Hybrid Region and Interest Points-based Active Contour for Object Tracking

W. Aitfares $^{1,2,3}$, E.H. Bouyakhf $^1$, A. Herbulot $^{2,3}$, F. Regragui $^1$ and M. Devy $^{2,3}$

$^1$ LIMIARF-FSR, Université Mohammed V-Agdal, B.P 1014 RP, Rabat, Maroc
$^2$ CNRS ; LAAS ; 7 avenue du Colonel Roche, F-31077 Toulouse Cedex 4, France
$^3$ Université de Toulouse ; UPS, INSA, INP, ISAE ; UT1, UTM, LAAS ; F-31077, Toulouse, France

Abstract

In this paper, we propose a hybrid object tracking approach based on region intensities and the motion vector of the interest points of the target object. After the heterogeneous region is identified (i.e foreground or background), the information extracted from the region intensities is computed through a pre-processing selection technique using local and global statistics. The motion vector of the interest points is obtained by computing the mean displacement of these points between consecutive frames. Therefore, for each received frame, the displacement vector is applied to the initial active contour that will be evolved using the region information. The main advantages of our approach which combines the information gathered from region intensities and motion of the interest points are first, a priori knowledge of the heterogeneous region is not required and second, the initial active contour in each frame is adjusted much closer to the true boundary of the object of interest. Experiments with synthetic and real-world images validate the efficiency of the proposed hybrid approach.

Keywords: Active contour, Local and global statistics, Interest Points, Object segmentation and tracking
1 Introduction

Tracking moving objects in video sequences has drawn increasing interest in recent years due to its wide variety of applications, such as robotics, video surveillance and scene monitoring. Generally, most approaches that allow tracking objects aim to identify the object of interest (i.e. Region Of Interest (ROI)) using background subtraction technique to identify moving objects in the video sequences [31-28]. An alternative approach consists to find the transformation of the object from frame to frame which is modeled by surrounding the object with a rectangle (or ellipse). The authors in [6] have proposed an approach toward target representation and localization of non-rigid objects. The histogram based target representations are regularized by spatial masking with an isotropic kernel while the target localization is formulated by using the basin of attraction of the local maxima optimized by using the mean shift approach. Additionally, Jepson and al. have proposed in [13] an approach for learning appearance models for motion-based object tracking. This approach uses an online EM-algorithm consisting of an Expectation step followed by a Maximization step as developed in [8]. Even if accurate tracking performances are achieved, these approaches allow tracking only the centroid or the orientation of the object and not the entire object.

Active Contour (AC) technique, which was primarily introduced by Kass and al [14] and has been widely used in image segmentation [1, 4, 7, 5, 14, 17, 19, 26, 33, 34] can be used to track the entire object region. This technique aims to segment a ROI by iteratively deforming the curve (i.e. AC) until it reaches the ROI boundary by minimizing certain energy functional. Existing AC approaches can be generally classified into two categories: edge-based techniques [4, 14] and region-based techniques [1, 7, 5, 17, 33, 34]. Edge-based AC techniques look for intensity discontinuities in the image and use the image gradients to guide the evolution of the contour. They assume that the ROI boundary can be detected in certain areas presenting rapid intensity changes and consider only the information located close to the evolving contour rather than the information located inside and outside it. Their main drawbacks consist on their sensitivity to image noise and initial contour location, due to the highly localized image information. Region-based AC techniques aim to identify the desired regions by using a certain region descriptor to guide the motion of the AC. The region-based technique developed by Zhu and Yuille in [34] searches for uniformity within the ROI and the background based on the user’s perception of the image content (e.g. intensity, color, or texture). Generally, the region-based techniques outperform the edge-based methods for their robustness to image noise and their low sensitivity to the initial contour location. However, they represent some limitations due to the assumption on the image intensities to be statistically homogeneous in the ROI as in the background. In fact, intensity heterogeneity often occurs in many real-world images due to different modalities; therefore it will be difficult to maintain the global constraint on the image data. To handle intensity heterogeneity problems,
Lankton and Tannenbaum [17] have considered local instead of global image statistics where the AC is evolved based on local information. This local region-based approach outperforms the standard global methods by its capability of segmenting heterogeneous objects. However, for conventional AC methods, neither the global region-based methods [7, 19, 33] nor their localizing versions [17] resolve completely the problems encountered in the object segmentation, mainly in the case of noisy image suffering from heterogeneous characteristics while using inadequate contour initialization. Generally speaking, global region-based methods are robust to contour initialization and image noise, but are more sensitive to image heterogeneity. In contrast, the local region-based method is robust against the heterogeneity by allowing foreground and background to be described in terms of local regions. However, it is more sensitive to contour initialization and image noise and hence can lead to segmentation errors.

A combination of local and global information was addressed in [25, 27]. Paragios and Deriche have proposed in [25] a combined energy minimization based on edge-based and region-based energies. Likewise, Sum and Cheung have proposed in [27] a combined energy minimization based on the sum of a global region-based and energy based local image contrast. In order to overcome the limitations of the global region-based and the local region-based approaches, namely, the presence of the heterogeneity (in ROI or in background) in noisy image and the use of an inadequate contour initialization, we developed in [1] an approach that combines the local region-based and the global region-based image statistics. Although, our approach in [1] has been demonstrated to be robust against the above mentioned limitations, it requires prior knowledge of the heterogeneous region (ROI or background) in order to identify the region from where the local extraction of image statistics is carried out. Hence, a pre-processing identification of the heterogeneous region (foreground or background) is needed to choose the segmentation technique.

In this paper a novel hybrid region tracking scheme that jointly exploits region-based information and interest point motion information to track the target objects by the AC is proposed. A pre-processing identification of the heterogeneous region is employed to select one of the two segmentation techniques represented in [1]. This selection is based on computing the variance intensities of the foreground and the background regions and the local extraction of image statistics is carried out on the region providing high intensities variance. Introducing the displacement vector of the interest points in our region-based AC approach is motivated by problems encountered in tracking object undergoing a significant displacement between consecutive received frames or by the presence of a similarity in mean intensities between the object of interest and the image background. We chose the well-known SIFT (scale-invariant feature transform) descriptor [16] as a candidate for our object recognition between consecutive frames owing to its high performance and suitability for video vision application with a slow acquisition video frames. Hence, using our proposed hybrid method, an automatic identification of the heterogeneous region is carried out primarily to determine the region-based AC technique used to segment the ROI. Later, the displacement
vector of the interest points generated from SIFT descriptor is calculated between consecutive frames. For each received frame in the video sequence, the displacement vector is recomputed and applied to the initial AC in the purpose of adjusting much closer the contour to the real object of interest boundary.

This paper is organized as follows. The next section reviews the methods of region-based AC and the object tracking. Section 3 describes the proposed hybrid approach for object tracking. Section 4 shows the experimental results and Section 5 concludes the paper.

2 Background

2.1 Active contour method

Let I denote a given image defined on the domain Ω and I(x) the intensity of the pixel x where x ∈ Ω. One approach to implement the AC segmentation consists in using the Level-Set method which considers the original curve as the zero level of a surface [23, 24]. The distortion of the entire surface induces a deformation of the curve shape. This process stimulates the evolution of the AC and achieves, at the end, the object segmentation. Let C be a closed contour represented as the zero level set of a signed distance function Φ, i.e., C={x|Φ(x)=0}. The purpose of this process is to implicitly evolve the contour C such that, at the convergence, the distance functions Φ < 0 (inside of C) and Φ > 0 (outside of C) represent the ROI and the background, respectively. In Level-Set formulation, a smoothed Heaviside function H(Φ(x)) is used to determine the inside and the outside of C. The following approximation of the smoothed Heaviside function specifies the interior of C:

$$H(Φ(x)) = \begin{cases} 
1, & Φ(x) < -\varepsilon \\
0, & Φ(x) > \varepsilon \\
1/2 \left( 1 + \frac{Φ(x)}{\varepsilon} + \frac{1}{\pi} \sin \left( \frac{π \cdot Φ(x)}{\varepsilon} \right) \right), & \text{otherwise} 
\end{cases}$$

Similarly, the exterior of C is specified as \((1 - H(Φ(x)))\).

The energy functional is only computed in a narrow band around C as presented in Adalsteinsson and Sethian work [2] in order to decrease the computational complexity of the standard Level-Set method. To specify this area around the curve, a smoothed version of the Dirac delta (2) which presents the derivative of the Heaviside function is used:

$$δΦ(x) = \begin{cases} 
1, & Φ(x) = 0 \\
0, & |Φ(x)| < \varepsilon \\
\frac{1}{2 \cdot \varepsilon} \left( 1 + \cos \left( \frac{π \cdot Φ(x)}{\varepsilon} \right) \right), & \text{otherwise} 
\end{cases}$$
Hybrid region and interest points-based active contour

To perform AC segmentation, we initially define an objective that determines what we want to extract from the image, and then we develop an energy criterion that should be minimized to achieve this objective. The region-based energy is generally expressed by an integral domain of a region descriptor (energy criterion). The choice of a relevant energy expression is a significant aspect that has been widely addressed by several authors as presented in [7, 19, 33]. In this paper, we employ two region-based energies presented in [7] and [33] where the ROI is represented by its mean intensities. The first descriptor used in this paper, introduced by Chan and Vese [7], models the image foreground and background as constant intensities that are represented by their means. It is given by:

\[ k(x, \Omega) = (I(x) - \mu_{in}(\Omega_{in}))^2 + (I(x) - \mu_{out}(\Omega_{out}))^2. \]  

where \( \Omega_{in} \) and \( \Omega_{out} \) denote respectively the inside and the outside of the contour \( C \), and \( \mu_{in} \) and \( \mu_{out} \) denote the mean intensities, respectively inside and outside of \( C \).

The second descriptor used in this paper is introduced by Yezzi and al. [33] and expressed as:

\[ k(x, \Omega) = -\frac{1}{2} \left( \mu_{in}(\Omega_{in}) - \mu_{out}(\Omega_{out}) \right)^2. \]  

This descriptor assumes that foreground and background regions should have maximal separate mean intensities.

Therefore, according to the chosen descriptor, the energy functional in the Level-Set formulation takes the form:

\[ E(\Phi) = \int_{\Omega} \delta \Phi(x) \cdot k(x, \Omega) \, dx. \]  

The evolution equation of \( C \) is expressed by:

\[ \frac{\partial \Phi}{\partial t}(x) = \delta \Phi(x) \cdot \nabla k(x, \Omega) + \lambda \cdot \delta \Phi(x) \cdot \text{div} \left( \frac{\nabla \Phi(x)}{|\nabla \Phi(x)|} \right), \]  

where \( \nabla \) and \( \text{div} \) represent respectively the gradient and the divergence operators. The second term in (6) is used to keep the curve \( C \) smooth and is weighted by a small positive coefficient \( \lambda \). Our models are elaborated based upon techniques of curve evolution, statistical function, and Level-Set method.

2.2 Region-based active contour

The global region-based approaches present certain robustness against contour initialization and image noise. In fact, they use all information within an image where both image foreground and background are examined simultaneously as
illustrated in Fig. 1. In that figure, the object of interest is shown in black and the area examined is the entire interior and exterior regions of the AC (red curve). The yellow point presents a point on the AC. Generally, the techniques that use global statistics to model regions are very sensitive to heterogeneous characteristics and may produce segmentation errors. To address this issue, Lankton and Tannenbaum [17] have developed a segmentation technique based on information collected from local interior and exterior regions along the AC rather than global regions in order to overcome the assumption that the foreground and background regions should be separated based on their global statistics. Consequently, the only assumption considered is that the interior and exterior regions are locally different. The choice for the local image statistics inside and outside the AC is based on defining a ball function $B$ to mask local regions by overlapping $B$ with inside and outside the AC [17]. This ball function centered at $x$, where $x$ corresponds here to a point on the AC, is expressed as:

$$B(x, p) = \begin{cases} 1, & ||x - p|| < \text{rad} \\ 0, & \text{otherwise} \end{cases}$$ \(7\)

where $\text{rad}$ is the ball radius and $p$ is a point in $\Omega$ that assigns to $B$ the value 1 on the local region and 0 elsewhere. Thus, as illustrated in Fig. 2, the local selected area is the intersection of the interior of the disk (the green dotted curve) with the interior and the exterior of the AC.

For information extraction, reducing the selection from global regions to solely local regions leads to place constraints only into selected local regions rather than the whole ROI and the background. For instance, if such a constraint of homogeneity should be verified, then only those of local regions will be considered to validate this constraint and guide the AC evolution. Consequently fewer assumptions are required on the image. We examined here the homogeneity constraint simply as an example of assumptions that the image must verify. For this aim, we employ the energy criteria using the equations (3) and (4) that are seeking to segment homogenous regions by computing the average intensities of the selected regions inside and outside the AC. Other forms of constraints can also be investigated on the image verifying for example a certain variance of intensities within a region or imposing a certain distribution of these intensities,
etc. However, the approach presented in [17] is still sensitive to the AC initialization and noise. This sensitivity is due to the fact that the information is extracted locally. Thus, with the aim to enable segmenting the object in the case of presence of heterogeneity and using an inadequate AC initialization in noisy image, we proposed an approach in [1] that combines both benefits of local-based and global-based techniques using local-based selection in one side of the AC and global-based selection in its other side at each point along the contour. Indeed, depending on where the heterogeneity appears (in ROI or in background), two cases have been investigated in our previous work [1] as shown in Fig.3 and Fig.4. The approach in [1] used in this paper is especially applied for images presenting heterogeneity either on the object of interest or on the background. Depending on which region is heterogeneous, we use one of the two techniques shown in Fig.3 and Fig.4 where the local selection is illustrated by the half disk (green dotted arc). If the background is heterogeneous compared to the object of interest, the technique, referred to as Global IN–Local OUT based information [1], is employed and which consists to extract locally the image statistics outside the AC and globally inside it (Fig.3). If the object of interest is heterogeneous compared to the background, we use the technique, referred to as Local IN–Global OUT based information [1] that extracts the image statistics locally inside the AC and globally outside it (Fig.4).

Unlike the standard global region-based approaches or the local region-based approach, the approach presented in [1] shows more robustness both against noise and contour initialization thanks to the global statistical information and against heterogeneous appearances thanks to the use of local statistical information.

### 2.3 Object tracking

The object tracking methods allow assessing over time the parameters of a target object present in the field of view of the camera and initially detected in the first frame of the video sequence. These parameters can be the position of the target object in the image, its shape and apparent orientation to name few. An automatic tracking method must not only follow the target object but also make an automatic initialization by a detection method regardless of any occlusions or disappearance of the target object from the field of the camera. Tracking methods require an object detection mechanism either in every frame or during the object
first appearance in the video. A good review of the major existent tracking methods can be found in Yilmaz and al. work [32]. Authors in this work [32] have divided the tracking methods into three main categories: methods establishing point correspondence, methods using primitive geometric models, and methods using contour evolution. Recently, a considerable amount of research has been devoted to visual tracking for a variety of applications [18, 3, 29, 10]. Authors in [18] have developed a spatio-temporal segmentation technique by detecting the moving objects in video image sequences; while the authors in [3] have developed an absolute differentiating and threshold detect motion regions technique by tracking individual objects through the segmented region. Authors in [29] have developed a real-time human tracking system in video sequences acquired by a stationary camera. Authors in [10] have developed an object tracking tool suitable for MPEG-2 sequences.

In this paper, we focus on object segmentation and tracking by using the AC method based on combining information on both the region and the motion of the interest points of the target object. Points of interest have been long used in the context of many applications, like motion, stereovision, scene monitoring and tracking problems. A desirable quality of an interest point is its invariance to changes in illumination and camera viewpoint. For all of these applications, the extracted points usually represent sites where the information is considered as perceptually relevant.

On the other hand, point detectors are used to find interest points in images which have an expressive texture in their respective localities. The point detector should be able to repeat the extracted points from an image to another, whatever the involved transformations. The most popular one is the point extractor presented in [12]. A performance evaluation of interest point descriptors such as SIFT descriptor for object recognition [16], steerable filters [9], differential invariants [15], complex filters [30], moment invariants [11] and cross-correlation for different types of interest points [12, 16, 20, 21] are presented in [22] where the local descriptor SIFT proposed by Lowe [16] is shown to provide the best performance. That is the main reason why the SIFT is chosen in this work as descriptor of interest points.

3 Hybrid proposed approach

Let \( F_f \) denote a given video sequence of \( N \) frames where \( f = \{0, 1, 2, 3, \ldots N-1\} \). The AC on the first frame \( F_0 \) is manually initialized with an ellipse, surrounding the ROI (i.e. the target object), and is evolved to the boundary of the ROI based on the region-based technique automatically chosen as will be described below. The object segmentation in each frame \( F_f \) is realized by using the segmentation technique chosen in the first frame \( F_0 \). Then, the object tracking process by the AC using our proposed hybrid approach is achieved by following two steps. First by moving the initial AC within each received frame \( F_i \) \( \{i=1, 2, 3 \ldots N-1\} \) by the displacement vector, of the object’s interest points, calculated between the consecutive frames \( F_{i-1} \) and \( F_i \). Second step, the AC is evolved in the frame \( F_i \).
according to the region-based technique chosen in the first frame $F_0$. The initial AC, before displacement, in each frame $F_i$ is defined by the AC resulting at the convergence state in the frame $F_{i-1}$. Let our Hybrid Region and Interest Points-based Active Contour approach referred to as HRIP-AC for short.

### 3.1 Automatic selection for the region-based active contour technique

The object segmentation used in this work is based on the approach presented in [1] which is applied especially for images presenting heterogeneity either in the ROI or in the background. Local statistics are extracted from the heterogeneous region (ROI or background) and global statistics are extracted from the other region. Thus, a priori knowledge of which region (foreground or background) is heterogeneous, is required, which gives this approach a certain limitation. To address this drawback, we propose in this paper as a pre-processing step, a technique that allows object segmentation without requiring a prior knowledge about the heterogeneity place in the image. In fact, once the initial AC is defined in the image; the histograms of the pixels intensities are extracted and the variance of these intensities is calculated for the foreground and the background regions. Let $Var_{in}$ and $Var_{out}$ be the variances of the intensities, respectively, within the foreground and the background regions. Let $Tech$ be the variable that determines whether the chosen technique will be the *Global IN-Local OUT* technique or the *Local IN-Global OUT* technique [1], and expressed as:

$$Tech = \begin{cases} -1, & Var_{in}(\Omega_{in}) \leq Var_{out}(\Omega_{out}) \\ 1, & Var_{in}(\Omega_{in}) > Var_{out}(\Omega_{out}) \end{cases}.$$  

(8)

The heterogeneous region will have the higher value of intensities variance, thus, an automatic selection between the *Global IN-Local OUT* and the *Local IN-Global OUT* techniques is applied for object segmentation according to which value is assigned to the variable $Tech$ (e.g. -1 or 1, respectively).

### 3.2 The information on the motion of object’s interest points

As mentioned above the local descriptor SIFT [16] is used in our approach owing to its robustness against large transformations. This local descriptor is used in the aim to evaluate the motion of the object’s interest points in the AC method even if this object undergoes a large displacement between consecutive frames.

Fig.5 illustrates the principle of the interest points matching between consecutive frames $F_{i-1}$ and $F_i$. The ROI (i.e. object of interest) is shown in black where the interest points belonging to the ROI are depicted in blue and those belonging to the background are depicted in gray. The green lines present the matching of recognized interest points between the two frames $F_{i-1}$ and $F_i$. The red curve, in the frame $F_{i-1}$, shows the AC at the convergence state as surrounding the ROI boundaries.
Let \((x^p_{F_{i-1}(ROI)}, y^p_{F_{i-1}(ROI)})\) be the coordinates of the interest points of the ROI in the frame \(F_{i-1}\) and let \((x^p_{F_i(ROI)}, y^p_{F_i(ROI)})\) be the coordinates of the interest points of the ROI in the frame \(F_i\). The displacement vector of the ROI from the frame \(F_{i-1}\) to the frame \(F_i\) can be approximated by the mean displacement vector of the ROI’s interest points between these consecutives frames. Let \(\text{Disp}_{F_{i-1} \rightarrow F_i}(ROI)\) be this displacement vector of the ROI between \(F_{i-1}\) and \(F_i\), and expressed as:

\[
\text{Disp}_{F_{i-1} \rightarrow F_i}(ROI) = \left( \mu \left( x^p_{F_i(ROI)} - x^p_{F_{i-1}(ROI)} \right), \mu \left( y^p_{F_i(ROI)} - y^p_{F_{i-1}(ROI)} \right) \right)
\]

where \(\mu\) represents the mean operator.

Once the vector \(\text{Disp}_{F_{i-1} \rightarrow F_i}(ROI)\) is calculated, it is applied to each point along the initial AC in the frame \(F_i\) and the AC is evolved to the ROI boundary according to the region-based technique selected in the first frame \(F_0\) as described in Section 3.1.

![Fig.5: Matching interest points of the object (black shape) between consecutives frames \(F_{i-1}\) and \(F_i\). The interest points are depicted in blue if they belong to the object and in gray if not.](image)

### 3.3 Algorithm of the proposed hybrid approach for object tracking

As mentioned above, the proposed HRIP-AC approach has two main steps between each consecutives frames. The following algorithm describes the different steps of the object tracking process in a sequence of \(N\) frames using the proposed hybrid approach.
**Algorithm steps of the proposed HRIP-AC approach for object tracking**

1. Initialization: Manually initialize the AC surrounding the ROI in the first frame $F_0$.
2. Automatically define the segmentation technique that will be used for ROI segmentation (*as described in the Section 3.1*).
3. Evolve the AC in the first frame $F_0$ according to the technique, *selected in step 2 of this algorithm*, until it reaches the ROI boundaries.
4. Extract the interest points of the ROI in the frame $F_0$.
5. for $i = 1$ to N-1 do
   6. Make the final AC in frame $F_{i-1}$ as the initial AC in frame $F_i$.
   7. Extract the interest points in the whole frame $F_i$.
   8. Match the interest points between those of the ROI in frame $F_{i-1}$ with those of the whole frame $F_i$ to identify the ROI’s interest points in frame $F_i$.
   9. Compute the displacement vector $\text{Disp}_{F_{i-1} \rightarrow F_i}(\text{ROI})$ of the ROI between the consecutives frames $F_{i-1}$ and $F_i$ (*as described in the Section 3.2*).
10. Apply the displacement vector $\text{Disp}_{F_{i-1} \rightarrow F_i}(\text{ROI})$ to the initial AC (*obtained in step 6 of this algorithm*) within the frame $F_i$.
11. Evolve the AC to the ROI boundary in frame $F_i$ based on region information according to the selected segmentation technique (*technique defined in step 2 of this algorithm*).
12. Extract the interest points corresponding only to the ROI in frame $F_i$.
13. end for

### 4 Experimental results

In this section, we present the results obtained in synthetic and real images to assess the performance of the proposed method in static images and video streams used by other researchers.

#### 4.1 Object segmentation in static images

Fig.6 illustrates the results obtained in case of a homogeneous object in a heterogeneous background in synthetic and real images as shown in the first and third column, respectively. The second column shows the result in case of a heterogeneous object in a homogeneous background in a real image. The AC initialization for each image is shown in Figs.6 (a-c) by the red curve. Figs.6 (d-f) show the foreground image (area inside the initial AC) where the excluded area (outside the initial AC) is shown in black in Figs.6 (d-e) and in white in Fig.6 (f). In Contrast, Figs.6 (j-l) show the background image (area outside the initial AC) where the excluded area (inside the initial AC) is shown in Figs.6 (j-k) by a black disk and in Fig.6 (l) by a white ellipse. Figs.6 (g-i) and Figs.6 (m-o) show respectively, the foreground and the background histograms for each above corresponding image (the histograms represent the distribution of the pixels...
intensities for the area included in each case). As shown for the synthetic image (first column), the intensity distribution in the background (Fig. 6 (m)) shows more variance compared to the intensity distribution in the foreground (Fig. 6 (g)) which indicates that the image background is more heterogeneous than its foreground. By comparing automatically the variance of the foreground and the background regions, the region-based AC technique was automatically selected to be the *Global IN-Local OUT* technique [1] where the object segmentation result obtained is illustrated in Fig. 6 (p). For the real image in the second column, the intensity distribution in this case of the foreground (Fig. 6 (h)) shows more variance compared to the intensity distribution in the background (Fig. 6 (n)) indicating that the image foreground is more heterogeneous compared to its background. By comparing automatically the variance of the foreground and the background intensities, the selected region-based AC technique was defined in this case by the *Local IN-Global OUT* technique [1] where the object segmentation result obtained is illustrated in Fig. 6 (q). As shown for the second real image (last column), the intensity distribution in the background (Fig. 6 (o)) shows more variance compared to the intensities distribution in the foreground (Fig. 6 (i)). The region-based AC technique selected in this case was the *Global IN-Local OUT* technique [1] since the background shows more heterogeneity compared to the foreground. The object segmentation result obtained is illustrated in Fig. 6 (r).
Fig. 6: Object segmentation based on an automatic selection of the region-based AC technique.
4.2 Object tracking in video sequences

In order to assess the performance of the proposed hybrid HRIP-AC approach; experiments have been conducted on the “CAVIAR sequences” used by many researchers in the field. Fig.7 and Fig.8 illustrate the tracking results, of a mobile non-rigid object, obtained in a first video sequence where the target object is moving near some other mobile objects in the scene. The second video sequence used in our experiments is presented in Fig.9 and Fig.10 where the target mobile non-rigid object in this sequence shows some similarity with the background region.

4.2.1 Case of presence of other mobile objects in the scene

Fig.7 illustrates the object tracking result obtained by using only the region-based information, while Fig.8 illustrates the object tracking result obtained by using our HRIP-AC approach. The region-based segmentation technique is automatically selected in the first frame (for both Fig.7 and Fig.8) as described in the Section 3.1 while the object segmentation is performed, within this first frame, using the region-based information.

a. Tracking results by using only the region-based information

The first frame in Fig.7 illustrates the segmentation result of the object of interest (man walking in the crowd in the supermarket) obtained by the Global IN-Local OUT technique which was automatically chosen in this first frame. The tracking process in the whole sequence is then performed using only the region-based information with the related selected technique to guide the evolution of the AC to the object boundaries at each frame. The AC, in this case, was not able to correctly track the silhouette of the target object and it was trapped by undesired boundaries while including in the segmentation some part belonging to other objects in the scene as shown in Fig.7 (starting from Frame 25) or including some part of the floor (for the latest Frames). This is due to the fact that the information extracted only from the region doesn’t take into account the motion that the target object has experienced between frames. Consequently, some parts verifying the region-based criterion can attract the AC even if these parts do not belong to the target object which results in AC that diverts from the real object of interest boundaries.
Fig.7: Object tracking in case of presence of other mobile objects in the scene using only the region-based information

b. Tracking results by using the hybrid region and interest points-based information

By using our HRIP-AC approach relying on both the region and the motion of object’s interest points, the tracking process of the target object had improved substantially as shown in Fig.8 compared to the case of relying on information extracted only from the region. The comparison of the results in Figs.7 and 8 proves that considering the temporal information related to the displacement of the object between frames came to compensate the region information by displacing the AC in each received frame by the same displacement vector that the object of interest has experienced. In fact, by applying the displacement vector to the initial AC in each received frame, the AC is brought closer to the real object of interest boundaries and hence the AC will not be trapped by false boundaries belonging to neighboring moving objects. This is proven by the tracking results obtained in Fig.8 where the AC was able to track the silhouette of the object of interest more accurately than the result obtained in Fig.7.
4.2.2 Case of similarity between the object and the background regions

Fig. 9 illustrates the object (a man walking in supermarket) tracking result obtained by using only the region-based information; while Fig.10 illustrates the tracking result of this mobile object obtained by using our hybrid HRIP-AC approach. The region-based segmentation technique automatically selected in the first frame (for both Fig.9 and Fig.10) was the *Global IN- Local OUT* technique which is obviously expected since the background shows more heterogeneous characteristics compared to the object of interest. The object segmentation is realized in the first frame of the video sequence by using the region-based information.

**a. Tracking results by using only the region-based information**

Fig. 9 illustrates the tracking process of a mobile object in this video sequence. The object tracking in the whole sequence is performed by using only the region-based information to evolve the AC within each frame. Using only the region information, the AC was not able to accurately track the silhouette of the target object and it was trapped by undesired boundaries belonging to the image background (starting from Frame 50). In this case, the information used to track
the object silhouette discards the way in which the object moves from one frame to another. Thus, some parts within the image background (the iron fence) trapped the AC altering the latter to include object not belonging to the object of interest region. As a result, the AC is distorted by these undesired parts along the video sequence even if the object of interest moves far away from these zones since no information on the object motion is taken into account.

Fig.9: Object tracking in case of similarity between the object and the background regions using only the region-based information

b. Tracking results by using the hybrid region and interest points-based information

Fig.10 illustrates the tracking process of the object of interest by using the proposed hybrid HRIP-AC approach that relies on information extracted from both the region and the motion of interest points. By comparing the object tracking results in Figs.9 and 10, it is clear that the tracking process shows a remarkable improvement when the displacement of the object of interest through frames has been taken into account (Fig.10). From frame 50 in Fig.9, we can see the AC starts to be attracted by the iron fence for the reason of the similarity appearing between the iron fence and the object of interest regions. This AC attraction continues to grow and leads the AC to deviate from the object boundaries with the progress of frames when only the information on the region intensities is considered. In contrast, as shown in Fig. 10, when the object motion is known and the displacement vector of the object is applied to the initial AC
within each frame, the entire object silhouette is reached and tracked accurately by the AC along the video sequence without confusion with other zones belonging to the background image.

Fig. 10: Object tracking in case of similarity between the object and the background regions using the HRIP-AC approach

5 Conclusion

In this paper, a new hybrid approach using region and interest points-based AC for object segmentation and tracking has been proposed. This approach relies on local and global statistics to segment the object where an automatic identification of the heterogeneous region is carried out primarily to determine the region-based AC segmentation technique. This hybrid approach operates the displacement vector of the object’s interest points in the tracking process where the initial AC at each frame is adjusted much closer to the real object of interest boundary. By combining the information gathered from both region and motion of interest points in the AC approach, a good object tracking results are achieved. We have tested the proposed method on synthetic and real image database presenting heterogeneous characteristics and obtained promising results for object segmentation and tracking.
References


Received: August 31, 2013