

PAPER • OPEN ACCESS

Target Pedestrian Tracking Algorithm Based on Occlusion Scene

To cite this article: Guocai Du 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **533** 012059

View the [article online](#) for updates and enhancements.

Target Pedestrian Tracking Algorithm Based on Occlusion Scene

Guocai Du

Department of Computer Science and Engineering, Yunnan University, 650000
Kunming, China

495139052@qq.com

Abstract. The traditional incremental learning tracking algorithm tends to drift or lose track when the pedestrian is occluded and the background is similar to the pedestrian target. This paper proposes a pedestrian tracking algorithm based on improved incremental learning. According to the difference between the central position of the current observation model and that of the previous one, the reliability of the observation model can be judged, and then the forgetting factor can be adjusted, so that the observation model can adapt to the changes of pedestrian targets. The experimental results show that the improved algorithm can track the pedestrian target accurately and effectively in the occlusion scene.

1. Introduction

Pedestrian tracking is one hot research topic in computer vision field, its main task is to track the target in image sequence feature detection, to further achieve semantic layer analysis (action recognition, scene recognition, etc). At present, pedestrian tracking can be divided into three categories according to the design of the models. The tracking methods based on the generative model[1-4], the tracking methods based on the discriminant model[5-6], and the tracking methods based on the hybrid generative-discriminant model[7]. Generative model-based tracking methods focus on target self-modeling and learning. When the illumination intensity changes and the target moves quickly, these will seriously affect the appearance and background of the tracking target, and eventually lead to the failure or drift of the tracking target. Discriminant model-based tracking method emphasizes the separation of target and background, but when complex background appears, due to the similarity and ambiguity of background, it will eventually lead to the failure of tracking target. Literature[8] presents a method based on incremental learning, which can adapt to the changes of some tracking targets and the surrounding environment without drastic changes. However, the disadvantage of this algorithm is that it can not adapt to these complex environmental changes when the change of illumination intensity is the largest and the fast movement of the final target is encountered. Finally, the tracking target failed. In this paper, the incremental learning algorithm is used to improve the forgetting factors, and the stability and robustness of the method are verified by three video tests. The structure of this paper is as follows: the first section is the introduction, the second section is the detailed introduction of the algorithm, the third section is the experimental results of the algorithm, and the last section is the summary of the paper.



2. Our Algorithm

2.1. Singular Value Decomposition Based on Incremental Learning

The eigenvalues and eigenvectors of the training set can be obtained by the singular value decomposition of the training set image. The feature space constituted by feature vector contains the information of image in training set. In the process of target tracking, training set is constantly increasing. Therefore, when faced with the ever-increasing training set, it should be able to make some changes to the training set, so as to be able to learn the knowledge hidden in the new training set. The idea of incremental learning is to add the latest training set on the basis of the previous training set, and then calculate the mean value vector and singular value decomposition again to obtain the new eigenvalue and eigenvectors.

n frame image sequences $A = [I_1, \dots, I_n]$, m frame new image sequences $B = [I_{n+1}, I_{n+2}, \dots, I_{n+m}]$, let $C = [A \ B]$, \bar{I}_A, \bar{I}_B represent the mean of A and B , which $\bar{I}_A = \frac{1}{n} \sum_{i=1}^n I_i$, $\bar{I}_B = \frac{1}{m} \sum_{i=n+1}^{n+m} I_i$. First, image sequence A is centralized $A = [I_1 - \bar{I}_A, \dots, I_n - \bar{I}_A]$, and then singular value decomposition is performed to obtain the matrix U_A and diagonal matrix Σ_A . Then, the solving steps of matrix C and diagonal matrix are as follows:

1. Calculate the mean value of the new matrix $\bar{I}_C = \frac{fn}{fn+m} \bar{I}_A + \frac{m}{fn+m} \bar{I}_B$, where f is the forgetting factor.
2. Structure the matrix: $\hat{B} = \left[(I_{n+1} - \bar{I}_B), \dots, (I_{n+m} - \bar{I}_B), \sqrt{\frac{nm}{n+m}} (\bar{I}_B - \bar{I}_A) \right]$.
3. The QR decomposition is applied to the matrix $(\hat{B} - U_A U_A^T \hat{B})$ to obtain the orthogonal matrix Q and $R = \begin{bmatrix} f\Sigma & U_A^T \hat{B} \\ 0 & Q(\hat{B} - U_A U_A^T \hat{B}) \end{bmatrix}$.
4. Compute the singular value decomposition of R : $R \stackrel{SVD}{=} \tilde{U} \tilde{\Sigma} \tilde{V}^T \approx \tilde{U}_k \tilde{\Sigma}_k \tilde{V}_k^T$. Retain the first k singular values and update the mean $\bar{I}_A = \bar{I}_C$.
5. Finally $U_C = [U_A \ Q] \tilde{U}_k$ and $\Sigma_C = \tilde{\Sigma}_k$.

2.2. Target Tracking Based on Incremental Learning

The problem with video tracing can be thought of as reasoning in a markov model with hidden state variables, the state variables X_t describe the target affine motion parameters and position at time t . Given a series of observation images $I_t = \{I_1, \dots, I_n\}$, the purpose is to estimate the value of hidden state variables X_t . According to Bias theorem and Markoff model[9], the formula can be obtained as follows:

$$p(X_t | I_t) \propto p(I_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | I_{t-1}) dX_{t-1} \quad (1)$$

The motion of a tracking target in a continuous frame can be represented by an affine change. When at time t , the target state is composed of six affine parameters $X_t = (x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t)$. Where respectively represent x, y translation, rotation Angle, proportion, width and height ratio of coordinates, and tilt direction at the time t . Each parameter in the equation follows the gaussian distribution and is independent from each other. The model is established by Brown motion, and the system state transition model[10] is as follows:

$$p(X_t|X_{t-1}) = N(X_t; X_{t-1}, \Psi) \quad (2)$$

Where Ψ is the covariance matrix with corresponding changes of affine parameters, the diagonal elements of this matrix are the variance of each affine parameter.

Given the image block I_t , it is predicated by X_t , the hypothesis I_t is generated by the feature subspace of the target object, and the probability of sample generation is inversely proportional to the reference point μ of the subspace, and inversely proportional to the distance d between the sample points. The distance between the reference point of the feature subspace and the sample is d , and d can be decomposed into two parts, namely the distance d_i from the sample to the feature subspace, and the distance d_w from the observation sample to the feature subspace. According to the gaussian distribution[11], the probability of generating the sample of the subspace is as follows:

$$p_{d_i}(I_t|X_t) = N(I_t; \mu, UU^T + \varepsilon I) \quad (3)$$

where I is the identity matrix, μ is the mean of the image's subspace, εI is the Gauss noise added to the observation process, U represents the eigenvector of the subspace after the t time is added to the new image.

According to the markov distance[12], the probability of affine sample is expressed as follows:

$$p_{d_w}(I_t|X_t) = N(I_t; \mu, U\Sigma^{-2}U^T) \quad (4)$$

Where Σ is the singular value matrix corresponding to U .

According to the above, the similar function of observation image and feature space can be deduced[13]. The formula is as follows:

$$p(I_t|X_t) = p_{d_i}(I_t|X_t)p_{d_w}(I_t|X_t) = N(I_t; \mu, UU^T + \varepsilon I)N(I_t; \mu, U\Sigma^{-2}U^T) \quad (5)$$

2.3. Adaptive Adjusted Forgetting Factor

Incremental learning algorithm is with the addition of new data, will continue to update the eigenvalue and eigenvector, used to express the appearance of the pedestrian targets, finally can adapt to the target part of the change of appearance, but if the pedestrian target or background under the condition of relatively complex, the incremental learning algorithm can't deal with these problems, eventually leading to the pedestrian tracking failure. There are some improvements to incremental learning algorithm, such as target tracking by sparse representation, which only uses a small number of samples to describe the shape of the tracking object. If there are very similar pedestrian targets and backgrounds, the tracking results will also fail. In this paper, a new idea is proposed. According to the difference between the central position of the current observation model and the previous frame, the magnitude of the forgetting factor is finally adjusted. It is necessary to adjust the proportion of the observation value, increase the proportion of the new observation value and reduce the proportion of the former observation value, so that the past samples can be forgotten to adapt to the current environmental changes, so the forgetting factor can be increased. The difference of the center position is higher than that of the novel observation model. The proportion of the former observation value and the new observation value is maintained. This can improve the accuracy of the observation model in tracking the target and reduce the value of the forgetting factor.

The central position of the target $C_t^n = (x_t^n, y_t^n)$ is calculated according to the current observation model, the observation model of the previous frame calculates the central position of the target $C_t^o = (x_t^o, y_t^o)$, the maximum difference d_t can be calculated according to formula (6). and the effect of tracking target can be judged according to d_t . Finally, the size of the forgetting factor can be adjusted according to formula (7), where E_m and E_M represent the thresholds of the accepted and rejected subspace models.

It was found that the forgetting factor was the most robust between 0.5 and 0.95, so parameters were set as follows: $E_M = 30$, $E_m = 10$, $FF_M = 0.95$, $FF_m = 0.5$.

$$d_t = \max(|x_t^n - x_t^o|, |y_t^n - y_t^o|) \quad (6)$$

$$f = \begin{cases} FF_M, & d_t < E_m \\ FF_M - \frac{(FF_M - FF_m)}{(E_M - E_m)} d_t, & E_m < d_t < E_M \\ FF_m, & d_t > E_M \end{cases} \quad (7)$$

2.4. Summary of the Algorithm

The main steps of the algorithm described above are as follows:

1. Initialize the relevant parameters and the size of the tracking window, and calculate the U and Σ in the tracking image subspace.
2. According to formula (2), the particle set generated by the state transfer function is $x_t^i \sim P(X_t | X_{t-1}^i)$.
3. According to the particle set generated in step 2 and formula (5), the similarity function is calculated and the particle weight is calculated by the weight formula of particle filter.
4. The weight is normalized, and if the effective value of the particle is less than the threshold value, the resampling is performed.
5. Locate the location of the pedestrian target to be tracked according to the particle weight calculated in steps 3 and 4.
6. The increment singular value decomposition proposed in 2.1 and the dynamically adjusted forgetting factor proposed in 2.3 updated U and Σ .
7. Read the next frame and return to step 2, otherwise, until the end of the track.

3. Experiment Results

In order to verify the reliability of the proposed algorithm, several experiments were carried out on pedestrian video. The CPU is Intel (R) Core (TM) i5-7500 CPU@3.40GHz, and the memory is 8GB. The algorithm in this paper was verified based on Matlab R2016a, and three groups of pedestrian video were used for experiment. The total frame number of video was 330, 256 and 208, including simple background, complex background and target occlusion.

3.1. Qualitative Comparison

Fig.1, the target pedestrian (marked in red boxes) has a zebra crossing at about 330 frames, but the algorithm in this paper is still capable of real-time tracking. Fig.2 shows that the algorithm in this paper can effectively track the target pedestrian when multiple people are side by side and the pedestrian turns from left to right. At 166 frames and 252 frames, there are pedestrian, motorcycle, zebra crossing and tree on the right side as background interference, which can still effectively track the target. Fig.3, target pedestrians appeared in the process of tracking other pedestrians sheltered, the algorithm still can effectively track, also the 80th frame, appear on the left side of the pedestrian walking from left to right, completely hides are tracking target pedestrian, pedestrian and target the right there are some side by side to interference, eventually to be able to effectively track pedestrians.

3.2. Quantitative Comparison

This section is a quantitative analysis of the tracking results of the target pedestrians. The average pixel error Average Pixel Error (APE) and the average overlap rate Average Overlap Rate (AOR)[14] are used to analyze the results. It is measured by the area. Table 1 and Table 2 compare the tracking results of different algorithms in different videos. It can be seen that the proposed algorithm has better

tracking accuracy and robustness than other algorithms. According to the above experimental results, it has a small error on the whole and better robustness than the other two algorithms.

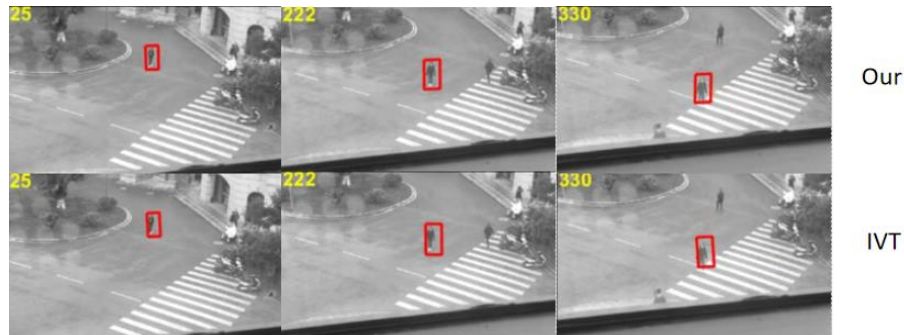


Figure 1. Tracking Results of A Target Pedestrian Against A Simple Background



Figure 2. Tracking Results of Target Pedestrians in Complex Background



Figure 3. Tracking Results of A Target Pedestrian in Occlusion

Table 1. Our Algorithm is Based on the APE Of 3 Test Videos.

Test video	Video 1	Video 2	Video 3
Our algorithm	4.6	11.5	8.0
IVT	10.0	20.2	25.1
LOT	11.0	18.6	16.7

Table 2. Our Algorithm is Based on the AOR in 3 Test Videos.

Test video	Video 1	Video 2	Video 3
Our algorithm	0.88	0.75	0.80
IVT	0.72	0.56	0.62
LOT	0.59	0.65	0.70

4. Conclusion

Traditional incremental learning tracking algorithms focus on the similarity between image samples, and do not take into account the past samples sometimes negligible, but sometimes must take into account the past samples, so the forgetting factor is set to a fixed value. Therefore, this paper improves the size of the forgetting factor by frame difference method, balances the proportion between new observation data and old observation data, and improves the accuracy of the subspace model in describing the appearance of the target at the current moment. Therefore, when the subspace model deviates from the current target appearance, the tracking algorithm in this paper can still maintain the tracking accuracy and adapt to the changes of the target appearance and illumination intensity. By adaptively adjusting the value of forgetting factor, the accuracy of the observation model in describing the appearance is enhanced. The algorithm can still maintain stability and robustness in the case of similar background or other pedestrian occlusion. The qualitative and quantitative results of experiments also verify the effectiveness of the proposed algorithm.

References

- [1] Tkachenko, Maksim, Lauw, IEEE Computer Society, **29(4)**, 771-783 (2017).
- [2] Feng Ping, Xu Chunyan, Elsevier B.V, **308**, 245-254 (2018).
- [3] Kawamoto Kazuhiko, Yonekawa Tatsuya, *IEEE Computer Society*, 711-714 (2012).
- [4] Bao C L, WuY, Linh H Bet, Proceeding soft the IEEE Conferenceon Computer Vision and Pattern Recognition, Rhode Island, 1830-1837 (2012).
- [5] Yang M, Zhang C X, Wu Y W, Electronic Proceedings of the 2013 IEEE International Conference on Multimedia and Expo Workshops, San Jose, 1-4 (2013).
- [6] Babenko B, Yang M H, Belongie S, IEEE Transactions on Pattern Analysis and Machine Intelligence, **33(8)**, 1619-1632 (2011).
- [7] Zhong Wei, Lu Hu-chuan, IEEE Transactions on Image Processing, **23(5)**, 2356-2368 (2014).
- [8] Ross D A, Lim J W, Lin R S, International Journal of Computer Vision, **77(1-3)**, 125-141 (2008).
- [9] MIHAYLOVA L, CARMIA Y, Digital signalprocessing, **25(1)**, 1-16 (2013).
- [10] MUTHUSWAMY K, IET computer vision, **9(3)**, 428—438 (2015).
- [11] Zhao Jing, Li Zhiyuan. Expert Systems with Applications, **37(12)**, 8910-8914 (2010).
- [12] Aughenbaugh, Jason Matthew, IEEE Transactions on Aerospace and Electronic Systems, **47(1)**, 503-523 (2011).
- [13] Xiao Jingjing, Stolkin Rustam, IEEE Sensors Journal, **16(8)**, 2639-2649 (2016).
- [14] Gao Shan, Ye Qixiang, IEEE Transactions on Image Processing, **18(6)**, 5575-5589 (2017).