Pedestrians Crowd Segmentation Based on Bayesian Formulation

Yun Liu, Guangpeng Ma

School of Information Science & Technology, Qingdao University of Science & Technology, Qingdao 266000,China

Keywords: Bayesian Formulation, head detection, Pedestrian Crowd Segmentation.

Abstract. Crowd people segmentation algorithm in complex environments is proposed based on Bayesian formula. In this method solved the case of the occlusion of crowd pedestrian which are difficult to detect one pedestrian from each other. This algorithm uses the Bayesian formula to convert the problem that how to segmentation the crowd people into seeking maximum a posterior probability. Using the assumption segmentation model of the pedestrian to match the foreground area, and then modify segmentation model for converging to the maximum posterior probability through multiple iterations. Experiments result that using the PETS 2009 video database show that this algorithm can effectively and accurately detect the pedestrian whose occlusion is less half and the part of severely blocked pedestrians can also be identification.

Introduction

Pedestrian detection is to identify the video object whether it is pedestrian, and to recording its location. Correct detection of the pedestrian is the basis of tracing process. In general, the pedestrian detection algorithm can be divided into three categories: shape-based matching method[1][2], based on the feature vector descriptor method[3][4] and motion detection method based on foreground segmentation[5][6]. For only a single object, or the less number of pedestrian in sparse scene, we can get a blob by using foreground segmentation, and each blob is a pedestrian. But when the number of people increase in the scene or the pedestrians walking abreast, it couldn't detect each person that the blob of the foreground contains a lot of number pedestrians. To solve this issue, in[7] make a projection of the foreground region on the vertical direction, and get the vertical histograms. To locate the position of pedestrian head by analysing the fluctuation of histogram. This method could effectively solve the simple occlusion situation, but in the pedestrian crowd environments, high overlap between pedestrians, the valley of the object's vertical number of pixels would not be obvious, so it couldn't achieve a accurate segmentation in that situation. [9] using a multiple cameras to solve the serious occlusion, however due to the large number of cameras it is limited to applications of small spatial areas.

In this paper, we proposed a pedestrian crowd segmentation algorithm based on Bayesian formulation to solve this issue. This algorithm converted the segmentation to maximise the posterior probability, and the segmentation hypothesizing that would maximise posterior probability are the segmentation results of pedestrian crowd .

Bayesian formulation

Convert the problem of segmentation. Using θ to express the pedestrian crowd segmentation model, and if we want to get the optimal segmentation result θ^*. $\theta^* = \arg\max P(\theta|I)$, where I is foreground image included pedestrian. so we could use Bayesian formulation :

$$P(\theta|I) = \frac{P(I|\theta)P(\theta)}{P(I)} \tag{1}$$

$$P(\theta|I) \propto P(I|\theta)P(\theta) \tag{2}$$

In Equ.1, P(I) is a fixed constant value in a gaving frame image, so if we want to make $P(\theta|I)$ maximized, we can ignore P(I) to get Equ.2. $P(I|\theta)$ indicates the probability of the foreground

image under the segmentation model θ, P(θ) represents the probability of the division segmentation model θ.

Pedestrian model. Pedestrian's motion is non-rigid, so establish an accurate model to describe the pedestrian is needed. For application scenarios in this paper, the pedestrians only have two states, walk or stand. Although there are a variety of walking direction, they wouldn't affect the human body shape obviously. Through the study , we found that the ellipse model can effectively match the shape of the pedestrian's body. In Fig.1, (a) using three ellipse models to represent head, torso and legs, this model can be more accurately matched pedestrian, (b) using a single ellipse model roughly to simulate the entire pedestrian. By experiment attempts, a single ellipse model could also obtain better segmentation results, and it has some good , quickly computing and simplicity of the model characteristics.
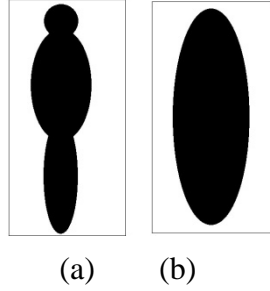

(a)   (b)

Fig.1 Pedestrian model

We assume that the pedestrians move on a ground plane, as the different distance of pedestrians from the camera, each pedestrian would have a different size in the image. Using O to represent each pedestrian and p、h、w represent pedestrian's position, height and width respectively.

Prior probability of segmentation model. The prior probability of P(θ) is the product of the probability of each pedestrian model, and the probability of each pedestrian model P(O) is dependent on the pedestrian's position, height and width. The formulation as follow:

$$P(\theta) = \prod_{i=1}^{n} P(O_i) \tag{3}$$

$$P(O_i) = P(p_i)P(h_i)P(w_i) \tag{4}$$

where $P(p_i)$ is the possibility of pedestrians' position, the height probability $P(h_i)$ obeys to the Gaussian distribution and different people have different height, but the heights are mainly obey to the Gaussian distribution and tend to a mean value which should be set lower for the place far away from the camera, the width possibility $P(w_i)$ also obeys to the Gaussian distribution.

Conditional probability of foreground. Foreground region is formed by overall the pedestrian, so the conditional probability couldn't be computed by product of each single pedestrian model due to the overlap between pedestrians. we use the likelihood between foreground region and the region covered by pedestrian model of segmentation model θ. It can be used as the conditional probability of the foreground region P(I|θ).
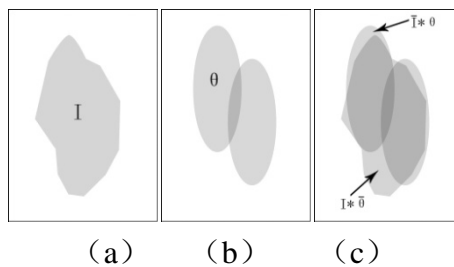

（a）   （b）   （c）

Fig. 2 Image of the likelihood of segmentation model and foreground

Fig.2(a) is foreground region I, Fig.2(b) is segmentation model θ, the overlap between them are shown in Fig.2(c). As shown in Fig.2 (c), the likelihood P(I|θ) is determined directly by the region that isn't overlap between foreground and model. We use follow parameters to represent these parts:

p_{10} denotes the probability that a pixel in the segmentation model does not in the foreground, p_{01} represents the probability that a pixel in the foreground does not in the segmentation model, N_{10}, N_{01} represents the number of pixels corresponding in the above region respectively. To prevent the over fitting of foreground region, N_{re} is the number of pixels that overlap between each segmentation model. Conditional probability of foreground region $P(I|\theta)$ is:

$$P(I|\theta) = e^{-(\alpha_{10}N_{10}+\alpha_{01}N_{01}+\alpha_{re}N_{re})} \qquad (5)$$

where $\alpha_{10}, \alpha_{01}, \alpha_{re}$ are weights of N_{10}, N_{01}, N_{re} respectively. By combining Equ.3 and Equ.5, the posterior probability Equ.2 becomes:

$$P(\theta|I) \propto (\prod_{i=1}^{n} P(p_i)P(h_i)P(w_i))e^{-(\alpha_{10}N_{10}+\alpha_{01}N_{01}+\alpha_{re}N_{re})} \qquad (6)$$

Computation of maximum posterior probability

Detect head position. We use the "Ω" shape of head and shoulder contour, this special shape could quickly detect all the head position. At first, extracting the contour of head and shoulder from sample, as Fig.3(a). The normal vector of each point are calculated by the edge direction, the center point of head is set as the origin of coordinates. The edge points set is $\{\vec{p_1}, ..., \vec{p_n}\}$ and the corresponding normal vectors set is $\{\vec{v_1}, ..., \vec{v_n}\}$. A Canny detector is applied to the foreground as Fig.3(d). For all points in the Canny image searching the nearest point to the edge, the nearest pixel position is recorded in matrix C and the distance is recorded in the matrix D. The normal vectors of Canny is recorded in a matrix of G.

The procedure of detection is like a filter that to scan all points one by one in Canny image, M is the likelihood between point(x, y) and the filter:

$$M(x,y) = \frac{1}{n}\sum_{i=1}^{n}(\vec{v_i} * G(C(\vec{p_i} + (x,y))))e^{-\alpha D(\vec{p_i}+(x,y))} \qquad (7)$$

where $e^{-\alpha D(\vec{p_i}+(x,y))}$ is the unlikelihood between the point in Canny image and head point, α could be set between 0.3 and 0.6. The point is the head candidate when its likelihood M greater than threshold, the result is shown in Fig.(f).

（a）　　　　　　　　（b）　　　　　　　　（c）
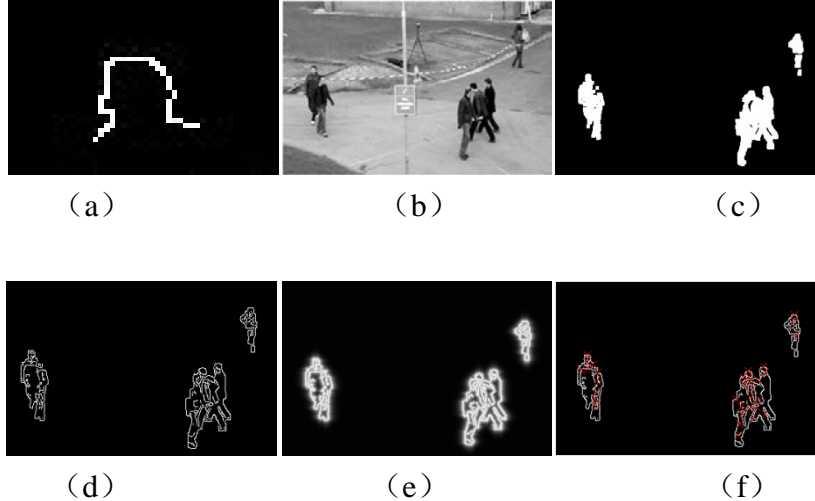

（d）　　　　　　　　（e）　　　　　　　　（f）

Fig.3 Detection head of pedestrian in image. (a)head shoulder model;(b) original image;(c)foreground image;(d)Canny edge;(e)nearest point's distance from Canny edge;(f)head candidates(red cross)

Maximise posterior probability. Randomly selected some points from head candidates to build the segmentation model $\theta = \{n, \{O_1, ..., O_n\}\}$. θ_t represents the t step, the steps follow as:

1. Randomly select one pedestrian model $O_i \in \theta_t$ to remove, then $\theta' = \{n-1, \{O_1, ..., O_{n-1}\}\}$. if $P(\theta'|I)$ is greater than $P(\theta_t|I)$, $\theta_{t+1} = \theta'$, otherwise $\theta_{t+1} = \theta_t$.

2. Randomly select one point from the remaining head candidate and build a pedestrian model O. Then add the new one to θ_t, then $\theta' = \{n + 1, \{O_1, ..., O_{n+1}\}\}$, if $P(\theta'|I)$ is greater than $P(\theta_t|I)$, $\theta_{t+1} = \theta'$, otherwise $\theta_{t+1} = \theta_t$.

3. Randomly select one pedestrian model $O_i \in \theta_t$, moving its position p_i to the nearest head point , then $\theta' = \{n, \{O_1, ..., O_n\}\}$. if $P(\theta'|I)$ is greater than $P(\theta_t|I)$, $\theta_{t+1} = \theta'$, otherwise $\theta_{t+1} = \theta_t$.

Iterating the above three steps to modify the segmentation model, the first and second steps are significant changes to the segmentation model. To detect the pedestrian which have not been included in model by directly increasing the pedestrian model, and then remove the mismatch and the over fitting for foreground region by reducing the segmentation model. Due to the inaccuracy of the head detection, there is a high possibility that the pedestrians' shoulders and necks are detected to be the head. The added pedestrian model in the second step would be in wrong position in a high possibility. So we should transfer these points to the right position by moving the position of the pedestrian model to the nearest point in the last step.

When iterations are taken about five to ten times as many as the number of pedestrian's head points, the posterior probability would become to the maximum. End the iteration, and achieve the accurate segmentation results.

Experiments and analysis

We have tested our algorithm on PETS2009 video datasets, which has some videos which have pedestrian crowd in the scene. Select three video from datasets, their length are 412, 220, 200 frames respectively and their degree of occlusion are different. The degree of occlusion are summarized in Fig.4, the abscissa is degree of occlusion i.e. the proportion of the part that couldn't be observed in the total pedestrian, the ordinate is the ratio of occlusion pedestrian number to total number in video. As the figure shown, the degree of occlusion of three videos are gradually increasingly.
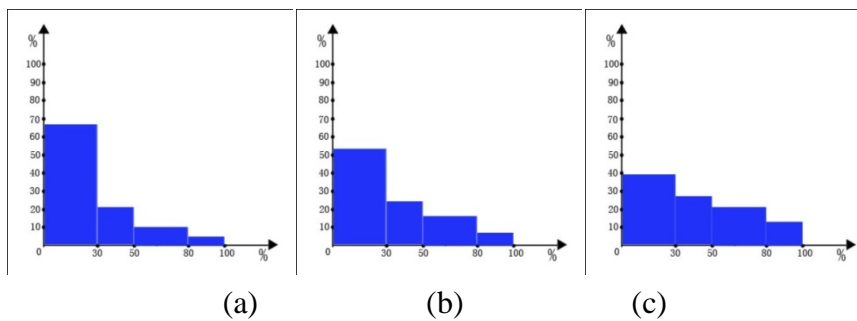

(a)   (b)   (c)

Fig. 4 Three videos of occlusion histogram.(a)video 1;(b)video 2;(c)video 3.

In our experiments, the parameters are fixed as the following: $\alpha_{10} = 1.0, \alpha_{01} = 1.5, \alpha_{re} = 0.3$. At first an image that was selected from video 2 is processed as shown Fig.5(a), this image contains 26 pedestrians, six of them were occluded half, as well as four of them couldn't be observed almost. We have explicitly marked the errors on the image when they occur: the red arrows are indicted the pedestrians which have not been detected in image and the red cross are indicted the wrong position which aren't the pedestrians. Almost all pedestrians are detected, and pedestrians which aren't detected could only been observed part of the edge contour. The wrong detection in the centre of the image segmentation is caused by the noise of foreground segmentation errors. Fig.5(b) is the curve of posterior probability $P(\theta|I)$ at each iteration. We compare three ways how to initialize the segmentation model, randomly select 40 points from head candidates, selection of all the head points and null initial state. There are 415 head points that detected in the image, although showing some difference at the beginning, they could converge to the maximum before 2000 iterations. As graph shown, all the initial method could achieve the maximum posterior probability, but the speed of convergence are affected by the initial state, the method that randomly select 40 points from head candidates could converge to the maximum quickly.

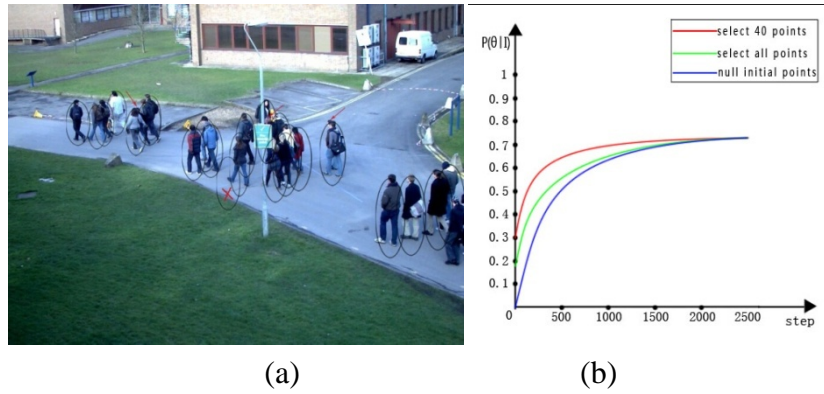<p align="center">(a)   (b)</p>

Fig.5 Detection result of one image.(a)The segmentation result;(b)The graph of three initial method.

The results of the evaluation of the three videos are summarized in Tab.1. Comparing the results, almost all of the pedestrians had been detected, and the pedestrians which hadn't been detected are occluded very severely. The degree of occlusion of one-third of pedestrians in video three are above fifty percent, some of them only could be observed the contour of head or torso and others are known by observing adjacent frames. Experimental results show that for these severely occluded pedestrians can effectively identify, which could not be identified are mostly whose whole body couldn't be directly observed in a single frame.

<p align="center">Tab.1 Segmentation result of three videos</p>

	Video 1	Video 2	Video 3
Frames	412	220	200
Pedestrians	4644	4816	4774
Correct detections	4411	4264	3925
False detections	137	157	173
Missed detections	233	552	849
Detection rate	94.98%	88.54%	82.22%
Missed detections of below half occlusion	34	76	141
Missed rate of below half occlusion	14.59%	13.77%	16.6%

Conclusions

To solve the problem which are difficult to detect one pedestrian from each other in a pedestrians crowd scene, this paper have presented an approach which use Bayesian formulation convert the problem to maximise posterior probability. At first pedestrian a head detection have been used to check the head position in the image, establish the segmentation model through these head positions. Foreground region could be matched by the segmentation model, and then iterate through changing segmentation model. The segmentation results could be achieved when posterior probability converges to maximum. By comparing the experiments on three videos whose pedestrians' occlusion degree are different from each other, the results shown that this algorithm can accurately detect the pedestrians whose occlusion degree less than fifty percent and the severely occluded pedestrians could be identified partially.

Acknowledgements

References

[1] Fieguth, P and Terzopoulos,D. Colo-based tracking of heads and other mobile objects at video frame rates. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 1997. 21-27.

[2] Gavrila D M,A Bayesian, Exemplar pedestrian detection: survey and experiments. IEEE Pattern Analysis and Machine Intelligence,2009,29(8);1-14.

[3] Papageorgiou C,Evgeniou T,Poggio T. A Trainable Pedestrian Detection System. Proceedings of IEEE Intelligent Vehicles Symposium, 1998,241-246.

[4] Dalal N, TriggsB. Histograms of oriented gradients for human detection. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA,2005,2:886 一 893

[5] Stauffer C, Grimson WEL. Adaptive background mixture models for real-time tracking. IEEE International Conference on Computer Vision and Pattern Recognition 1999;2:246–52.

[6] C. Stauffer, W.E.L Grimson. Adaptive Background Mixture Models for Real-Time Tracking. Proc IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado,1999,pp. 246-252.

[7] S. Haritaoglu, D. Harwood and L. S. Davis, W4: Real-Time Surveillance of People and Their Activities. IEEE Trans. on PAMI, Vol. 22, No.8, 2000.

[8] Hu W, Hu M, ZhouX. Principal axis-based correspondence between multiple cameras for people tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4) : 663-671.

[9] A. Mittal and L. S. Davis. M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene Using Region-Based Stereo. Proc. of European Conf. on Computer Vision,LNCS 2350, pp. 18-33, 2002.

[10] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground-background segmentation using codebook model. Real-Time Imaging, vol. 11 ,2005, pp. 172-185.