Object Tracking Based on Corrected Background-Weighted Histogram Mean Shift and Kalman Filter

Yu Yang[1,a] Yongxing Jia[1,a] Chuanzhen Rong[1,a] Ying Zhu[1,a] Yuan Wang[1,a]
Zhenjun Yue[1,a] Zhenxing Gao[2,b]

[1]College of Communication Engineering, PLA University of Science and Technology, Nanjing 210007, China

[2]Tangshan Caofeidian Port Limited Liability Company, Tangshan 063200, China

[a]idsp@qq.com, [b]feixiangdejimo@163.com

Keywords: object tracking; mean shift; background information; Kalman filter

Abstract. The classical mean shift (MS) algorithm is the best color-based method for object tracking. However, in the real environment it presents some limitations, especially under the presence of noise, objects with partial and full occlusions in complex environments. In order to deal with these problems, this paper proposes a reliable object tracking algorithm using corrected background-weighted histogram (CBWH) and the Kalman filter (KF) based on the MS method. The experimental results show that the proposed method is superior to the traditional MS tracking in the following aspects: 1) it provides consistent object tracking throughout the video; 2) it is not influenced by the objects with partial and full occlusions; 3) it is less prone to the background clutter.

Introduction

In computer vision, tracking refers to the task of generating the trajectories of the moving objects by computing its motion in a sequence of images. Object motion recorded in the form of a trajectory commonly contains translational motion [1]. Numerous approaches have been dedicated to computing the translation of an object in consecutive frames, among which the mean shift (MS) method is one of the most common methods. The MS algorithm was originally proposed by Fukunaga in 1975 [2], and with the concept of kernel function, which became the practical foundation of MS [3]. Later, Comaniciu imported the MS to visual object tracking field satisfactorily [4]. The popularity of the MS method is due to its ease of implementation, real time response and robust tracking performance.

The MS is a typical deterministic method, which compares a model with current frame to find out the most promising region. However, the deterministic methods are hard to deal with the partial and full occlusions, in an appropriate manner, since the tracking algorithm is based on the previous investigated results. If the tracked object is lost or occluded completely, the deterministic searching methods will correspondingly fail. So, the MS iterative procedure needs to be improved. To eliminate this problem, a Kalman filter (KF) is used here. The present KF has been widely used for tracking in so many areas of control theory, signal processing, computer vision, and other related fields [5]. The KF estimates the state of dynamic system, even if the precise form of the system is completely unknown. This approach is so powerful in the sense that it supports estimations of the past, the present, and even the future states. As soon as the object is partially or fully overlapping with another one, the approach is superimposed to the algorithm.

The second limitation of the traditional MS method is prone to local minima when some of the target features present in the background. Therefore, in [4], Comaniciu *et al.* further proposed the background-weighted histogram (BWH) to decrease background interference in target representation. Such an idea is reasonable and intuitive, and some works have been proposed to follow this idea [6,7]. Unfortunately, all of them do not notice that the BWH transformation formula proposed in [4] is actually incorrect and it is proved in [8]. And the corrected background-weighted histogram (CBWH) is proposed,which can truly achieve what the original BWH method wants: reduce the interference of background in target localization.

This paper is organized as follows. Firstly, provides an overview of the classical MS algorithm and introduces the CBWH scheme in detail. Secondly, explains the formulation of KF and describes the proposed object tracking algorithm using the CBWH algorithm and KF. Finally, the experimental results and concluding remark are given.

The CBWH-MS Algorithm

Mean Shift Tracking: An overview. MS is a semi-automatic tracking method. The desired object is manually selected in the initial frame, and it is shown as a rectangle. The inside pixel location of the rectangles is presented as $\{x_i^*\}_{i=1,\ldots,n}$. The selected area is considered as the object model [9]. The color histogram of the object model is then calculated by:

$$\hat{q} = \{\hat{q}_u\}_{u=1,\cdots,m}; \ \hat{q}_u = C \sum_{i=1}^{n} k\left(\left\|x_i^*\right\|^2\right) \delta\left[b(x_i^*) - u\right] \tag{1}$$

where u represents the eigenvalue of the target model, $b(x_i^*)$ associates the pixel x_i^* to the histogram bin, $k(x)$ is an isotropic kernel function, and constant C is $C = 1/\sum_{i=1}^{n} k\left\|x_i^*\right\|^2$.

Similarly, the probability of the feature $u=1,2,\ldots,m$ in the target candidate model from the target candidate region centered at position y is given by:

$$\hat{p} = \{\hat{p}_u\}_{u=1,\cdots,m}; \ \hat{p}_u = C_h \sum_{i=1}^{n_k} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \delta\left[b(x_i) - u\right] \tag{2}$$

Based on the Bhattacharyya coefficient, a metric which aims to estimate the similarity between the target model and the candidate model is defined as:

$$\rho(y) = \sum_{n=1}^{m} \sqrt{\hat{p}_u(y) q_u} \tag{3}$$

More similar between the two density distribution, ρ is greater. Find the maximum of the density estimation to obtain the new location of y by using MS vector.

$$y_1 = \sum_{i=1}^{n_k} x_i \omega_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right) \bigg/ \sum_{i=1}^{n_k} \omega_i g\left(\left\|\frac{y_0 - x_i}{h}\right\|^2\right) \tag{4}$$

where

$$\omega_i = \sum_{u=1}^{m} \delta\left[b(x_i) - u\right] \sqrt{\frac{q_u}{p_u(y_0)}} \ ; \ g(x) = -k'(x) \tag{5}$$

The Corrected Background Weighted Histogram Scheme. In target tracking, the background information is often included in the detected target region. If the correlation between target and background is high, the localization accuracy of the object will be decreased. To reduce the interference of salient background features in target localization, a representation model of background features was proposed by Comaniciu et al. as background-weighted histogram (BWH) MS algorithm [4] to select discriminative features from the target region and the target candidate region.

In [4], the new target model is:

$$\hat{q}'_u = C' v_u \sum_{i=1}^{n} k\left(\left\|x_i^*\right\|^2\right) \delta\left[b(x_i^*) - u\right] \tag{6}$$

And the new target candidate model is:

$$\hat{p}'_u(y) = C'_h v_u \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)\delta[b(x_i) - u] \tag{7}$$

where

$$\left\{v_u = \min\left(\hat{o}^*/\hat{o}_u, 1\right)\right\}_{u=1,\cdots,m} \tag{8}$$

and $\{\hat{o}_u\}_{u=1,\cdots,m}$ $\left(with \sum_{i=1}^{m} \hat{o}_u = 1\right)$ is the background, which is three times the size of the target as suggested in [4]. $\hat{o}^*$ is the minimal non-zero value in $\{\hat{o}_u\}_{u=1,\cdots,m}$.

The weight of point x_i computed by the BWH-MS in the target candidate region can be derived by Eq. 5 that is

$$\omega'_i = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}'_u}{\hat{p}'_u(y)}}\delta[b(x_i) - u] \tag{9}$$

The above BWH-MS transformation aims to reduce the effects of prominent background features in the target candidate region on the target localization. However, in [8] prove that BWH-MS cannot achieve this goal, because it is equivalent to the usual target representation under the mean shift tracking framework. Then the corrected BWH (CBWH) MS algorithm is proposed.

In CBWH-MS, Eq. 8 is employed to transform only the target model but not the target candidate model. A new weight formula is defined as:

$$\omega''_i = \sqrt{\hat{q}'_{u'}/\hat{p}_{u'}(y)} \tag{10}$$

And simplify Eq. 10 as:

$$\omega''_i = \sqrt{v_{u'}\omega_i} \tag{11}$$

Eq. 11 clearly reflects the relationship between the weight calculated by using the usual target representation (i.e. ω_i) and the weight calculated by exploiting the background information (i.e. ω''_i). If the color of point i in the background region is prominent, the corresponding value of $v_{u'}$ is small. Hence in Eq. 11 this point's weight is decreased and its relevance for target localization is reduced. This will then speed up mean shift's convergence towards the salient features of the target. Note that if we do not use the background information, $v_{u'}$ will be 1 and ω''_i will degrade to ω_i with the usual target representation.

Object Tracking

The Kalman Filter (KF). The CBWH-MS algorithm is not well suited for tracking objects in the presence of occlusions as MS algorithm does. In order to improve this problem a KF is used. The KF [10] is a state estimation algorithm which compares state prediction against state measurements to get an accurate estimation of the true state.

Let the state prediction of the filter be $X = [x_c, y_c, \dot{x}_c, \dot{y}_c]$, where $\dot{x}_c$ and $\dot{y}_c$ represent the velocity in x_c and y_c, respectively. The discrete-time process model will be given by:

$$X_t = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} X_{t-1} + N(0, Q) \tag{12}$$

where Δt is the sampling time and $N(0,Q)$ represents a normal distribution for the model with a covariance given by Q. For initialization, the initial positions are set to the initial centers of the object (computed in the very first frame) and the velocities are arbitrarily initialized (e.g. to 1). The discrete-time measurement model is given by:

$$Y_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} X_t + N(0,R) \tag{13}$$

Where $N(0,R)$ is the zero-centered normal distribution corresponding to the uncertainty in the measurement, with R as the covariance matrix for the measurement.

The KF approach is applicable with the appropriate Δt, Q and R.

The Tracking Algorithm Based on MS with CBWH and KF. When there is a partial or full occlusion, the Bhattacharyya coefficient drops [11]. Therefore, one of the best methods for detecting this error of tracking is studying changes of the Bhattacharyya coefficient during the tracking.

Based on the above analyses, the occlusion of the target can be identified by Bhattacharyya coefficient and then the proposed tracking algorithm, i.e. the KF-CBWH-MS algorithm, can be developed. The implementation of the whole algorithm is summarized as follows:

1) Initialize the position y_0 of the target candidate region in the previous frame. Calculate the target model $\hat{q}$ by Eq. 1 and the background-weighted histogram $\{\hat{o}_u\}_{u=1,\cdots,m}$, and then compute $\{v_u\}_{u=1,\cdots,m}$ by Eq. 8 and the transformed target model $\hat{q}'$ by Eq. 6.

2) Let $k \leftarrow 0$.

3) Calculate the target candidate model $\hat{p}(y_0)$ using Eq. 2 in the current frame.

4) Calculate the weights according to Eq. 11.

5) Calculate the new position y_1 of the target candidate region using Eq. 4.

6) Let $d \leftarrow \|y_1 - y_0\|$, $y_0 \leftarrow y_1$, $k \leftarrow k+1$. Set the error threshold ε_1 (default value: 0.1), the maximum iteration number N.

If $d < \varepsilon_1$ or $k > N$ Stop and go to step 7;

Otherwise $k \leftarrow k+1$ and go to step 3.

7) If $\rho[\hat{p}(y_1),\hat{q}'] < \varepsilon_2$, estimate object position using KF by Eq. 13. Then go to step 3.

Experiments

This section evaluates the proposed KF-CBWH-MS algorithm in comparison with the MS algorithm [4] and the CBWH-MS algorithm [8]. Experiments are carried out on two sequences clipped from PETS2001 dataset in which the objects being occluded in complex background. And the programming environment is MATLAB 7.01. The resolution of video is 720×576 and frame rate is 25fps. In the experiments, the RGB color space was used. Each color band was equally divided into 16 bins (16×16×16).

The first clipped video (59 frames) is a car sequence where the object (a blue car) is partially occluded by a pine tree (frames: 14 - 47). Fig. 1 shows the tracking results of MS algorithm is not good, since the influence of the similar object features background and the occlusion. Like the MS algorithm the CBWH-MS algorithm diverges when occlusion occurs, but this algorithm tracks the car well before the occlusion occurs due to CBWH scheme. The proposed KF-CBWH-MS algorithm tracks the car well through out the sequence without divergency since the KF prediction.

(a) The MS algorithm

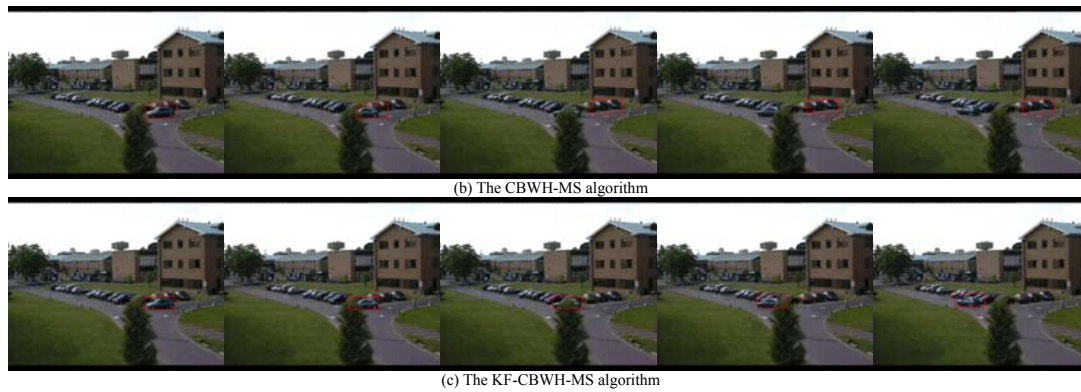(b) The CBWH-MS algorithm

(c) The KF-CBWH-MS algorithm

Fig. 1: Tracking results of the car sequence by different tracking algorithms. The frames 10, 20, 30, 40 and 55 are displayed.

The second experiment is on the same scene with a people sequence (168 frames), in where the target is fully occluded by a pine tree (frames: 62 - 113) until he gets out of it. Since the full occlusion, the tracking results are much more different. As can be seen in Fig. 2, only the proposed KF-CBWH-MS algorithm can effectively track the target over the whole sequence, but neither MS algorithm nor CBWH-MS algorithm. However, the CBWH-MS scheme still works better than the original MS algorithm before the occlusion occurs.

(a) The MS algorithm

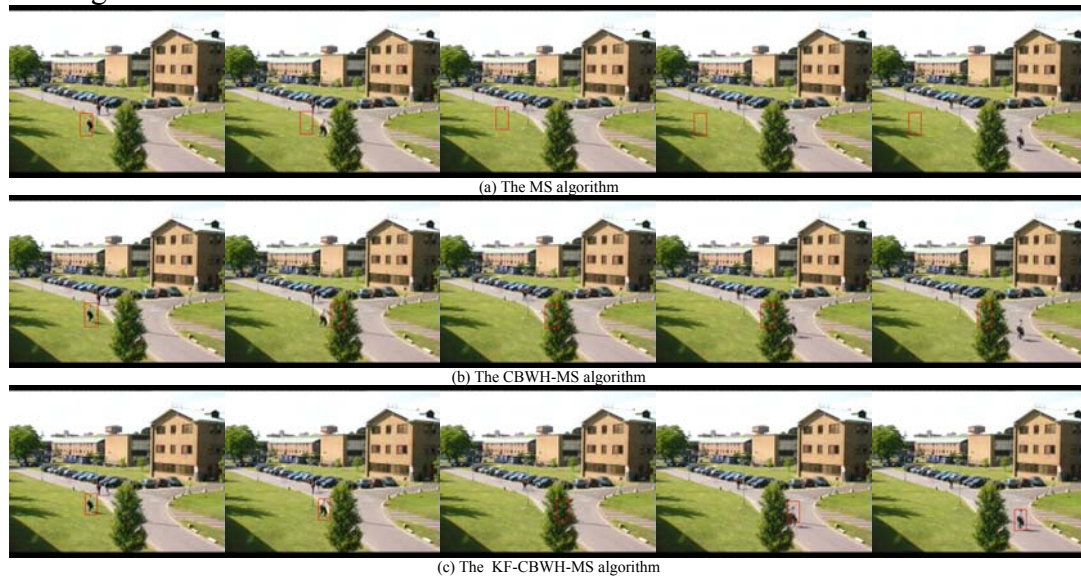(b) The CBWH-MS algorithm

(c) The KF-CBWH-MS algorithm

Fig. 2: Tracking results of the people sequence by different tracking algorithms. The frames 10, 46, 88, 118 and 138 are displayed.

Summary

The MS tracking algorithm works well when there is a lot of information in the frames, a notorious difference in contrast, and illumination without presence of noise. However, this algorithm performance is not unfortunately acceptable when facing severe or complete occlusions, especially in complex environment. The way to solve this problem, proposed in this paper, is a combination of CBWH and KF introduced to the MS framework. The proposed algorithm (i.e. KF-CBWH-MS) is employed to provide an optimum solution to object tracking problems. The CBWH could reduce the relevance of background information and improve the target localization. The KF yields to very good results when dealing with total occlusion due to the prediction of the target's motion based on the previous frames measurements. Experimental results show KF-CBWH-MS algorithm performs robustly in complex scenes where occlusion occurs.

References

[1] Alper Yilmaz, Object tracking by asymmetric kernel mean shift with automatic scale and orientation selection, IEEE Conf.. Computer Vision and Pattern Recognition, June 2007, pp.1-6.

[2] K. Fukunaga, L. D. Hostetler, The estimation of the gradient of a density function, with application in pattern recognition, IEEE Trans. Information Theory, 1975, 21 (1), pp.32-40.

[3] Y. Cheng, Mean-shift, mode seeking, and clustering, IEEE Transactions On pattern Analysis and Machine Intelligence,1995, 17 (8), pp. 790-799.

[4] D. Comaniciu, P. Meer, Kernel-based object tracking, IEEE Trans. Pattern Anal. Machine Intell., 2003, 25, (2), pp. 564-577.

[5] M. Pietikinen, A. Hadid, A. Zhao and T. Ahonen Local Binary Patterns for Still Images in Computer Vision using Local Binary Patterns, 2011, pp.13-43

[6] Jeyakar J., Babu R., and Ramakrishnan K. R, "Robust object tracking with background-weighted local kernels," Computer Vision and Image Understanding, 2009, 112, (3), pp. 296–309.

[7] Li L., and Feng Z., An efficient object tracking method based on adaptive nonparametric approach, Opto-Electronics Review, 2005, 13, (4), pp. 325-330.

[8] Jifeng Ning, Lei Zhang and Chengke Wu, Robust mean shift Tracking with corrected background-weighted histogram, IET Computer Vision. January 2012, vol. 6, pp.62- 69 .

[9] Motavalli S., A part image reconstruction system for reverse engineering of design modifications, J Manuf Syst, 1991,10,(5) ,pp.383-395.

[10] K. Nickels and S. Hutchinson, Estimating Uncertainty in SSD Based Feature Tracking, Image and Vision Computing, 2002, vol. 20, pp. 47-58.

[11] A. H. Mazinan and A. Amir-Latifi, A new algorithm to rigid and non-rigid object tracking in complex environments, Int. J. Adv. Manuf. Technol., 2012, pp. 51-60.