Visual Tracking Algorithm Based on Color Name Histogram

Chenchen Yue¹,², Wangsheng Yu³,*, Zhiqiang Hou¹,² and Sugang Ma¹

¹College of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an, China
²Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an, China
³Information and Navigation Institute of Air Force Engineering University, Xi'an 710077, China

*Corresponding author

Abstract. Aimed at the problem that traditional histogram is sensitive to illumination changes in visual tracking, combined with the CN(Color Name) feature, we proposed a new feature(denotes CNH, Color Name Histogram) based on color name. Firstly, the method projected the original RGB image to CN space to obtain robust 11 feature layers. Then, we counted the each pixel numbers of feature layers. Finally, normalizing the amount of pixels in each layer. In addition, we adopted a feature adaptive fusion method to combine CNH and HOG(Histogram of Oriented Gradient). In order to prove validity of the proposed algorithm, we use Staple(Sum of Template And Pixel-wise Learners) algorithm frame to make a controlled trial. In contrast with the reference algorithms, the success of our algorithm increases by 1.5% and the precision increases by 1.7%. The results show that this method retains the advantages of traditional histogram which is insensitive to target deformation, but also enhances the robustness to illumination change.

Keywords: Visual tracking; Color histogram; Color name; Feature fusion.

1. Introduction
How to describe the image is a crucial step in image applications, and research show that image features can directly influence the algorithm performance [1]. In 2010, Blome proposed the Mosse [2]algorithm, which adopts gray-scale features to represent the target. In 2014, Henriques proposed KCF [3]algorithm based on CSK. In this algorithm, they use HOG to represent the target. In 2014, Danelljan proposed the CN [4]algorithm, which uses the color name(CN) feature [5]to represent the target. In 2016, Bertinetto used CN features and HOG features to learn two complementary filtering templates [6]. As the deep learning developed, it has made great achievements in visual tracking [7] [8] [9]. Goutam Bhat studied the influence between deep features and manual features on the algorithm in reference [10], proved they have different function for target and the deep features should be treated separately from the manual features. However, deep learning need abundant data to train the model and require a high-spec computer environment. Traditional manual features can be extracted on any existing environment with high speed and the feature is simple and intuitive. So it's very valuable to study manual features further [11].

Aiming at the problem that the traditional histogram is sensitive to illumination, we proposes CNH feature based on CN features. This feature retains the robustness of the original histogram to the target deformation, but also enhances the robustness to the illumination change. At the same time, we uses
an adaptive fusion method to fuse CNH features with the HOG. The experiment show that the performance of visual tracking algorithm has an obvious improvement.

2. Relation Work

We select classical algorithm Staple, proposed in reference[6], as our baseline. In correlation filtering framework, they combines traditional histogram and HOG features, which own the complementary attributes, to represent the target. In \( n \)th frame, we denote the location of target in image \( x \) as \( p \) to maximize the equation (1), which is chosen from the set \( S_t \):

\[
p_t = \arg \max_{p \in S_t} f\left(T(x_t, p); \theta_{t-1}\right)
\]

Making an transformation for image by function \( T \) for the \( f\left(T(x_t, p); \theta_{t-1}\right) \) to score for window \( p \) on the basis of model parameters \( \theta \) in image \( x \), in which \( \theta \) should be determined by the loss function \( L(\theta; x_t) \).

And, the loss function is decided by preceding object location in images \( x_t = \{(x_i, p_i)\}_{i=1}^n \):

\[
\theta_t = \arg \min_{\theta \in \Omega} \left\{L(\theta; x_t) + \lambda R(\theta)\right\}
\]

\( \Omega \) is space of parameter. We restrict the model complexity and avoid over-fitting by adding a regularization term \( R(\theta) \) with relative weight \( \lambda \).

Considering the real-time of tracking algorithm, the response learned from these two features with complementary attributes is linearly fused to locate the target.

3. Visual Tracking Based on Color Name Histogram

3.1. The Traditional Histogram

Traditional histogram divides pixels into several intervals according to the fixed pixel value span. The pixel value at \((x, y)\)is \(f(x, y)\), \( R \) is the range of pixels gray value, typically is 256, and \( B \) is number of intervals divided, then the pixel set of \( u \)th is:

\[
u = \left\{(x, y) \mid R \left(\frac{B - 1}{B} \right) \leq f(x, y) \leq R \times u - 1\right\}
\]

When light changes, dividing the region according to gray value will make the feature different and weaken feature robustness, which is not conducive to locate the target during visual tracking.

3.2. Color Name Histogram

When target deformed or rotated, the CN features of the original target will get completely different feature representations, which is not conducive to the target representation. CN is robust to illumination, and its color distribution is more suitable for human. Figure 1 reveals the contrast of target CN features. It can be noticeable that the CN features of same object are quite different when they are deformed. Therefore, we proposes a histogram feature of color name.
(a)

(b)

(c)

(d)

Figure 1. Comparison of CN feature before deformation ((a)-(b) is the CN feature before deformation, (c)-(d) is the CN feature after deformation).

For the input RGB image $I$, mapping the pixel at $x$ to CN space, we get a 11 dimensional probability eigenvector $a$, that is to say, the color name features of the location are as follows:

$$a = \{p(k_1|I), p(k_2|I), \ldots, p(k_{11}|I)\}$$  \hspace{1cm} (4)

in which,

$$p(k_i|I) = \frac{1}{N} \sum_{x \in I} p(k_i|I(x))$$  \hspace{1cm} (5)

$k_i (i = 1,2,..11)$ represent $i$th color attribute label, $x$ denote the pixel location, $N$ is the amount of image $I$, $p(k_i|I(x))$ represent the probability of $i$th color name label. After mapping the input RGB image to CN space, we count the layer number of each pixel. The pixel set of first color name layer denote as follow:

$$l = \{(x,y)|p(k_i|f(x,y)) \text{ is the maximum among } a\}$$  \hspace{1cm} (6)

in which, we mark the pixel probability with $p(k_i|f(x,y))$ at location $(x,y)$ mapped to CN space.

3.3. Adaptive Feature Fusion

In different scenes, the response of target location calculated by different features will have different confidence. Thus, we calculates the adaptive fusion coefficient by the peak side lobe ratio (PSR):

$$PSR = \frac{\max(f) - \mu(f)}{\sigma(f)}$$  \hspace{1cm} (7)

in which, $f$ represent response, $\max(f)$ represent the max value of response, $\mu(f)$ is the mean value of response, and $\sigma(f)$ is the mean square deviation.

Then, the fusion coefficient denote as follow:

$$w_{hog} = \frac{PSR_{hog}}{PSR_{hog} + PSR_{cnh}} \quad \text{and} \quad w_{cnh} = \frac{PSR_{cnh}}{PSR_{hog} + PSR_{cnh}}$$  \hspace{1cm} (8)
According to the PSR of responses, we determine the fusion coefficient to locate the target.

3.4. Algorithm Flow

Figure 2 is the algorithm flow chart. The top row is extracting the HOG features and another is CNH extraction process. The filtering model based on HOG feature learning is very sensitive to target deformation, but robust to motion blur and illumination change. The CNH is a global statistical feature, which is very robust to target deformation, thus we calculate the appropriate fusion coefficient according to the respective response, giving full play to the complementary advantages of the two features to locate target.

\[
\text{PSR}_{\text{PSRw}} = \frac{\text{PSR}_{\text{HOGresponse}}}{} + \frac{\text{PSR}_{\text{CNHresponse}}}{}
\]

Figure 2. Algorithm flow.

4. Experiment

So as to verify the effectiveness of our algorithm (denotes Ours), we make a comparison with five classical algorithm including original staple algorithm, DSST, Siamesefc, SAMF and CN.

4.1. Illumination Change

38 video sequences with illumination change attribute are selected from OTB100 dataset. It can be visible clearly that the success and accuracy of our algorithm increased by 0.8% and 1.2% respectively from figure 3, which is higher than original Staple algorithm. Our histogram feature is proposed based on CN feature, which retain robustness to illumination change.

4.2. Deformation

44 video sequences with target deformation attribute are selected from OTB100 dataset. We can see clearly that the success and accuracy of this algorithm increases by 1.9% and 1.7% respectively from figure 4. Our algorithm is still a global statistical feature, and retains the advantages of the original histogram to the target deformation.
4.3. Overall Performance of the Algorithm
It can be shown distinctly in figure 5, the performance of our algorithm are improved by 1.7% and 1.5% respectively in accuracy and success, which represent that the our method can achieve high performance in visual tracking.

Furthermore, we select VOT2016 standard data platform measure. Compared with the original algorithm, all indexes except FPS have been improved.

| Algorithm | Overlap | Failures | EAO   | FPS  |
|-----------|---------|----------|-------|------|
| Ours      | 0.5548  | 22.2730  | 0.2964| 9.76 |
| Staple    | 0.5400  | 22.3808  | 0.2950| 10.11|

*a The data in bold is the optimal algorithm.

5. Conclusion
In this paper, we propose an improved color name feature, and in the framework of the Staple algorithm, adopt an adaptive fusion method to fuse the response of CNH and HOG feature. The traditional histogram is a kind of global statistical feature, which has the advantages of fast and robust
to target deformation in the visual tracking. However, when the illumination changes, this method will weaken the robustness of the feature. CN space is a kind of color space which is more consistent with human, and it has strong robustness to light changes. Therefore, we proposes the CNH feature. The experimental results show that this method not only retains the advantages of traditional histogram, but also enhances the robustness to illumination. In the follow-up work, we will further optimize CNH, and consider how to integrate deep feature and manual feature, giving full play to their advantages, and improve the algorithm performance.

Acknowledgments
This research is supported by the National Natural Science Foundation of China (No.61703423, 61473309, 61403414, 41601436).

Reference
[1] Nikolova I. Performance analysis of robust image features detection algorithms[J]. Information Technologies & Control, 2014, 11(3): 2-15.
[2] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]// The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010. IEEE, 2010.
[3] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 37(3): 583-596.
[4] Danelljan M, Khan F S, Felsberg M, et al. Adaptive Color Attributes for Real-Time Visual Tracking[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014.
[5] Van de Weijer J, Schmid C, Verbeek J, et al. Learning color names for real-world applications[J]. IEEE Transactions on Image Processing, 2009, 18(7): 1512-1523.
[6] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: Complementary Learners for Real-Time Tracking [C]// CVPR, 2016.
[7] Ma C, Huang J B, Yang X, el al. Hierarchical convolutional features for visual tracking [C]//Proceedings of the IEEE international conference on computer vision. 2015:3074-3082.
[8] Valmadre J, Bertinetto L, Henriques J, et al. End-to-end representation learning for correlation filter based tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2805-2813.
[9] Danelljan M, Bhat G, Khan F S, et al. ECO: efficient convolution operators for tracking[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA: IEEE, 2017: 6931-6939.
[10] Bhat G, Johander J, Danelljan M, et al. Unveiling the power of deep tracking. arXiv preprint arXiv:1804.06833, 2018.
[11] Yue C C, Hou Z Q, Yu W S, Ma S G. Hierarchical structure histogram combined with color name [J]. Journal of Image and Graphics, 2019, 24(07) : 1188-1196.