


Article

Visual Object Tracking Robust to Illumination Variation Based on Hyperline Clustering

Senquan Yang ^{1,2} , Yuan Xie ¹, Pu Li ^{2,*}, Haoxiang Wen ², Huan Luo ^{1,2} and Zhaoshui He ¹

¹ School of Automation, Guangdong University of Technology, Guangzhou 510006, China; sqyangg@sgu.edu.cn (S.Y.); yuanxiemath@163.com (Y.X.); lh525785441@163.com (H.L.); zhshhe@gdut.edu.cn (Z.H.)

² School of Physics and Electromechanical Engineering, Shaoguan University, Shaoguan 512026, China; xiangxiang_0@163.com

* Correspondence: lipu_300@sgu.edu.cn; Tel.: +86-0751-8120085

Received: 7 December 2018; Accepted: 10 January 2019; Published: 14 January 2019



Abstract: Color histogram-based trackers have obtained excellent performance against many challenging situations. However, since the appearance of color is sensitive to illumination, they tend to achieve lower accuracy when illumination is severely variant throughout a sequence. To overcome this limitation, we propose a novel hyperline clustering based discriminant model, an illumination invariant model that is able to distinguish the object from its surrounding background. Furthermore, we exploit this model and propose an anchor based scale estimation to cope with shape deformation and scale variation. Numerous experiments on recent online tracking benchmark datasets demonstrate that our approach achieve favorable performance compared with several state-of-the-art tracking algorithms. In particular, our approach achieves higher accuracy than comparative methods in the illumination variant and shape deformation challenging situations.

Keywords: visual tracking; hyperline clustering; illumination variation; discriminant model; scale estimation

1. Introduction

Visual object tracking, which aims at estimating locations of a target object in an image sequence, is an important problem in computer vision. It plays a critical role in many applications, such as visual surveillance, robot navigation, activity recognition, intelligent user interfaces, and sensor networks [1–5]. Despite significant progress has been made in recent years, it is still a challenging problem to develop a robust tracker for complex scenes due to appearance changes caused by partial occlusions, background clutter, shape deformation, illumination changes, and other variations.

For visual tracking, an appearance model based on feature is of prime importance for representing and locating the object of interest in each frame. Most state-of-the-art trackers rely on different features such as color [6,7], intensity [8,9], texture [1], Haar feature [10,11], and HOG feature [12,13]. Color feature is insensitive to shape variation and robust to object deformation. Numerous effective color-based representation schemes have been proposed for robust visual tracking. One common method is to adopt color statistics as an appearance description. Color histogram is the most commonly used descriptor representing object [14]. Distractor-Aware Tracker (DAT) uses the color histograms of distractor to distinguish object and background pixels [7]. Another successful method is to transform color space. Adaptive Color attributes Tracker (ACT) [6], on the other hand, maps the RGB value of pixel to a probabilistic 11 dimensional color representation and learns a kernelized classifier to locate the target using multi-dimensional color feature.

Although these color-based trackers have performed state-of-the-art results on recent tracking benchmark datasets [15], they fail to cope with the scene where color features vary significantly,

particularly in illumination changes. To solve this problem, we propose an illumination invariant Hyperline Clustering-based Tracker (HCT). The main components of the proposed HCT are shown in Figure 1. We exploit the observation that the color distributions of the same object under different illuminations locate in an identical line [16]. Using an hyperline clustering algorithm which is able to identify the direction vectors of all hyperlines, we can track the object throughout the sequence where the illumination is not consistent. Moreover, to distinguish the object pixels from surrounding background regions, a Bayes classifier is trained to suppress the background hyperlines, reducing the drift problem. Due to the favorable robustness of the proposed approach, it is well suited for illumination vary scenes such as *Singer1*, *Singer2*, *Trans*, and so on.

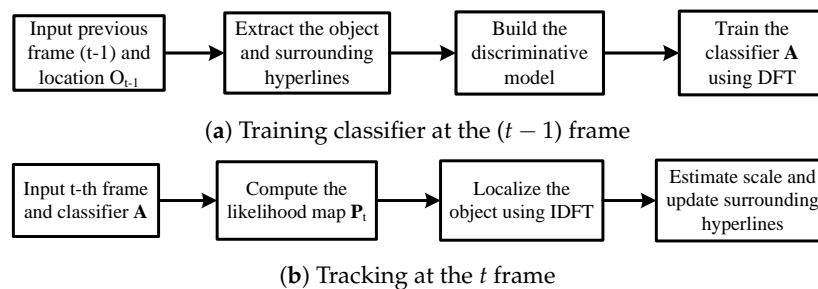


Figure 1. Main components of the proposed hyperline clustering-based tracker (HCT).

The contributions of this paper are as follows. First, we present a light yet discriminative object observation model in which the representation of object is formulated as the direction of hyperlines. Although it relies on direction initialization, this representation is able to distinguish the object of interest from background and achieves competitive performance on many challenging tracking sequences. Second, a Bayes classifier is trained in advance to identify and suppress the the background hyperlines, which improve the tracking robustness. Third, we adopt an anchor box scale estimation which allows us to cope with large variations of target scale and appearance (i.e., *Trans* sequence). Finally, we evaluate our approach on multiple tracking benchmarks demonstrating favorable results against several state-of-the-art trackers.

Notation: A boldface capital letter Y denotes a matrix and a boldface lowercase letter y a vector. $\|Y\|$ denotes the Euclid norm of Y . The transpose and complex conjugate are denoted by Y^T and Y^* , respectively. The inner product is denoted by $\langle \cdot, \cdot \rangle$. The element-wise product is denoted by \odot . \mathcal{F} denotes Discrete Fourier Transform (DFT) and \mathcal{F}^{-1} denotes the Inverse Discrete Fourier Transform (IDFT).

2. Related Work

Based on object appearance models, visual tracking approaches can be classified into two families: Generative and discriminative approaches. Generative approaches tackle the tracking problem by searching the image region that is the most similar to the target template. Such trackers either rely on templates or subspace models. Comaniciu et al. [17] present a histogram-based generative model with attraction of the local maxima to handle appearance change of object. Sevilla-Lara et al. [18] present a distribution field based generative tracking method. Shen et al. [19] propose a generalized Kernel-based mean shift tracker whose template model can be built from a single image and adaptively updated during tracking. In [20], a sparse representation based generative model is adopted to locate the target using a sparse linear combination of the templates. Although it is robust to various occlusions, a lot of manipulations of time-consuming sparse representation leads to low frame-rate. Additionally, these generative approaches fail to use the background information which is likely to alleviate drifts and improve the tracking accuracy.

Discriminative approaches typically train a classifier to separate the target object from background. For example, Zhang et al. [10] train a naive Bayes to locate the target in compressive projection where the features of the target appearance are efficiently extracted. Liu et al. [21] propose a robust tracking algorithm using sparse representation based voting map and sparsity constraint regularized mean shift. Yang et al. [22] present a discriminative appearance model based on superpixels, which is able to distinguish the target from the background with mid-level cues.

Recent benchmark evaluations [15,23,24] have demonstrated that Discriminative Correlation Filter (DCF) based visual tracking approaches achieve state-of-the-art results while operating at real-time [25]. Circulant Structure with Kernels (CSK) tracker employs a dense sampling strategy while exploiting circulant structure with Fast Fourier Transform to learn and track the object [26]. Its extension, called kernelized correlation filter (KCF) [12], incorporate multi-channel features via a linear kernel achieving excellent performance while running at more than 100 frames per second. However, the standard KCF is only robust to linearly scale changes. This implies inferior performance when the target encounters with large scale variations. To address this problem, Danelljan et al. [25] propose Discriminative Scale Space Tracking (DSST), which is capable of learning explicit scale filter using variant scales of target appearance. Despite of its competitive performance and efficient implementation, DSST starts to drift from object that is non-rigidly deforming. To further improve the robustness to deformation of tracker, both Distractor-Aware Tracker (DAT) [7] and Adaptive Color attributes Tracker (ACT) [6] adopt color-based representation that is invariant to significant shape deformation. Sum of Template And Pixel-wise LEarners (Staple) [14] combines the template to discriminate the object and the color-based model to cope with deformation in a ridge regression framework, outperforming many sophisticated trackers. However, color distribution is sensitive to varying illumination. Thus, the color-based trackers are likely to drift when the illumination significantly changes throughout a sequence.

3. Hyperline Clustering-Based Tracking

The proposed Hyperline Clustering-based Tracking (HCT) is motivated by the observation that the RGB values' distribution of the same kind color under different illuminations locate in the same lines [16]. Thus, the representation of object can be cast to hyperline clustering problem as shown in Section 3.1. Furthermore, a Bayes classifier based discriminative model, which is capable of separating the target from background, is proposed in Section 3.2. Section 3.3 demonstrates the capability of accurate object localization and update. Inspired by recent state-of-the-art object detection [27,28], we propose a Anchor Box based scale estimation to achieve accurate tracking, as described in Section 3.4.

3.1. Hyperline Clustering Representation

Hyperline clustering have been successfully applied in sparse component analysis [29] and image segmentation [16]. Given a set of observed data points $\{y_i\}_{i=1}^T$, which respectively locate on K hyperlines $L(l_k)$, where l_k is the directional vector of the corresponding hyperline and $k = 1, \dots, K$ (see Figure 2). K-HLC aims to estimate K hyperlines $L(l_k)$ by identifying $l_k, k = 1, \dots, K$.

Mathematically, K-HLC can be cast into the following optimization problem [30]:

$$\begin{aligned} & \min_{l_k, \Omega_k, k=1, \dots, K} J(l_k, \Omega_k, k=1, \dots, K) \\ &= \min_{l_k, \Omega_k, k=1, \dots, K} \sum_{i=1}^T \sum_{k=1}^K d^2(y_i, l_k) \times I_{i \in \Omega_k}, \end{aligned} \quad (1)$$

where $\mathbf{l}_k = [l_{k1}, \dots, l_{km}]^T \in \mathbb{R}^m$ is the directional vector of hyperline $L(\mathbf{l}_k)$. The indicator function $I_{i \in \Omega_k}$ is given by

$$I_{i \in \Omega_k} = \begin{cases} 1, & i \in \Omega_k \\ 0, & i \notin \Omega_k \end{cases}, \quad (2)$$

where Ω_k denotes the k -th cluster set. The distance $d(\mathbf{y}, \mathbf{l})$ from \mathbf{y} to $L(\mathbf{l}_k)$ is

$$d(\mathbf{y}, \mathbf{l}) = \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle - \frac{\langle \mathbf{y}, \mathbf{l} \rangle^2}{\langle \mathbf{l}, \mathbf{l} \rangle}}. \quad (3)$$

A robust K-hyperline clustering algorithm was proposed in [30], where it is implemented in a similar way to K-means clustering by two steps after initialization: The cluster assignment and the cluster centroid update. For the cluster assignment step, the observed data $\{\mathbf{y}_i\}_{i=1}^T$ are assigned to Ω_k by choosing k such that

$$k = \arg \min_{j=1, \dots, K} d^2(\mathbf{y}_i, \mathbf{l}_j). \quad (4)$$

For the second step, the cluster centroid is obtained by eigenvalue decomposition (EVD).

In RGB color model, three primary colors (red, green, and blue) are exploited together to reproduce a array of color [31]. Thus, the RGB vector of a color pixel can be represented by

$$\mathbf{y} = [r, g, b]^T. \quad (5)$$

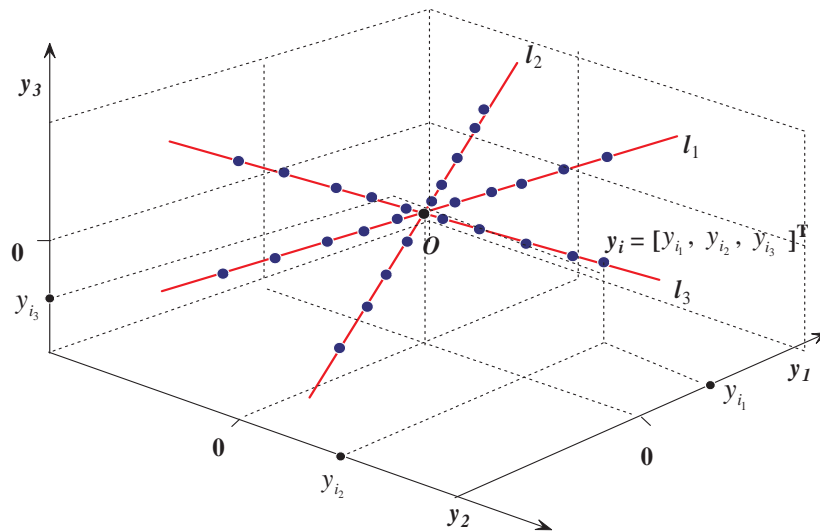


Figure 2. Hyperline clustering, observed data points $\{\mathbf{y}_i\}_{i=1}^T$ (blue points), the hidden hyperlines (red lines).

To illustrate the color represent ability of hyperline, we perform a scatter plot of three regions of a image from Trans sequences (see Figure 3). The foreground and two background regions are represented by yellow, blue, and black arrow lines, respectively. As we can observe, distributions of the pixels from three different regions approximatively locate in three different hyperlines. Hence, the representation of color can be considered as hyperline clustering problem. Moreover, when suffered illumination change, the distribution of the same color still locates in the same hyperline (see the red arrow).

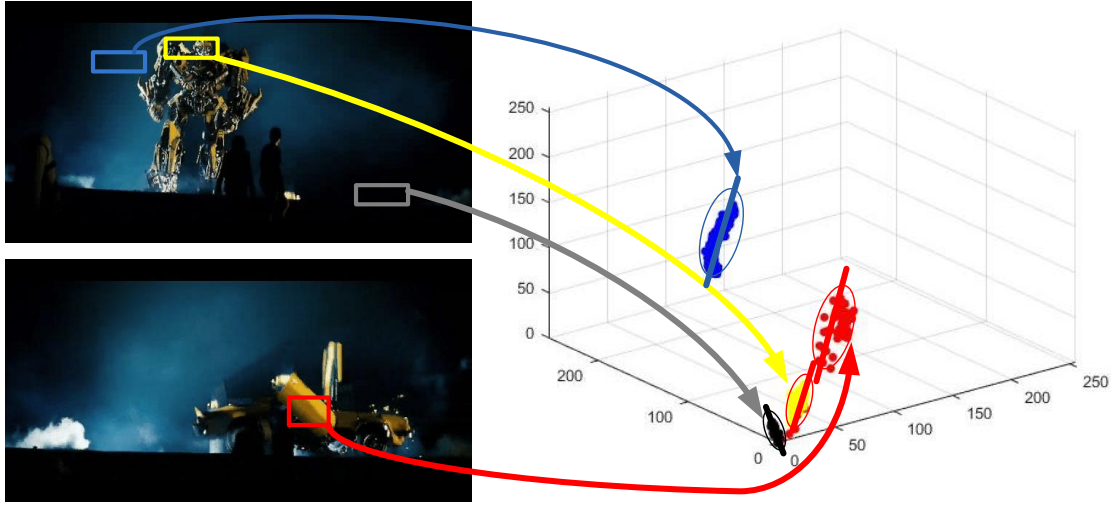


Figure 3. Scatter plot of three different regions, which are represented by blue yellow and black arrow respectively. The red arrow represents the foreground region in different illumination. The directional vectors of hyperlines that represent the foreground in different illuminations are approximately the same, as shown by the red lines.

3.2. Discriminative Model

To build a discriminative model which is capable of distinguishing the object from the surrounding, we propose a hyperline clustering based Bayes classifier for visual tracking. Let x denote the object pixels in a rectangular object region O and S denote the surrounding region of object. Additionally, let $b_{I,k}^x$ denote pixel x assigned to the k -th hyperline of image I . Thus, we formulate the object likelihood at location x by

$$P(x \in O|O, S, b) \approx \frac{P(b_{O,k}^x|x \in O)P(x \in O)}{\sum_{\Omega \in \{O,S\}} P(b_{\Omega,k}^x|x \in \Omega)P(x \in \Omega)}. \quad (6)$$

Particularly, the likelihood terms are estimated directly from the distance using Equation (3), i.e., $P(b_{O,k}^x|x \in O) \approx d(x, l_{ko})/|O|$ and $P(b_{S,k}^x|x \in S) \approx d(x, l_{ks})/|S|$, where $|\cdot|$ denotes the cardinality; ko and ks represent the ko -th and ks -th of hyperline of image O and S , respectively. Besides, the prior probability can be approximated as $P(x \in O) \approx |O|/(|O| + |S|)$. Then, Equation (6) can be simplified as follows:

$$P(x \in O|O, S, b) = \frac{d(x, l_{ko})}{d(x, l_{ko}) + d(x, l_{ks})}. \quad (7)$$

Applying the model (7), we are able to distinguish object pixels from background region, as illustrated in Figure 4. The proposed model is capable of eliminating the effect of background and reducing the risk of drifting. In addition, because model (7) estimates the likelihood using the hyperline directional vector of object and background regions, it requires fewer memory to obtain accurate estimation than other color-based algorithms.

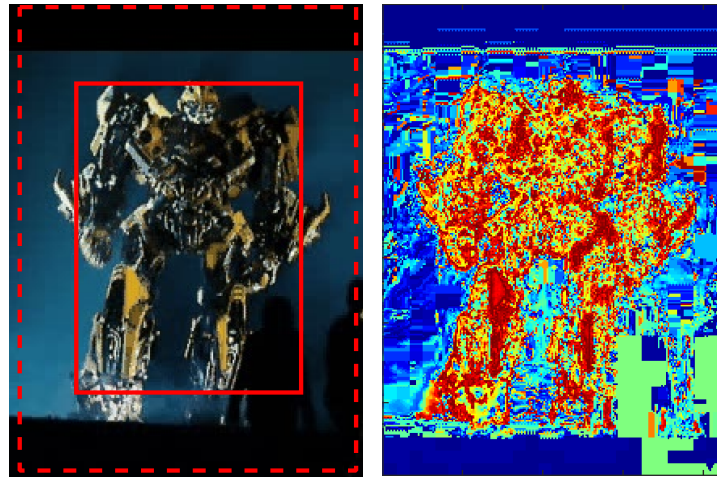


Figure 4. Exemplary object likelihood map for the discriminant model illustrating the object region O and surrounding region S .

However, it is dangerous to learn the model directly from the first frame image regions. To adaptively represent the changing object appearance and capture the object in different illuminations, we develop an update scheme in which the object and surrounding image hyperlines are updated independently. Because the color distribution discards the spatial position of image pixels, the proposed object hyperline representation is robust to shape deformation. Thus, the object hyperlines are fixed during the tracking process. For the surrounding hyperlines, the update scheme is summarized Algorithm 1.

Algorithm 1 The update scheme of surrounding hyperlines.

Require: Surrounding image pixels $\{y_i\}_{i=1}^T$, the surrounding hyperlines $L_S(l_k)$ and the number of hyperline K .

- 1: Assign image pixels $\{y_i\}_{i=1}^T$ to K different clustering sets $\Omega_k, k = 1, \dots, K$. Data index $i \in \Omega_k$ if and only if it satisfies $k = \operatorname{argmin}_{i=1, \dots, K} \{d^2(y, l_i)\}$.
- 2: Obtain a set of submatrices $Y_k = [y_1^k, \dots, y_{T_k}^k]$, where \tilde{T}_k data $y_1^k, \dots, y_{T_k}^k$ are assigned to Ω_k .
- 3: Compute the first eigenvector u_{1k} of each matrix $Y_k \cdot (Y_k)^\top / T, k = 1, \dots, K$.
- 4: Update direction vectors l_k using $u_{1k}, k = 1, \dots, K$.

Ensure: The updated surrounding hyperlines $L_S(l_k)$.

3.3. Localization

Similar to the state-of-the-art trackers adapting tracking-by-detection principle [6,7,12], we iteratively localize the object in a new frame after initializing the tracker in the first frame.

At the frame t , we use a trained classifier to predict the object location O_t basing on the previous object location O_{t-1} . Instead of utilizing representation of gray [19] or HOG [12], we perform correlation filtering on the likelihood map directly. The training of classifier is achieved by find a function $f(x)$ that minimizes the squared error

$$\min_w \sum_i (f(s_i) - g_i)^2 + \lambda \|w\|^2, \quad (8)$$

where s are the sample likelihood patches, which are cyclic shifts of previous location O_{t-1} , g are the regression targets and λ is a regularization constant. The work of [12] shows that the Ridge Regression (8) can be simplified by calculating

$$w = \sum_i \alpha_i \varphi(s_i), \quad (9)$$

where φ is the Hilbert space mapping, which is induced by the Gaussian kernel κ . Let $\alpha = [\alpha_1, \dots, \alpha_s]$, then we derive

$$A = \mathcal{F}(\alpha) = \frac{\mathcal{F}(g)}{\mathcal{F}(K) + \lambda}, \quad (10)$$

where K is the kernel matrix. The elements of K are

$$k_{ij} = \kappa(s_i, s_j) = \exp \left(-\frac{1}{\sigma^2} \left(\|s_i\| + \|s_j\| - 2\mathcal{F}^{-1}(\mathcal{F}(s_i^*) \odot \mathcal{F}(s_j)) \right) \right). \quad (11)$$

Since κ is shift invariant, K can be computed efficiently using Fast Fourier Transform (FFT).

In the detection step, the base patch z is firstly cropped out from the likelihood P_t using the previous location O_{t-1} . The candidate patches are cyclic shifts of z . Then, the kernel matrix K^z is calculated by $k_{ij}^z = \kappa(z_i, z_j)$. The detection scores are obtained by

$$\hat{y} = \mathcal{F}^{-1}(AK^z) \quad (12)$$

Finally, the target location in frame t is achieved by maximizing the score \hat{y} .

3.4. Anchor Box Based Scale Estimation

Motivated by recent advances of anchor box for object detection [27,28], as well as visual tracking [32], we present an anchor box based scale estimation to predict the resolution of object. A standard strategy to localize the object at different scales is to perform scale estimation at multiple resolutions [33]. To account for the scale change and geometrical deformation of target, a feature pyramid is first extracted from a rectangular likelihood map centered around the target. At each scale of pyramid, we predict multiple region proposals, which is called anchors. An anchor is centered at the target location and is associated with an aspect ratio. As shown in Figure 5, we first sample the likelihood map around the previously estimated target location at S different scales. At each scale, we use R aspect ratios, yielding $a = S \times R$ anchors at the feature pyramid. However, sampling the feature into multiple resolution is computationally demanding. To boost the speed of the tracker, we perform a multi-scale of anchors instead of a feature map.

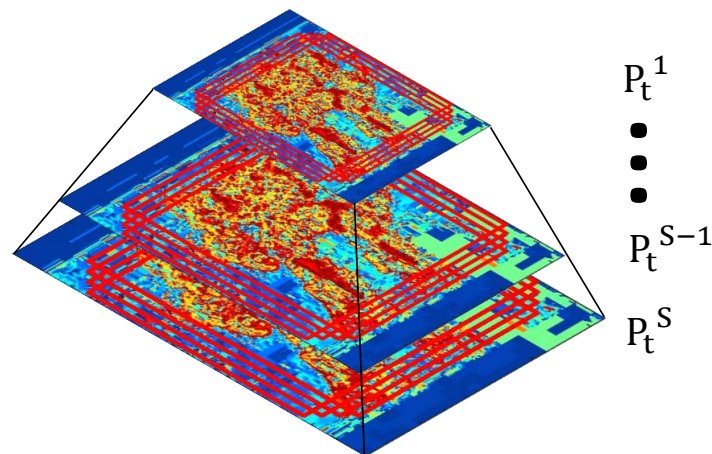


Figure 5. Visualization of anchors region based on feature pyramid. Red rectangles represent anchors at different aspect ratios.

The scale s and aspect ratio r of the target at current likelihood map P_t is obtained by searching the anchor with the highest vote score as following:

$$p_t^{s,r} = \arg \max_{s,r} \frac{\sum_{y \in A(s,r)} P_t(y)}{|A(s,r)|}, \quad (13)$$

where y denotes the location of pixel and $A(s,r)$ denote the anchor region in scale s aspect ratio r . This formulation calculates the average likelihood of anchor region whilst penalizing the maximal region likelihood.

4. Experiments

We validate the effectiveness of our Hyperline Clustering based Tracker (HCT) on the recent tracking evaluation benchmark [15]. In Section 4.1, we describe the details about the parameters and machine used in our experiments. Section 4.2 presents the used benchmark datasets and evaluation protocols. Section 4.3 shows a comprehensive comparison of our HCT with state-of-the-art color based methods.

4.1. Implementation Details

We set the detection region three times the size of the previous object hypothesis O_{t-1} and the surrounding regions is twice the size of O_t . The scale S and aspect ratio R are set 12 and 0.8:1:1.2 respectively, thus we have $12 \times 3 = 36$ anchors each frame. Additionally, we use the regularization parameter $\lambda = 0.001$. All algorithms are tested in MATLAB R2015b, and run on a Lenovo laptop with Intel I7 CPU 3.4 GHz under Windows 7 Professional.

4.2. Experiment Setup

To test the ability of HCT on handling illumination change problem, we employ the color sequences posing illumination variation challenging from OTB-100 dataset [15], namely *Basketball*, *Singer1*, *Singer2*, *CarScale*, *Woman*, and *Trans*. These sequences are also suffered other challenging, such as scale variation, occlusion, deformation, and background clutters.

To report the result of HCT, We use two standard evaluation metrics: Precision and success plots. The precision plot contains the distance precision over a range of center location error threshold. Given the center location of tracked object (x_t, y_t) and ground truth (x_g, y_g) , the location error is defined as

$$\text{location error} = \sqrt{(x_t - x_g)^2 + (y_t - y_g)^2}. \quad (14)$$

In the success plot, the overlap precision is plotted over a range of overlap thresholds. Given the tracked bounding box r_t and the ground truth bounding box r_g , the overlap is defined as

$$\text{Overlap} = \frac{|r_t \cap r_g|}{|r_t \cup r_g|}, \quad (15)$$

where \cap and \cup represent the intersection and union of two regions, respectively, and $||$ denotes the number of pixels in the region.

4.3. Comparison with State-of-the-Art

We compare the proposed method with mentioned state-of-the-art trackers including three color based trackers, namely Staple [14], DAT [7], and ACT [6], and three gray pixel based correlation filtering trackers, namely DSST [25], KCF [12], and CSK [26]. For all methods, we use the publicly available code and suggested parameters corresponding to the authors.

Figure 6a shows the precision plot illustrating the mean location error over 6 sequences. For clarity, we only report the result of one pass evaluation that the trackers are initialized at the first frame. From the figure, we can observe that the proposed HCT performs favorably compared to existing trackers. Staple, which has been demonstrated to acquire best performance in a recent benchmark [24], also outperform other trackers in our experiment. However, HCT outperforms the Staple tracker by 16% in average precision.

Figure 6b shows the success rate plots containing the overlap precision. The mean precision scores for each tracker are presented in the legends. Again, our proposed HCT outperforms Staple by 30% and the baseline KCF tracker by 90% in mean success rate.

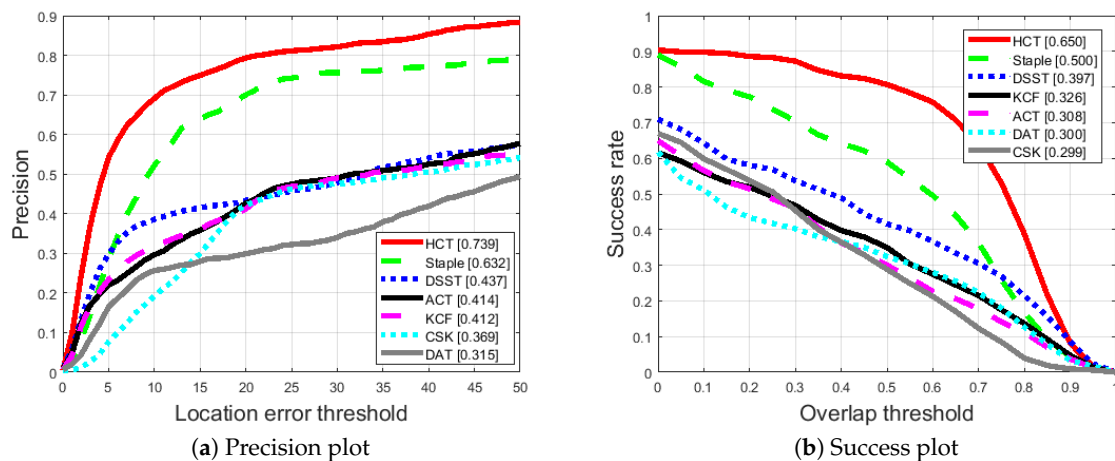


Figure 6. Quantitative comparison of the proposed HCT with several state-of-the-art trackers.

Figure 7 shows a qualitative comparison of the proposed approach with existing trackers. In the *Basketball* sequence, all of compared trackers perform well on this sequence. However, only HCT and Staple are able to estimate the size of object accurately. For the *Singer1* sequence, the target undergoes severe illumination change and scale variation. The color-based trackers, like ACT and DAT, start drifting from the target when the singer suffers severe illumination change in frame #100. Staple works well due to combination of HOG and color-based models. HCT is able to estimate the size of target. This can be attributed to the anchor based scale estimation. In the *Singer2* sequence, the ACT and DAT are less effective in handling illumination change while both HCT and Staple are able to track the target accurately. For the *CarScale* and *Woman* sequences, the targets undergo occlusion and illumination change. HCT and Staple perform well on these sequences with higher overlap scores than other methods. The target object in the *Trans* sequence undergoes shape deformation and severe illumination change. Both the color-based and the gray-based trackers including Staple fail to cope with shape deformation and illumination change simultaneously. However, HCT is able to track the object accurately despite appearance change owing to the deformation and illumination variant. This further confirms the effectiveness of the proposed hyper-line clustering discriminative model and anchor based scale estimation.

Finally, we analyze the running time performance of HCT. For hyperline clustering, the most time-consuming calculation is multilayer initialization [30], which involves a lot of manipulations of matrix decomposition. However, due to the efficient localization and scale estimation, our pure MATLAB prototype of HCT runs at 15 frame per second. Additionally, since HCT only stores the hyperline vectors of object and background, it desires less memory than other trackers. Thus, HCT is able to be utilized in time-critical application and embedded development platform.



Figure 7. Comparison of the proposed approach with state-of-the-art trackers in illumination changes sequence. The results of distractor-aware tracker (DAT) [7], adaptive color attributes tracker (ACT) [6], kernelized correlation filter (KCF) [12], Staple [14], and the proposed approach are represent by green, yellow, blue, magenta, and red respectively.

5. Conclusions

In this work, we investigate the RGB values' distribution of color pixels under different illuminations and propose a novel Hyperline Clustering based Tracker (HCT). Unlike other color based trackers that predominantly apply simple color histogram and are sensitive to illumination changes, the proposed HCT directly extracts hyperlines from both object and surrounding regions to build the likelihood map, enhancing its robustness to illumination changes. The location of an object is estimated by implementing correlation filtering. Furthermore, we propose an anchor based scale estimation to deal with the problem of scale variation and shape deformation. Numerous experiments on the Online Tracking Benchmark demonstrate the favorable performance of the proposed HCT compared with several state-of-the-art trackers. An interesting direction of future work is to apply HCT to multi-object tracking [34] and multi-camera tracking [35], which require an illumination invariant representation of object.

Author Contributions: Conceptualization, S.Y. and P.L.; Data curation, Y.X., H.W. and H.L.; Formal analysis, S.Y. and Y.X.; Funding acquisition, Z.H.; Investigation, Y.X. and H.L.; Methodology, P.L.; Software, S.Y. and P.L.; Supervision, Z.H.; Validation, H.W.; Writing—original draft, S.Y.; Writing—review & editing, P.L.

Funding: This research was funded in part by the National Natural Science Foundation of China under Grants 61773127, 61773128 and 61727810, the Science and Technology Program of Shaoguan City of China under Grant No. SK201644, Natural science foundation of HeBei province under Grant E2016106018 and the Natural Science Foundation of Guangdong Province of China under Grant 2016KQNCX156 and 2018A030307063.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Avidan, S. Ensemble tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 261–271. [[CrossRef](#)] [[PubMed](#)]
2. Grabner, H.; Leistner, C.; Bischof, H. Semi-supervised On-Line Boosting for Robust Tracking. In Proceedings of the European Conference on Computer Vision (ECCV), Marseille, France, 12–18 October 2008; pp. 234–247.
3. Babenko, B.; Yang, M.H.; Belongie, S. Robust object tracking with online multiple instance learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1619–1632. [[CrossRef](#)] [[PubMed](#)]
4. Ciunzo, D.; Buonanno, A.; D’Urso, M.; Palmieri, F.A.N. Distributed classification of multiple moving targets with binary wireless sensor networks. In Proceedings of the 14th International Conference on Information Fusion, Chicago, IL, USA, 5–8 July 2011; pp. 1–8.
5. Buonanno, A.; D’Urso, M.; Prisco, G.; Felaco, M.; Meliadò, E.F.; Mattei, M.; Palmieri, F.; Ciunzo, D. Mobile sensor networks based on autonomous platforms for homeland security. In Proceedings of the 2012 Tyrrhenian Workshop on Advances in Radar and Remote Sensing (TyWRRS), Naples, Italy, 12–14 September 2012; pp. 80–84. [[CrossRef](#)]
6. Danelljan, M.; Khan, F.S.; Felsberg, M.; Weijer, J.V.D. Adaptive Color Attributes for Real-Time Visual Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1090–1097.
7. Possegger, H.; Mauthner, T.; Bischof, H. In defense of color-based model-free tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 2113–2120. [[CrossRef](#)]
8. Ross, D.A.; Lim, J.; Lin, R.S.; Yang, M.H. Incremental Learning for Robust Visual Tracking. *Int. J. Comput. Vis.* **2008**, *77*, 125–141. [[CrossRef](#)]
9. Zhong, W.; Lu, H.; Yang, M.H. Robust object tracking via sparse collaborative appearance model. *IEEE Trans. Image Process.* **2014**, *23*, 2356–2368. [[CrossRef](#)] [[PubMed](#)]
10. Zhang, K.; Zhang, L.; Yang, M.H. Fast compressive tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 2002–2015. [[CrossRef](#)] [[PubMed](#)]
11. Chen, Z.; Liu, P.; Du, Y.; Luo, Y.; Zhang, W. Correlation Tracking via Self-Adaptive Fusion of Multiple Features. *Information* **2018**, *9*, 241. [[CrossRef](#)]
12. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-Speed Tracking with Kernelized Correlation Filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 583–596. [[CrossRef](#)] [[PubMed](#)]
13. Wei, J.; Liu, F. Online Learning of Discriminative Correlation Filter Bank for Visual Tracking. *Information* **2018**, *9*, 61. [[CrossRef](#)]
14. Bertinetto, L.; Valmadre, J.; Golodetz, S.; Miksik, O.; Torr, P.H.S. Staple: Complementary Learners for Real-Time Tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1401–1409. [[CrossRef](#)]
15. Wu, Y.; Lim, J.; Yang, M. Object Tracking Benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1834–1848. [[CrossRef](#)] [[PubMed](#)]
16. Yang, S.; Li, P.; Wen, H.; Xie, Y.; He, Z. K-Hyperline Clustering-Based Color Image Segmentation Robust to Illumination Changes. *Symmetry* **2018**, *10*, 610. [[CrossRef](#)]
17. Comaniciu, D.; Ramesh, V.; Meer, P. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 564–577. [[CrossRef](#)]
18. Sevilla-Lara, L.; Learned-Miller, E. Distribution fields for tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1910–1917.
19. Shen, C.; Kim, J.; Wang, H. Generalized kernel-based visual tracking. *IEEE Trans. Circuits Syst. Video Technol.* **2010**, *20*, 119–130. [[CrossRef](#)]

20. Mei, X.; Ling, H.; Wu, Y.; Blasch, E.; Bai, L. Minimum error bounded efficient l_1 tracker with occlusion detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011; pp. 1257–1264.
21. Liu, B.; Huang, J.; Kulikowski, C.; Yang, L. Robust visual tracking using local sparse appearance model and k-selection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2968–2981. [[CrossRef](#)] [[PubMed](#)]
22. Yang, F.; Lu, H.; Yang, M.H. Robust superpixel tracking. *IEEE Trans. Image Process.* **2014**, *23*, 1639–1651. [[CrossRef](#)] [[PubMed](#)]
23. Kristan, M.; Leonardis, A.; Matas, J.; Felsberg, M.; Pflugfelder, R.; Čehovin Zajc, L.; Vojir, T.; Häger, G.; Lukežič, A.; Fernandez, G. The Visual Object Tracking VOT2016 Challenge Results. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 191–217.
24. Kristan, M.; Matas, J.; Leonardis, A.; Felsberg, M.; Čehovin, L.; Fernandez, G.; Vojir, T.; Hager, G.; Nebel, G.; Pflugfelder, R. The Visual Object Tracking VOT2017 Challenge Results. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; pp. 1949–1972. [[CrossRef](#)]
25. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Discriminative Scale Space Tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1561–1575. [[CrossRef](#)] [[PubMed](#)]
26. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. Exploiting the circulant structure of tracking-by-detection with kernels. In Proceedings of the European Conference on Computer Vision (ECCV), Florence, Italy, 7–13 October 2012; pp. 702–715.
27. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.
28. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
29. Xie, K.; He, Z.; Cichocki, A.; Fang, X. Rate of Convergence of the FOCUSS Algorithm. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 1276–1289. [[CrossRef](#)] [[PubMed](#)]
30. He, Z.; Cichocki, A.; Li, Y.; Xie, S.; Sanei, S. K-hyperline clustering learning for sparse component analysis. *Signal Process.* **2009**, *89*, 1011–1022. [[CrossRef](#)]
31. Swain, M.J.; Ballard, D.H. Color indexing. *Int. J. Comput. Vis.* **1991**, *7*, 11–32.
32. Čehovin, L.; Leonardis, A.; Kristan, M. Robust visual tracking using template anchors. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–8. [[CrossRef](#)]
33. Li, Y.; Zhang, Y.; Xu, Y.; Wang, J.; Miao, Z. Robust Scale Adaptive Kernel Correlation Filter Tracker with Hierarchical Convolutional Features. *IEEE Signal Process. Lett.* **2016**, *23*, 1136–1140. [[CrossRef](#)]
34. Milan, A.; Leal-Taixé, L.; Reid, I.D.; Roth, S.; Schindler, K. MOT16: A Benchmark for Multi-Object Tracking. *arXiv* **2016**, arXiv:1603.00831.
35. Ristani, E.; Solera, F.; Zou, R.; Cucchiara, R.; Tomasi, C. Performance Measures and a Data Set for Multi-target, Multi-camera Tracking. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 17–35.

