Global Journal of Computer Science and Technology: F
Graphics & Vision
Volume 20 Issue 1 Version 1.0 Year 2020
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-4596 & Print ISSN: 0975-5861

Object Detection and Tracking using Watershed Segmentation and KLT Tracker

By Tunirani Nayak & Nilamani Bhoi
Veer Surendra Sai University of Technology

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.

Keywords: video object segmentation, morphological operator, watershed algorithm, KLT tracker.

GJCST-F Classification: I.4.8

Object Detection and Tracking using Watershed Segmentation and KLT Tracker

Tunirani Nayak [α] & Nilamani Bhoi [σ]

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.

Keywords: video object segmentation, morphological operator, watershed algorithm, KLT tracker.

I. 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 disassociated 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] Nishu 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.

Author α σ: Department of Electronics & Telecommunication Engineering, Veer Surendra Sai University of Technology, Burla, Sambalpur-768018, Odisha, India. e-mails: tuniraninayak@gmail.com, nilamanib@gmail.com

II. 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.

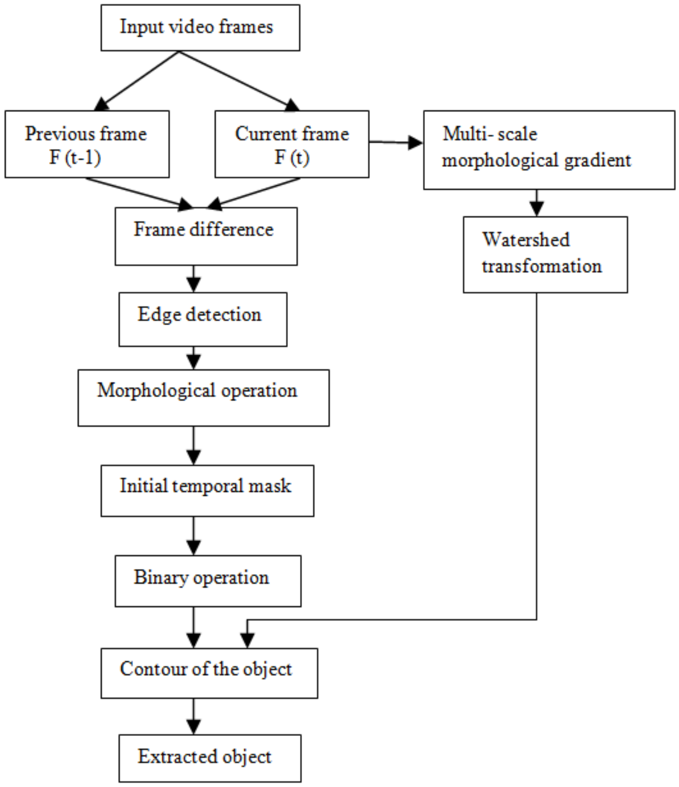


Fig.1: Block diagram of object detection

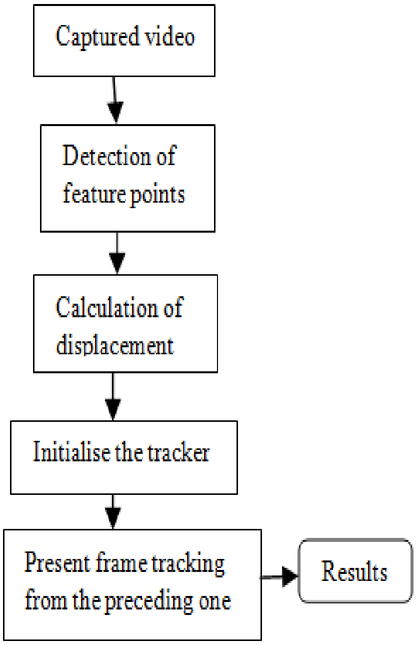

Captured video
Detection of feature points
Calculation of displacement
Initialise the tracker
Present frame tracking from the preceding one
Results

Fig. 2: Block diagram of KLT tracker

III. 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.

i. *Temporal Segmentation*

The initial step or method of video segmentation is temporal segmentation. The objective is to detect the object from back ground. 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.

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.

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:

$$\text{Edge} = canny(f_{t-1} - f_t) \tag{1}$$

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.
3. 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.

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;

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \emptyset \tag{2}$$

On the off chance that the set B is symmetric about its origin and changes with z. so $\hat{B}$ and A have at least one common component. At that point it can be characterized as;

$$A \oplus B = \{z \mid [(\hat{B})_z \cap A] \subseteq A\} \tag{3}$$

The erosion operation is just the reverse of dilation operation. The erosion operation shrinks the object. The erosion can be characterized as;

$$A \Theta \mathrm{B} = \{z \mid (B)_z \subseteq A\} \tag{4}$$

The erosion of an object A by structuring element B can also be defined as

$$A \Theta \mathrm{B} = \{z \mid (B)_z \cap A^c = \emptyset\} \tag{5}$$

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 disk-shaped for proper detection.

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, hence water 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 lines 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 picture handling, counting parts of therapeutic imaging, due to a few of the benefits given below.

a. *Multi-scale morphological gradient*

The input gray scale object is denoted as f, the structuring element is B and ⊕ , Θ are the dilation and erosion morphological operations, then with the standard operator with a single dimensional morphological gradient is characterized as $G(f) = (f \oplus \mathrm{B}) - (f \Theta \mathrm{B})$. 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 a 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;

$$MG(f) = \frac{1}{3} \times \sum_{i=1}^{3}[(f \oplus B_i) - (f \Theta B_i) \Theta B_i - 1] \qquad (6)$$

Bi is the disk-shaped structuring component of ith groups, and its span 2*i* +1, 0 ≤*i*≤ 3. The multi-scale morphological gradient has preferred to apply individually for both large and small structuring components. It is safe to clamours and intelligently edges due to the normal operation utilized within the calculation. It makes strides the obscured edge and decreases the number of neighbourhood minima that cannot be performed.

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.

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.

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:

$$E_t(dx,dy) = \sum [I(x+dx, y+dy, t+dt) - I(x,y,t)]^2 \qquad (7)$$

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.

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:

$$G = \sum_{x \in W(p)} \nabla I(x).\nabla I(x) \qquad (8)$$

Its eigen values λ1, λ2 is always≥ 0 because the matrix is positive semi-definite represent the neighbourhood region. Based on the values of λ1 and λ2, W is defined. If λ1= λ2=0, W is completely homogenous. If λ1 > 0, λ2 = 0 then W indicates an edge and λ1 > 0, λ2 > 0 indicates a corner. Strong corners that have higher λ1, λ2 values are extracted by KLT tracer.

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

$$\varepsilon(v) = \sum_{x \in W(p)} (J(x+v) - I(x))^2 \qquad (9)$$

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 v1, take the first derivative of ε(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 3. Fig 3 (a) is the current frame 46. Fig 3 (b) is the subtracted image from current frame with previous frame, 46 and 47. The subtracted output carries the less information about the object. By using the canny operator, we got the edge map as shown in fig 3(c). Here still some edge information's are missing. Fig 3 (d) gives the initial temporal mask by using some morphological operation like dilation and erosion. In order to avoid over segmentation then we found watershed transform on the gradient image of the current frame as shown in fig 3(e). To properly extract the object with exact boundary, a contour of the object is found out after executing binary operation on the initial binary mask as shown in fig 3 (f). Here we got the exact contour. The moving object is extracted as appeared in Fig 3 (g) after applying some post processing operation. The proposed algorithm ia also applied over Claire, daria_walk and momson database as shown in Figure 4, 5 and 6 respectively. Still some background information is available in the extracted video object. In Claire video sequence, some background information are available in the head portion. In daria_walk sequence the path walk of the human being is perfectly identified .But still some background parts are detected with object. In momson sequence the object is perfectly detected.

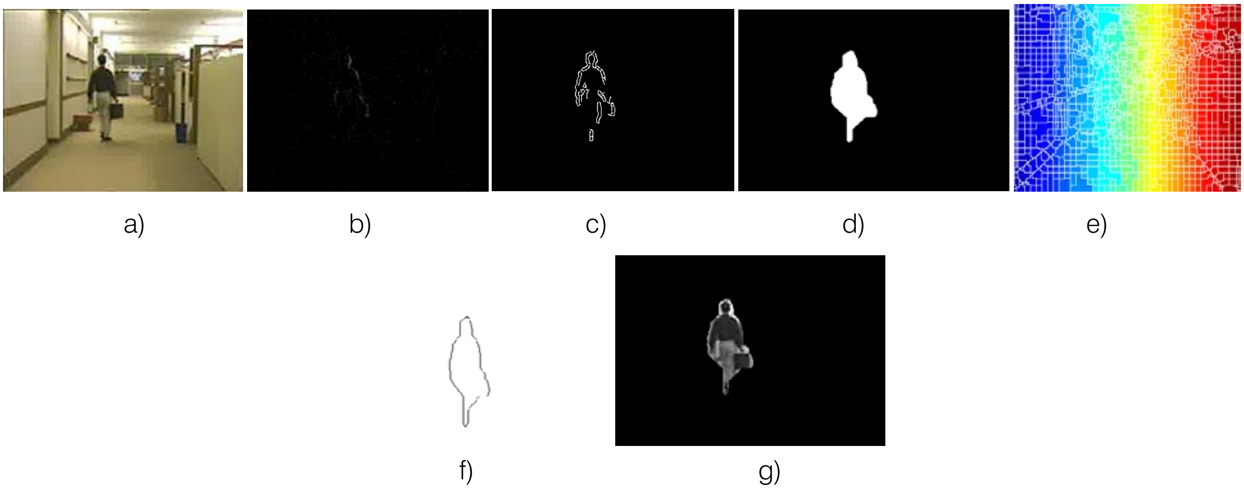


Fig. 3: Object detection result of hall_monitor image (a) original frame-46, (b) Frame difference image 46, 47, (c) edge mask, (d) binary mask, (e) watershed segmentation output, (f) contour of the object, (g) final output.

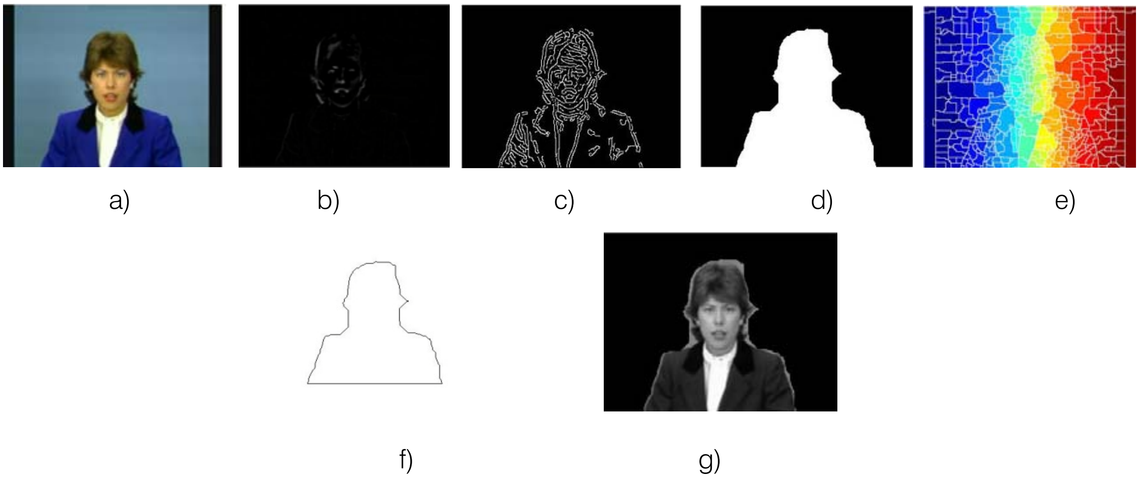
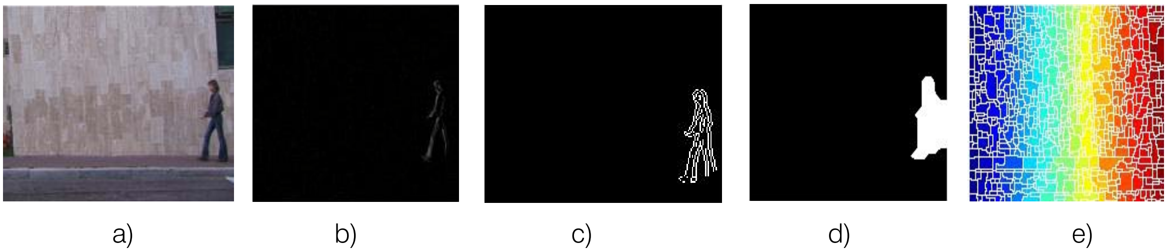
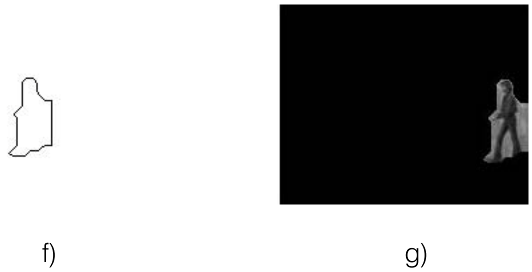


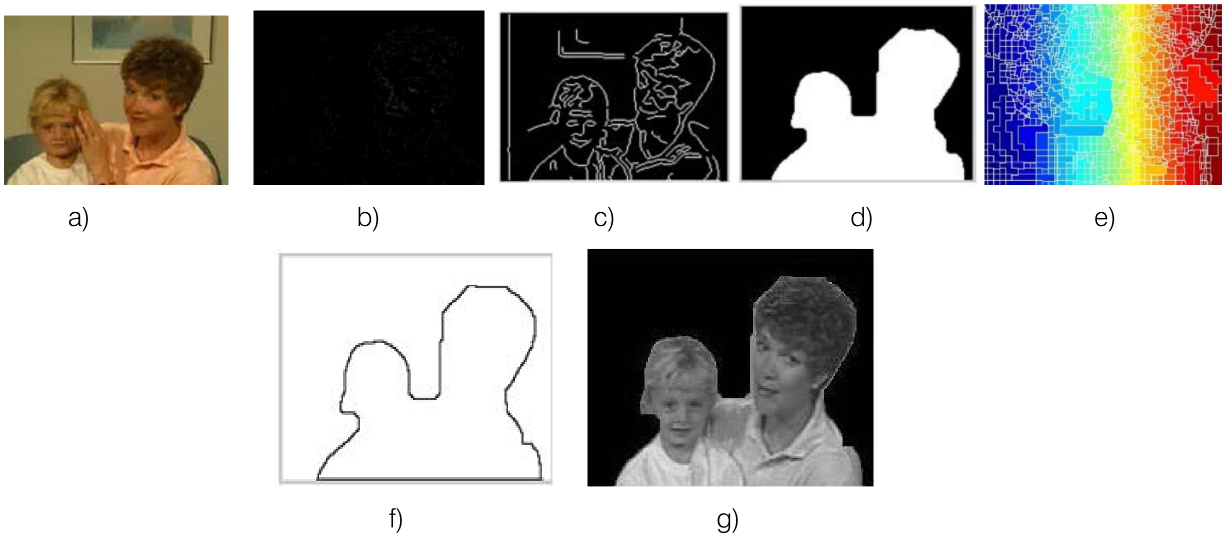
Fig. 4: Object detection result of Claire image (a) original frame-1, (b) Frame difference image 1,2, (c) edge mask, (d) binary mask, (e) watershed segmentation output, (f) contour of object, (g) Extracted moving object.





f) g)

Fig. 5: Object detection result of daria_walk image (a) original frame-46, (b) Frame difference image 46, 47, c) edge mask, (d) binary mask, (e) watershed segmentation output, (f) contour of the object, (g) final output.



a) b) c) d) e)

f) g)

Fig. 6: Object detection result of momson image (a) original frame-33,(b) Frame difference image 33,34 (c) edge mask, (d) binary mask, (e) watershed segmentation output, (f) contour of object, (g) Extracted moving object.

By using TP, FP, TN, FN statistics, different evaluation metrics are calculated as follows.

$$FPR = \frac{\text{number of FP}}{(\text{number of FP} + \text{number of TN})}$$

$$TPR = \frac{number\ of\ TP}{number\ of\ TP + number\ of\ FN}$$

$$\text{Accuracy(Ac)} = \frac{\text{number of TP} + \text{ number of TN}}{(\text{number of }(\text{TP} + \text{TN} + \text{FP} + \text{FN})}$$

Table 1: Performance measures (FPR, TPR, Ac) of Hall_monitor, Claire, Momson Image Sequence

Sl. No.	Image Sequence	FPR	TPR	Ac
1	Hall monitor(46)	0.0054	0.9443	0.9927
2	Claire(1)	0.0123	0.0326	0.9818
3	Momson(33)	0.0241	0.9914	0.9829

The performance of the object detection method can be quantified by using FPR, TPR, Ac. 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.

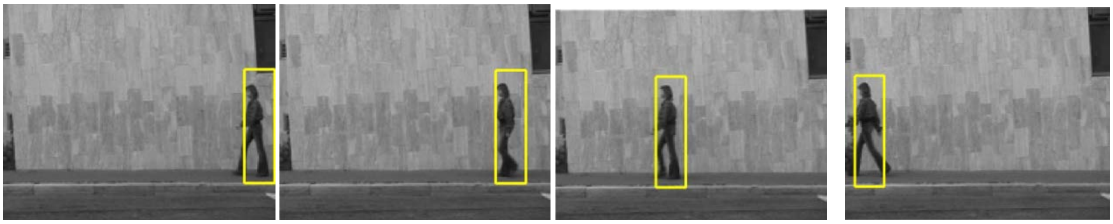


Fig. 7: Object tracking result a) Frame No.1, (b) Frame No.10, (c) Frame No.50, (d) Frame No.80.

V. 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.

References Références Referencias

1. Sikora T, The MPEG-4 video standard verification model, IEEE Transactions on Circuits System for Video Technology, Vol.7, 19~31, 1997.
2. King Ngi Ngan, Hongliang Li, Video segmentation and its applications [electronic resource], New York: Springer, c2011.
3. Neri, A, Colonnese, S, Russo, G, Talone, P, (1998), Automatic moving object and background separation, Signal Processing, APR, Vol.66, no.2, p.219-p232.S.K. Sharma, Performance Analysis of Reactive and Proactive Routing Protocols for Mobile Ad-hoc N/W, World Academics Journal of Engineering Sciences, Vol.1, No.5, pp.1-4, 2013.
4. Renjie Li, Songyu Yu, Xiaokang Yang, (2007), Efficient Spatio-temporal Segmentation for Extracting Moving Objects in video Sequences, IEEE Transactions on Consumer Electronics, Vol.53, Issue 3, p.1161-1167.
5. Chinchkhede, D.W.; Uke, N. J,(2002),Fast and Automatic Video Object Segmentation and Tracking for Content- Based Application ,IEEE Transactions on Circuits & Systems for Video Technology, Vol. 12, Issue 2, p.122-129.
6. K Ganesan, S Jalla,(2009)Video Object Extraction Based on a Comparative Study of Efficient Edge Detection Technique International Arab Journal of Information Technology(IAJIT), Vol. 6, Issue 2, p.107-115
7. Gao Hai, Siu Wan2Chi, Hou Chao2Huan,(2001),Improved techniques for automatic image segmentation, IEEE Transactions on Circuits and Systems for video technology, Vol.11 ,no.12,pp. 1273 – 1280.
8. Muthukrishnan. R and M. Radha, (2011), International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6.
9. D. Wang, (1998), unsupervised video segmentation based on watersheds and temporal tracking, IEEE Transactions on Circuits and Systems for Video Technology, vol. 8, pp.539 -546.
10. JOHN CANNY, (1986),A Computational Approach to Edge Detection, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. PAMI-8, NO. 6.
11. L. Vincent and P. Soille, (1991),Watersheds in digital spaces: An efficient algorithm based on immersion simulations, IEEETrans. On Pattern Analysis and Machine Intelligence, vol. 13,pp. 583–598.
12. Thomas Sikora,(1997),The MPEG-4 video standard verification model, IEEE Transactions on Circuits System for Video Technology,Vol.7,19~31,1997.
13. R. C. Gonzalez and R. E. Woods, (2002) Digital Image Processing, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition,. H.R. Singh, “Randomly Generated Algorithms and Dynamic Connections,” International Journal of Scientific Research in Biological Sciences, Vol.2, Issue.1, pp.231-238, 2014.
14. Hanqing Jiang, Guofeng Zhang, Huiyan Wang , and Hujun Bao, (2015), Spatio-Temporal Video Segmentation of Static Scenes and Its Applications, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 17, NO. 1.
15. Nishu Singla, (2014), Motion Detection Based on Frame Difference Method International Journal of

Information & Computation Technology. ISSN 0974-2239 Volume 4, Number 15, pp. 1559-1565.

16. Yasira Beevi C P and Dr. S. Natarajan, (2009), An efficient Video Segmentation Algorithm with Real time Adaptive Threshold Technique, International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 2, No.4.
17. Arindrajit Seal, Arunava Das, Prasad Sen, (2015), Watershed: An Image Segmentation Approach, Arindrajit Seal et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2295-2297.