Accepted in

IEEE TENCON -2008 November 18-21 University of Hyderabad, Hyderabad

Detection of Slow Moving Video Objects 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

Detection of Slow Moving Video Objects 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 Electronics and Telecommunication
C. V. Raman College of Engineering
Bhubaneswar, India
pknanda_d13@yahoo.co.in

Abstract— **Often, moving object detection in a video sequence has been achieved a variant of temporal segmentation methods. For slow moving video objects, a temporal segmentation method fails to detect the objects. In this paper, we propose a Markov random Field (MRF) model based scheme to detect slow movements in a video sequence. The proposed scheme is a combination of a proposed spatio-temporal segmentation scheme and temporal segmentation method. A compound MRF model is used in spatio-temporal framework. In this framework, the a priori distribution is MRF and this takes care of spatial distribution of current frame, temporal frames and the Change Detection Masks (CDM) of the temporal frames. The spatio-temporal segmentation problem is formulated as a pixel labeling problem in Maximum a posteriori (MAP) framework. The MAP estimates are obtained using a hybrid algorithm. These estimated labels are used to obtain the Video Object Plane (VOP) and hence the detection of objects. The results are compared with joint segmentation scheme (JSEG). Results presented demonstrate that the proposed scheme with CDM model could detect slow moving video objects.**

I. 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 [1]-[4]. The video segmentation methodologies have extensively used stochastic image models, particularly Markov Random Field (MRF) model, as the model for video sequences [5]-[7]. MRF model has proved to be an effective stochastic model for image segmentation [8]-[10] 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 [5] 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 [11] with different filtering techniques. Extraction of moving object and tracking of the same has been achieved in spatio-temporal framework [12] with Genetic algorithm serving as the optimization tool for image segmentation. Recently, MRF model has been used to model spatial entities in each frame [12] and Distributed Genetic algorithm (DGA) has been used to obtain segmentation. Modified version of DGA has been proposed [6] to obtain segmentation of video sequences in spatio-temporal framework. Besides, video segmentation and foreground subtraction has been achieved using the spatio-temporal notion [13]-[14] 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 [15]. 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 [18]. The problem of extraction of moving target from the background has been investigated [19] where adaptive thresholding based scheme has been employed to segment the images.

In this paper, we propose a scheme to detect slowly moving objects in a video sequence. The movement could be slow enough to be missed by different existing temporal segmentation. A spatio-temporal scheme is proposed to obtain spatial segmentation of a given frame and, in the sequel, use the same results for temporal segmentation. The spatio-temporal scheme is formulated as a pixel labeling problem and the pixel labels are estimated using MAP criterion. MRF model is used to model the label process. In this model the prior distribution takes into account the spatial distribution of a given frame, interactions in a temporal direction, edge maps in temporal direction. The edge maps helps in preserving the edges of the moving objects. In order to detect slow movement we take in to account the changes in the different frames, slow moments in a video could be obtained. In spatio-temporal framework, observed frame is viewed as a degradation of the label process. This degradation of the label process is assumed to be Gaussian. The spatio-temporal segmentation results thus obtained are used to obtain temporal segmentation, which in turn used to construct the video objects plane and hence detection of objects. The MRF model parameters have been selected on trial and error basis. It is found that spatial segmentation for every frame of the sequence is computationally intensive. In order to reduce the computational burden, we obtain the spatial segmentation of the initial frame and use it as the initials one for the next frame. ICM (Iterative Conditional Mode) algorithm is used to obtain these spatial segmentation of the next frame. The spatial segmentation, thus obtained is used as the initial one for the subsequent frames. The proposed scheme has been tested for a wide variety of sequences and it is observed that the model incorporating changes could detect the slow moving objects successfully. The ground truth image constructed manually. The results obtained by the proposed method are compared with the JSEG [17] method and it is found that the proposed method outperformed JSEG in terms of misclassification error.

II. Spatio temporal image modeling and Moving Object Detection

In the proposed scheme moving objects are determined as follows. In a given frame, the segmentation is obtained using spatio-temporal framework. The segmentation problem is formulated as a pixel labeling problem. The pixel labels obtained from spatio-temporal segmentation are used to obtain the temporal segmentation. Subsequently, the temporal segmentation thus obtained is used to obtain the Video Object plane (VOP). The VOP represents the moving object of the given scene or video frame.

A. 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 y_t 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 y.

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. The a priori distribution takes care of the spatial model of X_t, the temporal modeling taking care of X_t, X_{t-1} and X_{t-2} for second order, edge feature modeling in temporal directions. In order to detect slow changes of the object position, we also incorporate the change model into account. We compute the changes from consecutive changes frames and the changes are also incorporated in the a priori model. We compute the changes finding out the change detection mask. 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, four MRF models are used as the spatio-temporal image model. The two temporal direction MRF models are shown in Fig. 1. (a) and (b). Fig.1 (a) correspond to the interaction of pixel x_{st} with the corresponding pixels of x_{t-1} and x_{t-2} respectively. The MRF model taking care of changes in temporal directions 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

$$P\left(X_{st} = x_{st} \middle| X_{qt} = x_{qt}, \forall q \varepsilon S, s \neq q\right)$$
$$= P\left(X_{st} = x_{st} \middle| X_{qt} = x_{qt}, (q,t) \varepsilon \eta_{s,t}\right)$$

where η_{st} is denoted the neighborhood of (s, t) and S denotes spatial Lattice of the frame X_t. For temporal MRF, the following Markovianity is satisfied

$$P\left(X_{st} = x_{st} \middle| X_{pq} = x_{pq}, q \neq t, p \neq s, \forall (s,t) \varepsilon V\right)$$
$$= P\left(X_{st} = x_{st} \middle| X_{pq} = x_{pq}, (p,q) \varepsilon \eta_{s,t}\right)$$

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 $P(X_t)$ 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 $U(x_t) = \sum_{c \varepsilon C} V_c(x_t)$ 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.

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

Analogously in the temporal direction

$$V_{tec}(x) = \begin{cases} +\beta, if\ x_{st} \neq x_{qt} and\ (s,t),(q,t-1) \varepsilon S \\ -\beta, if\ x_{st} = x_{qt} and\ (s,t),(q,t-1) \varepsilon S \end{cases}$$

$$V_{teec}(x) = \begin{cases} +\gamma, if\ x_{st} \neq x_{et} and\ (s,t),(e,t-1) \varepsilon S \\ -\gamma, if\ x_{st} = x_{et} and\ (s,t),(e,t-1) \varepsilon S \end{cases}$$

For the change model, the CDM for different frames are determined with the CDM, the clique potential function is defined as

$$V_{tch}(x) = \begin{cases} +\delta, if\ x_{st} \neq x_{ct} and\ (s,t),(c,t-1) \varepsilon S \\ -\delta, if\ x_{st} = x_{ct} and\ (s,t),(c,t-1) \varepsilon S \end{cases}$$

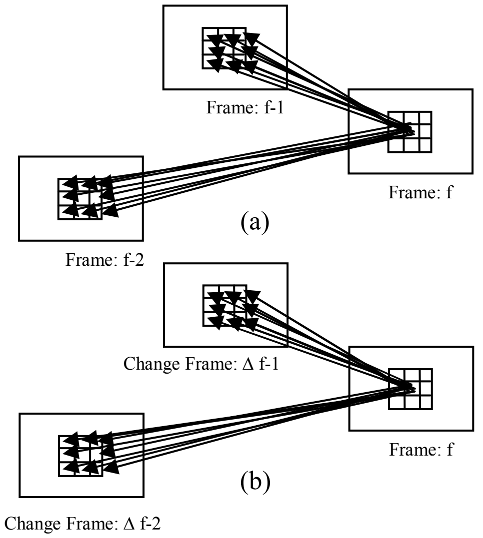


Figure 1. (a) MRF modeling taking two previous frames in the temporal direction (b) MRF with two change frames to incorporate changes

B. Spatio-temporal segmentation in MAP framework

The Segmentation problem is cast as a pixel labeling problem. Let y be the observed video sequence and y_t 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 denote the segmentation of an image at time t. Let X_t denote the random field in the spatial domain at time t. X_t is assumed to be MRF and for proper spatial segmentation we model the prior probability incorporating the following,

(i) Clique potential function in the temporal direction are incorporated.

(ii) The edge maps of each frames is computed and the edge feature in the temporal direction is considered to preserve the edges.

Since, we focus on the detection of slow moving video objects. We have modeled the changes from frame to frame in the MRF-MAP framework. The Change Detection Mask (CDM) of consecutive frames has been determined and the changes are denoted as ΔX_{t-1}. In the prior model of X_t, the changes at Δ_{t-1} and Δ_{t-2} at frames $t-1$ and $t-2$ are incorporated. The corresponding clique potential function is included in the prior distribution of X_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 X_t 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

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

In (4) $V_{sc}(x)$ denotes the clique potential function in the spatial domain at time t, $V_{tec}(x)$ denotes the clique potential in the temporal domain, $V_{teec}(x)$ denotes the clique potential in the temporal domain incorporating edge feature and $V_{tch}(x)$ denotes clique potential incorporating change feature. We have proposed this additional feature in the temporal direction. (4) is called the edgebased change model. 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

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

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

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

In (6) Variance σ^2 corresponds to the Gaussian degradation.

Hence (3) can be expressed as

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

The a priori model having the three components is attributed as the edgebased model. In the following the clique potential corresponding to CDM of different frames have been introduced. This is called the change based model.

$$\hat{x} = \arg\max e^{-\left[\frac{\|y-x\|^2}{2\sigma^2}+\sum_{c\epsilon C} V_{sc}(x)+V_{tc}(x)+V_{tec}(x)+V_{tch}(x)\right]} \qquad (7)$$

Maximizing (7) is tantamount to minimizing the following

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

$\hat{x}$ 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

C. *Hybrid Algorithm*

It is observed that SA algorithm takes substantial amount of time for convergence. This algorithm also helps to come out of the local minima and converge to the global optimum 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. Subsequently, it is assumed that the solution is locally available and hence a local convergent based strategy is adopted for quick convergence. We have used the Iterated Conditional Mode (ICM) [9] as the locally convergent algorithm. A specific number of iterations is fixed by trial and error. This avoids the undesirable time taken by SA when the solution is close to the optimal solution. The steps of the 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 the energy ***ΔU = U' - U***.
5. If ***ΔU < 0,*** accept the perturbed system as the new configuration. Else accept the perturbed system as the new configuration with a probability exp ***(-ΔU / t),*** Where t is the temperature of the cooling schedule.
6. Decrease the temperature according to the cooling schedule**.**
7. Repeat steps 2-7 till some pre specified number of epochs are completed.
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 the energy ***ΔU = U' - U.***
11. If ***ΔU < 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 criteria is the energy U< threshold.

D. *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

$$H = \{h_s \mid 0 \le s \le (M_1 - 1)(M_2 - 1)\}$$

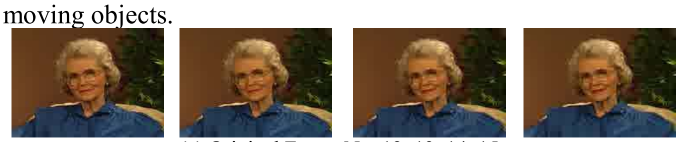
Where H is a matrix of size of a frame. If a pixel is found to have $h_s = 1$, then it belongs to moving object in the previous frame; otherwise it belonged to the background in the previous frame. Based on this information, CDM is modified as follows. If it belongs to a moving object part in the previous 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.

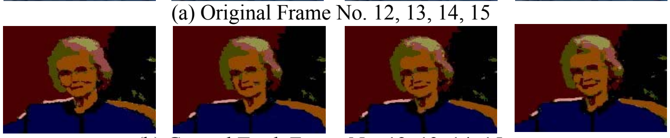
E. *VOP Generation*

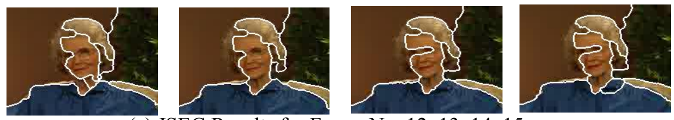
The Video Object Plane (VOP) is obtained by the combination of temporal segmentation result and the original video image frame. In a given scene we consider objects as one class and background as the other thus having a two class problem of foreground and background. Therefore, the temporal segmentation 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.

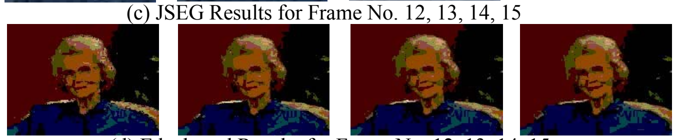
III. RESULTS AND DISCUSSION

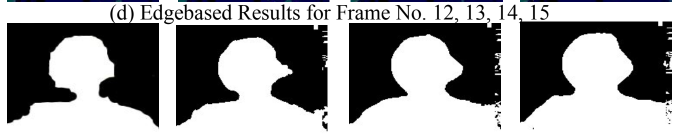
In simulation, we have considered several video sequences, however for the sake of illustration; we have presented the results obtained for two sequences shown in Fig. 2 and 3. Fig. 2(a) shows the original frame with 12,13,14,15 frame numbers that corresponds to Grandma video sequence. It can be observed in these video, there are very slow movements. The manually constructed ground truth images are shown in Fig. 2 (b). The results obtained by JSEG method is shown in Fig. 2 (c). Using edgebased model of section III, the results are shown in fig. 2(d). The image model parameters are $\alpha = 0.05, \beta = 0.009, \gamma = 0.007, \sigma = 5.2$. As seen from the table, the percentage of misclassification error is much less as compared to JSEG method. Temporal segmentation using original frames are shown in Fig. 2(e) and the corresponding VOPs are presented in Fig.2 (f). It is observed that in the VOPs, the part of the silhouette of the image that is reflected at the back of the shoulders is retained. We have used the segmented results of Fig.2 (d) to obtain the temporal segmentation. The temporal segmentation and the corresponding VOPs are shown in Fig.2 (g) and (h) respectively. Comparing Fig.2 (f) and (h) it is observed that inclusion of the edge feature has enhanced the edge preserving capability. But, there are some effects of the silhouette on the background. This effect has been eliminated while using the change based model. The segmented images obtained using the change based model is shown in Fig.2 (i). The temporal segmentation and the corresponding VOPs are shown in Fig.2 (j) and (k). The model parameter used for the change model are $\alpha = 0.05, \beta = 0.009, \gamma = 0.007, \delta = 0.1, \sigma = 5.2$ It has been observed from the Fig. 2(k) that there is no silhouette effect on the background that was present in earlier cases. This could be achieved due to the introduction of changes in the a priori distribution. As seen from Table I, the segmented results shown in Fig. 2(i) is having less percentage of misclassification error as compared to JSEG method and comparable with edge based model. Similar observations are also made for the second Akiyo video sequence shown in Fig.3 (a). The corresponding ground truth images are in Fig.3 (b). The segmentation results obtained by JSEG method are shown in Fig.3 (c) which shows that there are more misclassification. The edge model based segmentation is shown in Fig.3 (d). The model parameters are $\alpha = 0.009, \beta = 0.008, \gamma = 0.007$ and $\delta = 0.1, \sigma = 2$ .The corresponding temporal segmentation and VOPs are shown in Fig.3 (e) and (f) that there are some misclassification near the head of the news reader image. This has been removed partially when the temporal segmentation and VOPs constructed using the segmented results as opposed to using original frames. The misclassification error is also much less as compared to JSEG method. Temporal segmentation and VOPs obtained using change model are shown in Fig. 3(j) and (k). It is clear that there is no silhouette in the background of the image and thus the slow moving object could be detected. Thus the change based model could effectively detect the slow moving objects.

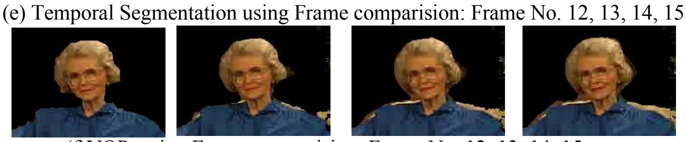

(a) Original Frame No. 12, 13, 14, 15

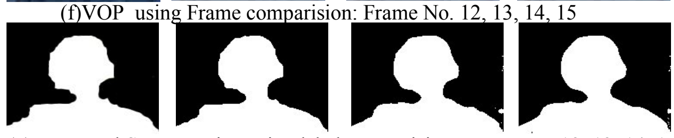

(b) Ground Truth Frame No. 12, 13, 14, 15

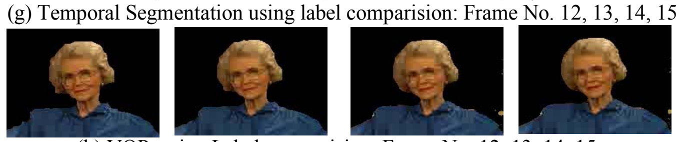

(c) JSEG Results for Frame No. 12, 13, 14, 15

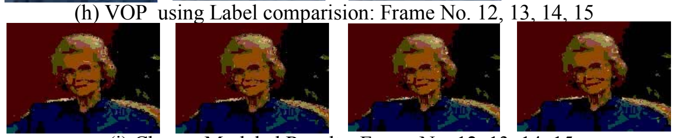

(d) Edgebased Results for Frame No. 12, 13, 14, 15

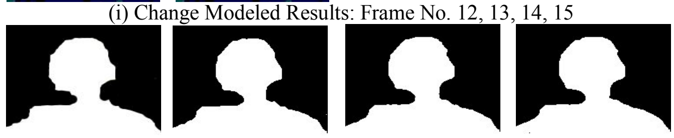

(e) Temporal Segmentation using Frame comparision: Frame No. 12, 13, 14, 15

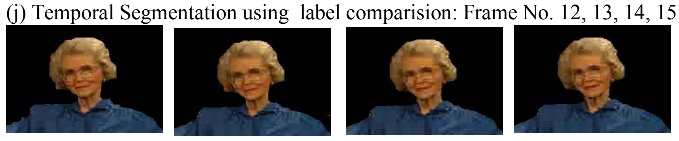

(f)VOP using Frame comparision: Frame No. 12, 13, 14, 15


(g) Temporal Segmentation using label comparision: Frame No. 12, 13, 14, 15


(h) VOP using Label comparision: Frame No. 12, 13, 14, 15

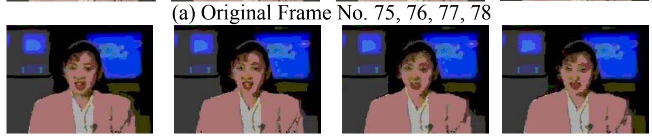

(i) Change Modeled Results: Frame No. 12, 13, 14, 15

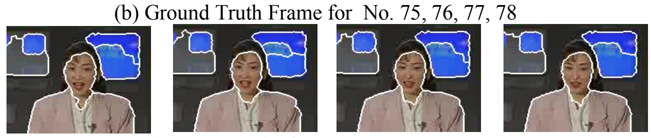

(j) Temporal Segmentation using label comparision: Frame No. 12, 13, 14, 15

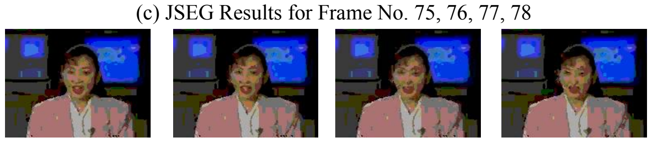

(k) VOP using Label comparision: Frame No. 12, 13, 14, 15

(Figure.2 Results for Grandma video sequence)

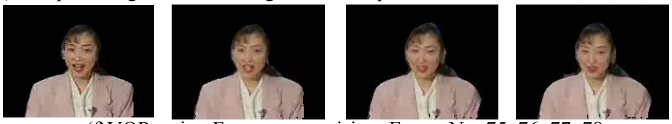

(a) Original Frame No. 75, 76, 77, 78


(b) Ground Truth Frame for No. 75, 76, 77, 78

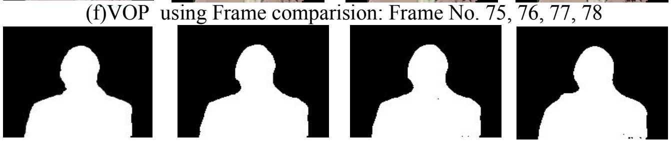

(c) JSEG Results for Frame No. 75, 76, 77, 78


(d) Edgebased Results for Frame No. 75, 76, 77, 78

(e) Temporal Segmentation using Frame comparision: Frame No. 75, 76, 77, 78

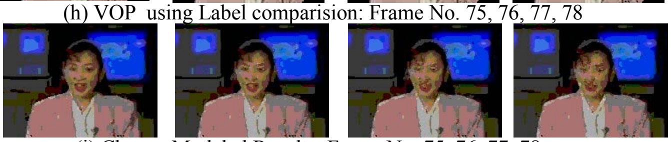


(f)VOP using Frame comparision: Frame No. 75, 76, 77, 78

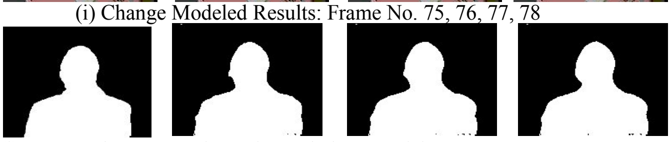


(g) Temporal Segmentation using label comparision: Frame No. 75, 76, 77, 78

(h) VOP using Label comparision: Frame No. 75, 76, 77, 78



(i) Change Modeled Results: Frame No. 75, 76, 77, 78



(j) Temporal Segmentation using Label comparision: Frame No. 75, 76, 77, 78

(k) VOP using Label comparision: Frame No. 75, 76, 77, 78
(Figure.3 Results for Akiyo video sequence)

TABLE I.

	Percentage of Misclassification Error (Spatial segmentation results)							
	grandma				***Akiyo***			
Frame No	12	13	14	15	75	76	77	78
With JSEG	6.82	3.77	4.29	4.20	6.2	4.9	5.5	5.4
Edge Based	0.24	0.30	0.25	0.29	0.12	0.12	0.10	0.10
Changemodel	0.30	0.40	0.40	0.13	0.12	0.18	0.19	0.22

ACKNOWLEDGMENT

The authors acknowledge the facility provided at IPCV Lab of N. I. T, Rourkela and IACV Lab of C. V. Raman College of Engineering, Bhubaneswar.

IV. CONCLUSION

In this paper, our work focused on the detection of slow moving objects. It could be observed that temporal segmentation alone could detect fast moving objects but failed to detect the slow moving objects. Hence, we have proposed a scheme to detect slow moving objects. The scheme consists of spatio-temporal segmentation. We have proposed a compound MRF model with change features that could detect slow movements in a video sequence. Our scheme does not need to obtain spatial segmentation at each frame and hence reduce computational burden. We have selected the model parameters on trial and error basis. Currently attempts are made to estimate model parameters and hence propose an unsupervised scheme.

REFERENCES

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

[2] P. Salember and F. Marques, "Region based representation of image and video segmentation tools for multimedia services,*"IEEE Trans. Circuit systems and video Technology*, vol. 9, No. 8, pp. 1147-1169, Dec.1999.

[3] E. Y. Kim, S. H. Park and H. J. Kim, "A Genetic Algorithm-based segmentation of Random Field Modeled images,*" IEEE Signal processing letters*, vol. 7, No. 11, pp. 301-303, Nov. 2000.

[4] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, No. 6, pp. 721-741, Nov. 1984.

[5] 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*, vol. 4, pp. 2427-2430, May. 1995.

[6] E. Y. Kim and K. Jung, "Genetic Algorithms for video Segmentation," *Pattern Recognition*, vol. 38, No. 1, pp.59-73, 2005.

[7] E. Y. Kim and S. H. Park, "Automatic video Segmentation using genetic algorithms,*" Pattern Recognition Letters*, vol. 27, No. 11, pp. 1252-1265, Aug. 2006.

[8] Stan Z. Li, *Markov field modeling in image analysis*. Springer: Japan, 2001.

[9] J. Besag, "on the statistical analysis of dirty pictures," *Journal of Royal Statistical Society Series B (Methodological)*, vol. 48, No. 3, pp.259-302, 1986.

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

[11] G. K. Wu and T. R. Reed, "Image sequence processing using spatiotemporal segmentation,*" IEEE Trans. on circuits and systems for video Technology,* vol. 9, No. 5, pp. 798-807, Aug. 1999.

[12] 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*, vol.2, pp. 383-386, Oct. 2001.

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

[14] 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, Hawaii, USA, pp. 1065-1068, April 2007.

[15] S. S. Huang and L. Fu,*"* Region-level motion-based background modeling and subtraction using MRFs," *IEEE Transactions on image Pocessing*, vol.*16, No. 5, pp.1446-1456, May. 2007.*

[16] S. C. Kirkpatrick, C. D. Gelatt Jr. and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, No. 4598, pp. 671-680, 1983.

[17] Y. Deng and B. S. Manjunath, "Unsupervised Segmentation of color texture regions in image and video,*" IEEE Transaction on pattern Analysis and Machine Intelligence,* vol. 23, No. 8, pp. 800-810, 2001.

[18] B.N.Subudhi & P.K.Nanda, "Compound Markov random field Model based video segmentation, *" Proc. of SPIT-IEEE colloquium and international conference 2007-2008*, vol.1, pp-97-102, 2008.

[19] B. G. Kim, D. J. Kim & D. J. Park, Noval precision target detection with adaptive thresholding for dynamic image segmentation, *MachineVision and Applications, Springer-Verlag*, vol. 12, pp- 259-270, 2001.