Hindawi
Mathematical Problems in Engineering
Volume 2017, Article ID 5295601, 13 pages
https://doi.org/10.1155/2017/5295601

Research Article

Visual Tracking Based on Complementary Learners with Distractor Handling

Suryo Adhi Wibowo, Hansoo Lee, Eun Kyeong Kim, and Sungshin Kim

Department of Electrical and Computer Engineering, Pusan National University, Busan, Republic of Korea

Correspondence should be addressed to Sungshin Kim; sskim@pusan.ac.kr

Received 26 October 2016; Revised 18 January 2017; Accepted 30 January 2017; Published 13 April 2017

Academic Editor: Simone Bianco

The representation of the object is an important factor in building a robust visual object tracking algorithm. To resolve this problem, complementary learners that use color histogram- and correlation filter-based representation to represent the target object can be used since they each have advantages that can be exploited to compensate the other's drawback in visual tracking. Further, a tracking algorithm can fail because of the distractor, even when complementary learners have been implemented for the target object representation. In this study, we show that, in order to handle the distractor, first the distractor must be detected by learning the responses from the color-histogram- and correlation-filter-based representation. Then, to determine the target location, we can decide whether the responses from each representation should be merged or only the response from the correlation filter should be used. This decision depends on the result obtained from the distractor detection process. Experiments were performed on the widely used VOT2014 and VOT2015 benchmark datasets. It was verified that our proposed method performs favorably as compared with several state-of-the-art visual tracking algorithms.

1. Introduction

Given the initial state (e.g., position and other information) of a target object in the first frame, the goal of visual tracking is to predict the states of the target in subsequent frames. Visual tracking has an important role in several applications in the areas of computer vision, such as motion analysis, visual surveillance, human computer interaction, and robot navigation. Although this issue has been studied for several decades and considerable progress has been made, it still presents challenges, in particular, the development of a robust algorithm for overcoming problems such as occlusions, camera motion, illumination changes, motion changes, and size changes.

An important factor in creating a robust visual tracking algorithm is the representation of the target object. Several decades ago, to solve challenging problems in visual tracking, researchers used a color histogram [1] to represent the target object. A generative approach combined with an optimization method, such as the Lukas-Kanade algorithm, Kalman filter [2], and particle filter [3], was usually applied. The Lukas-Kanade algorithm usually utilized a differential method to handle optical flow. Unfortunately, the computation involved in this method is expensive and it has many disadvantages for addressing challenging problems in visual tracking. The Kalman filter also has some limitations for challenging problems in that it assumes that both the system and observation model equations are linear and that the distribution of the state uses Gaussian distribution. These assumptions are not realistic in many real conditions. The particle filter was proposed for overcoming the limitations of the Kalman filter. Although it has been shown that a particle filter significantly improves the results and can handle nonlinear problems, it has some issues related to the relationship between accuracy, the number of particles, and computation time [4]. Furthermore, the generative approach is focused only on learning an appearance model. It does not take the information from the background model into consideration, although such information is very valuable for developing a more robust visual tracking algorithm. Moreover, although color histogram-based representation has advantages which are robust to deformations, it has a disadvantages or a drawback when illumination changes occur. It is also sensitive to motion blur.

Later, the discriminative approach was proposed for improving the performance of the generative approach. The main difference between the discriminative and the generative approach is in the utilization of a classifier method to determine the location of the target object. The generative approach, on the one hand, does not need a classifier method to determine the output; the output is determined by the nearest distance according to a one-by-one distance comparison with the target. For this reason, the computation time of the generative approach is expensive. On the other hand, the discriminative approach uses a classifier method for determining the output and takes the information from the background model into consideration. Therefore, positive and negative samples should be used for representing the target object and the background, respectively. For example, Grabner et al. proposed an online feature selection method using an AdaBoost algorithm for visual tracking. This method has online training capability [5]. Although it operates quickly, online learning is problematic, in particular, when each update of the tracker may introduce an error, which finally can lead to tracking failure (drifting). Semisupervised online boosting alleviates the drifting problem in tracking applications [6]. Another method for visual tracking, called multiple instance learning (MIL), was proposed by Babenko et al. to replace traditional supervised learning [7]. This method treats positive and negative samples as a positive and negative bag, respectively. Then, to determine the output, a boosting classifier is used. This method operates faster and more accurately than traditional supervised learning. Kalal et al. proposed a tracking-learning-detection framework [8]; however, unfortunately this framework needs a large memory for computation. These methods can be termed tracking-by-detection methods.

Recently, a correlation filter has been used, which provides efficient computation, since the operator is transformed into the Fourier domain. Further, it also produces good results, although a limited amount of training data is used. For these reasons, researchers introduced the correlation filter into the tracking-by-detection method for visual tracking. An example is the method, called minimizing the output sum of squared error (MOSSE) tracker, that was introduced by Bolme et al. [9]. For training the correlation filter, this method used only grayscale samples. To improve the method, according to the results of recent studies multidimensional features such as histogram of Gaussian (HOG) features can be used [10–13]. Although the correlation filter provides efficient computation, all the circular shifts should be learned during the process. To resolve this issue, Danelljan et al. proposed the spatially regularized discriminative correlation filter (SRDCF) [14]. Although it achieves excellent results, this method needs a computational time longer than the original one. Moreover, although the correlation filter has the advantages which show excellent robustness to challenging problems, such as illumination changes and motion blur, it has a disadvantages or a drawback when problems such as deformation arise.

To compensate the advantages and disadvantages of color histogram-based representation and correlation filter-based representation, respectively, a representation of the target object based on complementary learners was proposed [10, 11, 15]. In this study, we adopted complementary learners and we propose an object-aware method based on them. These representations are computed in parallel, where each representation produces a color histogram response and correlation filter response, respectively. Since the tracking algorithm can fail because of the distractor, a method to handle the distractor is proposed to minimize tracking failures. First, distractor detection should be performed. This can be achieved by calculating the distance between the maximum value of the color histogram response and the maximum value of the correlation filter response. Then, the location of the target object can be determined from either the maximum value of the correlation filter response or the maximum value of the merged responses of the color histogram and the correlation filter; the value selected depends on the results of the distractor detection process. We demonstrate our proposed method on the widely used VOT2014 and VOT2015 benchmarks. According to the results of our experiments, the proposed method performs favorably as compared to state-of-the-art visual tracking algorithms.

The rest of this paper is organized as follows. We describe our object-aware method based on complementary learners in Section 2. The distractor detection method is explained in Section 3. The proposed method is detailed in Section 4. In Section 5, the experimental results with comparisons to the state-of-the-art methods are presented. Finally, conclusions are presented in Section 6.

2. Object-Aware Method Based on Complementary Learners

One important factor in building a robust visual tracking algorithm is determining the model representation of the target object. Color histogram-based object representation has been used widely. Unfortunately, this representation is not robust when the color of the distractor is similar to that of the tracked object. In addition, this representation has disadvantages or the drawback when illumination changes occur and is also sensitive to motion blur. Recently, a correlation filter has been used for representing the object. Although it is robust to challenges such as motion changes and illumination changes, it has a drawback when deformation occurs. Complementary learners, in which the results of a collaboration between the correlation filter and color histogram are used to represent the target object in visual tracking, were inspired by these ideas [10, 11, 15]. The representations should be computed in parallel to produce each response before the distractor is analyzed based on these responses.

Given frame t, we can calculate the color histogram of the object, h_o, and the color histogram of the background, h_b, from the previous frame to obtain the response of the color histogram, r_t^{ch}. First, this response is computed from the pixel at location x in the location of the search area of the target object $|a_{\mathrm{search},t}|$, which has the same bin index id_x. Then, following Bayes' theorem, we calculate $P(x \in o \mid \mathrm{id}_x)$ by using

$$
\begin{aligned}
P\left(x \in o \mid \mathrm{id}_x\right) &= \frac{P\left(\mathrm{id}_x \mid x \in o\right) P\left(x \in o\right)}{P\left(\mathrm{id}_x \mid x \in o\right) P\left(x \in o\right) + P\left(\mathrm{id}_x \mid x \in b\right) P\left(x \in b\right)} \\
&= \frac{\left(h_o\left(\mathrm{id}_x\right) / \left|a_o\right|\right)\left(\left|a_o\right| / \left(\left|a_o\right| + \left|a_b\right|\right)\right)}{\left(h_o\left(\mathrm{id}_x\right) / \left|a_o\right|\right)\left(\left|a_o\right| / \left(\left|a_o\right| + \left|a_b\right|\right)\right) + \left(h_b\left(\mathrm{id}_x\right) / \left|a_b\right|\right)\left(\left|a_b\right| / \left(\left|a_o\right| + \left|a_b\right|\right)\right)} \approx \frac{h_o\left(\mathrm{id}_x\right)}{h_o\left(\mathrm{id}_x\right) + h_b\left(\mathrm{id}_x\right)},
\end{aligned}
\tag{1}
$$

where $|a_o|$ and $|a_b|$ are the rectangle area of the object and the background, respectively. Finally, the response of the color histogram r_t^{ch} can be obtained by using the integral image from $P(x \in o \mid \mathrm{id}_x)$.

On the other hand, as in [10–14], HOG features are used as multidimensional features. They produce n-dimensional feature map representation of an image. Based on this representation, the optimal correlation filter q is obtained by using

$$
\arg\min_{q} \sum_{i=1}^{n} \left| q^i \otimes f^i - d \right|^2 + \lambda \left\| q^i \right\|^2, \tag{2}
$$

where f, d, $\otimes$, and λ are the rectangle patch of the feature map that represents the target, the desired correlation output, the circular correlation, and the parameter that controls the effect of the regularization term, respectively. Further, the correlation filter operates in the Fourier domain, and, therefore, we can use the discrete Fourier transform (DFT), which produces a complex variable. Because the results of the DFT take a complex form and we need to solve (2), we follow the method presented in [16] and then we obtain

$$
Q^i = \frac{\overline{D} \odot F^i}{\sum_{k=1}^{n} \overline{F}^k \odot F^k + \lambda}, \tag{3}
$$

where $\overline{D}$ is the complex conjugate of the DFT of d, F^i is the DFT of f^i, $\overline{F}^k$ is the complex conjugate of the DFT of f^k, F^k is the DFT of f^k, $\odot$ represents element-wise multiplication, and Q^i is the result in the Fourier domain.

An inexpensive computation is required to develop a visual tracking algorithm. This is because, to handle the appearance changes in the target object, online learning is effective, as was proved in [5–8]. Further, based on (3), $n \times n$ linear system of equations per pixel needs to be solved and this requires expensive computation. Thus, rather than performing expensive computation, where robust approximation is needed, an online update of the numerator β_t and denominator γ_t at frame t, which was adopted from [16], is used:

$$
\begin{aligned}
\beta_t^i &= \left(1 - \alpha^{\mathrm{cf}}\right) \beta_{t-1}^i + \alpha^{\mathrm{cf}} \widehat{\beta}_t^i, \\
\gamma_t &= \left(1 - \alpha^{\mathrm{cf}}\right) \gamma_{t-1} + \alpha^{\mathrm{cf}} \widehat{\gamma}_t,
\end{aligned}
\tag{4}
$$

where $\widehat{\beta}_t^i = \overline{D}_t \odot F_t^i$, $\widehat{\gamma}_t = \sum_{k=1}^{n} \overline{F}_t^k \odot F_t^k$, α^{cf} is a learning rate parameter, β_{t-1}^i is the numerator at frame $t-1$, and γ_{t-1} is the denominator at frame $t-1$. Moreover, a response of the correlation filter r_t^{cf} can be calculated using the inverse DFT:

$$
r_t^{\mathrm{cf}} = \mathscr{F}^{-1}\left(\frac{\sum_{i=1}^{n} \overline{\beta}_t^i \odot \phi_t^i}{\gamma_t + \lambda} \right), \tag{5}
$$

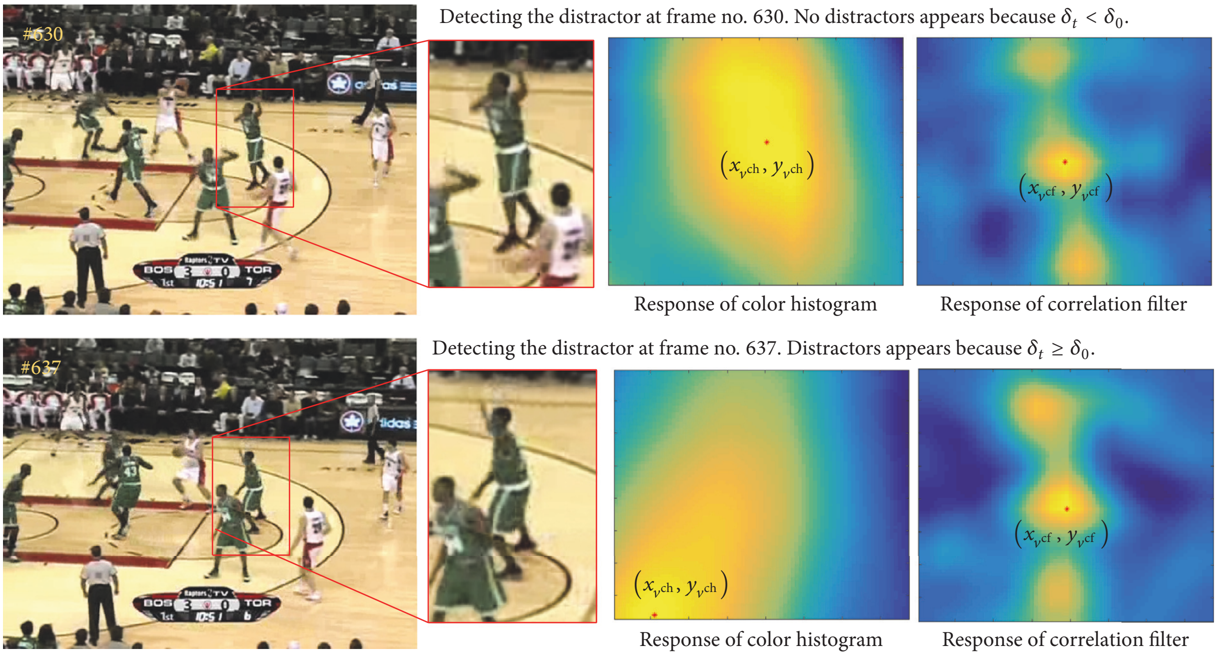
where ϕ_t^i is the feature map from $|a_b|$ which has been multiplied by hanning window and $\overline{\beta}_t^i$ is the complex conjugate from β_t^i.

3. Distractor Detection

Visual tracking algorithms usually fail because of the distractor, in particular when the distractor has a representation similar to that of the target object. To overcome this problem, Kalal et al. [8] proposed a learning method assisted by positive and negative constraint to distinguish a target object from the background. In addition, they used optical flow for motion model. Unfortunately, this approach needs a large memory for computation. Recently, Possegger et al. [17] proposed foreground and background modeling based on the color histogram. Unfortunately, the drawback or the disadvantages of the color histogram features still influence their approach and makes less robust than shape HOG correlation filter-based tracker. In this section, we describe our proposed distractor detection method. Given the responses from color histogram r_t^{ch} and correlation filter r_t^{cf}, the maximum value of r_t^{ch} and r_t^{cf} can be determined. The maximum value of r_t^{ch} is represented by v^{ch} and that of r_t^{cf} by v^{cf}. Because these responses take a two-dimensional form, these maximum values have coordinate information indicating their respective positions. Distractor detection can be achieved by using the Euclidean distance between v^{ch} position $(x_{v^{\mathrm{ch}}}, y_{v^{\mathrm{ch}}})$ and v^{cf} position $(x_{v^{\mathrm{cf}}}, y_{v^{\mathrm{cf}}})$:

$$
\begin{aligned}
\delta_t &= \sqrt{\left(x_{v^{\mathrm{ch}}} - x_{v^{\mathrm{cf}}}\right)^2 + \left(y_{v^{\mathrm{ch}}} - y_{v^{\mathrm{cf}}}\right)^2}, \\
\omega &= \begin{cases} 1, & \delta_t \geq \partial_0 \\ 0, & \text{otherwise}, \end{cases}
\end{aligned}
\tag{6}
$$

where δ_t and δ_0 represent the distance at frame t and the distance threshold, respectively, and 1 indicates that a distractor appears and 0 that no distractor appears. The distractor detection procedure is illustrated in Figure 1. Moreover, compared with [8], our proposed distractor detection method does not need a large memory for computation.


Detecting the distractor at frame no. 630. No distractors appears because $\delta_t < \delta_0$.

#630
$(x_{v\mathrm{ch}}, y_{v\mathrm{ch}})$
$(x_{v\mathrm{cf}}, y_{v\mathrm{cf}})$

Response of color histogram

Response of correlation filter

Detecting the distractor at frame no. 637. Distractors appears because $\delta_t \geq \delta_0$.

#637
$(x_{v\mathrm{ch}}, y_{v\mathrm{ch}})$
$(x_{v\mathrm{cf}}, y_{v\mathrm{cf}})$

Response of color histogram

Response of correlation filter

Figure 1: Illustrations of detecting the distractor at frames numbers 630 and 637.

4. Proposed Method

In this section, our proposed method for visual tracking is described. Given frame $t-1$, the rectangle area of an object $|a_{o,t-1}|$, and that of the background $|a_{b,t-1}|$, we calculate certain parameters that are related to each representation before we proceed to frame t, since the proposed method uses a color histogram and correlation filter for representing the target object. First, considering the color histogram-based representation, the parameters $h_{o,t-1}$ and $h_{b,t-1}$ can be calculated based on the pixels in the observation area and the number of bins η_{bins} that are needed. For the correlation filter-based representation, the numerator $\beta_{\mathrm{trans},t-1}$ and denominator $\gamma_{\mathrm{trans},t-1}$ parameters for translation estimation and the numerator $\beta_{\mathrm{scale},t-1}$ and denominator $\gamma_{\mathrm{scale},t-1}$ parameters for scale estimation should be determined. Parameters $\beta_{\mathrm{trans},t-1}$ and $\gamma_{\mathrm{trans},t-1}$ can be calculated by $\beta^i_{\mathrm{trans},t-1} = \overline{D}_{\mathrm{trans},t-1} \odot F^i_{\mathrm{trans},t-1}$ and $\gamma_{\mathrm{trans},t-1} = \sum_{k=1}^{n} \overline{F}^k_{\mathrm{trans},t-1} \odot F^k_{\mathrm{trans},t-1}$, respectively. On the one hand, parameters $\overline{D}_{\mathrm{trans},t-1}$, $F^i_{\mathrm{trans},t-1}$, $\overline{F}^k_{\mathrm{trans},t-1}$, and $F^k_{\mathrm{trans},t-1}$ are the complex conjugate of the DFT of $d_{\mathrm{trans},t-1}$, the DFT of $f^i_{\mathrm{trans},t-1}$, the complex conjugate of the DFT of $f^k_{\mathrm{trans},t-1}$, and the DFT of $f^k_{\mathrm{trans},t-1}$, respectively. On the other hand, parameters $\beta_{\mathrm{scale},t-1}$ and $\gamma_{\mathrm{scale},t-1}$ can be calculated by $\beta^i_{\mathrm{scale},t-1} = \overline{D}_{\mathrm{scale},t-1} \odot F^i_{\mathrm{scale},t-1}$ and $\gamma_{\mathrm{scale},t-1} = \sum_{k=1}^{n} \overline{F}^k_{\mathrm{scale},t-1} \odot F^k_{\mathrm{scale},t-1}$, respectively. Parameters $\overline{D}_{\mathrm{scale},t-1}$, $F^i_{\mathrm{scale},t-1}$, $\overline{F}^k_{\mathrm{scale},t-1}$, and $F^k_{\mathrm{scale},t-1}$ are the complex conjugate of the DFT of $d_{\mathrm{scale},t-1}$, the DFT of $f^i_{\mathrm{scale},t-1}$, the complex conjugate of the DFT of $f^k_{\mathrm{scale},t-1}$, and the DFT of $f^k_{\mathrm{scale},t-1}$, respectively.

After these parameters for frame $t-1$ have been calculated, the search for the target object in frame t can proceed. To search the target object in frame t, the response from color histogram r^{ch}_t and the response from correlation filter r^{cf}_t are needed. Given the search area of the target object $|a_{\mathrm{search},t}|$ at frame t, where $|a_{\mathrm{search},t}| = |a_{b,t}|$, to obtain r^{ch}_t, we use $h_{o,t-1}$ and $h_{b,t-1}$ and, then, implement these parameters in (1), where id_x is related to the pixel at $|a_{\mathrm{search},t}|$. Further, the results of this step are computed by using an integral image in order to obtain r^{ch}_t. On the other hand, translation estimation is used to estimate the location of the target object when the correlation filter-based target object representation is used. Given frame t, translation sample $\phi^i_{\mathrm{trans},t}$ is extracted from $|a_{\mathrm{search},t}|$ within the scale estimation from the previous frame s_{t-1}. After $\phi^i_{\mathrm{trans},t}$ is extracted, the parameters $\beta_{\mathrm{trans},t-1}$ and $\gamma_{\mathrm{trans},t-1}$ are used together with $\phi^i_{\mathrm{trans},t}$ to obtain r^{cf}_t by implementation in (5). Figure 2 shows the proposed method framework.

When the parameters r^{ch}_t and r^{cf}_t have been obtained, in order to minimize the tracking failure due to the distractor, the distractor must be detected prior to the final location estimation of the target object. To detect the distractor, we use (6). The final location estimation of the target object can be obtained by maximizing the score r_t, where

$$r_t = \begin{cases} r^{\mathrm{cf}}_t, & \omega = 1 \\ \theta^{\mathrm{ch}} r^{\mathrm{ch}}_t + \theta^{\mathrm{cf}} r^{\mathrm{cf}}_t, & \text{otherwise,} \end{cases} \tag{7}$$

where θ^{ch} and θ^{cf} are the coefficients related to r^{ch}_t and r^{cf}_t, respectively. According to (7), when the distractor appears, the response from correlation filter r^{cf}_t is selected in order to get final location estimation. This is because color histogram-based representation is less discriminative than correlation

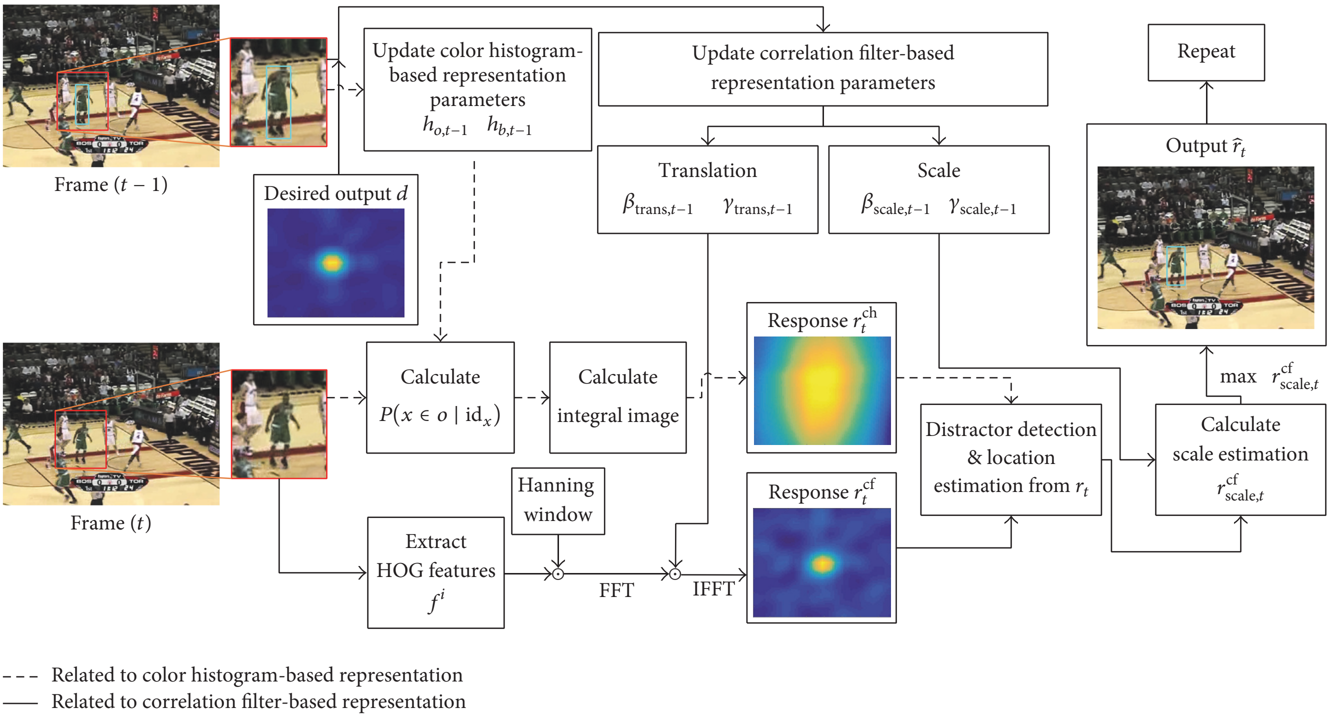

Frame $(t-1)$
Update color histogram-based representation parameters
$h_{o,t-1}$ $h_{b,t-1}$
Update correlation filter-based representation parameters
Repeat
Output $\hat{r}_t$
Translation
$\beta_{\text{trans},t-1}$ $\gamma_{\text{trans},t-1}$
Scale
$\beta_{\text{scale},t-1}$ $\gamma_{\text{scale},t-1}$
Desired output d
Response r_t^{ch}
Calculate $P(x \in o \mid \text{id}_x)$
Calculate integral image
max $r_{\text{scale},t}^{\text{cf}}$
Distractor detection & location estimation from r_t
Calculate scale estimation $r_{\text{scale},t}^{\text{cf}}$
Frame (t)
Hanning window
Extract HOG features f^i
FFT
IFFT
Response r_t^{cf}
- - - Related to color histogram-based representation
—— Related to correlation filter-based representation

Figure 2: Framework of the proposed method.

filter-based representation. This reason is based on the disadvantages of color histogram-based representation, where this representation is often inadequate to discriminative target object from the background, sensitive to motion blur, and can not handle the variation of the illumination well. Besides that, this reason is made based on the benchmark results of the VOT2014 dataset [18, 19] and VOT2015 dataset [15, 20]. The DSST tracker [16], SAMF tracker [21], and KCF tracker [22] occupied the top three rank in the benchmark results of the VOT2014 dataset. These trackers are developed based on shape HOG correlation filter. Furthermore, shape HOG correlation filter-based tracker is always dominant and leading in the accuracy-robustness rank compared to color-based tracker of the VOT2015 benchmark dataset.

Scale changes of the target object also can cause tracking failure. For this reason, scale estimation is required, for which a correlation filter can be used, as proved in [16]. The process is almost the same as for translation estimation. Scale sample $\phi_{\text{scale},t}^{i}$ is extracted from r_t, considering the scale estimation from the previous frame s_{t-1}. After $\phi_{\text{scale},t}^{i}$ is extracted, the parameters $\beta_{\text{scale},t-1}$ and $\gamma_{\text{scale},t-1}$ are used together with $\phi_{\text{scale},t}^{i}$ to obtain $r_{\text{scale},t}^{\text{cf}}$ by implementation in (5). Scale estimation s_t at frame t can be calculated by maximizing the score $r_{\text{scale},t}^{\text{cf}}$. The parameter $r_{\text{scale},t}^{\text{cf}}$ that has the maximum score is represented by the output $\hat{r}_t$ of the proposed method.

Since appearance changes always occur and influence the target object, they also can cause tracking failure. Certain parameters need to be updated to handle this problem. Six parameters should be updated: the parameters $h_{o,t}$ and $h_{b,t}$ for color histogram-based representation and the parameters $\beta_{\text{trans},t}$, $\gamma_{\text{trans},t}$, $\beta_{\text{scale},t}$, and $\gamma_{\text{scale},t}$ for correlation filter-based representation. The parameters $h_{o,t}$ and $h_{b,t}$ can be obtained as

$$\begin{aligned} h_{o,t} &= \left(1-\alpha^{\text{ch}}\right) h_{o,t-1} + \alpha^{\text{ch}} \hat{h}_{o,t}, \\ h_{b,t} &= \left(1-\alpha^{\text{ch}}\right) h_{b,t-1} + \alpha^{\text{ch}} \hat{h}_{b,t}, \end{aligned} \tag{8}$$

where $\hat{h}_{o,t}$ is the color histogram for the target object, $\hat{h}_{b,t}$ is the color histogram for the background, and α^{ch} is a coefficient related to the color histogram-based representation. On the other hand, the samples f_{trans} and f_{scale} should be extracted from frame t at $\hat{r}_t$ and s_t to update the parameters in the correlation filter-based representation, respectively. After the samples have been extracted, the updates of parameters $\beta_{\text{trans},t}$ and $\gamma_{\text{trans},t}$ are determined by using (4) with f_{trans}. Parameters $\beta_{\text{scale},t}$ and $\gamma_{\text{scale},t}$ are also updated by using (4) with f_{scale}.

5. Experimental Results and Discussions

In this section, a comprehensive evaluation of the proposed method is presented. The proposed method is compared on two recently published benchmarks that are widely used: VOT2014 [18, 19] and VOT2015 [15, 20]. The method was implemented in MATLAB 2016A, and the experiment was performed on an Intel(R) Core(TM) i5 2.60 GHZ CPU with 8 GB RAM. For color histogram-based representation, the number of bins η_{bins} that was used was 32 for each channel of a red green blue (RGB) image color format. The value of the parameter α^{ch} for updating the color histogram was 0.01. Further, for the correlation filter-based representation, we used a HOG cell size of 8 × 8. The values of parameters α^{cf}, λ, and δ_0 were 0.01, 0.01, and 20, respectively. When a distractor did

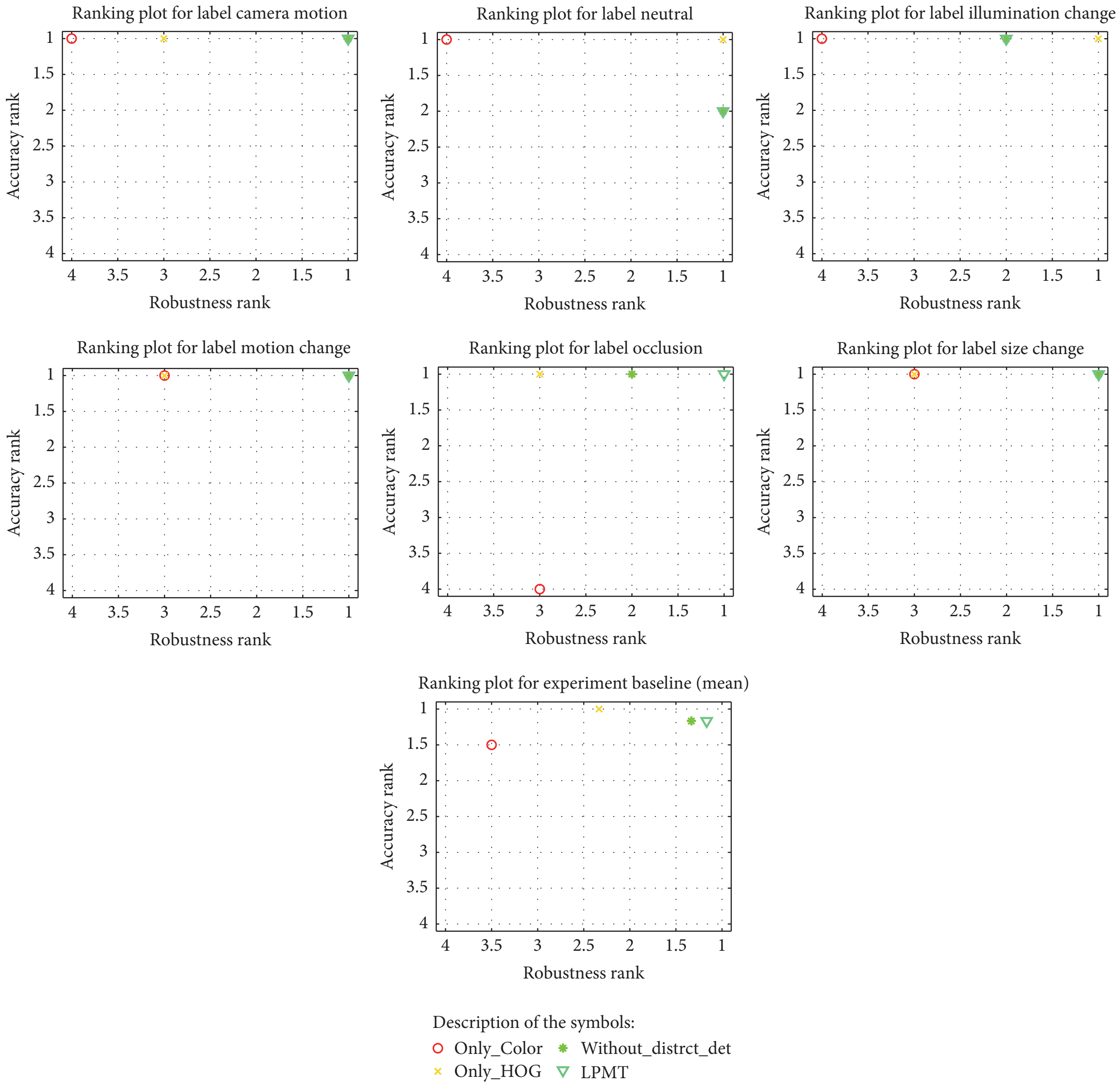


Figure 3: Accuracy-robustness rank plot based on VOT2014 benchmark dataset. The proposed tracker which only uses color histogram features is symbolized by the red circle. The proposed tracker which only uses shape HOG correlation filter is symbolized by the yellow cross. The proposed tracker without distractor detection is symbolized by the green asterisk. The proposed tracker is symbolized by the green triangle.

not appear, parameter r_t was constructed from the merged responses of r_t^{ch} and r_t^{cf}. Thus, coefficients θ^{ch} and θ^{cf} were required. According to the results of our experiments, these coefficients θ^{ch} and θ^{cf} are equal to 0.3 and 0.7, respectively.

The VOT2014 benchmark dataset includes 25 sequences that represent several challenging problems in visual tracking: camera motion, illumination change, motion change, occlusion, and size change. For this benchmark dataset, two performance measures were used: accuracy and robustness. The accuracy parameter was determined as the average per-frame overlap between the bounding box output of the system $\widehat{r}_t$ and the ground truth $\mathrm{BB}_{\mathrm{GT}}$ using the area under curve (AUC) criterion $\mathrm{AUC} = (\widehat{r}_t \cap \mathrm{BB}_{\mathrm{GT}})/(\widehat{r}_t \cup \mathrm{BB}_{\mathrm{GT}})$. Further, the robustness parameter was expressed as the number of failures over the sequence, where a failure is the condition that the AUC is equal to zero. By using this benchmark dataset and in order to justify the design choice of the proposed method which uses compLementary learners for rePresentation Model of the target object and detecTing the distractor (LPMT), this proposed method is compared with the proposed method without distractor detection, the proposed method which uses only shape HOG features, and the proposed method which uses only color histogram features. Figure 3 shows the results of these comparisons.

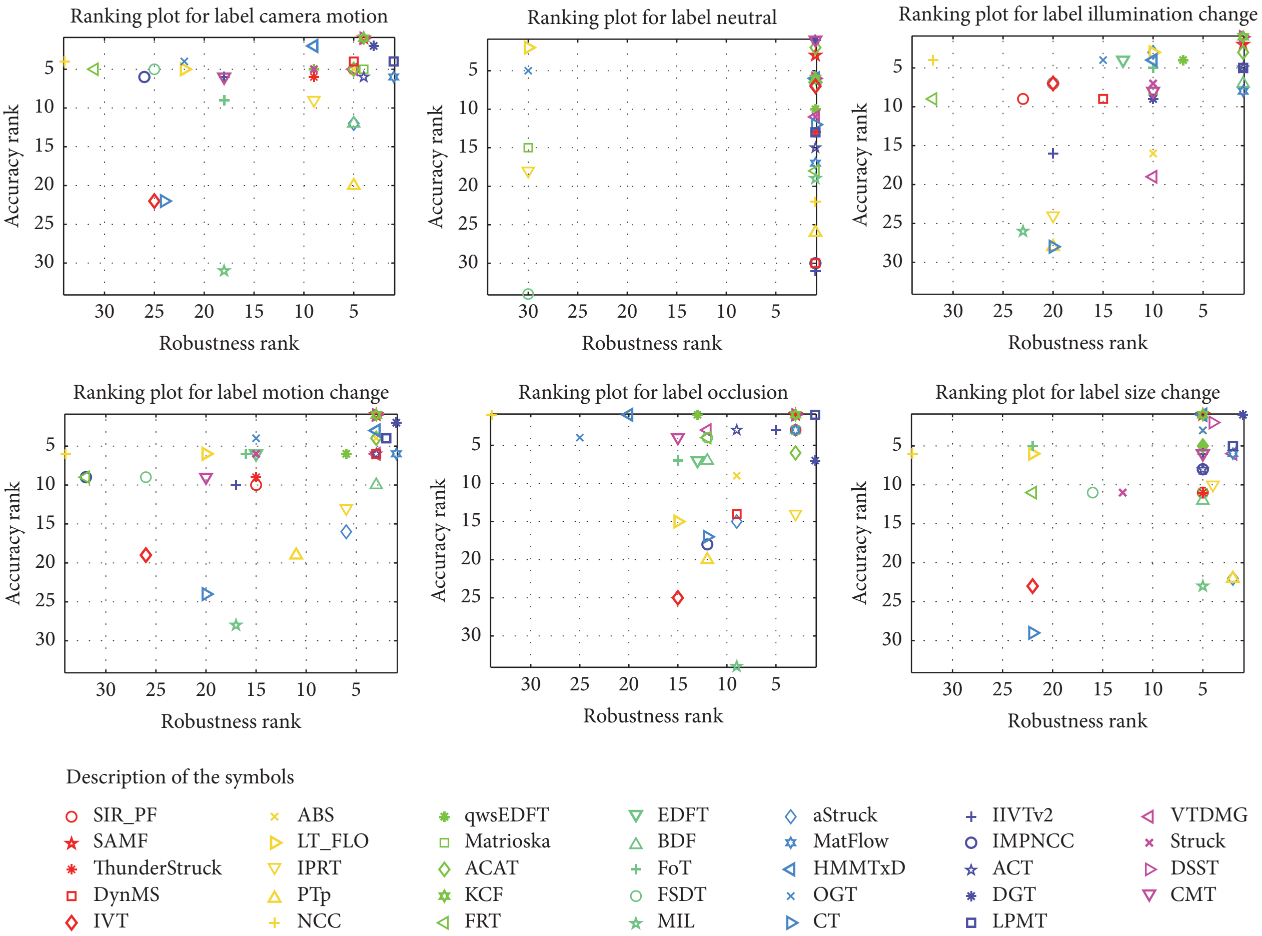
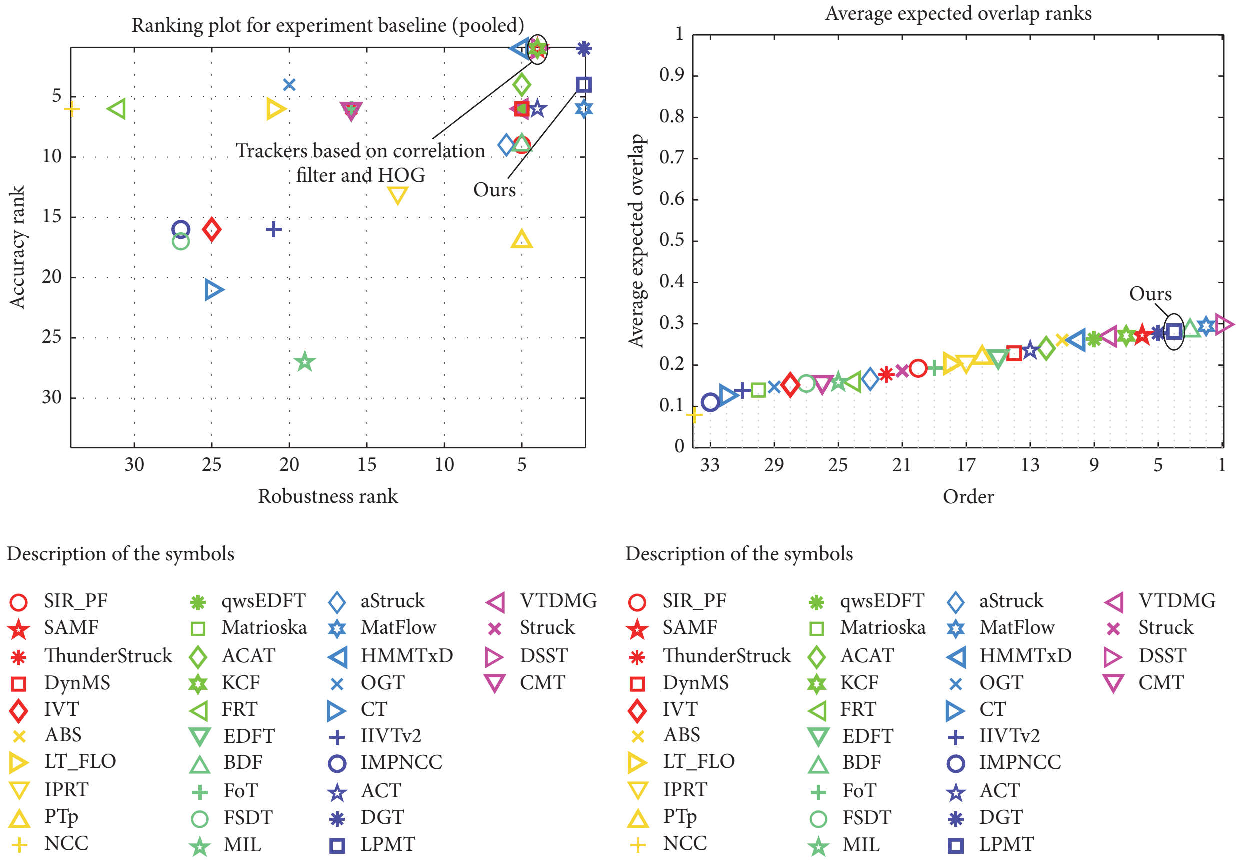


Figure 4: Accuracy-robustness rank plots of LPMT and the state-of-the-art tracker on the VOT2014 benchmark datasets for the experimental baseline of the following challenges: camera motion, neutral, illumination change, motion change, occlusion, and size change. The accuracy and robustness ranks are plotted along the vertical and horizontal axis, respectively. LPMT is represented by the purple square.

Using the VOT2014 benchmark dataset, the proposed method is also compared with the state-of-the-art visual tracking algorithms: SIR_PF [23], SAMF [21], ThunderStruck [24], DynMS [23], IVT [25], ABS - [23], LT_FLO [26], IPRT [23], PTp [23], NCC [27], qwsEDFT [28], Matrioska [29], ACAT [23], KCF [22], FRT [30], EDFT [31], BDF [32], FoT [33], FSDT [23], MIL [7], aStruck [23], MatFlow [23], HMMTxD [23], OGT [34], CT [35], IIVTv2 [23], IMPNCC [23], ACT [36], DGT [37], VTDMG [23], Struck [24], DSST [16], and CMT [38]. Based on the accuracy parameter and the robustness parameter, the accuracy-robustness (AR) rank plot is used to determine the comparative rank of the methods.

Figure 4 shows the AR rank plots of LPMT and the state-of-the-art methods for the challenges of camera motion, illumination change, motion change, occlusion, and size change. For each challenge, LPMT shows a good performance: it is always ranked in the top 5 among all the 33 trackers. In particular, in the occlusion challenge, where most trackers fail because of this problem and the problem is coupled with the disruption caused by the presence of an object similar to the target object, LPMT outperforms the other state-of-the-art algorithms. This proves that the proposed method meets these challenges effectively. The definition of neutral in this figure is that no challenge exists in the sequence frame.

Figure 5 shows the AR rank plots of LPMT and the state-of-the-art trackers on the VOT2014 benchmark dataset for all the challenges combined and the average expected overlap rank. Since LPMT showed a good performance according to the AR plot rank for each challenge, where it was always ranked in the top five, this method also ranked in the top five for the overall challenges. Based on the average expected overlap, the LPMT was ranked fourth, where the average expected overlap is almost 0.3. In the average expected overlap parameter of this benchmark dataset, DSST [16] achieved the top rank, which has an average expected overlap equal to 0.3. This method uses HOG and grayscale features. For detailed information about the VOT2014 benchmark dataset and its performance parameters, please refer to [18, 19].



(a)

(b)

Figure 5: Accuracy-robustness rank plot of LPMT and the state-of-the-art trackers on the VOT2014 benchmark dataset for the baseline experiment of the overall challenges (a) and the average expected overlap rank (b). In the accuracy-robustness rank plot, the accuracy and robustness rank are plotted along the vertical and horizontal axis, respectively. Our LPMT is represented by the purple square.

In the VOT2015 benchmark dataset, there are 60 sequences that represent more challenging problems than those in the VOT2014 dataset. As for the VOT2014 benchmark dataset, the accuracy and robustness performance parameters were used, which are represented by the AR rank plot. By using this benchmark dataset and in order to justify the design choice of the proposed method LPMT, this proposed method is compared with the proposed method without distractor detection, the proposed method which uses only shape HOG features, and the proposed method which uses only color histogram features. Figure 6 shows the results of these comparisons.

Based on the VOT2015 benchmark dataset, for all of the proposed trackers, the proposed tracker without distractor detection, the proposed tracker which only uses shape HOG features, and the proposed tracker which only uses color histogram features achieve the accuracy rank of 1.00. Furthermore, the robustness rank baseline mean of the proposed tracker, the proposed tracker without distractor detection, the proposed tracker which only uses shape HOG features, and the proposed tracker which only uses color histogram features are 1.00, 1.33, 2.83, and 3.33, respectively. According to the results, these prove that color histogram features are less robust than shape HOG features. It indicates that color histogram features are less discriminative than shape HOG features. Furthermore, these results also prove that the proposed tracker which uses distractor detection can improve the robustness compared to the proposed tracker without distractor detection.

Using the VOT2015 benchmark dataset, the proposed LPMT method was compared with the state-of-the-art visual tracking algorithms: ACT [31], CT [35], ggt [15], L1APG [39], mkcf_plus [15], RobStruck [15], STC [40], amt [15], DAT [17], HMMTxD [15], LGT [41], muster [42], s3Tracker [15], sumshift [43], AOGTracker [15], DFT [15], HT [15], loft_lite [15], mvcft [15], samf [21], TGPR [44], ASMS [45], DSST [16], IVT [25], LT_FLO [26], ncc [27], SCBT [46], tric [15], dtracker [15], kcf_mtsa [15], matflow [15], OAB [15], sKCF [15], zhang [15], bdf [15], fct [15], KCF2 [15], MCT [15], OACF [11], sme [15], cmil [15], fot [33], kcfdp [15], MEEM [47], PKLTF [15], SODLT [48], CMT [38], FragTrack [15], kcfv2 [15], MIL [7], rajssc [15], and srat [15].

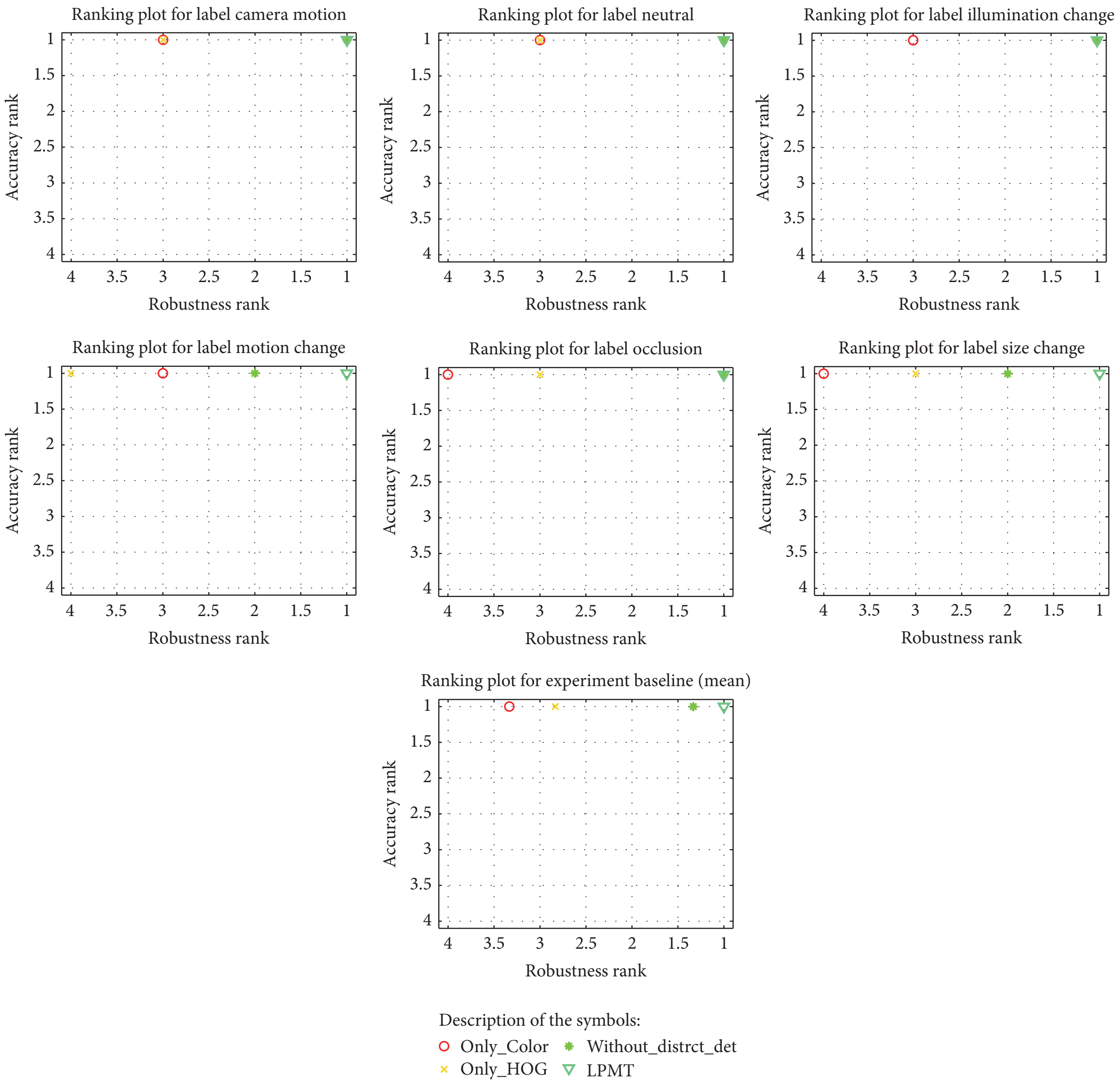


Figure 6: Accuracy-robustness rank plot based on VOT2015 benchmark dataset. The proposed method which only uses color histogram features is symbolized by the red circle. The proposed method which only uses shape HOG correlation filter is symbolized by the yellow cross. The proposed method without distractor detection is symbolized by the green asterisk. The proposed method is symbolized by the green triangle.

Figure 7 shows the AR rank plots of LPMT and the state-of-the-art methods for the challenges of camera motion, illumination change, motion change, occlusion, and size change. For each challenge, surprisingly, LPMT shows a good performance, being always ranked in the top two. This proves that the proposed method addresses these problems effectively, where these problems are more challenging than those in the VOT2014 benchmark dataset and the number of sequences is also greater. This condition is inversely proportional to DSST, which in this experiment achieved a rank considerably below that of LPMT.

Figure 8 shows the AR rank plots of LPMT and the state-of-the-art trackers on the VOT2015 benchmark dataset of the overall challenges and the average expected overlap rank. Since LPMT shows a good performance in the AR rank plot for each challenge, where it was always ranked in the top two, for the overall challenges, this method outperforms the other state-of-the-art tracker. Based on the average expected

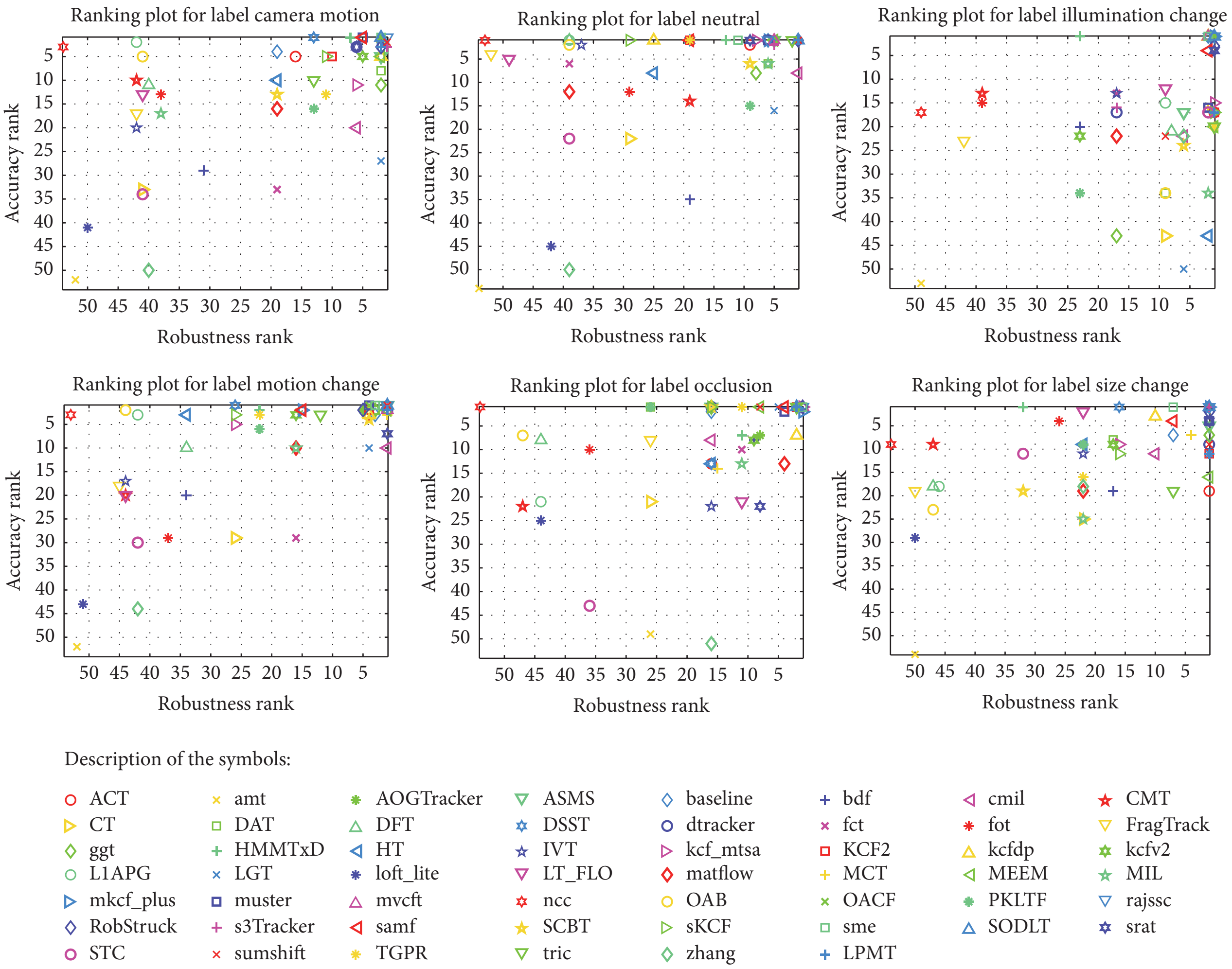


Figure 7: Accuracy-robustness rank plot for the baseline experiment of the challenges of camera motion, neutral, illumination change, motion change, occlusion, and size change on the VOT2015 benchmark dataset. The accuracy and robustness rank are plotted along the vertical and horizontal axis, respectively. LPMT is represented by the blue plus-sign.

overlap, the LPMT achieves the first rank, where the average expected overlap is equal to 0.25. In the average expected overlap parameter of this benchmark dataset, DSST [16], which achieved the top rank on the VOT2014 benchmark dataset, was ranked the thirtieth. The second rank is achieved by the Rajssc tracker, which is based on a correlation filter. For detailed information about the VOT2015 benchmark dataset and its performance parameters, please refer to [15, 20].

6. Conclusions

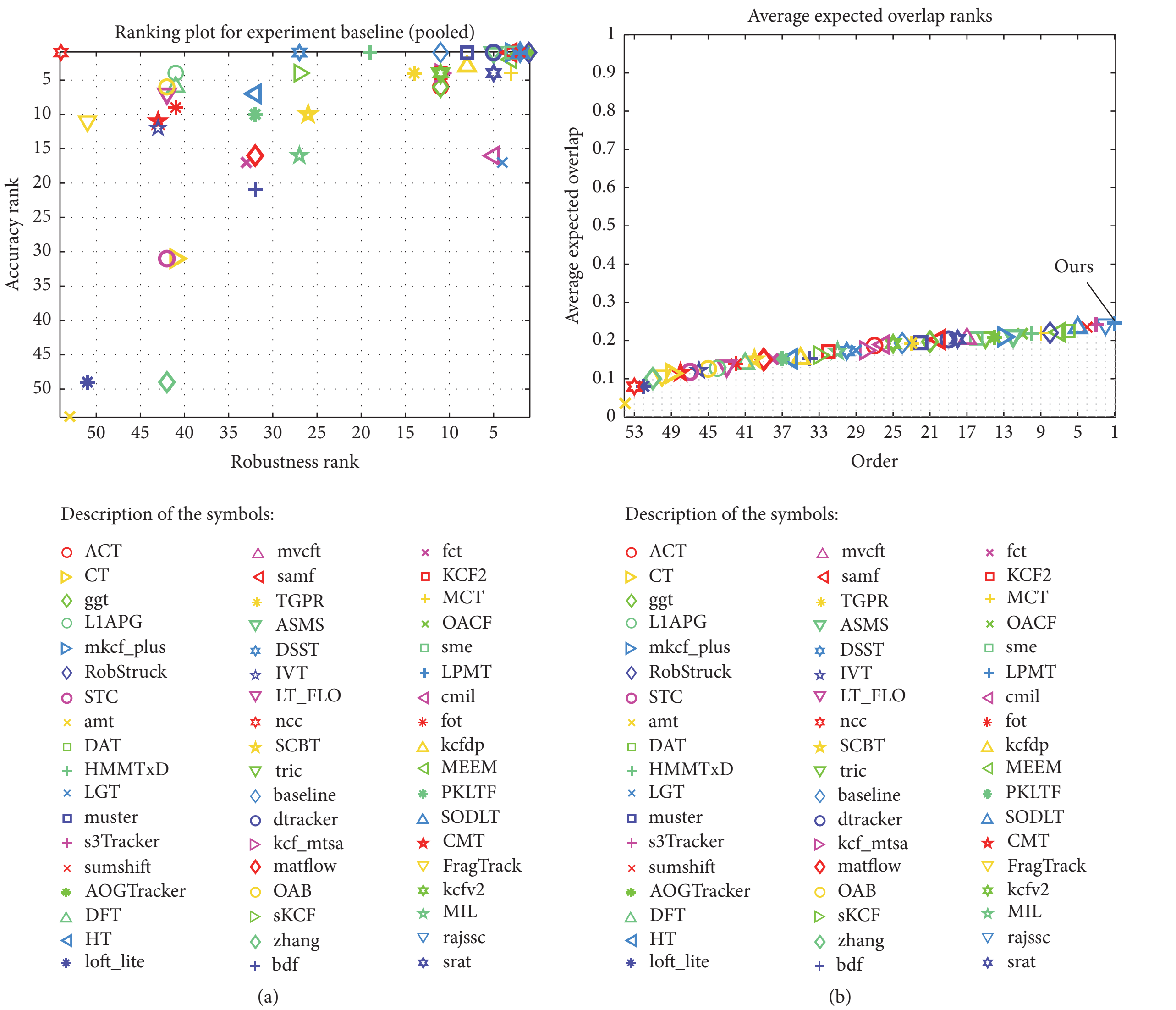
This paper presented a method that uses complementary learners, which consist of the response of the color histogram and the response of the correlation filter, for representing the target object. To overcome a distractor that has a representation similar to that of the target object, the proposed method also detects the distractor based on the response of the color histogram and correlation filter. Based on evaluations on the VOT2014 and VOT2015 benchmark datasets, the proposed method yields a favorable performance as compared to several state-of-the-art visual tracking algorithms.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by BK21PLUS, Creative Human Resource Development Program for IT Convergence, and was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MEST) (no. 2010-0024110).



(a)

(b)

Figure 8: Accuracy-robustness rank plot for the baseline experiment of the overall challenges (a) and the average expected overlap rank (b) on the VOT2015 benchmark dataset. In the accuracy-robustness rank plot, the accuracy and robustness rank are plotted along the vertical and horizontal axis, respectively. Our LPMT is represented by the blue plus-sign.

References

[1] M. Kim, “Adaptive Bayesian object tracking with histograms of dense local image descriptors,” *The International Journal of Fuzzy Logic and Intelligent Systems*, vol. 16, no. 2, pp. 104–110, 2016.

[2] S.-K. Weng, C.-M. Kuo, and S.-K. Tu, “Video object tracking using adaptive Kalman filter,” *Journal of Visual Communication and Image Representation*, vol. 17, no. 6, pp. 1190–1208, 2006.

[3] K. Nummiaro, E. Koller-Meier, and L. Van Gool, “An adaptive color-based particle filter,” *Image and Vision Computing*, vol. 21, no. 1, pp. 99–110, 2003.

[4] H. Liu, L. Wang, and F. Sun, “Mean-shift tracking using fuzzy coding histogram,” *International Journal of Fuzzy Systems*, vol. 16, no. 4, pp. 457–467, 2014.

[5] H. Grabner, M. Grabner, and H. Bischof, “Real-time tracking via on-line boosting,” in *Proceedings of the 17th British Machine Vision Conference (BMVC ’06)*, pp. 47–56, Edinburgh, Scotland, September 2006.

[6] H. Grabner, C. Leitsner, and H. Bischof, “Semi-supervised on-line boosting for robust tracking,” in *Proceedings of the European Conference on Computer Vision (ECCV ’08)*, pp. 234–247, Marseille, France, October 2008.

[7] B. Babenko, M.-H. Yang, and S. Belongie, “Robust object tracking with online multiple instance learning,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1619–1632, 2011.

[8] Z. Kalal, K. Mikolajczyk, and J. Matas, “Tracking-learning-detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, 2012.

[9] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, “Visual object tracking using adaptive correlation filters,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’10)*, pp. 2544–2550, San Francisco, Calif, USA, June 2010.

[10] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, “Staple: complementary learners for real-time tracking,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’16)*, pp. 1401–1409, IEEE, Las Vegas, Nev, USA, June 2016.

[11] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, “The importance of estimating object extent when tracking with correlation filter,” Report for the Visual Object Tracking Workshop, 2015.

[12] L. Zhang, Y. Wang, H. Sun, Z. Yao, and S. He, “Robust visual correlation tracking,” *Mathematical Problems in Engineering*, vol. 2015, Article ID 238971, 13 pages, 2015.

[13] J. F. Henriques, J. Carreira, R. Caseiro, and J. Batista, “Beyond hard negative mining: efficient detector learning via block-circulant decomposition,” in *Proceedings of the 14th IEEE International Conference on Computer Vision (ICCV ’13)*, pp. 2760–2767, Sydney, Australia, December 2013.

[14] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, “Learning spatially regularized correlation filters for visual tracking,” in *Proceedings of the 15th IEEE International Conference on Computer Vision (ICCV ’15)*, pp. 4310–4318, Santiago, Chile, December 2015.

[15] M. Kristan, J. Matas, A. Leonardis et al., “The visual object tracking vot2015 challenge results,” in *Proceedings of the IEEE International Conference on Computer Vision Workshop (ICCV ’15)*, pp. 1–23, IEEE, Santiago, Chile, December 2015.

[16] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, “Accurate scale estimation for robust visual tracking,” in *Proceedings of the British Machine Vision Conference (BMVC ’14)*, pp. 1–11, BMVA Press, Nottingham, UK, September 2014.

[17] H. Possegger, T. Mauthner, and H. Bischof, “In defense of color-based model-free tracking,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’15)*, pp. 2113–2120, Boston, Mass, USA, June 2015.

[18] M. Kristan, J. Matas, A. Leonardis et al., “A novel performance evaluation methodology for single-target trackers,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 11, pp. 2137–2155, 2016.

[19] L. Cehovin, A. S. Leonardis, and M. Kristan, “Visual object tracking performance measures revisited,” *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1261–1274, 2016.

[20] http://www.votchallenge.net/.

[21] L. Yang and Z. Jianke, “A scale adaptive kernel correlation filter tracker with feature integration,” in *Proceedings of the European Conference on Computer Vision, Workshop on Visual Object Tracking Challenge (ECCV ’14)*, pp. 254–265, Zurich, Switzerland, 2014.

[22] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, “High-speed tracking with kernelized correlation filters,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015.

[23] M. Kristan, R. Pflugfelder, A. Leonardis et al., “The Visual Object Tracking VOT2014 challenge results,” in *Proceedings of the European Conference on Computer Vision Workshops (ECCV ’14) and the Workshop on Visual Object Tracking Challenge*, pp. 98–111, Zurich, Switzerland, September 2014.

[24] S. Hare, A. Saffari, and P. H. S. Torr, “Struck: structured output tracking with kernels,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV ’11)*, pp. 263–270, Barcelona, Spain, November 2011.

[25] 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.

[26] K. Lebeda, S. Hadfield, J. Matas, and R. Bowden, “Long-term tracking through failure cases,” in *Proceedings of the 14th IEEE International Conference on Computer Vision Workshops (ICCVW ’13)*, pp. 153–160, Sydney, Australia, December 2013.

[27] K. Briechle and U. D. Hanebeck, “Template matching using fast normalized cross correlation,” in *Optical Pattern Recognition XII*, vol. 4387 of *Proceedings of SPIE*, pp. 95–102, Orlando, Fla, USA, April 2001.

[28] K. Ofjall and M. Felsberg, “Weighted update and comparison for channel-based distribution field tracking,” in *Proceedings of the European Conference on Computer Vision Workshop on Visual Object Tracking Challenge (ECCV ’14)*, pp. 218–231, Zurich, Switzerland, September 2014.

[29] M. E. Maresca and A. Petrosino, “Matrioska: a multi-level approach to fast tracking by learning,” in *Proceedings of the International Conference on Image Analysis and Processing (ICIAP ’13)*, pp. 419–428, Naples, Italy, September 2013.

[30] A. Adam, E. Rivlin, and I. Shimshoni, “Robust fragments-based tracking using the integral histogram,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’06)*, pp. 798–805, New York, NY, USA, June 2006.

[31] M. Felsberg, “Enhanced distribution field tracking using channel representations,” in *Proceedings of the 14th IEEE International Conference on Computer Vision Workshops (ICCVW ’13)*, pp. 121–128, IEEE, Sydney, Australia, December 2013.

[32] M. Maresca and A. Petrosino, “Clustering local motion estimates for robust and efficient object tracking,” in *Proceedings of the European Conference on Computer Vision Workshops (ECCV ’14) and the Workshop on Visual Object Tracking Challenge*, pp. 244–253, Zurich, Switzerland, September 2014.

[33] T. Vojir and J. Matas, “Robustifying the flock of trackers,” in *Proceedings of the Computer Vision Winter Workshop*, pp. 91–97, Styria, Austria, February 2011.

[34] H. Nam, S. Hong, and B. Han, “Online graph-based tracking,” in *Proceedings of the European Conference on Computer Vision Workshops (ECCV ’14) and the Workshop on Visual Object Tracking Challenge*, pp. 112–126, Zurich, Switzerland, September 2014.

[35] K. Zhang, L. Zhang, and M.-H. Yang, “Real-time compressive tracking,” in *Proceedings of the European Conference on Computer Vision (ECCV ’12)*, pp. 864–877, Florence, Italy, October 2012.

[36] M. Danelljan, F. S. Khan, M. Felsberg, and J. Van De Weijer, “Adaptive color attributes for real-time visual tracking,” in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR ’14)*, pp. 1090–1097, Columbus, Ohio, USA, June 2014.

[37] Z. Cai, L. Wen, Z. Lei, N. Vasconcelos, and S. Z. Li, “Robust deformable and occluded object tracking with dynamic graph,” *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5497–5509, 2014.

[38] G. Nebehay and R. Pflugfelder, “Consensus-based matching and tracking of keypoints for object tracking,” in *Proceedings of the IEEE Winter Conference on Application of Computer Vision (WACV ’14)*, pp. 862–869, IEEE, Steamboat Springs, Colo, USA, March 2014.

[39] C. Bao, Y. Wu, H. Ling, and H. Ji, “Real time robust L1 tracker using accelerated proximal gradient approach,” in *Proceedings of*

the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12), pp. 1830–1837, Providence, RI, USA, June 2012.

[40] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *Proceedings of the European Conference on Computer Vision, Workshop on Visual Object Tracking Challenge (ECCV '14)*, pp. 127–141, Zurich, Switzerland, September 2014.

[41] L. Čehovin, M. Kristan, and A. Leonardis, "Robust visual tracking using an adaptive coupled-layer visual model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 941–953, 2013.

[42] Z. Hong, Z. Chen, C. Wang, X. Mei, D. Prokhorov, and D. Tao, "MUlti-Store Tracker (MUSTer): a cognitive psychology inspired approach to object tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '15)*, pp. 749–758, Boston, Mass, USA, June 2015.

[43] J.-Y. Lee and W. Yu, "Visual tracking by partition-based histogram backprojection and maximum support criteria," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '11)*, pp. 2860–2865, IEEE, Phuket, Thailand, December 2011.

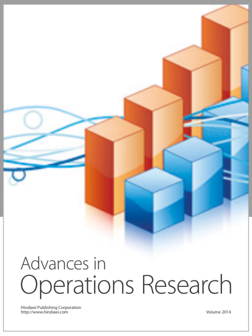
[44] J. Gao, H. Ling, W. Hu, and J. Xing, "Transfer learning based visual tracking with gaussian processes regression," in *Proceedings of the European Conference on Computer Vision Workshops (ECCV '14) and the Workshop on Visual Object Tracking Challenge*, pp. 188–203, Zurich, Switzerland, September 2014.

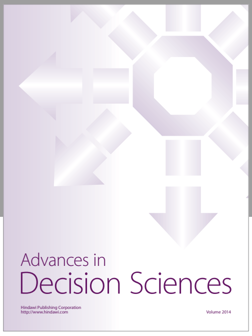
[45] T. Vojir, J. Noskova, and J. Matas, "Robust scale-adaptive mean-shift for tracking," *Pattern Recognition Letters*, vol. 49, pp. 250–258, 2014.

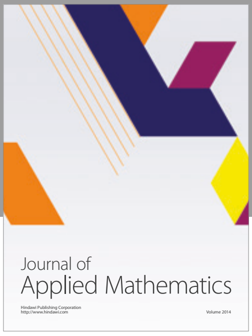
[46] S. Moujtahid, S. Duffner, and A. Baskurt, "Classifying global scene context for on-line multiple tracker selection," in *Proceedings of the British Machine Vision Conference (BMVC '15)*, pp. 163.1–163.12, Swansea, UK, September 2015.

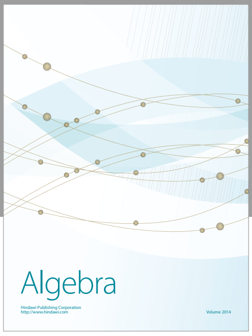
[47] J. Zhang, S. Ma, and S. Sclaroff, "Meem: robust tracking via multiple experts using entropy minimization," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Patern Recognition (CVPR '14)*, pp. 188–203, IEEE, Columbus, Ohio, USA, June 2014.

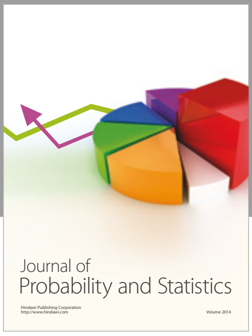
[48] N. Wang, S. Li, A. Gupta, and D.-Y. Yeung, "Transferring rich feature hierarchies for robust visual tracking," https://arxiv.org/abs/1501.04587.


Advances in
Operations Research

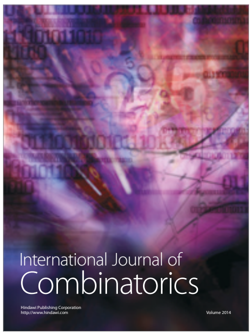

Advances in
Decision Sciences

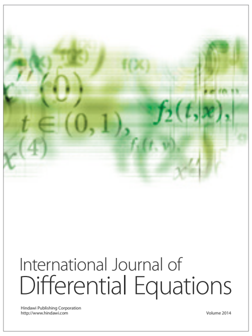

Journal of
Applied Mathematics

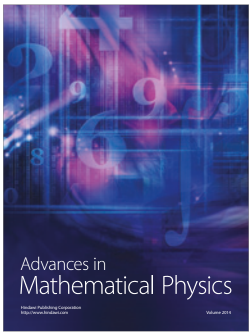

Algebra

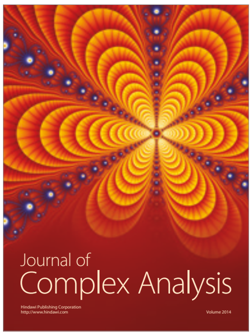

Journal of
Probability and Statistics

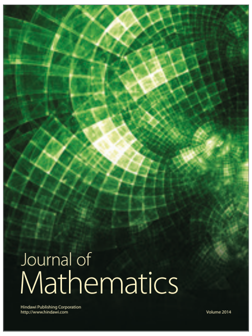
The Scientific
World Journal

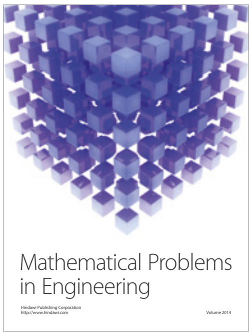

International Journal of
Combinatorics

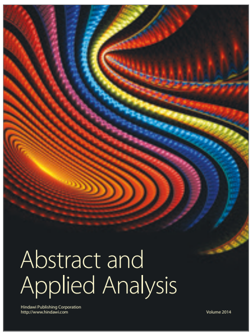

International Journal of
Differential Equations

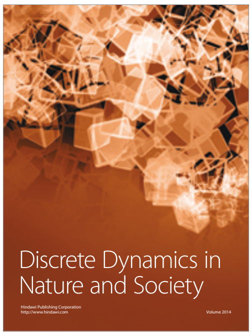

Advances in
Mathematical Physics

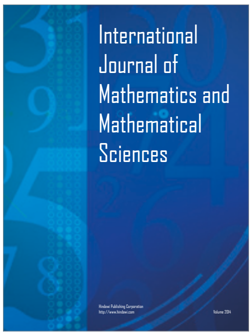

Journal of
Complex Analysis


Journal of
Mathematics


Mathematical Problems
in Engineering

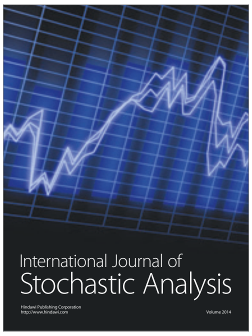

Abstract and
Applied Analysis


Discrete Dynamics in
Nature and Society


International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces


International Journal of
Stochastic Analysis

Journal of
Optimization