Fragments based tracking with adaptive multi-cue integration

Lichuan Gu*, Jianxiao Liu, Chengji Wang

School of Computer and Information, Anhui Agriculture University, Hefei Anhui, No.130 West Changjiang Road China, 230036

Received 1 January 2014, www.tsi.lv

Abstract

In this paper, we address the issue of part-based tracking by proposing a new fragments-based tracker with multi-cue integration. First the target template is divided into multiple fragments to get the contribution of each fragment to the possible positions of the target in the current frame, similarity measure is used in edge histogram and HSV histogram in every fragment, and the weights of cues integration are computed adaptively. Second we present a fragment template update mechanism with the discrimination between occlusions and appearance changes. The template is unchanged when the target is occluded and some fragments of template are updated in the case of appearance changes. In the experiments we use the indoor and outdoor test videos which contain the illumination changes, occlusions, and the appearance changes of targets. The experimental results show that our approach has strong robustness and high tracking accuracy.

Keywords: Target tracking, Multi-cue integration, Fragments tracking, Template update

1 Introduction

The object tracking in an accurate way is essential for applications like activity analysis, man-machine interaction and visual surveillance. However, in the actual process of tracking, target appearance changes (as shown in Figure 1 (a)) and occlusions (as shown in Figure 1 (b)) will affect the stability of the tracking algorithm, which leads to inaccurate target tracking and even target lost.
As a result, how to effectively deal with target appearance changes and occlusions has always been one of the difficulties in target tracking problems.

A common solution to the above problems is tracking target by the integration of multiple visual cues, namely each cue provides a likelihood value for possible positions of the target in the next frame, and these cues are integrated according to the likelihood values to locate the final output (Figure 2). In recent ten years, a few of tracking algorithms using multi-cue integration have been proposed, say, Birchfield proposed using intensity gradient and color histograms to track people head [1]. The disadvantage of this method is that during tracking the cues do not always provide reliable information about the target object, since the equal weights are assigned to each cue and the cues' weights are fixed in the whole video sequence. In order to solve this problem, Triesch and von der Malsburg proposed a new integration framework [2], which can take advantage of the uncertainty of each cue to adaptively adjust each cue contribution to the tracking results. Following this framework appeared many adaptive cues integration algorithms, as shown in [3-8]. These algorithms integrated a variety of cues, complemented each other and improved the accuracy of the tracking process to some extent.

(a)

(b)

FIGURE 1 Difficulties in tracking conditions: (a) target's variable appearance; (b) occlusions

Another solution to the above problems is to track moving targets based on fragments. In the tracking of

* Corresponding author - E-mail: glc@ahau.edu.cn

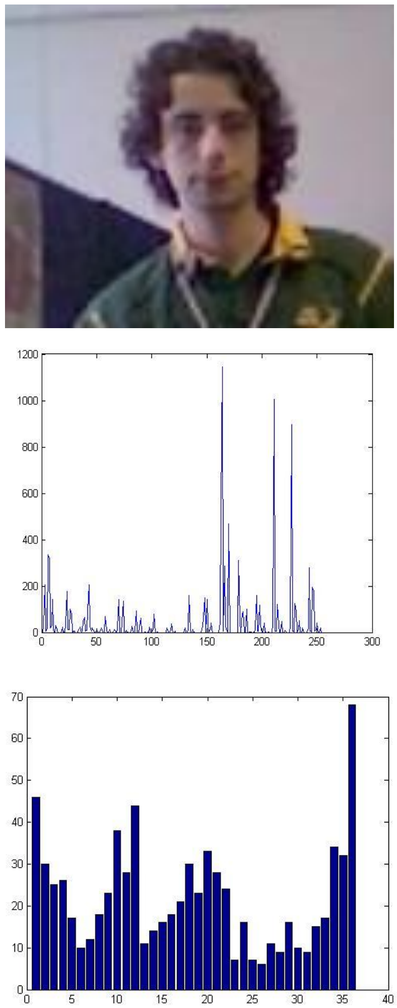

FIGURE 2 Cues based on different features: (a) Original image; (b) HSV histogram; (c) Edge histogram

human, for example, human is divided into the head, torso and limbs [9-11]. This approach generally required target model is known or a priori premise and is not suitable for generic target tracking. In order to solve this problem, some scholars put forward the general fragment tracking algorithms [12, 13]. This kind of method divided the target into different parts, but the division was arbitrary without considering any reference target model. In the process of tracking, this method weighted each fragment, and the contributions of different fragments are combined through statistical method to get the final output location of target. If there were target appearance changes or occlusions, the weight of corresponding fragment would be small and the impact on the overall goal would be small, thus the resolution of target would be improved.

The target tracking method based on template matching has attracted increasing attention [14-17]. Traditional image template matching based tracking method is still widely used because of its simple computations. This kind of method extracted some cues as a template which will be kept unchanged, and then looked for a region whose cues were most similar to the template in current frame. However, the target may be occluded during motion, and may also change its appearance due to own motion. In these cases the template needs to be updated online to track the target accurately. The existing main template update algorithms update template once the changes of grey are detected in target regions. These algorithms are different from one another only in that the update rate or updated components are not the same, without discrimination of appearance changes and occlusions. The comparison of aforementioned algorithms is given in Table 1.

This paper's contributions lie in: 1) our algorithm employs an adaptive cue integration scheme in each fragment of the target. Similarity measure in edge histogram and HSV histogram are used, the vote of each fragment contributes to the joint tracking result according to adaptive integration, and the ones having small values have little effect on the outcome. This allows us to achieve a better tracking accuracy in handling partial occlusions and appearance changes. 2) Our algorithm can update the fragment template online during the tracking process. In the small fragments the method detects the possible fragments where appearance changes or occlusions occur, and then takes corresponding update strategies according to the matching of the small fragments. The fragment template overcomes the shortcoming of the traditional template, which modelled the whole target without the spatial information.

The rest of this paper is organized as follows. The subsequent section describes the fragment tracking algorithm. Then an adaptive multi-cue integration model is given in Section 3. Section 4 presents a new tracking algorithm of multi-cue integration based on fragment, which is the main contribution of this paper. Section 5 is devoted to algorithm analysis, including experimental results on videos with different tracking conditions, and gives qualitative and quantitative analysis. The final section concludes this paper and suggests the future direction of our research.

2 Fragment tracking algorithm

Fragment tracking is a kind of target tracking algorithm based on cues proposed by Adam [12]. The basic idea is dividing target window into multiple sub-fragments and represents the target according to the joint sub-fragment histograms. This kind of algorithm has good anti-interference and anti-occlusion, and the computation has nothing to do with tracked target size. Detailed description is as follows:

The division pattern is shown in Figure 3. The target template is:

$$T=\left\{(h_{p_k}(b),\lambda_{p_k}\right\}_{p_k=p_1,\cdots,p_n} ,$$

p_k is the index of fragment *k*, the target is divided into *n* fragments, $h_{p_k}(b)$ and λ_{p_k} are a histogram and weight respectively, where $b=1,\ldots B$, and fragment weight equals 1 for every fragment.

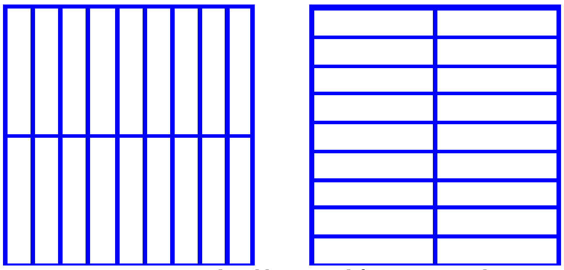

FIGURE 3 Vertical and horizontal fragment template

Fragment tracking is to find the most similar region to the template in current frame. The tracking principle is shown in Figure 4. Assume that a previous estimate of the position has been got, and each point in the neighbourhood of this estimate is searched, then every coordinate point in the searching window is a candidate for the target in current frame.

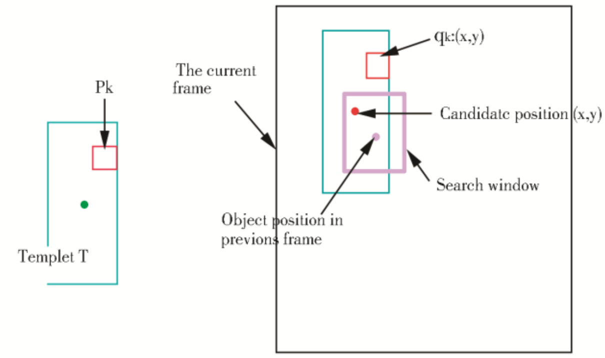

FIGURE 4 Fragment tracking principle

Let $p_k=(dx,dy,h,w)$ be a rectangular patch in the template, whose centre is displaced (dx,dy) from the template centre, and whose half width and height are w and h respectively. Let (x,y) be a hypothesized position of the object in the current frame. Then the patch p_k defines a corresponding rectangular patch in the image $q_{k;(x,y)}$ whose centre is at $(x+dx,y+dy)$ and whose half width and height are w and h, respectively. Calculating the histogram matching distances of each $q_{k:(x,y)}$ in the candidate region and the corresponding fragment in the template, we can get Formula (1) using the linear weighting of all the matching distances to represent the similarity between candidates and target template.

$$S(x,y)=\sum_{k=1}^{n} d(p_k,q_{k:(x,y)})\lambda_{p_k}, \quad (1)$$

where $S(x,y)$ is the similarity between the candidate at the location (x,y) and the template, and $d(p_k,q_{k:(x,y)})$ is the histogram matching distance between fragment p_k and fragment $q_{k:(x,y)}$, which is defined as

$$d(p_k,q_{k:(x,y)})=\sum_{b=1}^{B}\frac{(h_{p_k}(b)-h_{q_{k:(x,y)}}(b))^2}{h_{p_k}(b)+h_{q_{k:(x,y)}}(b)}. \quad (2)$$

The smaller $d(p_k,q_{k:(x,y)})$ means more similarity between the histograms. After (x,y) has traversed all the candidate points, we obtain the similarity between each candidate target and the current template. Then the current target location $(\hat{x},\hat{y})$ is defined as:

$$(\hat{x},\hat{y})=\arg\min_{(x,y)\in\Theta}(S(x,y)), \quad (3)$$

where Θ is the set of all the candidate locations. To distinguish easily between occlusions and appearance changes, after finding the target location in current frame according to the matching distance between fragment $q_{k:(\hat{x},\hat{y})}$ in the target region and the corresponding fragment p_k, we update the weight of p_k as

$$\lambda_{p_k}=\exp\left(\frac{-(d(p_k,q_{k:(\hat{x},\hat{y})}))}{\sigma_d^2}\right), \quad (4)$$

where σ_d is the variance of $d(p_k,q_{k:(\hat{x},\hat{y})})$.
For convenience $q_{k:(\hat{x},\hat{y})}$ is represented as q_k in the sequel.

3 Adaptive multi-cue integration

In complicated scene, it is difficult to obtain good tracking performance by single visual cue. To further improve tracking robustness, we need integrate other cues. Suppose F_j is a visual cue which has been selected and the fused cue is f_{fusion}, then f_{fusion} can be represented as:

$$\begin{cases} f_{fusion}=\sum_{j=1}^{n}\pi_j F_j \\ \sum_{j=1}^{n}\pi_j=1 \end{cases}, \quad (5)$$

where π_j is the weight of F_j. It is clear that different cues have different contributions to the ability of identifying the whole target due to the impact of the environment, so different cues should have different weights. Since there may be target's appearance changes or occlusions in real tracking, the weight of each cue in the each frame of video sequence should be updated dynamically.

There are two problems to be solved: one is how to evaluate the weights according to different contribution of each cue in different tracking environment, the other is how to update the weights in a steady and real-time way.

4 Proposed tracking algorithm

4.1 TRACKING PROCESS

This paper proposes a new method of fragments based tracking with multi-cue integration and template updating. The algorithm describes the target object by a template, and carries out tracking by searching for the image region in each frame of the tracking sequence with multiple cues similar to the template. This estimation process is done by splitting the target object into multiple arbitrary patches with each of them describing a different part of the target and accordingly providing the multi-cue information. The template is adaptively updated in accordance with whether occlusions or appearance changes occur to alleviate the template drift. The tracking schedule is shown in Figure 5.

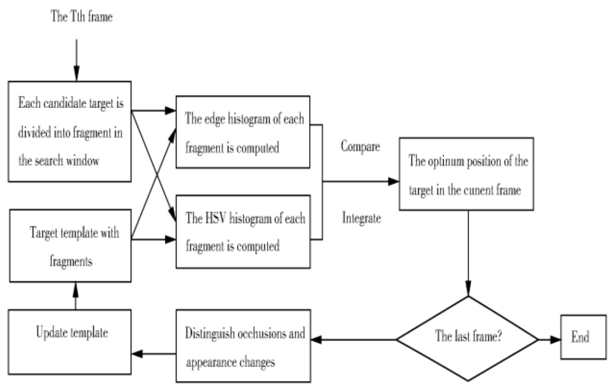

FIGURE 5 The flow chart of our algorithm

The steps of our algorithm are as follows.

Step1: Target template is divided into multiple fragments using the patch layouts given in Figure 2, and then each of these fragments is represented with histograms of different cues (colour, edge), and Integral Histogram data structure [18] is used to estimate them in a fast way.

Step2: The search window size is set to be $(2M+1)\times(2M+1)$ in current frame, namely the search radius is set to be M pixels from the previous target position (M =7 in this paper, and the search window is 15×15). Matching each template patch and the corresponding image patch is then carried out by comparing their histograms using the formula (13), and then we obtain the values $\{d(p_k, q_{k:(x,y)})\}$ by adaptive weight integration. Each $d(p_k, q_{k:(x,y)})$ indicates every fragment votes on the possible positions of the target. The detail is given in section 4.3.

Step3: The individual votes of the template patches are combined to obtain the joint tracking result. The weights of the template fragments are computed according to the formula (4), if $\lambda_{p_k} < 0.5$, the histograms of q_k changed a lot compared to corresponding p_k, that means there might be occlusions or appearance changes. We call such q_k as invalid fragment. The joint similarity $S(x,y)$ is defined in formula (6), and then we can get the optimum position of the target in current frame using formula (3):

$$S(x,y)=\sum_{k=1}^{n}\begin{cases} d(p_k,q_{k:(x,y)})\lambda_{p_k}, & \lambda_{p_k}\geq 0.5,\\ 0, & \lambda_{p_k}<0.5.\end{cases} \quad (6)$$

Step4: We judge each invalid fragment to see if it is caused by occlusion or appearance change. If there is occlusion, then go to step 2 directly without template update and start the next frame tracking; otherwise update the template and then go to step 2. The detail is described in section 4.4.

4.2 CUES HISTOGRAM COMPUTATION

In complicated scene, it is difficult to obtain good tracking performance by single visual cue. To improve tracking robustness, we need integrate other cues. It is very important to select the strong discriminate cues. This paper uses the integration of colour and edge cues to construct likelihood function for the purpose of scattering the impacts of all kinds of changes.

Colour is a major cue to describe the target. There have been many studies about the description of target colour cues before. Ref. [19] proposed a colour histogram to describe the target colour cues, which has simple calculation and fast processing, and is relatively robust in solving problems such as partial occlusion, rotation, etc. HSV histogram is employed to extract colour cue in each fragment in this paper. In our experiments, the Hue, Saturation and Value are quantified to be 16, 4, and 4. The three colour components are combined into one dimensional cue according to the quantitative level:

$$C=16H+4S+V, \quad (7)$$

where $C\in[0,1,\cdots,255]$, so we can get 256 bins HSV histogram.

Edge is another commonly used cue to describe the target. This paper chooses gradient direction histogram to describe the edge cue of the target [19]. The calculation formulas are as follows:

Horizontal gradient

$$G_x(x,y)=f(x+1,y)-f(x-1,y), \quad (8)$$

Vertical gradient

$$G_y(x, y) = f(x, y+1) - f(x, y-1), \quad (9)$$

Gradient magnitude

$$m(x, y) = \sqrt{G_x^2(x, y) + G_y^2(x, y)}, \quad (10)$$

Gradient direction

$$\theta(x, y) = \arctan(G_y(x, y) / G_x(x, y)), \quad (11)$$

where $\theta(x, y) \in [-\pi/2, \pi/2]$.

θ(x, y) should be mapped into $0 \sim 2\pi$.

$$\sigma(x, y) = \begin{cases} \pi + \theta(x, y) & G_x < 0 \\ 2\pi + \theta(x, y) & G_x > 0 \text{ and } G_y < 0 \\ \theta(x, y) & G_x \geq 0 \text{ and } G_y \leq 0 \end{cases} . \quad (12)$$

$[0, 2\pi)$ is divided into k bins with the width of each bin being $2\pi / k$ (k=9 in the paper). Each bin value is got by counting the pixels number in it and we can obtain the edge histogram.

4.3 ADAPTIVE WEIGHT INTEGRATION

In order to adapt to the environment and the target appearance changes, the weight of each feature in each frame must be updated in real time. Ref. [21] proposed Kolmogorov-Smirnov similarity measure criteria, K-S distance for short, to calculate the degree of similarity between the candidate object and target template. Its value represents the difference of two histograms; the smaller K-S value means smaller difference and higher similarity. Based on K-S distance, the matching distance for one cue is defined as follows:

$$d(p_k^j, q_{k:(x,y)}^j) = \sum_{b=1}^{B} \frac{(h_{p_k^j}(b) - h_{q_{k:(x,y)}^j}(b))^2}{h_{p_k^j}(b) + h_{q_{k:(x,y)}^j}(b)} . \quad (13)$$

$d(p_k^j, q_{k:(x,y)}^j)$ is the histogram matching distance for j cue between fragment p_k and fragment $q_{k:(x,y)}$. We can change the weight of corresponding cue by formula (14), so as to automatically change the influence of different cues to global functions. An adaptive weight integration formula is defined as in (15):

$$\omega_{q_{k:(x,y)}}^j = \frac{d(p_k^j, q_{k:(x,y)}^j)}{\sum_{j=1}^{n} d(p_k^j, q_{k:(x,y)}^j)} , \quad (14)$$

$$d(p_k, q_{k:(x,y)}) = \sum_{j=1}^{n} \omega_{q_{k:(x,y)}}^j d(p_k^j, q_{k:(x,y)}^j) . \quad (15)$$

It can be seen that automatically changing the different cues weights can better reflect the contributions of cues to the result. In experiments j=1, 2, colour is the first cue, edge is the second, $d(p_k^1, q_{k:(x,y)}^1)$ and $d(p_k^2, q_{k:(x,y)}^2)$ are the histograms matching distances for colour and edge cues between fragment p_k and fragment $q_{k:(x,y)}$ respectively. When the object begins to enter the occlusion or changes appearance, the edge will change greatly, the weight of $d(p_k^2, q_{k:(x,y)}^2)$ is larger, the weight of $d(p_k^1, q_{k:(x,y)}^1)$ is smaller, and then the contribution of edge cue to $d(p_k, q_{k:(x,y)})$ is larger. Therefore, this adaptive integration method can better reflect actual situation during tracking.

4.4 TEMPLATE ONLINE UPDATE

In order to ensure real-time tracking effect, this paper tracks the target by dividing the target into fragments, judges states of fragments, and adjusts update strategy according to whether there is occlusion or appearance change. The distance between the fragment and the template will be increased because of occlusion or appearance change. The template does not have to be updated in the case of occlusion while the template needs to be updated in the case of appearance change.

Fragment weights of the template are given in Section 2. After the experiments we found that if $\lambda_{p_k} < 0.5$, then q_k is an invalid fragment. Through the analysis we found that if the invalid fragment is caused by occlusion, the H component in HSV space of the fragment will be distributed with greater probability in the previous frame in the background region; if it is caused by target's own changes, the H component in HSV space of the fragment will be distributed with greater probability in the previous frame in the target region. Because obstacles only appear around the target, we select a "ring" background region around the target to determine whether the target is occluded or not, where the area of the annular region is about 2 times of target area. The distribution of the H component in HSV space of the invalid fragment is back projected onto the target and annular background in the previous frame, so we can obtain the distribution region in the previous frame, and then calculate respectively the probabilities of the invalid fragment belonging to the target region and the background region. This paper uses the main H component in HSV space of the invalid fragment for back projection, and the projection method is the same as the one used in Camshift [2]. As shown in Figure 6, back projection image is a binary image, where white pixels denote the pixels whose grey values are equal to those of the main H component of invalid fragments.

In back projection image, we count the number of the foreground pixels in the target area S_0 and the number of the foreground pixels in the whole image S_t. If the probability p_0 of the H component of the invalid fragment belonging to the target satisfies formula (16), we think the target appearance changes. th is set to 0.8 after many experiments. If the invalid fragment is determined as occlusion, the template will not be updated; if the invalid fragment is determined as appearance change, the information of the current frame will be updated to the template. So that we can rapidly get the target change information and suppress background disturbance to the template at the same time, which obviously improves the tracking performance:

$$p_0 = \log \frac{S_0}{S_t - S_0} > th \quad . \tag{16}$$

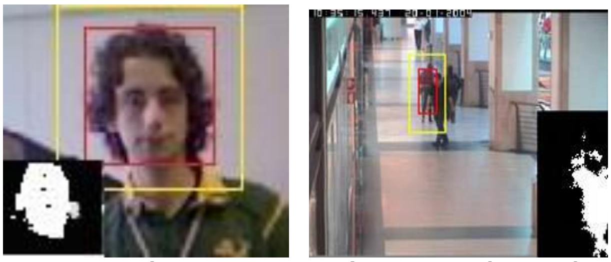

FIGURE 6 Back projection images under appearance change or the occlusion

5 Experiments and analysis

We use the following four groups of experiments to verify the effectiveness of the algorithm. The experimental environments are Microsoft visual studio 2008 and Opencv2.0. The results are analysed in Matlab 2008. Qualitative and quantitative analysis are made for the results of the four methods, the fragment-based tracker [12], the colour-texture based mean-shift tracker [7], the decentralized template tracker [17], and the method proposed in this paper (our method). In the figures, the window of our method is labelled as 1, the tracking window of the decentralized template tracker is labelled as 2, the tracking window of the colour-texture based mean-shift tracker is labelled as 3, and the tracking window of the fragment-based tracker is labelled as 4. The qualitative analysis is done through observing test images and the quantitative analysis is done through measuring location errors.

Let

$$err(x_{obj}, y_{obj}) = \sqrt{\frac{(x_{obj} - x_{true})^2 + (y_{obj} - y_{true})^2}{2}} \ . \tag{17}$$

Then $err(x_{obj}, y_{obj})$ is the location error between the target location (x_{obj}, y_{obj}) and the true location (x_{true}, y_{true}).

5.1 THE QUALITATIVE ANALYSIS

Test Sequence 1. Figure 7 shows the simulation results of a video in a campus named "campus". The video sequence contains 500 384×288-pixel colour images. In the video, from frame 272 to frame 377, the target enters into dark region from bright region and later returns to the bright region. The illumination has a great change. At frame 347, the tracking window of the fragment-based tracker loses the target; the colour-texture based mean-shift tracker does not lose the target, but the estimation location of the target has already deviated from the actual location. At frame 387, the tracking window of the colour-texture based mean-shift tracker loses the target too while the tracking windows of our method and the decentralized template tracker never lose the target.

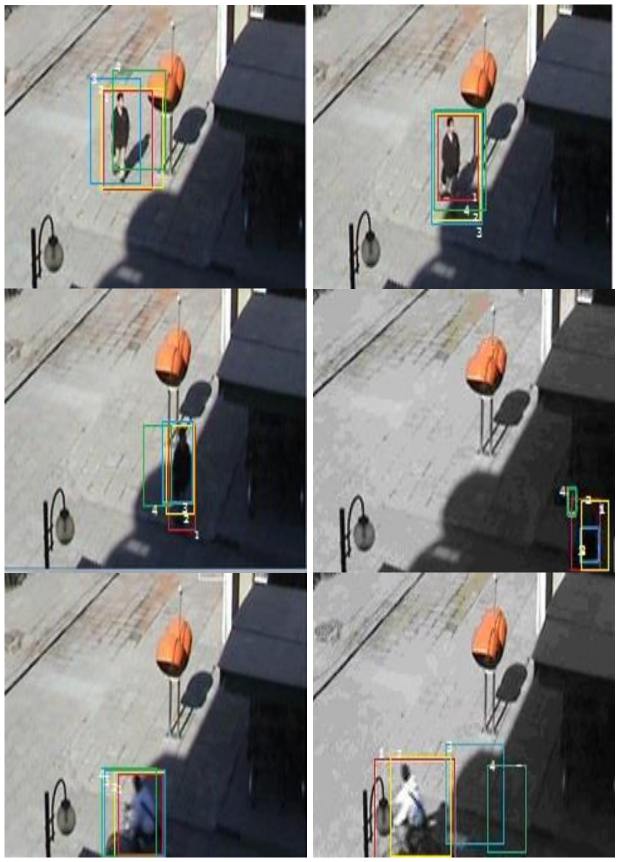

FIGURE 7 Tracking results of video "Campus" (Frames: 246, 272, 302, 347, 377, 387)

Test Sequence 2. Figure 8 shows the simulation results of video "hall", which comes from CAVIAR database. The video sequence contains 300 384×288-pixel colour images. The main difficulty with this sequence is that the target is occluded by another people. When the target is occluded at the video frame 191, both the fragment-based tracker and the decentralized template tracker lose the target from frame 208 to frame 221, and later on the colour-texture based mean-shift tracker loses the target too, while our method has kept tracking the target. To sum up, our method can always track the target no matter if there is occlusion or not.

FIGURE 8 Tracking results of video "hall"
(Frames: 163, 191, 208, 221, 240, 267)

Test Sequence 3. Figure 9 shows the simulation results of video "man1" which comes from SPEVI database.

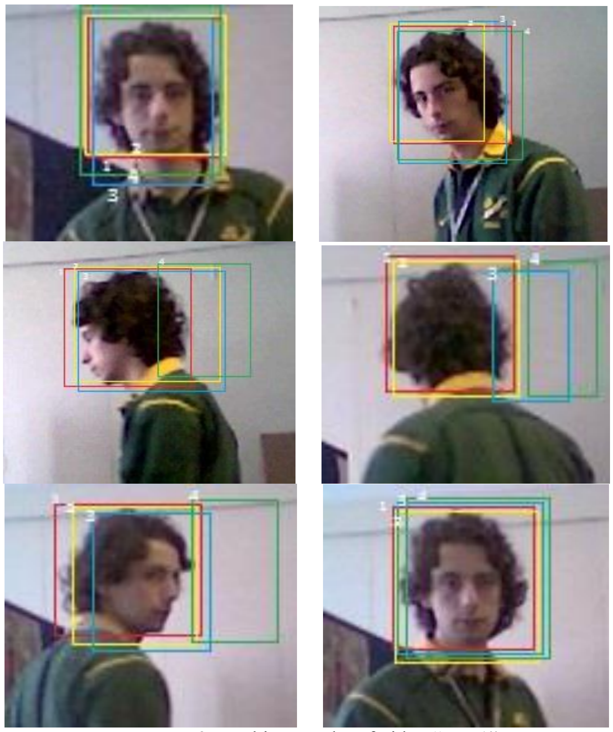

FIGURE 9 Tracking results of video "man1"
(Frames: 100, 110, 117, 130, 141, 145)

The video sequence contains 150 320×240-pixel colour images, where the man's head in the sequence has irregular translation and rotation movements. At frame 100 and frame 110 when the face is approximately front face, the four methods track the target accurately. When the back face appears at frame 130, the fragment-based tracker and the colour-texture based mean-shift tracker lose the target immediately, and the fragment-based tracker cannot recover from the error even for a side face at frame 131. Our method and the decentralized template tracker have more robustness.

Test Sequence 4. Figure 10 shows the simulation results of video "man2" which comes from SPEVI database.

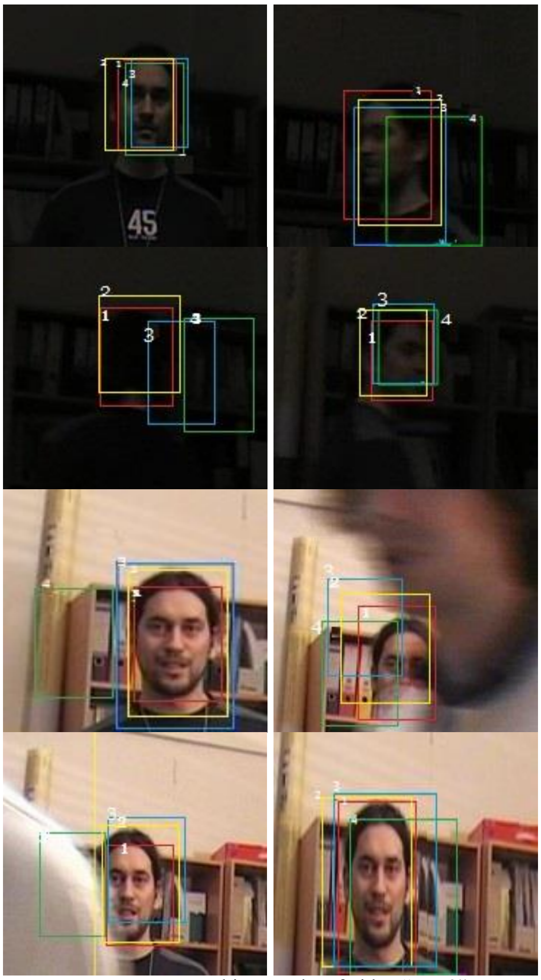

FIGURE 10 Tracking results of video "man2"
(Frames: 368, 384, 398, 410, 474, 580, 590, 601)

The video sequence contains 600 320×240-pixel colour images. The target first rotates and then the illumination changes sharply, followed by the part occlusion in the sequence. At frame 398 when the target rotates, both the colour and gradient of the target change, and the fragment-based tracker with one single cue tracking loses the target when the target turns back, the colour-texture based mean-shift tracker with fixed template tracking loses the target too, and the decentralized template tracker brings about a certain tracking drift while our method achieves accurate tracking. At frame 474, the lamp is turned on, which makes the illumination change sharply and the fragment-based tracker lose the target again, while the other three algorithms keep tracking the target. At frame 580 when the target is occluded by another moving object, the other three algorithms fail to track the target accurately while our method still tracks well, since our method not only uses adaptive cue integration scheme during fragment tracking, but also employs the online updating template by distinguishing appearance changes and occlusions.

5.2 THE QUANTITATIVE ANALYSIS

Figures 7-10 show the qualitative analysis results. We also provide the error plots for each sequence in Figure 11.

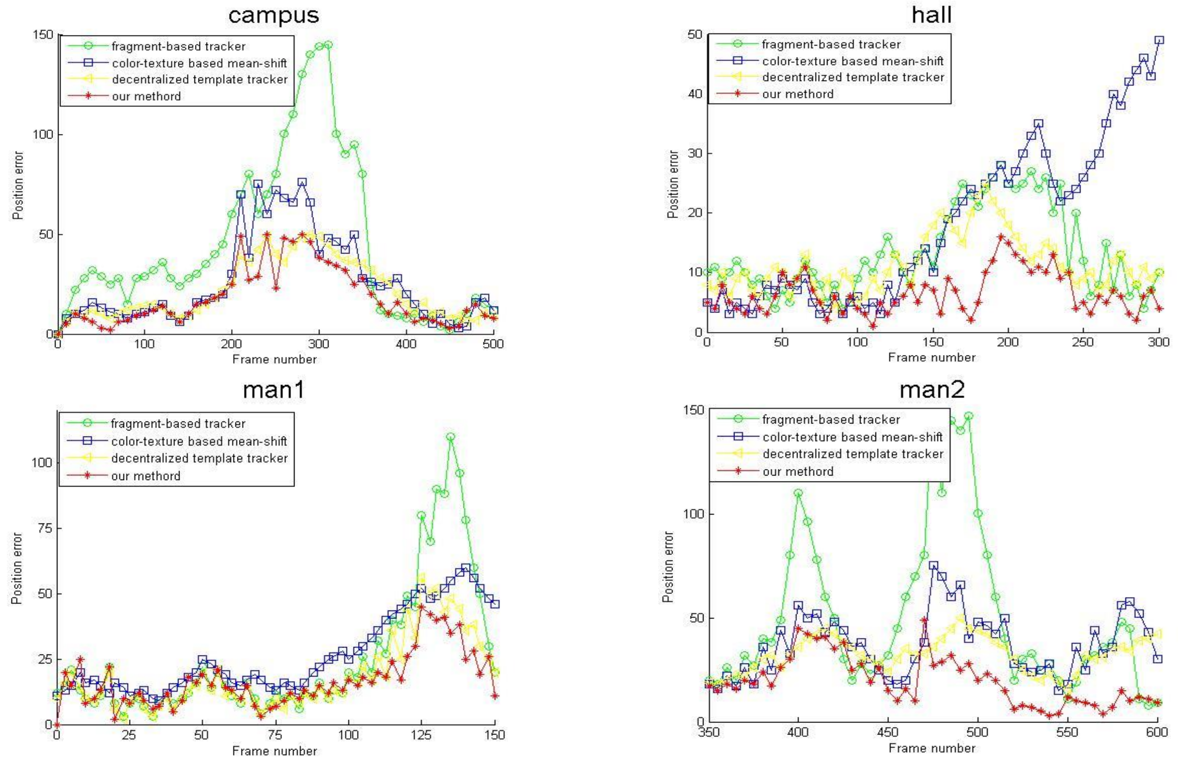

FIGURE 11 Error plots for the video sequences used in the quantitative analysis

It can be seen from these results that the proposed algorithm outperforms the fragment-based tracker in terms of tracking accuracy. The reason for this mainly stems from our adaptive cue integration scheme, which removes the assumption on the degree of potential occlusions for the fragment-based tracker, and replaces its competitive approach to cue integration with a more cooperative strategy.

The colour-texture based mean-shift tracker, which uses the joint colour and LBP cues to track target, outperforms the single cue tracking approaches and is robust to the complex environment. However, it yields, in many cases, poor tracking results when the target becomes occluded by the other objects in the scene. Our method and the fragment-based tracker cope with these kinds of occlusion better than most of the multi-cue integration trackers. The reason stems from tracking the target in fragments.

The decentralized template tracker, which uses online updating template to track target, is proved to be very robust against severe appearance changes since it can learn a new object model at each frame. Our method gives better results than the decentralized template tracker for the video sequences because of the fact that the decentralized template tracker updates target template at each frame without discrimination of appearance changes and occlusions, and cannot prevent the template from drifting, even if it uses short-term and long-term appearance models

As a result, for most of the video sequences, our algorithm provides the best results.

6 Conclusions

Video target tracking algorithm is faced with many problems, such as illumination change, occlusions, appearance changes, etc. A new tracking algorithm of multi-cue integration based on fragment has been proposed in this paper. The new algorithm associates the image fragments describing the different parts of the target object with multi-cue integration, where the weight of each cue is dynamically adjusted during tracking. In this way, the vote of each fragment contributes to the joint tracking result according to its adaptive weight integration. Furthermore, the algorithm uses the template updating for target tracking, designs the corresponding template update algorithm, and successfully implements the target tracking under complex background. We have demonstrated the potential and the effectiveness of the proposed approach on various challenging video sequences with different tracking scenarios. As revealed by our experiments, the proposed approach works generally better than the fragment-based tracker, the multi-cue integration trackers and template-based trackers for the sequences in the complex background of illumination, occlusion, and the appearance change.

Acknowledgements

The authors wish to thank the helpful comments and suggestions from my teachers and colleagues. This work was supported by the National Natural Science Foundation of china(Grant No.31371533),National Science and Technology Support Program(Grant No. 2013BAJ10B12), he Key Technologies R & D Program of AnHui Province (Grant No.1301032169) and by the Natural Science Foundation of AnHui Province(Grant No.1308085MF89).

References

[1] Birchfield S 1998 *Computer Vision and Pattern Recognition (CVPR) Santa Barbara, CA:IEEE Computer Society* 232-7

[2] Triesch J, von der Malsburg C 2001 *Neural Computation* **13**(9) 2049-74

[3] Yuru Wang, Xianglong Tang, Qing Cui 2012 *Pattern Recognition* **45**(12) 4510-24

[4] Brasnett P, Mihaylova L, Bull D, Canagarajah N 2007 *Image Vision Computation* **25**(8) 1217-27

[5] Wang J Q, Yagi Y 2008 *IEEE Transaction on Image Processing* **17**(2) 235-40

[6] Jia G, Tian Y, Wang Y 2010 Dynamic multi-cue tracking with detection responses association *Proceedings of the ACM Multimedia* NY USA: ACM New York 1171-74

[7] Ning J, Zhang L, Zhang D, Wu C 2009 *International Journal of Pattern Recognition and Artificial Intelligence* **23**(7), 1245-63

[8] Li Ying, Liang Shi, Bai Bendu, Feng David 2012 *Multimedia Tools and Applications* **6**(3) 1-21

[9] Nejhum S M S, Ho J, Yang M H 2008 Visual tracking with histograms and articulating blocks *Proc. Conf. Computer Vision and Pattern Recognition (CVPR), Anchorage,* AK :CVPR 2008 1-8

[10] Nickel K, Stiefelhagen R 2008 Dynamic integration of generalized cues for person tracking *Proc. Eur. Conf. Computer Vision (France)* (ECCV): Marseille 514-26

[11] Chockalingam P, Pradeep N, Birchfield S 2009 Adaptive fragments-based tracking of non-rigid objects using level sets *Proc. Int. Conf. Computer Vision (Japan)* (ICCV) :Kyoto 1530-7

[12] Adam A, Rivlin E, Shimshoni I 2006 Robust fragments-based tracking using the integral histogram *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)* 798-805

[13] Erdem E, Dubuisson S 2012 *Computer Vision and Image Understanding* **116**(7) 827-41

[14] Porikli F, Tuzel O, Meer P 2006 Covariance tracking using model update based on Lie algebra. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition:New York, USA* 728-35

[15] Collins R T, Liu Y X, Leordeanu M 2005 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(10) 1631-43

[16] Wang Yong-Zhong, Liang Yan, Zhao Chun-Hui, Pan Quan 2008 *Acta Automatica Sinica* **34** (4) 393-9

[17] Firouzi Hadi, Najjaran Homayoun 2012 *Pattern Recognition* **45**(12) 4494-509

[18] Porkili F 2005 Integral histogram: A fast way to extract histograms in Cartesian spaces *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* 439-47

[19] Comaniciu D, Ramesh V, Meer P 2003 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(5) 564-77

[20] Yang C, Duraiswami R, Davis L 2005 Fast multiple target tracking via a hierarchical particle filter *Proceedings of the Tenth IEEE International Conference on Computer Vision: Beijing, China* 212-9

[21] Simard R, L'Ecyuer P 2011 *Journal of Statistical Software* **39**(11) 1-18

[22] Fashing M, Tomasi C 2005 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(3) 471-4

Authors

Lichuan Gu, October 1974, Sichuan, China

Current position, grades: Associate Professor of Computer Science Department of Anhui Agricultural University
University studies: BSc, MSc degrees in computer science in HeFei University of technology, China, in 1997 and 2005.
Scientific interest: research interests, mainly including machine learning, image processing, and pattern recognition.

Chengji Wang

Current position, grades: master studies in agricultural informatization in AHAU University
University studies: BE in Logistics Engineering, AHAU University (2013)
Scientific interests: research interests include cognitive computing and logistics system optimization

Jianxiao Liu

Current position, grades: master studies in Agricultural Informatics at University.
University studies: BE in Logistics Engineering (2013) from AHAU University.
Scientific interests: research direction includes different aspects of Agricultural logistics