COLLABORATIVE MEAN SHIFT TRACKING BASED ON MULTI-CUE INTEGRATION AND AUXILIARY OBJECTS

Hong Liu, Lin Zhang, Ze Yu, Hongbin Zha, Ying Shi

State Key Laboratory of Machine Perception
Peking University, Shenzhen Graduate School, P.R.China
Email: hongliu@pku.edu.cn, zhanglin@cis.pku.edu.cn, angeloyz@hotmail.com
zha@cis.pku.edu.cn, shiying@cis.pku.edu.cn

ABSTRACT

Colour-based Mean Shift is an effective and fast algorithm for tracking colour blobs. However, it is vulnerable to full occlusion and target out of range for a few frames. This paper proposes a tracking method based on multi-cue integration and auxiliary objects to deal with these problems. A colour-location-prediction integration Mean Shift method is proposed to track each auxiliary object. Motivated by the idea of tuning weight of each cue according to their performances, these three cues are integrated adaptively according to their quality functions. Moreover, auxiliary objects get effective relative information with targets automatically, and update the information ceaselessly. When the target disappears, auxiliary objects will export useful information to estimate the location of the target. Experiments show that this method can adapt the weight of multi-cue efficiently, reinitialize the targets after long time disappearance, and increase the robustness of tracking in various conditions.

Index Terms—Auxiliary Objects, Multi-Cue Integration, Mean Shift

1. INTRODUCTION

Tracking algorithms fall into two categories. The first category is probabilistic methods[1][2][3] which view the tracking algorithm as a state solving problem under the Bayesian framework, model uncertainty and propagate conditional densities through the tracking process. The second category is deterministic methods [4][5][6]which compare a model with current frame and find out the most probable region. The deterministic methods can hardly handle complete occlusion very well.

Up to now, most of the Mean Shift related literatures employ only single color probability distribution which is vulnerable to complex conditions. Some researchers are focusing on establishing a multi-cue integration mechanism under the probabilistic framework, including Monte Carlo method [7], Particle Filter [8], etc. There are four reasons to integrate location cue and prediction cue into Mean Shift. First, location cue and prediction cue can offer necessary position information and moving information to enhance the robustness of the original color-base Mean Shift methods in various conditions. Second, location and prediction detection results can be integrated with color distribution naturally. Third, as Mean Shift is robust to small distractors, we can employ preliminary location and predict detection algorithm, which saves computation. Last, Mean Shift is a fast mode searching algorithm which saves computational resources for multi-cue integration methods.

At present, both the occlusion problem and target out of the range problem are difficult to solve for tracking. In Mean Shift algorithm, if the object is totally occluded for a couple of frames, the algorithm has no mechanism to continue tracking. Ming Yang' method[9] integrates into the tracking process a set of auxiliary objects that are automatically discovered in videos on the fly by data mining to verify whether the tracker is following the true target and avoid drifting due to short-term occlusion or tracking lost.

In this paper, probability distributions from location, prediction and color cues are integrated under the Mean Shift framework. These cues can enhance the robustness of the original color based Mean Shift methods in various conditions like color clutter in background and target confused by similar objects. Moreover, motivated by the idea of tuning weight of each cue according to their performances, these three cues are integrated adaptively according to their quality functions, which are used to evaluate their performance in the adaptive integration scheme. Based on this, collaborative tracking using auxiliary object is proposed. Auxiliary objects get effective relative information with the target automatically, when target is occluded or out of range, auxiliary objects will export useful information to help estimate the location of the target. This method is robust to long term full occlusion and target out of range.

The rest of this paper is organized as follows: section 2 brings out the adaptive color-location-prediction integration method, and

This work is supported by National Natural Science Foundation of China (NSFC, No. 60675025) and National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247).

section 3 illustrates the collaborative tracking with auxiliary objects. Experiment results and conclusions are presented in section 4 and section 5, respectively.

2. INTEGRATING MULTIPLE CUES

2.1. Color, Location and Prediction Cues:

A color probabilistic map $p_c(X, t)$ is calculated by histogram back projection: first, the color histogram of the object's color is calculated and stored in a look-up table. When each new frame comes in, the table is looked up for each pixel's color, and a probability value is assigned to each pixel. Hence, a probabilistic distribution map is obtained.

However, there are two deficiencies in color-based tracking. Firstly, when pixels' color has a low saturation near zero, hue is not defined or inaccurate, which results in inaccuracy and noise in the back projection image. The second is similarly colored backgrounds. Mean Shift is robust to small distractors, but if the distractor is larger than the object color area, the object may be lost when it moves near the similar distractor.

It should be noticed that the distrators are all from the backgrounds. When the camera is static, the background is assumed not moving, which can be used as a priori to eliminate those noisy areas in the back projection image. Therefore, it is believed that location information can be employed to solve the deficiencies.

First, adaptive background subtraction technique is employed to get the entering object's silhouette. Suppose $I(X,t)$ is the incoming frame, $B(X,t)$ is the updated background image, th is an appropriate threshold, then:

$$D(X,t) = \begin{cases} 1, & |I(X,t) - B(X,t)| > th \\ 0, & else \end{cases} \quad (1)$$

The location image $D(X, t)$ can be regarded as representing the pixels' probabilistic distribution of location, so it is convenient to be integrated into other 2D discrete distributions. As background subtraction results always have distractors, combining with other complementary cue may be helpful to eliminate distractors and noise. For each image I, denote $p_m(X, t)$ to be the location probability of the pixel at X at time t. Let

$$p_m(X,t) = D(X,t) \quad (2)$$

Here, $p_m(X,t)$ is a binarized distribution that represents the probability of location for each pixel.

Moreover, if there are other moving objects near the target, when the location probability increases, location cue may enhance the influence of moving objects, too. As a result, it will confuse the target and enlarge the search area easily which will lead to failure results. Therefore, it is necessary to bring in prediction cue to enhance the probability in the target area. The target may move at a steady velocity in short time. The position in the next frame can be estimated by velocity from former frames, and the tracking window size in the next frame can be assumed to change little. A table is created to store target positions and tracking window size in former N frames. Suppose T_t is the table at time t, $T_t = \{s_t^i, p_t^i \mid i \in (1, 2\cdots N)\}$, s_t^i ($i \in (1\cdots N)$) stores tracking window size at time (t-1-N+i) and p_t^i ($i \in (1\cdots N)$) stored tracking window position at time (t-1-N+i). We get the average velocity $\overline{v}_t$ and tracking window size $\overline{s}_t$ from P_t. As a result, we can predict the position $\hat{p}_{t+1}$ and tracking window size $\hat{s}_{t+1}$ in the next frame:

$$\hat{p}_{t+1} = p_t + \overline{v}_t, \quad \hat{s}_{t+1} = \overline{s}_t \quad (3)$$

Denote $N(X, t)$ to be the prediction image,

$$N(X,t) = \begin{cases} 1, & pixel \in prediction - window \\ 0, & else \end{cases} \quad (4)$$

The image $N(X, t)$ can also be integrated into other 2D discrete distributions easily. For each image I, denote $p_e(X,t)$ to be the prediction probability of the pixel at X at time t. Let

$$p_e(X,t) = N(X,t) \quad (5)$$

Here, $p_e(X,t)$ is a binarized distribution that represents the probability of the target in the next frame.

How to combine these three maps is of our interest. These three cues have different reliability in various conditions, i.e. the contribution of each cue is not the same. Hence, we employ a weighted Multi-Cue integration technique. Our work differs from McKenna' work[10] in the point that we employ the adaptive integration mechanism under the framework of Mean Shift, with a new quality function suitable for blob tracking.

Suppose $p_i(X, t)$ is the probability distribution map of cue i, $p(X, t)$ is the combined probability distribution map, the cues are integrated as a weighted sum of probability distribution,

$$p(X,t) = \sum_i \omega_i(t) \times p_i(X,t), \quad \sum_i \omega_i(t) = 1 \quad (6)$$

The weighted integration method adapts each cue's weight according to the reliability of each cue in previous frame. Suppose the performance of individual cue i can be evaluated using a quality function $q_i(t)$. The relation between the quality and weight of cue i can be defined as follows,

$$\tau\omega_i'(t) = \overline{q}_i(t) - \omega_i(t), \quad \overline{q}_i(t) = q_i(t) / \sum_i q_i(t) \quad (7)$$

τ is a time constant controlling the speed of update. Formula (7) can be used to update individual weight of each cue. It can be regarded as a running average, and $\omega_i(t)$ is adapted according to $q_i(t)$, which brings in the information about the performance of cue i in the last frame. Quality functions $q_i(t)$ can be viewed as feedback of tracking results. In this paper, the quality function can be defined as the ratio between the numbers of non-zero pixels inside and outside the tracking window on individual probability distribution map.

After the combined probability distribution map is obtained, a region detection algorithm should be operated on the combined map and find the objects. The color-location-prediction integration based Mean Shift algorithm includes following steps:

Step1. Calculate the color probabilistic distribution map $p_c(X, t)$ through back projection.

Step2. Calculate the location distribution map $p_m(X, t)$ through location detection.

Step3. Calculate the prediction distribution map $p_e(X, t)$ through prediction detection.

Step4. Integrate these three maps to the combined distribution map $p(X, t)$ according to their weights, respectively.

Step5. Choose a search window scale s_0 and initial location P_0 on the map $p(X, t)$.

Step6. Compute the mean location $\hat{P}$ and zeroth moment M_{00} of the pixels in the search window (P, s).

Step7. Set the new window parameters as $P = \hat{P}$, $s = k\sqrt{M_{00}}$. (k is a constant).

Step8. If convergence, repeat steps 6 and 7, else stop the algorithm.

3. COLLABORATIVE TRACKING

Mean Shift algorithm is vulnerable to full occlusion and target out of range for a few frames because current iteration is initialized according to the previous one. If an object totally disappears for a couple of frames, the tracking window will drift away and the algorithm is not appropriate to continue tracking. Based on the multi-cue integration mechanism, a collaborative tracking method can also be advanced, which can be used to deal with full occlusion and target out of range. This method can reinitialize tracking automatically when the object reappears. In this method, auxiliary objects are used to offer useful information related to the target. At first, auxiliary objects get effectively relative information with the target automatically, and update the information ceaselessly. When the target is difficult to track or disappear, auxiliary objects will export useful information to estimate the location of the target, which can help to reinitialize the position of the target correctly. What's more, the location of the target can be estimated for several frames until the target recurs, therefore, this method is advancement to methods which can only be predicted for the next frame. On the other hand, as the relative information is stored between the target and the auxiliary objects, the target can also be regarded as an auxiliary object offering information to other auxiliary objects when they disappear.

In this paper, we chose auxiliary objects which have high location correlation with the target by hand in the first several frames, which is propitious to track the target in real time. Then, calculate each candidate auxiliary object's correlation with the target. The ones with high correlation will be kept and used as auxiliary objects.

For each auxiliary object, create a table to memorize their relative distance to the target and the size of their track windows in the past N frames.

$$D_j = \{d_j^i, i = t+1-N, \cdots, t-1, t\} \quad (8)$$

$$S_j = \{s_j^i, i = t+1-N, \cdots, t-1, t\} \quad (9)$$

Suppose D_j to be the table of auxiliary object j, d_j^i represents the distance between the target and object j in the frame i. Moreover, S_j denotes to be the size of auxiliary object j's tracking window. Both of them will be updated all the time. Therefore, we can obtain the effective information between the auxiliary objects and targets.

When the target is fully occluded or out of range at frame T_{dis}, get the average relative distance for each auxiliary object to the target by these tables. For object j, denote $\overline{D_j}$ to be the average relative distance:

$$\overline{D_j} = \sum_{i=t+1-N}^{N} d_j^i / N \quad (10)$$

The target and each auxiliary object have highly location relation, and the distance between them may be steady, so the position of the target can be estimated by the average distance of each auxiliary object before the target disappeared. Each auxiliary object offers information about the likely position of the target. Assume mp_j to be the likely position deduced by auxiliary object j, and p_t^j represents the position of auxiliary object j at frame t. From all the information the auxiliary objects supply, we can conclude that the target may emerge in the average likely position:

$$mp_j = \overline{D_j} + p_t^j, \quad P_{target} = \sum_{i=1}^{M} mp_i / M \quad (11)$$

Here, M means the number of auxiliary objects. The size of searching area is defined a bit larger than the former tracking window of the target. By this method, we can focus on area where the target may reappear and exclude other similar objects easily. In the next step, large non-zero region is searched in the estimated place according to the combined probability distribution to find the reappearing object.

Fig.1 shows the flow chart of the collaborative tracking when target disappeared for several frames.

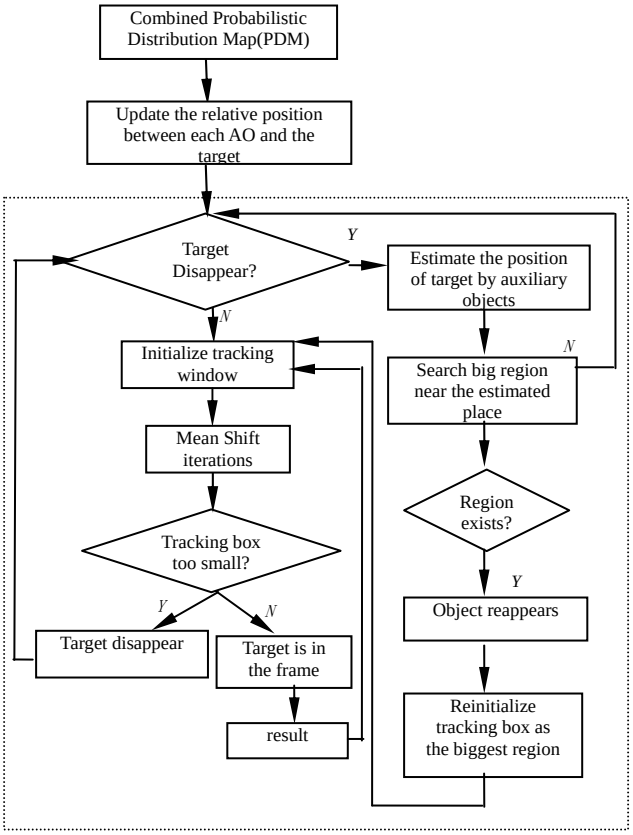


Fig.1 Multi-cue integration based Mean Shift collaborative tracking when target disappear (emphasized by the dotted box).

4. EXPERIMENTS

Algorithm is tested with 9 video sequences including 2300 frames. A 2D histogram with 16 bins hue and 10 bins saturation is used. In order to enhance the target, the length and width of prediction window is reduced to 0.6 times. Target in 93.6% frames can be tracked successfully. Four samples are explained thoroughly as follows.

S1 is the sequence with multi-people objects. S2 is the sequence in which target are confused by other objects with similar feature. S3 is the sequence in which auxiliary objects are used for collaborative tracking. S4 is the sequence in which targets are out of range. All the results are real-time.

Sequence S1 shows that integrating prediction information helps multi-cue Mean Shift algorithm tracking the target correctly when the location cue increases which leads to holes of other moving objects.

(a) the right result (b) the wrong result

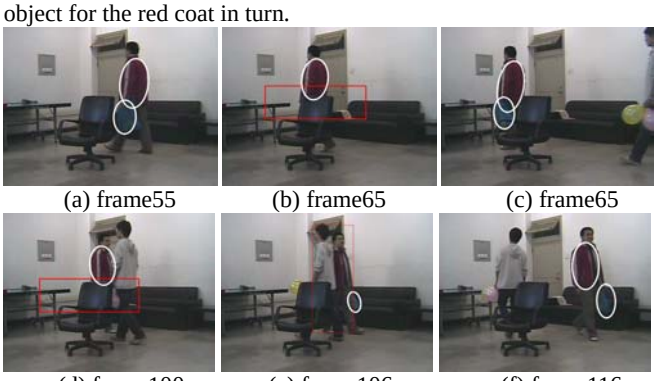
Fig.2. (a) shows the result with prediction information, (b) shows the wrong result without prediction information from sequence S2

In sequence S2, two men wearing red clothes walk face to face, and go in opposite direction. Prediction cue supplies important information enhancing the probability of the target, which is helpful to track the target successfully.

(a) Frame75 (b) Frame81

Fig.3. Walking face to face with red clothes from sequence S2.

The occlusion handler is tested through multiple-human video sequences S3. Fig.4 shows a full occlusion case in sequence S3. The red coat and blue bag are regarded as auxiliary objects to each other. When the bag is occulted by the sofa from frame 59 to 67, the red coat is regarded as the auxiliary object for the blue bag, the red rectangle shows the likely position of the bag. When the boy is occluded by the other boy, the blue bag is regarded as the auxiliary object for the red coat in turn.



(a) frame55 (b) frame65 (c) frame65

(d) frame100 (e) frame106 (f) frame116

Fig.4. Full occlusion case from video sequence S3.

Fig.5 samples the results on the sequence which is very challenging due to a serious occlusion, target out-of-range. In this sequence, the boy in red coat and blue bag are selected as auxiliary objects to each other and the auxiliary object can still reinitialize tracking. When the boy goes out of the target for 56 frames, the blue bag is used to estimate the position the boy may recur, and help reinitialize the tracking of the boy successfully.

(a)frame85 (b)frame 95

(c)frame 115 (d)frame170

Fig.5. Target out of range from video sequence S4. (a)-(d) are tracking results from S4.

5. CONCLUSIONS

In this paper, we propose a collaborative tracking method based on multi-cue integration and auxiliary objects. Location and prediction cues integration can deal with tracking problems in complex conditions, and the weight of multi-cue can be adapted according to their quality functions. Auxiliary objects are able to reinitialize the Mean Shift algorithm efficiently by estimating the possible position of the target with stored useful information. Experiments show that this method can increase the robustness of tracking in various conditions.

REFERENCES

[1] M. Isard and A. Blake, “CONDENSATION - conditional density propagation for visual tracking”, Int. Journal of Computer Vision, pp.5-28, 1998.

[2] K. Nummiaro, E. Koller-Meier and L. Van Gool, “Object tracking with an adaptative color-based Particle Filter”, Image and Vision Computing, pp.99-111, 2002.

[3] P. Perez, C. Hue, J. Vermaak and M. Gangnet, “Color-based probabilistic tracking”, European Conference on Computer Vision, pp 661-675, 2002.

[4] G.R. Bradski, “Computer vision face tracking for use in a perceptual user interface”, IEEE Workshop on Applications of Computer Vision, pp.214-219, 1998.

[5] D. Comaniciu, V. Ramesh and P. Meer, “Real-time tracking of non-rigid objects using mean shift”, IEEE Conference on Computer Vision and Pattern Recognition, Proceedings, pp.142-149, vol.2, 2000.

[6] T.-L. Liu, and H.-T. Chen, “Real-time tracking using trust-region methods”, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 397 - 402 , vol. 26, Issue 3, 2004.

[7] Y. Wu and T.S. Huang, “A Co-inference Approach to Robust Visual Tracking”, Proc. of Int. Conference on Computer Vision, pp. 26-33, Vol.2, 2001.

[8] M. Spengler and B. Schiele, “Toward robust multi-cue integration for visual tracking”, Machine Vision and Applications, pp.50-58, vol.14, 2003.

[9] Ming Yang, Ying Wu and Shihong Lao, "Intelligent Collaborative Tracking by Mining Auxiliary Objects", Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on Volume 1, pp. 697 – 704, 17-22 June 2006.

[10] S. J. McKenna, Y. Raja and Shaogang Gong, “Tracking colour objects using adaptive mixture models”, Image Vision Computing. pp. 225-23, 1999.