Accurate moving object segmentation in unconstraint videos based on robust seed pixels selection

Wenlong Zhang¹,², Xiaoliang Sun¹,² and Qifeng Yu¹,²

Abstract
Due to the clutter background motion, accurate moving object segmentation in unconstrained videos remains a significant open problem, especially for the slow-moving object. This article proposes an accurate moving object segmentation method based on robust seed selection. The seed pixels of the object and background are selected robustly by using the optical flow cues. Firstly, this article detects the moving object’s rough contour according to the local difference in the weighted orientation cues of the optical flow. Then, the detected rough contour is used to guide the object and the background seed pixel selection. The object seed pixels in the previous frame are propagated to the current frame according to the optical flow to improve the robustness of the seed selection. Finally, we adopt the random walker algorithm to segment the moving object accurately according to the selected seed pixels. Experiments on publicly available data sets indicate that the proposed method shows excellent performance in segmenting moving objects accurately in unconstraint videos.

Keywords
Moving object segmentation, local difference, seed pixels selection, random walker

Date received: 11 February 2020; accepted: 13 July 2020

Introduction
Moving object segmentation has been widely used in monitoring systems, unmanned aerial vehicles, autonomous vehicles, and so on. Accurate segmentation of moving objects, consisting of precise position and shape information, is of great significance for subsequent object tracking and recognition. However, due to the motion of the clutter background, existing moving object detection algorithms generally obtained a rough segmentation of the moving object (see Figure 1(c)). The optical flow between adjacent frames was used by most motion segmentation algorithms to segment the moving object from the background. However, these algorithms will fail in segmenting slow-moving objects, in which case the differences between the optical flow corresponding to the moving object and the background are subtle (see Figure 1(d)). The algorithms adopted a long-term analysis of video to tackle the problem (see Figure 1(e)). However, they are not suitable for the online process because of using the frames after the current moment.

¹ College of Aerospace Science and Engineering, National University of Defense Technology, Changsha, China
² Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha, China

Corresponding author:
Xiaoliang Sun, College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China; Hunan Provincial Key Laboratory of Image Measurement and Vision Navigation, Changsha 410073, China.
Email: alexander_sxl@nudt.edu.cn
The orientation cues of the optical flow are robust to the depth of the scene. This article selected the object and background seed pixels according to the local difference in the optical flow orientation cues. The temporal consistency is used to solve the problem of the slow-moving objects. Firstly, the proposed method detected the moving object’s rough contour according to the local difference of orientation cues of the optical flow between adjacent frames. Different from the optical flow cues used in the existing algorithms, this article adopts the weighted orientation cues. The weight is determined adaptively based on the local magnitude cues of the optical flow. Then, the detected rough contour is used to guide the object and the background seed selection via the point-in-polygon (PIP). To further improve the robustness of the object seed selection, the object’s seed pixels in the previous frame are propagated to the current frame according to the optical flow. Lastly, this article used the random walk algorithm to segment the moving object accurately according to the selected seeds. The proposed method only uses the information of the current and previous frames for the accurate moving object segmentation. Thus, the proposed method can achieve online accurate moving object segmentation. We evaluate the proposed method on publicly available data sets and perform comparisons with typical algorithms. Experiment results indicate that the proposed method performs better than existing algorithms in segmenting the moving object in unconstraint videos.

This article’s main contributions are twofold: (1) The proposed method uses the weighted local difference in orientation of the optical flow to detect the rough contour of the moving object. The detected rough contour is used to select the object and background seed pixels for accurate segmentation by the random walker algorithm; (2) the object seed pixels in the previous frame are propagated to the current frame for improving the robustness of seed selection.

**Related works**

Existing moving object segmentation in unconstraint video methods can be classified into two major categories: background-modeling-based methods and motion-segmentation-based methods. The background-modeling-based methods segment the moving object by subtracting the background component from the input frame. The background is often modeled as a codebook, Gaussian mixture model, and so on. Such methods have difficulties in dealing with quick-changing or clutter backgrounds. They are more suitable for constraint monitoring systems, in which the background is fixed. The motion-segmentation-based methods use the differences between the motion of the object and the background to segment the moving object. The motion information is usually obtained by estimating the optical flow of the image sequence. They are widely used for moving object segmentation in unconstraint videos. This article focuses on the latter and summarizes motion-segmentation-based methods as follows.

**Moving object segmentation based on sparse optical flow**

Such methods estimate the motion between adjacent frames based on the sparse optical flow. Estimating the sparse optical flow of feature points between adjacent frames is efficient. However, they can only distinguish the object from the background sparsely. Thus, post-processing is needed for a dense result. The algorithms defined the moving objects as groups of pixels that are salient in motion and color, which were segmented directly through clustering by mean-shift. The center of clustering for the tracking pixels was computed by the feature of position, color, and optical flow calculated by Lucas and Kanade. Then, the markov random fields (MRF) was used to segment the moving object densely based on the clustering centers. Kim et al. found that the background feature points are scattered more widely than that of the moving object, so the scatteredness of clustered optical flow vectors was used to classify the object and background. Nonaka et al. used spatial closeness, amplitude, and direction similarity to cluster the sparse optical flow. The cluster results were classified into different labels according to their shape and size. Finally, the graph cut was used to obtain a dense segmentation based on the sparse label. The algorithms proposed by Malik and Brox, Ochs et al., and Sheikh et al. used the long trajectories of the feature points to segment moving objects to enhance the robustness of the segmentation algorithm. The basis vectors of
trajectories were extracted based on the rank constraint of the background motion trajectory matrix in the study of Sheikh et al.\textsuperscript{17} Then, the trajectories belonging to the moving object were segmented by the difference with the trajectories constructed by basis vector. In the studies of Malik and Brox\textsuperscript{5} and Ochs et al.,\textsuperscript{6} the spectral clustering was used to segment the moving object based on the affinities between long trajectories. Then, the information in the spatial-temporal domain was used to obtain dense segmented object based on the sparse segmented trajectories.

**Moving object segmentation based on dense optical flow**

Motion segmentation based on sparse optical flow cannot obtain the complete object directly. Thus, some algorithms\textsuperscript{2,7–9,18} turned to dense optical flow for segmenting moving objects. Huang et al.\textsuperscript{2} modeled the background regions in the optical flow field. They computed the homograph matrix based on the dense optical flow estimated by Ilg et al.,\textsuperscript{19} which was used to construct a new optical flow of background. Then, the moving objects were detected by comparing the original optical flow with the constructed optical flow. A context-aware motion descriptor was designed by Chen and Lu\textsuperscript{18} based on the histogram of oriented optical flow to measure the inconsistency between the optical flow of the object and its surrounding background. Sajid et al.\textsuperscript{20} used a second-order function to approximate the background motion based on its low-rank characteristics. Then, they estimated the foreground probability by comparing the motion field with background motion approximation. Narayana et al.\textsuperscript{8} considered that the magnitude of the optic flow is depth-dependent, that the objects at different depths from the camera can exhibit different optical flow although they share the same real-world motion. Thus, they segmented moving objects in the orientation field of optical flow. The orientation fields of background were modeled by discrete sampling on different camera movements, which cannot cover all the motion of the camera. Bideau and Learned-Miller\textsuperscript{9} also used the orientation of optical flow to segment moving objects, where the optical flow caused by camera rotation was subtracted by a modified algorithm.\textsuperscript{21} Then, the Bayesian probability model was used to measure the information on how objects are moving differently from the background, where the segmented result in the previous frame served as the prior, and the difference of the orientation fields served as the likelihood. Papazoglou and Ferrari\textsuperscript{7} used the optical flows of long sequence to solve the problem of object motion slowing down. The difference in optical flow in the spatial domain was used to detect the coarse moving object. Then, the appearance model and MRF were used to refine the coarse result based on the information in the temporal and spatial domain. Wu et al.\textsuperscript{22} utilized the dense particle trajectories to segment moving objects coarsely by reduced singular value decomposition. They segmented fine objects according to the reconstructed background motion, which is obtained through image inpainting based on the coarse segmentation. Zhu and Elgammal\textsuperscript{23} formulated the problem as a multi-label segmentation problem by modeling moving objects and background in different layers. They assigned an independent processing layer to each moving object and background and estimated both motion and appearance models. Koh and Kim\textsuperscript{24} augmented the initial object regions, which are generated by using both color and motion edges, with missing parts and reduced them by excluding noisy parts.

A summary and comparison of these two types of moving object segmentation algorithms are shown in Table 1.

| Algorithm                          | Classification model                                      | Dense segmentation /post-processing                                      |
|------------------------------------|----------------------------------------------------------|-------------------------------------------------------------------------|
| Sparse-optical-flow-based methods  |                                                          |                                                          |
| Bugeau and Pérez\textsuperscript{3,4} | Mean-shift                                              | Graph cut                                                              |
| Malik and Brox\textsuperscript{5} and Ochs et al.\textsuperscript{6} | Long trajectory + spectral clustering                    | Potts model                                                            |
| Sheikh et al.\textsuperscript{17} | Long trajectory + rank constraint                        | Graph cut                                                              |
| Nonaka et al.\textsuperscript{16} | Cluster growing algorithm                                | Graph cut                                                              |
| Dense-optical-flow-based methods   |                                                          |                                                          |
| Huang et al.\textsuperscript{2}    | Homograph approximation                                  | Dual-mode judge mechanism + adaptive thresholding                      |
| Papazoglou and Ferrari\textsuperscript{7} | Motion boundaries + inside–outside maps                  | Spatial-temporal smoothness + graph cut                                |
| Narayana et al.\textsuperscript{8} | Orientation field + color + prior                        | Gradient descent                                                       |
| Chen and Lu\textsuperscript{18}    | Prior + translation angle likelihood                     | Maximum posterior                                                      |
| Sajid et al.\textsuperscript{20}   | Second-order approximation + appearance module           | Support vector machines                                                |
| Wu et al.\textsuperscript{22}      | Dense particle trajectories + reduced singular value decomposition + background motion reconstruction | Graph cut                                                              |

Table 1. Comparison of some moving object detection/segmentation methods.
The sparse-optical-flow-based methods firstly adopt a classification model to label the sparse points based on their optical flow and other characteristics. Then, they usually turn to the graph model for a dense segmentation based on the sparse seed points. The dense-optical-flow-based methods can distinguish all pixels directly based on different classification models according to the pixel’s features. Usually, the prior information is introduced to eliminate noise in the final segmentation, such as the temporal-spatial domain continuity.

The background optical flow is caused by the motion of camera platform, while object optical flow is caused by the motion itself plus the camera. There may be a large difference in the optical flow in different background regions. However, it has the property of local smoothness. In other words, the difference in the optical flow is small for two adjacent pixels in the background region for most camera motion types. When a moving object appears in the background, there will be a difference between the optical flow of objects and that of its surrounding background to some extent. So that the local difference of orientation and magnitude field in optical flow, calculated by algorithm, is used to detect the rough contours of the moving object in this article, which is used to select the seed pixels of the object and background. Then, we use the random walker algorithm to segment the moving object based on the spatial distribution continuity of the object and background. To further improve the robustness of the proposed algorithm, the temporal domain movement is utilized to propagate the seed pixels of the object in the previous frame to the current frame.

The flowchart of the proposed method is shown in Figure 2. There are three main steps in our method: the rough contour detection using the weighted local orientation cues of the optical flow, the robust seed pixels selection, and the moving object segmentation.

**Rough contours detection using the weighted local orientation difference of optical flow**

The distribution of the optical flow has an apparent discontinuity next to the contour of the moving object. Thus, the contour of the moving object can be detected by calculating the local difference of the optical flow. The background motion projected on the image depends on the distance to the camera, that is, the depth of the scene. The optical flow of background regions may be different, although they share the same real-world motion. Thus, the optical flow vector is not a robust cue for moving object segmentation, especially for the absence of clutter backgrounds. The orientation of the optical flow is independent of the depth of the scene in the case of camera translation. The change of orientation field in the background region is smooth in the spatial domain. To avoid the error detection caused by the clutter background, this article calculates the local difference in orientation and magnitude field of the optical flow, respectively.

The local orientation difference $OD_i$ of the optical flows for the pixel $p_j$ is defined as equation (1). This article calculates the cosine of orientation difference, and the minimum cosine value obtains the map of local orientation difference.

$$
OD_i = \min_j \left( \cos \left( \langle (u_i, v_i), (u_j, v_j) \rangle \right) \right) \\
= \min_j \left( \frac{u_i u_j + v_i v_j}{\sqrt{u_i^2 + v_i^2} \times \sqrt{u_j^2 + v_j^2}} \right) \quad p_j \in N_i
$$

**Metrology**

This article segments the moving object accurately in the unconstraint videos based on robust seed pixels selection.
where \((u_i, v_i)\) is the optical flow vector of the pixel \(p_i\), \(\angle\) denotes the angle between the two vectors, \(N_i\) is the neighborhoods of pixel \(p_i\) and set to \(3 \times 3\) in this study.

The local difference between the optical flow orientation cues may be small when the motion direction of the object and the background are similar. The orientation difference of the pixels near the optic axis will be dramatically significant when the camera moves along the optic axis, which often happens in autonomous driving. There may also be some changes in the orientation field within the object and the background region because of the inaccurate estimated optical flow. Therefore, this article uses the magnitude of optical flow to weight the orientation in detecting the contours of the moving object. The maximum of the magnitude difference between adjacent pixels is used as the weight as in equation (2)

\[
MD_i = \max_j \left( |u_j^2 + v_j^2 - (u_i^2 + v_i^2)| \right) \quad j \in N_i
\]  

(2)

As shown in equation (3), the weighted orientation difference (WOD) is calculated by multiplying OD with the normalized magnitude difference \(MD\)

\[
\text{WOD}_i = \text{OD}_i \times \bar{MD}_i
\]  

(3)

The maps of local differences in optical flow obtained by different ways are shown in Figure 3. We can see that there are some wrong boundaries inside the object region in the map of the local difference of orientation because of the inaccurate optical flow, such as Figure 3(b). There are also some wrong boundaries occurring in background regions in the map obtained by the magnitude difference. By weighting the orientation difference with that of amplitude, we can get a cleaner local difference map, as shown in Figure 3(d).

The rough contour of the moving object is detected from the WOD map via a two-step thresholding method. Firstly, we detect the definite contours through thresholding on WOD with a larger value \(T_{c1}\). Then, a smaller threshold \(T_{c2}\) is used for the pixels adjacent to the definite contours to get much more complete contours. The value of the threshold is determined by the mean \(m\) and standard deviation \(\sigma\) of WOD, as in equation (4)

\[
T_{c1} = m + 4 \cdot \sigma
\]

\[
T_{c2} = m + 2 \cdot \sigma
\]  

(4)

Robust seed pixels selection

Based on the rough contours detected in the “Rough contours detection using the weighted local orientation difference of optical flow” section, this article selects the seed pixels of the object and the background using the PIP.\(^{10}\) The pixels belonging to the moving object should be located inside the contours, while the pixels of the background should be located outside of the contours. The algorithm PIP\(^{10}\) determines whether a point is inside or outside the contours simply.

As shown in Figure 4(a), the PIP\(^{10}\) draws a ray from the current point to infinity. The point locates inside the contours if the number of intersections between the ray and contours is odd (as the red points in Figure 4(a)). If the number of intersections is even, the point locates outside of the contours (as the blue point in Figure 4(a)). To improve the robustness of the seed selection to the incomplete rough contour, this study emits rays from the current pixel to multiple directions, as shown in Figure 4(b). We count the number of rays with an odd number of intersections for each pixel. The number for the pixel \(p_i\) is donated as \(I_{n_i}\). Then, this study selects the seed pixels of the object and background according to the number \(I_{n_i}\). As shown in Figure 4(b), this study emits rays to...
eight directions. The probability of the red point inside the contours is \( \frac{2}{3} \) in Figure 3(b), and it is \( \frac{2}{8} \) for the blue point.

Let \( R \) represents the areas inside the bounding box of rough contours. The definite background pixels cannot locate inside the bounding box of the detected contours, and the definite object pixels cannot locate outside the bounding box. So, this study selects the seed pixels of background and object as

\[
S_t = \begin{cases} 
0 & p_i \in R \text{ and } I_{n_i} \geq T_1 \\
1 & p_i \notin R \text{ and } I_{n_i} \leq T_2 \\
\text{unknown} & \text{otherwise}
\end{cases}
\]  

(5)

where threshold \( T_1 \) is a larger constant to ensure the pixels are the definite object, and \( T_2 \) must be small enough for particular background pixels. They are set 7 and 1 in this study, respectively. The tuple \( S_t \) corresponds to the pixel \( p_i \) consisting of the probabilities that \( p_i \) belongs to the background and the moving object.

The slow moving object brings excellent difficulty for us to get seed pixels robustly by the incomplete rough contour. This study propagates the object seed pixels in the previous frame to the current frame through the optical flow to tackle this problem. The object seed pixel \( p_{j,t-1} \) in the previous frame is propagated to the current frame \( p_{j,t} \) by the optical flow vector \( (u_{j,t-1}, v_{j,t-1}) \). The corresponding object seed pixel is obtained by equation (6)

\[
\begin{align*}
  x_{j,t} &= x_{j,t-1} + u_{j,t-1} \\
  y_{j,t} &= y_{j,t-1} + v_{j,t-1}
\end{align*}
\]  

(6)

Further, \( p_{j,t} \) is validated as in equation (7). The \( p_{j,t} \) is an object seed pixel only if it locates inside the bounding box of the object contours and \( I_{n_{j,t}} \) is larger than a threshold \( T_3 \)

\[
S_{t,t} = \begin{cases} 
[10] & p_{j,t} \in R \text{ and } I_{n_{j,t}} \geq T_3 \\
\text{unknown} & \text{otherwise}
\end{cases}
\]  

(7)

where threshold \( T_3 \) is smaller than \( T_1 \) because of the temporal constraint and is set to 4 in this study.

The moving object segmentation

In this section, this study segments the moving object accurately through the random walker algorithm\(^{11}\) based on the seed pixels selected in the “Robust seed pixels selection” section.

Based on the seed pixels, the random walker algorithm segments the pixels according to the continuity of spatial intensity distribution. The random walker\(^{11}\) is defined on a graph model \( G = (V, E) \) consisting of nodes \( v \in V \) and edges \( e \in E \subseteq V \times V \). The edge \( e_{ij} \) connects two adjacent nodes \( V_i \) and \( V_j \) is assigned a weight \( w(e_{ij}) \) for measuring the similarity of \( V_i \) and \( V_j \). The graph is defined 4-links and undirected in this study. We define the weight as

\[
w(e_{ij}) = \exp\left(-\lambda \left( (R_i - R_j)^2 + (G_i - G_j)^2 + (B_i - B_j)^2 \right) \right)
\]  

(8)

where \( R_i, G_i, \) and \( B_i \) are the red, green, and blue intensity of the pixel \( P_i \). \( \lambda \) is a positive coefficient and is set to 90.\(^{11}\)

The degree of \( v_i \) is defined as the sum of weight for all edges \( e_{ij} \) connected to \( v_i \)

\[
d_i = \sum_j w(e_{ij})
\]  

(9)

The element \( l_{ij} \) of the combinatorial Laplacian matrix \( L \) can be calculated as follows

\[
l_{ij} = \begin{cases} 
d_i & i = j \\
-w(e_{ij}) & v_i \text{ and } v_j \text{ are adjacent nodes} \\
0 & \text{otherwise}
\end{cases}
\]  

(10)

The \( L \) is a symmetric matrix. We rearrange it and let the elements correspond to the seed pixels located in the top left part \( L_M \). The unknown pixels located in the bottom right part \( L_U \)

\[
L \rightarrow \begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix}
\]  

(11)

The probability of the unknown pixels belonging to the moving object and backgrounds can be calculated by solving the linear equation (12)

\[
L_U S_U = -B^T S_M
\]  

(12)

where \( S_M \) is the label tuple of the seed pixels, which can be obtained by equations (4) and (6). \( S_U \) is the probability vector for the unknown pixels. The mask \( M \) of the moving object can be obtained by comparing of the two elements in \( S_t \) as in equation (13)

\[
M_{i} = \begin{cases} 
1 & S_{i}(1) < S_{i}(2) \\
0 & \text{otherwise}
\end{cases}
\]  

(13)

Algorithm 1 summarizes the overall procedure of the proposed moving object segmentation method.

**Experiment**

**Experiment settings**

To evaluate the performance of the proposed method, this study conducts experiments on eight standard benchmark sequences from the Moseg_dataset.\(^5\) Besides, we compared our method with three existing typical methods.\(^1,2,7\) The results of the algorithm\(^1,2,7\) are all produced by the codes downloaded from the authors’ home page. The optical flow used in previous methods\(^2,7\) and the proposed method is calculated by using the same algorithm.\(^{25}\) The parameters in the compared algorithms are set as the authors recommend. Moreover, all the segmentation results of the proposed method are obtained with the same parameters.
**Algorithm 1. Accurate moving object segmentation based on robust seed selection.**

**Input:** adjacent images \( I_t \) and \( I_{t+1} \)

1: Estimating dense optical flow \( f_{t,t+1} \) using \( I_t \) and \( I_{t+1} \) by\(^{25}\);
2: Calculating rough contours and the probability of moving object \( \Pi_t \) by algorithm\(^{10}\);
3: Selecting the seed pixels of object and background by (5);
4: if \( t > 0 \) then

    Adding seed pixels of the object from the previous frame utilizing (7);

end if
5: Establishing a random walker model on the ROI region by (8) and seed pixels;
6: Solving the random walker model by (12) and obtaining object mask \( M \) utilizing (13).

**Output:** Object mask \( M \)

---

**Experiment results and analysis**

The sample experiment results are shown in Figure 5. The results show that the proposed method can obtain a more accurate moving object than the three compared methods. MCD\(^1\) tends to obtain incomplete objects because only a single Gaussian model is used to model the background, which is not enough and does not make use of the spatial information. At the same time, the homograph transformation cannot accurately compensate for background motion in some scenes. The algorithm of Huang et al.\(^2\) cannot deal with the slow-motion challenges because only the optical flow is used to segment moving objects, and the homograph transformation is used to denote the background movement. The contour of the object cannot be extracted accurately in Papazoglou and Ferrari’s method.\(^7\)

To compare the robustness of the four algorithms for the slow-moving object, the segmentation results of three consecutive frames in two sequences are shown in Figure 6. The algorithm of Huang et al.\(^2\) only uses the optical flow information of the current frame. When the object’s speed slows down, the algorithm cannot accurately detect or even lose the object. The algorithm of Papazoglou and Ferrari\(^7\) makes use of the information in the temporal domain of the image sequence. However, the segmentation result is greatly affected by the motion intensity of the object, so that the objects may not be segmented in some consecutive frames. The proposed method propagates the object’s seed pixels in the previous frame to the current frame according to the optical flow, which can segment the object effectively even when the motion of the object becomes weak.

The overlap area of the detected object and ground truth is used to evaluate the precision of our algorithm quantificationally. Given the segmented mask of object \( M_t \) and ground truth \( M_g \), the overlap score \( S \) is defined as

\[
S = \frac{|M_t \cap M_g|}{|M_t \cup M_g|}
\]

where \( \cap \) and \( \cup \) are the intersection and union operators, respectively, and \( |\cdot| \) is the number of pixels in the mask.

The average overlap scores of each sequence by compared and proposed algorithms are shown in Table 2, from which we can find that the proposed method outperforms the other three methods in most sequences. It illustrates that our algorithm can segment the moving object completely and has less false alarm due to utilizing the continuity of object and background distribution in spatial domain based on the random walker.

To evaluate the performance of the algorithm more precisely, this study uses another two metrics (precision rate and recall rate) to measure the proposed method. These two metrics are used to evaluate the missing and false alarm of the methods, respectively. In this study, the precision \( P \) and recall rate \( R \) are defined as equation (15). A larger \( P \) means the complete object segmentation and less missing. The large \( R \), the less false positive in the segmentation result indicates the algorithm has a lower false alarm

\[
P = \frac{|M_t \cap M_g|}{|M_t|}
\]

\[
R = \frac{|M_t \cap M_g|}{|M_g|}
\]

The average \( P \), \( R \), and \( S \) of all the sequences by compared and proposed algorithms are calculated and shown as the bar graph in Figure 7. As shown in Table 2, the proposed method performs better than the compared algorithms\(^1\)\(^,\)\(^2\)\(^,\)\(^7\) in terms of precision and recall rate. Compared with the algorithm of Huang et al.\(^2\), which has the best performance among the compared methods, the proposed algorithm improves by 8.0\%, 4.37\%, and 7.28\% in precision, recall rate, and overlap score, respectively.

**The computation efficiency of the proposed method**

We measure the computation time to evaluate the efficiency of the proposed and the compared
Table 3 shows the average computation time per frame measured by an Intel Core i5-6200U, 2.4 GHz PC with a resolution of 640 × 480 of the four algorithms, where the one by Yi et al.\textsuperscript{1} is implemented by C++ and other three by MATLAB R2018a. The time spent on calculating optical flow is excluded in the methods of Huang et al.\textsuperscript{2} and Papazoglou and Ferrari\textsuperscript{7} and the proposed method.

The algorithm of Huang et al.\textsuperscript{2} takes the least amount of time because the core is to calculate the homograph on the downsampled optical flow. Moreover, the low computational efficiency of the algorithm by Papazoglou and Ferrari\textsuperscript{7} is that it needs to calculate the superpixels of each frame and graph cut is used in each frame for segmentation. To improve the efficiency of the proposed method, this study extracts the region of interest (ROI) from the original image for segmenting the object, where the ROI must include the bounding box of the rough contours, and the pixels on its boundaries are all set to seed pixels of background. Thus, the computation time of the moving object segmented process has little correlation with the resolution of the image in the proposed method.

![Figure 5](image)

\textbf{Figure 5.} The segmentation results from different sequences of the proposed method and the compared methods.\textsuperscript{1,2,7} (a) Original image, (b) ground truth, (c) MCD,\textsuperscript{1} (d) Huang et al.’s method,\textsuperscript{2} (e) Papazoglou and Ferrari’s method,\textsuperscript{7} and (f) proposed method.
Discussion

This study adopts a coarse-to-fine strategy to segment moving objects under a moving camera. The proposed method utilizes the local difference of the optical flow field to detect the rough object contours, which is used to select the seed pixels of the object and background. Then, this study processes the object extraction locally by the random walker, which can segment the moving object efficiently and accurately. The moving object segmentation algorithm should make full use of the information in the spatial and

![Figure 6. The segmentation results of the slow-moving object by the proposed method and the compared methods.](image)

(a) The consecutive from two sequences, (b) ground truth, (c) MCD,1 (d) Huang et al.’s method,2 (e) Papazoglou and Ferrari’s method,7 and (f) proposed method.

Table 2. The overlap scores (%) of the proposed method in comparison to the methods of Yi et al.,1 Huang et al.,2 and Papazoglou and Ferrari.7

| Image sequences | MCD1 | Huang et al.2 | Papazoglou and Ferrari7 | Proposed method |
|-----------------|------|---------------|-------------------------|-----------------|
| Cars 1          | 26.36| 64.59         | 63.14                   | 79.17           |
| Cars 2          | 20.90| 48.28         | 38.18                   | 79.34           |
| Cars 3          | 19.21| 87.56         | 74.61                   | 94.42           |
| Cars 4          | 25.05| 61.73         | 43.69                   | 91.60           |
| Cars 5          | 29.35| **81.31**     | 70.89                   | 73.65           |
| Cars 6          | 36.77| 88.65         | 59.71                   | **93.53**       |
| People 1        | 41.31| **76.90**     | 50.55                   | 71.41           |
| People 2        | 39.64| **78.38**     | 73.38                   | 62.58           |
| Average         | 29.82| 73.42         | 59.52                   | **80.71**       |

"The figures in boldface represent the highest overlap score".
temporal domain. However, a strong constraint will lead to missing detection. In the temporal domain, the proposed algorithm propagates the object seed pixels to the next frame to improve the robustness. Because of using a broad range of frames for the object continuity in the temporal domain, the algorithm of Papazoglou and Ferrari\(^7\) may not segment the moving object for a long time due to the missing detection in several frames. This study just uses the seed pixels in the last frame as the supplementary information. Moreover, the proposed method can serve as the moving object detection method be it only uses the information before the current frame.

**Conclusion**

This study proposes an accurate moving object segment method in unconstrained video based on robust seed selection. The seed pixels of the object and background are selected robustly using the optical flow cues. The proposed method uses the weighted local difference in orientation of the optical flow to detect the moving object’s rough contours. Furthermore, the detected rough contours are used to guide the object and background seed pixels selection. To further improve the robustness for the slow-moving object, the object seed pixels in the previous frame are propagated to the current frame. Experiments conducted on the publicly available data sets indicate that the proposed method shows excellent performance in accurate moving object segmentation, especially for the slow-moving object.

We will try to select the seed pixels of the object and background based on the sparse optical flow for higher calculation efficiency for future work. Besides, the algorithm will be optimized and transplanted to the C++ program to be accelerated by CUDA to improve the computational efficiency.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iDs**

Wenlong Zhang https://orcid.org/0000-0003-3756-5235

Xiaoliang Sun https://orcid.org/0000-0003-3018-8043

**References**

1. Yi KM, Yun K, Kim SW, et al. Detection of moving objects with non-stationary cameras in 5.8 ms: bringing motion detection to your mobile device. In: *Computer vision and pattern recognition workshops*, Portland, OR, USA, 23–28 June 2013, pp. 27–34. Washington, DC, USA: IEEE.

2. Huang J, Zou W, Zhu J, et al. Optical flow based real-time moving object detection in unconstrained scenes. *CoRR*, vol. abs/1807.04890, 2018.

3. Bugeau A and Pérez P. Detection and segmentation of moving objects in highly dynamic scenes. In: *2007 IEEE conference on computer vision and pattern recognition*, Minneapolis, MN, USA, 17–22 June 2007, pp. 1–8. Washington, DC, USA: IEEE.

4. Bugeau A and Pérez P. Detection and segmentation of moving objects in complex scenes. *Comput Vis Image Underst* 2009; 113(4): 459–476.

5. Malik J and Brox T. Object segmentation by long term analysis of point trajectories. In: *European conference on computer vision*, Heraklion, Greece, 5–11 September 2010, pp. 282–295. Berlin, Heidelberg: Springer.
6. Ochs P, Malik J, and Brox T. Segmentation of moving objects by long term video analysis. *IEEE Trans Pattern Anal Mach Intell* 2014; 36(6): 1187–1200.

7. Papazoglou A and Ferrari V. Fast object segmentation in unconstrained video. In: *2013 IEEE international conference on computer vision*, Sydney, NSW, Australia, 1–8 December 2013, pp. 1777–1784. Washington, DC, USA: IEEE.

8. Narayana M, Hanson A, and Learned-Miller E. Coherent motion segmentation in moving camera videos using optical flow orientations. In: *2013 IEEE international conference on computer vision*, Sydney, NSW, Australia, 1–8 December 2013, pp. 1577–1584. Washington, DC, USA: IEEE.

9. Bideau P and Learned-Miller E. It’s moving! A probabilistic model for causal motion segmentation in moving camera videos. In: *European conference on computer vision*, Amsterdam, the Netherlands, 11–14 October 2016, pp. 433–449. Cham: Springer.

10. Foley JD, van Dam A, Feiner SK, et al. *Computer graphics: principles and practice*. 2nd ed. 1990. Boston, MA, United States: Addison-Wesley Longman Publishing Co., Inc.

11. Grady L. Random walks for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2006; 28(11): 1768–1783.

12. Kim K, Chalidabhongse TH, Harwood D, et al. Real-time foreground–background segmentation using codebook model. *Real-Time Imaging* 2005; 11(3): 172–185.

13. Elgammal A, Harwood D, and Davis L. Non-parametric model for background subtraction. In: *European conference on computer vision*, Dublin, Ireland, 26 June–1 July 2000, pp. 751–767. Berlin, Heidelberg: Springer.

14. Lucas BD and Kanade T. An iterative technique of image registration and its application to stereo. In: *Proceedings of the seventh international joint conference on artificial intelligence*, Vancouver, BC, Canada, 24–28 August 1981, pp. 674–679. New York: Elsevier.

15. Kim J, Wang X, Wang H, et al. Fast moving object detection with non-stationary background. *Multimed Tools Appl* 2012; 67(1): 311–335.

16. Nonaka Y, Shimada A, Nagahara H, et al. Real-time foreground segmentation from moving camera based on case-based trajectory classification. In: *2013 2nd IAPR Asian conference on pattern recognition*, Naha, Japan, 5–8 November 2013, pp. 808–812. Tokyo, Japan: IEEE.

17. Sheikh Y, Javed O, and Kanade T. Background subtraction for freely moving cameras. In: *2009 IEEE 12th international conference on computer vision (ICCV)*, Kyoto, Japan, 29 September–2 October 2009, pp. 1219–1225. Tokyo, Japan: IEEE.

18. Chen T and Lu S. Object-level motion detection from moving cameras. *IEEE Trans Circuits Syst Video Technol* 2016; 27(11): 2333–2343.

19. Ilg E, Mayer N, Saikia T, et al. FlowNet 2.0: evolution of optical flow estimation with deep networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, Honolulu, HI, USA, 21–26 July 2017, pp. 2462–2470. Piscataway, NJ: IEEE.

20. Sajid H, Cheung SCS, and Jacobs N. Motion and appearance based background subtraction for freely moving cameras. *Signal Process Image Commun* 2019; 75: 11–21.

21. Bruss AR and Horn BKP. Passive navigation. *Comput Vis Graph Image Process* 1983; 21(1): 3–20.

22. Wu Y, He X, and Nguyen TQ. Moving object detection with a freely moving camera via background motion subtraction. *IEEE Trans Circuits Syst Video Technol* 2017; 27(2): 236–248.

23. Zhu Y and Elgammal A. A multilayer-based framework for online background subtraction with freely moving cameras. In: *2017 IEEE international conference on computer vision (ICCV)*, Venice, Italy, 22–29 October 2017, pp. 5132–5141. Piscataway, NJ: IEEE.

24. Koh YJ and Kim C. Primary object segmentation in videos based on region augmentation and reduction. In: *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017, pp. 3442–3450. Piscataway, NJ: IEEE.

25. Sun D, Roth S, and Black MJ. Secrets of optical flow estimation and their principles. In: *IEEE conference on computer vision and pattern recognition, CVPR 2010*, San Francisco, CA, USA, 13–18 June 2010, pp. 2432–2439. San Francisco, CA: IEEE.