An Evolutionary Based Slow and Fast Moving Video Object Detection Scheme Using Compound Markov Random Field Model

Badri Narayan Subudhi
Department of Electrical Engineering
National Institute of Technology
Rourkela, India
subudhi.badri@gmail.com

Pradipta Kumar Nanda
Department of Electonics and Telecommunication
C. V. Raman College of Engineering
Bhubaneswar, India
pknanda_d13@yahoo.co.in

Abstract

We propose an evolving scheme to detect slow as well as fast moving objects in a video sequence. The proposed scheme employ both spatio-temporal and temporal segmentation to obtain the Video Object plane and hence detection. We propose a Compound Markov Random Field Model as the a priori image model that takes into account the spatial distribution of the current frame, temporal frames and the edge maps of the temporal frames. The spatio-temporal segmentation is cast as a pixel labeling problem and the labels are the MAP estimates. The MAP estimates of a frame are obtained by a hybrid algorithm. The spatial segmentation of a given frame evolves to generate the spatial segmentation of the subsequent frames. The evolved spatial segmentation together with the temporal segmentation produces the Video Object Plane (VOP) and hence detection. Our scheme does require the computation of spatio-temporal segmentation of the initial frame thus speeding up the whole process. The results of the proposed scheme are compared with JSEG method are found to be better in terms of the misclassification error.

1. Introduction

There has been a growing research interest in video image segmentation over the past decade and towards this end, a wide variety of methodologies have been developed [18],[16],[13],[6]. The video segmentation methodologies have extensively used stochastic image models, particularly Markov Random Field (MRF) model, as the model for video sequences [7],[11],[12]. MRF model has proved to be an effective stochastic model for image segmentation [15],[3],[4] because of its attribute to model context dependent entities such as image pixels and correlated features. In Video segmentation, besides spatial modeling and constraints, temporal constraints are also added to devise spatio-temporal image segmentation schemes. An adaptive clustering algorithm has been reported [7] where temporal constraints and temporal local density have been adopted for smooth transition of segmentation from frame to frame. Spatio-temporal segmentation has also been applied to image sequences [19] with different filtering techniques. Extraction of moving object and tracking of the same has been achieved in spatio-temporal framework [9] with Genetic algorithm serving as the optimization tool for image segmentation. Recently, MRF model has been used to model spatial entities in each frame [9] and Distributed Genetic algorithm (DGA) has been used to obtain segmentation. Modified version of DGA has been proposed [11] to obtain segmentation of video sequences in spatio-temporal framework. Besides, video segmentation and foreground subtraction has been achieved using the spatio-temporal notion [1],[2] where the spatial model is the Gibbs Markov Random Field and the temporal changes are modeled by mixture of Gaussian distributions. Very recently, automatic segmentation algorithm of foreground objects in video sequence segmentation has been proposed [8]. In this approach, first region based motion segmentation algorithm is proposed and thereafter the labels of the pixels are estimated. A compound MRF model based segmentation scheme has been proposed in spatio-temporal framework [17]. The problem of extraction of moving target from the background has been investigated [10] where adaptive thresholding based scheme has been employed to segment the images.

In this piece of work, we propose a scheme that detects slow as well as fast moving objects. The proposed scheme is a combination of spatio-temporal segmentation and temporal segmentation. In this approach, we obtain spatio-temporal segmentation once for a given frame and thereafter, for subsequent frames, the segmentation is obtained

978-0-7695-3476-3/08 $25.00 © 2008 IEEE
DOI 10.1109/ICVGIP.2008.38

by the evolution of the initial spatio-temporal segmentation. We have proposed a Compound MRF model that takes care of the spatial distribution of the current frame,temporal frames, edge maps in the temporal direction. The MRF model parameters are selected on a trial and error basis. This problem is formulated using MAP estimation principles. The pixel labels are obtained using the proposed hybrid algorithm.For the subsequent frames the initial segmentation evolves to obtain the spatial segmentation. This spatio-temporal segmentation combined with temporal segmentation yields the VOP and hence Video Object detection. In our scheme for temporal segmentation, we use the segmented frames as opposed to the original frames. The results obtained by proposed methods are compared with that of the JSEG method and it is observed that the proposed method is found to be better than former one in the context of misclassification error.

2. Spatio temporal image modeling

Let the observed video sequences y be considered to be 3-D volume consisting of spatio-temporal image frames. For video, at a given time t, y_t represents the image at time t and hence is a spatial entity. Each pixel in y_t is a site s denoted by y_{st} and hence, y_{st} refers to a spatio-temporal representation of the 3-D volume video sequences. Let x denote the segmented video sequences and x_t denote the segmentation of each video frame y_t. Instead of modeling the video as a 3-D model we adhere to a spatio-temporal modeling. We model X_t as a Markov random Field Model and the temporal pixels are also modeled as MRF. We model X_t as Markov Random Field model and the temporal pixels are also modeled as MRF. In particular for second order modeling in the temporal directions, we take X_t, X_{t-1} and X_{t-2}. In order to preserve the edge features, another MRF model is considered for the pixel of the current frame x_{st} and the line fields of X_{t-1} and X_{t-2}. Thus, three MRF models are used as the spatio-temporal image model. The MRF model taking care of edge features, in other words the line fields of frame x_{t-1} and x_{t-2} together with x_t are modeled as MRF. It is known that if X_t is MRF then, it satisfies the markovianity property in spatial direction.

$$\begin{aligned}&P(X_{st} = x_{st} \mid X_{qt} = x_{qt}, \forall q \epsilon S, s \neq q)\\ &= P(X_{st} = x_{st} \mid X_{qt} = x_{qt}, (q,t) \epsilon \eta_{s,t})\end{aligned}$$

where $\eta_{s,t}$ is denotes the neighborhood of (s,t) and S denotes spatial Lattice of the frame X_t. For temporal MRF, the following markovianity is satisfied.

$$\begin{aligned}P(X_{st} = x_{st} \mid X_{pq} = x_{pq}, q \neq t, p \neq s, \forall (s,t) \epsilon V)\\ = P(X_{st} = x_{st} \mid X_{pq} = x_{pq}, (p,q) \epsilon \eta_{s,t})\end{aligned}$$

where V denotes the 3-D volume of the video sequence. In spatial domain X_t is modeled as MRF and hence the prior probability can be expressed as Gibb's distributed which can be expressed as $P(X_t) = \frac{1}{z} e^{\frac{-U(X_t)}{T}}$ where z is the partition function which is expressed as $z = \sum_x e^{\frac{-U(x_t)}{T}}$, $U(X_t)$ is the energy function and expressed as $U(X_t) = \sum_{c \in C} V_c(x_t)$ and $V_c(x_t)$ denotes the clique potential function,T denotes the temperature and is considered to be unity. We have considered the following clique potential function.

$$V_c(x) = \begin{cases} +\alpha : if x_{st} \neq x_{pt} and (s,t),(p,t) \epsilon S \\ -\alpha : if x_{st} = x_{pt} and (s,t),(p,t) \epsilon S \end{cases}$$

$$V_{tec}(x) = \begin{cases} +\beta : if x_{st} \neq x_{qt} and (s,t),(q,t) \epsilon S \\ -\beta : if x_{st} = x_{qt} and (s,t),(q,t) \epsilon S \end{cases}$$

Analogously in the temporal direction

$$V_{teec}(x) = \begin{cases} +\gamma : if x_{st} \neq x_{et} and (s,t),(e,t) \epsilon S \\ -\gamma : if x_{st} = x_{et} and (s,t),(e,t) \epsilon S \end{cases}$$

2.1. MRF-MAP Based Framework

The Segmentation problem is cast as a pixel labeling problem. Let y be the observed video sequence and be an image frame at time t and s denote the site of the image y_t. Correspondingly Y_t is modeled as a random field and y_t is a realization frame at time t. Thus, y_{st} denotes as a spatio-temporal co-ordinate of the grid (s,t). Let x denotes the segmentation of the video sequence and let x_t denotes the segmentation of an image at time t. Let X_t denote the random field in the spatial domain at time t. The observed image sequences Y are assumed to be the degraded version of the segmented image sequences X. For example at a given time t, the observed frame Y_t is considered as the degraded version of the original label field X_t . This degradation process is assumed to be Gaussian Process. Thus, the label field can be estimated from the observed random field Y_t . The label field is estimated by maximizing the following posterior probability distributions.

$$\hat{x} = \arg \max_x P(X = x|Y = y) \qquad (1)$$

Where $\hat{x}$ denotes the estimated labels. Since, x is unknown it is very difficult to evaluate (1), hence, using Baye's theorem (1) can be written as

$$\hat{x} = \arg \max_x \frac{P(Y = y|X = x)P(X = x)}{P(Y = y)} \qquad (2)$$

Since y is known, the prior probability $P(Y = y)$ is constant. hence (2) reduces to

$$\hat{x} = \arg \max_x P(Y = y|X = x, \theta)P(X = x, \theta) \qquad (3)$$

Where θ is the parameter vector associated with x. According to Hammerseley Clifford theorem, the prior probability $P(X = x, \theta)$ is Gibb's distributed and is of the following form

$$P(X = x) = e^{-U(x,\theta)} \tag{4}$$
$$= e^{\left[-\sum_{c\epsilon C}[V_{sc}(x)+V_{tec}(x)+V_{teec}(x)]\right]}$$

In (4) $V_{sc}(x)$ is the clique potential function in the spatial domain at time t, $V_{tec}(x)$ denotes the clique potential in the temporal domain and $V_{teec}(x)$ denotes the clique potential in the temporal domain incorporating edge feature. We have proposed this additional feature in the temporal direction. (4) is called the edgebased model. In the absence of the edge feature in temporal direction it has been observed that many classes are merged and the accuracy of the segmentation degrades. This could be attribued to the missing edge feature in the temporal directions. The corresponding edgeless model is

$$P(X = x) = e^{-U(x,\theta)} = e^{\left[-\sum_{c\epsilon C}[V_{sc}(x)+V_{tec}(x)]\right]}$$

The likelihood function $P(Y = y|X = x)$ can be expressed as

$$P(Y = y|X = x) = P(y = x + n|X = x + \theta)$$
$$= P(N = y - x|X = x + \theta)$$

Since n is assumed to be Gaussian and there are three components present in color, $P(Y = y|X = x)$ Can be expressed as

$$P(N = y - x|X, \theta) \tag{5}$$
$$= \frac{1}{\sqrt{(2\pi)^n det\,[k]}} e^{-\frac{1}{2}(y-x)^T K^{-1}(y-x)}$$

Where k is the covariance matrix. Assuming decorrelation of the three RGB planes and the variance to be same among each plane, (5) can be expressed as

$$P(N = y - x|X, \theta) = \frac{1}{\sqrt{(2\pi)^3\sigma^3}} e^{-\frac{1}{2\sigma^2}(y-x)^2} \tag{6}$$

In (6) Variance σ^2 corresponds to the Gaussian degradation. Hence (3) can be expressed as

$$\hat{x} = \arg\,\max_x \frac{1}{(2\pi)^3\sigma^3} \times \tag{7}$$
$$e^{\frac{\left[-\|y-x\|^2\right]}{2\sigma^2}\left[-\left[\sum_{c\epsilon C}[V_{sc}(x)+V_{tec}(x)+V_{teec}(x)]\right]\right]}$$

The apriori model having the three components is attributed as the edgebased model. Maximizing (7) is tantamount to minimizing the

$$\hat{x} = \arg\,\min_x \left[\frac{\| y - x \|^2}{2\sigma^2}\right] + \tag{8}$$
$$\left[\sum_{c\epsilon C} V_{sc}(x) + V_{tec}(x) + V_{teec}(x)\right]$$

$\hat{x}$ in (8) is the MAP estimate and the MAP estimate is obtained by the proposed hybrid algorithm. The associated clique potential parameters and the noise standard deviation σ are selected on trial and error basis

2.2. Evolutionary Approach Based Segmentation Scheme

In order to detect fast moving objects, temporal segmentation usually used and for slow moving objects spatio-temporal segmentation has to be coupled with temporal segmentation. Spatio-temporal segmentation in MRF-MAP frame work is computational intensive and hence computing spatial segmentation of each frame would incur high computational burden. Hence, we suggest the following evolutionary approach to obtain spatial segmentation. In this scheme, the temporal changes in the spatio-temporal segmentation is replaced with the changes occuring in the respective original frames. In other words we obtain the spatio-temporal segmentation, and subtract the temporal changes. Thereafter, we we add the respective changes of the original frame. This improves the accuracy of segmentation while taking care of the moving objects. Let y_t denotes the current frame and x_t denotes the corresponding spatial segmentation. The next frame is denoted by y_{t+d} and $x_{(t+d)i}$ denotes the initial spatial segmentation for the y_{t+d}th frame. $x_{(t+d)i}$ is obtained as follows,

$$x_{(t+d)i} = x_t - \mid y_{t+d} - y_t \mid + y_{t+d(y_{t+d}-y_t)} \tag{9}$$

Where $y_{t+d(y_{t+d}-y_t)}$ denotes the change portion of the tth frame to be replaced in the tth segmented frame x_t. $x_{(t+d)i}$ serves as the initial spatial segmentation for $(t+d)$th frame. Iterated Conditional Mode (ICM) is run on the $(t + d)$th frame starting from $x_{(t+d)i}$ to obtain the $x_{(t+d)}$. This process repeated to obtain spatio-temporal segmentation of any other frame.

2.3. Hybrid Algorithm

It is observed that SA algorithm takes substantial amount of time to converge to the global optimum solution. SA algorithm has the attribute of coming out of the local minima and converging to the global optimal solution. This feature could be attributed to the acceptance criterion(acceptance with a probability). We have exploited this feature, that is the proposed hybrid algorithm uses the notion of acceptance criterion to come out of the local minima and to be near the global optimal solution. Thus, in the hybrid algorithm, SA algorithm produces an intermediate solution that can be local to the optimal solution. In order to obtain the optimal solution, a local convergence based strategy is adopted for quick convergence. Towards this end, we have used Iterated Conditional Mode (ICM) [3] algorithm as the locally convergent algorithm. Thus, the proposed algorithm is a hybrid of both SA algorithm and ICM algorithm. The hybrid algorithm's working principle is as follows. Initially, a specific number of time steps of SA algorithm, fixed by trial and error, are executed to achieve the near optimal solution. Thereafter, ICM is run to converge to the desired optimal solution. This avoids the undesirable time taken by SA algorithm when the solution is close to the optimal solution. The steps of proposed hybrid algorithm are enumerated as below :

1. Initialize the temperature T_{in}.
2. Compute the energy U of the configuration.
3. Perturb the system slightly with suitable Gaussian disturbance.
4. Compute the new energy $U^{'}$ of the perturbed system and evaluate the change in energy $\Delta U = U^{'} - U$.
5. If $(\Delta U < 0)$, accept the perturbed system as the new configuration Else accept the perturbed system as the new configuration with a probability $exp(-\Delta U)/t$ (where t is the temperature of cooling schedule).
6. Decrease the temperature according to the cooling schedule.
7. Repeat steps 2-7 till some prespecified number of epochs.
8. Compute the energy U of the configuration.
9. Perturb the system slightly with suitable Gaussian disturbance.
10. Compute the new energy $U^{'}$ of the perturbed system and evaluate the change in energy $\Delta U = U^{'} - U$.
11. If $(\Delta U < 0)$, accept the perturbed system as the new configuration, otherwise retain the original configuration.
12. Repeat steps 8-12, till the stopping criterion is met. The stopping criterion is the energy $(U < threshold)$.

2.4. Temporal Segmentation

In temporal segmentation, a change detection Mask (CDM) is obtained and this CDM serves as a precursor for detection of foreground as well as background. This CDM is obtained by taking the label difference of two consecutive frames followed by thresholding. We have adopted a global thresholding method such as Otsu's method for thresholding the image. The results, thus obtained are verified and compensated by historical information, to enhance the segmentation results of the moving object. Thus the results obtained are compared with that of the CDM constructed with taking intensity difference of two consecutive frames. Where we found that label difference as that of intensity difference give better results. The historical information of a pixel means whether or not the pixel belongs to the moving object parts in the previous frame. This is represented as follow

$$H = \{h_s | 0 \leq s \leq (M_1 - 1)(M_2 - 1)\} \qquad (10)$$

Where H is a matrix of size of a frame. If a pixel is found to have $h_s = 1$, then it belongs to moving object in the previous frame; otherwise it belonged to the background in the previous frame. Based on this information, CDM is modified as follows. If it belongs to a moving object part in the previous frame and its label obtained by segmentation is same as one of the corresponding pixels in the previous frame, the pixel is marked as the foreground area in the current frame.

2.5. VOP Generation

The Video Object Plane (VOP) is obtained by the combination of temporal segmentation result and the original video image frame. In a given scene we consider objects as one class and background as the other thus having a two class problem of foreground and background. Therefore, the temporal segmentation results yield two classes. We denote FM_t and BM_t as the foreground and background part of the CDM_t respectively. The region forming foreground part in the temporal segmentation is identified as object and is obtained by the intersection of temporal segmentation and original frame as $VOP = num(FM_t \cap y_t)$ .Where the num (.) is the function counting the number of pixel forming the region of interest.

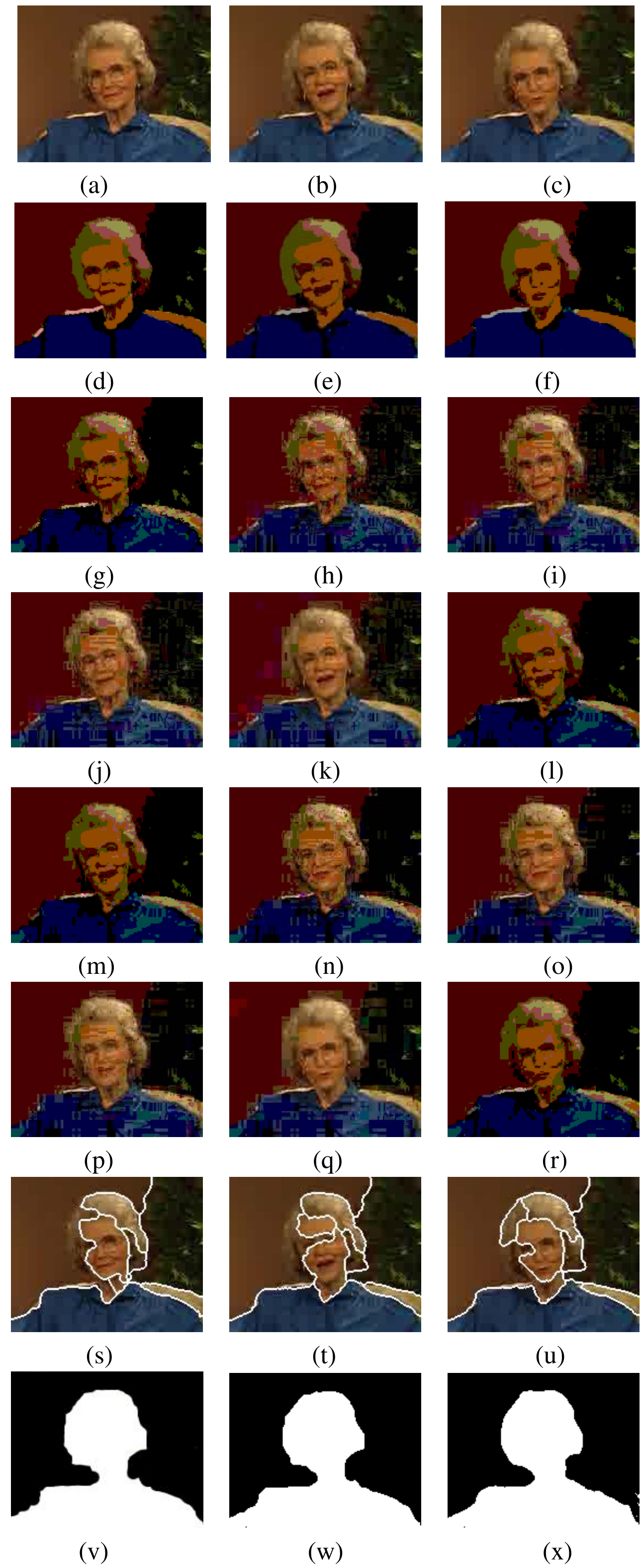
3. Simulation

We have considered three types of video sequences as shown in Fig. 1, Fig. 2, Fig. 3. Fig. 1 corresponds to slow movements of the sequence where as Fig. 2 and Fig. 3 corresponds to video sequences with fast

Table 1. Percentage of Misclassification Error

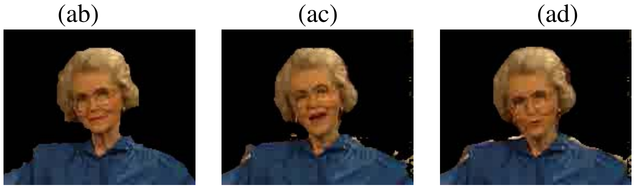
$Video$	$FrameNo.$	$Evolving$	$JSEG$
$Grandma$	12	0.24	6.82
	37	0.15	0.65
	62	0.15	4.5
$Akiyo$	75	0.12	6.20
	95	0.10	1.62
	115	0.15	1.65
$Container$	4	0.10	2.55
	12	0.11	1.51
	20	0.13	2.08

moving objects. Fig. 1(a) shows Grandma image of 12th frame and 1(b) and 1 (c) corresponds to 37th and 62 nd frame. It is observed from these frames that there are slow changes. The corresponding ground truth image constructed manually are shown in Fig. 1(d), (e) and (f). Fig. 1(g) shows the spatial segmentation obtained using the CMRF Model (Compound markov Random Field Model) and hybrid algorithm. The MRF model parameters chosen are $\alpha = 0.05, \beta = 0.009, \gamma = 0.007, \sigma = 5.2$. Fig.1(g) evolves to produce the initial segmentation results corresponding to 18, 24, 30 and 37th frame as shown in Fig. 1(h), (i), (j) and (k) respectively. Using 1(k) as the crude segmentation ICM is run to obtain the segmentation of 37th frame as shown in Fig. 1(l). Analogously for the 62nd frame segmentation result of 37th frame evolves to obtain crude segmentation of 62nd frame as shown in Fig. 1(q). ICM is run starting Fig. 1(q) and the segmented results obtained for 62nd frame is shown in Fig. 1(r). The temporal segmentation result obtained using the segmented result instead of original frames are shown in Fig. 1(v) to (x) and the corresponding VOPs are shown in Fig. 1(y) to (aa). It is observed from these VOP that the objects (i.e Grandma with slow moments) in different frames have been detected. Temporal segmentation using the original frames are shown in 1(ae), (af) and (ag). It is observed from these figures that there are some white portion appearing near the solder of the Grandma that leads to misclassification. Thus, temporal segmentation obtained using the segmented frame yields better VOPs than that of using the original frames. The results obtained by JSEG method is shown in Fig. 1(s), (t) and (u). The %age of misclassification error is given in Table. 1 and it can be observed that the proposed method has less misclassification error as compared to JSEG method.



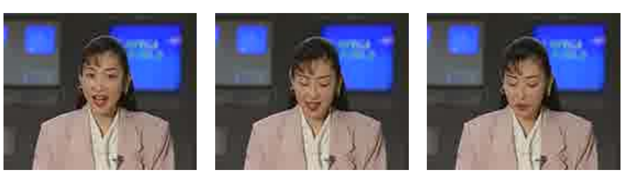
(y) (z) (aa)

(ab) (ac) (ad)

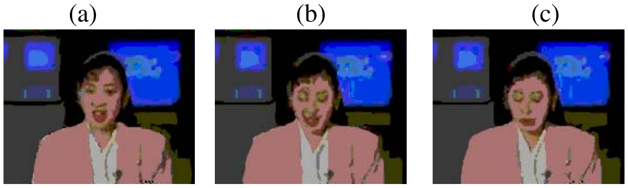


(ae) (af) (ag)

Figure 1. Grandma Video (a)-(c)Original Frame No.12,37,62, (d)-(f) Ground truth of Frame No.12,37,62, (g)segmentation of Frame No.12 with Edge based Compound MRF Model, (h)-(k) Evolving Crude result of Frame No. 18,24,30,37, (l)-(m) Segmentation of Frame No.37 using Evolving scheme, (n)-(q) Evolving Crude Result of Frame No. 41,47,53,62, (r) Segmentation of Frame No.62 using Evolving Scheme, (s)-(u) Segmentation Result using JSEG Scheme, (v)-(x) Temporal Segmentation Result using Segmented Result CDM, (y)-(aa) VOP Extracted using Temporal Segmentation Result (v) to (x), (ab)-(ad) Temporal Segmentation Result using Original Frame CDM, (ae)-(ag) VOP Extracted using Temporal Segmentation Result (ab) to (ad)



(a) (b) (c)

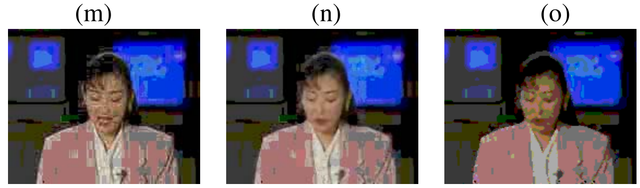


(d) (e) (f)

(g) (h) (i)

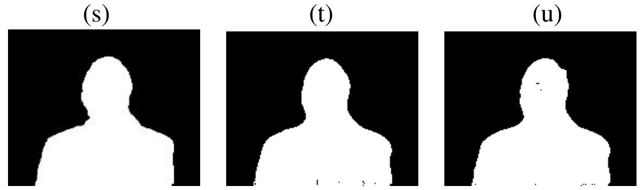
(j) (k) (l)

(m) (n) (o)



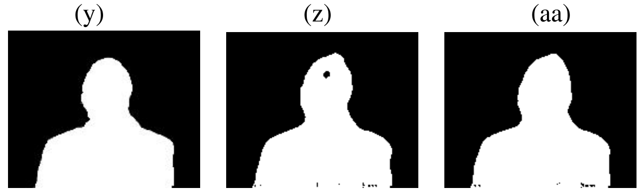
(p) (q) (r)

(s) (t) (u)



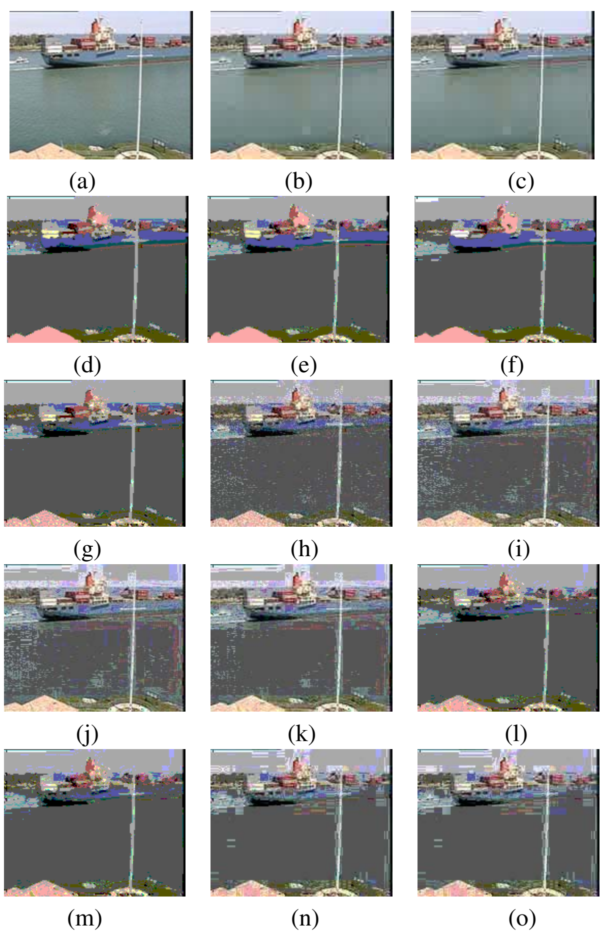
(v) (w) (x)

(y) (z) (aa)



(ab) (ac) (ad)

Figure 2. Akiyo Video (a)-(c)Original Frame No.75,95,115, (d)-(f) Ground truth of Frame No.75,95,115, (g) Segmentation of Frame No.75 with Edge based Compound MRF Model, (h)-(k) Evolving Crude Result of Frame No. 79,83,87,95, (l)-(m) Segmentation of Akiyo video Frame No.95 using Evolving scheme, (n)-(q) Evolving Crude Result of Frame No. 100,105,110,115, (r) Segmentation of Akiyo video Frame No.115 using Evolving Scheme, (s)-(u) Segmentation Resul using JSEG Scheme, (v)-(x) Temporal Segmentation Result using Segmented Result CDM, (y)-(aa) VOP Extracted by Evolving Scheme using Temporal Segmentation Result (v) to (x), (ab)-(ad) Temporal Segmentation Result using Original Frame CDM, (ae)-(ag) VOP Extracted using Temporal Segmentation Result (ab) to (ad)



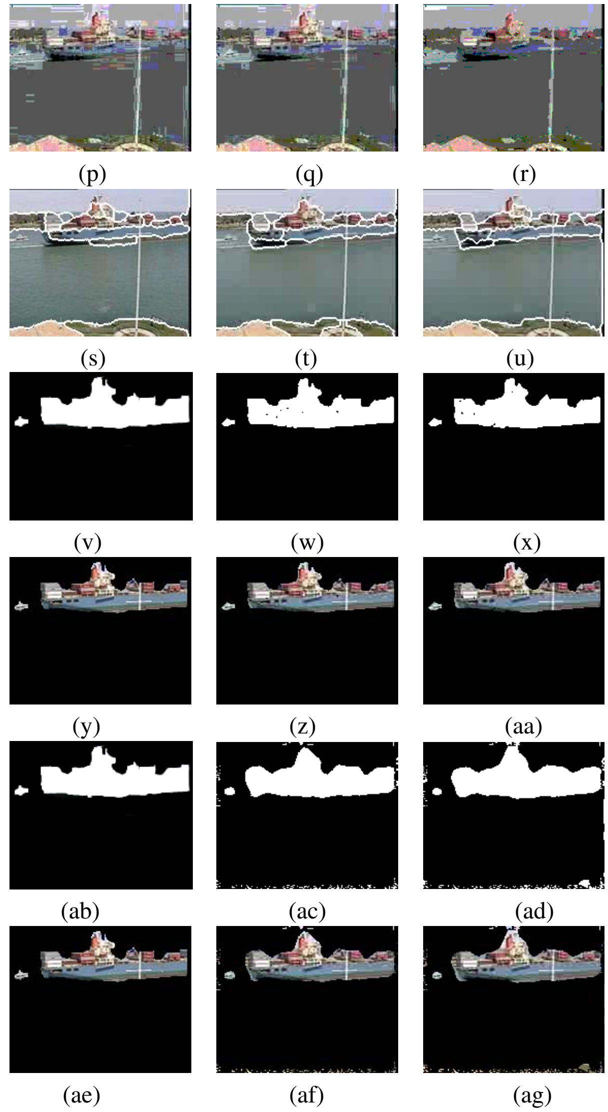


Figure 3. Container Video(a)-(c)Original Frame No.4,12,20, (d)-(f) Ground truth of Frame No.4,12,20, (g) Segmentation of Frame No.4 with Edge based Compound MRF Model, (h)-(k) Evolving Crude Result of Frame No. 6,8,10,12, (l)-(m) Segmentation of Frame No.12 using Evolving scheme, (n)-(q) Evolving Crude Result of Frame No. 14,16,18,20, (r) Segmentation of Frame No.20 using Evolving Scheme, (s)-(u) Segmentation Result using JSEG Scheme, (v)-(x) Temporal Segmentation Result using Segmented Result CDM, (y)-(aa) VOP Extracted by Evolving Scheme using Temporal Segmentation Result (v) to (x), (ab)-(ad) Temporal Segmentation Result using Original Frame CDM, (ae)-(ag) VOP Extracted using Temporal Segmentation Result (ab) to (ad)

Table 2. Time required for execution of the programme in Second

$Video$	$FrameNo.$	$EdgeBased$	$Evolving$
$Grandma$	37	104	9
	65	104	9
$Akiyo$	95	82	8
	115	82	8
$Container$	12	112	12
	20	112	12

The second video is Akiyo video sequence as shown in Fig. 2 where it is observed that there are more changes in the moving part. The evolving frames are also shown in Fig. 2. The VOPs obtained using the segmented results are shown in Fig. 2(y), (z) and (aa). It is observed that the moving parts have been detected. The temporal segmentation using original frame is shown in Fig. 2(ab), (ac) and (ad) and the corresponding VOPs are shown in Fig. 2(ae), (af) and (ag). As observed some background portion have been reflected here. Similar observation is also made for the container video sequence of Fig. 3. In both the cases the $\%$age of misclassification error is less than that of JSEG method and is given in Table. 1. Thus the proposed method proved to be very effective in both slow as well as fast moving objects. We have implemented this algorithm in a $Pentium4(D)$, $3GHz$, $L2$ *cache* $4MB$, $1GBRAM$, $667FSB$ PC. The execution time for different video image frames using the above configuration is tabulated in Table 2

4. Conclusion

We propose a scheme of spatial segmentation based on the notion of evolution. The spatio-temporal frame of the first frame is obtained and this segmented one evolves to generate the segmentation of subsequent. This avoids to compute the spatio-temporal segmentation of each frame and thus the proposed scheme is 13 times faster than that of the scheme computing spatio-temporal segmentation of each frame. The compound Markov Random Field Model is used to model the image. The parameters are selected on trial and error basis. The proposed scheme is found to be better than JSEG method.

References

[1] S. D. Babacan and T. N. Pappas. Spatiotemporal algorithm for joint video segmentation and foreground detection. *Proc. of EUSIPCO, Florence, Italy*, September 2006.

[2] S. D. Babacan and T. N. Pappas. Spatiotemporal algorithm for background subtraction. *Proc. of IEEE International Conf. on Acoustics, Speech, and Signal Processing (ICASSP 07)*, pages 1065–1068, April 2007.

[3] J. Besag. On the statistical analysis of dirty pictures. *Journal of Royal Statistical Society Series B (Methodological)*, 48(3):259–302, 1986.

[4] A. L. Bovic. *Image and Video Processing*. Academic Press, New York, 2000.

[5] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 23:800–810, 2001.

[6] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6):721–741, November 1984.

[7] R. O. Hinds and T. N. Pappas. An adaptive clustering algorithm for segmentation of video sequences. *Proc. of International Conference on Acoustics, speech and signal Processing (ICASSP)*, 4(6):2427–2430, May 1995.

[8] S. S. Huang and L. Fu. Region-level motion-based background modeling and subtraction using mrfs. *IEEE Transactions on image Pocessing*, 16(5):1446–1456, May 2007.

[9] S. W. Hwang, E. Y. Kim, S. H. Park, and H. J. Kim. Object extraction and tracking using genetic algorithms. *Proc. of International Conference on Image Processing ,Thessaloniki, Greece*, 2:383–386, 7-10th October 2001.

[10] B. G. Kim, D. J. Kim, and D. J. Park. Noval precision target detection with adaptive thresholding for dynamic image segmentation. *Machine Vision and Applications, Springer-Verlag*, 12:259–270, February 2001.

[11] E. Y. Kim and K. Jung. Geenetic algorithms for video segmentation. *Pattern Recognition*, 38(1):59–73, 2005.

[12] E. Y. Kim and S. H. Park. Automatic video segmentation using genetic algorithms. *Pattern Recognition Letters*, (11), 2006.

[13] E. Y. Kim, S. H. Park, and H. J. Kim. A genetic algorithm-based segmentation of random field modeled images. *IEEE Signal Processing letters*, 7(11):301–303, 2000.

[14] S. C. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[15] S. Z. Li. *Markov field modeling in image analysis*. Springer, Japan, 2001.

[16] P. Salember and F. Marques. Region based representation of image and video segmentation tools for multimedia services. *IEEE Trans. Circuit systems and video Technology*, 9(8):1147–1169, December 1999.

[17] B. N. Subudhi and P. K. Nanda. Compound markov random field model based video segmentation. *Proc. of SPIT-IEEE colloquium and International conference 2007-2008*, 1:97–102, February 2008.

[18] A. M. Teklap. *Digital Video Processing*. Prentice Hall, NJ, 1995.

[19] G. K. Wu and T. R. Reed. Image sequence processing using spatiotemporal segmentation. *IEEE Trans. on circuits and systems for video Technology*, 9(5):798–807, August 1999.