Stable Feature Mining Based Object Representation for Tracking Applications

Hong Lu[1,2], Ke Gu[2]
1 School of Automation, Nanjing Institute of Technology, Nanjing, 211167, China
2 School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore

Abstract—**Tracking the object with stable features is an important and challenging task in real scenes where the object appearance constantly changes or is disturbed by the background etc. In this paper, a novel frame work of object representation and tracking based on stable feature mining is presented. Firstly, the object region is adaptively detected and then the peak contour of V color component histogram in the object region is extracted to calculate candidate and region peaks which are used for acquiring the number of clusters and further classifying the object appearance. Secondly, the connection subregions belonging to every cluster are described with observation and increment models, and subregions association between the object template and the current observation is then utilized to mine stable subregion pairs and get feature change ratios. Finally, stable subregion displacements are weighted fused to locate the object in the current frame, and the object template is updated in terms of average increment variations. Experimental results show the excellent performance of the proposed algorithm in cluster number adaptivity, robustness in stable feature mining and accurateness in object tracking.**

Keywords- adaptively clustering; observation and increment representations; subregions association; stable features mining; object tracking

I. INTRODUCTION

Modeling the object with sparse and reliable features to enhance the object representation power and efficiency is very critical in many computer vision applications such as action retrieval [1], object recognition [2], detecting and tracking [3][4]. However, object features are not invariant in the unconstrained environment, which make the current object observation be slightly or severely inconsistent with the predetermined object model, lead to the object missing, and ultimately degrade the accurateness of real applications such as intelligent traffic surveillance [5], medical image understanding [6] and Unmanned Aerial Vehicle surveillance [7] systems. Therefore, robustly and efficiently modeling and tracking the object are still an area of active research.

To cope with these problems, a number of elegant algorithms have been established to endow the object model with strong discriminative power, such as color histogram [8] and correlogram [3], shape [9], spatio-temporal [10] and edge-color [4] contexts, and local dynamic sparse model [11] based object feature representations. Among of them, image and video data mining based feature extracting algorithms attract much notice and obtain remarkable results. Image mining deals with the acquirement of implicit knowledge, image data relationship or other patterns not explicitly stored in the image [6][12]. Clustering as one of the unsupervised image mining techniques, can group unlabelled raw images into meaningful classes, dispatch the heterogeneous data into different groups [13] or mine visual patterns according to the image content without the priori knowledge [14-15], so clustering is very useful in finding robust features for further efficiently modeling the target object. K-means clustering [16] is very effective in highlighting the main color or texture features etc while providing dimensionality reduction on the original data, but it need set the cluster number in advance. Plant et al. [6] presented the interaction K-means to cluster the functional magnetic resonance image with the pre-set cluster number for brain function understanding. Video mining can extract the moving object features, spatial or temporal correlations of those features [2], discover the object activity and event [5], and track the object [17-18] etc without little assumption about video contents. Association as an important video mining method can get related information or discover two features, or objects that always occur simultaneously etc. Yang et al. [17] 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. And then these auxiliary observations were used to track the target object whose current observation was unreliable due to occlusion or background disturbance etc. Grabner et al. [18] mined supporters which were temporally but useful for tracking the object from the embedded context and dealing with the occlusion in real scenarios. Zhang et al. [5] 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. Nevertheless, every method mainly deals with one or several special problems, so to realize accurately and robustly feature mining and object tracking under any scenes still has a long way to go.

In this paper, clustering and association are combined to mine stable object features and track the object in image sequences. The S and V color components (in HSV color space) in the object region detected with adaptive background difference are classified by the adaptive cluster number based K-means clustering. The connection subregions of every cluster are then labelled and modeled. To mine stable subregions, the observation model is associated with the object template in terms of the multi-feature distances and change ratios. The increment model is utilized to real-timely calculate the feature variations and update the object template. To obtain the object position in the current frame, the center displacements of stable subregions are weighted fused. The results from the experiments illustrate that the proposed method can continuously mine stable subregions and robustly track the object in real unconstrained scenarios.

This research was funded partially by the National Natural Science Foundation of China (61305011), the Jiangsu Government Scholarship for Overseas Studies (JS-2014-126), the Natural Science Foundation of Jiangsu Higher Education Institutions of China (14KJB460016), the Innovation Foundation of Nanjing Institute of Technology (CKJB201405) and the Open Fund of the Key Laboratory of remote measurement and control technology in Jiangsu Province (YCCK201401).

Four main contributions are summarized below: 1) we establish a new stable feature mining based object representation and tracking framework. 2) a new cluster number acquiring technology is explored. 3) local subregion observation and increment representations are proposed which are more efficient in handling the partial appearance variation and loss. 4) our tracker performs better than the color histogram based Mean shift [8] and recently developed the spatio-temporal context [10] based trackers on relevant image databases. Compared with the previous works, to the best of our knowledge, this paper is the first to propose a clustering and association based stable feature mining and object tracking framework. In addition, our algorithm obtains a high performance.

The rest of this paper is organized as follows: Section 2 summarizes the presented object representation and tracking method. Section 3 proposes object representation and stable feature mining algorithms. Section 4 describes the object tracking algorithm. Section 5 gives experimental results and comparisons. Finally some conclusion and discussion are presented in Section 6.

II. The Overview of the Proposed Algorithm

The flow chart of the proposed algorithm is shown in Figure 1.

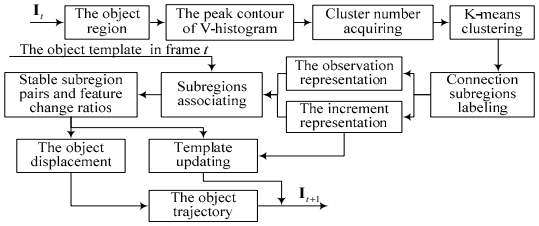


Figure 1. The flow chart of the presented algorithm.

In video based automatic object recognition and tracking, the object region is often extracted to decrease the searching cost and restrain the background disturbance etc. Similarly, the adaptive background difference here is firstly employed to extract the object region. And then the peak contour of V component histogram in the object region is utilized to compute the cluster number. K-means clustering is then adopted for classifying the object appearance. The 8-connection subregions of every cluster are used to build observation and increment representations. Furthermore, subregions association between the observation model and the object template is used to mine the stable subregion pairs and get feature change ratios. Finally, the center displacements of the stable subregions are weighted combined to derive the object displacement and trajectory in the current frame, and the object template is real-timely updated to adapt to gradual illumination change, and object pose and scale variations etc.

III. Object Representation and Stable Feature Mining

A. Cluster Number Acquiring and Intraframe Clustering

Let $\mathbf{I}_t(\mathbf{x})$ be the current image as in Figure 2(a). The object image $\mathbf{O}_t(\mathbf{x})$ is detected with the adaptive threshold based background difference [19], and shown in Figure 2(b). $\mathbf{O}_t(\mathbf{x})$=$\mathbf{I}_t(\mathbf{x})$ and 0 represent the object pixel and the background pixel respectively. t is the frame number and $\mathbf{x}$ describes the pixel coordinate.

Since S and V components are independent of color information and much more easily processed than in RGB space, we use them to cluster the object appearance. Figure 2(c) shows S and V component histogram contours of $\mathbf{O}_t(\mathbf{x})$. Obviously V contour only reflects one main peak while S contour embodies the distinctions of different clusters with multiple main peaks which represent car outside and shadow etc. Hence, the V component histogram is adopted to compute the cluster number K_t as follows.

Step 1. Smooth the V component contour to highlight the main peak values ($MPVs$) (red solid line in Figure 2(d)).

Step 2. Judge $MPVs$ as the candidate ones when $MPVs \geq \beta \cdot MEAN$ (blue solid line). β is a scale factor and $MEAN$ (blue dash line) represents the average value of MPVs.

Step 3. Sum the $MPVs$ which are 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 peak values difference $APVD$ between the adjacent candidate peaks, and judge the candidate peaks as the region peaks when $APVD > \alpha_1$, otherwise retain the maximum peak to compare with the next candidate peak and repeat Step 4.

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

To classify $\mathbf{O}_t(\mathbf{x})$, the S and V components $\mathbf{V}_t(\mathbf{x})$ and $\mathbf{S}_t(\mathbf{x})$ in $\mathbf{O}_t(\mathbf{x})$ are reshaped line by line into a $M \times 2$ sample intensity matrix $\mathbf{\Gamma}_t(n)$. Then, $\mathbf{\Gamma}_t(n)$ is clustered using K-means algorithm with K_t and the Euclidean distance as a dissimilarity metric to gain a $M \times 2$ clustered matrix $\mathbf{D}_t(n)$. M is the total number of object pixel and $n \in [1, M]$.

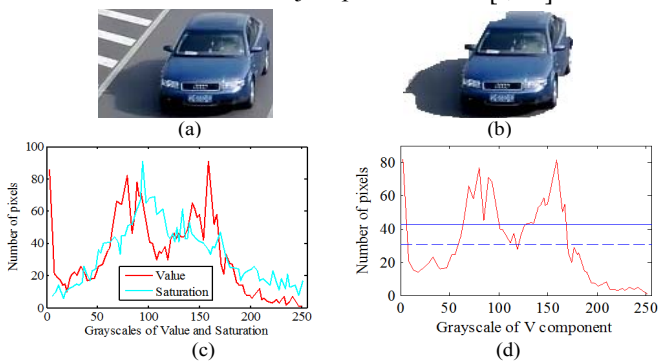


Figure 2. The S and V component contours of the object image. (a) the original image; (b) the object region; (c) the component contours; (d) the filtered peaks contour of the V component.

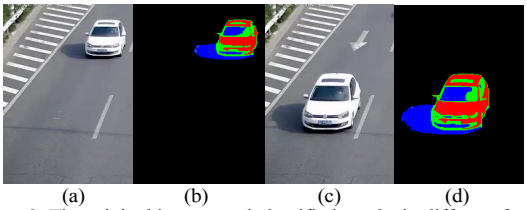


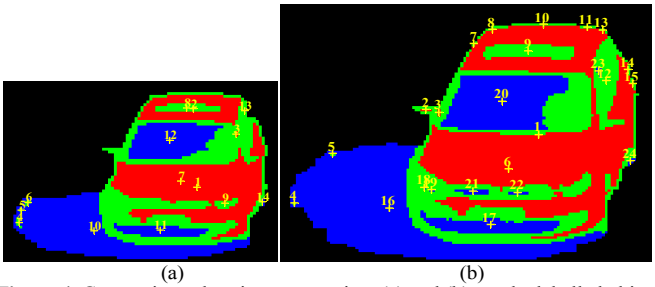
Figure 3. The original images and classified results in different frames.

Finally, $\mathbf{D}_t(n)$ is mapped to $\boldsymbol{O}_t(\mathbf{x})$ according to the one-to-one relationship between n and $\mathbf{x}$, and very cluster is represented with a pseudo-color as in Figure 3. Figure 3(a) and (c) illustrate the original images, and (b) and (d) are the clustered results of the object regions in frames 230 and 239 respectively. The cluster numbers in two frames are both 3.

B. Subregion Representation and Stable Feature Mining

In real scenarios, illumination, object pose and scale variations etc, can alter the clustered result, such as the cluster area increasing and the cluster center moving as in Figure 3(b) and (d). Thus, it is very important to extract the stable and robust object features even if they are sparse or exist momentarily. Figure 4 illustrates the local subregion comparisons between frames 230 and 239, where the 8-connection subregions belonging to each cluster are calculated and labelled with the numerical orders, and every subregion center position is marked with the '+' (plus) character. In Figure 4(a), the cluster marked with red, green and blue correspond to subregions 1~3, subregions 4~9 and subregions 10~14 respectively. Meanwhile, the cluster marked with red, green and blue in Figure 4(b) correspond to subregions 1~3, subregions 4~15 and subregions 16~24 respectively.

There are obviously associations between Figure 4(a) and (b), e.g. subregion 8 (skylight glass labelled with green) in frame 230 and subregion 9 in frame 329, and subregion 10 (shadow) in frame 230 and subregion 16 in frame 239. These related subregions contain the rules of object appearance, position and scale variations etc. Thus, they can be regarded as important clues for modeling and tracking the object.



(a) (b)

Figure 4. Connection subregions comparing. (a) and (b) are the labelled object regions in frames 230 and 239 respectively.

Let $\mathbf{E}_t^{r1}=\left\{E_{\mathbf{x},t}^{r1},E_{A,t}^{r1},E_{\mathbf{C},t}^{r1}\right\}$ be the $r1$-th subregion template, $\mathbf{F}_t^{s1}=\left\{F_{\mathbf{x},t}^{s1},F_{A,t}^{s1},F_{\mathbf{C},t}^{s1}\right\}$ denote the $s1$-th subregion observation model, and $\Delta\mathbf{E}_t^{r1}=\left\{\Delta E_{A\_ratio,t}^{r1},\Delta E_{\mathbf{C}\_ratio,t}^{r1}\right\}$ describe the $r1$-th subregion increment model. $\hat{\mathbf{F}}_t^m=\left\{\hat{F}_{\mathbf{x},t}^m,\hat{F}_{A,t}^m,\hat{F}_{\mathbf{C},t}^m\right\}$ and $\hat{\mathbf{E}}_t^m=\left\{\hat{E}_{\mathbf{x},t}^m,\hat{E}_{A,t}^m,\hat{E}_{\mathbf{C},t}^m\right\}$ stand for the m-th stable subregion pair $\left(\hat{\mathbf{F}}_t^m,\hat{\mathbf{E}}_t^m\right)$ mined from $\mathbf{F}_t^{s1}$ and $\mathbf{E}_t^{r1}$. The subscripts $\mathbf{x},A,\mathbf{C}=\{S,V\},A\_ratio$ and $\mathbf{C}\_ratio$ represent the subregion center coordinate, area, color (S and V), 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. (1). Similarly, $E_{S,t}^{r1},F_{S,t}^{s1}$ and $F_{V,t}^{s1}$ are obtained.

To automatically find the associated subregions in image sequence, the template difference $\boldsymbol{\Delta}_t^{r1}=\mathbf{E}_t^{r1}-\mathbf{F}_t^{s1}$ is firstly computed. And the center distance $d_{\mathbf{x},t}^{r1}$, the area change ratio $\Delta E_{A\_ratio,t}^{r1}$ and the color change ratio $\Delta E_{\mathbf{C}\_ratio,t}^{r1}$ are further calculated via Eqs. (2)~(4) respectively.

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

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

$$\Delta E_{A\_ratio,t}^{r1}=\frac{\Delta_{A,t}^{r1}}{E_{A,t}^{r1}} \tag{3}$$

$$\Delta E_{\mathbf{C}\_ratio,t}^{r1}=\left(\frac{\Delta_{\mathbf{C},t}^{r1}\cdot\left(\Delta_{\mathbf{C},t}^{r1}\right)^T}{E_{\mathbf{C},t}^{r1}\cdot\left(E_{\mathbf{C},t}^{r1}\right)^T}\right)^{\frac{1}{2}} \tag{4}$$

Then, subregions association is used to mine $\hat{\mathbf{F}}_t^m$ and $\hat{\mathbf{E}}_t^m$, and obtain the stable increment $\Delta\hat{\mathbf{E}}_t^m=\left\{\Delta\hat{E}_{A\_ratio,t}^m,\Delta\hat{E}_{\mathbf{C}\_ratio,t}^m\right\}$ via Eq. (5). The stable subregion difference $\hat{\boldsymbol{\Delta}}_t^m=\hat{\mathbf{E}}_t^m-\hat{\mathbf{F}}_t^m$ is further computed. $\mu_t=\max\left\{\left(E_{A,t}^{r1}\right)^{\frac{1}{2}},\left(F_{A,t}^{s1}\right)^{\frac{1}{2}}\right\}$ is an adaptive threshold, λ_1 and λ_2 are constant thresholds.

$$\begin{aligned}&\left\{\hat{\mathbf{E}}_t^m,\hat{\mathbf{F}}_t^m,\Delta\hat{\mathbf{E}}_t^m\right\}=\\&\left\{\mathbf{E}_t^{r1},\mathbf{F}_t^{s1},\Delta\mathbf{E}_t^{r1}\right\}\Big|\left\{\mathbf{E}_t^{r1},\mathbf{F}_t^{s1},\Delta\mathbf{E}_t^{r1}\right\}:=\\&\arg\left(d_{\mathbf{x},t}^{r1}\le\mu_t\cap\left|\Delta_{A\_ratio,t}^{r1}\right|\le\lambda_1\cap\Delta_{\mathbf{C}\_ratio,t}^{r1}\le\lambda_2\right)\end{aligned} \tag{5}$$

IV. OBJECT TRACKING

It is assumed that the current object template $\mathbf{E}_t^{r1}$ is updated with the stable subregions mined in frame t-1, the center displacement $\hat{\Delta}_{\mathbf{x},t}^m$ between $\hat{\mathbf{E}}_t^m$ and $\hat{\mathbf{F}}_t^m$ will directly reflects the motion rules of the object in neighbor frames. To obtain the object trajectory $\mathbf{x}_t$, the average template increment $\overline{\boldsymbol{\Delta}}_t=\left\{\overline{\Delta}_{\mathbf{x},t},\overline{\Delta}_{A,t},\overline{\Delta}_{\mathbf{C},t}\right\}$ is computed by weighted fusing $\hat{\boldsymbol{\Delta}}_t^m$ via Eq. (6), where $\hat{E}_{A,t}^m$ is used to increase the voting power of the subregion being with larger area and vice versa. Then $\mathbf{x}_t$ is acquired in terms of $\overline{\Delta}_{\mathbf{x},t}$, the historical trajectory $\mathbf{x}_{t-1}$ and the detected center trajectory $\mathbf{x}_t^{\det ect}$ of $\mathbf{O}_t(\mathbf{x})$ as Eq. (7).

$$\overline{\boldsymbol{\Delta}}_t=\frac{\sum_{m=1}^{N3}\hat{E}_{A,t}^m\cdot\hat{\boldsymbol{\Delta}}_t^m}{\sum_{m=1}^{N3}\hat{E}_{A,t}^m} \tag{6}$$

$$\mathbf{x}_t=\gamma\cdot\left(\mathbf{x}_{t-1}-\overline{\Delta}_{\mathbf{x},t}\right)+(1-\gamma)\cdot\mathbf{x}_t^{\det ect} \tag{7}$$

In addition, the subregion area variations in adjacent frames are almost with similar rule, i.e. simultaneously increasing or decreasing, hence we utilize the average area change ratio $\overline{\Delta}_{A\_ratio,t}$ of stable subregions to update $E_{A,t}^{r1}$ including of those unstable subregions without directly data source for self-renewaling. $\overline{\Delta}_{A\_ratio,t}$ is calculated via Eq. (8).

$\mathbf{E}_t^{r1}$ is updated according to Eqs. (9)~(11), where *SIGN* denotes the plus-minus sign of $\overline{\Delta}_{A\_ratio,t}$. Since $\hat{\Delta}_{\mathbf{x},t}^m$ has larger variation in those subregions being with larger areas and is roughly proportional to $\left|\Delta\hat{E}_{A\_ratio,t}^m\right|^{\frac{1}{2}}$, we update $E_{\mathbf{x},t}^{r1}$ by taking into account $\overline{\Delta}_{\mathbf{x},t}$ and $\overline{\Delta}_{A\_ratio,t}$ as Eq. (9). Similarly, $\overline{\Delta}_{A,t}$ and $\overline{\Delta}_{A\_ratio,t}$ are applied to acquiring $E_{A,t+1}^{r1}$ as Eq. (10). To deal with gradual illumination variation, the average color translation $\overline{\Delta}_{\mathbf{C},t}$ here is employed to update $E_{\mathbf{C},t}^{r1}$ via Eq. (11).

$$\overline{\Delta}_{A\_ratio,t}=\frac{\sum_{m=1}^{N3}\hat{\Delta}E_{A\_ratio,t}^m\cdot\hat{E}_{A,t}^m}{\sum_{m=1}^{N3}\hat{E}_{A,t}^m} \tag{8}$$

$$E_{\mathbf{x},t+1}^{r1}=E_{\mathbf{x},t}^{r1}-\overline{\Delta}_{\mathbf{x},t}\cdot\left(1-SIGN\cdot\left|\overline{\Delta}_{A\_ratio,t}\right|^{\frac{1}{2}}\right) \tag{9}$$

$$E_{A,t+1}^{r1}=E_{A,t}^{r1}-\overline{\Delta}_{A,t}\cdot\left(1-\overline{\Delta}_{A\_ratio,t}\right) \tag{10}$$

$$E_{\mathbf{C},t+1}^{r1}=E_{\mathbf{C},t}^{r1}-\overline{\Delta}_{\mathbf{C},t} \tag{11}$$

V. Experimental Results

To evaluate our method, two image sequences in traffic scenes, containing gradual illumination and object scale variations, lower resolution appearance and background disturbence etc are employed to show robustness of feature mining and object tracking. As to the control parameters, all simulations were obtained empirically setting $\beta\in[1.3,1.4]$, $\alpha_1=15$, $\alpha_2\in[0.35,0.45]$, $\lambda_1=0.3$, $\lambda_2=0.1$ and $\gamma=0.8$. We run the proposed tracker, the Mean shift (MS) tracker [8], and the spatio-temporal context (STC) tracker [10] on those testing data, respectively. Some examples are presented in Figures 5~10. The round marks in Figures 5 and 8 are the initial object trajectories, and the round marks in Figures 6, 7, 9 and 10 denote the tracked object locations.

A. Stable Features Mining and Object Tracking

Figure 5 illustrates the process of acquiring the initial object template in image sequence 1. Figures 5(a) and (b) are the original image and the detected object region of frame 230 respectively, where the object is with the size 85×59, 3 clusters and 14 subregions. The object template is scaled up in Figure 5(c) and the subregions are numbered so as to conveniently describe the processes of stable features mining and object tracking as shown in Figure 6. The initial object trajectory (the white round mark) in Figure 5(b) and (c) is the center position of the object region including of the shadow. Since the shadow has the same moving with the object and may be an important clue under the case of the main body of the object is occluded or disturbed, here it is regarded as a part of the object.

The results of stable feature mining and object tracking under gradual illumination variation, and object scale and resolution changes are presented in Figure 6. Row 1 shows the clustered results of frames 231, 235, 239, 241 and 243. The object scale changes gradually when the object approaches the camera and more appearance details are displayed, which leads to the corresponding changes in clustered results and subregions numbers. The 8-connection subregions in row 1 are labelled from left to right with the total numbers of subregions being, in order, 21, 24, 38 and 33 respectively. The subregions number decreases in frame 243 due to the object being about to leave from the field of view and only the partial object body being detected. Row 2 demonstrates the mined stable subregions marked with the same colors and '+' characters as row 1. The total numbers of stable subregions from left to right are, in order, 7, 5, 2, 2 and 1 respectively.

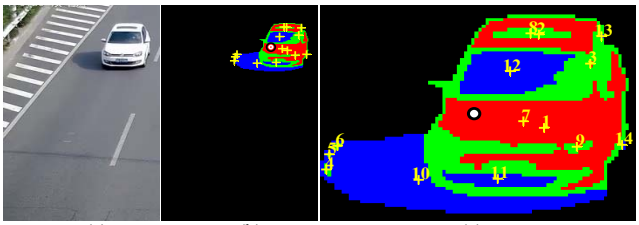


Figure 5. The initial object template in image sequence 1. (a) the original image; (b) the labelled subregions; (c) the scaled up and numbered subregions.

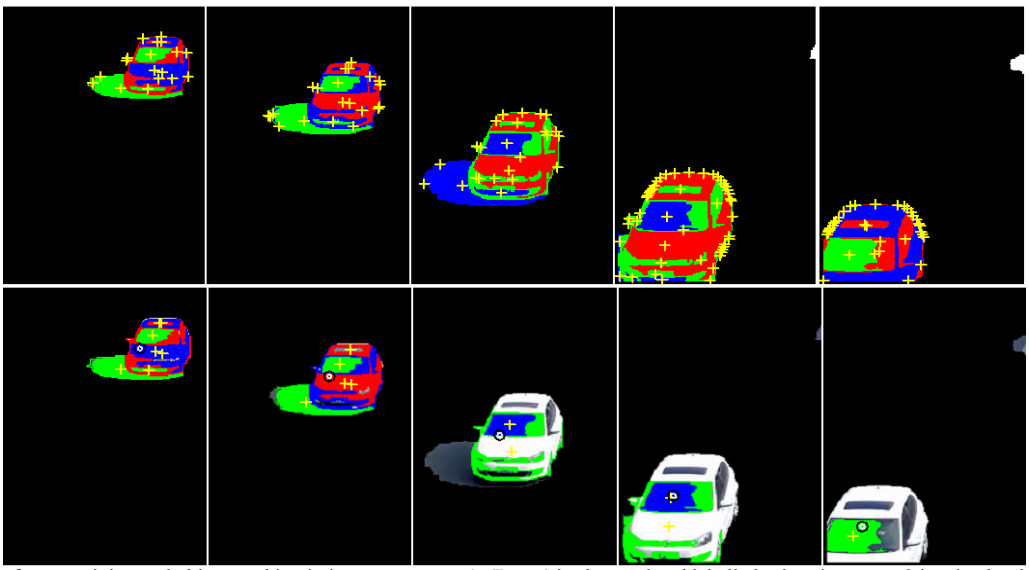


Figure 6. Stable feature mining and object tracking in image sequence 1. (Row 1 is clustered and labelled subregions; row 2 is mined stable subregions and tracked results.)

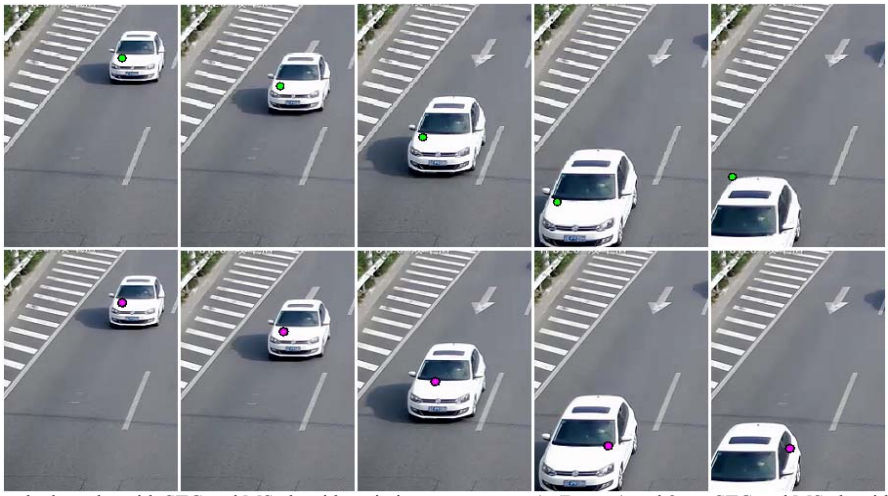
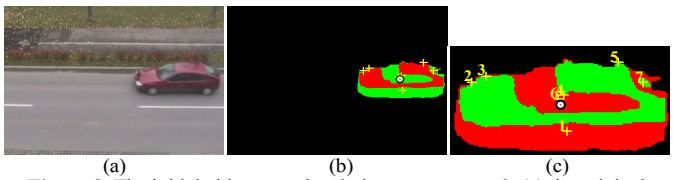


Figure 7. The tracked results with STC and MS algorithms in image sequence 1. (Rows 1 and 2 are STC and MS algorithms respectively.)

The stable subregion centers are further fused to locate the object trajectory (the white round dot) in the image sequence as in row 2. In frames 231 and 235, the stable subregions such as subregions 10 (shadow) and 12 (windshield), almost covering the total object region and most of the stable subregions, provide reliable and powerful supports for object locating. In frames 239 and 241, the object scale and velocity increase quickly, only two stable subregions (such as the car body labelled with subregion 7) are mined. However, the object still is tracked robustly in absence of most of object appearance. The tracking effectiveness is further reflected in frame 243, where the detected object is incomplete and only one stable subregion (windshield) is utilized to locate the object.

The tracked object trajectories of STC and MS algorithms are given in Figure 7. STC tracker can accurately locate the specific object position (the green round dot) during the object moving (except frame 243), but it could not perceive the object appearance loss and adjust the object model to obtain the valid object center in the partial object region. Since the object center changes simultaneously along with the object moving and the partial body loss, the scale and template updates in MS tracker cannot pursue these accelerated variations and ultimately cause trajectory drift (the purplish red round dot). The same problem happens to STC tracker in frame 243, where the retarded scale update causes the trajectory lag. By comparison, our tracker can robustly locate the object trajectory in terms of clues provided by the stable subregion features under this case.



(a) (b) (c)

Figure 8. The initial object template in image sequence 2. (a) the original image; (b) the labelled subregions; (c) the scaled up and numbered subregions.

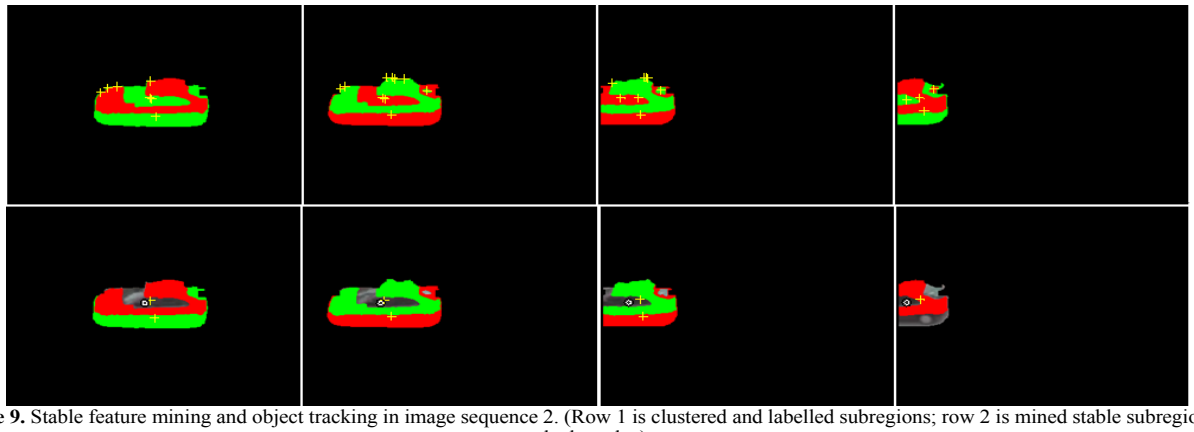


Figure 9. Stable feature mining and object tracking in image sequence 2. (Row 1 is clustered and labelled subregions; row 2 is mined stable subregions and tracked results.)

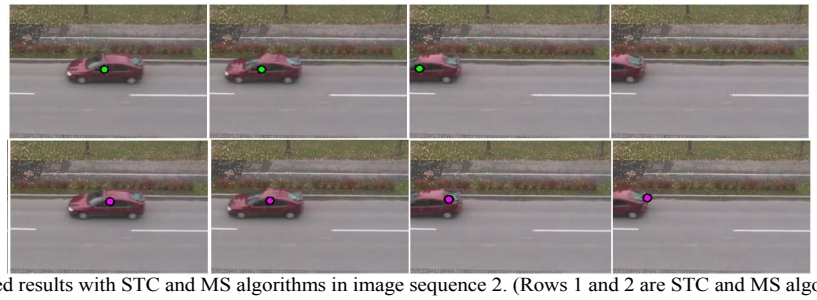


Figure 10. The tracked results with STC and MS algorithms in image sequence 2. (Rows 1 and 2 are STC and MS algorithms respectively.)

Figure 8 shows the initial object template with the size 154×70, 2 clusters and 7 subregions, extracted from frame 120 (Figure 8(a)) in image sequence 2. Similarly, the clustered subregion centers and the initial object trajectory (the round mark) are given in Figure 8(b) and (c). Figure 8(c) presents the scaled up and numbered object subregions.

The results of stable feature mining and object tracking under lower resolution appearance and background disturbance are described in Figure 9. Row 1 illustrates the clustered results of frames 124, 127, 130 and 131. In spite of the object being blurry and influenced by the shadow and reflection of the plant in the scene, such as the car roof in frames 130 and 131, our method can robustly cluster the object motion region. Furthermore, the subregion areas and numbers constantly change along with object moving as shown in row 1, where the total numbers of subregions are, in order, 7, 10, 8 and 4 respectively. Row 2 gives the mined stable subregions marked with the same colors and '+' characters as row 1 with the total numbers of stable subregions from top to bottom being, in order, 2, 2, 2 and 1 respectively. In frames 124 and 127, the clustered result of car glass is unstable due to the background reflection and the object pose change, but stable subregions 1 and 4 (car body) combined to achieve robust tracking results. In frames 130 and 131, although the object gradually moves away from the field of view, our method still keeps robustly tracking the partial car body, which further shows the high efficiency of the proposed method.

The tracked results of STC and MS algorithms are illustrated in Figure 10. STC tracker still maintains the steadily tracking performance (the green round dot) in frames 124 and 127. However, it could not locate the new center determined by the partial object body in frame 130, and gives the object trajectory outside of the scene in frame 131. Under MS tracker (the purplish red round dot), there are obvious lags in frames 130 and 131. This is because the local object appearance loses and the background disturbance directly influences the color histogram model and the scale update. And trajectory drifting ultimately occurs. In contrast, the proposed tracker can obtain relatively accurate object trajectory depending on the stable subregion center displacements mined with the interframe association.

B. Quantitative Comparisons and Errors Analyzing

Figure 11 shows the trajectory comparisons between the tracked result $\mathbf{x}_t = \{x_t, y_t\}$ and the ground truth $\hat{\mathbf{x}}_t = \{\hat{x}_t, \hat{y}_t\}$ (red dash and solid lines). $\hat{\mathbf{x}}_t$ and the object scale $\mathbf{h}_t = \{h_{x,t}, h_{y,t}\}$ are acquired by manual method and indicate the center location and scale of the object region. The trajectory error $\mathbf{x}_{error,t} = \mathbf{x}_t - \hat{\mathbf{x}}_t$ and $\mathbf{h}_t$ (red dash and solid lines) are illustrated in Figure 12. Although our trajectory slightly drifts from frame 240 in image sequence 1 (Figure 11(a)) and from frame 129 in image sequence 2 (Figure 11(b)), the relative error growths in both sequences are small with respect to the real-time object size (Figure 12). Our trajectory and STC trajectory along y direction (vertical) in image sequence 2 almost coincide with the true one, while the x (horizontal) trajectory in STC method has larger drift than our method from frame 128 where the object is about to leave the field of view and has a sharp size decrease. MS tracker achieves a less trajectory error in y direction in image sequence 1, but makes higher errors in x direction in both image sequences.

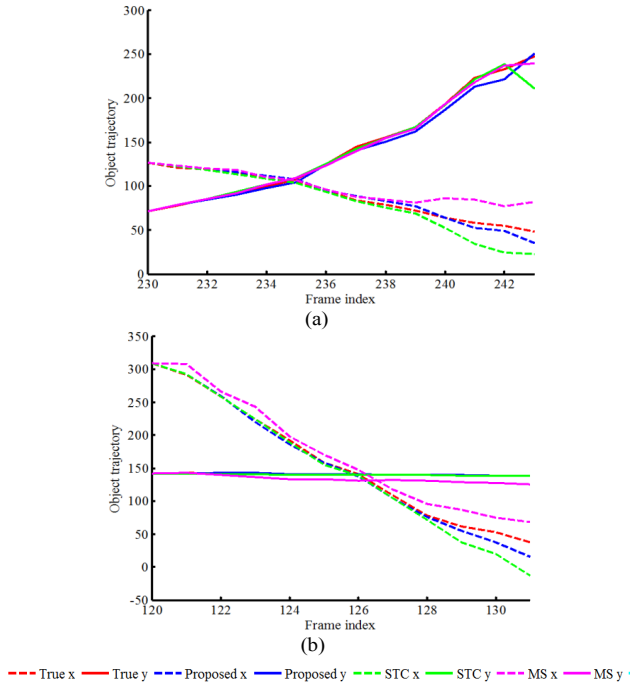


Figure 11. Trajectory tracking comparisons. (a) image sequence 1; (b) image sequence 2.

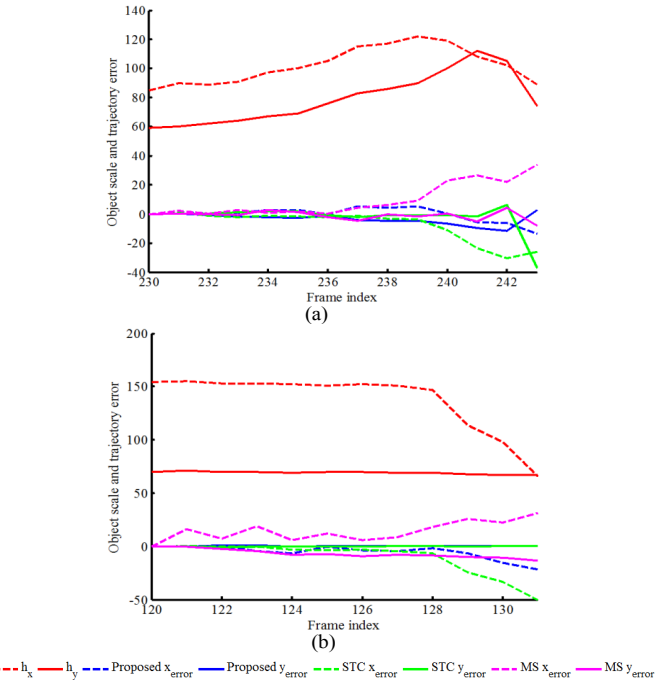


Figure 12. Trajectory errors comparing with the object scale variations. (a) image sequence 1; (b) image sequence 2.

For each testing sequence, the trajectory errors statistics and comparisons are shown in Table 1. The average trajectory error $\bar{x}_{error} = \frac{1}{N}\sum_{t=1}^{N}|x_t - \hat{x}_t|$ is calculated according to the absolute difference between x_t and $\hat{x}_t$. N is the total number of image frames. Similarly, the average trajectory error $\bar{y}_{error}$ is derived. The synthetical error is acquired via $T_{error} = \left(\bar{x}_{error}^2 + \bar{y}_{error}^2\right)^{\frac{1}{2}}$

The average relative error $\overline{x}_{relat\_error}$ of the object trajectory comparing with the object size $h_{x,t}$ in x direction is computed as Eq. (12). Similarly, the average relative error $\overline{y}_{relat\_error}$ is obtained by comparing with $h_{y,t}$. The synthetical relative error is obtained according to $T_{relat\_error} = \left(\overline{x}^2_{relat\_error} + \overline{y}^2_{relat\_error}\right)^{\frac{1}{2}}$. The bold numbers in Table 1 highlight the smallest errors among three algorithms. It is obvious that $\overline{y}_{error}$ and $\overline{y}_{relat\_error}$ under the proposed and STC methods are almost zeroes in image sequence 2, and MS tracker gets the best y trajectory in image sequence 1. However, the proposed method obtains the smallest T_{error} and T_{relat_error} in both sequences.

$$\overline{x}_{relat\_error} = \frac{1}{N}\sum_{t=1}^{N}\frac{\left|x_t - \hat{x}_t\right|}{h_{x,t}} \times 100\% \tag{12}$$

TABLE 1. THE TRAJECTORY ERRORS STATISTICS AND ANALYSIS (UNITS: PIXELS).

Image sequence	Algorithm	$\overline{x}_{error}$	$\overline{y}_{error}$	T_{error}	$\overline{x}_{relat\_error}$	$\overline{y}_{relat\_error}$	T_{relat_error}
Image sequence 1	Proposed	**3.41**	3.78	**5.09**	**3.35%**	4.20%	**5.37%**
	STC	7.68	3.96	8.64	7.56%	5.16%	9.15%
	MS	9.55	**2.28**	9.82	9.23%	**2.75%**	9.63%
Image sequence 2	Proposed	**5.43**	0.43	**5.45**	**5.69%**	0.63%	**5.72%**
	STC	10.75	**0.33**	10.76	12.09%	**0.48%**	12.10%
	MS	14.37	6.73	15.87	12.86%	9.82%	16.18%

VI. CONCLUSIONS

In this paper, we have presented a new approach for adaptively clustering and mining the stable feature to deal with object appearance changes in unconstrained environment, and eventually enhance the robustness of object tracking in image sequence. The motion detecting, adaptive cluster number based K-means clustering, connection subregions modeling and updating, and subregions association are combined to improve the stable feature mining and object tracking performances. Experimental results demonstrate that the proposed algorithm shows favourable cluster number adaptivity, and achieves robustly feature mining and efficiently object tracking. Although the results are promising in certain situations, further evaluation is anticipated in more complicated image sequences. We have implemented all of the experiments without code optimization. Real-timely tracking will be realized in further development.

REFERENCES

[1] L. J. Cao, R. R. Ji, Y. Gao, W. Liu, and Q. Tian, "Mining Spatiotemporal Video Patterns towards Robust Action Retrieval," Neurocomputing, vol. 105, pp. 61-69, 2013.

[2] 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, pp. 1832-1840, 2013.

[3] Q. Zhao, and H. Tao, "A Motion Observable Representation Using Color Correlogram and Its Applications to Tracking," Comput. Vis. Image Und, vol. 113, pp. 273-290, 2009.

[4] 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, pp. 148-152, 2015.

[5] T. Z. Zhang, S. Liu, C. S. Xu, and H. Q. Lu, "Mining Semantic Context Information for Intelligent Video Surveillance of Traffic Scenes," IEEE Trans. Ind. Inf.,vol. 9, pp. 149-160, 2013.

[6] C. Plant, A. Zherdin, C. Sorg, and A. Meyer-Baese, "Wohlschläger A. M. Mining Interaction Patterns among Brain Regions by Clustering," IEEE Trans. Knowl. Data En.,vol. 26, pp. 2237-2249, 2014.

[7] 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, pp. 8214-8231, 2015.

[8] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based Object Tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, pp. 564-575, 2003.

[9] Z. Liu, H. Shen, G. Y. Feng, D. W. Hu, "Tracking Objects Using Shape Context Matching," Neurocomputing, vol. 83, pp. 47-55, 2012.

[10] K. H. Zhang, L. Zhang, Q. S. Liu, D. Zhang, and M.H. Yang, "Fast Visual Tracking via Dense Spatio-Temporal Context Learning," In Proceedings of the 2014 European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6-12 September 2014, pp. 1-15.

[11] Z. J. Ji, and W.Q. Wang, "Object Tracking Based on Local Dynamic Sparse Model," J. Vis. Commun. Image R., vol. 28, pp. 44-52, 2015.

[12] C. A. Bhatt, and M. S. Kankanhalli, "Multimedia Data Mining: State of the Art and Challenges," Multimed. Tools Appl., vol. 51, pp. 35-76, 2011.

[13] A. Chemchem, and H. Drias, "From Data Mining to Knowledge Mining: Application to Intelligent Agents," Expert Syst. Appl., vol. 42, pp. 1436-1445, 2015.

[14] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: a Review," ACM Comput. Surv., vol. 31, pp. 264-323, 1999.

[15] 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, pp. 91-106, 2011.

[16] R. O. Duda, P. E. Hart, and D. G. Stork, in: Pattern Classication, 2nd Edition, New Jersey: Wiley, 2001.

[17] M. Yang, Y. Wu, and G. Hua, "Context-Aware Visual Tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, pp. 1195-1209, 2009.

[18] H. Grabner, J. Matas, L.V. Gool, and P. Cattin, "Tracking the Invisible: Learning Where the Object Might be," In Proceedings of the 2010 IEEE conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 13-18 June 2010, pp. 1285-1292.

[19] H. Lu, H. S. Li, W. L. Zou, S. M. Fei, "A Robust Algorithm for Tracking Object under Occlusion and Illumination Change," In Proceedings of the 2012 IEEE conference on Control, Automation, Robotics and Vision (ICARCV), Guang Zhou, China, 5-7 December 2012, pp. 1354-1459.