Robust visual tracking with channel weighted color ratio feature

Shan Jiang
University of Chinese Academy of Sciences
Institute of Automation, Chinese Academy of Sciences
Beijing, China
jiangshan2017@ia.ac.cn

Shuxiao Li, Chengfei Zhu, Xiaosong Lan
Institute of Automation, Chinese Academy of Sciences
Beijing, China
{shuxiao.li, chengfei.zhu, lanxiaosong2012}@ia.ac.cn

***Abstract*—Robust visual tracking is an important and challenging problem due to various challenging factors and computational constraints. Recent studies have shown taking advantage of color information is a simple and effective way to improve correlation-based tracker performance. In this paper, we propose a 1-channel color feature called color ratio (CR) feature inspired by mean-shift-based tracking algorithms, which is more efficient and effective than currently widely used 10-channel color-naming features. We then concatenate 1-channel CR, 13-channel HOG and 1-channel gray together to get totally 15-channel features for efficient DCF tracking. During feature concatenation process, we find that weighting between different feature channels can improve the tracking performance notably. Finally, correlation-based responses and CR-based responses are fused to further boost tracker robustness. Experimental results demonstrate that our feature and fusion strategy can achieve superior performance while attaining real-time performance.**

Keywords-visual tracking, correlation filter, color ratio feature, channel weighting, feature fusion

I. INTRODUCTION

Generic single object visual tracking is to estimate the target state in a video given only the initial state in the first frame. It has a widespread applications such as robotics, UAV monitoring and human-computer interaction. In most applications, visual tracking can be considered as an online complement of object detection. It has many challenging factors such as illumination variation, motion blur, occlusion and background clutters [1]. Extensive research has been performed in this field and considerable progress has been made in the past decade. However, robust and efficient tracking algorithm remains an open problem.

Although in recent years, deep learning based trackers have gained outstanding performance in various tracker benchmarks [2], correlation filter based trackers with handcrafted features still play an important role in vision applications, for their real-time accurate performance and low computational cost. Correlation filters are first introduced to visual tracking by Bolme *et al.* [3]. They are further extended by Henriques *et al.* with kernel trick [4] and multi-channel HOG features [5] to achieve state-of-the-art performance with high speed. Danelljan *et al.* [6] and Li *et al.* [7] enable scale estimation by searching the optimal scale in a multi-scale pyramid. To enhance tracker robustness, some works focus on boundary effects [8-10], other works model the target by parts [11, 12] at the cost of tracking speed. On the other hand, Danelljan *et al.* [13] propose color-naming (CN) feature to incorporate into feature channels and Bertinetto *et al.* [14] fuse the correlation filter response with a dense color histogram response. The performance of these trackers demonstrate that taking advantage of color information is a simple and effective way to improve tracker robustness.

Most current correlation based trackers employ 42-channel features (31-channel HOG, 10-channel CN, and 1-channel gray), which limits the algorithm efficiency to some extent. In early mean-shift based trackers, target model and target candidate are modelled by color weighted histograms to get reliable weight image, which reflects the existence of the tracked object at each pixel. In this paper, we try to borrow the essential idea from mean-shift-based trackers to get a more compact color feature representation, thus reducing the required feature channels for DCF tracker. Our main contributions can be summarized as follows:

(1)We borrow the essential idea from mean-shift based trackers and propose a 1-channel color feature called color ratio (CR) feature, which is an alternative to CN [13] feature to incorporate into feature channels. Experiments have shown this feature is more compact and discriminative than CN feature.

(2)We use an additional weight for different types of features in multi-channel feature integration and further exert the discrimination power of each type of features in the total feature representation.

II. OUR APPROACH

In this section, we first describe our color ratio feature in details and discuss the manner to incorporate the feature into correlation filter tracking framework. Next, channel weighted feature integration strategy is discussed. Finally we present the overall process of our tracker and describe some details.

A. Color ratio feature

In early mean-shift based trackers, Li *et al.* [15] employ spatial context information and global tracking skills to generate reliable weight image and achieved superior reliability and accuracy to classical mean-shift trackers. The weight image is generated by comparison between target model and target candidate to give a high weight on the target and a low weight on the background. The target

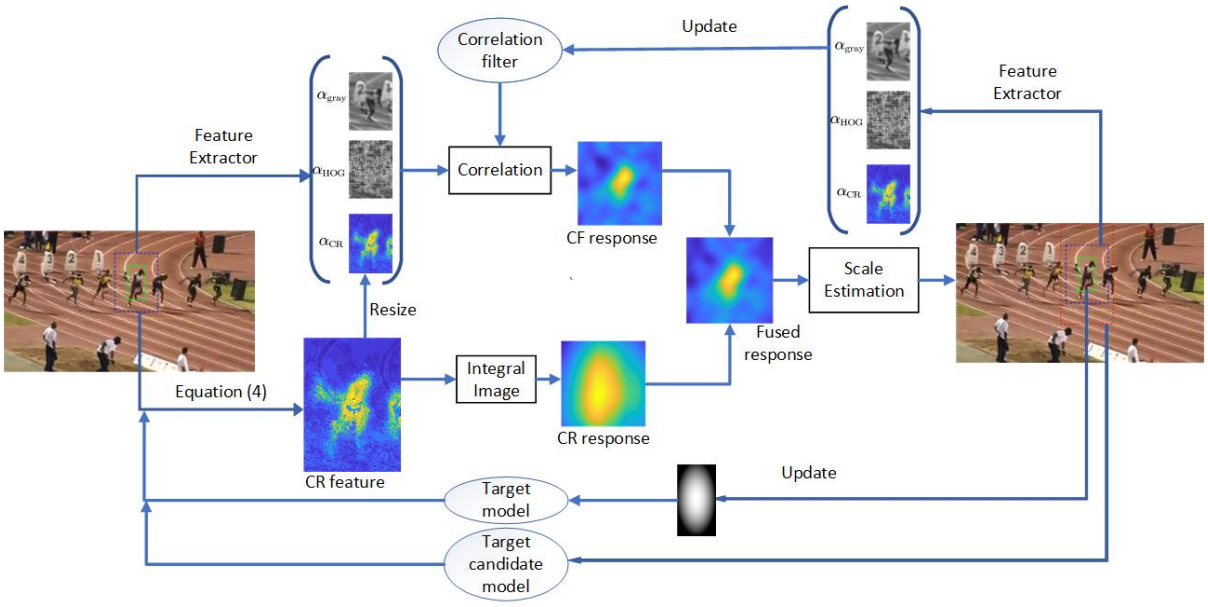


Figure 1 Visualization of overall procedure of our tracking approach

model is represented by a color histogram of $16\times16\times16$ bins in RGB space, which is computed from the target region. Let $\{\mathbf{x}_i\}_{i=1,\ldots,n}$ denote pixel locations centered at target location $\mathbf{x}_0$ within bandwidth h, then the target model is computed as

$$q[u]=C\sum_{i=1}^{n}k\left(\left\|\frac{\mathbf{x}_i-\mathbf{x}_0}{h}\right\|^2\right)\delta[b(\mathbf{x}_i)-u],\ u=1,2,\ldots,m \tag{1}$$

Here, m is the number of bins which equals to 4096. C is the normalization factor to ensure the probabilities sum up to 1. $k(\cdot)$ is the kernel profile function. This function is adopted to weight over pixel locations to reduce background distraction on the edges of target region and Epanechnikov profile is adopted in our paper. δ is Kronecker delta function and $b(\mathbf{x}_i)$ is the serial number of histogram bin for $\mathbf{x}_i$. For the target candidate, the region is enlarged with a scale factor s. Let $\{\mathbf{x}_i\}_{i=1,\ldots,n_s}$ denote pixel locations of the target candidate centered at target location $\mathbf{x}_0$ within bandwidth $s\cdot h$, the target candidate model is then computed as

$$p_s[u]=C_s\sum_{i=1}^{n_s}k\left(\left\|\frac{\mathbf{x}_i-\mathbf{x}_0}{s\cdot h}\right\|^2\right)\delta[b(\mathbf{x}_i)-u],\ u=1,2,\ldots,m \tag{2}$$

where C_s is the normalization factor. Here uniform profile is adopted for $k(\cdot)$, which is equivalent to unweighted histogram. This selection is for two reasons. First, it has been proved in [15] that a less precise candidate model is preferred in generating a more reliable weight image. Secondly, this is more computationally efficient. Then the weight for a given color index u can be computed by

$$w[u]=\sqrt{\frac{q[u]}{p_s[u]}} \tag{3}$$

For arbitrary colors the weight are approximately bounded by $[0,\ s]$. Therefore, the weight to get a novel color feature for DCF tracking can be obtained by dividing with scale factor s

$$CR_s[u]=\frac{1}{s}\sqrt{\frac{q[u]}{p_s[u]}} \tag{4}$$

Thus, the novel color feature for each pixel in an image patch can be computed according to (4) by finding its color index u. This feature image is a more compact color feature representation to be integrated to DCF feature channels and we call it color ratio (CR) feature.

To handle scale and illumination changes in the tracking process, the target model and target candidate model are updated linearly,

$$q^t=(1-\eta_{CR})q^{t-1}+\eta_{CR}q \tag{5}$$

$$p_s^t=(1-\eta_{CR})p_s^{t-1}+\eta_{CR}p_s \tag{6}$$

where η_{CR} is the model learning rate.

On the other hand, it has been proved in [14] that fusing deformation invariant color histogram response with template based correlation filter response is a simple and effective way to improve tracker performance. Since the value of CR feature is approximately bounded by [0, 1] and can reflect the target existence at each pixel, this feature representation can also be utilized to compute dense color response in addition to being integrated to feature channels.

B. Channel weighted feature

In our approach, gray intensity, HOG and CR features are employed. Note that gray intensity feature and CR feature are of 1 channel and HOG feature has much more channels than the former two features. Intuitively, the importance of each type of feature in the final feature representation should not be decided simply by channel number. Therefore, we use an additional weight for different types of features to control their importance in multi-channel feature representation, and the feature representation can be summarized as following.

$$\{\alpha_{\text{gray}} f_{\text{gray}}, \alpha_{\text{HOG}} f_{\text{HOG}}, \alpha_{\text{CR}} f_{\text{CR}}\} \tag{7}$$

where α_{gray}, α_{HOG} and α_{CR} are the weights for each type of features and they sum up to 1.

The HOG feature in most current correlation filter trackers is 31 channel HOG feature proposed in [16]. However, we find that an alternative 13 channel HOG feature in [16] can also be employed in visual tracking, with a minor loss of performance and a 58% reduction of computational burden. This 13 channel feature is an analytic dimensionality reduction of 36 channel HOG feature proposed in [17], which is an approximation of PCA. Therefore, the final feature representation is of 15 channels.

C. CRCF tracker

The overall procedure of our CRCF tracker is visualized in Figure 1. The CR feature map is simultaneously utilized to compute dense CR-based response and resized to be integrated to feature channels of the correlation filter. The correlation filter response and CR-based response are fused linearly and the target translation is estimated from the fused response.

$$R = (1-\gamma) R_{cf} + \gamma R_{CR} \tag{8}$$

where γ is the response fusing factor.

For scale estimation, we choose to follow [6] to learn a one-dimensional correlation filter on scale pyramid to perform scale search after translation search and only HOG feature is employed in the scale filter. For translation search, compared to padding the target region with a fixed ratio to get the sample patch, padding the target region equally on width and height following the manner of [14] should be a better choice. This is because the target has an equal possibility to translate in any direction and the padding of the sample patch should not be determined by the target aspect ratio. In addition, we normalize the sample patch to a predefined area to balance the computational burden and performance on large and small targets.

III. EXPERIMENT

A. Experiment setup

We implemented our tracker with MATLAB and mex and evaluate our approach on OTB-2015 [18] which contains 100 sequences with various challenges. All the experiments are conducted on an Intel i7-6700 CPU (3.4GHz) PC with 4 GB memory. The correlation filter learning rate η_{cf} is set to 0.01 and the CR model learning rate η_{CR} is set to 0.04. The padding of sample patch is set to $(m+n)/2$ where m and n are the target width and height. The sample patches are normalized to a predefined area of 150^2. The bandwidth σ for desired Gaussian response is set to $mn/16$. The response fusing factor γ is set to 0.3 and the regularization factor λ is set to 10^{-3}.

OTB-2015 uses precision plot and success plot to evaluate trackers. The precision plot measures the ratio of the frames whose center location error are under a series of threshold. The success plot measures the ratio of the successful frames whose overlap are over a series of threshold. In precision plot, the precision at the threshold of 20 pixels is regarded as the representative precision score while in success plot, AUC (area under curve) is commonly used to rank trackers.

B. Comparison between CN and CR feature

To validate the discriminative power of our CR feature representation, we replace the CN feature in the tracker proposed in [13] with our proposed CR feature and compare the performances.

Table I summarizes our experimental results. The CN tracker employs gray and CN_2 feature, which is a PCA compression of 10 channel CN feature and CN10 tracker employs gray and 10 channel CN feature. CR2, CR3, CR4 tracker are the trackers employing CR feature with context factor $s = 2, 3, 4$ respectively. We can see from the results that CR4 and CR3 tracker have a superior performance to CN feature trackers with only one channel color feature while CR2 has only a slightly inferior performance to CN tracker. This shows our CR feature is a compact and discriminative feature representation to incorporate to correlation filer tracking. From the above, we choose CR3 feature to balance the tracker performance and efficiency.

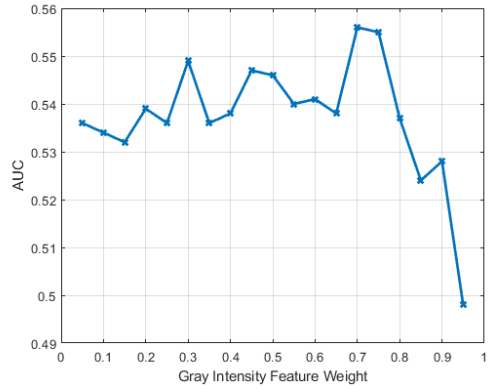
TABLE I. PRECISION AND AUC SCORE OF CN TRACKER AND TRACKERS WITH CR FEATURE

Method	Precision(20px)	AUC
CN	0.574	0.411
CN10	0.598	0.430
CR2	0.570	0.407
CR3	0.616	0.439
CR4	**0.626**	**0.450**

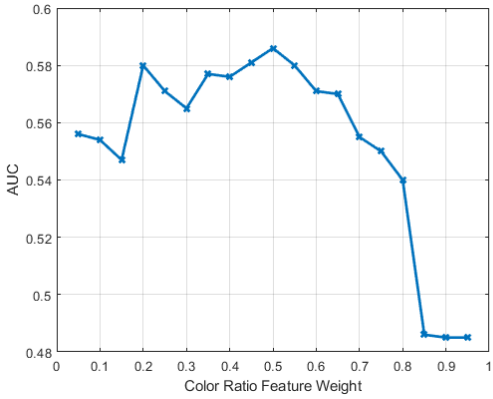
C. Experiments of channel weighted feature

To validate the effect of channel weighted feature representation, we first conduct experiments on the feature weight between gray intensity feature and HOG feature, then on the weight of CR feature.

Experiment on gray intensity feature weight is conducted using our CRCF tracker without CR feature channel. As is shown in Figure 2, the tracker gains the best performance with a gray intensity weight of 0.7. Experiment on CR feature weight is conducted with this gray feature weight and the best performance is achieve with the weight of 0.5. Therefore, the optimal weight for gray, HOG and CR feature are $\alpha_{\text{gray}} = 0.35$, $\alpha_{\text{HOG}} = 0.15$, $\alpha_{\text{CR}} = 0.5$ respectively. With this channel weighted feature representation, our tracker have a 6.9% gain in precision score and a 4.6% gain in AUC score compared with the tracker without channel weighting. This shows our channel weighted feature is a better feature integration strategy to improve tracker performance.



(a)



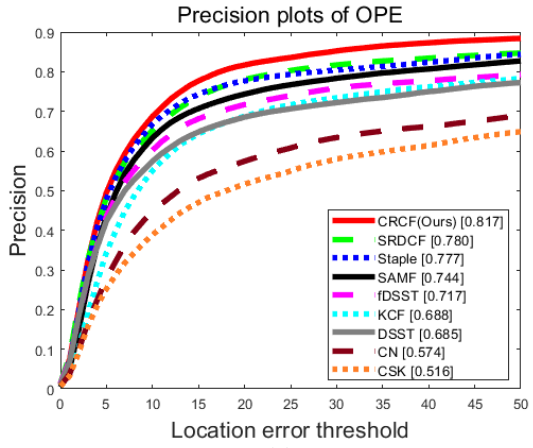
(b)

Figure 2 AUC in relation to gray intensity feature weight (a) and CR feature weight (b)

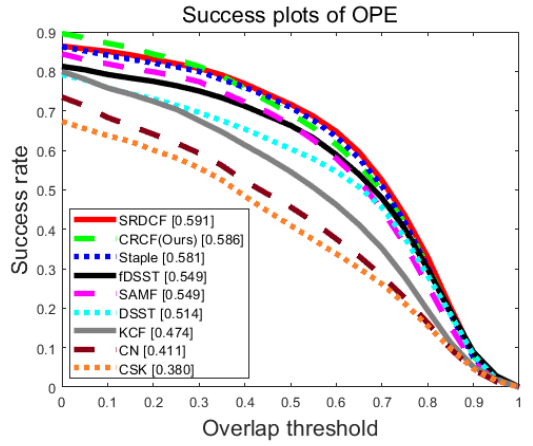
D. Comparison with other correlation filter trackers

To validate our approach, we compare our tracker with some representative correlation filter trackers, including CSK [4], CN [13], KCF [5], DSST [6], SAMF [7], fDSST [19], Staple [14] and SRDCF [8]. The one-pass evaluation (OPE) is employed in our experiment.

As is shown in Figure 3 and table II, our proposed tracker has the best performance on the precision plot and the second best performance on the success plot. It should be noted that our proposed tracker outperforms SRDCF by 4.75% in location precision, which indicates that our tracker has less tracking failures and better robustness. Furthermore, our tracker is over 10 times faster than SRDCF and slightly faster than Staple without optimization thanks to using more compacted CR and HOG features. The speed can be further boosted with optimized CR and HOG feature extractor.



(a)



(b)

Figure 3 Precision and success plots of our tracker and other correlation filter trackers on OTB-2015

TABLE II. PERFORMANCE COMPARISON BETWEEN OUR TRACKER AND OTHER CORRELATION FILTER TRACKERS. THE BEST PERFORMANCES ARE SHOWN IN BOLD AND THE SECOND BEST ARE SHOWN UNDERLINED. THE FPS OF THE TRACKERS WITH REAL-TIME PERFORMANCE ARE SHOWN IN BLUE.

Method	Precision(20px)	AUC	FPS
CRCF(Ours)	**0.817**	0.586	86.6
SRDCF	0.780	**0.591**	6.42
Staple	0.777	0.581	85.5
SAMF	0.744	0.549	25.6
fDSST	0.717	0.549	115
DSST	0.685	0.514	50.9
KCF	0.688	0.474	296
CN	0.574	0.411	273
CSK	0.516	0.380	528

IV. Conclusion

In this paper, we borrow ideas from mean-shift based trackers and propose a more compact color feature representation and a channel weighted feature integration strategy. Experimental results demonstrate that these strategies improve tracker performance by a large margin. This indicates that taking advantage of color information is vital in visual tracking and better approach should be investigated in the future.

Acknowledgment

This work is supported by the National Science Foundation of China (NSFC) with granting No. 61573350.

References

[1] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2013.

[2] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pfugfelder, L. C.Zajc, T. Vojir, G. Bhat, A. Lukezic, A. Eldesokey, G. Fernandez, and et al., "The sixth visual object tracking vot2018 challenge results," 2018.

[3] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Computer Vision & Pattern Recognition*, 2010.

[4] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," *in proceedings of the European Conference on Computer Vision*, 2012.

[5] J.F.Henriques, R. Caseiro, P. Martins, and J.Batista, "High-speed tracking with kernelized correlation filters," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2015.

[6] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," 2014.

[7] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," pp. 254–265, 2014.

[8] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," *international conference on computer vision*, pp. 4310–4318, 2015.

[9] H. K. Galoogahi, A. Fagg, and S. Lucey, "Learning backgroundaware correlation filters for visual tracking," *international conference on computer vision*, pp. 1144–1152, 2017.

[10] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," *computer vision and pattern recognition*, pp. 4847–4856, 2017.

[11] T. Liu, G. Wang, and Q. Yang, "Real-time part-based visual tracking via adaptive correlation filters," pp. 4902–4912, 2015.

[12] Y. Li, J. Zhu, and S. C. H. Hoi, "Reliable patch trackers: Robust visual tracking by exploiting reliable patches," pp. 353–361, 2015.

[13] M. Danelljan, F. S. Khan, M. Felsberg, and J. V. De Weijer, "Adaptive color attributes for real-time visual tracking," pp. 1090–1097, 2014.

[14] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, "Staple: Complementary learners for real-time tracking," *computer vision and pattern recognition*, pp. 1401–1409, 2016.

[15] S. Li, O. Wu, C. Zhu, and H. Chang, "Visual object tracking using spatial context information and global tracking skills," *Computer Vision and Image Understanding*, vol. 125, pp. 1–15, 2014.

[16] P. F. Felzenszwalb, R. B. Girshick, D. A. Mcallester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," vol. 1, pp. 886–893, 2005.

[18] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 9, pp. 1834–1848, 2015.

[19] M. Danelljan, G. Hger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 8, pp. 1561–1575, 2017.

Authors' background (This form is only for submitted manuscript for review)

Your Name	Title*	Affiliation	Research Field	Personal website
Shan Jiang	Master Student	University of Chinese Academy of Sciences	Visual tracking	
Shuxiao Li	Associate Professor	Institute of Automation, Chinese Academy of Sciences	Computer Vision UAV Vision	Personal Website
Chengfei Zhu	Associate Professor	Institute of Automation, Chinese Academy of Sciences	Computer Vision UAV Vision	
Xiaosong Lan	Assistant Professor	Institute of Automation, Chinese Academy of Sciences	Computer Vision UAV Vision	

*This form helps us to understand your paper better, the form itself will not be published.

*Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor