International Journal of Engineering & Technology, 7 (4) (2018) 2598-2602

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.17610
Research paper

Video object extraction using optimized smoothed dirichlet process multi-view learning with improved adaptive modified Markova random field

G. S. Gowri [1] *, Dr. P. Ponmuthuramalingam [1]

[1] Department of Computer Science, Government Arts College, Coimbatore-641046
**Corresponding author E-mail: gowri_jana@yahoo.co.in*

Abstract

Video object extraction (VOE) using segmentation from a video sequence is a very important task in editing and multimedia analysis for film making. Most of the VOE approaches required prior knowledge about background and foreground to extract target objects. In this paper, an Optimized smoothed Dirichlet Process Multi-view learning with improved adaptive Modified Markov Random Field which is enhanced by adaptive shape prior modified graph cut (OsDPMVL-IASMMRF) model has been extended for video-based object extraction. The contour tracking has been additionally included OsDPMVL-IASMMRF for VOE. The Teh–Chin algorithm has been used with OsDPMVL-IASMMRF for predicting the contour in the current frame by matching the extracted object contour from the previous segmented frame. The contour tracking propagates the shape of the target object, whereas the OsDPMVL-IASMMRF segmentation refined the object boundary and the shape for enhancing the accuracy of video segmentation. The experimental outcomes show that the proposed approach provides better segmentation results in terms of accuracy, precision and recall.

Keywords: *Video-Based Object Extraction; Contour Tracking; Adaptive MRF; Video Segmentation; TEH-Chin Algorithm*

1. Introduction

Numerous computer vision applications need the segmentation of the moving object, which is an essential process in target identification, traffic observation and interloper activity analysis [1]. The Background Subtraction (BS) was the foremost best strategy to figure out moving objects from the video sequences [2-4].The tracking of objects using BS provided high precision at pixel, video frame and patch level processing. The usage of BS was computationally very cheap and the prior input for object tracking was not required. The extraction of moving objects from the video sequences using BS method provided additional statistical information about the extracted region. The estimation of object location using the Kalman filter improved the accuracy of BS method for object tracking [5-6]. The initial estimation provided the adequate size of tracking samples than simple BS method.

The image matting method was used for reducing classification error and wrongly detected shadows while object tracking [7]. The correlation and phase angle methods were utilized in template matching for object tracking [8]. The motion vector was found by using Block Matching Algorithm (BMA) [9]. Scale-invariant feature transform and Speeded up Robust Feature [10] matching and image matting were used to detect false shadow. However, more inputs were required for these methods for object tracking.

A Markov Random Field (MRF) based contour tracking and graph-cut image segmentation for VOE (MRFCT-GCS-VOE) was proposed [11]. The contour tracking propagates the shape of the target object, whereas the graph-cut refines the shape and improves the accuracy of video segmentation. However, the MRFCT-GCS-VOE has the limitation, if the target object contained holes. The color intensity of the hole in the background was preventing to distinguish target object from holes. The context based segmentation has been required to solve this issue.

In this paper, OsDPMVL-IASMMRF model has been utilized for VOE along with contour tracking. OsDPMVL-IASMMRF was proposed for contextual integrative image segmentation [12]. OsDPMVL-IASMMRF was simulated annealing optimized Dirichlet Multi-view learning based modified graph cut segmentation method, capable of distinguishing holes from the target image. The Teh-Chin algorithm [13] has been used to predict the contour in the current frame by matching the extracted object contour from the previous segmented frame. The contour tracking propagates the shape of the target object, whereas the OsDPMVL-IASMMRF segmentation refined object boundary and the shape to improve the accuracy of video segmentation.

The remainder of the article has been prepared as follows: Section 2 describes about the existing VOE methods. Section 3 explains about the proposed approach. Section 4 shows the performance evaluation of the proposed approach. Section 5 concludes the proposed work of this paper.

2. Literature survey

An approach [14] was designed for extracting objects from video sequences by using spatiotemporal independent component analysis (stICA) and multiscale analysis. The stICA extracted the preliminary source images from video sequences for extracting moving objects. Then, the wavelet-based multiscale image segmentation and region detection method extracted the object from primary source images. However, the more iteration was required to extract possible video objects. An object extraction method [15] was designed to extract the required motion pattern from the vide-

os which consist of different motion patterns. The incorporation of color and texture features which were semantically relevant to objects through multiple views achieved various motion pattern based object extraction. Video coding approaches [16] were used for object extraction and motion estimation to support multiple objects extraction. The objects undergone occlusion and merging in different illumination conditions was extracted accurately.

Three different types of background subtraction approaches [17] were proposed for tracking the player's activity in sports which helps to improve the performance of the player. The Mixture of Gaussians provided a better result than frame differencing and approximates median filtering. The three methods were not performing well when environment noise in the video. Automatic video object segmentation approach [18] detected object like region and basic structures initially. The primary objects and strictures were used to extract diverse set of object from remaining frames. The time complexity was the main drawback of this approach. An illumination-invariant color-texture feature based online video object segmentation technique [19] was introduced. The location of the object of interest was initialized through user-specified markers. The marker pixels were called Super Pixels (SP). In the next available frame, the object marker prediction located the object of interest through SP motion prediction using optical flow.

A video object segmentation using recurrent neural network [20] was proposed. The recurrent structure was capable of extracting long term temporal objects from the video while rejecting outliers. An object tracking through movement prediction [21] used the optical flow between video sequences. Once the flow field determined, the Gabor feature was extracted to estimate the object tracking in the flow filed. The Gaussian mixer model of extracted features was used for background subtraction using Expectation Maximization.The boundary of objects was refined using Adaboost classifier.

3. Proposed methodology

In this section, OsDPMVL-IASMMRF model is extended with VOE. The image segmentation through this model is incorporated contour tracking for VOE. The Teh–Chin algorithm is used with this model for predicting the contour in the current by matching the extracted object contour from the previous segmented frame. The contour tracking propagates the shape of the target object, whereas this model segmentation refined object boundary and the shape to improve the accuracy of video segmentation.

3.1. OsDPMVL-IASMMRF with video object extraction (OsDPMVL-IASMMRF-VOE) model

In this model, the contour prediction of the current frame in the video, the Teh Chin algorithm (Teh, C. H., & Chin, R. T. 1989) is adopted for extracting the object contour from the previous segmented frame. The contour is indicated as a polygon,

$$C^{t-1} = \{V^{t-1}, \varepsilon^{t-1}\} \tag{1}$$

In the above equation, the $t-1$ represents the index of the previous frame, V^{t-1} denotes the vertices of the polygon and ε^{t-1} indicates the edges of the polygon. So as to determine the object contour in the current frame the motion estimation method is explained in below is applied. This approach is contained the motion vector of every vertex $v \in V^{t-1}$.

After that, design of the issue of motion estimation as the issue of decreasing the posterior probability. Assume D^t represents the set which integrated the motion vector d_v of every vertex v in V^{t-1}. The posterior probability of D^t is,

$$P(D^t|I^t, C^{t-1}) \propto P(I^t|D^t, C^{t-1})P(D^t|C^{t-1} \tag{2}$$

In the above equation, the I^t represents the image of frame t. After that, take the negative algorithm of the left and right sides of the equation (2) at the same time. Then, the issue of motion estimation becomes the issue of decreasing the posterior energy.

$$\underset{D^t}{\arg\max}\, P(D^t|I^t, C^{t-1}) = \underset{D^t}{\arg\min}\, E(I^t|D^t, C^{t-1}) + E(D^t|C^{t-1}) \tag{3}$$

In equation (3), the energy operator $E(.) = -\log P(.)$. The initial term $E(I^t|D^t, C^{t-1})$ on the right hand side is the likelihood energy determined from the observed pixel. It is utilized for examining the vertex motion and is described as follows,

$$E(I^t|D^t, C^{t-1}) = \omega_L L + \omega_G G \tag{4}$$

In the above equation, ω_L and ω_G denotes the weights of the coefficients L and G respectively. L is employed for computing the block difference among the current frame and the previous frame.

$$L = \sum_{v \in V^{t-1}} \sum_{p \in W_v^{t-1}} M(\propto_p) \left\| I^{t-1}(z_p) - I^t(z_p + d_v) \right\| \tag{5}$$

Where

$$M(\propto) = \begin{cases} 1, \text{if } \propto_p = F \\ 0, \text{Otherwise} \end{cases} \tag{6}$$

It is the mask function, W_v^{t-1} represents a 11 X 11 window in frame $t-1$ centered at the vertex v, z_p denotes the position of pixel p and z_p represents the position of pixel p and $\propto_p$ denotes the label of pixel p. The evaluation of L is a block matching process masked through the labels of the previous frame due to only the foreground object is tracked. After that, run the block matching process on Graphing Processing Unit (GPC) for accelerating its speed. For avoiding the contour from shrinking inward because of the mask function, another coefficient G is utilized for guiding the contour vertices to lay at the image edges. This coefficient is described as follows,

$$G = \sum_{v \in V^{t-1}} \exp(-\max_{c \in \{R, G, B\}} \|g_c^t(z_v + d_v)\|) \tag{7}$$

In the above equation, z_v represents the position of vertex v and g_c^t denotes the Sobel gradient of the RGB channel c of the image I^t.

The next term $E(D^t|C^{t-1})$ can be the prior energy that is independent of the observed pixels. We utilize it for managing the contour shape. It can be described as follows,

$$E(D^t|C^{t-1}) = \omega_F F + \omega_S S \tag{8}$$

In the above equation, the ω_F and ω_S are the weights of coefficients F and S respectively. F is utilized for penalizing the large vertex motion and constrain the contour velocity.

$$F = \sum_{v \in V^{t-1}} 1 - \exp[-(\|d_v\|^2 / (2\sigma_f^2))] \tag{9}$$

In the above equation, the σ_f represents the parameter associated with the variance of the variance motion. The coefficient S manages the shape variance through penalize the relative motion among neighboring vertices.

$$S = \sum_{\{v,u\} \in \breve{\varepsilon}^{t-1}} \exp[-(\|z_v - z_{uu}\|^2) / (2\sigma_d^2) \sqrt{\|d_v - d_u\|^2}] \tag{10}$$

In equation (10), $\breve{\varepsilon}^{t-1} = \varepsilon^{t-1} \cup \varepsilon_a^{t-1}$ denotes the edge set which consists of polygon edges ε^{t-1} and auxiliary edges ε_a^{t-1} and σ_d is a parameter associated with the variance of edge length. The auxiliary edges connected the preceding vertex and the succeeding vertex of every vertex. After that, utilize these edges for constraining the internal angle of every contour vertex and make the system more robust.

The adaptive MRF approach is decreased the posterior energy through assuming each vertex in the contour C^{t-1} as a node in an adaptive MRF system. The hidden state of every node corresponds to the motion vertex of its vertex. For instance, if the search range for every vertex is 5 X 5 after that every node has a state selected from a 25 element set $I = (l_1, l_2, \ldots\ldots\ldots l_{25})$ for indicating its motion vector. The posterior energy is described as follows,

$$E(D^t|I^t, C^{t-1}) = \omega_L L + \omega_G G + \omega_F F + \omega_S S \quad (11)$$

In the above equation, the first three terms represents the data energy and the last term denotes the link energy of the adaptive MRF system.

In Figure 1, at first, the user denotes the key frame of the input video and utilizes an interactive image segmentation approach (modified graph cut approach) and then the system automatically segmented the target object frame by frame. In the automatic segmentation process, the user can be allowed for interrupting the system and editing unsatisfied frames through modifying the object boundary by a boundary editing tool. Then, the user may continue the automatic segmentation for processing the remaining frame. The proposed approach consists of two components (contour tracking and OsDPMVL-IASMMRF segmentation). The object is predicted the contour of current frame through extracting the object contour of previous segmented frame after that estimating the motion of polygon vertices. The predicted contour is utilized as the primary constraint of the last component, where the modified graph cut approach is applied to the regions near the predicted contour.

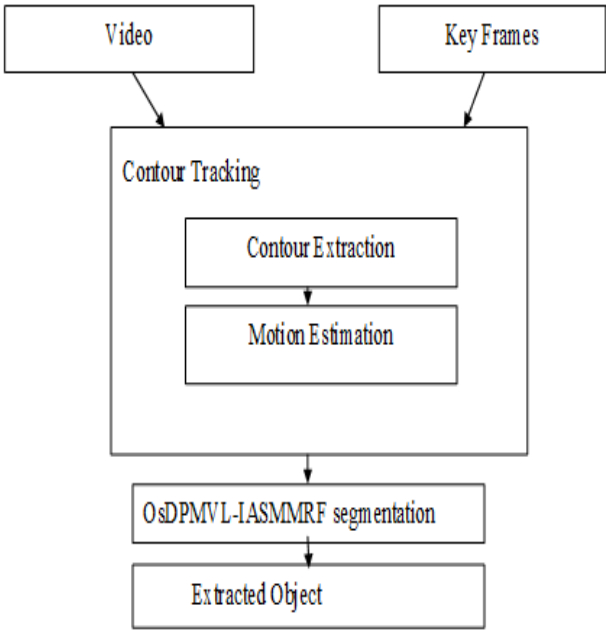

Fig. 1: Video Object Extraction.

4. Results and discussion

In this section, the experimental results are conducted on vide to evaluate the performance of the proposed OsDPMVL-IASMMRF-VOE model and existing MRF-based contour tracking and graph-cut image segmentation (MRFCT-GCS-VOE) model in terms of precision, accuracy and recall.

4.1. Dataset description

The dataset for video object extraction was taken Mexico Wildlife video gallery. The Brown Pelican video has been used for VOE. The frame size of the video is 1174 × 1086 .The video clipping time is 16 secs. The MATLAB 8.6 has been used to extract, totally 60 frames from this video. In Figure 2, the original video frame and extracted objects from video frames are shown in Figure 2. These segmentation procedures separate an object of interest from the remaining image region. The VOE is measured in terms of accuracy, precision and recall.

Fig. 2: Comparison Output of Segmentation Images.

4.2. Accuracy

It is computed the percentage of true positives (TPs) and true negatives among the total number of features clustered.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+False\ positive+False\ negative)}$$

The comparison of accuracy values is given in Table 1.

Table 1: The Accuracy Comparison of VOE (in %)

Number Of Frames	MRFCT-GCS-VOE	Osdpmvl-IASMMRF-VOE
15	92.53	94.31
30	93.21	95.67
45	95.42	96.22
60	96.86	98.79

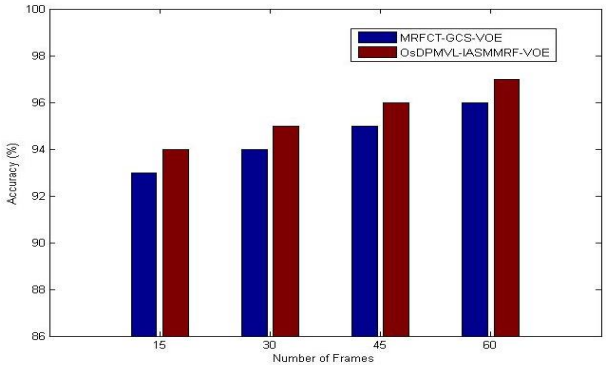

Fig. 3: Comparison Based on Accuracy.

In Figure 3, the comparison results of the OsDPMVL-IASMMRF-VOE and MRFCT-GCS-VOE in terms of accuracy. The result shows that the accuracy is increased for proposed OsDPMVL-IASMMRF-VOE model compared to existing MRFCT-GCS-VOE model.

4.3. Precision

It is computed to the clustering level at TP prediction, false positive.

$$Precision = \frac{TP}{(TP+False\ Positive)}$$

The comparison of precision values is given in Table 2.

Table 2: The Precision Comparison of VOE

Number of frames	MRFCT-GCS-VOE	OsDPMVL-IASMMRF-VOE
15	0.932	0.946
30	0.961	0.982
45	0.943	0.972
60	0.957	0.961

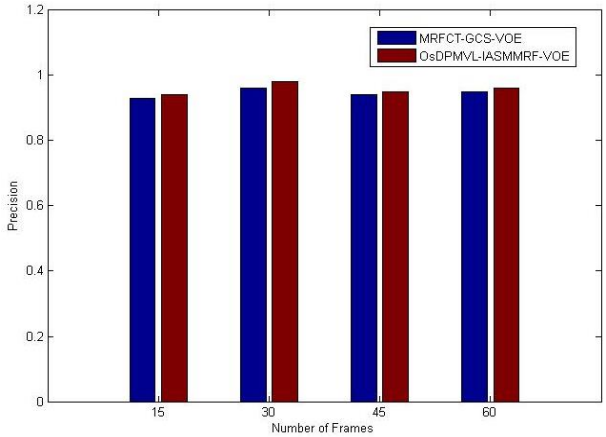

Fig. 4: Comparison Based on Precision.

Figure 4 shows that the comparison of OsDPMVL-IASMMRF-VOE and MRFCT-GCS-VOE models in terms of Precision. From the analysis, the number of video frames and the precision values are represented in X and Y-axis, respectively. The precision value of the proposed OsDPMVL-IASMMRF-VOE method obtained high precision values for the various numbers of frames compared to MRFCT-GCS-VOE models.

4.4. Recall

It value is measured to the clustering level at TP prediction, false negative.

$$Recall = \frac{TP}{(TP+False\ negative)}$$

The comparison of recall values is given in Table 3.

Table 3: The Recall Comparison of VOE

Number of frames	MRFCT-GCS-VOE	Osdpmvl-IASMMRF-VOE
15	0.945	0.962
30	0.974	0.982
45	0.928	0.941
60	0.971	0.992

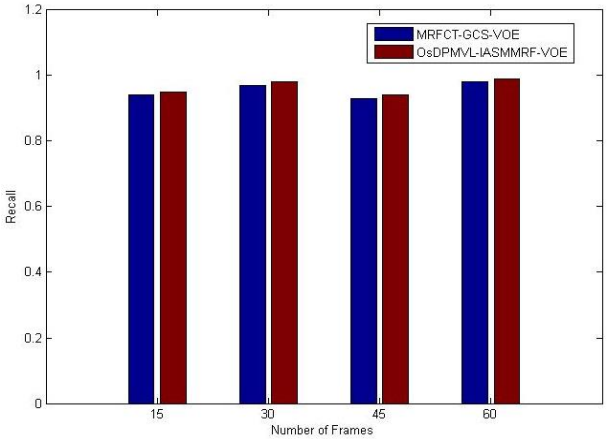

Fig. 5: Comparison Based on Recall.

The comparison of existing MRFCT-GCS-VOE model and proposed OsDPMVL-IASMMRF-VOE model for metric Recall is shown in Figure 5. From the graph, the no of frames used for object extraction s is denoted in X-axis and the recall values are indicated from Y-axis. From this graph, it shows that the proposed model has obtained high recall value compared to existing model.

5. Conclusion

In this paper, OsDPMVL-IASMMRF model is expanded for VOE. The image segmentation by this model is integrated contour tracking for VOE. The Teh–Chin algorithm is utilized with this model to predict the contour in the current by matching the extracted object contour from the previous segmented frame. The contour tracking propagates the shape of the target object, whereas the proposed model segmentation refined object boundary and the shape for enhancing the accuracy of video segmentation. The experimental results show that the proposed OsDPMVL-IASMMRF-VOE model provided better results in terms of recall, precision and accuracy.

References

[1] Bouwmans T (2014). Traditional and recent approaches in background modeling for foreground detection: An overview.*Computer Science Review*, *11*, 31-66. https://doi.org/10.1016/j.cosrev.2014.04.001.

[2] Vosters L, Shan C & Gritti T (2012). Real-time robust background subtraction under rapidly changing illumination conditions. *Image and Vision Computing*, *30*(12), 1004-1015. https://doi.org/10.1016/j.imavis.2012.08.017.

[3] Nikolov B & Kostov N (2014). Motion detection using adaptive temporal averaging method. *Radioengineering*, *23*(2), 652-658.

[4] Xue G, Sun J & Song L (2012). Background subtraction based on phase feature and distance transform. *Pattern Recognition Letters*, *33*(12), 1601-1613. https://doi.org/10.1016/j.patrec.2012.05.009.

[5] Fu Z & Han Y (2012). Centroid weighted Kalman filter for visual object tracking. *Measurement*, *45*(4), 650-655. https://doi.org/10.1016/j.measurement.2012.01.004.

[6] Weng SK, Kuo CM & Tu SK (2006). Video object tracking using adaptive Kalman filter. *Journal of Visual Communication and Image Representation*, *17*(6), 1190-1208. https://doi.org/10.1016/j.jvcir.2006.03.004.

[7] Zhang L, He X & Wang H (2012). Shadow Verification Based on Feature Matching and Image Matting.

[8] Ahuja K & Tuli P (2013). Object recognition by template matching using correlations and phase angle method.*International Journal of Advanced Research in Computer and Communication Engineering*, *2*(3), 1368-1373.

[9] Ruri S. Basuki, M. Hariadi, Eko M. Yuniarno, Mauridhi H. Purnomo. Spectral-Based Temporal-Constraint Estimation for Semi-Automatic Video Object Segmentation. International Review on Computers and Software (I.RE.CO.S.), Vol. 10, N. 9, 2015.

[10] ZHANG L, WANG H, DENG T & HE X (2014). Improving integrality of detected moving objects based on image matting. *Chinese Journal of Electronics*, *23*(4).

[11] Chung CY & Chen HH (2010). Video object extraction via MRF-based contour tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, *20*(1), 149-155. https://doi.org/10.1109/TCSVT.2009.2026823.

[12] Gowri GS & Ponmuthuramalingam P (2018). A SMOOTHED DPMVL FOR INTERACTIVE IMAGE SEGMENTATION AND ENHANCED ADAPTIVE MRF FOR SEGMENTATION REFINEMENT. PARIPEX - INDIAN JOURNAL OF RESEARCH, *7*(4).

[13] Teh CH & Chin RT (1989). On the detection of dominant points on digital curves. *IEEE Transactions on pattern analysis and machine intelligence*, *11*(8), 859-872. https://doi.org/10.1109/34.31447.

[14] Zhang XP & Chen Z (2006). An automated video object extraction system based on spatiotemporal independent component analysis and multiscale segmentation. *EURASIP Journal on Advances in Signal Processing*, *2006*(1), 045217. https://doi.org/10.1155/ASP/2006/45217.

[15] Lu Y & Li ZN (2008). Automatic object extraction and reconstruction in active video. *Pattern Recognition*, *41*(3), 1159-1172. https://doi.org/10.1016/j.patcog.2007.07.015.

[16] Mohammed US & Abd-Elhafiez WM (2009). A new video coding approach based on object-extraction. *International Journal of Video & Image Processing and Network Security IJVIPNS-IJENS*, *9*(10), 62-70.

[17] Manikandan R & Ramakrishnan R (2013). Video object extraction by using background subtraction techniques for sports applications. *Digital Image Processing*, *5*(9), 435-440.

[18] Wang H & Wang T (2016). Primary object discovery and segmentation in videos via graph-based transductive inference.*Computer Vision and Image Understanding*, *143*, 159-172. https://doi.org/10.1016/j.cviu.2015.11.006.

[19] Pun CM & Huang G (2016). On-line video object segmentation using illumination-invariant color-texture feature extraction and marker prediction. *Journal of Visual Communication and Image Representation*, *41*, 391-405. https://doi.org/10.1016/j.jvcir.2016.10.017.

[20] Hu YT, Huang JB & Schwing A (2017). MaskRNN: Instance Level Video Object Segmentation. In *Advances in Neural Information Processing Systems* (pp. 324-333).

[21] Kanagamalliga S & Vasuki S (2018). Contour-based object tracking in video scenes through optical flow and gabor features. *Optik-International Journal for Light and Electron Optics*, *157*, 787-797. https://doi.org/10.1016/j.ijleo.2017.11.181.