Multi-invariance Appearance Model for Object Tracking

Guicong Xu, Xiangmin Xu*, Xiaofen Xing, Bolun Cai, Chunmei Qing
School of Electronic and Information Engineering
South China University of Technology, Guangzhou, China
guicongxu@foxmail.com, {xmxu,xfxing}@scut.edu.cn, caibolun@gmail.com, qchm@scut.edu.cn

***Abstract*—Visual tracking is a challenging problem in computer vision. Most state-of-the-art visual trackers either rely on intensity information, texture information, or use simple color representations for image description, which cannot provide all-around invariance to different scene conditions. Meanwhile there exists no single tracking approach that can successfully handle all scenarios. Due to the complexity of the tracking problem, the combination of multiple features should be computationally efficient and possess a certain amount of robustness while maintaining high discriminative power. This paper combine intensity information (cross-bin distribute field, CDF), texture information (enhance histograms of oriented gradients, EHOG) and color information (color name, CN) in a tracking-by-detection framework, in which a simple tracker called CSK is extended for multi-dimension and multi-cue fusion. The proposed approach improves the baseline single-cue tracker by 4.4% in distance precision. Furthermore,we show that our approach achieving 75.4% is better than most recent state-of-the-art tracking algorithms.**

***Index Terms*—multi-invariance; object tracking**

I. Introduction

Object tracking is one of the most challenging problems in computer vision. It plays a vital role in different kinds applications, especially for human-computer interaction, video surveillance and robotics. The performance of a tracking algorithm is affected by illumination variation, occlusion, background clutters, etc. Object tracking has achieved significant progress recently, but it is still a challenging problem. Recent tracking framework can be split into two main modules generally: appearance model and tracking model. We review target appearance representation schemes in recent state-of-the-art tracking frameworks.

Object representation is one of the major components in any visual tracker and numerous schemes have been presented. According to previous researches, most state-of-the-art trackers depend on intensity information [1], [2]. Holistic templates based on raw intensity values have been extensively used for tracking since Lucas and Kanade's work [3]. Later, intensity histogram [4] is used to model the object appearance, which describes integrate information over a large patch of the target. However, the loss of spatial information when building the histogram makes it sensitive to noise. Multi-kernel [5] or multi-patch [6] descriptors including some spatial information are proposed to address this problem. The fragment-based tracker [7] splits the target into multiple regions and describes them using multiple local histograms, increasing the time and computation even the use of integral histograms [8]. In order to balance a pair of contradictions between the specificity of the descriptor and the landscape smoothness criterion, a distribution field descriptor [9] is proposed as a novel intensity appearance model. A The appearance of an object changes drastically when illumination varies significantly. Since intensity histogram features is easily affected by lights, numerous tracking frameworks based on illumination-invariant features have been proposed. Previous researches on tracking describe objects with contours [10] when satisfy the brightness constancy assumption. Many other texture features are utilized to model object appearance for tracking, such as histograms of oriented gradients (HOG) [11], covariance region descriptor [12], local binary patterns (LBP) [13] and Haar-like features [14]. Considering local directional edge information, Tang et al. [11] adopts HOG with the integral histogram [8] for tracking. Lately, online subspace models have been used for object tracking in dealing with large lighting variation [15], [16].

In addition, color feature has gained more attention since it contains abundant information. Color videos become more and more popular in computer vision. Therefore, tracking algorithms based on color appearance model have much progress recently [17], [18]. Color-histogram-based mean-shift algorithm is used for tracking [1]. Color Name (CN) [19] uses many color attributes and proposes an adaptive dimensionality reduction technique for tracking. However, the use of color leads to the algorithm unstable to similarly colored backgrounds and low saturation objects.

Most algorithms only use an individual cue and cannot provide all-around invariance to different scene conditions. Meanwhile, there exists no single tracking approach that can successfully handle all scenarios. So it's appealing to integrate multiple cues into one observation model for tracking. Many researchers focus on employing probabilistic approaches to model interactions among multiple cues, such as Dynamic Bayesian Network[20], Monte Carlo method[12], and Particle Filters[21]. However the use of Bayesian framework makes them difficult to be used in deterministic tracking methods. Recently, tracking associated with detection has become popular because detection responses could help to locate object exactly and alleviate drift [22], [23], [24]. Breitenstein et al.

*Xiangmin Xu is the corresponding author.

978-1-4799-8058-1/15/$31.00 ©2015 IEEE.

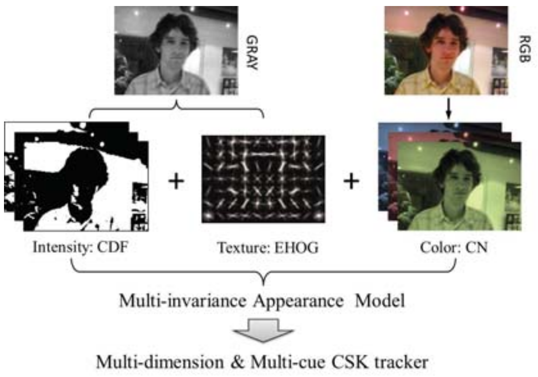
GRAY
RGB
Intensity: CDF
Texture: EHOG
Color: CN
Multi-invariance Appearance Model
Multi-dimension & Multi-cue CSK tracker

Fig. 1. Multi-invariance Appearance Model for Object Tracking

[23] integrates detector into tracker by monitoring continuous detection confidence in particle filtering framework. Through online-trained classifier, prior knowledge is introduced and reliable detections are associated in the end. Xing et al. [24] collects detection responses in a temporal sliding window and associates them with potential tracklets.

Most multi-cue tracking algorithms cannot choose the most effective feature in different application environments. Considering this problem, we combine three different features including intensity, texture and color information to obtain the advantages of different features in diverse scenarios. In addition, direct integration of the three features in particle filter framework leads to low accuracy and unsatisfying real-time performance. In this paper, we extend CSK[25] for multi-dimension and multi-cue fusion. In extended CSK framework, we fuse multiple features in response layer to attain a compact system and get an effective multi-cue tracking algorithm. Finally, based on the evaluation of online tracking benchmark[26], we show that our tracker combined with multi-invariance appearance model and extended CSK achieves outstanding performance in a comprehensive evaluation over 35 color image sequences.

II. Multi-invariance Appearance Model

In this paper, a multi-invariance appearance model is proposed for visual object tracking showed as Fig. 1. Firstly, a cross-bin distribution field (CDF) feature describes the intensity properties. Then an enhance histograms of oriented gradients (EHOG) is used to model texture variety. Moreover, a color name (CN) method extracting color attribute keeps color invariance. In the end, a CSK tracker extended multi-cue combines multi-invariance appearance model.

A. Intensity: Cross-bin Distribution Field

In this paper, we employ cross-bin distribution field (CDF) to describe the intensity properties of objects. A distribution field (DF) [9] is simply an array of probability distributions, which is a matrix with 3 dimensions including the width, the height of the image, and the gray intensity feature space. In gray intensity space, an image of size $w * h$ yields a 3D distribution field of size $w * h * b$, b is the number of intensity feature bins.

- Divide the original intensity $I(m,n)$ into different distribution filed with respect to feature layers.

  $$F_{df}(m,n,k) = \begin{cases} 1, if\ I(m,n) \in D_k \\ 0, otherwise \end{cases} \tag{1}$$

  ,where m and n represent the row and column of the intensity image, and $D_k = [255k/b, 255(k+1)/b)$ $, k \in \{0, 2, ..., b-1\}$ represents the k-th bin's intensity region.
- It is widely accepted that cross-bin metrics are generally more suitable for measuring histogram similarity. Here a cross-bin metric is a concatenation of histogram smoothing to content in each bin depends on multiple bins [27].

  $$F_{cdf}(m,n,k) = F_{df}(m,n,k) * h_{\sigma_s} * h_{\sigma_k} \tag{2}$$

  ,where h_{σ_s} is a spatial space Gaussian kernel of standard deviation σ_s, h_{σ_k} is a 1D Gaussian kernel of standard deviation σ_k over feature space, and $*$ is the convolution operator.

B. Texture: Enhance Histograms of Oriented Gradients

Histograms of oriented gradients (HOG) [28] is a classical texture descriptor, which is widely applied to object detection, action recognition, visual tracking. In this paper, an enhance histograms of oriented gradients (EHOG) proposed by [29] is used to model texture variety. Using finite difference filters $[-1, 0, +1]$ and its transpose, orientations $\theta(m,n)$ and magnitudes $\alpha(m,n)$ of the intensity gradient at a pixel (m,n) is computed. The gradient orientation $\theta(m,n)$ is discretized into one of the values B using either a contrast sensitive (B_1)or insensitive (B_2), where $(p+q)$ is the number of orientation bin.

$$\begin{aligned} B_1(m,n) &= \text{round}\left(\frac{p\theta(m,n)}{2\pi}\right) \bmod p \\ B_2(m,n) &= p + \text{round}\left(\frac{q\theta(m,n)}{\pi}\right) \bmod q \end{aligned} \tag{3}$$

Then, a feature map specifies a sparse histogram of gradient magnitudes at each 2×2 dense grid of rectangular cell. Let $k \in \{0, 1, \ldots, p+q-1\}$ range over orientation bins, and the cell feature map at (m,n) is

$$F_{hog}(m,n,k) = \begin{cases} \alpha(m,n), if\ k = B(m,n) \\ 0, otherwise \end{cases} \tag{4}$$

Gradient invariance can be achieved via normalization [28] using $F_{nhog}(m,n,k) = F_{hog}(m,n,k)/N_{\delta_x,\delta_y}(m,n,k)$ by four different factors $N_{\delta_x,\delta_y}(m,n,k), \delta_x, \delta_y \in \{+1, -1\}$.

$$\begin{aligned} N_{\delta_x,\delta_y}(m,n) = (&F_{hog}^2(m,n) + F_{hog}^2(m+\delta_x, n+\delta_y) \\ &+F_{hog}^2(m+\delta_x, n) + F_{hog}^2(m, n+\delta_y))^{0.5} \end{aligned} \tag{5}$$

In this paper, the cell-based feature map F_{hog} is nine contrast insensitive orientations ($p = 9$) and 18 contrast sensitive orientations ($q = 18$). Therefore, four normalization factors for each cell obtain $4*(9+18) = 108$ normalize dimensional map F_{nhog}. According to [29], an analytic projection of these 108-dimensional vectors is defined by 27 sums over different normalizations (one for each orientation channel) , and 4 sums over the 27 contrast insensitive/sensitive orientations(one for each normalization factor) . The final feature map F_{ehog} has 31-dimensional vectors, including 27 dimensions corresponding to different orientation channels (9 contrast insensitive and 18 contrast sensitive) and 4 dimensions capturing the overall gradient energy.

C. Color: Color Name

The selection of color feature is vital for the performance of a visual tracker. Inspired by recent progress in tracking based on color [19], rather than use the simple color histogram, we employ many color features to describe objects. Color names (CN) [30] are linguistic color labels assigned by humans to represent colors in the real world, including eleven basic color: black, brown, green, pink, red, yellow, blue, grey, orange, purple and white. However, RGB color using in computer vision usually can be mapped to a probabilistic 11 dimensional color attributes by the mapping matrix, which is automatically learned from images retrieved with Google-image search. Therefore, color name probabilities can be describe by (6).

$$F_{cn}\left(m,n,k\right)=Map\left(R\left(m,n\right),G\left(m,n\right),B\left(m,n\right),k\right) \tag{6}$$

, where $R\left(m,n\right)$, $G\left(m,n\right)$, $B\left(m,n\right)$ corresponding to RGB color value of images, and Map indicates a mapping matrix from RGB to 11 dimensional color probabilities.

D. Multi-invariance Tracking Model based on CSK

We apply our multi-invariance appearance model on a simple tracker called CSK[9], which provides the highest speed among the top ten trackers in the recent benchmark [25]. In this paper, CSK is extended for multi-dimension and multi-cue fusion. Considered tracking as a pixel classification problem to distract foreground and background. Each pixel considered to be the probability of foreground or background is determined by all the cues. Every cue has a saliency map. We combine all the saliency maps to get final tracking result.The original CSK tracker optimize a kernelized least squares classifier of a target from a single dimensional gray-image patch. The CSK tracker exploits the circulant structure to realize dense sampling that appears from the periodic assumption of the local patch, and employs Fast Fourier Transform (FFT) to speed up. Here we provide a brief overview of CSK tracker and extend it for multi-invariance appearance model.

A multi-invariance appearance model $x\left(m,n\right)=F_f\left(m,n\right)$ around the target's center $\left(m_o,n_o\right)$ is used to train CSK classifier. $F_f\left(m,y\right)$ is labelled with a Gaussian function $y(m,n)$. The classifier is trained by minimizing the cost function over w.

$$\varepsilon=\sum_{m,n}\left|\left\langle\phi\left(x\left(m,n\right)\right),\phi\left(\omega\right)\right\rangle-y\left(m,n\right)\right|^2+\lambda\left\langle\phi\left(\omega\right),\phi\left(\omega\right)\right\rangle \tag{7}$$

Here the constant $\lambda>0$ is a regularization parameter, and ϕ is the mapping to the Hilbert space induced by the kernel κ, defining the inner product as $\left\langle\phi\left(x_1\right),\phi\left(x_2\right)\right\rangle=\kappa\left(x_1,x_2\right)$.

For the sake of multi-dimension extension, multi-channel correlation filters[31] is used for allowing multiple dimension k feature by summing over them in the Fourier domain with Gaussian kernel (variance σ) as (8).

$$\begin{aligned}&\mathcal{K}\left(\mathcal{F}\left(x_1\right),\mathcal{F}\left(x_2\right)\right)=\mathcal{F}\left(\kappa\left(x_1,x_2\right)\right)\\&=\mathcal{F}\left(\exp\left(-\frac{\|x_1\|^2+\|x_2\|^2-2\mathcal{F}^{-1}\left(\sum_k\mathcal{F}(x_1(\cdot,k))\mathcal{F}(x_2(\cdot,k))\right)}{\sigma^2}\right)\right)\end{aligned} \tag{8}$$

Then, the cost function in (7) is minimized by ω, where the coefficients ω are

$$A=\mathcal{F}\left(\phi\left(\omega\right)\right)=\frac{\mathcal{F}\left(y\left(m,n\right)\right)}{\mathcal{K}\left(\mathcal{F}\left(x_1\right),\mathcal{F}\left(x_2\right)\right)+\lambda} \tag{9}$$

In addition, the weighted superposition of different cues respond maps applies to multi-cue fusion shown as (10).

$$r=\sum\nolimits_f w_f r_f=\sum\nolimits_f w_f\mathcal{F}^{-1}\left(A_f\mathcal{K}\left(\mathcal{F}\left(F_f\right),M_f\right)\right) \tag{10}$$

,where F_f is the appearance feature , M_f is the frequency domain model and w_f is the multi-cue weight. Based on the CSK framework, the multi-invariance tracker extending CSK for multi-dimension and multi-cue can be summarized in Algorithm 1.

Algorithm 1 Multi-invariance Tracker based on CSK

```
Input: Multi-feature $F_f, f \in \{cdf, ehog, cn\}$
Output: Tracking result $(\hat{m},\hat{n})$
Set $Y=\mathcal{F}\left(\exp\left(-\frac{1}{2\sigma_s^2}\left(\left(m-m_o\right)^2+\left(n-n_o\right)^2\right)\right)\right)$
for $t=1,2,\ldots$ do
    Multi-dimension extension:
    $K_f^t=\mathcal{K}\left(\mathcal{F}\left(F_f^t\left(m,n,k\right)\right),M_f^t\right)$
    Multi-cue extension:
    $r=\sum_f w_f r_f=\sum_f w_f\mathcal{F}^{-1}\left(A_f^t\times K_f^t\right)$
    Find target: $(\hat{m},\hat{n})=\arg\max_{(m,n)} r\left(m,n\right)$
    Calculate Model:
    $M'_f=\mathcal{F}\left(F_f^t\left(\hat{m},\hat{n},k\right)\right)$
    $A'_f=Y\Big/\left(K'_f+\lambda\right)=Y\Big/\left(\kappa\left(M'_f,M'_f\right)+\lambda\right)$
    Update Model:
    $A_f^{t+1}=\gamma A'_f+\left(1-\gamma\right)A_f^t$
    $M_f^{t+1}=\gamma M'_f+\left(1-\gamma\right)M_f^t$
end for
```

III. Experimental results

The proposed tracker is implemented in MATLAB 2013A on a PC with Intel Core2 CPU (2.66 GHz) with 2 GB memory, and runs about 10 frames per second (fps) in this platform. In this paper, we set $w_{cdf}=1/6$, $w_{ehog}=1/2$ and $w_{cn}=1/3$.

We compare the proposed method with 10 state-of-the-art trackers (KCF [32], Struck [33], SCM [34], TLD [35], VTD [15], VTS [16], CSK [25], LSK [36], OAB [37], RS-V [38]) on 35 color sequences 2 in the CVPR2013 benchmark [26]. The best way to evaluate trackers is still a debatable subject. Averaged measures like mean center location error or average bounding box overlap penalize an accurate tracker that fails for short-time more than an inaccurate tracker. According to

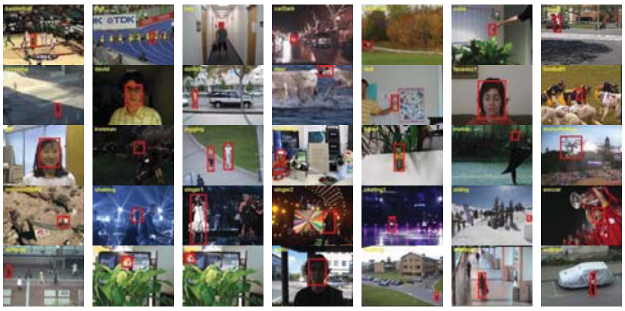

Fig. 2. Tracking sequences for evaluation

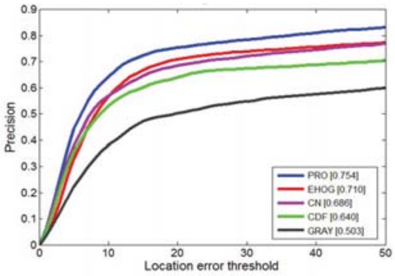

Fig. 3. Precisions plots between different appearance models

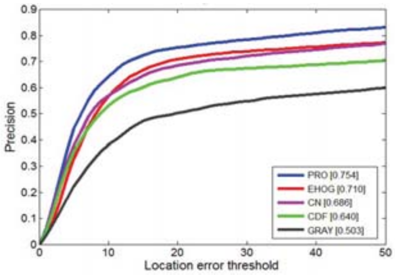

Fig. 4. Precisions plots between different appearance models

[26], the precision plot shows the percentage of frames on which the Center Location Error (CLE) of a tracker is within a given threshold e, where CLE is defined as the center distance between tracker output $(\hat{m}, \hat{n})$ and ground truth (m_g, n_g).

Fig.3 shows the precision plots containing the mean error over all the 35 sequences, and a representative precision score $(e = 20)$ is used for ranking. In the precision plot, the proposed tracker outperforms KCF [32] by 4.7% and Struck (the top tracker on [26]) by 18.0% in mean CLE at the threshold of 20 pixels. KCF tracker is same as the single texture feature (EHOG) used in CSK tracker. Fig. 4 and Table I summarize the performances between multi-invariance appearance model and single appearance model over 35 color sequences. Fig. 4 shows that multi-invariance appearance model is better than most recent state-of-the-art algorithms. According to the experimental result, the texture model (EHOG) and the color model (CN) are the most important features of object expression. In addition, the intensity feature supplies the correction of tracking accuracy.

TABLE I
TRACKER PRECISIONS $(e = 20)$ OVER 35 SEQUENCES
(**BLACK** INDICATES THE BEST)

Sequence	Gray	CDF	EHOG	CN	Proposed
Basketball	0.284	0.022	0.923	**0.997**	0.974
Bolt	0.034	0.991	0.989	1.000	**1.000**
Boy	1.000	1.000	1.000	1.000	**1.000**
CarDark	1.000	1.000	1.000	1.000	**1.000**
CarScale	0.694	0.663	**0.806**	0.651	0.647
Coke	0.866	**0.918**	0.838	0.904	0.876
Couple	**0.621**	0.593	0.257	0.107	0.536
Crossing	1.000	1.000	1.000	1.000	**1.000**
David	0.652	1.000	1.000	1.000	**1.000**
David3	0.421	0.937	1.000	1.000	**1.000**
Deer	**1.000**	0.901	0.817	0.972	0.873
Doll	0.733	0.990	0.967	0.992	**0.992**
FaceOcc1	0.915	**0.935**	0.878	0.785	0.841
Football1	0.743	0.892	0.959	0.959	**0.973**
Girl	0.916	0.600	0.864	0.918	**1.000**
Ironman	0.145	0.078	0.217	0.169	**0.446**
Jogging.1	0.228	0.228	0.235	0.225	**0.977**
Jogging.2	1.000	0.186	0.163	1.000	**1.000**
Lemming	0.382	**0.528**	0.495	0.275	0.405
Liquor	0.196	0.189	0.423	0.407	**0.435**
Matrix	0.100	0.110	**0.170**	0.030	0.150
MotorRolling	0.037	0.037	0.043	**0.061**	0.049
MountainBike	1.000	1.000	1.000	1.000	**1.000**
Shaking	0.008	0.164	0.025	0.282	**0.811**
Singer1	0.476	0.855	0.980	1.000	**1.000**
Singer2	0.036	0.612	**0.945**	0.036	0.036
Skating1	0.700	1.000	1.000	0.517	**1.000**
Skiing	0.136	0.136	0.074	**1.000**	0.136
Soccer	0.258	0.151	**0.793**	0.204	0.214
Subway	0.240	1.000	1.000	1.000	**1.000**
Tiger1	0.480	0.797	**0.975**	0.410	0.941
Tiger2	0.110	0.244	0.356	0.455	**0.501**
Trellis	0.225	0.963	1.000	0.982	**1.000**
Walking	0.816	1.000	1.000	1.000	**1.000**
Walking2	0.402	0.388	**0.440**	0.414	0.408
Woman	0.248	0.940	0.938	0.940	**0.940**
Average	0.503	0.640	0.710	0.686	**0.754**

IV. Conclusion

In this work, we demonstrate that it is possible to build multi-invariance appearance model to track targets successfully. By combining features including intensity, texture and color information, our method achieves excellent result in complicated and diverse environments. Extensive experiments demonstrate that the proposed multi-invariance appearance model and extended CSK algorithm perform quite well in the unconstrained tracking situations. Meanwhile, there are many improvements that could be explored. We expect to find an adaptive weight selection strategy. We also hope to see that the multi-invariance appearance model has useful applications outside of tracking.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (#61171142, #61401163), the Science and technology Planning Project of Guangdong Province of China (#2011A010801005, #2012B061700102).

References

[1] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 5, pp. 564–577, 2003.

[2] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008.

[3] S. Baker and I. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International journal of computer vision*, vol. 56, no. 3, pp. 221–255, 2004.

[4] K. J. Cannons, J. M. Gryn, and R. P. Wildes, "Visual tracking using a pixelwise spatiotemporal oriented energy representation," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 511–524.

[5] G. D. Hager, M. Dewan, and C. V. Stewart, "Multiple kernel tracking with ssd," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1. IEEE, 2004, pp. I–790.

[6] Z. Fan, Y. Wu, and M. Yang, "Multiple collaborative kernel tracking," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 502–509.

[7] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 798–805.

[8] F. Porikli, "Integral histogram: A fast way to extract histograms in cartesian spaces," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 829–836.

[9] L. Sevilla-Lara and E. Learned-Miller, "Distribution fields for tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1910–1917.

[10] D. Freedman and T. Zhang, "Active contours for tracking distributions," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 518–526, 2004.

[11] F. Tang, S. Brennan, Q. Zhao, and H. Tao, "Co-tracking using semi-supervised support vector machines," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[12] Y. Wu, J. Cheng, J. Wang, H. Lu, J. Wang, H. Ling, E. Blasch, and L. Bai, "Real-time probabilistic covariance tracking with efficient model update," *Image Processing, IEEE Transactions on*, vol. 21, no. 5, pp. 2824–2837, 2012.

[13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[14] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting." in *BMVC*, vol. 1, no. 5, 2006, p. 6.

[15] J. Kwon and K. M. Lee, "Visual tracking decomposition," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 1269–1276.

[16] J. Kwon and Lee, "Tracking by sampling trackers," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1195–1202.

[17] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1940–1947.

[18] R. T. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1631–1643, 2005.

[19] M. Danelljan, F. S. Khan, M. Felsberg, and J. v. d. Weijer, "Adaptive color attributes for real-time visual tracking," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1090–1097.

[20] G. Jia, Y. Tian, Y. Wang, T. Huang, and M. Wang, "Dynamic multi-cue tracking with detection responses association," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1171–1174.

[21] X. Sun, H. Yao, and X. Lu, "Dynamic multi-cue tracking using particle filter," *Signal, Image and Video Processing*, vol. 8, no. 1, pp. 95–101, 2014.

[22] C. Huang, B. Wu, and R. Nevatia, "Robust object tracking by hierarchical association of detection responses," in *Computer Vision–ECCV 2008*. Springer, 2008, pp. 788–801.

[23] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Robust tracking-by-detection using a detector confidence particle filter," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1515–1522.

[24] J. Xing, H. Ai, and S. Lao, "Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1200–1207.

[25] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 702–715.

[26] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Computer vision and pattern recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 2411–2418.

[27] I. Leichter, "Mean shift trackers with cross-bin metrics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 4, pp. 695–706, 2012.

[28] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

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

[30] J. Van De Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *Image Processing, IEEE Transactions on*, vol. 18, no. 7, pp. 1512–1523, 2009.

[31] H. K. Galoogahi, T. Sim, and S. Lucey, "Multi-channel correlation filters," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3072–3079.

[32] 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*, vol. 37, no. 3, pp. 583–596, 2014.

[33] S. Hare, A. Saffari, and P. H. Torr, "Struck: Structured output tracking with kernels," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 263–270.

[34] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Computer vision and pattern recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1838–1845.

[35] Z. Kalal, J. Matas, and K. Mikolajczyk, "Pn learning: Bootstrapping binary classifiers by structural constraints," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 49–56.

[36] B. Liu, J. Huang, L. Yang, and C. Kulikowsk, "Robust tracking using local sparse appearance model and k-selection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1313–1320.

[37] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *Computer Vision–ECCV 2008*. Springer, 2008, pp. 234–247.

[38] R. Collins, X. Zhou, and S. K. Teh, "An open source tracking testbed and evaluation web site," in *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2005, pp. 17–24.