with white pixels. If an intensity of the image is smaller than fixed constant value, so-called intensity threshold, each pixel in an image replaced with a black pixel, if bigger that intensity threshold, then replaced with a white pixel.

Background Subtraction

For improving productivity, BackgroundSubtractorGMG , offered by OpenCV 3.4, which performs Bayesian based foreground and background segmentation was used [Fig. 4].

Fig. 4. Left: Original image - Right: Background subtraction

Dilatation

Dilatation is one of the primary operations in mathematical morphological transformations. Is widely being used in varied contexts, for instance, eliminating the noise of the image, isolating individual elements, and joining disparate elements of the image. This morphological transformation can also be utilized in finding intensity peaks in a picture, and to determine a particular form of an image gradient. For expanding the shapes in the input image, usually, dilation operation uses a structuring element. Dilation is a convolution of the image with the kernel (usually, "solid" square kernel, or sometimes, a disk). In the kernel, each given pixel is replaced with the local maximum of all of the kernel covered pixel. Actually, the exact result depends on the kernel, but generally, dilation expands a bright region and tend to fill concavities in the image [Fig. 5].

3.3 Kalman filter

Kalman filter, also known as Linear quadratic estimation, is a series of mathematical equations, which, by minimizing the mean of the squared error, provides an efficient computational means to determine the state of a process given the

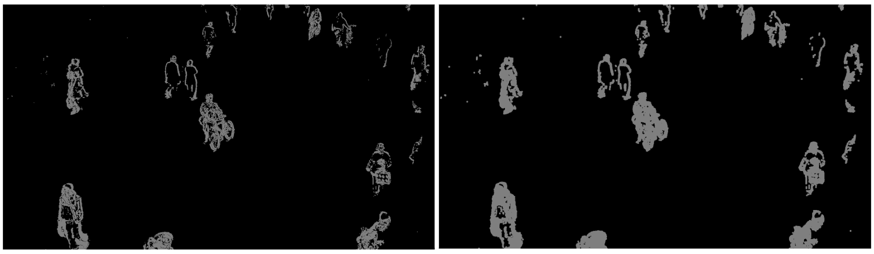

Fig. 5. Left: Original image - Right: Dilated image

previous state. This filter can estimate states of future time, even when the precise nature of the modeled system is unknown. The Kalman filter has numerous applications in technology, such as navigation, guidance, vehicles, aircraft. Moreover, The Kalman filter is a widely used in signal processing, econometrics, and robotic motion planning and control fields. The algorithm of this filter works in

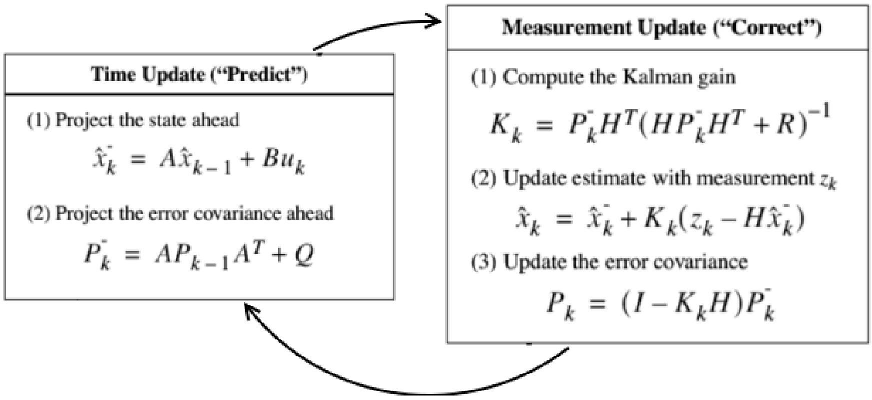

Fig. 6. Kalman filter algorithm

two steps, such as prediction and update [Fig. 6]. At first, the filter estimates the current state variables with their uncertainties in the prediction step. After obtaining the estimates of the next measurement, by utilizing a weighted average, these estimates are updated. The estimates with higher certainty, get more weight. As the algorithm of the filter is recursive without any additional past information, only by utilizing the present input measurements and the previously calculated state and its uncertainty matrix, it can run in real time. These filters built on linear operators, perturbed by errors, which Gaussian noise may include. The Kalman filter does process estimation by feedback control. At some time, the filter estimates the process state, then gets feedback as noisy measurements. Kalman filter works in two groups, such as time update equations and measurement update equations. For obtaining the priori estimates for the next time step, the time update equations predict the current date, and error covariance estimates. To obtain a corrected a posteriori estimate, the measurement update equations include a new measurement into the a priori estimate. The time update equations are also known as predictor equations and the measurement update equations as corrector equations.

Model

The state space consists of the six parameters vector $x_{k-1} = [x, y, v_x, v_y, h, w]$.
Transition matrix:

$$A = \begin{bmatrix} 1&0&1&0&0&0 \\ 0&1&0&1&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \end{bmatrix}$$

Covariance matrix:

$$P = \begin{bmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&10^3&0&0&0 \\ 0&0&0&10^3&0&0 \\ 0&0&0&0&0&0 \\ 0&0&0&0&0&0 \end{bmatrix}$$

Process noise matrix:

$$Q = \begin{bmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&0.1&0 \\ 0&0&0&0&0&0.1 \end{bmatrix}$$

Measurement noise matrix:

$$R = \begin{bmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&10&0 \\ 0&0&0&10 \end{bmatrix}$$

Measurement function:

$$H = \begin{bmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \end{bmatrix}$$

4 Results

The tracking method depends on the output of the detection process, the system can not track a pedestrian without detection. But the system can still keep track when it loses the detection. The average tracking performance for each state measurement variable is presented in [Fig. 7]. Those errors are at pixel scale. The difference between lines is 10px, except for the w state, where the difference is 5px. Results were computed by comparing real per detection features (x, y, h, w) average *vs.* average of the characteristics predicted by the Kalman filter.

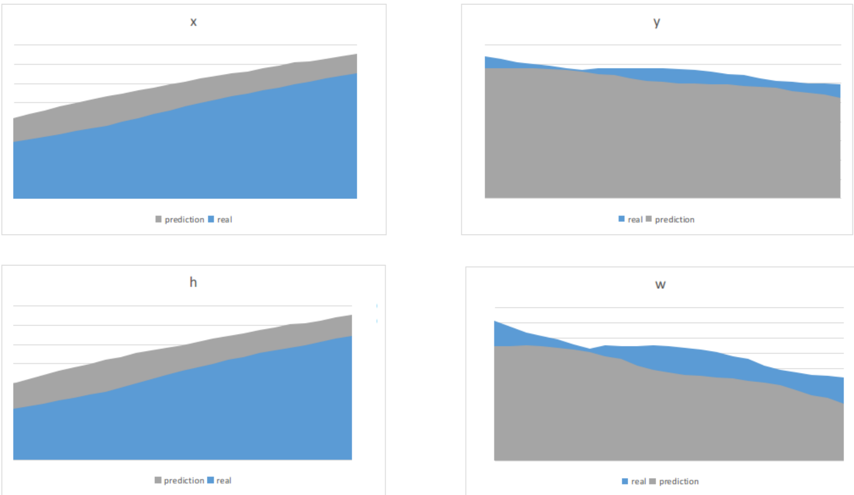

Fig. 7. Kalman filter average tracking performance

5 Conclusion

In this paper was introduced a human detection and tracking method. The general architecture and most important components of the system were explained. All obtained results were discussed.

5.1 Observations

- Background Subtraction approach (without the classifier) is pretty fast and suitable for real-time applications but the problem is that it is very sensitive to any change in illumination and consequently produce lots of false positives and wrong detections, which make it a non-robust approach for most real-world applications.

- Kalman filter is good for keeping track of occluded persons since it keeps predicting their current position -based on their previous dynamics- even when thier detections disappear for some frames.

5.2 Future enchancement

- Use Extended / Unscented Kalman Filter, given that people's motion is highly nonlinear.
- Use Kalman filter along another tracker (CAMShit / MeanShift).
- Use a more robust human detector (Yolo).

References

[Bewley et al., 2016] Bewley, A., Ge, Z., Ott, L., Ramos, F., and Upcroft, B. (2016). Simple online and realtime tracking. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 3464–3468. IEEE.

[Dalal and Triggs, 2005] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.

[Kaewtrakulpong and Bowden, 2001] Kaewtrakulpong, P. and Bowden, R. (2001). An improved adaptive background mixture model for realtime tracking with shadow detection.

[Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45.

[Sobral, 2013] Sobral, A. (2013). Bgslibrary: An opencv c++ background subtraction library. In *IX Workshop de Visao Computacional*, volume 2, page 7.

[Wan and Van Der Merwe, 2000] Wan, E. A. and Van Der Merwe, R. (2000). The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158. Ieee.

[Zivkovic and Van Der Heijden, 2006] Zivkovic, Z. and Van Der Heijden, F. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, 27(7):773–780.