Moving object detection using Markov Random Field and Distributed Differential Evolution

Ashish Ghosh\textsuperscript{a,*}, Ajoy Mondal\textsuperscript{b}, Susmita Ghosh\textsuperscript{b}

\textsuperscript{a} Machine Intelligence Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India
\textsuperscript{b} Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India

\textbf{A R T I C L E   I N F O}

Article history:
Received 16 March 2012
Received in revised form 22 July 2013
Accepted 25 October 2013
Available online 9 November 2013

Keywords:
Markov Random Field
MAP estimation
Differential Evolution
Distributed Differential Evolution
Object detection

\textbf{A B S T R A C T}

In this article, we present an algorithm for detecting moving objects from a given video sequence. Here, spatial and temporal segmentations are combined together to detect moving objects. In spatial segmentation, a multi-layer compound Markov Random Field (MRF) is used which models spatial, temporal, and edge attributes of image frames of a given video. Segmentation is viewed as a pixel labeling problem and is solved using the maximum a posteriori (MAP) probability estimation principle; i.e., segmentation is done by searching a labeled configuration that maximizes this probability. We have proposed using a Differential Evolution (DE) algorithm with neighborhood-based mutation (termed as Distributed Differential Evolution (DDE)) algorithm for estimating the MAP of the MRF model. A window is considered over the entire image lattice for mutation of each target vector of the DDE; thereby enhancing the speed of convergence. In case of temporal segmentation, the Change Detection Mask (CDM) is obtained by thresholding the absolute differences of the two consecutive spatially segmented image frames. The intensity/color values of the original pixels of the considered current frame are superimposed in the changed regions of the modified CDM to extract the Video Object Planes (VOPs). To test the effectiveness of the proposed algorithm, five reference and one real life video sequences are considered. Results of the proposed method are compared with four state of the art techniques and provide better spatial segmentation and better identification of the location of moving objects.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Detection and tracking of moving objects from a given video sequence are challenging tasks in video processing and have been an active research area for the last few decades [1,2]. It has wide applicability in video surveillance, event detection, anomaly detection, dynamic scene analysis, activity recognition, activity based human recognition [1,2] etc. To detect moving objects for a given video sequence, various object regions which are moving with respect to their background are identified [2]. It can be performed using (i) change/motion detection and (ii) motion estimation [2] and the detected objects, in turn, can be tracked. Tracking provides the velocity, acceleration and position of moving objects at different time instants.

Literature on detection of moving objects using different stochastic models is quite rich. Markov Random Field (MRF) model is one such approach which represents spatial continuity among adjacent pixels and is robust to degradation. Since the last few decades, MRF models [3–7] and Hidden MRF models [8] are used for segmenting images and these models are quite popular for segmenting video sequences [9–20]. In [9,21], to detect moving objects, MRF model was considered in the spatial direction only. Whereas, in a video, two image frames are not independent of each other. Hence, temporal coherence can be explored to obtain better segmented output. Considering this temporal coherence, a multi-layer MRF modeling is adhered to in [10–16] (one frame in temporal direction). Kim and Park [11] showed that a combination of spatial segmentation and temporal segmentation provides better results for detecting moving objects. To preserve the boundary of the objects, Sududhi et al. [12] proposed a multi-layer compound MRF model which takes care of the spatial distribution of color, temporal color coherence, and edge map/line-field in the temporal frames (two frames in temporal direction) to obtain a spatial segmentation. Temporal coherence helps in preserving the region information in segmentation. The use of edge map/line-field can preserve the accurate object boundary in segmentation.

To the best of our knowledge, in the literature no method is available which exploits the merits of Differential Evolution (DE) scheme with MRF model. In the present article, the applicability of DE has been explored with MRF model for moving object detection. A preliminary experiment of this work was reported in [17].
A robust analysis of the results with application in different complex video sequences is reported in the present article. Comparison with recent state-of-the-art techniques along with its quantitative analysis has also been made in the present version.

The present work considers, a spatial segmentation scheme where multi-layer compound MRF model [12] has been used for detecting moving objects in a given video sequence. Spatial segmentation as well as temporal segmentation are combined for detecting moving objects. The assumed MRF model takes care of spatial distribution of color, temporal color coherence, and edge map/line-field in temporal direction (two frames in the temporal direction) of a given video frame. The considered MRF model is found to preserve accurate object boundary. The spatial segmentation using the considered multi-layer compound MRF model is equivalent to a pixel labeling problem and is solved using the MAP estimation principle. A Distributed Differential Evolution (DDE) algorithm is proposed for searching the maximum a posteriori probability of MRF modeled frame. In the proposed search technique, a population represents a segmented output of the considered video frame and size of the population is equal to the dimension of the video frame. A population consists of a set of target vectors. Here, each target vector encodes the red (R), green (G), and blue (B) components of the segmented output and a set of such target vectors constitutes each population. The population is initialized randomly. It is to be noted that the segmentation of the MRF modeled frame corresponds to the MAP estimation and the proposed DDE algorithm maximizes this probability. This maximization process has been done iteratively by evolving the target vectors (i.e., possible solutions) of the population using the following three operations: mutation, crossover, and selection. The algorithm continues until some stopping criterion is reached. At the end of the evolutionary process, a stable labeled configuration is obtained. Such a configuration is considered as the required segmented output of the given input video frame.

In the proposed DDE, to mutate each target vector, instead of considering the whole population, a small window centered at the target vector is considered thereby reducing computational time as compared to the conventional Differential Evolution algorithm. It also provides a better spatial segmentation as the considered window maintains the spatial regularity of the lattice in MRF.

In temporal segmentation, the difference images of each of the R, G, and B channels are obtained by taking absolute difference of the respective R, G, and B components of the two consecutive spatially segmented image frames. Thereafter, Change Detection Masks (CDMs) of each of the R, G, and B channels are obtained by thresholding the corresponding difference images. The changed and the unchanged regions in the current frame are obtained by considering the union of CDMs of these three channels. To obtain the exact location of the moving objects in the current frame, we modify the obtained CDM by considering the information of pixels belonging to the moving objects in the previous frame and result of the corresponding current spatial segmentation. Video Object Plane (VOP) is extracted by superimposing the intensity/color values of original pixels of the current frame on the changed regions of the modified CDM.

To test the effectiveness of the proposed algorithm investigation was carried out on five reference and one real life video sequences. Results obtained by the proposed method are compared with those of MRF with DGA scheme [10,11], MRF with DGA scheme with multi-layer compound MRF model as used in the proposed method (henceforth, called as modified MRF with DGA scheme), MRF with Graph Cut scheme [7,22], and MRF with SA and ICM scheme [12].

Organization of the article is as follows. Section 2 describes the related work. In Section 3, the proposed methodology for moving object detection where spatial segmentation (spatio-temporal MRF based image modeling, MAP estimation framework for MRF modeling), DE algorithm, DE algorithm with neighborhood-based mutation, the proposed DDE algorithm for MAP estimation, segmentation of subsequent frames based on change information, and temporal segmentation are described. Experimental results and discussion are given in Section 4 and conclusive remarks are put in Section 5.

2. Related work

For change/motion detection initially a reference frame is considered where no objects are present [2]. Detection of moving objects becomes very difficult using temporal segmentation if the reference frame is absent or movement of objects is either very slow, or very fast, or they halt for some time and then move again [1,2].

In order to solve these problems a region based robust video frame segmentation algorithm is needed [1,2]. Salembier and Marques [23] had proposed a computationally efficient watershed based spatial segmentation scheme. To detect the object’s boundary, they have used both spatial and temporal segmentations. But this scheme sometimes produces over-segmented results.

Image segmentation using MRF model can be achieved by estimating maximum a posteriori (MAP) probability of the concerned model [3,10,11,24]. For a given observed image or a video frame Y, segmentation is obtained by searching a configuration X which maximizes this probability. This maximization process is quite complex owing to the large search space. Hence, efficient and effective optimization algorithms are required to maximize the posterior probability.

The MAP of the MRF model has been estimated using various optimization algorithms such as Simulated Annealing (SA) [4,24,25], Iterated Conditional Mode (ICM) [3,24], and Genetic Algorithm (GA) [5,24]. Though the convergence speed of SA is less but it has the capability of finding better solutions [5,24]. On the other hand, ICM consumes less time as compared to other optimization techniques; but sometimes it may get stuck to local optima [5,24]. In [12], Subudhi et al. has proposed an MRF model based spatial segmentation scheme by combining both SA and ICM to estimate the MAP of the corresponding model. It has the ability to provide comparable results quickly. In [12], the SA was initially executed for a few iterations to get a suboptimal solution and thereafter ICM (with suboptimal solution as initialization) was used to obtain the optimal solution. As SA is executed for a few iterations, the suboptimal solutions obtained using SA sometimes correspond to the local optima or nearer to the local optima and thereafter ICM may get stuck to the same and may not reach the global optimum or near to the global optimum. Hence, segmentation using the same algorithm may not be a good choice. Moreover, the setting of the number of iterations in SA, is not easier.

GA was used as an alternative optimizer to SA and ICM in [5] due to suitable values of the parameter, its ability to explore and exploit the search space quickly using the concept of natural evolution and selection [26]. It is robust for identifying solutions to combinatorial optimization problems. In [5,26,27], it is mentioned that the GA sometimes converges to a premature solution and has low searching speed. To overcome these problems, GA has been modified in [28,29]. In [28], a deterministic hill-climbing procedure was embedded into GA. Although the computational cost was low, it still landed up to a suboptimal solution. Zhang et al. developed evolutionary algorithm based on a combination of GA and SA [30,31]. These algorithms produce better segmentation than those obtained using GA alone, but had a lower convergence speed. A combination of ICM and GA was proposed by Lai and co-worker [5]. For quick convergence and to obtain better segmentation, a new version of GA called, Distributed Genetic Algorithm (DGA) has been proposed.
for segmenting textured images in [32] and video sequences in [10,11]. In DGA [11], the chromosome with the highest fitness value among its neighborhood was selected. During crossover, a neighboring chromosome within the window was randomly picked up and a component of the feature vector was chosen and replaced by the corresponding value of the feature vector of the selected chromosome. For this reason, DGA merges smaller regions to a large region and produces over-segmented results at the boundary of an object.

Unlike evolutionary approaches, some deterministic methods such as Belief Propagation (BP) and Graph Cuts (GC) have been used for MAP estimation. In [14], Yin and Collins had presented BP approach for moving object detection using a 3D MRF model. In this method, each hidden state in the 3D MRF model represented a pixel’s motion likelihood and was estimated using message passing in a 6-connected spatio-temporal neighborhood. Huang et al. [15] had proposed a framework for segmenting moving objects using a new spatio-temporal MRF model and Loopy Belief Propagation (LBP) was used to estimate MAP of MRF model. In [22], Szeliski et al. studied different energy minimization methods such as ICM, swap-move, expansion-move, max-product LBP, and Tree-Reweighted Message passing (TRW) for stereo, image stitching, interactive segmentation, and denoising. The authors mentioned that the TRW and expansion-move GC algorithms achieve minimum energy. Feng and Jian-ya [7] had proposed a novel MRF based image segmentation scheme where MAP of MRF model was estimated using expansion-move algorithm in GC. However, these approaches [7,14,15,22] are found to provide over-segmented results.

3. Proposed method for moving object detection

Fig. 1 represents the block diagram of the proposed method. As mentioned earlier, in the present article spatial and temporal segmentations are integrated together to detect moving objects. Spatial segmentation accurately provides the boundary of the regions in a given scene, whereas temporal segmentation determines the changed and the unchanged regions. MRF model with spatio-temporal framework is used for spatial segmentation as it was used in [12]. Color features, both in spatial and temporal directions, and edge map/line-field in temporal direction are taken into consideration while modeling the image frame. The edge map has been obtained using a $3 \times 3$ Laplacian mask. Segmentation using the MRF model is usually done through the MAP estimation principle. Using a local mutation based DE termed as Distributed Differential Evolution (DDE) is proposed to estimate the MAP of such a model. In DDE unlike conventional DE algorithm, a small window is considered over the image lattice for mutating each of the target vectors. Random initialization is done for segmenting the initial frame. A change information based initialization is utilized for segmenting the subsequent frames.

In temporal segmentation, the difference images for each of the R, G, and B channels are obtained by computing the absolute differences of the corresponding R, G, and B components of two consecutive spatially segmented image frames. Then, the CDMs of each of the R, G, and B channels are obtained by thresholding the corresponding difference images. These CDMs represent the changed and the unchanged regions in the respective channel. Union of CDMs of these three channels is taken to obtain the changed and the unchanged regions in the current frame. The changed regions correspond to moving objects whereas the unchanged regions correspond to stationary objects and background. To obtain the exact locations of the moving objects, the CDM is modified. Information of the pixels belonging to the moving objects in the previous frame and output of the current spatial segmentation are considered to modify it. The intensity/color values of the original pixels of the current frame are then superimposed on the changed regions of the modified CDM to extract the VOP.

3.1. Spatial segmentation

Spatial segmentation can be done by combining the spatial and temporal features of a given video image frame in a single framework [11,33]. In the present article, MRF (with both spatial and temporal features) based image modeling is considered for spatial segmentation.

3.1.1. Spatio-temporal MRF based image modeling

Here, the observed video sequence $y$ is assumed to be a 3D volume, consisting of spatio-temporal image frames. The $t$th frame of the video sequence $y$ is represented as $y_t$. Each pixel of $y_t$ is a spatial site, denoted by $y_{st}$, $q$ is a temporal site, and $e$ is an edge site. Let $S$ be the set of all the sites. Thus $S = \{s\} \cup \{q\} \cup \{e\}$. Let $Y_t$ represent a random field and $y_t$ be a realization of it at time $t$. Thus, $y_{st}$ denotes a pixel at site $s$ at time $t$. Let $x$ denote the segmentation of video sequence $y$ and $X_t$ the segmented version of $y_t$. Let $X_t$ be the MRF and $x_t$ represents one of its realization. In the present investigation, the second order MRF modeling in spatial as well as in temporal directions are considered to take care of the spatial distribution of color and temporal color coherence, respectively. Another MRF model (with the edge/line-field of the current frame $x_t$ and the edges/line-fields of the $(t-1)$th frame $x_{t-1}$ and the $(t-2)$th frame $x_{t-2}$) is considered to preserve the accurate object boundary in segmentation. As $X_t$ is an MRF, it satisfies the following Markovian property:

$$P(X_{st} = x_{st} | X_q = x_q, \forall q \in S, s \neq q) = P(X_{st} = x_{st} | X_{qt} = x_{qt}, (q,t) \in \eta_{st}).$$
where $\eta_{st}$ denotes the neighborhood of site $s$ at time $t$. For temporal MRF, the following Markovian property is also satisfied:

$$P(X_{st} = x_{st}|X_{qr} = x_{qr}, s \neq q, t \neq r, \forall (s, t), (q, r) \in V)$$

$$= P(X_{st} = x_{st}|X_{qr} = x_{qr}, s \neq q, t \neq r, (q, r) \in \eta_{st}).$$

Here $V$ represents the 3D volume of the video sequence and $\eta_{st}$ denotes the neighborhood of $(s, t)$ in temporal direction.

The novelty of the MRF model lies in the fact that it takes into account the spatial and temporal characteristics of a region [24]. The regions are assumed to have uniform intensities. The considered image frame is assumed to come from an imperfect imaging modality. It is also assumed that some noise has corrupted the actual image to produce the observed image $y_t$. Given this observed image $y_t$, our aim is to find out the image $x_t$, which maximizes the posterior probability. An estimate of $x_t$ is denoted by $\hat{x}_t$. Hence, it is assumed that due to the presence of noise $x_t$ cannot be observed properly. But a noisy version of $x_t$ (i.e., $y_t$) is observed. Noise is assumed to be white (additive i.i.d., i.e., independent and identically distributed). Hence $y_t$ can be expressed as [24]

$$y_t = x_t + N(0, \sigma^2),$$

where $N(0, \sigma^2)$ is the i.i.d. noise with zero mean and $\sigma^2$ variance.

Fig. 2 represents the 2nd order MRF modeling. In Fig. 2(a), the $(i, j)$th pixel with its spatial neighbors are used to construct the

**Fig. 2.** (a) MRF model in spatial direction, (b) MRF model in temporal direction, and (c) MRF model with line-field in temporal direction.

where $\eta_{st}$ denotes the neighborhood of site $s$ at time $t$. For temporal MRF, the following Markovian property is also satisfied:

$$P(X_{st} = x_{st}|X_{qr} = x_{qr}, s \neq q, t \neq r, \forall (s, t), (q, r) \in V)$$

$$= P(X_{st} = x_{st}|X_{qr} = x_{qr}, s \neq q, t \neq r, (q, r) \in \eta_{st}).$$

Here $V$ represents the 3D volume of the video sequence and $\eta_{st}$ denotes the neighborhood of $(s, t)$ in temporal direction.

The novelty of the MRF model lies in the fact that it takes into account the spatial and temporal characteristics of a region [24]. The regions are assumed to have uniform intensities. The considered image frame is assumed to come from an imperfect imaging modality. It is also assumed that some noise has corrupted the actual image to produce the observed image $y_t$. Given this observed image $y_t$, our aim is to find out the image $x_t$, which maximizes the posterior probability. An estimate of $x_t$ is denoted by $\hat{x}_t$. Hence, it is assumed that due to the presence of noise $x_t$ cannot be observed properly. But a noisy version of $x_t$ (i.e., $y_t$) is observed. Noise is assumed to be white (additive i.i.d., i.e., independent and identically distributed). Hence $y_t$ can be expressed as [24]

$$y_t = x_t + N(0, \sigma^2),$$

where $N(0, \sigma^2)$ is the i.i.d. noise with zero mean and $\sigma^2$ variance.

**Fig. 2.** (a) MRF model in spatial direction, (b) MRF model in temporal direction, and (c) MRF model with line-field in temporal direction.

where $\eta_{st}$ denotes the neighborhood of site $s$ at time $t$. For temporal MRF, the following Markovian property is also satisfied:

$$P(X_{st} = x_{st}|X_{qr} = x_{qr}, s \neq q, t \neq r, \forall (s, t), (q, r) \in V)$$

$$= P(X_{st} = x_{st}|X_{qr} = x_{qr}, s \neq q, t \neq r, (q, r) \in \eta_{st}).$$

Here $V$ represents the 3D volume of the video sequence and $\eta_{st}$ denotes the neighborhood of $(s, t)$ in temporal direction.

The novelty of the MRF model lies in the fact that it takes into account the spatial and temporal characteristics of a region [24]. The regions are assumed to have uniform intensities. The considered image frame is assumed to come from an imperfect imaging modality. It is also assumed that some noise has corrupted the actual image to produce the observed image $y_t$. Given this observed image $y_t$, our aim is to find out the image $x_t$, which maximizes the posterior probability. An estimate of $x_t$ is denoted by $\hat{x}_t$. Hence, it is assumed that due to the presence of noise $x_t$ cannot be observed properly. But a noisy version of $x_t$ (i.e., $y_t$) is observed. Noise is assumed to be white (additive i.i.d., i.e., independent and identically distributed). Hence $y_t$ can be expressed as [24]

$$y_t = x_t + N(0, \sigma^2),$$

where $N(0, \sigma^2)$ is the i.i.d. noise with zero mean and $\sigma^2$ variance.

**Fig. 2.** (a) MRF model in spatial direction, (b) MRF model in temporal direction, and (c) MRF model with line-field in temporal direction.

where $\eta_{st}$ denotes the neighborhood of site $s$ at time $t$. For temporal MRF, the following Markovian property is also satisfied:

$$P(X_{st} = x_{st}|X_{qr} = x_{qr}, s \neq q, t \neq r, \forall (s, t), (q, r) \in V)$$

$$= P(X_{st} = x_{st}|X_{qr} = x_{qr}, s \neq q, t \neq r, (q, r) \in \eta_{st}).$$

Here $V$ represents the 3D volume of the video sequence and $\eta_{st}$ denotes the neighborhood of $(s, t)$ in temporal direction.

The novelty of the MRF model lies in the fact that it takes into account the spatial and temporal characteristics of a region [24]. The regions are assumed to have uniform intensities. The considered image frame is assumed to come from an imperfect imaging modality. It is also assumed that some noise has corrupted the actual image to produce the observed image $y_t$. Given this observed image $y_t$, our aim is to find out the image $x_t$, which maximizes the posterior probability. An estimate of $x_t$ is denoted by $\hat{x}_t$. Hence, it is assumed that due to the presence of noise $x_t$ cannot be observed properly. But a noisy version of $x_t$ (i.e., $y_t$) is observed. Noise is assumed to be white (additive i.i.d., i.e., independent and identically distributed). Hence $y_t$ can be expressed as [24]

$$y_t = x_t + N(0, \sigma^2),$$

where $N(0, \sigma^2)$ is the i.i.d. noise with zero mean and $\sigma^2$ variance.
determined during execution. Hence, in the priori image model, the clique potential function is a combination of the above mentioned three terms and the energy function takes the form

$$U(X_t) = \sum_{c \in C_0} V_{cpl}(x_t, x_q) + \sum_{c \in C_1} V_{cpl}(x_{st}, x_q) + \sum_{c \in C_2} V_{cpl}(x_{st}, x_{tr}) \tag{2}$$

In Eq. (2), \( C \) is the set of cliques (here, 2nd order cliques are considered). Three clique potentials \( V_{cpl}(x_t, x_q), V_{cpl}(x_{st}, x_q), \) and \( V_{cpl}(x_{st}, x_{tr}) \) over all possible cliques are added to obtain the energy function \( U(X_t) \).

### 3.1.2. MAP estimation framework for MRF modeling

In MRF modeling, the observed image sequence \( y \) is considered as a degraded version of the actual image sequence \( x \). For example, at \( t \)th instant of time, the observed image frame \( y_t \) is considered as a degraded version of the true labeled \( x_t \). It is already mentioned that the spatial segmentation using MRF model considers the principle of MAP estimation and it is done by maximizing the posterior probability given in Eq. (3).

$$\hat{x}_t = \arg \max_{x_t} P(X_t = x_t | Y_t = y_t), \tag{3}$$

where, \( \hat{x}_t \) is an estimated label. Using Bayes’ theorem, Eq. (3) can be written as

$$\hat{x}_t = \arg \max_{x_t} \frac{P(Y_t = y_t | X_t = x_t) P(X_t = x_t)}{P(Y_t = y)} \tag{4}$$

Since the prior probability \( P(Y_t = y_t) \) is constant, Eq. (4) can be reduced to

$$\hat{x}_t = \arg \max_{x_t} P(Y_t = y_t | X_t = x_t, \theta) P(X_t = x_t, \theta), \tag{5}$$

where \( \theta \) is the parameter vector associated with \( x_t \). According to Hammersley Clifford theorem [24], the prior probability follows Gibbs’s distribution as follows

$$P(X_t = x_t) \propto e^{-U(x_t)} \tag{6}$$

This model is an edge based MRF model as used in [12]. The corresponding likelihood function \( P(Y_t = y_t | X_t = x_t) \) is as follows:

$$P(Y_t = y_t | X_t = x_t) = P(Y_t = y_t | x_t, \theta) = P(N = y_t - x_t | x_t, \theta),$$

Here, \( n \) is the realization of the Gaussian noise \( N(0, \sigma^2) \). Assuming decorrelation among the three RGB planes for the color image and the variance to be the same among each plane [34], Eq. (7) can be expressed as

$$P(N = y_t - x_t | x_t, \theta) = \frac{1}{\sqrt{(2\pi)^3 \sigma^3}} e^{-[(y_t - x_t) / (2\sigma^2)]^2 / 2}. \tag{8}$$

Using Eqs. (6)–(8), Eq. (5) can be rewritten as

$$\hat{x}_t = \arg \min_{x_t} \left\{ \frac{1}{\sqrt{(2\pi)^3 \sigma^3}} \left[ |y_t - x_t|^2 / 2\sigma^2 \right] + \sum_{c \in C_0} V_{cpl}(x_{st}, x_q) + V_{cpl}(x_{st}, x_q) + V_{cpl}(x_{st}, x_{tr}) \right\}, \tag{9}$$

where \( \hat{x}_t \) is the MAP estimate. The MAP of MRF model is estimated by the proposed Distributed Differential Evolution algorithm. In the following sections, Differential Evolution, Differential Evolution with neighborhood-based mutation, and the proposed Distributed Differential Evolution are briefly described.

### 3.2. Differential Evolution (DE) algorithm

Differential Evolution (DE) [35,36] is a parallel stochastic search method which uses a number of parameter vectors, called the target vectors, as a population in each generation. The parameter vectors of initial population are chosen randomly. DE generates new parameter vectors (called, mutant vectors) by adding the weighted difference of the two parameter vectors to a third vector. This operation is called mutation. With this operation, DE explores the search space. Thereafter, crossover operation is applied to increase the diversity of the mutated vectors. In crossover operation, parameters of the mutated vector and those of the target vector (a predetermined one) are swapped, based on crossover probability, to yield a trial vector. During selection process, if the trial vector yields better objective function than the target vector, then the trial vector replaces the target vector for the next generation.

DE searches for a global optimum solution in a \( d \)-dimensional real parameter space \( \mathbb{R}^d \). It begins with a randomly initiated population consisting of \( N \) numbers of \( d \)-dimensional real valued parameter vectors at a generation. Each parameter vector forms a candidate solution to the multidimensional optimization problem. Let at generation \( G \), the population consists of \( N \) numbers of \( d \)-dimensional parameter vectors

$$\tilde{x}_t, G = (x_{t,1}, G, x_{t,2}, G, \ldots, x_{t,d}, G), \tag{10}$$

where \( i = 1, 2, \ldots, N \). The initial parameter vectors of the population (at \( G = 0 \)) is chosen randomly and should cover the entire parameter space constrained by minimum and maximum bounds: \( \bar{x}_{min} = \left( x_{min,1}, x_{min,2}, \ldots, x_{min,d} \right) \) and \( \bar{x}_{max} = \left( x_{max,1}, x_{max,2}, \ldots, x_{max,d} \right) \). Hence we may initialize the \( j \)th component of the \( i \)th vector as:

$$x_{i,j} = x_{min,j} + rand_{ij} (x_{max,j} - x_{min,j}), \tag{11}$$

where \( j = 1, 2, \ldots, d \) and \( rand_{ij} \) is a uniformly distributed random number lying between 0 and 1.

#### 3.2.1. Mutation

In DE, parameter vector at the current generation is called target vector. For each target vector \( \tilde{x}_t \), at generation \( G \), for \( i = 1, 2, \ldots, N \), a mutant vector is generated as

$$\tilde{v}_t, G + 1 = \tilde{x}_i, G + F \times (\tilde{x}_{r_2}, G - \tilde{x}_{r_3}, G), \tag{12}$$

The indices \( r_1, r_2, \) and \( r_3 \) are mutually exclusive integers and randomly chosen from the range \([1, N]\). The chosen indices \( r_1, r_2, \) and \( r_3 \) are different from the index \( i \). The difference of two target vectors \( \tilde{x}_{r_2}, G - \tilde{x}_{r_3}, G \) is amplified by the scaled factor \( F \in [0, 2] \).
3.2.2. Crossover

It is used to increase the diversity of the population. The mutated vector generated through the mutation operation exchanges its parameters (components) with the target vector to obtain trial vector. The trial vector at generation \(G + 1\) is formed as:

\[
\vec{u}_{i,j} \equiv (G + 1) = (u_{i,1}, (G + 1), u_{i,2}, (G + 1), \ldots, u_{i,d}, (G + 1)).
\]

where

\[
\begin{align*}
    \vec{u}_{i,j} \equiv (G + 1) = & \begin{cases} 
    v_{i,j}, (G + 1) & \text{if } rand_{i,j} \leq CR \text{ or } j = j_{\text{rand}}, \\
    x_{i,j}, G & \text{otherwise}
    \end{cases}
\]
\]

where \(j = 1, 2, \ldots, d\) and \(j_{\text{rand}} \in [1, d]\) is a randomly chosen index, which ensures that \(u_{i,j}(G + 1)\) gets at least one component from the mutant vector \(\vec{u}_{i}, (G + 1). CR \in [0, 1]\) is the crossover probability.

3.2.3. Selection

The selection operation is used to decide whether the target or the trial vector survives for the next generation \((G + 1)\). The trial vector \(\vec{x}_{i,j} \equiv (G + 1)\) is compared to the target vector \(x_{i,j}\), using some objective function. If the vector \(\vec{x}_{i,j} \equiv (G + 1)\) yields better (objective) functional value than \(x_{i,j}\), then the vector \(\vec{x}_{i,j} \equiv (G + 1)\) goes to the next generation \((G + 1)\) as a member of the population; otherwise \(\vec{x}_{i,j}\). G. The selection operation is described as:

\[
\begin{align*}
    \vec{x}_{i,j} \equiv (G + 1) = & \begin{cases} 
    \vec{u}_{i,j}, (G + 1) & \text{if } f(\vec{u}_{i}) \leq f(\vec{x}_{i,j}), \\
    \vec{x}_{i,j}, G & \text{otherwise}
    \end{cases}
\]
\]

where \(f(\vec{x}_{i,j})\) is the objective function to be minimized.

3.3. Differential Evolution (DE) algorithm with neighborhood-based mutation

The efficiency of most of the Evolutionary Algorithms (EAs) depends on their exploration and exploitation ability during searching. Using exploitation, the searching algorithm exploits information and moves toward better solutions; whereas exploration allows to introduce new information into the population and helps to search new regions [36]. In DE, three vectors are randomly chosen from the whole search space to mutate each target vector. Sometime, such mutation mechanism delays the convergence and may get stuck to local optima. In literature, some modifications of DE have been done to increase the convergence speed and to get better solutions. In such a context, Tasoulis et al. [37] proposed a topological index based neighborhood DE where each population is divided into different sub-populations. Each sub-population evolves independently toward a solution. The best individual of each sub-population is moved to other sub-populations according to ring topology. The best individuals of each sub-population are allowed to migrate to the next sub-population of the ring. On the other hand Das et al. [38] proposed DE with neighborhood-based mutation to achieve better balancing of both exploration and exploitation ability of DE. The authors have proposed two kinds of neighborhood model for mutation of DE. The first one is the local neighborhood model where each vector is mutated using the best position found so far in a small neighborhood of it. For every target vector, a neighborhood of radius \(k\) is defined. They assumed that the vectors within the neighborhood can be organized on a ring topology with respect to their indices. For each target vector, the mutant vector is generated using combination of the best vector in the neighborhood and two other vectors chosen from the same neighborhood. On the other hand, the second one is referred to as the global mutation model where the globally best vector of the entire population is used to generate the mutant vector. They combined local and global mutant vectors using scalar weight to generate actual mutant vector. Instead of index based neighborhoods, few works using distance based neighborhood have also been proposed for single global optimization. Epitropakis et al. [39] proposed a proximity based mutation operator which selects the vectors to perform mutation operation based on distance related probability.

All the above discussed methods are for finding solutions of single objective functions. To solve multi-modal optimization problems, Qu et al. [40] proposed a neighborhood mutation strategy and integrated it with various niching DE algorithm. For each target vector, mutation is performed within its Euclidean neighborhood. The proposed method is able to maintain the multiple optima found during the evolution and evolve toward global/local optimum. Decomposition based multi-objective evolutionary algorithm decomposes the approximation of the Pareto front into a number of single objective optimization sub-problems. Each sub-problem is optimized using information from its neighboring sub-problems. In such cases, the neighborhood size (NS) plays a major role. In this regard, Zhao et al. [41] proposed an ensemble of different NSs and dynamically adjust their selection probabilities based on their previous performance.

From the above discussion we may conclude that in DE algorithms, each parameter (target) vector of the population is considered to be a solution. It also may be noted that all the local neighborhood based DE algorithms have defined the neighborhood either depending on the index of target vectors (ring topology) or Euclidean distances among them. However use of such an approach directly for image/video frame segmentation rarely gives satisfactory results. It may produce a segmentation result with many number of isolated pixels in a region. Hence, it is found not to provide a context based segmentation result that maintain the spatial regularity. In an image/video frame, belongingness of pixels to a particular class is more likely to be the same as that of its neighboring pixels. Hence, this motivated us to include the neighboring pixels (defined by a window or mask) at a particular pixel location of an image for mutation in DE. Please note that the target vectors present inside a window may not be topologically close. Hence, we modified the use of DE algorithm with neighborhood-based mutation for image/video frame segmentation and have termed it Distributed Differential Evolution (DDE) algorithm.

3.4. Proposed Distributed Differential Evolution (DDE) algorithm for the MAP estimation

In conventional DE algorithm, to mutate each target vector, three vectors are randomly selected from the whole population (total search space). Using the mutation operation, it explores new search regions. DE needs large computational time for convergence. Since the mutated vector may not lie within its neighborhood, the solution might not have spatial regularity. Hence, its use for video image segmentation might not carry proper contextual information.

In this article, we have proposed using a DE algorithm with neighborhood-based mutation and called it Distributed Differential Evolution (DDE) algorithm. In the proposed DDE, each population is divided into a number of sub-populations (overlapping \(w \times w\) windows). To mutate each target vector, a small window (sub-population) centered at the target vector is considered instead of considering the whole population. This mechanism accelerates the searching speed of the DE algorithm. As the considered window maintains the spatial regularity of the MRF model, the proposed DDE algorithm is expected to provide better segmentation.

In the proposed DDE based search procedure a population corresponds to a segmented output of the considered video frame and
size of the population is equal to the dimension of the video frame. Each population (same as that of conventional DE) consists of a set of target vectors. Here, each target vector $\mathbf{x}_{st}$ ($x^G_{st}$, $x^C_{st}$, and $x^K_{st}$) encodes the R, G, and B channels of a pixel of the segmented output. Initial population is chosen randomly. Here, it is also noted that the segmentation of the MRF modeled frame is achieved by estimating the MAP of the said model. The proposed DDE algorithm maximizes this probability by iteratively evolving the target vectors (a possible solution) of the population using three operations: mutation, crossover, and selection until some stopping criterion is satisfied. At the end of this evolutionary process, a stable labeled configuration is obtained. Such a configuration is considered as the required segmented output of the given input video frame. Fitness of the target vector is inversely proportional to the cost function. The cost function of a target vector is defined in terms of its local energy and formulated as

$$E_{st} = \frac{1}{2\sigma^2} \sum_{i \in c} (|y_{st_i} - \bar{y}_{st_i}|^2) + \sum_{c \in c} |V_{cst}(\bar{x}_{st}, \bar{x}_{qt}) + V_{cst}(\bar{x}_{st}, \bar{x}_{qr})$$

$$+ V_{cst}(\bar{x}_{st}, \bar{x}_{et})].$$

The energy values of all the target vectors are summed up to obtain the energy value of the population. The population having a lower energy value is considered as a better solution. Each target vector is then evolved by mutation, crossover, and selection (thereby yielding new target vectors) operations through a number of generations/iterations. Mutation, crossover, and selection operators of the proposed DDE model and its stopping criterion are briefly described in the following sections.

### 3.4.1. Mutation

Using mutation operation, a mutant vector is generated by adding the weighted difference between the two vectors to a third vector. The search space is explored quickly using this operation. A mutant vector corresponding to sth target vector $\mathbf{x}_{st}(G)$ (where $s = 1, 2, \ldots, (M \times N)$) of the population at generation $G$ is generated according to the following equation

$$\bar{u}_{st}(G + 1) = G_{st}(G) + F \times (\bar{x}_{qt}(G) - \bar{x}_{st}(G)),$$

where $\bar{x}_{qt}$, $\bar{x}_{st}$, and $\bar{x}_{et}$ are the three randomly chosen vectors from the $w \times w$ neighborhood centered around the target vector $\bar{x}_{st}$; and $F$ is a random number drawn from a Gaussian distribution with mean $\mu$ and variance $\sigma^2$.

### 3.4.2. Crossover

Crossover operation is needed to increase the diversity of the mutated vector. The parameters of the mutated vector $\bar{u}_{st}(G + 1)$ and the target vector $\bar{x}_{st}(G)$ are swapped, depending on the crossover probability (say, $CR$), to yield a trial vector $\bar{u}_{st}(G + 1)$ at generation $(G + 1)$. The trial vector at generation $(G + 1)$ (denoted as $\bar{u}_{st}(G + 1)$) consists of three components $u^n_{st}(G + 1)$, $u^K_{st}(G + 1)$, and $u^C_{st}(G + 1)$; and is obtained by

$$u^n_{st}(G + 1) = \begin{cases} u^n_{st}(G + 1) & \text{if } (R_j \leq CR) \text{or } j = R_3 \\ \bar{x}_{st}(G) & \text{if } (R_j > CR)\text{and } j \neq R_3. \end{cases}$$

Here, $R_j$ (with $j = 1, 2, $ and $3$) is a random number in $[0, 1]$ and $R_3$ is an integer which denotes randomly chosen index in $[1, 3]$. It ensures that the trial vector $\bar{u}_{st}(G + 1)$ will contain at least one parameter from the mutated vector $\bar{u}_{st}(G + 1)$.

### 3.4.3. Selection

The selection operation helps the better fitted vectors to survive for the next generation. If the cost function of the trial vector $\bar{u}_{st}(G + 1)$ is less than that of the target vector $\bar{x}_{st}(G)$, then $\bar{x}_{st}(G + 1)$ is replaced by $\bar{u}_{st}(G + 1)$, otherwise the old value of $\bar{x}_{st}(G)$ is retained.

### 3.4.4. Stopping criterion

As mentioned earlier, the target vectors of the population are evolved using three DE operators (namely, mutation, crossover, and selection) through a number of generations until the termination criterion is satisfied. In the present work, execution of the algorithm is terminated when the class labels of 90% target vectors of the population remain unchanged between the two consecutive generations.

In the proposed DDE, the population corresponds to segmentation of the considered video image frame. The population consists of target vectors which encode red, green, and blue components of each pixel of the segmented image frame. Here, size of the population is equal to the dimension of the segmented image frame. Instead of considering the whole search space for mutation, in the proposed DDE we have considered a sub-set of it. A window is defined at each pixel location and the pixels from its neighbor are used for mutation to obtain the estimated label of corresponding pixel. Since this mechanism uses a small subspace for mutation it helps in saving the large computational time of the DE and also maintain the spatial regularity as mutation is obtained from the neighboring pixels. This is also expected to provide better segmented result.

In the proposed search technique, each population is divided into a number of overlapping $w \times w$ windows. Here, each $w \times w$ window corresponds to a sub-population. Three (mutation, crossover, and selection) operations of conventional DE algorithm are applied within each window centering the target vector, yielding new sub-populations. They are merged to generate an improved population. This is similar to a distributed evolutionary algorithm where the whole population is divided into a number of sub-populations and several sub-populations evolve independently. Afterwards, migration operation is performed to generate an improved population. The proposed algorithm is termed as distributed differential evolution algorithm.

For existing DE algorithms, the population consist of a number of target vectors. Each target vector is considered to be a solution. For DE with neighborhood based mutation algorithms, neighborhood is defined either depending on the index of the target vectors or the Euclidean distances among them. On the other hand, in the proposed DDE, the whole population is considered to be a solution (segmented output of the considered video frame). Each target vector of the population represents one pixel of the segmented frame. Here, neighborhood is defined based on the spatial regularity of the target vector (pixel).

### 3.5. Segmentation of subsequent frames based on change information

It is already mentioned that the spatial segmentation using multi-layer compound MRF model is obtained by estimating the MAP of that model. To segment the initial frame, random initialization was made in [11]. Due to random initialization, segmentation using MRF model becomes computation intensive. Though in case of video sequences, random initialization is highly redundant in the temporal direction as most of the information in the current frame is already present in the previous frame and only some parts of any frame changes over time. Accordingly, it can be inferred that the current segmentation results will basically be the same as those for the previous frame, except for the changed parts. Therefore, based on this assumption, Subudhi et al. [12] proposed a change
information based heuristic initialization for segmenting the subsequent frames. In [12], initialization is done by combining the change information between the current frame and the previously considered frame with the previously segmented one. This mechanism reduces the computational cost for segmenting the subsequent frames. The complete procedure is now described in detail.

Let $x_i$ be the spatial segmentation of the frame $y_i$ at $t$th instant of time. At time $(t + d)$, $y_{t+d}$ represents the input frame and $x_{t+d}$ is its segmented version. Let $x_{(t+d)ij}$ be its initialization and is obtained as follows:

1. Obtain the changes corresponding to the frame $y_{t+d}$ by thresholding the absolute difference of the intensities between $y_{t+d}$ and $y_t$ frames. The original pixel values of $y_{t+d}$ frame corresponding to the changed pixels at $y_{t+d}$ frame is denoted by $y_{(t+d)|y_{t+d} - y_t}$. 
2. The pixels in the previously segmented frame corresponding to the changed pixels at $(t + d)$th frame can be found as

$$x_{(t+d)ij} = x_t - x_{(t+d)ij}. \quad (20)$$

These regions in $x_{ij}$ are replaced by original intensities of $y_{t+d}$ frame for initializing the segmentation of $(t + d)$th frame and can be written as

$$x_{(t+d)ij} = x_t + y_{(t+d)ij}. \quad (21)$$

To obtain $x_{t+d}$ (segmentation of the $y_{t+d}$ frame), DDE algorithm with $x_{(t+d)ij}$ (i.e., change information based) as an initialization is used to segment $y_{t+d}$ frame.

3.6. Temporal segmentation

In case of temporal segmentation, Change Detection Mask (CDM) is obtained by thresholding the absolute differences of the two consecutive spatially segmented image frames. The obtained CDM represents the changed regions and the unchanged regions in the current frame. The changed regions correspond to the moving objects whereas the unchanged regions correspond to stationary objects and background. In order to obtain the locations of the moving objects in the absence of any reference frame, some information from previous frame is used to update the obtained CDM.

This information is represented as:

$$Q = \{q_{ij} | 0 \leq i \leq (M - 1), 0 \leq j \leq (N - 1) \}, \quad (22)$$

where $Q$ is a matrix having the same size of the video frame. Each element $q_{ij}$ of the matrix $Q$ represents the value of the VOP at location $(i,j)$. If $q_{ij} = 1$, then the pixel belongs to the moving object in the previous frame; otherwise to the background or to the still object in the previous frame. Based on this information, the obtained CDM is modified. If the pixel (in CDM) belongs to a moving object in the previous frame and its label is the same as that of the corresponding pixel in the previous frame, then the said pixel belongs to the
moving objects in the current frame. Subsequently, the modified CDM represents the final temporal segmentation.

With the help of temporal segmentation, the given video frames are divided into two regions: changed (moving objects) and unchanged (stationary objects and background). The changed regions and the unchanged regions of the CDM$_t$ at tth instant of time, are represented as CR$_t$ and UR$_t$, respectively. The changed regions in the modified CDM are identified as moving objects in the current frame. As mentioned, the Video Object Planes (VOPs) are extracted by superimposing the intensity/color values of original pixels of the considered current frame $y_t$ in the changed regions CR$_t$ of the modified CDM, thereby detecting the moving objects in the current frame $y_t$.

4. Experimental results and discussion

To test the effectiveness of the proposed algorithm, five reference video sequences (Claire, Grandma, Container, Deadline [42], and Karlsruhe-taxi [43]) and one real life video sequence (Rahul) are considered. These are well known benchmark video sequences. Claire, Grandma, Deadline, and Rahul video sequences have only one moving object, whereas Container and Karlsruhe-taxi video sequences have multiple moving objects.

Since the main objective of the present work is to handle the challenges for detection of moving objects from video sequences where an object in the scene moves very slow, or very fast, or stops, and moves further with better accuracy (less effect of silhouettes).

Claire is a news reader video sequence. The movement of object Claire in the scene is unpredictable. The movement of Claire is different from one frame to another. Since, the scene has undergone illumination variation, it is very difficult to identify the detailed information of collar, tie, face, and hair regions. In Grandma video sequence, the object Grandma moves for some times, then temporarily stops her movement for some time, and again resumes. This is a poorly illuminated sequence. Different regions in the face and hair of the Grandma are very similar with that of the background wall. The face region has undergone illumination variation. In Container video sequence, there are three moving objects with different speeds. The boundaries of the objects in the scene are blurred. So it is very difficult to identify the exact boundaries of the objects in the scene. Similarly, in Deadline sequence, a complex background exists, where hair and dress of the object are similar to some part of the background. The movement of the object in Deadline sequence is not constant throughout the sequence. In Karlsruhe-taxi sequence, there are three objects having different sizes and speeds. In its background, many similar kinds of confusing static objects are present. Rahul video sequence is captured by a low resolution camera. Here, the movement of Rahul’s head is very fast than the other parts of the body; and face and hand are having similar color with that of the background.

![Fig. 4. Results on 12th and 87th frames of Grandma video sequence: (a) Original frames. Spatial segmentations obtained using (b) MRF with DGA scheme [10,11], (c) modified MRF with DGA scheme, (d) MRF with Graph Cut scheme [7,22], (e) MRF with SA and ICM scheme [12], (f) the proposed method. Temporal segmentations obtained using (g) MRF with DGA scheme [10,11], (h) modified MRF with DGA scheme, (i) MRF with Graph Cut scheme [7,22], (j) MRF with SA and ICM scheme [12], (k) the proposed method. VOPs obtained using (l) MRF with DGA scheme [10,11], (m) modified MRF with DGA scheme, (n) MRF with Graph Cut scheme [7,22], (o) MRF with SA and ICM scheme [12], (p) the proposed method.](image-url)
In the present article, all the considered video sequences are captured by a fixed camera. Figs. 3(a)–8(a) show a few considered original frames of the above mentioned video sequences. In our experiment, the first video sequence considered is \textit{Claire} which has a single moving object. Here, the 3rd frame is an initial frame and all other frames are the subsequent ones. Initial frame of the video sequence \textit{Grandma} is the 12th one whereas the remaining frames are the subsequent frames. In \textit{Container} video sequence, multiple objects such as ship, trawler, and flying flag are moving. The 4th frame of this sequence is considered as initial frame and all other frames are subsequent ones. In case of \textit{Deadline} video sequence, we considered 3rd frame as initial and the remaining frames as subsequent frames. \textit{Rahul}, the real life video sequence is considered to assess the robustness of the proposed approach. Here, we considered the 11th frame as the initial one and all other frames as the subsequent ones. In this work, we have done experiment on all frames of the considered video sequences. But due to the page constraint, we have reported only few sample results for visual illustration.

As mentioned in Section 3.4, for mutation of each target vector, three parameter vectors may be randomly selected either (i) from the whole search space in case of conventional DE or (ii) from a $w \times w$ neighborhood of the corresponding target vector in case of the proposed DDE. It is observed through our experiments that the latter method provides better spatial segmentation results with respect to average segmentation error with less computational time. Here, we have reported results corresponding to the latter method. For illustration, results of six video sequences \textit{Claire}, \textit{Grandma}, \textit{Container}, \textit{Deadline}, \textit{Karlsruhe-taxi}, and \textit{Rahul} are reported in this article. Figs. 3–8 represent the results for the \textit{Claire}, \textit{Grandma}, \textit{Container}, \textit{Deadline}, \textit{Karlsruhe-taxi}, and \textit{Rahul} video sequences, respectively.

Values of the parameters for MRF with DGA scheme [10,11], modified MRF with DGA scheme, MRF with Graph Cut scheme [7,22], MRF with SA and ICM scheme [12], and the proposed method are put in Tables 1–5, correspondingly.

In the present work, ground truths of the considered video sequences are generated manually. Figs. 3(b)–8(b), 3(c)–8(c), 3(d)–8(d) and 3(e)–8(e), respectively show the spatial segmentations of the considered frames of \textit{Claire}, \textit{Grandma}, \textit{Container}, \textit{Deadline}, \textit{Karlsruhe-taxi}, and \textit{Rahul} video sequences using MRF with DGA scheme [10,11], modified MRF with DGA scheme, MRF with Graph Cut scheme [7,22], and MRF with SA and ICM scheme [12]. Similarly, the spatial segmentations of the considered frames for different video sequences using the proposed method are displayed in Figs. 3(f)–8(f). The corresponding temporal segmentation results of the considered frames for the same video sequences using these methods are shown respectively in Figs. 3(g)–8(g), 3(h)–8(h), 3(i)–8(i), 3(j)–8(j), 3(k)–8(k), 3(l)–8(l), 3(m)–8(m), 3(n)–8(n), 3(o)–8(o) and 3(p)–8(p), correspondingly show the extracted VOPs of the considered frames.
using these methods. Fig. 9 displays the average segmentation errors produced by different methods: MRF with DGA scheme [10,11], modified MRF with DGA scheme, MRF with Graph Cut scheme [7,22], MRF with SA and ICM scheme [12], and the proposed method. From this figure, it is observed that the proposed method provides better segmentation for all the considered frames (with less average segmentation error) than all other methods.

The first video sequence considered for the experiment is Claire. Fig. 3(a) shows few original image frames of this sequence. Fig. 3(b) shows the spatial segmentation results obtained by MRF with DGA scheme [10,11]. Here, the frames of this sequence are modeled with MRF and corresponding MAP is estimated using DGA algorithm. For subsequent frame segmentation, chromosomes having more energy are allowed to evolve through DGA. The parameters used

Table 1
<table>
<thead>
<tr>
<th>Video</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>Crossover probability (CR)</th>
<th>Mutation probability (MU)</th>
<th>Window size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claire</td>
<td>0.05</td>
<td>0.02</td>
<td>0.05</td>
<td>0.005</td>
<td>5 × 5</td>
<td></td>
</tr>
<tr>
<td>Grandma</td>
<td>0.03</td>
<td>0.01</td>
<td>0.07</td>
<td>0.001</td>
<td>5 × 5</td>
<td></td>
</tr>
<tr>
<td>Container</td>
<td>0.08</td>
<td>0.04</td>
<td>0.04</td>
<td>0.005</td>
<td>5 × 5</td>
<td></td>
</tr>
<tr>
<td>Deadline</td>
<td>0.05</td>
<td>0.009</td>
<td>0.03</td>
<td>0.002</td>
<td>5 × 5</td>
<td></td>
</tr>
<tr>
<td>Karlsruhe-taxi</td>
<td>0.009</td>
<td>0.008</td>
<td>0.02</td>
<td>0.001</td>
<td>5 × 5</td>
<td></td>
</tr>
<tr>
<td>Rahul</td>
<td>0.05</td>
<td>0.004</td>
<td>0.01</td>
<td>0.005</td>
<td>5 × 5</td>
<td></td>
</tr>
</tbody>
</table>

Parameters chosen for different video sequences of MRF with DGA scheme.

Table 2
<table>
<thead>
<tr>
<th>Video</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>Crossover probability (CR)</th>
<th>Mutation probability (MU)</th>
<th>window size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claire</td>
<td>0.05</td>
<td>0.02</td>
<td>0.008</td>
<td>0.05</td>
<td>0.005</td>
<td>5 × 5</td>
</tr>
<tr>
<td>Grandma</td>
<td>0.03</td>
<td>0.01</td>
<td>0.001</td>
<td>0.07</td>
<td>0.001</td>
<td>5 × 5</td>
</tr>
<tr>
<td>Container</td>
<td>0.08</td>
<td>0.04</td>
<td>0.02</td>
<td>0.04</td>
<td>0.005</td>
<td>5 × 5</td>
</tr>
<tr>
<td>Deadline</td>
<td>0.05</td>
<td>0.009</td>
<td>0.007</td>
<td>0.03</td>
<td>0.005</td>
<td>5 × 5</td>
</tr>
<tr>
<td>Karlsruhe-taxi</td>
<td>0.009</td>
<td>0.008</td>
<td>0.005</td>
<td>0.02</td>
<td>0.001</td>
<td>5 × 5</td>
</tr>
<tr>
<td>Rahul</td>
<td>0.05</td>
<td>0.004</td>
<td>0.003</td>
<td>0.01</td>
<td>0.005</td>
<td>5 × 5</td>
</tr>
</tbody>
</table>

Parameters chosen for different video sequences of modified MRF with DGA scheme.
by MRF with DGA scheme [10,11] for Claire sequence are shown in Table 1. The results obtained by modified MRF with DGA scheme are shown in Fig. 3(c). Here, multi-layer compound MRF is used to model the frames and MAP is estimated using DGA algorithm. The considered parameters for this scheme are reported in Table 2. In MRF with Graph Cut scheme [7,22], the considered frames are modeled with MRF and corresponding MAP is estimated using Graph Cut scheme. Parameters of this approach are provided in Table 3. Fig. 3(d) shows the spatial segmentation of this scheme. In MRF with SA and ICM scheme [12], frames are segmented by modeling it with multi-layer compound MRF model followed by the hybrid algorithm for MAP estimation. Figure 3(e) shows spatial segmentation results. The MRF parameters chosen for this sequence are displayed in Table 4. For the proposed method, frames of this sequence are modeled with the multi-layer compound MRF model.

Segmentation of the initial frame is obtained by estimating MAP of the MRF model using the proposed DDE algorithm. Change information based subsequent frame segmentation scheme is used to segment the subsequent frames of this sequence. Fig. 3(f) shows spatial segmentation results of the considered Claire frames. The parameters used by the proposed scheme are reported in Table 5. Temporal segmentation is obtained using original frame difference

Fig. 7. Results on 4th and 6th frames of Karlsruhe-taxi video sequence: (a) Original frames. Spatial segmentations obtained using (b) MRF with DGA scheme [10,11], (c) modified MRF with DGA scheme, (d) MRF with Graph Cut scheme [7,22], (e) MRF with SA and ICM scheme [12], (f) the proposed method. Temporal segmentations obtained using (g) MRF with DGA scheme [10,11], (h) modified MRF with DGA scheme, (i) MRF with Graph Cut scheme [7,22], (j) MRF with SA and ICM scheme [12], (k) the proposed method. VOPs obtained using (l) MRF with DGA scheme [10,11], (m) modified MRF with DGA scheme, (n) MRF with Graph Cut scheme [7,22], (o) MRF with SA and ICM scheme [12], (p) the proposed method.

Table 3
Parameters chosen for different video sequences of MRF with Graph Cut scheme [7,22].

<table>
<thead>
<tr>
<th>Video</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \gamma )</th>
<th>( k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claire</td>
<td>0.09</td>
<td>0.5</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>Grandma</td>
<td>0.05</td>
<td>0.3</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>Container</td>
<td>0.08</td>
<td>0.5</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>Deadline</td>
<td>0.06</td>
<td>0.4</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>Karlsruhe-taxi</td>
<td>0.03</td>
<td>0.4</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>Rahul</td>
<td>0.07</td>
<td>0.5</td>
<td>550</td>
<td></td>
</tr>
</tbody>
</table>

Table 4
Parameters chosen for different video sequences of MRF with SA and ICM scheme [12].

<table>
<thead>
<tr>
<th>Video</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \gamma )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claire</td>
<td>0.009</td>
<td>0.008</td>
<td>0.007</td>
<td>1.0</td>
</tr>
<tr>
<td>Grandma</td>
<td>0.05</td>
<td>0.009</td>
<td>0.007</td>
<td>5.19</td>
</tr>
<tr>
<td>Container</td>
<td>0.01</td>
<td>0.009</td>
<td>0.001</td>
<td>2.44</td>
</tr>
<tr>
<td>Deadline</td>
<td>0.009</td>
<td>0.006</td>
<td>0.003</td>
<td>1.3</td>
</tr>
<tr>
<td>Karlsruhe-taxi</td>
<td>0.01</td>
<td>0.008</td>
<td>0.007</td>
<td>2.44</td>
</tr>
<tr>
<td>Rahul</td>
<td>0.009</td>
<td>0.005</td>
<td>0.001</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Table 5
Parameters chosen for different video sequences of the proposed method.

<table>
<thead>
<tr>
<th>Video</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \gamma )</th>
<th>( \sigma )</th>
<th>Window size</th>
<th>Crossover probability (CR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claire</td>
<td>0.009</td>
<td>0.008</td>
<td>0.003</td>
<td>1.0</td>
<td>5 \times 5</td>
<td>0.9</td>
</tr>
<tr>
<td>Grandma</td>
<td>0.03</td>
<td>0.01</td>
<td>0.001</td>
<td>0.01</td>
<td>5 \times 5</td>
<td>0.8</td>
</tr>
<tr>
<td>Container</td>
<td>0.08</td>
<td>0.01</td>
<td>0.001</td>
<td>0.05</td>
<td>5 \times 5</td>
<td>0.9</td>
</tr>
<tr>
<td>Deadline</td>
<td>0.009</td>
<td>0.008</td>
<td>0.003</td>
<td>0.1</td>
<td>5 \times 5</td>
<td>0.7</td>
</tr>
<tr>
<td>Karlsruhe-taxi</td>
<td>0.01</td>
<td>0.007</td>
<td>0.005</td>
<td>0.15</td>
<td>5 \times 5</td>
<td>0.5</td>
</tr>
<tr>
<td>Rahul</td>
<td>0.005</td>
<td>0.004</td>
<td>0.01</td>
<td>0.9</td>
<td>5 \times 5</td>
<td>0.7</td>
</tr>
</tbody>
</table>
followed by thresholding technique for both MRF with DGA scheme [10,11] and its modified version. Temporal segmentations are shown in Fig. 3(g) and (h). Label frame difference followed by thresholding scheme is used for temporal segmentation for MRF with Graph Cut scheme [7,22], MRF with SA and ICM scheme [12], and the proposed method. Fig. 3(i)–(k) shows the temporal segmentation of MRF with Graph Cut scheme [7,22], MRF with SA and ICM scheme [12], and the proposed method, respectively. Fig. 3(l)–(p) correspondingly show the extracted VOPs of the considered frames using these methods.

To demonstrate the effectiveness of the proposed approach over other methods, for typical illustration, let us consider the output of the Claire video sequence. From Fig. 3(b)–(f), it is noticed that the nose, eye, some parts of face, and few parts of shirt are not properly segmented using MRF with DGA scheme [10,11] and its modified version. It produces more small patches in the shirt region. MRF with Graph Cut scheme [7,22] merges eyes, nose, and lip regions to the face region of Claire. It is also seen from the figures that some part of the head, face and the body of the Claire are not properly segmented using MRF with SA and ICM scheme [12]. This may be due

![Fig. 9. Average Segmentation Error](image-url)
to the premature convergence of ICM used in [12]. The proposed strategy based on DDE, on the other hand, can segment these parts properly. Fig. 9 displays the average segmentation error of different considered video sequences. From this figure, it is observed that the proposed method produces better results than those obtained by all other methods.

From the segmentation results of Grandma video sequence reported in Fig. 4(b)–(f), it is seen that few parts of the face region are segmented as background using both MRF with DGA scheme [10, 11] and its modified version. It is also noticed that few parts of noise and eye are merged to face region and more patches present in the body part. MRF with Graph Cut scheme [7, 22] merges the face, hair, and neck regions into a single region. MRF with SA and ICM scheme [12] segments these parts into small patches. The proposed technique, on the other hand, properly segments these parts with less amount of average segmentation error.

Incase of Container video sequence, MRF with DGA scheme [10, 11], its modified version, MRF with Graph Cut scheme [7, 22] and MRF with SA and ICM scheme [12] segment more parts of the container as water region. On the other hand the proposed method produces better segments.

From Fig. 6(b)–(f), it is also visually seen that the proposed method provides better segmentation results of Deadline video sequence than all other schemes. From the segmentation results of Karlsruhe-taxi video sequence displayed in Fig. 7(b)–(f), it is found that few small cars and tree are merged to road for both MRF with DGA scheme [10, 11] and its modified version. MRF with Graph Cut scheme [7, 22] merges all static small cars into a single region and unable to segment the tree. Whereas, MRF with SA and ICM scheme [12] properly segment static small cars but merges the tree with the road. On the other hand, the proposed method properly segments all static small cars as well as the tree. If we visually analyze the segmentation results of Rahul video sequence as depicted in Fig. 8(b)–(f), it is noticed that the face and the neck region of Rahul are not properly segmented by MRF with DGA scheme [10, 11] and its modified version, whereas face, neck, and hair regions are merged into a single region by MRF with Graph Cut scheme [7, 22]. On the other hand MRF with SA and ICM scheme [12] and the proposed method provide similar results except a few parts of face and neck regions.

All simulations have been done using visual C++ on a machine with Pentium(R)4 CPU 2.40 GHz and 512 MB RAM. Table 6 shows the average computational time (±standard deviation) required for segmenting the considered video frames using different methods. From this table, it is seen that MRF with SA and ICM scheme [12] requires maximum average time for segmenting frames of the considered video sequences whereas MRF with DGA scheme [10, 11] takes the least amount of average execution time. In MRF with SA and ICM scheme [12], SA was used initially to find out the suboptimal solutions and then ICM was executed to find out the near optimal solutions, thereby making it computation intensive. MRF with Graph Cut scheme [7, 22] takes less average time than MRF with SA and ICM scheme [12] and more average time than other methods for segmenting frames. On the other hand, MRF with DGA scheme [10, 11] takes the least amount of average time for segmenting the image frames as elitism is used and mutation may not occur always. The proposed method explores the search space quickly using DDE; thereby produces the output with moderate average execution time. The minimum and maximum average execution times required for segmenting all considered frames among different methods are highlighted in Table 6.

The bold value indicates that the method takes maximum execution time among all the techniques for segmenting frames of the considered video. Whereas the italic value indicates that the method takes minimum execution time among all the techniques for segmenting frames of the considered video.

The results reported in this article were obtained with the optimal values of the MRF model bonding parameters α, β, and γ. The value of α affects the number of regions in segmentation. Large value of α merges contextually less variate regions. On the other hand, small value of α produces smaller regions which are contextually less variate. β encourages pixels to have the same label in consecutive image frames. Fragmentation can occur due to high value of β. Similarly, small value of γ produces thicker boundary and high value of γ will result in thinner boundary. Fig. 10(a)–(c) displays the spatial segmentations of Walk-Run [44] (it is a high resolution video of size 640 × 480) video sequence obtained by the proposed method with large values of α, β, and γ, respectively. Fig. 10(d)–(f) shows the spatial segmentations obtained by the proposed method with small values of α, β, and γ, respectively.

Selection of optimal values of α, β, and γ is critical. To better assess the effects of α, β, and γ on spatial segmentation, we obtained different results by considering different combination of α, β, and γ values in the ranges [0, 2.5], [0, 1.0], and [0, 0.7], respectively. The spatial segmentation, temporal segmentation, and VOP of Walk-Run video sequence obtained by the proposed method with optimal values of parameters α, β, and γ are shown in Fig. 11(b)–(d), respectively.

The crossover probability (CR) and scaling factor (F) are the control parameters of the proposed DDE. The CR controls how many parameters of the target vector and the mutated vector are exchanged. A large value of CR speeds up the convergence of the DDE and solution may stuck to a local optimal; whereas a small value of CR degrades the convergence speed but may obtain near global optimum solution. So good choice of CR is a critical issue for balancing both convergence speed and solution quality. In this regard, we empirically found that CR ∈ [0.5, 0.9] is a good choice for video frame segmentation. In the proposed DDE, the scaling factor F is randomly drawn from a Gaussian distribution with mean 0 and variance 1. Segmented output obtained using the proposed DDE also depends on the size of the neighborhood. If we increase the neighborhood size, the convergence speed of the DDE may decrease and may merge more small regions to a single region. On the other hand, if we decrease the neighborhood size, the convergence speed may increase and may produce more small regions which are not distinct with respect to their spatial continuity. Hence a proper choice of neighborhood size is an important factor. From the experiment it is found that the neighborhood size between 5 × 5 to 7 × 7 produces better results.

Parameters of both MRF with DGA scheme [10, 11] and its modified schemes are determined on trial and error basis. It is found

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Claire</td>
<td>11.00 ± 2.18</td>
<td>17.50 ± 4.33</td>
<td>28.50 ± 13.43</td>
<td>33.50 ± 37.75</td>
<td>24.50 ± 12.99</td>
</tr>
<tr>
<td>Grandma</td>
<td>15.50 ± 2.60</td>
<td>25.25 ± 5.63</td>
<td>37.50 ± 18.19</td>
<td>35.75 ± 85.30</td>
<td>26.50 ± 16.45</td>
</tr>
<tr>
<td>Container</td>
<td>10.50 ± 2.41</td>
<td>17.25 ± 3.90</td>
<td>29.50 ± 17.92</td>
<td>41.25 ± 45.47</td>
<td>23.50 ± 14.72</td>
</tr>
<tr>
<td>Deadline</td>
<td>38.25 ± 21.22</td>
<td>51.25 ± 28.58</td>
<td>64.50 ± 37.24</td>
<td>75.00 ± 77.94</td>
<td>60.25 ± 35.18</td>
</tr>
<tr>
<td>Karlsruhe-taxi</td>
<td>22.50 ± 7.40</td>
<td>32.50 ± 16.54</td>
<td>55.50 ± 28.57</td>
<td>55.25 ± 42.80</td>
<td>49.50 ± 25.97</td>
</tr>
<tr>
<td>Rahul</td>
<td>7.25 ± 0.87</td>
<td>12.25 ± 1.95</td>
<td>25.00 ± 15.39</td>
<td>25.25 ± 22.22</td>
<td>22.25 ± 15.81</td>
</tr>
</tbody>
</table>

The bold value indicates that the method takes maximum execution time among all the techniques for segmenting frames of the considered video. Whereas the italic value indicates that the method takes minimum execution time among all the techniques for segmenting frames of the considered video.
from the experiments that the parameters set to similar values as reported in [45] produce better results. For MRF with Graph Cut scheme [7,22], parameters are also empirically determined. From the experiment it is found that it produces better segmented results with these set of parameter values. Parameters of MRF with SA and ICM schemes [12] are set according to those values reported in [46]. Tables 1–5 display the parameter values for different considered video sequences by various methods.

5. Conclusions and future works

In this article, moving objects from a given video sequence are detected using a multi-layer compound MRF model with DDE as optimization technique. In the proposed method, considered MRF model preserves the boundary information of the segmented regions. The proposed DDE algorithm explores the search space quickly for finding out the optimum segmentation. From the results it is found that the proposed DDE based algorithm provides better segmented results (with respect to average segmentation error) than those obtained using MRF with DGA scheme [10,11], its modified version, MRF with Graph Cut scheme [7,22], and MRF with SA and ICM scheme [12]. The present approach could also identify the locations of the moving objects well. The proposed method is less computation intensive than MRF with Graph Cut scheme [7,22] and MRF with SA and ICM scheme [12]. Here, the parameters of the considered MRF model are chosen on a trial and error basis. To make the system more robust and automatic, these parameters could be determined using some parameter estimation algorithms. Mallipeddi et al. discussed some heuristic techniques to determine the control parameters of DE algorithm in [47]. These mechanisms may also be incorporated to determine the control parameters of the proposed DDE. The performance of DE algorithm also depends on selection of vectors for mutation operation. Recent literature [48–52] briefly discuss about the selection procedure of vectors for mutation operation. In future, instead of random selection we will incorporate such kind of mechanism for selecting vectors to improve the performance of the DDE.

Acknowledgements

The authors like to thank the reviewers for their thorough and constructive comments, which helped a lot to enhance the quality of the manuscript. Funding by U.S. Army through the project “Processing and Analysis of Aircraft Images with Machine Learning Techniques for Locating Objects of Interest” (Contract No. FA5209-08-P-0241) is also gratefully acknowledged.

References


