TELKOMNIKA, Vol. 11, No. 6, June 2013, pp. 3072 ~ 3079
e-ISSN: 2087-278X

Object Descriptor Combining Color and Detection

Yun Gao*, Guo-wu Yuan, Hao Zhou, Xue-jie Zhang
School of Information Science and Engineering, Yunnan University, Kunming, China
*Corresponding author, e-mail: gausegao@163.com

Abstract

Object descriptor has become one of the key factors for a robust and accurate tracker. In this paper, we propose an object descriptor combining color information and motion detection. A tracked object can be described by its hue histogram excluding the background pixels around the tracked object for restraining the disturbing of complex background environments. During the tracking process, we model the object descriptor by Gaussian Mixture Model for adapting the appearance variation of the tracked object. Tracking experiments in the frame of particle filter show that our proposed object descriptor can effectively improve the robustness and accuracy of object tracking under the situations of complex environments and appearance variations.

Keywords*: object descriptor, motion detection, color, complex environments, appearance variations*

1. Introduction

Object tracking can be considered as the process of selecting candidate objects and measuring likelihoods with object descriptor, which has been an essential component of several vision applications such as intelligent surveillance, human-computer interface, traffic control and vehicle navigation. Still, for a deforming and non-rigid object without getting restricted to particular model assumptions, robust and accurate tracking presents a major challenge [1].

Here we briefly describe the conventional tracking methods and their latent shortcomings. Mean-shift [2] is a non-parametric density gradient estimator to find the image window that is most similar to the object's color histogram in the current frame. Similar to Mean-shift, yet Cam-shift [3] can adapt the scale variation of tracked objects. The success of the two methods highly depends on the discriminating ability of the color histogram model. Tracking can also be considered as estimation of the state given all the measurements up to that moment, which commonly employs predictive filtering and uses the statistics of object's color and location in the distance computation while updating the object model by constant weights [4]. Kalman Filter [5] assumes the measurement noise to be linear Gaussian. And Particle Filter is based on Monte Carlo integration methods. The current density of the state (which can be location, size, speed, boundary [6], etc.) is represented by a set of random samples with associated weights and the new density is computed based on these samples and weights. Particle Filter has a broad application in the video tracking field, since it is not restricted by the linear Gaussian assumption.

Accuracy and effective object descriptor has become one of key factors for a success tracker. Object descriptor can be used to describe the tracked object and match the candidate regions. In the literature, many different descriptor, from aggregated statistics to appearance models, have been widely used to track objects, such as color [7], shape [1], texture [1], SIFT [8], and combination of color and SIFT[9]. Color Histogram [7] is a popular object model of nonparametric density, which can denote the color distribution state in a certain region. However, color histogram model can not discriminate the disturbing of complex environments (clutter, illumination, etc.) and appearance variations (view, deformation, etc.). Then, many improved methods were presented. Reference [8] proposed to track objects using SIFT features and Mean-shift, where Mean-shift makes color histogram as the object model. Reference [9] also proposed to track objects using SIFT and color histogram. Reference [10] proposed to track objects using color histogram and texture features. Reference [11] proposed a multi-cue method to track objects. Compared to the single color histogram, these methods have better tracking effects with appearance variations to a certain extent.

Received January 23, 2013; Revised March 26, 2013; Accepted April 15, 2013

However, via a large number of experiments, we find that most existed tracking methods use a rectangular region as the tracked object area. Thus, it is necessary to include some background pixels in the rectangular region for non-rigid objects. Under this circumstance, the object descriptor necessarily contain a certain background disturbing, using no matter what feature, color, SIFT, HOG, multi-cue feature, and so on. Inaccurate object descriptor also necessarily lead to unsuccessful tracking. As shown in Figure 1, we test the difference between the hue histogram of a candidate object excluding the background pixels around the tracked object and that including the background pixels. It is clear that two hue histogram are difference from the intensities in many bin indexes. Even if the tracked object self has not any variation, it is possible that a true candidate object can not match with the object descriptor in Figure 1(b) when the background environments around the tracked object are changed.

Motivated by the influence of the background pixels to the accurate descriptor, we propose an object descriptor combining color information and motion detection. The descriptor can exclude the background pixels around the tracked object and the candidate objects and restain the influence of complex environments to object tracking. It is worth to be noted that the color feature can be replaced by other common feature after the background pixels are excluded, such as texture, SIFT, color and SIFT, color and texture, HOG, and so on, that the motion detection method of Gaussian Mixture Model used in our descriptor can only adapt the situation with fixed cameras, and that it can also be replaced by other better detection methods for complex environments or dynamic scenes. For adapting the appearence variations of the tracked object itself with deforming, pose, illumination or view angle, it is intuitively reasonable that the object variations is modeled by Gaussian Mixture Model when the background pixels are excluded.

In the rest of this paper, we explain how to construct and model the object descriptor of the tracked object and measure the likelihood of a candidate region with the object descriptor, and give the tracking algorithm in the frame of particle filter with our proposed object descriptor in Section 2. Experimental results and analysis in both visual evaluation and quantitative evaluation are reported in Section 3. We conclude this paper in Section 4.

2. Object Descriptor

For accurately describing a given tracked object or a candidate region excluding the background pixels around the object, we first detect the motion regions in the given scenes, and then we can calculate the hue histogram of the tracked object or the candidate region only including the foreground pixels in the results of motion detection. During the process of object tracking, we model the object descriptor to adapt some unpredictable situations such as the appearance variations due to deforming, pose, illumination or view angle.

2.1. Construct and Model Object Descriptor

For our proposed object descriptor, both motion detection and modeling descriptor are realized by the method of Gaussian Mixture Model. For motion detection, Gaussian Mixture Model (abbreviated as GMM) [12] is a popular detection method, which can handle these situations of shadows, illumination change, repetitive motion, and so on. The idea of motion detection with GMM motivates us to model our object descriptor by GMM for restaining the factors of complex background.

GMM can build a background model for each pixel with video frame sequences and model the changing process of our descriptor combining hue histogram and motion detection during object tracking. Let $X = \{I_1, ..., I_t\}$ be a history sample for a pixel or a descriptor. We assume that the samples satisfy a Normal distribution $N(\mu, \delta^2)$, where μ and δ^2 can be initialized by Formula (1) and Formula (2).

$$\mu = E(X) = \sum_{k=1}^{t} I_k p_k = (\sum_{k=1}^{t} I_k)/t \tag{1}$$

$$\delta^2 = D(X) = E\{[X - E(X)]^2\} = E(X^2) - (E(X))^2 \tag{2}$$

For motion detection, GMM judges a foreground pixel by comparing a new pixel with every Gaussian model. The regions formed from foreground pixels signify motion regions. For modeling the descriptor, GMM judges a new object Gaussian model by comparing a with every Gaussian model. The probability densities of a new sample can be estimated by Formula (3) for each Gaussian distribution. In Formula (4), Φ denotes the aggregation according to the regulation $\varphi_i(I_t) \geq th$, i and m show the order and number of Gaussian. If Φ is equal to $\varnothing$, a new Gaussian is created by the new sample, otherwise, it belongs to the existed Gaussian corresponding the maximal element in Φ, $\varphi_{\max}$, as Formula (5).

$$\varphi_i(I_t) = exp(-(I_t - \mu_i)^2 / (2\delta_i^2)) / (\sqrt{2\pi}\delta_i), i = 1..m \tag{3}$$

$$\Phi = \{\varphi_i(I_t) : \varphi_i(I_t) \geq th, i = 1..m\} \tag{4}$$

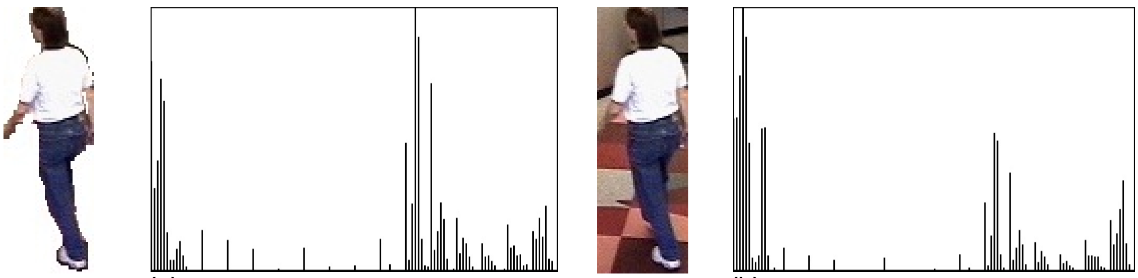
$$\varphi_{\max} = \max\{\varphi_i(I_t) : \varphi_i(I_t) \in \Phi, i = 1..m\} \tag{5}$$

The mean and variance of the matched Gaussian can be updated as Formula (6) and Formula (7), using an Expectation Maximization (EM) algorithm, which estimates an unknown parameter by an incomplete value sequence.

$$\mu_i(t+1) = (1-\alpha)\mu_i(t) + \alpha x_i(t) \tag{6}$$

$$\delta_i^2(t+1) = (1-\alpha)\delta_i^2(t) + \alpha(x_i(t) - \mu_i(t))(x_i(t) - \mu_i(t))^T \tag{7}$$

Only for the positions of the foreground pixels with the above GMM method, we can calculate its hue histogram, which is our object descriptor combining color infomation and motion detection, such as Figure 1(a). Furthermore, we can model our object descriptor by the above GMM method.



(a) (b)

Figure 1. The difference of the hue histograms under two descriptors for a candidate object: (a) Excluding the background pixels; (b) Including the background pixels.

2.2. Measuring Descriptor Likelihood

For measuring the likelihood of our descriptor, we assess a best candidate by comparing the hue histogram of a candidate with the object descriptor. We calculate Bhattacharyya Distance of the hue histogram of a predicted candidate H_p and that of the object descriptor H_M to measure their likelihood, as the Equation 8. The smaller the distance about a predicted candidate is, the more similar it is with the object descriptor.

$$d_B(H_p, H_M) = \sqrt{1 - \sum_i \frac{\sqrt{H_p(i) \cdot H_M(i)}}{\sqrt{\sum_i H_p^2(i) \cdot H_M^2(i)}}} \tag{8}$$

2.3 Tracking Flowchart with Our Descriptor

The tracking flowchart with our proposed object descriptor in the frame of particle filter is given in Figure 2. At the beginning of object tracking, we first detect motion object regions based on the method of GMM. After *p* video frames, detection regions are gradually stable, and here we set the number of *p* as 5. Then we manually select a tracked object region and construct our object descriptor based on the method in Section 2.1. When the number of video frame is more than *p*, we predict *q* candidate regions, and here we set the number of *q* as 100. According to the method of measuring likelihood in Section 2.2, we can obtain a best candidate as the current state of the tracked object. By the method of modeling our descriptor in Section 2.1, we can update our descriptor model. When any input frame can not be captured, the tracking process is finished.

3. Experimental Results and Analysis

In order to validate the effectiveness of our proposed object descriptor, we conduct experiments on more than 10 challenging video frame sequences from CAVIAR datasets, which involve severe appearance variations with drastic illumination changes and clutter backgrounds with other motion objects' disturbing or complex environments. The proposed descriptor is compared with the traditional color-based descriptor and our former object model. The comparison is performed from both visual evaluation and quantitative evaluation in Section 3.1 and 3.2, respectively. Experiments shows that our descriptor can restrain the disturbing of clutter backgrounds and adapt the situations of appearance variations.

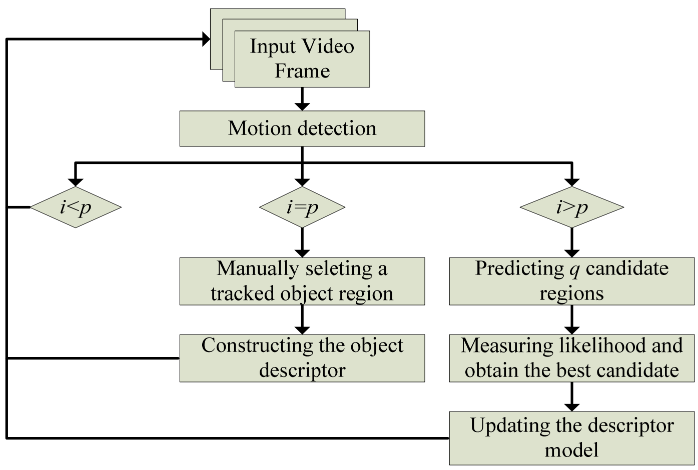


Figure 2. Tracking Flowchart with our Proposed Object Descriptor.

3.1. Visual Evaluation

The first test sequence is the *JPEGS* sequence with drastic illumination changes. There is a person going through a strong illuminant stretch before the glass window of a hall. During the moving process, both the surrounding environment and his appearance are variational with the influence of the strong illuminant. The tracking results of six representative frames with indices 6, 20, 31, 63, 88, 135 are shown in Figure 3 where rows 1 and 2 are for our proposed descriptor and the traditional color-based descriptor respectively. When the person starts into the strong illumination in Frame 20, both the background environment and his shirt abruptly change from dark gray to close white. For the traditional color-based descriptor, the object model is suddenly invalidated, and the tracked person is completely lost in Frame 31, which makes the continuous tracking failed. For the proposed descriptor, the disturbing of background variations are restrained due to describing the tracked person combining its hue histogram and motion detection, and the likelihood between the object model and the predicted particle is accurately measured due to modeling and updating our descriptor in time. The timely updating of the object descriptor ensures the continuous tracking. When the person leaves the strong illuminant region in Frame 88 and Frame 135, we can find the person is still successfully tracked.

The second test sequence is the *ThreePastShop2cor* sequence including more motion objects with drastic illumination changes and the disturbing of other motion object. There is a person going through a strong illuminant stretch before the glass window of a corridor. During the moving process, the appearance of her sweater is variational from gray to pink with the influence of the strong illuminant, and another motion person is near her. The tracking results of six representative frames with indices 1403, 1452, 1489, 1492, 1500, 1511 are shown in Figure 4 where rows 1, 2 and 3 are for our proposed descriptor, our former FFH descriptor and the traditional color-based descriptor respectively. For the traditional color-based descriptor, when she moves into the strong illumination area, the tracking results include more background pixels. For our former descriptor, due to referencing the area with foreground pixels in the stage of likelihood measurement, the tracking results are influenced by her near motion object and are partial to her near object from Frame 1489. For our proposed descriptor, the tracking results can better follow her avoiding the disturbing of her near person and illumination changes.

The third test sequence is the *IndoorGTTest2* sequence with complex backgrounds. There are a person going through a block of disturbing grid carpet region, where the carpet color is brightly contrary with the appearance of the tracked person and the starting background environment. The tracking results of six representative frames with indices 694, 709, 712, 719, 724, 732 are shown in Figure 5 where rows 1 and 2 are for our proposed descriptor and the traditional color-based descriptor respectively. For the traditional descriptor, when the tracked persons start through the carpet region in Frame 709, though they are influenced by the background carpet, they can still keep tracked. At the same time, we can also find that there are some better red particles have started to depart from the tracked object, and some red particles in Frame 709 are behind the pillar, whose color is close to that of the background in the started tracked region. Then, when the tracked person comes into the carpet region, the object models completely invalidate, all particles depart from them in Frame 724. For the proposed descriptor, tracking only foreground object pixels in the rectangular particle region effectively restrains the disturbing of the grid carpet, which makes the continuous tracking.

The fourth test sequence is the *OneStopMoveNoEnter1front* sequence with complex backgrounds. There are a person going through a corridor before a shop, where the shop environment is complex and is brightly contrary with the appearance of the tracked person. The tracking results of six representative frames with indices 468, 495, 498, 507, 522, 551 are shown in Figure 6 where rows 1 and 2 are for our proposed descriptor and the traditional color-based descriptor respectively. For the traditional descriptor, though they are influenced by the shop background in Frame 495, they can still keep tracked. At the same time, we can also find that there are some better red particles have started to depart from the tracked object, and some red particles are at his starting position, whose color is close to that of the background in the started tracked region. Then the object models completely invalidate, all particles depart from them in Frame 498. For the proposed descriptor, tracking only foreground object pixels in the rectangular particle region effectively restrains the disturbing of the shop background, which makes the continuous tracking.

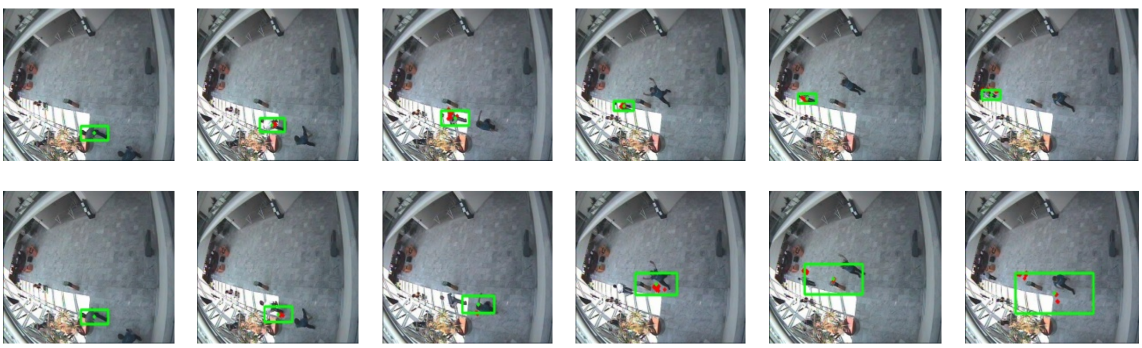


Figure 3. Tracking Results of the *JPEGS* Sequence. Rows 1 and 2 are the Results of our Descriptor and Traditional Color Feature Respectively

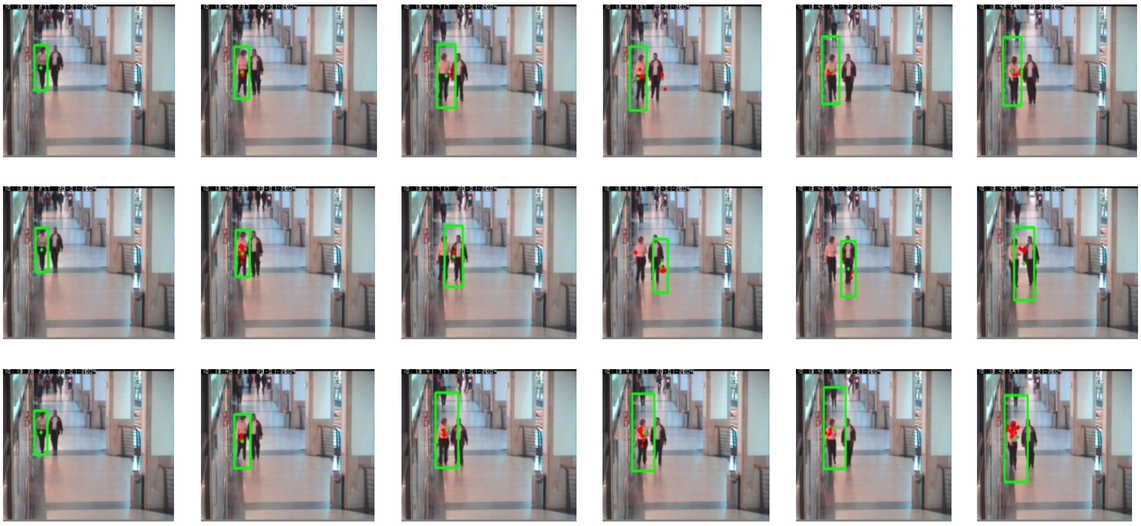


Figure 4. Tracking Results of the *ThreePastShop2cor* Sequence. Rows 1, 2 and 3 are the Results of our Descriptor, FHH and Traditional Color Feature Respectively

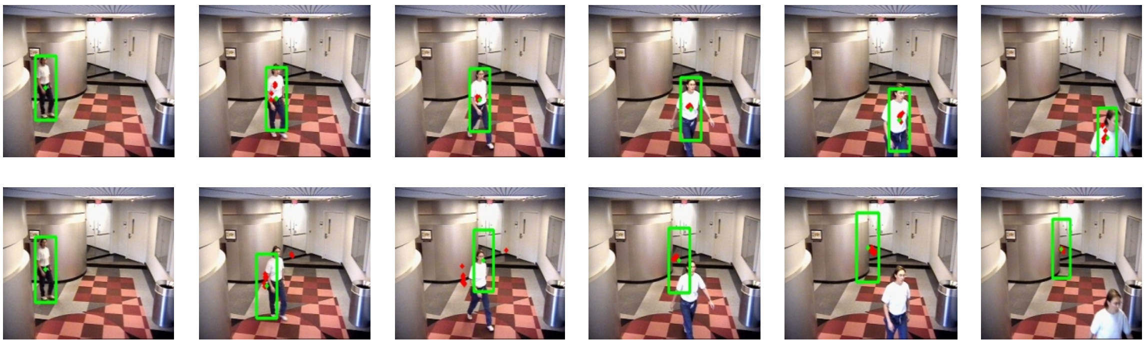


Figure 5. Tracking Results of the *IndoorGTTest2* Sequence. Rows 1 and 2 are the Results of our Descriptor and Traditional Color Feature Respectively

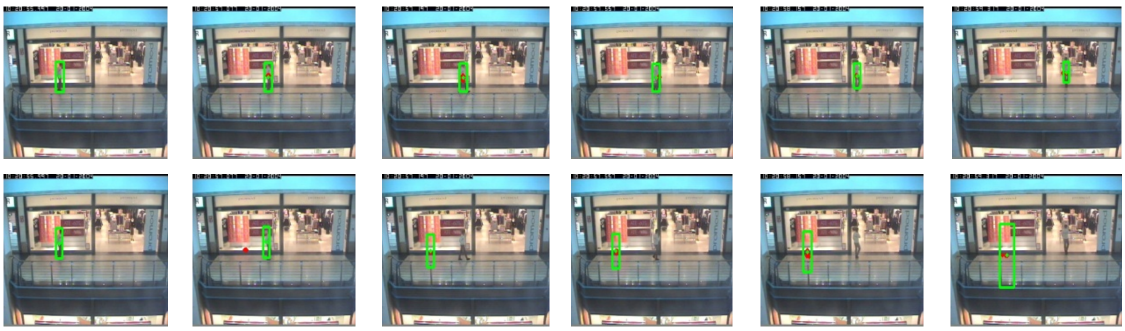


Figure 6. Tracking Results of the *OneStopMoveNoEnter1front* Sequence. Rows 1 and 2 are the Results of our Descriptor and Traditional Color Feature Respectively

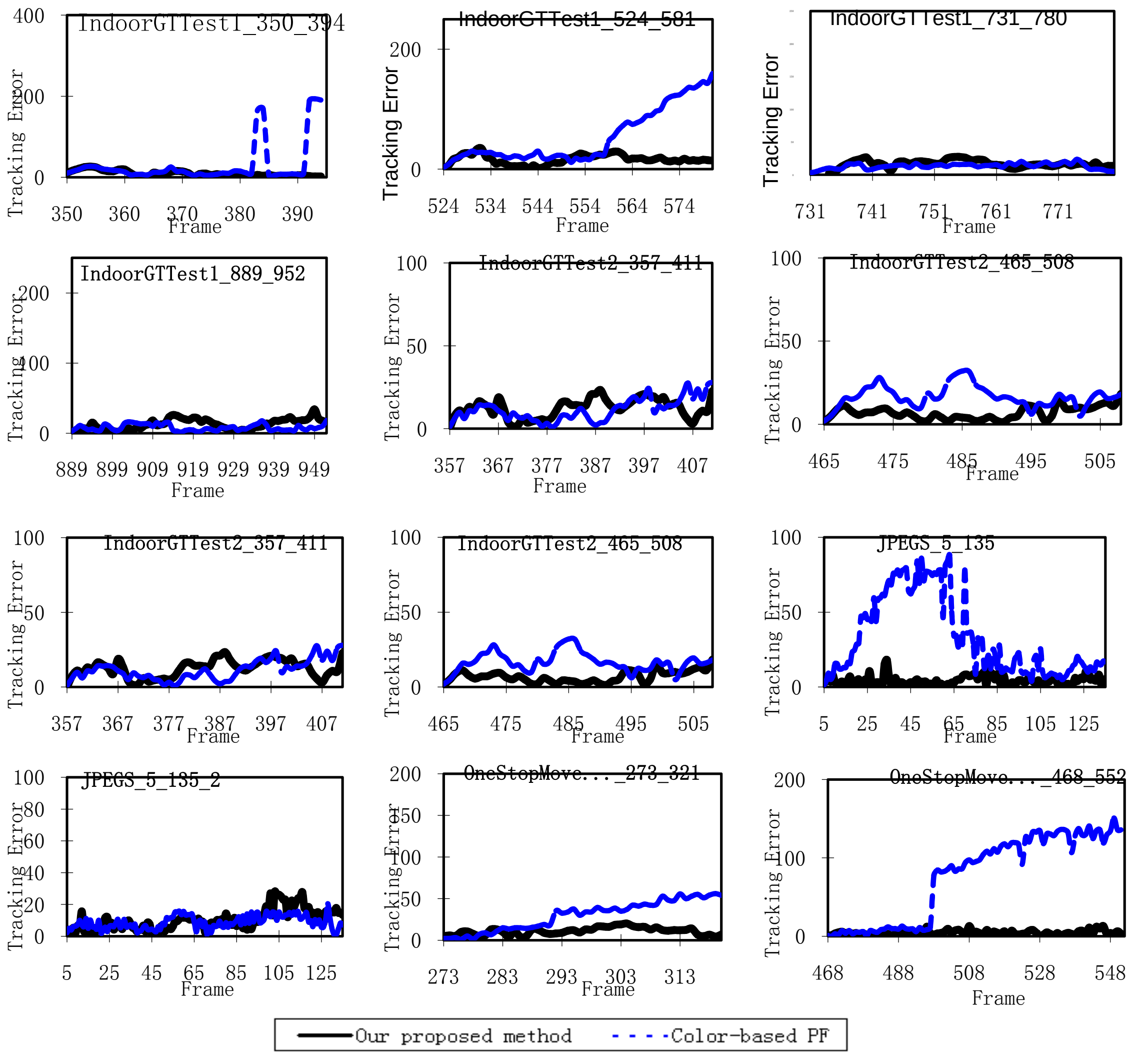
3.2. Quantitative Evaluation

In order to evaluate the overall performance of the proposed descriptor, we perform quantitative evaluation which is based on the tracking error e in each frame. We use a simple measure method, e is the offset of the center of the tracking result from the ground truth. The smaller e is, the better the tracking performance is obtained. In Figure 7, we present the tracking error curves of 12 video sequences, and two curves respectively show the errors of our proposed descriptor and those of the traditional color-based descriptor in each error comparision. In Table 1, we present the tracking error averages of 12 video sequences.

Table 1. The Average Tracking Errors (in Pixels)

Video Sequences	Our descriptor	Color-based PF
JPEGS_5_135	4.33	33.41
JPEGS_5_135_2	10.18	8.80
IndoorGTTest1_350_394	9.30	31.11
IndoorGTTest1_524_581	16.13	53.99
IndoorGTTest1_731_780	16.97	13.21
IndoorGTTest1_889_952	12.85	8.22
IndoorGTTest2_357_411	12.35	11.41
IndoorGTTest2_465_508	6.90	16.62
IndoorGTTest2_569_599	10.63	84.65
IndoorGTTest2_693_733	8.29	54.96
OneStopMoveNoEnter1front_273_321	10.89	30.02
OneStopMoveNoEnter1front_468_552	4.25	76.74

4. Conclusion



We have presented a object descriptor combining color information and motion detection. The proposed descriptor can exclude the disturbing of the background pixels around a tracked object. During the process of object tracking, we model the descriptor by the method of Gaussian Mixture Model to adapt the appearance variations. Experimental results and analysis from both visual evaluation and quantitative evaluation show that our proposed descriptor can be used to robustly and accurately track a given motion object.

Acknowledgements

The work is supported by National Natural Science Foundation of China (Grant No.: 61262067, 61163024 and 61170222), Scientific Research Foundation from department of education of Yunnan Province of China (Grant No.: 2011Y118 and 2012C105), and Application Foundation Project of Yunnan Province of China (Grant No.: 2011FB019).

References

[1] Yilmaz O, Javed, M Shah. *Object tracking: A survey*, ACM Comput. Surv., 2006; 38(4): 13+.
[2] P Zhao, Z Liu, W Cheng. Object Robust Tracking Based an Improved Adaptive Mean-Shift Method. *Advanced Information Technology in Education.* 2012; 126: 169-177.
[3] Xin Chen, Xiang Li, Hefeng Wu, Taisheng Qiu. *Real-time object tracking via CamShift-based robust framework*. Proc. of International Conference on Information Science and Technology. 2012: 527-530.
[4] C. Wren, A. Azarbayejani, T. Darell, and A. Pentland. Pfinder: *Real-time tracking of the human body*. IEEE Trans. Pattern Anal. Machine Intell., 1997, 19:780–785.
[5] G Bishop, G Welch. *An Introduction to the Kalman Filter*. SIGGRAPH. 2001; Course 8.
[6] B Xie, Y Gu. An Object Tracking Method Based on an Improved Particle Filter Algorithm. *Recent Advances in Computer Science and Information Engineering*. 2012; 128: 309-314.
[7] P Pérez, C Hue, J Vermaak, M Gangnet. *Color-Based Probabilistic Tracking*. Computer *Vision—ECCV*. 2002; 2350: 661-675.
[8] Huiyu Zhou, Yuan Yuan, Chunmei Shi. Object tracking using SIFT features and mean shift. *Computer Vision and Image Understanding*. 2009; 113(3): 345-352.
[9] S Zhang, H Yao, P Gao. *Robust object tracking combining color and scale invariant features*. Visual Communications and Image Processing, Proceedings of the SPIE. 2012; 7744: 8 pages.
[10] F Bousetouane, L Dib, H Snoussi. Improved mean shift integrating texture and color features for robust real time object tracking. *The Visual Computer.* 2012: 1-16.
[11] M Spengler, B Schiele. *Towards robust multi-cue integration for visual tracking*. Machine Vision and Applications. 2003; 14(1): 50-58.
[12] Li, YB, Jia, HJ, Li, A. A Moving Objects Detection Method Base on Improved GMM. *Advanced Materials Research.* 2011; 225-226(1-2): 637-641.