Multi-pedestrian Tracking using Kalman Filter and Particle Filter

Jiangpeng Tao
jtao@kth.se

Wenji Yin
yinw@kth.se

January 18, 2018

Abstract — In this report, we propose two methods for multi-pedestrian tracking in the framework of Kalman filter and particle filter respectively. We combine the ACF detector with background subtraction to improve its accuracy, especially pedestrians with small size and large range of motion. To solve occasional drift in the measurement previous two states are used to correct the direction of velocity measurement. In the condition of splits and merges our algorithm reassigns trackers based on previous walking direction after data association. We evaluate the performance on the standard pedestrian dataset and show that our methods can track multi-pedestrian properly.

I. INTRODUCTION

Multi-pedestrian tracking has received much research interest in computer vision, since it can be applied in many fields, such as security monitoring, driverless cars, intelligent service robots etc. There are a lot of difficulties of this tracking problem. The measurement uncertainty of a certain detector might change dramatically due to changing background and occlusion. The motion of pedestrians depends on individual initiative, so it is hard to establish a general motion model. In addition, the similar appearance, splits and merges make association of detections much challenging.

In order to deal with these difficulties, researcher did a lot of works. Many detection approaches rely on background subtraction [1] [2] and can extract segmentations of pedestrians successfully. However, such methods can only be applied in unchanging background and cannot distinguish pedestrians and other moving objects. Recently, researchers focus more on methods using machine learning [3][4][5]. These methods are robust to changing background and improve detection accuracy greatly.

Most multi-pedestrian tracking algorithms are based on the Kalman filter and particle filter. Because the Markovian approaches only rely on information from past frames and is suitable for real-time applications. In [6], Bayesian networks and Kalman filter are used to combine multiple modalities for matching subjects. In [7], a two-dimensional color histogram is used to measure similarity between moving objects in a Kalman filter framework. However, this method might go wrong when associating pedestrians with similar dressing with detections. Xin Li et al. [8] proposed a treatment for merges and splits based on features matching to solve the occlusion problem. In addition, Kenji Okuma et al. [9] present a boosted particle filter which combine mixture particle filter and Adaboost. Michael et al. [10] uses the detectors' confidence and instance-specific classifiers as a graded observation model in a particle filtering framework.

The goal of this project is to detect and track a number of pedestrians and propose a tracking method which is robust to detection noise, merges and splits.

II. PEDESTRIAN DETECTION

A. *GMM method*

For a fixed camera, the background in the video usually keeps constant. A common approach for discriminating pedestrians from the background is detection by background subtraction. Gaussian Mixture Model (GMM) is a popular method that has been employed to tackle the problem of background subtraction. This method combines statistical background image estimation and per-pixel Bayesian segmentation, so some frames without any pedestrians are extracted from video to train the background model. Then morphological opening and closing are applied to remove noise in the foreground segmentation. The positions and sizes of pedestrians are detected by blob analysis. However, the detection result is sensitive to illumination variation, shadow of pedestrians and occlusion.

B. ACF method

Another common method for detecting pedestrians is aggregated channel features (ACF). We try to use the ACF people detector in MATLAB which was trained using the Caltech Pedestrian data set. Its accuracy rate to pedestrians' positions and sizes is higher than that of GMM detector. However, compared to GMM detector the defect of the ACF detector is it usually cannot detect pedestrians with too small sizes or too large range of motion.

C. Improvement

We combine the detection results of these two methods above. The GMM result $g_t = \{g_1, g_2, \ldots, g_n\}$ consists of n recognized states. every recognized state $g_n = \{x, y, l, w\}$ consists of the target's central position (x, y) and the target's size (l, w). Similarly, the ACF result is $a_t = \{a_1, a_2, \ldots, a_m\}$. We eliminate the states in g_t that represent the same targets in a_t due to ACF's higher accuracy rate to positions and sizes. The algorithm for combing GMM and ACF detection results is depicted in Table 1.

Algorithm_combine_GMM_ACF (g_t, a_t):

$S_t = \varnothing$

$for\ i = 1\ to\ m$

  $Dist(j) = \sqrt{(g_{j,x} - a_{i,x})^2 + (g_{j,y} - a_{i,y})^2}$

  $k = arg\min_j Dist(j)$

  $if\ Dist(k) > a_{i,w}$

    $S_t = S_t \cup g_k$

  $end\ if$

$end\ for$

$S_t = S_t \cup a_k$

$return\ S_t$

Table 1 the algorithm for combining GMM and ACF detection results

III. TRACKING USING KALMAN FILTER

The first tracking algorithm estimates the distribution of each target state by a Kalman filter. The state $\boldsymbol{x} = \{x, y, u, v, l, w\}$ consists of the target's central position (x, y), the velocity components (u, v) and the target's size (l, w).

A. Process model

Generally, the frame rate of videos is over 25 Hz, so we assume that pedestrians do not change their velocities and sizes in such short time. Therefore, we use a constant size and velocity process model:

$$(x, y)_t = (x, y)_{t-1} + \Delta t \cdot (u, v)_{t-1} + \varepsilon_{(x,y)} \quad (1)$$
$$(u, v)_t = (u, v)_{t-1} + \varepsilon_{(u,v)} \quad (2)$$
$$(w, l)_t = (w, l)_{t-1} + \varepsilon_{(w,l)} \quad (3)$$

The process noise $\varepsilon_{(x,y)}$, $\varepsilon_{(u,v)}$, $\varepsilon_{(w,l)}$ is independently white Gaussian. Δt depends on the frame-rate of the video. We apply a strategy to update the variance of the process noise [10]. The variance $\sigma^2_{(x,y)}$ for the position noise is directly proportional to the size of the tracking target, whereas the variance $\sigma^2_{(u,v)}$ and $\sigma^2_{(w,l)}$ is inversely proportional to the number of successfully tracked frames.

The process equation is given by (4) (5)

$$x_t = A_t x_{t-1} + \varepsilon_t \quad (4)$$

$$A_t = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

Where A_t is the transition matrix and ε_t is the Gaussian measurement noise with zero mean and covariance R_t.

B. Measurement data

The pedestrian detector generates measurement data of targets' positions and sizes as a sensor.

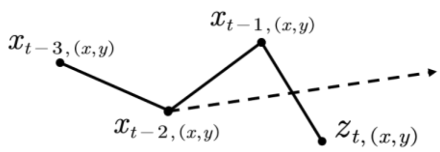


Figure 1. Connection diagram of measurement data and states

Generally, the connection diagram of measurement data and states is as shown Figure 1. If we only consider previous state x_{t-1} and detection data z_t to compute velocity measurement value, the direction of the velocity might deviate largely from the main tracking direction. Thus, we consider two previous states x_{t-1}, x_{t-2} and z_t to measure the velocity at time t. Its direction is as shown by the dotted line in Figure 1.

$$z_{t,u} = \left[\frac{(z_{t,x} + x_{t-1,x})}{2} - x_{t-2,x} \right] / \Delta t \tag{6}$$

$$z_{t,v} = \left[\frac{(z_{t,y} + x_{t-1,y})}{2} - x_{t-2,y} \right] / \Delta t \tag{7}$$

C. Observation model

The observation model for the detection is given by equation (8)

$$z_t = C_t x_t + \delta_t \tag{8}$$

Where C_t is the measurement matrix and an identity matrix. δ_t is the Gaussian measurement noise with zero mean and covariance Q_t.

D. Data association

In order to assign each detection to at most one corresponding tracker, we solve the data association problem by the Hungarian algorithm. A matching score matrix describes the matching likelihood between each detection and tracker and is defined as equation (9):

$$S(tr,d) = \lambda D_{pos}(tr,d) + (1-\lambda) D_{size}(tr,d) \tag{9}$$

Where $D_{pos}(tr,d)$ is the Euclidean distance between the predicted centroid of the tracker and the centroid of the detection. Similarly, $D_{size}(tr,d)$ is the Euclidean distance between the predicted size and the detection size. The constant $\lambda \in [0,1]$ represents the weight of $D_{pos}(tr,d)$. Then, the Hungarian algorithm iteratively select pairs (tr,d) with minimum score that is also under the threshold.

E. Split and merge problem

When two pedestrians cross with each other, the detector usually can only detect one person due to occlusion, which will cause one tracker cannot be assigned to the detection. Also, when these two pedestrians separate after merge the matching scores are very close, therefore the assignment easily go wrong.

We propose a method to solve this problem. Before applying the Hungarian algorithm, we copy the detection if two pedestrians would cross with each other according to the predicted states and only one pedestrian is detected. The algorithm is depicted in Table 2. We assumed that pedestrians would not walk back the way they come at the moment of crossing with another pedestrian. Therefore, after applying the Hungarian algorithm we exchange the two corresponding assignments if their current direction is much different from the previous main walking direction.

Algorithm_copy_detection $(d, \overline{x}_t)$**:**

m *is the number of trackers*
$for\ i = 1\ to\ m$
  $Dist(j) = \sqrt{(\overline{x}_{t,j,x} - \overline{x}_{t,i,x})^2 + (\overline{x}_{t,j,y} - \overline{x}_{t,i,y})^2}$
  $k = arg\min_{j \neq i} Dist(j)$
  $if\ Dist(k) < \overline{x}_{t,i,w}\ \&\ only\ one\ d_l$
    $d = d \cup d_l$
  $end\ if$
$end\ for$
$return\ d$

Table 2 Algorithm for copying the detection when one pedestrian is covered by another pedestrian and cannot be detected

IV. TRACKING USING PARTICLE FILTER

The particle filter is to represent the posterior belief $bel(x_t)$ by a set of random state samples drawn from $\overline{bel}(x_t)$. The main difference of multiple tracking from that in lab 2 is associating particle sets with the observations.

A. Motion model

We use a lower dimensional state $\boldsymbol{x} = \{x, y, u, v\}$. Without the size components, the motion model is the same as that in the Kalman filter. The motion model is used to predict particles' states.

$$\overline{x}_t^{[m]} = A_t x_{t-1}^{[m]} + N(0,R) \tag{10}$$

B. Observation model and weighting

Without the size components, the measurement z_t only consists of the central position.

$$z_t = C'_t x_t + \delta_t \tag{11}$$

$$C_t' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (12)$$

Where C_t' is the measurement matrix. δ_t is the Gaussian measurement noise with zero mean and covariance Q_t.

The importance weight $\omega_t^{[m]}$ for each particle m at time t is described by:

$$\omega_t^{[m]} \propto p(z_t | x_t^{[m]}) \quad (13)$$

C. Data association

Similarly, we solve the data association problem by the Hungarian algorithm. The matching score matrix is defined by the Euclidean distance between the predicted centroid of the tracker and the centroid of the detection.

$$S(tr, d) = D_{pos}(tr, d) \quad (14)$$

We use the Hungarian algorithm to assign each particle to the detection. The corresponding detection of each particle set depends on the assignment which is voted most.

D. Resampling procedure

We use systematic re-sampling method to erase those particles with low weights. The longer a target is tracked successfully, the smaller the variance of the particles is. If the target changes his velocity suddenly, such as stopping and turning back, particle deprivation easily occurs due to the low variance of the particles. So, in order to avoid particle deprivation, we add a small number of randomly generated particles into the set after resampling process when the variance is relatively small.

E. Density extraction

In order to get pedestrians' trace, we fit a single Gaussian model to the mean and variance of the particle sets. Each target's state x_t at time t equal to the mean of its particle set.

V. EXPERIMENTAL RESULTS

We evaluate our tracking algorithm on the sequence: "EPFL" Multi-camera pedestrian video data set [11].

A. Tracking multiple pedestrians

Figure. 2 shows the tracking results in an outdoor scene. In the video three persons walk from left in turn and the person in black make a turn in the middle. In Figure. 2, all three persons have been tracked properly. The yellow trace has a sharply zigzag part due to dramatic change of position measurement. The velocity measurement considers previous two states, which diminish the impact of some inaccurate measurement on the velocity direction.

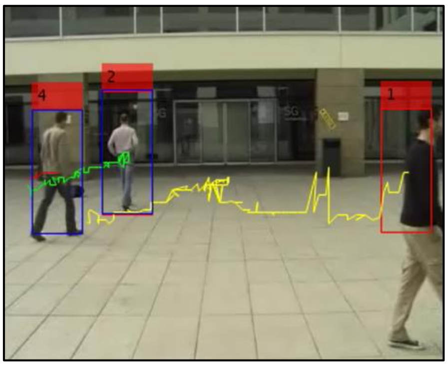
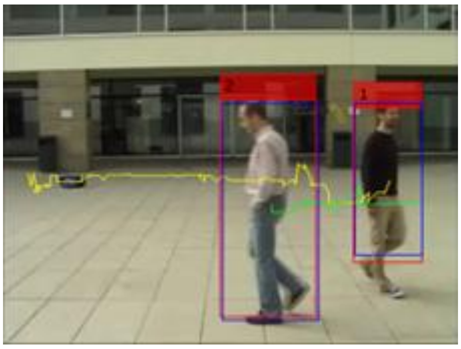


Figure 2 Tracking result using Kalman filter for three persons in scene A (The blue windows denotes the detection result, the red windows with tracker serial numbers denote the estimated result and the lines denote targets' traces)

B. Tracking pedestrians with splits and merges

Figure. 3 shows the tracking results in the same outdoor scene. In the video, the person in black walk from left to right and another person walk in opposite direction. At some point only the person in white can be detected due to occlusion. In Figure 3(a), the data association assigns wrong trackers due to neglecting previous walking direction. In Figure 3(b), both of them have been tracked properly in the assumed condition that pedestrians keep the previous direction after separate.

(a) Neglect previous walking direction



(b) Consider previous walking direction

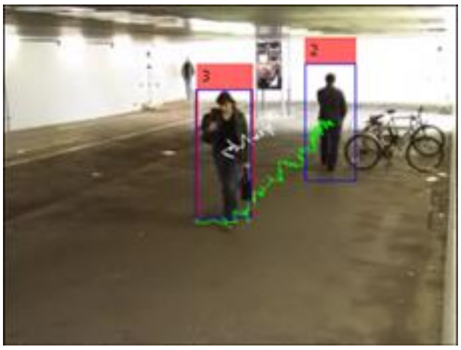
Figure 3 Tracking result using Kalman filter in scene A

C. Error analysis

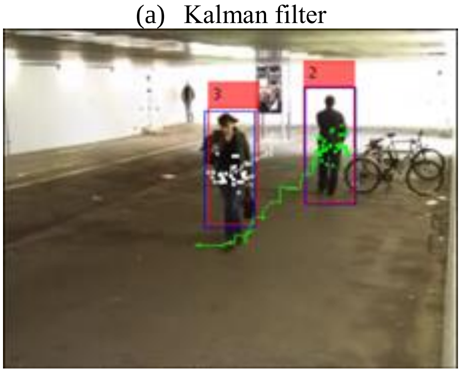
Figure. 4 shows the tracking result of a passageway sequence. Both filter has tracked the two persons successfully.

We created the ground truth data by locating the persons artificially every 5 frames. The tracking result is sampled every 20 frames and is evaluated using mean-square error (MSE). Figure. 5 shows the MSE of the two filters.

Both error curves are much close with each other. In the front part of video, the measurement is relatively steady and accurate. So, the estimated states coverage and get close to the ground truth as successfully tracking. However, in the later part of video the measurement become inaccurate, which cause the MSE increases.



(a) Kalman filter



(b) Particle filter (The points denote particle sets)

Figure 4 Tracking result for two persons in scene B

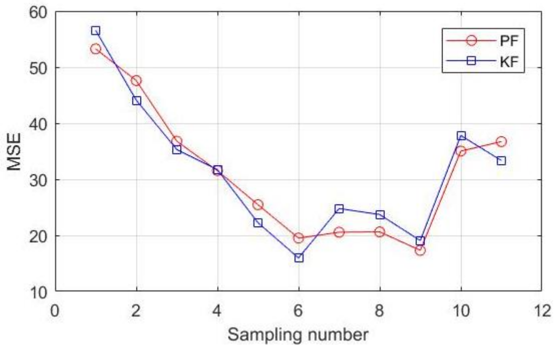


Figure 5 MSE of Kalman filter and particle filter (Sample every 20 frames)

VI. CONCLUSION

We have presented an approach for tracking multi-pedestrian using Kalman filter and particle filter. The main work of this report is: (1) improving ACF detection result by background subtraction; (2) establishing the constant velocity process model; (3) correcting direction of velocity measurement by considering past states; (4) combining previous walking direction with the Hungarian algorithm. As the experiments show, the approach tracks pedestrians successfully in spite of inaccurate measurements sometimes and conditions of splits and merges. Using similar models and data association method, the tracking results of Kalman filter and particle filter are quite similar. We believe results can be further improved if the matching score matrix includes color features and the observation model is based on detector's posterior confidence.

REFERENCES

[1] T. Zhao and R. Nevatia. Tracking multiple humans in complex situations. PAMI, 26(9):1208-1221, 2004.

[2] O. Lanz. Approximate Bayesian multibody tracking. PAMI, 28(9):1436-1449, 2006.

[3] P. Sermanet, K. Kavukcuoglu, S. Chintala, et al. Pedestrian detection with unsupervised multi-stage feature learning. CVPR, 3626-2633, 2013.

[4] P. Dollár, R. Appel, S. Belongie, et al. Fast feature pyramids for object detection. PAMI, 36(8): 1532, 2014

[5] S. Paisitkriangkrai, C. Shen, A. van den Hengel. Efficient pedestrian detection by directly optimize the partial area under the roc curve. ICCV, 1057-1064, 2013

[6] T. H. Chang, S. Gong. Tracking Multiple People with a Multi-Camera System. IEEE Workshop on Multi-Object Tracking, 19-26, 2001

[7] A. Czyzewski and P. Dalka. Examining Kalman Filters Applied to Tracking Objects in Motion. Ninth

International Workshop on Image Analysis for Multimedia Interactive Services, 175-178, 2008.

[8] X. Li, K. Wang, W. Wang, et al. A Multiple Object Tracking Method Using Kalman Filter. IEEE International Conference on Information and Automation, 1862-1866, 2010.

[9] K. Okuma, A. Taleghani, N. De Freitas, et al. A boosted particle filter: Multitarget detection and tracking. ECCV, 28-39, 2004.

[10] D. M. Breitenstein, F. Reichlin, B. Leibe, et al. Robust Tracking-by-Detection using a Detector Confidence Particle Filter. IEEE International Conference on Computer Vision, 1515-1522, 2010

[11] J. Berclaz, F. Fleuret, E. Turetken et al. Multiple Object Tracking using K-Shortest Paths Optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(9): 1806-1819, 2011.