Robust Tracking Using On-Line Selection of Multiple Features

Sihua Yi, Zhijun Yao, Juntao Liu, Jun Chen, Wenyu Liu
Department of Electronics and Information Engineering
Huazhong University of Science and Technology
Wuhan, China

Abstract—This paper presents a novel on-line feature selection mechanism based on mean shift tracking algorithm, which adjusts the weight for each feature and each bin in feature histograms during the tracking process, according to the discrimination between the appearance of object and background with different features. Then the tracking performance would be stable and reliable using those features combined by the weights. As the appearance model, we select features of gray level, Local Binary Patterns (LBP) texture and edge orientation. The gradient amplitude is supplement to the feature space for tracking. Experiments on two video sequences show the effectiveness of the proposed method.

Keywords—object tracking, mean shift, LBP texture, edge orientation, gradient amplitude, feature selection

I. INTRODUCTION

Real-time object tracking is a hot research topic in many computer vision applications, such as surveillance and human computer interaction. One of main factors which limit tracking performance is the lack of robustness.

Features of gray level [1, 2], LBP [3, 4, 5, 6] and edge orientation [7, 8] are respectively used for mean shift tracking. However, the target presented by individual feature, may have a similar appearance to the background in some complex environment, which makes the tracking unstable. By using a “center-surround” approach to sample pixels from object and background, Collins et al. [9] backprojects the log likelihood ratio between foreground and background into a weighted image, and assign high weight to those pixels corresponding to histogram bins which are discriminative. However, weights get from log likelihood ratio function are not always reliable, and sometimes appearance of pixels from foreground with high weight still looks like background, as Fig.1 (b) shows. This is because that, the histogram of some feature is multimodal, and two successive bins may be not independent, furthermore, peak is more stable than bin. A Gaussian Mixture Model (GMM) is proposed in [10], from which the stable Gaussian component of object model can be obtained, as shown in Fig.1 (c).

As the target moves from place to place, both foreground and background appearance will change, and those features corresponding to changed appearance may be ineffective, while the others may be still valid enough. To improve robustness of tracking performance, we use gray level, LBP, edge orientation and gradient amplitude feature to characterize color, texture and structure of the target. However, once background changes, the discrimination between object and background of features will also changes. Thus for promoting the stability of tracking performance, it is necessary to find which and how those features are discriminative on line. In this paper, we firstly calculate statistic histograms of gray level, LBP, edge orientation and gradient amplitude feature. Then we figure out the discrimination of each feature histogram and each bin based on the distinction between appearance of object and background, and assign each feature and bin a weight accordingly. Finally, we obtain a novel feature combination to make tracking more robust.

II. THE PROPOSED METHOD

During a tracking process, if only one feature is used to represent the target, the contribution of each bin in the feature histogram is various and continually changing. To track the target robustly, the higher contribution a bin provides, the higher weight should be assigned. We call these weights interior weight of features. Moreover, individual feature can not guarantee the accuracy of tracker. And if we use multiple features to track, the contribution of each feature is also obviously various. It is necessary to assign each feature a...
different weight accordingly, to make those discriminative features provide more contribution for tracking. We call these weights exterior weight of features. We will present how to calculate interior and exterior weight of features in the next two sections.

A. Interior Weight of Features

For an individual feature, bins that are discriminative between foreground and background can provide robust tracking performance. Assuming the distributions of gray level and gradient amplitude on image are multiple Gaussian, we use GMM to characterize these features. So their interior weight could be calculated by the stable Gaussian from \([10]\). While extracting the feature of LBP and edge orientation, pixels from flat area of image are discarded, so their distributions can’t be Gaussian. As a result, we still calculate their interior weight using log likelihood function. Moreover, the range of log likelihood ratio \(L_{u}\) obtained by \([9]\) is between \(\log \theta\) and \(-\log \delta\), which should be mapped into \([0,1]\) by

\[
w_e = \frac{L_u - \min\{L\}}{\max\{L\} - \min\{L\}},
\]

where \(\max\{L\}\) denotes the maximum of log likelihood ratio between foreground and background in current frame while \(\min\{L\}\) is the minimum of it.

B. Exterior Weight of Features

When tracking using combined multiple features, the contributions of different features are not the same for the change of scene. Those features characterizing the target, which have similar performance to the background once, is not reliable and should be assigned small weights. On the contrast, discriminative features should be assigned a high weight.

We still use the "center-surround" approach to evaluate the distance between foreground and background. There are three measures commonly used to calculate the distance between histograms: the Bhattacharyya distance, \(L_2\) distance and chi-square distance. (a) object with red bounding box, (b) distance over y-translation (in pixels), (c) distance over x-translation (in pixels).

![Figure 2. Comparison with three distance measures: Bhattacharyya distance, \(L_2\) distance and chi-square distance. (a) object with red bounding box, (b) distance over y-translation (in pixels), (c) distance over x-translation (in pixels).](image)

C. Histogram of Gradient Amplitude

For the instability of individual feature, gray level, LBP and edge orientation features are used to describe color, texture and structure information of target in this paper. When expressing the outline of object, edge orientation histogram loses the amplitude of edge points, which can not characterize the whole structure information of the target. So we put forward gradient amplitude histogram here. We have to solve two problems before computing the statistic distribution of gradient amplitude of the target in each frame:

1) The gradient amplitude of edges may change sharply, once the illumination in video changes.

2) Most pixels from video frames are located at flat area, which do not contain the information of structure. Furthermore, the number of them is so large that they ruin the statistical property of gradient amplitude histogram.

For the first problem, considering that gradient amplitude of edge points is always higher than that of points on flat area, we pick the maximum value of gradient amplitude in each frame as the upper bound, and allocate the gradient amplitude into \(l\) bins, so that pixels with high bin number are closed to edges. For the second problem, we set a gradient threshold to filter out those insignificant pixels. Denote \(grads(x_i)\) as the gradient amplitude of the pixel located on \(x_i\). The index of bin for the \(i^{th}\) pixel is computed as

\[
b_{\text{grad}}(x_i) = \begin{cases} \text{floor}(\max\{\text{grads}(x_i)\}/\text{threshold}) & \text{if } \text{grads}(x_i) \geq \text{threshold} \\ 0 & \text{otherwise} \end{cases}
\]

where \(\text{floor}(\cdot)\) is a rounding function, and \(\max\{\text{grads}\}\) is the maximum of gradient amplitude on image. In our experiments we set \(\text{threshold} = \max\{\text{grads}\}/5\). Gradient amplitude feature is invariant to the 2D translation, scale transform and rotation. Combined with edge orientation feature, it can describe the whole structure information of object.

D. Tracking

In conclusion, denote \(q_G, q_T, q_O\) and \(q_\delta\) as the gray level, LBP, edge orientation and gradient amplitude histogram of the
target respectively, while $p_G$, $p_T$, $p_O$ and $p_d$ as the feature histograms of the target candidate. The similarity between the target and candidate is

$$
\rho(y) = W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}
$$

$$+ W_O \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_d \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}},
$$

where $W_G$, $W_T$, $W_O$ and $W_d$ are the external weight of gray level, LBP, edge orientation and gradient amplitude features respectively, $w_{\omega}$, $W_T$, $w_{\omega}$ and $w_{\omega}$ are the interior weight of features, and $m_1$, $m_2$, $m_3$ and $m_4$ are the number of bins in each feature histogram.

Using Taylor expansion around the equation (7), the linear approximation of the value $\rho(y)$ is obtained after some manipulations as

$$\rho(y) = \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}})
$$

$$+ \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right) + \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right)
$$

$$- \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) + \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right)
$$

$$- \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) + \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right)
$$

$$+ \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right) + \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right)
$$

$$+ \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right) + \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right)
$$

$$- \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) + \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right)
$$

$$- \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) + \left( \frac{1}{2} (W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y) q_{\omega}}) \right)
$$

where $\omega_j$ is weight value of the pixel at location $y_j$, given by

$$\omega_j = W_G \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y_j) q_{\omega}} + W_T \sum_{\omega} w_{\omega} \sqrt{p_{\omega}(y_j) q_{\omega}} + \sum_{\omega} \omega_j k\left( \frac{y_j}{\lambda} \right),$$

In equation (8), the function $b: \mathbb{R}^2 \to \{1...m\}$ associates to the pixel at location $y_j$ the index $b(y_j)$ of its bin in the quantized feature space. Then the new location $y_{i+1}$ of the target center can be get from the current location $y_i$ according to

$$y_{i+1} = \sum_{x} g(x) \left[ \begin{array}{c} y_i - x \\ \lambda \end{array} \right],
$$

where $g(x) = -k'(x)$, and $k(x)$ is a kernel function defined in [2].

### III. EXPERIMENT

#### A. Tracking Time Complexity

Our tracking algorithm can be summarized as three steps: calculating feature histograms, weights of histogram and the center location of target in next frame. Let $N$ be the average number of iterations per frame in mean shift model and $n_i$ the number of pixels corresponding to the target. We conclude that the mean cost of the proposed algorithm given by

$$\text{Cost} = N(c_H + c_W + n_i c_{\text{MS}}),$$

where $c_H$ is the cost of histograms, $c_W$ is the cost of histogram weights and $c_{\text{MS}}$ is the cost of an addition, a square-root, and a

In the experiment, both algorithms are implemented in MATLAB 2010 and running on a PC with 2.40GHz CPU and 3GB RAM.

#### B. Experiment Result and Analysis

Taking two video sequence "woman" and "sylv" for examples, using features of gray level, LBP, edge orientation and gradient amplitude to track in experiments, we’ve compared the performance between three tracking algorithms: our algorithm, mean shift algorithm with interior weight of features but no exterior weight, and traditional mean shift tracking algorithm with four features, as shown in Fig.5 and Fig.9 respectively.

As shown in Fig. 3, the method proposed in this paper has a more stable performance. We have calculated exterior weight of four features in each frame, illustrated by Fig. 4.
interior weight of four features, and blue rectangle for [2]. Frames 13, 121, 145, 199 and 265 are displayed.

Figure 4. Exterior weights of four features in video sequence "woman".

As shown in Fig. 4, gray level, edge orientation and gradient amplitude feature of the target increase or decrease alternately with the change of scene, while LBP feature is affected by background so much that the exterior weight of LBP is small from begging to end. For three of the four features being effective, none of three methods misses the target, but the tracking performance of our method is most stable. Taking frame 292 in video "woman" as an example, we backproject exterior weight of four features into weight images, as shown in Fig. 5:

Figure 5. Weight image by backprojecting interior weight of four features in frame 292 of "woman".

From Fig. 5 (e) and (h), we can see that white points (with high weight values) concentrate on object "woman", which means that, for the features of gray level and gradient amplitude, GMM is able to find distinctive pixels on the target. As shown in Fig. 5 (f) and (g), the log likelihood function just filter out few points corresponding to background for the feature of LBP and edge orientation, which lacks recognition ability of the target. In frame 292, features of gray level, LBP, edge orientation and gradient amplitude are assigned exterior weight of 0.3953, 0.0898, 0.2266 and 0.2884 respectively. Because we have assigned high exterior weight to those discriminative features of gray level, edge orientation and gradient amplitude, and the interior weight separates the object from background for features of gray level and gradient amplitude, our method tracks the target stably on the current frame. Comparing the tracking results with the ground truth, we can see that, even though the center location error rate of each method is rather low, error rate of the method proposed in this paper is still 2.39 pixels per frame less than [2], as shown in Fig. 6:

Figure 6. Center location error rate for video sequence “woman”.

Then we check the tracking result of video sequence “sylv”, shown by

Figure 7. Tracking result of video sequence "sylv" for three methods: red rectangle for our method, green rectangle for mean shift tracking using interior weight of features, and blue rectangle for [2]. Frames 470, 858, 920, 1055 and 1124 are displayed.

As illustrated in Fig. 7, our method is able to track the target stably, while the other two failed on last frames. Exterior weight of four features are calculated, shown by

Figure 8. Exterior weights of four features in video sequence “sylv”.

As shown in Fig. 8, during the tracking process, the exterior weight of gray level is always higher than that of the others. For the most of time, gray level is the only feature that can distinguish the target well, while the other three are not so valid. Therefore, tracking without exterior weight of features fails. Taking frame 1055 as an example, weight images by backprojecting exterior weight of four features are shown as Fig. 9. In frame 1055, we see that, the target is distinguished by those weight points with high values corresponding to features except edge orientation. At this time, the exterior weights of gray level, LBP, edge orientation and gradient amplitude are 0.7781, 0.0163, 0.0944 and 0.1112 respectively, which means, in the current frame, only the feature of gray level is discriminative between object and background, and has a similar performance with the target template. In another word, the contribution of gray level feature to tracking is greatest. Compared our tracking rectangle in “sylv” with the ground truth, the center location error rate is given by Fig. 10.
As shown in Fig.10, feature selection using interior and exterior weight keeps running during the tracking process, which is able to reduce center location error rate of 15.1 pixels per frame.

IV. CONCLUSION

This paper presents a tracking algorithm using online selection of multiple features, which has a stable tracking performance. We use both interior and exterior weight to improve the discrimination between object and background. Moreover, we use features of gray level, LBP and edge orientation to describe the color, texture and structure of object, and supplement gradient amplitude feature to replenish the structure information of object, which makes tracking performance more robust.

The distribution of LBP and edge orientation for image cannot be represented by GMM, so our future work will focus on searching the discriminative part of those features and bringing more distinctive features into our tracking model.

REFERENCES

[1] D. Comaniciu and P. Meer, “Mean Shift: A Robust Approach toward Feature Space Analysis”. IEEE Trans on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619.
[2] D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-Based Object Tracking”. IEEE Trans on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-575.
[3] T. Ojala, K. Valkealahti, E. Oja and M. Pietikainen, “Texture Discrimination with Multi-Dimensional Distributions of Signed Gray Level Differences”, Pattern Recognition, 2001, 34(3): 727–739.
[4] T. Ojala, M. Pietikainen and T. Maenpaa, “Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns”, IEEE Trans on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971–987.
[5] M. Heikkila, M. Pietikainen, “A Texture-Based Method for Modeling the Background and Detecting Moving Objects”, IEEE Trans on Pattern Analysis and Machine Intelligence, 2006, 28(4): 657–662.
[6] J. Ning, L. Zhang and D. Zhang, “Robust Object Tracking Using Joint Color-Texture Histogram”, Pattern Recognition, 2009, 23(7): 1245–1263.
[7] W. Zhu, S.E.Levinson, “Edge orientation-based multiview object recognition”, ICPROO Barcelona: IEEE CS Press, 2000, 2: 1936-1939.
[8] W Liu, Y Zhang, “Edge-color-histogram and Kalman filter-based real-time object tracking”, Journal of Tsinghua University(Science and Technology), 2008, 48(7): 1104–1107.
[9] R.T. Collins, Y. Liu, M. Leordeanu, “Online selection of discriminative tracking features”, IEEE Trans on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1631–1643.
[10] Zhijun Yao, Bin Feng, Junwei Wang, Wenyu Liu. Building a Robust Appearance Model for Object Tracking. In: Proceeding of 2009 International Conference on Artificial Intelligence and Computational Intelligence, Shanghai: IEEE, 2009: 471-475.