Object Detection and Tracking using Watershed Segmentation and KLT Tracker

Nilamani Bhoi

Received: 13 December 2019 Accepted: 3 January 2020 Published: 15 January 2020

Abstract

In this paper, a moving object is extracted from a video using video object detection algorithm based on spatial and temporal segmentation. The technique begins with temporal segmentation in which edge map is extracted using edge operator. The initial binary mask is obtained by using morphological operation applied on initial edge map. The next phase is spatial segmentation where gradient image is obtained by multi-scale morphological operator. The modified gradient image is obtained by the operator applied over the current frame. At last, moving object is extracted by precisely and accurately by watershed segmentation which is performed on the modified gradient image. Again, morphological operation is applied on the output to get final binary mask. This binary mask is then complemented to yield the contour line of the video object. Using the binary mask, the video object is extracted from the video frames. After detection of video object, the object tracking is performed using Kanade-Lucas-Tomasi (KLT) feature tracker.

Index terms — video object segmentation, morphological operator, watershed algorithm, KLT tracker

1 Introduction

Video object segmentation describes to take out of the objects moving from the camera’s order. Segmentation is the process of dividing information fragments into essential elements that are defined as part. In terms of the still images, the segmentation means to split the image into a random number of areas that represent the main part of the image. The given video, word segmentation is used to describe the number of different processes for the division of videos into sections that have meaning into different granulations. This video can be temporarily dissociated to a scene or shots, and a background.

Spatial segmentation of pictures incorporates regional-based and boundary-based strategies. The region-based approach depends on the nearby highlights such as intensity, surface and position. In other words, a temporary segmental video segment is expected to separate the video into a component picture called scenes and shots. In other words, a temporary segmental video segment is expected to separate the video into a component picture called scenes and shots. This shot is characterized as frame that obtained without intrusion by a camera. Temporal division to shots is done by distinguishing the move from one to the next.

There are number of object detection techniques and algorithm based on video segmentation has proposed by many researchers. In [3] Neri, A, Colonnese, S, Russo, G, Talone, P, jointly presented a paper in which the method used for segmentation is of low computational complexity. This paper aimed at separating the moving objects from the background in video grouping. In [4] Renjie Li, Songyu Yu, Xiaokang Yang, proposed a paper which addresses an productive spatio-temporal division plot to extricate moving objects from video arrangements. The temporal segmentation yields a temporal mask that demonstrates moving regions and inactive region for each frame. In [5] Chinchkhede, D.W.; Uke, N. J, presented a paper based on the process for image segmentation in video sequence based on Expectation Maximization (EM) is used, which is a mixture of Gaussian classification model. In [6] K Ganesan, S Jalla, presented a paper in which a comparative study of video object extraction based on efficient edge detection techniques. In [7] Gao Hai, Siu Wan2Chi, Hou Chao2Huan presented a new progressed image segmentation scheme in which an unsupervised image-segmentation algorithm based on morphological tools.
has been presented. In [11] L. Vincent and P. Soille introduced powerful algorithm for computing watersheds in digital gray-scale images. Here a overview of watersheds is discussed then immersion simulation process is applied. In [12] Thomas Sikora implemented a video standard verification model to develop the algorithm. Here description of content-based scalability and content-based bit stream access and manipulation are given. In [16] Nisha Singla implemented a paper which presents a modern calculation for identifying moving objects from a inactive background scene based on frame difference. In [18] Arindrajit Seal, Arunava Das, Prasad Sen, described that a grey-level picture may be visualized as a topographic alleviation, where the grey-level of a pixel is thought as its stature within the relief.

The paper is organised as follows: session I we discussed about the proposed method. Temporal segmentation and spatial segmentation are discussed in session II. At last experimental result, discussion and conclusion is presented in session III.

2 II. Proposed Method

3 Proposed Method

Here the purpose of the paper is that the process of the video segmentation begins with the difference between the image between the preceding frame and the current frame. Because of the different picture, the frame contains all the information about changes between the frame and the noise. Working with differences is followed by the extraction of video frames in folders that are used more to play and read video frames. The edge of the image, which differs in the frame, is determined by the edge detection operators. The Canny operator has the ability to get high accuracy when detecting edges and constraining the false edges. Within the same way the edge of the current frame is detected by the edge operator. Then morphological operations are processed on the edge of difference image, resulting in the temporal mask of the image difference. Background noise has been extracted from the MATLAB image processing function. Here we use multi-scale morphological gradient on the current framework for watershed conversion. Then the paper reached the desired results i.e. contour of the video object followed by extracted video object.

After object detection or moving object extraction, then we can track the object by using tracking algorithm. Here we used Kanade-Lucas-Tomasi (KLT) Tracker. It works well suited for tracking objects in which it does not change shape and exhibit visual texture.

4 Methodology

Different methods used for object detection are described below:a) Object Segmentation

Object segmentation is one of the methods in which we can successfully segment the object and detect the moving object. It is performed by using temporal and spatial segmentation.

5 i. Temporal Segmentation

The initial step or method of video segmentation is temporal segmentation. The objective is to detect the object from background. The object may be moving or stationary based on the video taken while recording.

Here we have taken moving object video database. From the block diagram, temporal segmentation can proceed by using the following methods.

6 a. Frame Difference

Here we have started our work by extracting number of frames from the video and stored in a folder which is further used for our work. Then we have chosen a particular frame i.e. f (t) which is the current frame and its previous frame i.e. f (t-1). After converting these original images into their respective gray images we have to subtract one frame from another.

7 b. Edge detection of moving object

Edge location may be an essential apparatus for image segmentation. Most strategies are applied to image fragmentation based on changes in local intensity. The boundary between two regions with different gray properties is called as edge of an image. We get the edge map image by using canny operator as follows:Edge =??????????(Δ ??"Δ ??" ???1 ? Δ ??"Δ ??" ?? ) (1)

Canny approach is based on three goals:
1. Low error rate here no false response is found responses should be there. The edge detected must be close to the actual edge. 2. The edges must be well localized i.e., the distance between the edges marked by the detector should be as close as possible to the centre of the real edge.

8 Single point response-Only one point should come

back from the detector for the real edge, which means the number of peaks around the edges should be minimal.
9 c. Morphological Process

When images are processed for enhancement and while performing some operations like thresholding, more is
the chance for distortion of the image due to noise. As a result, imperfections exist in the structure of the image.
The primary goal of morphological operation is to remove this imperfection that mainly affects the shape and
texture of image. Dilation and erosion are two primary operations of morphological processing.

The dilation operation expands an object both horizontally and vertically. Hence number of structuring
elements of different shape and size are applied over the object. The dilation of an object A (set) striking by
structuring element B is characterized as: 
\[
\text{\texttt{A \ominus B}} = \{x | B(x) \subseteq A\}
\]
On the off chance that the set B is symmetric about its origin and changes with z. so B? and A have at least
one common component. At that point it can be characterized as:
\[
\text{\texttt{A \ominus B}} = \{x | B(x) \subseteq A\}
\]
The erosion operation is just the reverse of dilation operation. The erosion operation shrinks the object. The
erosion can be characterized as:
\[
\text{\texttt{A \ominus B}} = \{x | B(x) \subseteq A\}
\]
The erosion of an object A by structuring element B can also be defined as:
\[
\text{\texttt{A \ominus B}} = \{x | B(x) \subseteq A\}
\]

As described above dilation and erosion morphological methods are used to find out the initial binary mask
of the object. In order to eliminate the background noise, 'bwareaopen' Matlab function is used here to remove
all the connected components that have less than a certain number of pixels. The structuring element is of type

10 ii. Spatial Segmentation

We only get a rough part through the temporary breakdown due to the complex traffic information. Spatial
segmentation is required to achieve the precise boundary of the object. Watershed is one of the fast segmentation
algorithms of the mathematical morphology. A neighborhood least compares to the valley, while the most extreme
compares to the top. The water surface will be filled slowly from the minimum base. As water level will rise,
wherever the level of other regions also increases. On the off chance that the location where the assembly was
built was a dam that avoided this merger, at that point the geography is partitioned into distinctive regions,
known as the catchment basins. At the conclusion of each least submersion strategy, it is totally encompassed by
the dams -close the range where the building, called the watershed. So, it introduced catchment basins and edge
features due to watershed segmentation. Watershed algorithms are ordinarily executed on the gradient. The normal
temporary operator generates minimal local results on broken or abnormal errors. To mitigate this problem, due
to the fact that the morphological gradient of the image has increased more in scale, compared to the resulting
gradient image, by the spatial template of the image, the gradient of the morphological structure by a symmetric
structural element is less depending on the direction of the edges. Multi-scale morphological algorithms are
connected to the current frame, the video object and the foreground marker, and this background is utilized to
control watershed segmentation in arrange to attain way better spatial distribution. The watershed change is
broadly utilized in numerous zones of image handling, counting parts of therapeutic imaging, due to a few of
the benefits given below.

11 a. Multi-scale morphological gradient

The input gray scale object is denoted as f, the structuring element is B and ? , ? are the dilation and erosion
morphological operations, then with the standard operator with a single dimensional morphological gradient is
characterized as:
\[
\text{\texttt{f \odot B}} = \{x | B(x) \subseteq f\}
\]
Its execution depends on the estimate of the structuring component, in the event that the sort B is huge, it’ll lead to an overlap between the edges, which
can lead to a greatest gradient that does not coordinate an edge. In any case, if the structuring element is as
well small, with an incline, it produce ramp edges having low output values with high spatial resolution gradient
operator. The multi-scale morphological gradient operator is defined as below; Year 2020 © 2020 Global

12 b. Binary Operation

Here in this paper we have used sobel and canny edge detection techniques on different video sequence. Then
using morphological operations using matlab we got the initial binary mask of the object. After getting the
multi-scale gradient image, it is needed to apply watershed transformation to reduce the over segmentation. By
multiplying the initial binary mask with the original frame resulting the extracted video object. Then the paper
reached the desired results i.e. contour of the video object followed by extracted video object.
13  b) Object tracking

In computer vision, the motion of an object is tracked by one of the methods called optical flow. In this method, the velocity vectors for points in a series of images or frames calculated and it approximate positions of points in next image sequence. One of the challenges of computer vision is calculating optical flow or motion velocity. Hence next it describes an object tracking technique called KLT (Kanade-Lucas-Tomasi) feature tracker.

14  i. Kanade-Lucas-Tomasi (KLT) Tracker

One of the feature-tracking algorithm is Kanade-Lucas-Tomasi(KLT), which tracks a set of points. There are many applications of KLT such as camera motion estimation, video stabilization and object tracking. It works well suited for tracking objects in which it does not change shape and exhibit visual texture.

Kanade-Lucas-Tomasi method derives and calculates the difference between two frames of the video sequence. The object tracking is based on the criteria of Sum of Squared Difference (SSD) which is applied to find the feature point whose objective is to minimize the following energy function using window: \( E(v) \). The KLT algorithm can be of two main phases, (1) detection phase (2) tracking phase. In the detection phase, first step is searching for the salient feature points and next these feature points are added to the already existing ones. In the tracking phase, the motion vector is calculated for each corresponding feature point.

15  a. Feature Point Detection

In this method for a given object we have to detect new feature points and add these feature points to the already existing one. Basically, the feature points pixels neighborhood are highly structured. Hence it is more reliable and accurate to track feature points. So, the structure matrix \( G \) can be defined as: \( G = \mathbf{W}(\mathbf{W}) \mathbf{W} \). \( \mathbf{W}(\mathbf{W}) \)

Its eigen values \( ?1, ?2 \) is always \( > 0 \) because the matrix is positive semi-definite represent the neighbourhood region. Based on the values of \( ?1 \) and \( ?2 \), \( W \) is defined. If \( ?1 = ?2 = 0 \), \( W \) is completely homogenous. If \( ?1 > 0, ?2 = 0 \) then \( W \) indicates an edge and \( ?1 > 0, ?2 > 0 \) indicates a corner. Strong corners that have higher \( ?1, ?2 \) values are extracted by KLT tracker.

16  b. Feature Point Tracking

In this tracking phase, let \( I \) and \( J \) are denoted as the current frame and next frame respectively in the given video sequence. Our objective is to calculate the motion vector \( v \) for each corresponding feature point \( p \) in frame \( I \), so that its tracked position in frame \( J \) is \( p + v \). Hence the SSD error function is calculated as \( E(v) = \sum (I(x) - J(x + v))^2 \). The above equation defines or measures the deviation of intensity of frame between a neighbourhood of the feature point position in \( I \) and its potential position in \( J \) and should be zero in the ideal case. In order to better estimate for \( v \), take the first derivative of \( E(v) \) and set it to zero and approximating \( J(x + v) \) by its first order Taylor expansion around \( v = 0 \). It is an iterative method; hence by using no of iteration, we obtain the better result for \( v \). Here, a particular threshold value is set. The condition is that the feature point is removed from consideration if the quality of a tracked feature point decreases below a predefined threshold. Hence new features are introduced in the same window for compensation. The feature point with the minimum criteria should be retained if its SSD is below a certain threshold.

IV. Result and Discussion

The proposed technique used for video segmentation is executed in MATLAB program of version R2018. We have taken four standard video database called hall_monitor, Claire and daria_walk. These databases are well suited for object detection. Initially we have taken hall_monitor database for segmentation is illustrated in Figure ??.

The performance of the object detection method can be quantified by using FPR, TPR, Acc. We have taken the ground truth image of the respective video sequence for better comparison of the detection result with the binary mask. Among all the video sequence hall_monitor sequence gives good accuracy.

Here we have taken daria_walk video sequence. The KLT tracker is used to track the motion of the human, who is walking on the street. Here the object is moving and the background is stationary. The figure shows 1, 10, 50 and 80th frame of the video sequence.

18 Conclusion

This paper inquires the segmenting algorithm for a video based on the temporal and spatial data. Amid the temporal segmentation stage, Canny is utilized to discover the edge of the difference between the two adjoining frames. Initial binary segmentation mask is obtained by the morphological process. The erosion operation is chosen so as to work on a temporal mask for partial division to quote the foreground and background substrates for watershed calculation. Within the spatial division stage, a multi-scale morphological gradient operator with a high capacity in commotion concealment is applied to the current image frame to pick up gradient images. At long last, the watershed division is done on a modified gradient image. It is slightest influenced by commotion
and lighting changes. It overcomes the method of over-segmentation and partitioning algorithm, maintained a strategic distance from the present-day handle of joining this region to diminish computer complexity. The proposed procedure incorporates numerous parameters such as high and low levels within the canny operator and the measure of the structuring components that are characterized within the test. In order to exactly extract the moving object from stationary background historical information is required. One of the challenging tasks is how to extract the moving object if the background is also moving.

Figure 1: Fig. 1 : 2 :

---

1 © 2020 Global Journals Object Detection and Tracking using Watershed Segmentation and KLT Tracker
2 Year 2020 ( ) F © 2020 Global Journals Object Detection and Tracking using Watershed Segmentation and KLT Tracker
18 CONCLUSION

Figure 2: Fig. 3 : Fig. 4 : Fig. 5 : Fig. 6 :

\[ E_i(dx, dy) = \sum [I(x + dx, y + dy, t + dt) - I(x, y, t)]^2 \]

Figure 3: Fig. 7 :

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Image Sequence</th>
<th>FPR</th>
<th>TPR</th>
<th>Ac</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hall monitor(46)</td>
<td>0.0054</td>
<td>0.9443</td>
<td>0.9927</td>
</tr>
<tr>
<td>2</td>
<td>Claire(1)</td>
<td>0.0123</td>
<td>0.0326</td>
<td>0.9818</td>
</tr>
<tr>
<td>3</td>
<td>Momson(33)</td>
<td>0.0241</td>
<td>0.9914</td>
<td>0.9829</td>
</tr>
</tbody>
</table>

Figure 4: Table 1 :


[Singla ()] Motion Detection Based on Frame Difference Method International Journal of, Nishu Singla . 2014.


[King Ngi Ngan and Li ()] Video segmentation and its applications [electronic resource], Hongliang King Ngi Ngan, Li . c2011. New York: Springer.