IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 26, 2016, accepted December 14, 2016, date of publication March 30, 2017, date of current version April 24, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2673400

Object Tracking Based on Stable Feature Mining Using Intraframe Clustering and Interframe Association

HONG LU[1], (Member, IEEE), KE GU[2], WEISI LIN[3], (Fellow, IEEE), AND WENJUN ZHANG[4], (Fellow, IEEE)
[1]School of Automation, Nanjing Institute of Technology, Nanjing 211167, China
[2]BJUT Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[3]School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798
[4]Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding author: K. Gu (guke.doctor@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61305011 and in part by the Jiangsu Government Scholarship for Overseas Studies of China under Grant JS-2014-126.

ABSTRACT Extracting stable features to enhance object representation has proved to be very effective in improving the performance of object tracking. To achieve this, mining techniques, such as K-means clustering and data associating, are often adopted. However, K-means clustering needs the pre-set number of clusters. Real scenarios (heavy occlusion and so on) often make the tracker lose the target object. To handle these problems, we propose an intraframe clustering and interframe association (ICIA)-based stable feature mining algorithm for object tracking. The value (in HSV space) peak contour is employed to automatically estimate the number of clusters and classify value and saturation colors of the object region to get connected subregions. Every subregion is described with observation and increment models. Multi-feature distances-based subregion association, between the current object template and the current observation, is then utilized to mine stable subregion pairs and obtain feature change ratio. Stable subregion displacements, and current detected and historical trajectories are systematically fused to locate the object. And, stable and unstable subregion features are updated separately to restrain the accumulative error. Experimental comparisons are conducted on six test sequences. Compared with several relevant state-of-the-art algorithms, the proposed ICIA tracker most accurately locates objects in four sequences and shows the second-best performance in the other two sequences with only less 1 pixel distance difference than the best method.

INDEX TERMS Object tracking, stable features mining, intraframe clustering, interframe association, observation and increment models, template update.

I. INTRODUCTION

Visual tracking is a hard problem as many different and varying circumstances need to be reconciled in one algorithm [1]. Modeling the object with sparse and reliable features to increase object representation efficiency is a significant and challenging issue in object tracking [2]–[4], and also widely facilitates numerous other computer vision applications such as action retrieval [5] and recognition [6], [7]. Nonetheless, object features often vary along with object posture and resolution variations, occlusion and background clutters [8] etc in real-world conditions, which makes the current object observation depart from the predetermined object model. Recently, evolving the appearance model is widely adopted so as to adapt to changing imaging conditions. However, appearance model adaptation introduces several challenges, such as simultaneously fulfilling the contradicting goals of rapid learning and stable memory (referred to as the stability/plasticity dilemma) [9]. A target object may be missed due to the wrong observation and the model drift [10] problems during detecting, recognizing and tracking, and this ultimately degrades the accurateness of real applications such as intelligent traffic surveillance [11], image understanding [12] and UAV (Unmanned Aerial Vehicle) navigation [13]. The tracker is further challenged by heavy or total occlusion, similar color disturbance, blurry and ambiguous object appearance, fast movement and rotation etc. Therefore, efficiently

2169-3536 © 2017 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

modeling and tracking objects to improve the robustness, accurateness and real-time performance of the tracker are still an open problem attracting a broad range of researchers' attention.

To cope with these difficulties, a number of algorithms have been established to endow the object model with strong discriminative power, such as kernel [14], [15] and mask [16] color histograms, color correlogram [2], spatio-temporal [17]–[19] and edge-color [3] appearance contexts, context-aware object model [20]–[22], correlation filter model [23], [24], local dynamic sparse model [4], [25], centroid [26], block or patch [27], [28], structure [29], [30], shape [31] and linear combination of basis samples [32] based object feature representations. Among of them, image and video data mining based feature extracting algorithms [20]–[22], [33] play a more critical role towards obtaining remarkable results.

Image mining can acquire the implicit knowledge, image data relationship or other patterns not explicitly stored in the object image [12], [34]. Clustering, as one of the unsupervised image mining techniques, can dispatch the heterogeneous data into different groups [35], [36] or mine visual patterns according to the image content without the priori knowledge [22], [37], [38]. As thus, clustering is very useful in finding the block-level features on the object appearance by grouping unlabeled raw images into meaningful classes, and vital in getting the object observation even though the observable information is incomplete or ambiguous under the cases of occlusion and blurry object appearance. K-Means clustering is good at highlighting the main color or texture features etc and serving dimensionality reduction on the original data [39], but it needs the pre-set number of clusters. Plant et al. [12] utilized the interaction K-Means (IKM) to cluster the functional magnetic resonance image with the pre-set number of clusters for brain function understanding. Wang et al. [21] presented a regularized K-Means formulation where spatial co-occurrences as constraints are added to the conventional K-Means clustering to improve the pattern discovery results. Li [40] proposed a color model based on the K-Means clustering to automatically divide the color space of the object and get the histogram bins. Likewise, [33], [41], [42] also adopted the true or pre-set number of clusters. To adaptively acquire the number of clusters, Pelleg and Moore [43] extended K-Means with efficient estimation of the number of clusters. Kuncheva and Vetrov [44] studied the relationship between stability and accuracy with respect to the number of clusters, and presented that this relationship strongly depends on the data set. They further used the stability measures to select the number of clusters based on the hypothesis of a point of stability of a clustering algorithm corresponding to a structure found in the data. In [45], the number of clusters was adjusted dynamically to arrive at the correct number of clusters even when the number of clusters in the first frame is not correctly chosen.

Video mining can extract the moving object features, spatial or temporal correlations of those features [2]. When an object encounters occlusion, similar feature disturbance, ambiguous appearance, fast movement or rotation etc, these correlations are especially helpful in determining stable features and correcting the ambiguous observation. Video mining is widely used for discovering the object activity and event [11], [46], and tracking the object [20], [47], [48] etc without any assumption about video contents. Association [49] as an important video mining method can obtain related information or discover two features or objects that always occur simultaneously etc. Yang et al. [20] discovered auxiliary objects, i.e. a set of color regions which were temporally stable and spatially correlated to the target object in a video sequence, by learning their co-occurrence associations and estimating affine motion models in an unsupervised way. Next these auxiliary observations were fused to track the target object whose current observation was unreliable due to occlusion or background disturbance etc. Grabner et al. [47] mined supporters which were temporally but useful for tracking the object from the embedded context and dealing with the occlusion in real scenarios. Quan et al. [48] proposed a collaboration model in which the acceleration difference between two objects was used to calculate the motion correlation value based on the two-dimensional Gaussian function and the location of occluded target was estimated using the motion information from other objects. Zhang et al. [11] learned and mined the object-specific context and the scene-specific context informations to improve the robustness of objects (pedestrians and vehicles) classification and objects tracking based abnormal event detection. Wang and Yagi [50] selected reliable features from color and shape-texture cues according to their descriptive ability and extended the standard mean-shift tracking algorithm. Yang et al. [51] exploited the temporal consistency for visual tracking and incorporated the temporal consistency into the multi-graph learning framework to effectively improve the robustness of the tracker.

Nevertheless, a tracker may be good at handling several kinds of circumstances (such as appearance changes induced by different viewpoints), but may be hard to deal with other situations (such as accelerant movement) [1]. So there is still a long way to achieve accurately and robustly object tracking under general scenes.

In this paper, we focus on mining stable object features to elevate the performance of object tracking under the scenarios of heavy occlusion and multiple times occlusions, similar color disturbance and ambiguous appearance, and rapid movement along with scale and posture changes, etc, and hereby present an intraframe clustering and interframe association (ICIA) algorithm. A cluster number is real-time estimated according to the peak contour of V (value) component, and input to K-Means algorithm to classify the S (saturation) and V colors (in HSV space) of the object region. The connection subregions of every cluster are then represented with the observation model and the increment model. To mine stable subregions, we associate the subregion observations with the current object template in terms of the multi-feature distances and change ratios, and utilize the increment model to get the feature variations for further robustly updating the object template. The object trajectory in the current frame is located by weighted fusing the center displacements of stable subregions, and current detected and historical trajectories. We perform comparative experiments with the classical Mean Shift (MS) tracker [14], as well as state-of-the-art Kernelized Correlation Filter (KCF) [24], spatio-temporal context (STC) [17] and structure complexity coefficients (SCC) [29] based trackers. Experimental results indicate that the proposed method can continuously mine stable subregions and robustly track the object under unconstrained scenarios.

The remainder of this paper is organized as follows. Section 2 introduces the overview of the proposed method. Section 3 proposes the number of clusters estimating and intraframe classifying algorithm. Section 4 describes the subregion modeling and stable feature mining algorithm. Section 5 gives the object tracking and template updating algorithm. Section 6 presents experiments and associated results, and performance comparisons. Finally some conclusion and discussion are presented in Section 7.

II. OVERVIEW OF THE PROPOSED ALGORITHM

In video based automatic object recognition and tracking, the motion region containing a moving object is often extracted to decrease the searching cost and restrain background disturbance etc. Similarly, the adaptive background difference here is firstly employed to extract the region of moving object. And then we utilize the peak contour of V component histogram in the object region to estimate the number of clusters and adopt K-Means algorithm to classify the S and V colors of moving pixels. The 8-connection subregions of every cluster are further computed and described with the observation model and the increment model. Furthermore, the subregion association between the current observation model and the current object template is used to mine the stable subregion pairs and get the feature change ratios. Finally, the center displacements of the stable subregions, the current detected object trajectory $\mathbf{x}_t^{\det ect}$ and historical trajectories $\mathbf{x}_{t-1}$ are weighted combined to derive the object displacement and the trajectory $\mathbf{x}_t$ in the current frame. The stable and unstable subregion features in the object template are updated individually to restrain the accumulative error and adapt to gradual variations in object location, posture, scale, color and illumination etc. The flow chart of the proposed ICIA based stable feature mining for object tracking algorithm is shown in Fig. 1, where every subregion center position is marked with the '+' (plus) character and the round mark denotes the tracked object trajectory in frame *t*.

In comparison to previous works, such as [12], [14], [17], [21], [24], [29], [40], three main contributions of this paper are summarized below: 1) we establish a novel ICIA based stable feature mining and object tracking framework by automatically estimating the number of clusters, modeling subregions, mining stable subregions and employing stable feature pairs to derive object trajectory. 2) we develop a dynamic and local stable feature model and update scheme, which are more efficient than the global feature model and the uniform segmentation based feature model in handling heavy occlusion accompanied with similar color disturbance, ambiguous appearance or clutter background, and fast movement along with scale or posture changes etc. 3) our tracker performs better than the classical kernel color histogram based MS [14], and recently developed KCF [24], STC [17] and SCC [29] trackers on relevant image databases. Compared with the previous works, to the best of our knowledge, this paper is the first to propose the automatically clustering based connection subregion modeling, and subregion associating based stable feature mining and tracking framework. And furthermore, the proposed ICIA algorithm has acquired a substantially high performance.

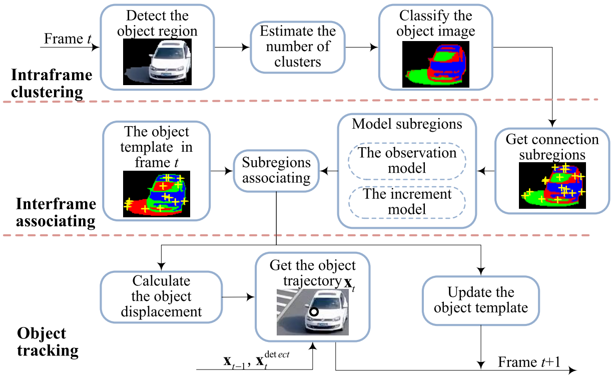

Frame *t*
Detect the object region
Estimate the number of clusters
Classify the object image
Intraframe clustering
The object template in frame *t*
Subregions associating
Model subregions
The observation model
The increment model
Get connection subregions
Interframe associating
Calculate the object displacement
Get the object trajectory $\mathbf{x}_t$
Update the object template
Object tracking
$\mathbf{x}_{t-1}$, $\mathbf{x}_t^{\det ect}$
Frame *t*+1

FIGURE 1. The flow chart of the proposed algorithm.

III. THE NUMBER OF CLUSTERS ESTIMATING AND INTRAFRAME CLUSTERING

A. ESTIMATING THE NUMBER OF CLUSTERS

Due to S and V components are independent of color information and much easier for processing than that in RGB color space, we use them to cluster the object appearance. To promote clustering adaptivity and efficiency, and restrain the background disturbance, we estimate the number of clusters according to the color feature of moving pixels and only cluster the motion region of the object. Here, we mainly deal with the object tracking under the stationary camera and detect the object region with the background difference algorithm [52]. The effectiveness of background suppression in object tracking has been illustrated in [16] and [26]. Beyan and Temizel [16] modeled the object with motion mask based kernel color histogram and located the individual object with MS in the case of multi-object merging, which effectively depressed the track drift. Lee and Kang [26] proposed a background feature elimination algorithm using the level set based bimodal segmentation to increase the robustness of object tracking. The background difference, with an adaptive threshold computed with iteration algorithm [53], here is adopted to obtain the binary motion region $\mathbf{R}_t(\mathbf{x})$. It is very important in dealing with the noise disturbance induced by illumination etc. And then the object image $\mathbf{O}_t(\mathbf{x})$ is obtained via Eq. (1) and shown in Fig. 2. $\mathbf{I}_t(\mathbf{x})$ denotes the current image, and $\mathbf{R}_t(\mathbf{x}) = 1$ and 0 represent the motion region (foreground) and the non-motion region (background) respectively. t is the frame number and $\mathbf{x}$ describes the pixel coordinate.

$$\mathbf{O}_t(\mathbf{x}) = \begin{cases} \mathbf{I}_t(\mathbf{x}) & if\ \mathbf{R}_t(\mathbf{x}) = 1 \\ 0 & otherwise \end{cases} \tag{1}$$

Fig. 2(a) shows the S and V component histograms of $\mathbf{O}_t(\mathbf{x})$. Obviously the S component only reflects one main peak while the V component embodies the distinction of different clusters with multiple main peaks which may represent car outside, windshield and shadow etc. Therefore, we employ the V component histogram to estimate the class number K_t in the current frame as follows.

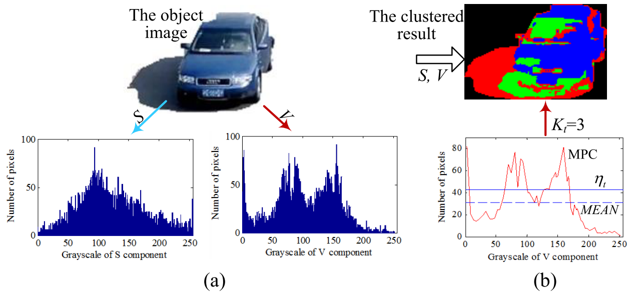


FIGURE 2. The flow chart of the proposed algorithm. (a) the S and V component histograms of the object image; (b) estimating the number of clusters with the MPC of the V component histogram.

Step 1: Extract the peak contour from the V component histogram and smooth obtain the main peak contour (MPC) (red solid line in Fig. 2(b)).

Step 2: Judge MPVs as candidate ones when MPVs are greater than or equal to an adaptive threshold η_t (blue solid line in Fig. 2(b)). η_t is calculated via (2) , where $\beta \in [1.3, 1.4]$ denotes a scale factor and *MEAN* (blue dash line) represents the average value of MPVs

$$\eta_t = \beta \cdot MEAN \tag{2}$$

Step 3: Sum the MPVs being lower than *MEAN* to obtain the residual energy, and then compute the percentage of the residual energy from the total number of MPVs to get the residual ratio RES_t.

Step 4: Calculate the absolute difference of adjacent V component grayscales *ADV* (corresponding to adjacent candidate peaks), and judge candidate peaks as region peaks when $ADV > \alpha_1$, otherwise retain the maximum peak to compare with the next candidate peak and repeat *Step* 4.

Step 5: Obtain the estimation value of the cluster number K_t by accumulating the number of region peaks. When $RES_t > \alpha_2$, add a cluster, i.e. $K_t \Leftarrow K_t + 1$.

$\alpha_1 \in [10, 15]$ and $\alpha_2 \in [0.20, 0.45]$ are the constant thresholds. In Fig. 2, we set $\beta = 1.38$, $\alpha_1 = 15$ and $\alpha_2 = 0.45$, and derive $K_t = 3$ according to above steps. Then the object image is classified with K-Means algorithm, and more details will be further illustrated in *Section B*. The clustered result is shown in Fig. 2(b), where the white background is replaced with black so as to highlight the foreground.

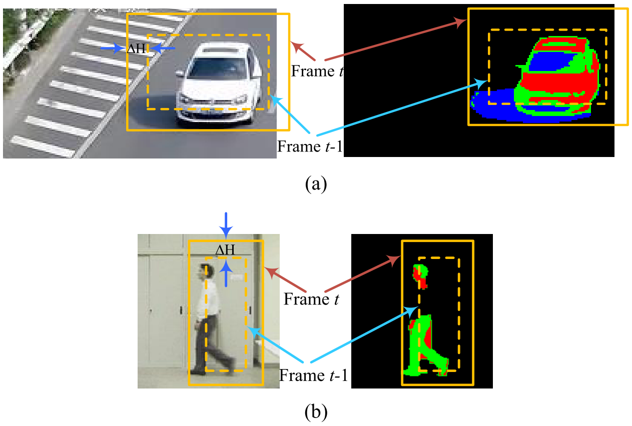


FIGURE 3. Classifying the object observation and obtaining connection subregions. (a) the object with scale change; (b) the object with incomplete detection.

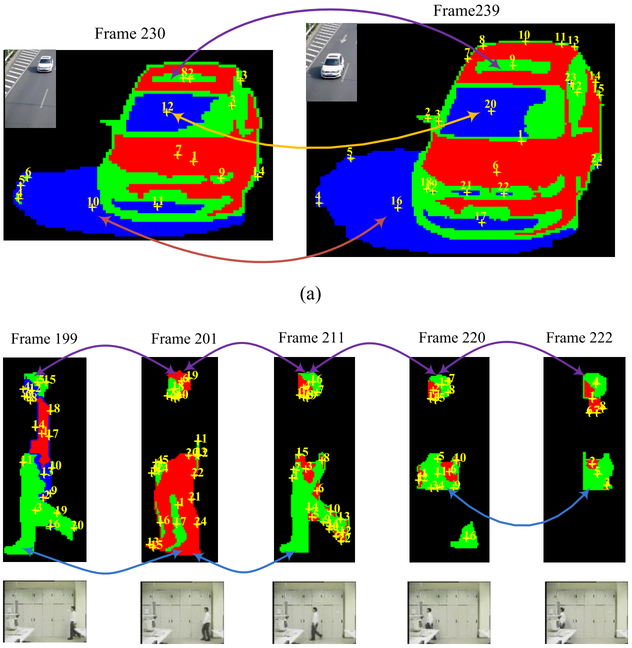
B. CLASSIFYING THE OBJECT OBSERVATION AND OBTAINING CONNECTION SUBREGIONS

When the object appearance is akin to the background or partially occluded etc, the detected object image may be incomplete or split into serval segments. Hereby, the object observation may include more than one motion region, and $\mathbf{O}_t(\mathbf{x})$ no longer can denote an object image. How to cluster the fractured informations and obtain the object observation under this case is vital. Furthermore, for the object being with scale change, how to determine its corresponding observation in next frame also is an important issue. To address these problems, we expand the tracked object region (yellow dash line) in frame$t$-1 with the scale increment ΔH to get the candidate object region (yellow solid line) in frame t (Fig. 3). ΔH is determined empirically to ensure that the candidate object region can cover the target object. To classify $\mathbf{O}_t(\mathbf{x})$ in the candidate object region, the S and V components $\mathbf{S}_t(\mathbf{x})$ and $\mathbf{V}_t(\mathbf{x})$ of moving pixel $\mathbf{x}$ firstly reshaped line by line into a $M \times 2$ sample intensity matrix $\mathbf{\Gamma}_t(n)$. M denotes the total number and $n \in [1, M]$. Let $\mathbf{C}_t(1) \cdots \mathbf{C}_t(K_t)$ be the cluster centroid intensities and initialized randomly, $\mathbf{\Gamma}_t(n)$ is clustered via (3) and a $M \times 1$ clustered matrix $\mathbf{D}_t(n)$ is gained. When a new $\mathbf{\Gamma}_t(n)$ is classed into the cluster k, the corresponding new centroid $\mathbf{C}_t(k)$ is updated via (4), where n iterates over all intensities and $k \in [1, K_t]$ iterates over all centroids. $\|\cdot\|_2$ is to compute the Euclidean distance and δ denotes Kronecker delta function.

$$\mathbf{D}_t(n) = k \left| \left\{ k := \arg\min_k \|\mathbf{\Gamma}_t(n) - \mathbf{C}_t(k)\|_2 \right\} \right. \tag{3}$$

$$\mathbf{C}_t(k) = \frac{\sum_{n=1}^{M} \mathbf{\Gamma}_t(n) \cdot \delta\left(\mathbf{D}_t(n) - k\right)}{\sum_{n=1}^{M} \delta\left(\mathbf{D}_t(n) - k\right)} \tag{4}$$

Then, $\mathbf{D}_t(n)$ is mapped to the candidate object region according to the one-to-one relationship between n and $\mathbf{x}$, and every cluster is represented with a pseudo-color as shown in Fig. 3. The numbers of clusters corresponding to the car (Fig. 3(a)) and the person (Fig. 3(b)) are 3 and 2 respectively. Every cluster is described with a unique color and includes multiple 8-connection subregions to reveal the spatial feature distribution of the object appearance.



(a)

(b)

FIGURE 4. Connection subregions comparison and analysis. (a) the *car* 1 sequence with scale and resolution changes. (b) the *man* 1 sequence with occlusion and similar color disturbance in the background.

IV. SUBREGION MODELING AND STABLE FEATURE MINING

A. SUBREGIONS COMPARISON AND ASSOCIATION RULES ANALYSIS

During tracking the object, the appearance feature often gradually changes along with illumination, object scale and resolution variations, background disturbance and occlusion etc. These variations affect the clustered results and are mainly reflected in the subregion center translation, area scaling and original color shift etc as shown in Fig. 4. The 8-connection subregions belonging to each cluster are located, and every subregion center position is marked with the '+' (plus) character. Several representative frames in *car* 1 and *man* 1 sequences are given in Fig. 4(a) and (b) respectively. The number of clusters is ranked in terms of classes. In frame #230 of Fig. 4(a), the class marked with red corresponds to subregions 1∼3, while the class marked with green and blue correspond to subregions 4∼9 and subregions 10∼14 respectively. Similarly, the subregions in frame #239 and the *man* 1 sequence in Fig. 4(b) are numbered. The total number of the subregions grows along with the increase of the object resolution (Fig. 4(a)), and reduces under occlusion and incomplete detection (Fig. 4(b)).

It is obvious that some subregions appear continuously in sequential frames, and they have one-to-one correspondence. In Fig. 4(a), the subregions 8 (skylight glass labeled with green), 10 (shadow labeled with blue) and 12 (windshield labeled with blue) in frame #230 correspond to subregions 9, 16 and 20 in frame #239 respectively. Similarly, in Fig. 4(b), subregion 3 (trousers labeled with green) in frame #199, subregion 1 (red) in frame #201 and subregion 14 (green) in frame #211 also have the association. Even though the partial body (shirt) of the object is not detected in frames #201 and #211, other subregions (such as trousers) can keep the association. The object observation is further damaged by occlusion in frames #220 and #222 (trousers are partially detected), but the hair maintains the correspondence. Our purpose in this paper is to mine the stable object features (such as the shadow and the trousers in above examples) and use them to improve the tracking performance. The stable features contain two level restrictions. One is that the subregion feature should exist both in the current frame and the real-time template simultaneously. The other is that the subregion feature variation in current frame should satisfy the coherence and continuity rule, i.e. stable subregions having similar displacement (coherence), and scale change and color shift being lower than certain thresholds (continuity). If the subregion feature conforms to both restrictions, they are stable and reliable in deriving a robust and accurate object trajectory. Thus, we need to build the continuity and coherence rule for mining stable subregions, and model the object with local subregion features. Compared with global features such as the color histogram [14], [16] and the uniform block-division based feature [27], our feature model appears sparse, local or exist momentarily but more robust. Since clustering segments the object region non-uniformly, it can restrain background disturbance in the marginal region of the object where some background features may be introduced under uniform segmentation.

B. SUBREGION MODELING AND STABLE FEATURE MINING

Let $\mathbf{E}_t^{r1} = \left\{E_{\mathbf{x},t}^{r1}, E_{A,t}^{r1}, E_{S,t}^{r1}, E_{V,t}^{r1}\right\}$ be the $r1$-th subregion template, $\Delta\mathbf{E}_t^{r1} = \left\{\Delta E_{\mathbf{x},t}^{r1}, \Delta E_{A,t}^{r1}, \Delta E_{A\_ratio,t}^{r1}, \Delta E_{SV\_ratio,t}^{r1}\right\}$ describe the $r1$-th subregion increment model, and $\mathbf{F}_t^{s1} = \left\{F_{\mathbf{x},t}^{s1}, F_{A,t}^{s1}, F_{S,t}^{s1}, F_{V,t}^{s1}\right\}$ denote the $s1$-th subregion observation model.

$\hat{\mathbf{F}}_t^{m} = \left\{\hat{F}_{\mathbf{x},t}^{m}, \hat{F}_{A,t}^{m}, \hat{F}_{S,t}^{m}, \hat{F}_{V,t}^{m}\right\}$ and $\hat{\mathbf{E}}_t^{m} = \left\{\hat{E}_{\mathbf{x},t}^{m}, \hat{E}_{A,t}^{m}, \hat{E}_{S,t}^{m}, \hat{E}_{V,t}^{m}\right\}$ stand for the m-th stable subregion

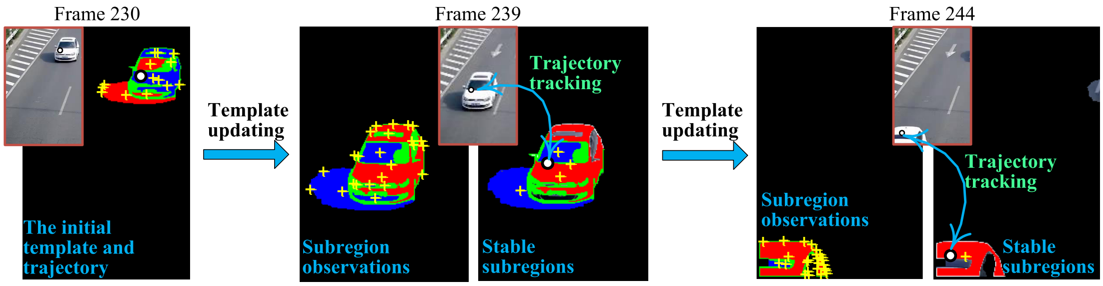


FIGURE 5. Stable subregions mining and trajectory tracking.

pair $\left(\hat{\mathbf{F}}_t^m, \hat{\mathbf{E}}_t^m\right)$ mined from $\mathbf{F}_t^{s1}$ and $\mathbf{E}_t^{r1}$ respectively. The subscripts $\mathbf{x}, A, S, V, A\_ratio$ and SV_ratio represent the subregion center coordinate, area, saturation, value, area and color change ratios respectively. $r1 \in [1, N1]$, $s1 \in [1, N2]$ and $m \in [1, N3]$ are the subregion numbers of $\mathbf{E}_t^{r1}$, $\mathbf{F}_t^{s1}$ and $\hat{\mathbf{F}}_t^m$ respectively. $E_{V,t}^{r1}$ is computed via Eq. (5), and $E_{S,t}^{r1}$, $F_{S,t}^{s1}$ and $F_{V,t}^{s1}$ are obtained in the same way.

$$E_{V,t}^{r1} = \frac{\sum_{n=1}^{M} \widehat{\mathbf{V}}_t^{r1}(n) \cdot \delta\left(\mathbf{D}_t(n) - k\right)}{\sum_{n=1}^{M} \delta\left(\mathbf{D}_t(n) - k\right)} \tag{5}$$

To automatically find the associated subregions in the image sequence, the template difference $\mathbf{\Delta}_t^{r1}$ between $\mathbf{E}_t^{r1}$ and $\mathbf{F}_t^{s1}$ is firstly computed using Eq. (6). The center distance $d_{\mathbf{x},t}^{r1}$, the distance increment $\Delta E_{\mathbf{x},t}^{r1}$, the area increment $\Delta E_{A,t}^{r1}$, the area change ratio $\Delta E_{A\_ratio,t}^{r1}$ and the color change ratio $\Delta E_{SV\_ratio,t}^{r1}$ are further calculated via Eqs. (7)∼(11) respectively.

$$\begin{aligned}\mathbf{\Delta}_t^{r1} &= \mathbf{E}_t^{r1} - \mathbf{F}_t^{s1} = \left\{\Delta_{\mathbf{x},t}^{r1}, \Delta_{A,t}^{r1}, \Delta_{S,t}^{r1}, \Delta_{V,t}^{r1}\right\} \\ &= \left\{E_{\mathbf{x},t}^{r1} - F_{\mathbf{x},t}^{s1}, E_{A,t}^{r1} - F_{A,t}^{s1}, E_{S,t}^{r1} - F_{S,t}^{s1}, E_{V,t}^{r1} - F_{V,t}^{s1}\right\}\end{aligned} \tag{6}$$

$$d_{\mathbf{x},t}^{r1} = \left(\Delta_{\mathbf{x},t}^{r1} \cdot \left(\Delta_{\mathbf{x},t}^{r1}\right)^T\right)^{\frac{1}{2}} \tag{7}$$

$$\Delta E_{\mathbf{x},t}^{r1} = \Delta_{\mathbf{x},t}^{r1} \tag{8}$$

$$\Delta E_{A,t}^{r1} = \Delta_{A,t}^{r1} \tag{9}$$

$$\Delta E_{A\_ratio,t}^{r1} = \frac{\Delta_{A,t}^{r1}}{E_{A,t}^{r1}} \tag{10}$$

$$\begin{aligned}&\Delta E_{SV\_ratio,t}^{r1} \\ &\quad = \left(\frac{\left(\Delta_{S,t}^{r1}, \Delta_{V,t}^{r1}\right) \cdot \left(\Delta_{S,t}^{r1}, \Delta_{V,t}^{r1}\right)^T}{\left(E_{S,t}^{r1}, E_{V,t}^{r1}\right) \cdot \left(E_{S,t}^{r1}, E_{V,t}^{r1}\right)^T}\right)^{\frac{1}{2}}\end{aligned} \tag{11}$$

In what follows, we use Eq. (12) to associate $\mathbf{E}_t^{r1}$ and $\mathbf{F}_t^{s1}$, mine $\left(\hat{\mathbf{F}}_t^m, \hat{\mathbf{E}}_t^m\right)$ and acquire the stable subregion increment $\Delta\hat{\mathbf{E}}_t^m$, where $d_{\mathbf{x},t}^{r1} \leq \mu_t \cap \left|\Delta_{A\_ratio,t}^{r1}\right| \leq \lambda_1 \cap \Delta_{SV\_ratio,t}^{r1} \leq \lambda_2$ is multi-feature distances based the coherence and continuity rule.

When this rule is satisfied, $\left(\mathbf{E}_t^{r1}, \mathbf{F}_t^{s1}\right)$ is judged to be stable and assigned to $\left(\hat{\mathbf{F}}_t^m, \hat{\mathbf{E}}_t^m\right)$ via Eq. (12). The stable subregion increment $\Delta\hat{\mathbf{E}}_t^m = \left\{\Delta\hat{E}_{\mathbf{x},t}^m, \Delta\hat{E}_{A,t}^m, \Delta\hat{E}_{A\_ratio,t}^m, \Delta\hat{E}_{SV\_ratio,t}^m\right\}$ is accordingly obtained. If $\mathbf{E}_t^{r1}$ is stable, we set the stable sign $\phi_t^{r1} = 1$, or else $\phi_t^{r1} = 0$, which will be further used in template updating. $\mu_t = \max\left\{\left(E_{A,t}^{r1}\right)^{\frac{1}{2}}, \left(F_{A,t}^{s1}\right)^{\frac{1}{2}}\right\}$ is an adaptive threshold, $\lambda_1 \in [0.3, 0.5]$ and $\lambda_2 \in [0.1, 0.3]$ are constant thresholds. The stable subregion difference $\hat{\mathbf{\Delta}}_t^m$ between $\hat{\mathbf{E}}_t^m$ and $\hat{\mathbf{F}}_t^m$ is acquired via Eq. (13).

$$\begin{aligned}&\left\{\hat{\mathbf{E}}_t^m, \hat{\mathbf{F}}_t^m, \Delta\hat{\mathbf{E}}_t^m\right\} \\ &\quad = \left\{\mathbf{E}_t^{r1}, \mathbf{F}_t^{s1}, \Delta\mathbf{E}_t^{r1}\right\} \Big| \left\{\mathbf{E}_t^{r1}, \mathbf{F}_t^{s1}, \Delta\mathbf{E}_t^{r1}\right\} \\ &\quad := \arg\left(d_{\mathbf{x},t}^{r1} \leq \mu_t \cap \left|\Delta_{A\_ratio,t}^{r1}\right| \leq \lambda_1 \cap \Delta_{SV\_ratio,t}^{r1} \leq \lambda_2\right)\end{aligned} \tag{12}$$

$$\hat{\mathbf{\Delta}}_t^m = \hat{\mathbf{E}}_t^m - \hat{\mathbf{F}}_t^m = \left\{\hat{\Delta}_{\mathbf{x},t}^m, \hat{\Delta}_{A,t}^m, \hat{\Delta}_{S,t}^m, \hat{\Delta}_{V,t}^m\right\} \tag{13}$$

V. OBJECT TRACKING AND TEMPLATE UPDATING

Here, we update the current object template $\mathbf{E}_t^{r1}$ with stable subregions mined in frame t-1, and employ the center displacement $\hat{\Delta}_{\mathbf{x},t}^m$ between $\hat{\mathbf{E}}_t^m$ and $\hat{\mathbf{F}}_t^m$ to derive the object displacement between neighbor frames. To obtain the object trajectory $\mathbf{x}_t$, the average template increment $\bar{\mathbf{\Delta}}_t$ is firstly computed by systematically fusing the stable subregion increment $\hat{\mathbf{\Delta}}_t^m$ via Eq. (14). The larger the subregion area is, the more reliable the stable increment $\hat{\mathbf{\Delta}}_t^m$ is. On this basis, we adopt $\hat{E}_{A,t}^m$ to weight $\hat{\mathbf{\Delta}}_t^m$ for increasing the voting power of the subregion being with larger area and vice versa. Then $\mathbf{x}_t$ is acquired via Eq. (15) in terms of the average center displacement $\bar{\Delta}_{\mathbf{x},t}$, the historical trajectory $\mathbf{x}_{t-1}$ and the detected center trajectory $\mathbf{x}_t^{\det ect}$ in object region. If none of the subregions are stable under some cases such as heavy or total occlusion, $\bar{\mathbf{\Delta}}_t$ is estimated with $\bar{\mathbf{\Delta}}_{t-1}$. $\gamma_1 \in [0.9, 1]$ is to endow $\bar{\Delta}_{\mathbf{x},t}$ higher weight in determining $\mathbf{x}_t$. $\mathbf{x}_t^{\det ect}$ is use to restrain the tracking drift introduced by accumulated

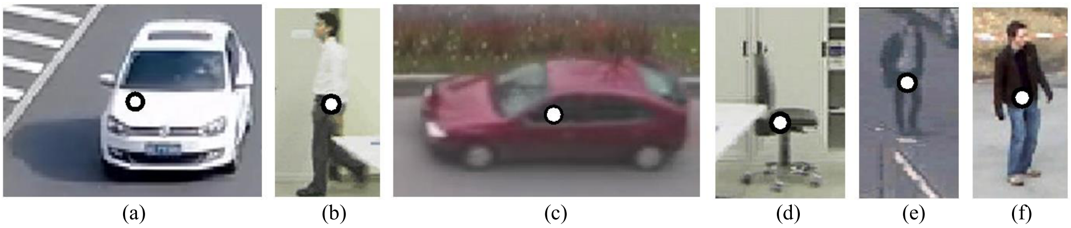


FIGURE 6. The initial object trajectory. (a) *car* 1; (b) *man* 1; (c) *car* 2; (d) *chair*; (e) *man* 2; (f) *man* 3.

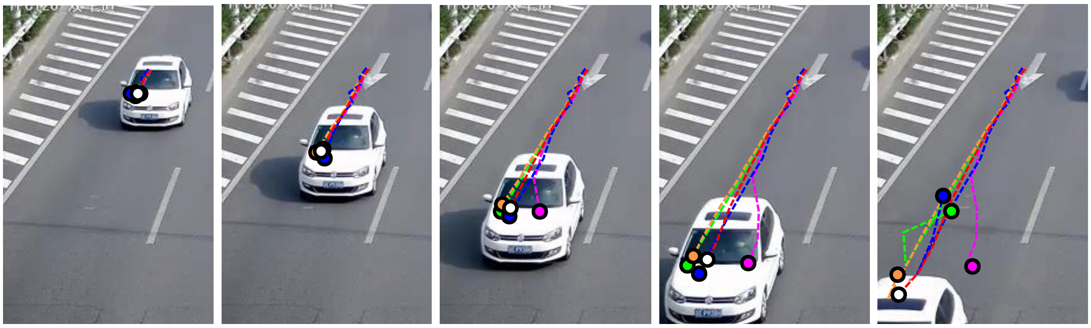


FIGURE 7. The tracked results with ICIA, STC SCC, MS and KCF algorithms in the *car* 1 sequence.

error etc.

$$\bar{\mathbf{\Delta}}_t = \left\{\bar{\Delta}_{\mathbf{x},t}, \bar{\Delta}_{A,t}, \bar{\Delta}_{S,t}, \bar{\Delta}_{V,t}\right\} = \frac{\sum_{m=1}^{N3} \hat{E}_{A,t}^m \cdot \hat{\mathbf{\Delta}}_t^m}{\sum_{m=1}^{N3} \hat{E}_{A,t}^m} \quad (14)$$

$$\mathbf{x}_t = \gamma_1 \cdot \left(\mathbf{x}_{t-1} - \bar{\Delta}_{\mathbf{x},t}\right) + (1-\gamma_1) \cdot \mathbf{x}_t^{\det ect} \quad (15)$$

Let $\bar{\Delta}_{A\_ratio,t}$ be the average area change ratio of stable subregions, *SIGN* denote the plus-minus sign of $\bar{\Delta}_{A\_ratio,t}$, and $\Phi_{\mathbf{x},t}$ and $\Phi_{A,t}$ describe the average displacement and area increments of the stable subregions as in Eqs. (16)~(19). $\bar{\Delta}_{A\_ratio,t}$ is calculated with the weighted average of $\hat{\Delta E}_{A\_ratio,t}^m$ via Eq. (16).

$$\bar{\Delta}_{A\_ratio,t} = \frac{\sum_{m=1}^{N3} \hat{E}_{A,t}^m \cdot \hat{\Delta E}_{A\_ratio,t}^m}{\sum_{m=1}^{N3} \hat{E}_{A,t}^m} \quad (16)$$

$$SIGN = \begin{cases} 1 & if\ \bar{\Delta}_{A\_ratio,t} > 0 \\ -1 & if\ \bar{\Delta}_{A\_ratio,t} < 0 \\ 0 & otherwise \end{cases} \quad (17)$$

$$\Phi_{\mathbf{x},t} = \bar{\Delta}_{\mathbf{x},t} \cdot \left(1 - SIGN \cdot \left|\bar{\Delta}_{A\_ratio,t}\right|^{\frac{1}{2}}\right) \quad (18)$$

$$\Phi_{A,t} = \bar{\Delta}_{A,t} \cdot \left(1 - \bar{\Delta}_{A\_ratio,t}\right) \quad (19)$$

Since stable subregion observations in the current frame can support a robust update of stable features, we separately update stable and unstable features. Specifically, we update $E_{\mathbf{x},t}^{r1}$ and $E_{A,t}^{r1}$ in terms of the stable sign ϕ_t^{r1} via Eqs. (20) and (21). As for the stable subregion, $E_{\mathbf{x},t}^{r1}$ is updated by weighting the subregion increment $\Delta E_{\mathbf{x},t}^{r1}$ and the average increment $\Phi_{\mathbf{x},t}$. We select $\gamma_2 \in [0.8, 1]$ to save more stable suregion observation informations and use $\Phi_{\mathbf{x},t}$ to relieve the excessive update. To the unstable subregion without directly observation for self-renewaling, $\Phi_{\mathbf{x},t}$ is adopted to estimate $E_{\mathbf{x},t+1}^{r1}$. $E_{A,t}^{r1}$ is updated in a similar way as in Eq. (21). To the S and V translational variations caused by gradual illumination, the average color translations $\bar{\Delta}_{S,t}$ and $\bar{\Delta}_{V,t}$ here are employed to update $E_{S,t}^{r1}$ and $E_{V,t}^{r1}$ (include stable and unstable subregions) via Eqs. (22)~(23). If none of the subregions are stable, $\mathbf{E}_{t+1}^{r1} = \mathbf{E}_t^{r1}$.

$$E_{\mathbf{x},t+1}^{r1} = E_{\mathbf{x},t}^{r1} - \phi_t^{r1} \cdot \left(\gamma_2 \cdot \Delta E_{\mathbf{x},t}^{r1} + (1-\gamma_2) \cdot \Phi_{\mathbf{x},t}\right) - \left(1 - \phi_t^{r1}\right) \cdot \Phi_{\mathbf{x},t} \quad (20)$$

$$E_{A,t+1}^{r1} = E_{A,t}^{r1} - \phi_t^{r1} \cdot \left(\gamma_2 \cdot \Delta E_{A,t}^{r1} + (1-\gamma_2) \cdot \Phi_{A,t}\right) - \left(1 - \phi_t^{r1}\right) \cdot \Phi_{A,t} \quad (21)$$

$$E_{S,t+1}^{r1} = E_{S,t}^{r1} - \bar{\Delta}_{S,t} \quad (22)$$

$$E_{V,t+1}^{r1} = E_{V,t}^{r1} - \bar{\Delta}_{V,t} \quad (23)$$

Fig. 5 presents the process and qualitative performance of the stable feature mining and object tracking scheme in the *car* 1 sequence. This sequence includes resolution and scale changes (frames #230 and #239), and accelerant movement along with fast appearance reduction due to the object being about to leave from the field of view (frame #244). Fig. 5 depicts the initial template (frame #230), clustered subregion observations, mined stable subregions and the tracked object trajectories over time in frames #239 and #244. The mined stable subregions are marked with the same colors and '+'

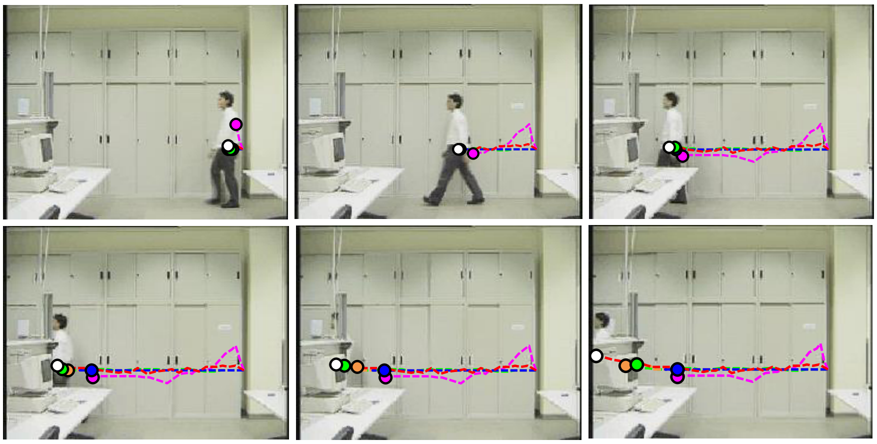


FIGURE 8. The tracked results with ICIA, STC, SCC, MS and KCF algorithms in the *man* 1 sequence.

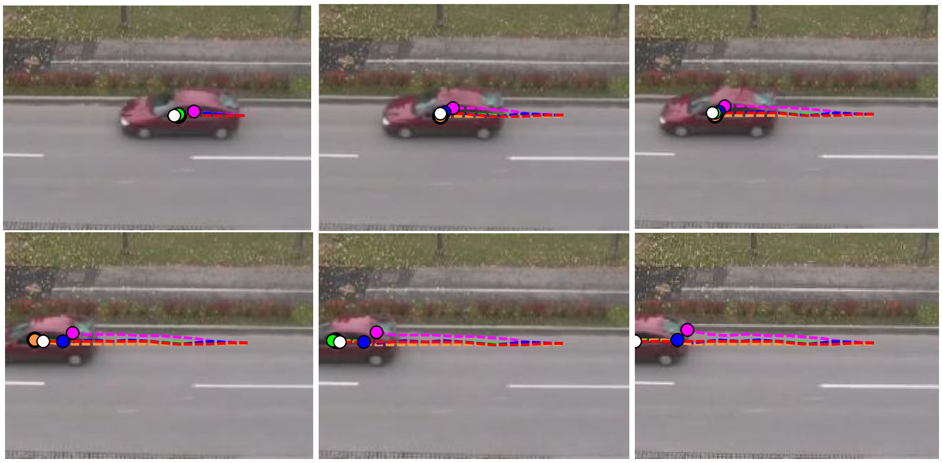


FIGURE 9. The tracked results with ICIA, STC, SCC, MS and KCF algorithms in the *car* 2 sequence.

characters as that done in the corresponding observations. The stable subregion centers are further fused to locate the object trajectory (the white round dot). Our method successfully tracks the object in terms of the initial trajectory in frame #230. In frame #239, stable subregions, such as shadow (blue), windshield (blue), hood (red) and carframe (green), almost cover the total object region and indeed provide reliable and powerful supports for object locating. In frame #244, the partial object body is detected and only one stable subregion (partial car roof) is mined. Since we use $\mathbf{x}_t^{\det ect}$ to correct the object trajectory, our method can stably locate the partial object even though most of the car body is out of the field of view.

VI. EXPERIMENTAL RESULTS

The proposed algorithm is implemented on a PC with a 2.1GHz AMD A8-5550M CPU, Windows 7 operation system and MatLab implementation. To evaluate our method, we conduct comparison experiments among the presented ICIA based tracking method, and one classical and three state-of-the-art methods, including MS [14], STC [17], SCC [29] and KCF [24] trackers. All these algorithms are tested on five challenging sequences with partial or heavy occlusions, similar color disturbance in the background, blurry and ambiguous appearance, scale and resolution changes, fast movement and rotation etc. And these tested sequences cover many tracking scenarios including both indoor and outdoor, and various object types such as rigid (vehicle and chair) and nonrigid (person) objects, and large and small objects. Some examples are presented in Figs. 6∼12, where the round marks illustrate the tracked object trajectories in every frame and the dash lines denote the historical trajectories.

A. COMPARISON ON QUALITATIVE TRACKING PERFORMANCE

The round marks in Fig. 6 are the initial center trajectories of the objects corresponding to *car* 1, *man* 1, *car* 2, *chair*, *man* 2 and *man 3* sequences respectively, which are automatically determined by motion detection in the first frame and

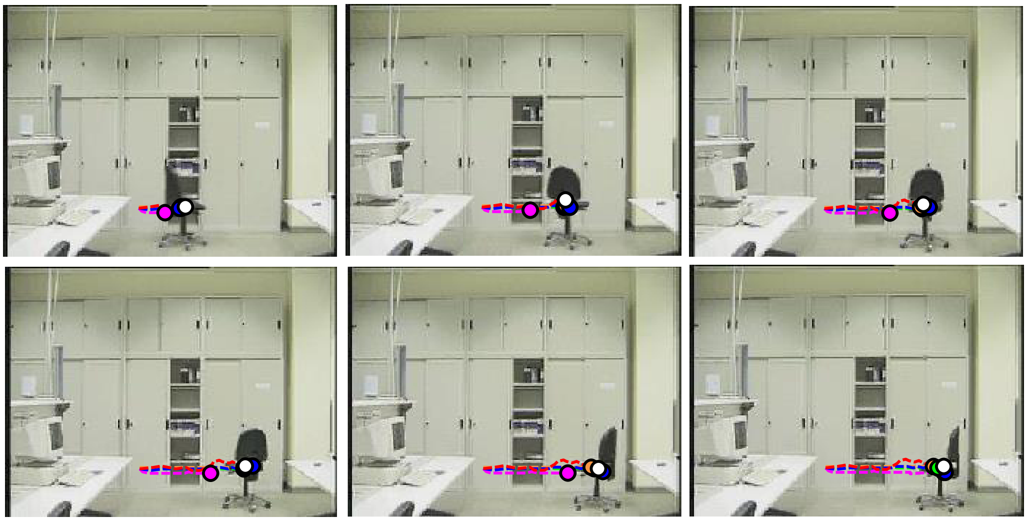


FIGURE 10. The tracked results with ICIA, STC, SCC, MS and KCF algorithms in the *chair* sequence.

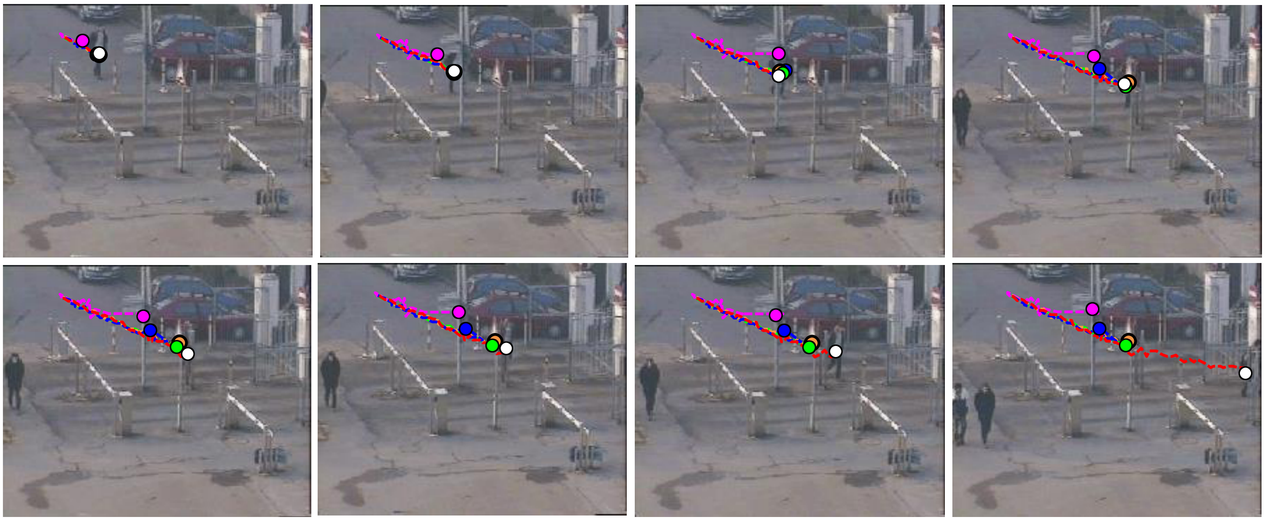


FIGURE 11. The tracked results with ICIA, STC, SCC, MS and KCF algorithms in the *man* 2 sequence.

coincide with the center position of the detected object region. Figs. 7∼12 present the comparisons on qualitative tracking performances of six sequences. The white, green, blue, purplish red and orange round marks denote the currently tracked object trajectories of ICIA, STC, SCC, MS and KCF trackers respectively. Meanwhile, the red, green, blue, purplish red and orange dash lines denote the corresponding historical trajectories. All the trackers are automatically initialized with same initial trajectories.

In Fig. 7, the *car* 1 sequence tested includes accelerant movement, and scale and resolution changes to the object. The outputs of trackers in frames #233, #237, #240, #242 and #244 (from left to right) are illustrated. After frame #240, significant appearance details appear along with the car accelerating towards the camera. Meanwhile, the partial car appearance rapidly loses due to its quickly moving away from the field of view. Under this sequence, ICIA outperforms other trackers and robustly tracks through the appearance and scale variations. We can see from the tracked results that KCF, STC and SCC algorithms can successfully track the object when most car appearance keeps in the field of view. But in frame #244, they show obvious trajectory drift. Nevertheless, the kernel based color histogram model cannot adapt to the dramatic changes in the appearance resolution and scale of the object, which makes the MS tracker converge to the local extremum and leads to failures (frames #240, #242 and #244).

The tested *man* 1 sequence is displayed in Fig. 8. The main challenges are heavy occlusion and similar color disturbance in the background (bookcase and desk) that occur as the man moving. Rows 1 and 2 give the tracked results in frames #201, #209, #219, #222, #224 and #228. Under SCC and

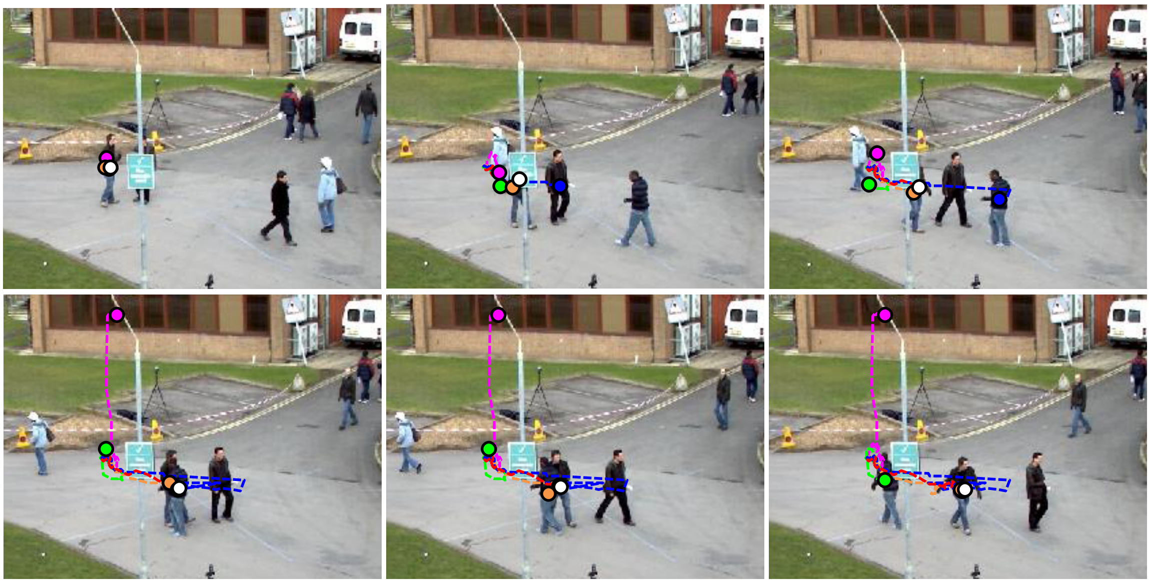


FIGURE 12. The tracked results with ICIA, STC, SCC, MS and KCF algorithms in the *man* 3 sequence.

MS methods, trajectories are hijacked by coexisting partial occlusion and background disturbance as illustrated in frame #222. STC and KCF methods can handle this case due to their robust confidence model and correlation filter model respectively. However, when the object is heavily occluded (frame #224), the STC and KCF trackers cannot find the credible object position and stays around the place where the object loses. In comparison, the proposed ICIA method can handle these difficulties as it replaces the global model using a local stable subregion model of the object, and effectively estimates the average center displacement under heavy or total occlusion (frame #224). Our ICIA tracker can stably follow the object until frame #228. We can see from the tracked results that a good strategy for feature modeling is needed for the tracker.

Fig. 9 provides the tested *car* 2 sequence which includes fast movement, background disturbance and low resolution appearance to the object. Rows 1 and 2 present the tracking outputs in frames #123, #125, #127, #129, #130 and #131. In spite of the object being of low resolution appearance and influenced by the shadow and reflection of the plant in the scene, the proposed ICIA tracker, the KCF tracker and the STC tracker can accurately locate the object before frame #130. In frame #131, KCF and STC methods give the trajectory outside the field of view where the object quickly moves and the partial object body goes out of the image bound. Our ICIA tracker handles this specific type of scenario easily and still keeps robustly tracking the partial car body with the help of the detected trajectory. However, under this case, SCC and MS trackers have apparent lags in frames #127, #129 and #131 due to no structure or histogram feature stably exists for more than two or three frames, and this raises the chance of tracking failure.

The tracked results of the tested *chair* sequence, which includes rapid posture change, fast movement and rotation, and similar color disturbance in the background, are illustrated in Fig. 10. Rows 1 and 2 describe the results of trackers in frames #823, #829, #832, #834, #838 and #841. Even though these kinds of objects are deformable, the stable subregion feature, the correlation filter, the confidence map and the structure models are enough robust. The tracked trajectories illustrate that ICIA, KCF, STC and SCC successfully accomplish the tracking task as the object quickly moves from left to right and rotates to show different postures. Since the MS tracker focuses on matching the color histogram, the iteration process is distracted by the opened bookcase being of the similar feature to the object (frame #823), and this makes the MS tracker gradually loses the object (frame #829). However, the MS tracker keeps searching in terms of pre-set iteration times and just finds the object again where the object comes back under the effect of the rotatory inertia (frame #841).

The tested results of the *man* 2 sequence are described in Fig. 11. The main challenges are clutter background, low resolution and ambiguous appearance, similar color disturbance in the background, heavy occlusion and multiple times occlusions. Rows 1 and 2 show the outputs of trackers in frames #282, #298, #301, #312, #315, #316, #320 and #344. Under these cases, the tracker that relies heavily on color or structure is more vulnerable to drifting. As shown in frames #282, #298 and #301, the MS tracker fails to track the object. Although the SCC tracker can deal with the partial occlusion in frames #298 and #301, it fails to track the ambiguous object in such low resolution and clutter scene as in frame #312 (the partial observable object body being intermixed with the background). In frames #315 and #316, occlusions are

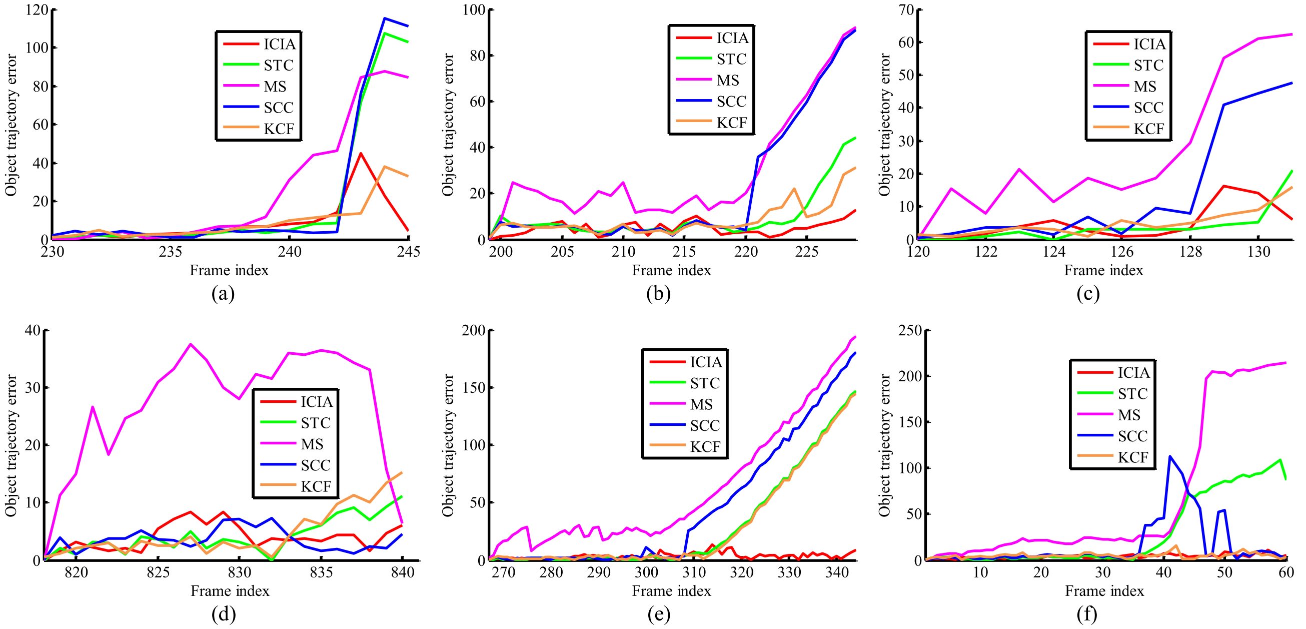


FIGURE 13. Trajectory error comparisons. (a) *car* 1; (b) *man* 1; (c) *car* 2; (d) *chair*; (e) *man* 2; (f) *man* 3.

simultaneously induced by the two obstructions being with the similar color to the object, which makes it difficult to identify the object. The STC and the KCF trackers mistake the static occluder for the optimal confidence observation, and keep locate this occluder (frame #315) in subsequent frames. However, the presented ICIA tracker succeeds under above difficult scenarios for the sake of its integration of stable subregions displacements, historical trajectory and motion detection towards handling the ambiguity in clutter background and occlusion.

In Fig. 12, the *man* 3 sequence (PETS 2009) tested includes partial and heavy occlusions along with similar color and similar object disturbances to the object in the multi-object scenario. The target man walks with varying body pose and pace. The outputs of trackers in frames #11, #40, #43, #51, #53 and #60 are illustrated (Rows 1 and 2). From frame #36 to #46, the upper body of the object is totally or partially occluded by a guidance sign. And from frame #51 to #55, the target object is heavily or partially occluded by another object with similar appearance feature. The ICIA and the KCF trackers outperform other trackers under these cases. The STC tracker can successfully track the object before frame #38. But After frame #38, it starts to drift away from the object as a result of occlusion. Meanwhile, the MS tracker shows evident trajectory drift from frame #40 due to the influence of the accumulated error induced by Kalman filter estimation, and stays far away from the target object. The SCC tracker mistakenly locates the other men with similar structure feature as in frames #40 and #43. However, it captures the location of the object again when the target man is out of occlusion after frame #51. Although both ICIA and KCF have small deviations in several frames as in #51 and #53, they achieve good performance throughout the video sequence.

B. COMPARISON ON QUANTITATIVE TRACKING PERFORMANCE

The trajectory error between the tracked result $\mathbf{x}_t = \{x_t, y_t\}$ and the ground truth $\hat{\mathbf{x}}_t = \left\{\hat{x}_t, \hat{y}_t\right\}$ is computed with Euclidean distance $\left\|\mathbf{x}_t - \hat{\mathbf{x}}_t\right\|_2$. The trajectory error versus frame graphs are shown in Fig. 13, where red, green, blue, purplish red and orange solid lines denote ICIA, STC, SCC, MS and KCF trackers respectively. For each testing sequence, the average trajectory error $\bar{x}_{error}$ is calculated via Eq. (24). N is the total number of image frames and $\hat{\mathbf{x}}_t$ is acquired by manual method.

$$\bar{\mathbf{x}}_{error} = \frac{1}{N}\sum_{t=1}^{N}\left\|\mathbf{x}_t - \hat{\mathbf{x}}_t\right\|_2 \tag{24}$$

Table 1 summarizes the tracking results of six different test sequences. One can see that the proposed ICIA tracker most accurately locates the objects in *car* 1, *man* 1, *man* 2 and *man* 3 sequences. In these sequences, larger trajectory drift occurs to other four trackers due to accelerant movement along with scale change, rapid appearance disappear, and occlusion companied by similar color and similar object disturbances etc. As shown in Fig. 13, these cases appear around frame #243 in Fig. 13(a), frame #224 in Fig. 13(b), frame #310 in Fig. 13(e) and frame #37 in Fig. 13(f) respectively. The ICIA tracker also robustly tracks the objects in *car* 2 (Fig. 13(c)) and *chair* (Fig. 13(d)) sequences and shows the second-best performance with only less 1 pixel distance difference than the best method. The STC tracker shows the best tracking performance in the *car* 2 sequence. Furthermore,

TABLE 1. Quantitative comparisons: the average trajectory errors (Units: Pixels). (The red mark represents the best results, and green and blue marks denote the second-best and the third-best results respectively.)

Sequence	*car* 1	*man* 1	*car* 2	*chair*	*man* 2	*man* 3
ICIA	8.33	4.76	4.67	3.95	3.67	4.09
KCF	9.60	8.28	4.94	4.75	29.22	4.84
STC	20.38	9.42	3.83	4.17	29.88	29.02
SCC	21.63	21.29	14.14	3.50	45.03	67.12
MS	25.93	29.58	26.36	26.66	62.70	16.30

the STC tracker achieves the third-best performance in other sequences and is with almost equivalent accuracy to ICIA, KCF and SCC trackers in the *chair* sequence. In *car* 1, *man 2* and *man 3* sequences, although the STC tracker obtains the third-best performance, its average trajectory error is much higher than that of the ICIA tracker. The KCF tracker demonstrates the second-best tracking performance in the *car* 1, *man* 1, *man* 2 and *man* 3 sequences and the third-best performance in the *car* 2 sequence. It also has less trajectory error in the *chair* sequence. The SCC tracker has an optimal performance in tracking the object with rotation and posture changes as in the *chair* sequence. However, we can find from the trajectory error versus frame graph that the SCC tracker do have some difficulties in handling the object with rapid structure variation, and blurry appearance along with occlusion induced by clutter background etc. In contrast, the MS tracker is with higher trajectory error which can be mainly attributed to its sensitive to similar color disturbance, and fast changes in object scale and appearance etc. In comparison with KCF, STC, SCC and MS trackers, the ICIA tracker produces better tracking results. It is mainly because we construct a good appearance model using the local stable subregion features, which are robust to occlusion (heavy occlusion and many times occlusions), ambiguous object in clutter scene along with similar color or object disturbance, scale and posture changes accompanied with accelerant and fast movements etc. With the motion detecting process, the ICIA method can prevent the similar color influence in the background and rectify the cumulative error in time. In addition, the ICIA method successfully tracks the objects both in benchmark (*man* 1, *chair*, *car* 2, *man* 2 and *man* 3) and the video captured by us (*car* 1). Note that the ICIA tracker shows a small trajectory error, which reveals that the proposed method obtains considerably stable tracking results.

VII. CONCLUSIONS

In this paper, we have proposed to mine stable features to improve object representation, and enhance the robustness of object tracking. This study is motivated by the idea that sparse but reliable features can achieve a more accurate and robust tracking. We have presented an ICIA based stable feature mining framework for object tracking. This is composed of motion detecting, the number of clusters estimating, intraframe clustering, connection subregion modeling, stable subregion mining and displacements fusing, as well as model updating. Experimental evaluations (both qualitative and quantitative) on several challenging image sequences show that our approach performs more robustly and reliably than several relevant state-of-the-art algorithms. Although the results are promising in certain situations, further evaluation is anticipated in more complicated data sets. We have implemented all of the experiments without code optimization. Real-time tracking will be the further development. A future work inclines to integrate the stable feature mining based tracking framework into multi-object tracking to handle the dynamic occlusion and mutual disturbance among similar objects.

REFERENCES

[1] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 36, no. 7, pp. 1442–1468, Jul. 2014.

[2] Q. Zhao and H. Tao, "A motion observable representation using color correlogram and its applications to tracking," *Comput. Vis. Image Understand.*, vol. 113, no. 2, pp. 273–290, Feb. 2009.

[3] H. Lu, W. L. Zou, H. S. Li, Y. Zhang, and S. M. Fei, "Edge and color contexts based object representation and tracking," *Optik*, vol. 126, no. 1, pp. 148–152, Jan. 2015.

[4] T. X. Bai, Y. Y. Li, and X. L. Zhou, "Learning local appearances with sparse representation for robust and fast visual tracking," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 663–675, Apr. 2015.

[5] L. J. Cao, R. R. Ji, Y. Gao, W. Liu, and Q. Tian, "Mining spatiotemporal video patterns towards robust action retrieval," *Neurocomputing*, vol. 105, no. 3, pp. 61–69, Apr. 2013.

[6] L. Wang, Y. Z. Wang, T. T. Jiang, D. B. Zhao, and W. Gao, "Learning discriminative features for fast frame-based action recognition," *Pattern Recogn.*, vol. 46, no. 7, pp. 1832–1840, Jul. 2013.

[7] L. Liu, L. Shao, X. L. Li, and K. Lu, "Learning spatio-temporal representations for action recognition: A genetic programming approach," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 158–170, Jan. 2016.

[8] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.

[9] S. Salti, A. Cavallaro, and L. D. Stefano, "Adaptive appearance modeling for video tracking: Survey and evaluation," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4334–4348, Oct. 2012.

[10] J. Zhang, S. Ma, and S. Sclaroff, "MEEM: Robust tracking via multiple experts using entropy minimization," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Zurich, Switzerland, 2014, pp. 188–203.

[11] T. Zhang, S. Liu, C. Xu, and H. Lu, "Mining semantic context information for intelligent video surveillance of traffic scenes," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 149–160, Feb. 2013.

[12] C. Plant, A. Zherdin, C. Sorg, A. Meyer-Baese, and A. M. Wohlschlager, "Mining interaction patterns among brain regions by clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 9, pp. 2237–2249, Sep. 2014.

[13] Y. N. Zhang, X. M. Tong, T. Yang, and W. G. Ma, "Multi-model estimation based moving object detection for aerial video," *Sensors*, vol. 15, no. 4, pp. 8214–8231, Apr. 2015.

[14] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–575, May 2003.

[15] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Hilton Head Island, SC, USA, Jun. 2000, pp. 142–149.

[16] C. Beyan and A. Temizel, "Adaptive Mean-shift for automated multi-object tracking," *IET Comput. Vis.*, vol. 6, no. 1, pp. 1–12, Jan. 2012.

[17] K. H. Zhang, L. Zhang, Q. S. Liu, D. Zhang, and M. H. Yang, "Fast visual tracking via dense spatio-temporal context learning," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Zurich, Switzerland, 2014, pp. 1–15.

[18] C. Ma, X. K. Yang, C. Y. Zhang, and M. H. Yang, "Long-term correlation tracking," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 5388–5396.

[19] M. Demi, "Contour Tracking with a spatio-temporal intensity moment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1141–1154, Jun. 2016.

[20] M. Yang, Y. Wu, and G. Hua, "Context-aware visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 7, pp. 1195–1209, Jul. 2009.

[21] H. Wang, J. Yuan, and Y. Wu, "Context-aware discovery of visual co-occurrence patterns," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1805–1819, Apr. 2014.

[22] T. Zhang, B. Ghanem, S. Liu, C. Xu, and N. Ahuja, "Robust visual tracking via exclusive context modeling," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 51–63, Jan. 2016.

[23] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Firenze, Italy, 2012, pp. 702–715.

[24] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[25] Z. Ji and W. Wang, "Object tracking based on local dynamic sparse model," *J. Vis. Commun. Image Represent.*, vol. 28, pp. 44–52, Apr. 2015.

[26] S. H. Lee and M. G. Kang, "Motion tracking based on area and level set weighted centroid shifting," *IET Comput. Vis.*, vol. 4, no. 2, pp. 73–84, Jun. 2010.

[27] W. Hu, X. Li, W. Luo, X. Zhang, S. J. Maybank, and Z. Zhang, "Single and multiple object tracking using log-Euclidean Riemannian subspace and block-division appearance model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2420–2440, Dec. 2012.

[28] J. Kwon and K. M. Lee, "Highly nonrigid object tracking via patch-based dynamic appearance modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2427–2441, Oct. 2013.

[29] Y. Yuan, H. Yang, Y. Fang, and W. Lin, "Visual object tracking by structure complexity coefficients," *IEEE Trans. Multimedia*, vol. 17, no. 8, pp. 1125–1136, Aug. 2015.

[30] Y. Yuan, J. Fang, and Q. Wang, "Online anomaly detection in crowd scenes via structure analysis," *IEEE Trans. Cybern.*, vol. 45, no. 3, pp. 562–575, Mar. 2015.

[31] Y. Yang and G. Sundaramoorthi, "Shape tracking with occlusions via coarse-to-fine region-based sobolev descent," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 1053–1066, May 2015.

[32] X. Li, C. H. Shen, A. Dick, Z. F. Zhang, and Y. T. Zhuang, "Online metric-weighted linear representations for robust visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 931–950, May 2016.

[33] F. Zhou, F. D. L. Torre, and J. K. Hodgins, "Hierarchical aligned cluster analysis for temporal clustering of human motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 582–596, Mar. 2013.

[34] C. A. Bhatt and M. S. Kankanhalli, "Multimedia data mining: State of the art and challenges," *Multimedia Tools Appl.*, vol. 51, no. 1, pp. 35–76, Jan. 2011.

[35] A. Chemchem and H. Drias, "From data mining to knowledge mining: Application to intelligent agents," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1436–1445, Feb. 2015.

[36] Y. Zhang, G. Pan, K. Jia, M. Lu, Y. Wang, and Z. Wu, "Accelerometer-based gait recognition by sparse representation of signature points with clusters," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1864–1875, Sep. 2015.

[37] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.

[38] C. R. Angel, C. C. Juan, and A. G. Fabio, "Visual pattern mining in histology image collections using bag of features," *Artif. Intell. Med.*, vol. 52, no. 2, pp. 91–106, Jun. 2011.

[39] R. O. Duda, P. E. Hart, and D. G. Stork, "Unsupervised learning and clustering," in *Pattern Classication*. Hoboken, NJ, USA: Wiley, 2001, pp. 648–649.

[40] P. H. Li, "A clustering-based color model and integral images for fast object tracking," *Signal Process, Image commun.*, vol. 21, no. 8, pp. 676–687, Aug. 2006.

[41] S. Kong, Z. L. Jiang, and Q. Yang, "Modeling neuron selectivity over simple midlevel features for image classification," *IEEE Trans. Image Process*, vol. 24, no. 8, pp. 2404–2414, Aug. 2015.

[42] A. K. Qin and D. A. Clausi, "Multivariate image segmentation using semantic region growing with adaptive edge penalty," *IEEE Trans. Image Process*, vol. 19, no. 8, pp. 2157–2170, Aug. 2010.

[43] D. Pelleg and A. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," in *Proc. Int. Conf. Mach. Learn. (ICML)*, San Francisco, CA, USA, Jun. 2000, pp. 727–734.

[44] L. I. Kuncheva and D. P. Vetrov, "Evaluation of stability of k-means cluster ensembles with respect to random initialization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1798–1808, Nov. 2006.

[45] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1450–1464, Sep. 2006.

[46] Z. N. Feng and Y. M. Zhu, "A survey on trajectory data mining: Techniques and applications," *IEEE Access*, vol. 4, pp. 2056–2067, 2016.

[47] H. Grabner, J. Matas, L. V. Gool, and P. Cattin, "Tracking the invisible: Learning where the object might be," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, Jun. 2010, pp. 1285–1292.

[48] W. Quan, J. X. Chen, and N. Y. Yu, "Adaptive collaborative tracking for multiple targets," *Meas. Sci. Technol.*, vol. 23, no. 12, pp. 3387–3395, Nov. 2012.

[49] Y.-B. Kang, S. Krishnaswamy, and A. Zaslavsky, "A retrieval strategy for case-based reasoning using similarity and association knowledge," *IEEE Trans. Cybern.*, vol. 44, no. 4, pp. 473–487, Apr. 2014.

[50] J. Wang and Y. Yagi, "Integrating color and shape-texture features for adaptive realtime object tracking," *IEEE Trans. Image Process*, vol. 17, no. 2, pp. 235–240, Feb. 2008.

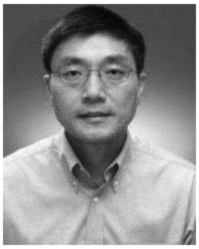
[51] X. Yang, M. Wang, and D. C. Tao, "Robust visual tracking via multi-graph ranking," *Neurocomputing*, vol. 159, no. 1, pp. 35–43, Jul. 2015.

[52] I. S. Kim, H. S. Choi, K. M. Yi, J. Y. Choi, and S. G. Kong, "Intelligent visual surveillance—A survey," *Int. J. Control Autom.*, vol. 8, no. 5, pp. 926–939, Oct. 2010.

[53] H. Lu, H. S. Li, W. L. Zou, and S. M. Fei, "A robust algorithm for tracking object under occlusion and illumination change," in *Proc. IEEE Conf. Control, Autom., Robot. Vis. (ICARCV)*, Guang Zhou, China, Dec. 2012, pp. 1354–1459.

HONG LU received the Ph.D. degree in control theory and control engineering from the School of Automation, and the M.S. degree from the School of electrical engineering, Southeast University, Nanjing, China, in 2009 and 2003, respectively. She is currently an Associate Professor with the School of Automation, Nanjing Institute of Technology, Nanjing, China. She is also a Visiting Scholar with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her current research interests include object detection, object tracking, and image and video data mining.

KE GU received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively. In 2014, he was with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. From 2014 to 2015, he was with the School of Computer Engineering, Nanyang Technological University, Singapore. In 2015, he was with the Department of Computer Science and Technology, Peking University, Beijing, China. His research interests include quality assessment, contrast enhancement, saliency detection and object tracking. He received the Best Paper Award at the IEEE International Conference on Multimedia and Expo in 2016. He is a Special Session Organizer for VCIP 2016. He is the Reviewer of the IEEE T-IP, the T-MM, the T-CSVT, the T-CYB, the T-BC, the J-STSP, the SPL, the *Information Sciences*, the *Neurocomputing*, the SPIC, the JVCI, and the DSP. He has reviewed over 50 journal papers each year.



WEISI LIN (F'16) received the Ph.D. degree from Kings College London. He is currently an Associate Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling, with over 340 refereed papers published in international journals and conferences. He has been on the Editorial Board of the IEEE Transactions on Image Processing, the IEEE Transactions on Multimedia from 2011 to 2013, the IEEE Signal Processing Letters, and the *Journal of Visual Communication and Image Representation*. He has been elected as an APSIPA Distinguished Lecturer in 2012 and 2013. He served as a Technical-Program Chair of Pacific-Rim Conference on Multimedia in 2012, the IEEE International Conference on Multimedia and Expo in 2013, and the International Workshop on Quality of Multimedia Experience in 2014. He is a fellow of Institution of Engineering Technology, and an Honorary Fellow of the Singapore Institute of Engineering Technologists.

WENJUN ZHANG (F'11) received the B.S., M.S., and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 1984, 1987, and 1989, respectively. From 1990 to 1993, he was a Post-Doctoral Fellow with Philips Kommunikation Industrie AG. Nuremberg, Germany, where he was actively involved in developing HD-MAC system. He joined the Faculty of Shanghai Jiao Tong University in 1993, where he was a Full Professor with the Department of Electronic Engineering in 1995. As the national HDTV TEEG Project Leader, he successfully developed the first Chinese HDTV Prototype System in 1998. He was one of the main contributors to the Chinese Digital Television Terrestrial Broadcasting Standard issued in 2006 and is leading team in designing the next generation of broadcast television system in China from 2011. He has authored over 90 papers in international journals and conferences and holds over 40 patents. His main research interests include digital video coding and transmission, multimedia semantic processing and intelligent video surveillance. He is a Chief Scientist of the Chinese National Engineering Research Centre of Digital Television, an industry/government consortium in DTV technology research and standardization ,and the Chair of the Future of Broadcast Television Initiative Technical Committee.