Video Image Segmentation and Object Detection Using Markov Random Field Model

Badri Narayan Subudhi

Department Of Electrical Engineering
National Institute of Technology
Rourkela
2008

Video Image Segmentation and Object Detection Using Markov Random Field Model

A Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Technology

(Research)

in

Electronic Systems and Communication

by

Badri Narayan Subudhi

under the guidance of

Dr. Pradipta Kumar Nanda

(Professor)

Department Of Electrical Engineering
National Institute of Technology
Rourkela

2008

National Institute of Technology

Rourkela

CERTIFICATE

This is to certify that the thesis entitled, "**Video Image Segmentation and Object Detection Using Markov Random Field Model**" submitted by **Badri Narayan Subudhi** in partial fulfillment of the requirements for the award of Master of Technology (Research) Degree in **Electrical Engineering** with specialization in "**Electronic systems and Communication**" at the National Institute of Technology, Rourkela (Deemed University) is an authentic work carried out by him under my supervision and guidance.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/Institute for the award of any degree or diploma.

Prof. Pradipta Kumar Nanda

Date:

Acknowledgment

I would like to thank all people who have helped and inspired me during my master study.

I especially want to thank my advisor, Prof. P. K. Nanda, for his guidance during my research and study. His perpetual energy and enthusiasm in research had motivated all his advisees, including me. In addition, he was always accessible and willing to help his students with their research. As a result, research life became smooth and rewarding for me.

In particular, I would like to thank Prof. S. Ghosh, H.O.D of Electrical Engineering who kept an eye on the progress of my work and always was available when I needed his help and advises.

Prof. B. Majhi, Prof. S. Pattanaik and Prof. P. K. Sahu deserve a special thanks as my thesis committee members and advisors.

I would like to express my thanks to Prof. D. Patra and Prof. S. Das for giving me their valuable time during my seminars and also giving me useful advise and supports during my study.

I would also like to gratefully acknowledge the support of a very special individual. He helped me immensely by giving me encouragement and friendly support. Prof. S. K. Patra whose help, stimulating suggestions and encouragement helped me in all the time of research for and writing of this thesis. I can only say a proper thank you.

All my lab buddies at the IPCV Lab of N.I.T. Rourkela and IACV Lab of C.V.Raman College of Engineering made it a convivial place to work. In particular, I would like to thank Research Scholar Mr. Priyadarshi Kanungo who understood me Basic Image Processing and has help a lot to debug small codes in image Processing. I also like to thanks Mrs. Sucheta Panda, Mr. Satya Swaroop

Pradhan, Md. Muzzamil Sani and Saibal Dutta for his friendship and help in the past two years. All other folks, including (just a list of names including all folks in *smile*), had inspired me in research and life through our interactions during the long hours in the lab. Thanks.

I would like to thanks Mr. Rahul Dey for his kind support at IPCV Lab of N.I.T, Rourkela, who has agreed to rectify my mistakes during initial debugging of my codes and also providing me knowledge of Markov Random Field Model.

I also like to thank Mr. Parthajit Mohapatra, with whom I stayed more than one year and shared my all problems. We use to discuss different aspects of problem and its solutions in Signal and Image Processing.

I like to thank Mr. Deepak Kumar Rout with whom I shares a lots of difficulties and discussed about the problem in my thesis work and who is also going to do future aspects of work in this problem. Also like to thanks Mr. C. R. Padhy who helped me a lots during my staying at Bhubaneswar.

I also like to thanks Mr. Prasant Kumar Pradhan and Sanatan Mohanty for his valuable suggestion, support and helping hand in every aspect starting from doing my thesis work to submitting of this thesis work in last two years at N.I.T, Rourkela. They realy needs a special thanks.

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

My deepest gratitude goes to my family for their unflagging love and support throughout my life; this dissertation is simply impossible without them. I am indebted to my father and Mother Mr. Ananda Chandra Subudhi and Subasini Subudhi, as traditional Father and Mother from a Kumuti Family who allow me to attain such a higher study and also provided me all supports during my study. I also like to thanks My Brother Mr. Rashmi Ranjan Subudhi for providing all financial and moral support during my study. Also like to thank my brother in

law Mr. Ranjan Kumar Prusty for being a supporting guardian to interact with my teacher.

Last but not least, thanks be to God for my life through all tests in the past years. You have made my life more bountiful. May your name be exalted, honored, and glorified.

Badri Narayan Subudhi

Roll No. 60602003

Contents

List of Figures

List of Tables

Abstract

In this dissertation, the problem of video object detection has been addressed. Initially this is accomplished by the existing method of temporal segmentation. It has been observed that the Video Object Plane (VOP) generated by temporal segmentation has a strong limitation in the sense that for slow moving video object it exhibits either poor performance or fails. Therefore, the problem of object detection is addressed in case of slow moving video objects and fast moving video objects as well. The object is detected while integrating the spatial segmentation as well as temporal segmentation. In order to take care of the temporal pixel distribution in to account for spatial segmentation of frames, the spatial segmentation of frames has been formulated in spatio-temporal framework. A Compound MRF model is proposed to model the video sequence. This model takes care of the spatial and the temporal distributions as well. Besides taking in to account the pixel distributions in temporal directions, compound MRF models have been proposed to model the edges in the temporal direction. This model has been named as edgebased model. Further more the differences in the successive images have been modeled by MRF and this is called as the change based model. This change based model enhanced the performance of the proposed scheme.

The spatial segmentation problem is formulated as a pixel labeling problem in spatio-temporal framework. The pixel labels estimation problem is formulated using Maximum a posteriori (MAP) criterion. The segmentation is achieved in supervised mode where we have selected the model parameters in a trial and error basis. The MAP estimates of the labels have been obtained by a proposed Hybrid Algorithm is devised by integrating that global as well as local convergent

criterion. Temporal segmentation of frames have been obtained where we do not assume to have the availability of reference frame. The spatial and temporal segmentation have been integrated to obtain the Video Object Plane (VOP) and hence object detection

In order to reduce the computational burden an evolutionary approach based scheme has been proposed. In this scheme the first frame is segmented and segmentation of other frames are obtained using the segmentation of the first frame. The computational burden is much less as compared to the previous proposed scheme.

Entropy based adaptive thresholding scheme is proposed to enhance the accuracy of temporal segmentation. The object detection is achieved by integrating spatial as well as the improved temporal segmentation results.

Chapter 1

INTRODUCTION

Video Segmentation and Object detection and tracking are quite challenging and active research areas in Video Processing and Computer Vision [38], [39]. The problem of segmentation and tracking a Video Object has wide applications such as video coding, video retrieval, video surveillance and video editing [6]-[11]. Temporal segmentation methods have been proposed to construct Video Object Planes (VOPs) [6]-[20]. Temporal Segmentation based on intensity difference has been proposed by M. Kim et al. [6], Which includes a statistical hypothesis test based on variance comparison. They have also introduced watershed based spatial segmentation and finally a combination of spatial as well as temporal segmentation is proposed to generate Video Object Plane (VOP) and hence object detection. The proposed scheme could satisfactorily separate background and moving objects of a video sequence. Automatic segmentation scheme with morphological method filter has been proposed [8] to detect moving objects. Subsequently the object track matcher using active contour model is proposed to track and match objects in the subsequent frames. Object detection and tracking becomes a hard problem when there is variation of illumination in the video sequence. A. Cavallaro and T. Ebrahimi [7] have proposed a color edge based detection scheme

for object detection. Specifically the color edge detection scheme has been applied to the difference between the current and a reference image. This scheme is claimed to be robust under illumination variation. In order to obtain refinement for the object boundary in the video sequence, a supervised video object segmentation has been proposed [9]. Where the algorithm consists of three steps (i) Semiautomatic first frame segmentation (ii) Automatic Object tracking and (iii) Boundary refinement. The algorithm has been claimed to have satisfactory results under semiautomatic framework. A novel method of separation of moving object from background has been proposed [10], for realtime implimentation. The algorithm is based on the notions of clustering. The algorithm also handles illumination variation of the whole sequence. There has been a wide variation to measure the quality of the object detected in a video sequence. Ç. E. Erdem et al. [11] have developed quantitative performance measures for video object tracking and segmentation that do not requires ground truth segmentation results. They have proposed several interesting quantitative measures for the quality of the video tracking. Edge based detection techniques has also been proposed by J. Zhang et al. [12], where pixel history and moving object masks are used to update background. Connected component analysis and morphological filtering are employed to obtain accurate VOPs. The computational time is reduced by a novel object tracking window. The object detection problem becomes quite challenging when the size of the object is very small as compared to the size of the background. S. Sun et al. [13] have proposed local adaptive threshold methods to determine salient areas in a frame. Thereafter local thresholding is proposed to the local region of interest. The second step segments the target silhouettes precisely and finally the notion of template matching is carried out to remove clutters and hence detection of small targets. Deng and Manjunath et al. [14] have proposed an unsupervised segmentation approach for video sequences. Their method

is known as Joint Segmentation (JSEG) method consists of two independent steps. (i) Color quantization and (ii) Spatial segmentation. Based on color quantization a class-map of the image is created and thereafter the spatial segmentation of the regions are obtained by a region growing approach. A Interpolated Bezier Curve Based Representation scheme [41] is also proposed to recognize the face. An object detection scheme using direct parametric approach in the tomographic images [40] are also proposed

Stochastic model [15] particularly Markov Random Field Models, have been extensively used [16]-[17] for image restoration and segmentation. MRF model, because of its attribute to model spatial dependency, proved to be better model for image segmentation. MRF model has also been used for video segmentation. R. O. Hinds and T. N. Pappas [19] have modeled the video sequence as a 3-D Gibbs Random Fields. In order to obtain smooth transition of segmentation results from frame to frame, temporal constraints and temporal local intensity adaptation are introduced. In order to reduce computational burden, multiresolution approach is adhered. Gibbs Markov Random Field Model has been used to obtain 3-D spatio-temporal segmentation [20]. The region growing approach is used to obtain segmentation. E. Y. Kim et al. [21] have used MRF to model each frame sequence and the observed sequence is assumed to be degraded by independent identically distributed (i.i.d) zero mean Gaussian white noise. The problem is formulated as a pixel labeling problem and the pixel labels are estimated using the MAP estimation criterion. The MAP estimates are obtained by Distributed Genetic Algorithm (DGA).

A novel target detection scheme is proposed by B. G. Kim et al. [22] where the adaptive thresholding scheme has been proposed to separate the foreground and background. The intensity distribution of the video sequence has been modeled by

Gaussian distribution and the parameters have been estimated. The background and objects have been classified and thereafter the object is tracked by a centroid algorithm. This has yielded quite satisfactory results.

Recently MRF modeling has been used to model the video sequences but the segmentation problem has been formulated using Spatio-temporal framework [23]. The segmentation obtained is combined with the temporal segmentation to detect the moving objects. The MAP estimates of the labels are obtained using Genetic Algorithm. S. W. Hwang et al. [24] have also proposed GA based object extraction scheme where spatial segmentation is obtained using Genetic Algorithm and the spatial segmentation thus obtained is combined with Change Detection Mask (CDM) to detect the objects. E. Y. Kim and K. jung [25] have proposed video segmentation scheme where MRF model is used to model the video sequence and the segmentation problem is formulated in spatio-temporal framework. Distributed Genetic algorithm has been used to obtain the MAP estimates. These MAP estimates are combined with temporal segmentation to obtain the video objects. The results are found to be quite promising. Recently E. Y. Kim and S. H. Park [26] have proposed a video segmentation scheme where the video sequences have been modeled as MRF and the segmentation problem is formulated in spatio-temporal framework. The estimates of the labels are obtained using Distributed Genetic algorithm (DGA). Thereafter temporal segmentation is obtained using CDM as well as the history of the label information of different frames. The object extraction and tracking has been successfully carried out. Quite promising results have been obtained in this scheme. S. Babacan and T. N. Pappas [27] have proposed a scheme where they have modeled video sequences as MRF and the changes in temporal direction have been modeled by a mixture of Gaussian. In this case also the spatial segmentation has been combined with

temporal segmentation to detect the foreground accurately. The authors have also improved the results by proposing a novel scheme for background modeling that exploits spatial and temporal dependency. Satisfactory results have been obtained for both indoor and outdoor surveillance videos. Recently S. S. Hwang et al. [29] have proposed a region based motion segmentation algorithm to obtain a set of motion coherence regions. They have also used MRFs for spatial segmentation and have integrated the spatial as well as temporal sequences to obtain the moving objects in the video sequences.

It has been observed that the spatio-temporal framework can together with temporal segmentation produced better results than that of using temporal segmentation. Thus the label fields play a crucial role for detection and tracking. P. M. jodoin et al. [30] have proposed a segmentation scheme where they have fused two label fields (i) a quickly estimated segmentation map and (ii) the spatial region map that exhibits the shape of the main objects. The scheme could be carefully employed for motion segmentation, motion estimation and occlusion detection. Very recently, Q. Shi and L. Wang [31] have attempted to recognize human actions under semi-markov model framework. The optimization problem is solved by them proposed algorithm analogous to viterbi-like algorithm. H. Zhao et al. [32] proposed a tracking algorithm to track objects in real time circumstances. This method presents a lagrangians based methods to improve the accuracy of tracking. The problem of object tracking in real time environment has been addressed by X. Pan and Y. Wu [33] where gaussian single model (GSM) and markov random Field (MRF) have been used. This method is found to be faster than many other methods and hence suitable for real time implementation. Another method has been proposed by C. Su and A. Amer [34] for real time tracking. The proposed method is computing block thresholds.

Temporal Segmentation

It has been attempted to address the moving object detection using the method of temporal segmentation. It was found that temporal segmentation could help to construct the video Object Plane (VOP) and detect the objects. In all these cases, it was assumed to have reference frames. This scheme produced poor results when the video has slow moving objects. This scheme also failed when reference frame is not available. This motivated to devise new methodologies to take care of slow as well as fast moving video objects in the absence of reference frames.

Often in practice reference frames may not be available. The available video may have slow moving objects and fast moving objects.

Spatio-temporal Framework

In order to address both the above problems, the video object detection problem is formulate in spatio-temporal framework using spatio-temporal formulation, Spatial segmentation is obtained. The problem is formulated as a pixel labeling problem in stochastic framework. Markov Random Field Model is proposed to take care of the spatial distribution of each frames and the distributions frames and the distributions of pixels of frames in temporal directions. The edges in the temporal directions have also been modeled as MRF and hence the a priori distributions of images take into account the distributions and pixels in spatial as well as temporal directions, edges in the temporal direction. In all these cases, the a priori MRF model parameters have been selected on trial and error basis. With this video modeling the label estimation problem has been cast as a Maximum a posteriori (MAP) estimation problem. These MAP estimates of the pixels have been obtained by Simulated Annealing (SA) algorithm. It has been observed that the SA is computationally involved and hence takes appriable amount of time to converge to solutions. In order to reduce the computational burden, the MAP estimates of the pixel labels are obtained by a proposed hybrid algorithm. The hybrid algo-

rithm has been designed based on the notion of the local and global convergence. The pixels labels thus obtained for each frames are being used for temporal segmentation. Temporal segmentation is obtained using the change detection masks and the history of the labels of different frames. Thereafter, Video Object Plane is constructed using the temporal segmentation and original frames. it has been observed that this scheme could detect moving object more efficiently than that of using only temporal segmentation. The edges of the moving objects could be preserved and this could be preserved and this could be attributed to the edge preserving property of the proposed model. The results of this scheme when compared with Joint Segmentation method (JSEG) of [14] are found to be superior to the later.

Spatio-temporal framework with Change based MRF Model

In order to enhance the efficacy of the earlier schemes, a new MRF model for video sequences is proposed. In the frame sequences, there are changes from frame to frame because of the object in the video. We assume these changes not to be abrupt ones and hence are expected to have a temporal neighborhood dependency. These changes in the consecutive frames are modeled as MRF. Therefore the proposed a priori MRF model of the video sequence takes in to account these changes of the frames together with the edges in temporal direction. This new MRF model is used to model the video sequences. The pixel label estimation, temporal segmentation and construction of Video Object planes are obtained as per the earlier scheme suggested.

Evolutionary approach based Object detection

It has been observed in the previous proposed scheme that spatial segmentation of each frame has to be obtained to find out temporal segmentation. Spatial segmentation of every frames is a time consuming procedure and hence the object detection scheme takes appreciable amount of time. This forbids the feasibility

of real time implementation. In order to reduce the computational burden, we compute the spatial segmentation of a given frame using the proposed spatio-temporal approach. The spatial segmentation of subsequent frames are obtained starting from the segmentation of given frame with adaptation strategy. Detection of video object at any frame is obtained using the frame together with the temporal segmentation. Spatial segmentation only one frame is obtained using spatio-temporal formulation of previous section.

Object Detection using Adaptive Thresholding

In temporal segmentation, CDM is obtained using the original frames and global thresholding. The performance detoriates when the frames are noisy or there are variation in conditions of illumination. hence, the notion of adaptive threshold has been adhered to and towards this end, we have proposed entropy based adaptive thresholding to obtain appropriate CDMs and hence the moving object parts of the video sequence. However, the spatio-temporal segmentation in MRF-MAP framework, as mentioned in the previous section is used to obtain the spatial segmentation. This spatial segmentation is combined with adaptive threshoilding based temporal segmentation to construct the VOPs and thus moving object detection. The results obtained using adaptive thresholding is found to be superior to that of using global thresholding method.

The major contribution of these can be summarized below

1. Proposed a compound Markov Random Field Model to obtain Video segmentation in spatio-temporal framework. This was combined with the temporal segmentation to detect object in video frames.

2. Proposed a MRF model based on the changes in the temporal direction and the spatio-temporal segmentation scheme. This scheme together with temporal segmentation could detect slow as well as fast moving video objects.

3. Evolutionary approach is proposed to obtain segmentation of *k th* frame evolving from the segmentation result of the initial frame. This is combined with the temporal segmentation method to detect slow as well as fast moving objects.

The organization of the thesis is as follows.

A brief background on MRF is provided in the Chapter 2. The proposed MRF model is described in Chapter 4. and the pixel label estimation problem is formulated in spatio-temporal framework. Hybrid algorithm is also presented in Chapter 4. The a priori MRF model with changes of different frames is also presented in chapter 5. Evolutionary approach based spatial segmentation is formulated in Chapter 6. Adaptive thresholding based temporal segmentation and the object detection scheme is dealt in Chapter 8. Conclusions for different chapters have been in Chapter 9.

Chapter 2

BACK GROUND ON MARKOV RANDOM FIELD MODEL

Random fluctuation in intensity, color, texture, object boundary, or shape can be seen in most real world images. The causes for these fluctuations are diverse and complex, and they are often due to factors such as non-uniform lighting, random fluctuations in object surface orientation and texture, complex scene geometry, and noise. Consequently, the processing of such images become a problem of statistical inference, which requires the definition of a statistical model corresponding to the image pixels.

Although simple image models can be obtained from image statistics such as the mean, variance, histogram and correlation function, a more general approach is to use random fields. Indeed, as a two dimensional extension of the one-dimensional random process, a random field provides a complete statistical characterization for given class of images. Combined with various frameworks for statistical inference, such as Maximum Likelihood (ML) and Bayesian estimation, random field models in recent years led to significant advances in many statisti-

cal image processing applications. A landmark paper by Geman and Geman in 1984 addressed Markov Random Field models and has attracted great attention and invigorated research in image modeling. Indeed the MRF, coupled with the Bayesian framework, has been the focus of many studies[16].

MRF theory provides a convenient and consistent way for modeling context dependent entities such as image pixels and correlated features. This is achieved through characterizing mutual influences among such entities using conditional MRF distributions. The MRF theory tells us how to model the a priori probability of contextual dependent patterns, such as textures and object features. A particular MRF model favors the class of patterns encoded by itself by associating them with larger probabilities than other pattern classes. MRF theory is often used in conjunction with statistical decision and estimation theories, so as to formulate objective functions in terms of established optimality principles. Maximum a posteriori (MAP) probability is one of the most popular statistical criteria for optimality and in fact, has been the most popular choice in MRF vision modeling. MRFs and the MAP criterion together give rise to the MAP-MRF framework. This frame work, advocated by Geman and Geman and others, enables us to develop algorithms for a variety of vising problems systematically using rational principles rather than relying on ad hoc heuristics.

An objective function is completely specified by its form, i.e. the parametric family, and the involved parameters. In the MAP-MRF framework, the objective is the joint posterior probability of the MRF labels. Its form and parameters are determined according to the Bayes formula, by those of the joint prior distribution of the labels and the conditional probability of the observed data[35].

2.1 MARKOV RANDOM FIELD AND GIBBS DISTRIBUTION

MRF theory is a branch of probability theory for analyzing the spatial or contextual dependencies of physical phenomena. It is used in visual labeling to establish probabilistic distributions of interacting labels.

2.1.1 Neighborhood System and Cliques

The site in S are related to one another via a neighborhood system. A neighborhood system for S is defined as

$$N = \{N_i \mid \forall i \in S\} \tag{2.1}$$

where N_i is the set of sites neighboring i. The neighboring relationship has the following properties:

1. a site is not neighboring to itself: $i \notin N_i$
2. the neighboring relationship is mutual: $i \in N_{i'} \Leftrightarrow i' \in N_i$

For a regular lattice S, the set of neighbors of i is defined as the set of sites within a radius of $\sqrt{r}$ from i.

$$N_i = \{i' \in S \mid [dist((x_{i'}, y_{i'}), (x_i, y_i))]^2 \leq r, \ i' \neq i\} \tag{2.2}$$

where $dist(A, B)$ denotes the Euclidean distance between A and B and r takes an integer value. The Fig 2.3 shows the first order (η^1) and second order (η^2) neighborhood system.

The pair $(S, N) = G$ constitutes a graph in the usual sense; S contains the nodes and N determines the links between the nodes according to the neighboring

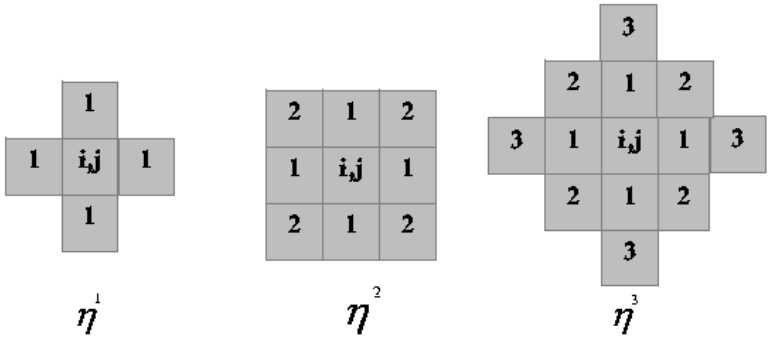


Figure 2.1: Figure showing first order (η^1), second order (η^2) and third order (η^3) neighborhood structure

relationship. A *clique* c for (S, N) is defined as a subset of sites $c = \{i, i'\}$, or a triple of neighboring sites $c = \{i, i', i"\}$, and so on. The collections of single-site, pair-site and triple-site cliques will be denoted by C_1, C_2, C_3, respectively, where

$$C_1 = \{i \mid \ i \ \in \ S\} \tag{2.3}$$

$$C_2 = \{\{i, i'\} \mid \ i' \in N_i, \ i \in \ S\} \tag{2.4}$$

$$C_3 = \{\{i, i', i"\} \ \mid \ i, i', i'' \in \ S \ \ are\, neighbors\, to\, one\, another\} \tag{2.5}$$

The sites in a clique are ordered, and $\{i, i'\}$ is not the same clique as $\{i', i\}$, and so on. The collection of all cliques for (S, N) is

$$C = C_1 \cup C_2 \cup \ C_3 \cup ..... \tag{2.6}$$

The type of a clique for (S, N) of a regular lattice is determined by its size, shape and orientation. Fig 2.4 shows the clique types for the first order and second order neighborhood systems for a lattice[35][16].

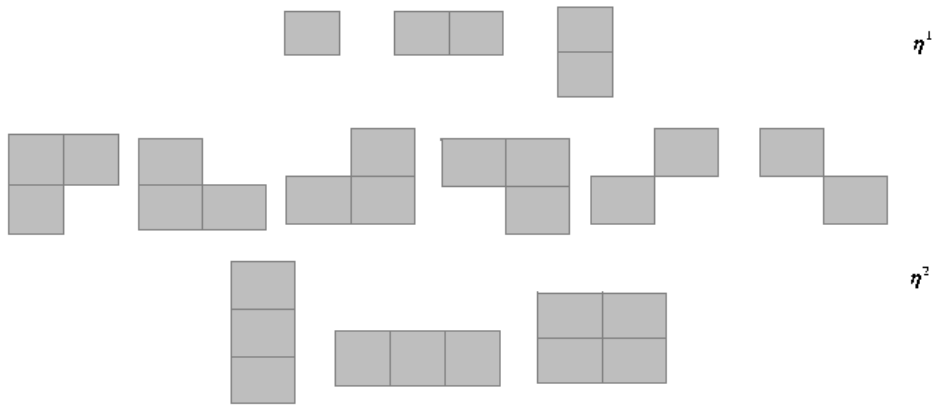


Figure 2.2: Cliques on a lattice of regular sites

2.1.2 Markov Random Field(MRF)

Let $Z = \{Z_1, Z_2, ..., Z_m\}$ be a family of random variables defined on the set S, in which each random variable Z_i takes a value z_i in L. The family Z is called a *random field*. We use the notion $Z_i = z_i$ to denote the event that Z_i takes the value z_i and the notion $(Z_1 = z_1, Z_2 = z_2, ...Z_m = z_m)$ to denote the joint event. For simplicity a joint event is abbreviated as $Z = z$ where $z = \{z_1, z_2, ...\}$ is a configuration of z, corresponding to realization of a field. For a discrete label set L, the probability that random variable Z_i takes the value z_i is denoted $P(Z_i = z_i)$,abbreviated $P(z_i)$, and the joint probability is denoted as $P(Z = z) = P(Z_1 = z_1, Z_2 = z_2, ...Z_m = z_m)$ and abbreviated $P(z)$.
F is said to be a Markov random field on S with respect to a neighborhood system N if and only if the following two conditions are satisfied:

$$P(Z = z) > 0, \quad \forall z \in \mathbf{Z} \qquad (Positivity) \tag{2.7}$$

$$P(z_i|z_{S-i}) = P(z_i|z_{N_i}) \qquad (Markovianity) \tag{2.8}$$

where $S - i$ is the set difference, z_{S-i} denotes the set of labels at the sites in $S - i$ and

$$z_{N_i} = \{z_{i'} | i' \in N_i\} \tag{2.9}$$

stands for the set of labels at the sites neighboring i.

The positivity is assumed for some technical reasons and can usually be satisfied in practice. The Markovianity depicts the local characteristis of Z. In MRF, only neighboring labels have direct interactions with each other[35][16].

The concept of MRF is a generalization of that of Markov processes(MPs) which are widely used in sequence analysis. An MP is defined on a domain of time rather than space. It is a sequence of random variables $.....Z_1, ..., Z_m$ defined in the time indices $....1, ..., m, ...$ It is generalized into MRFs when the time indices are considered as spatial indices.

There are two approaches for specifying an MRF:

1. Conditional probability

2. Joint probability

According to Besag, the conditional approach has the following disadvantages:

1. No obvious method is available for deducing the joint probability from the associated conditional probability.

2. The conditional probability themselves are subject to some non-obvious and highly restrictive consistency conditions.

3. The natural specification of an equilibrium of statistical process is in terms of the joint probability rather than the conditional distribution of the variables.

A theoretical result about the equivalence between MRF and Gibbs distribution (Hammersley and Clifford Theorem) provides a mathematical tractable means of specifying the joint probability of an MRF[35].

2.1.3 MRF models

In the 20's, mostly inspired by the Ising model, a new type of stochastic process appeared in the theory of probability called Markov Random Field. MRF's become rapidly a broadly used tool in a variety of problems not only in statical mechanics. Its use in image processing became popular with the famous paper of S.Geman and D.Geman in 1984 but its first use in the domain dates in the early 70's. Here we briefly give introduction to the theory of some MRF models.

Weak membrane model

The weak membrane model have been introduced in image restoration by A.Blake and A.Zisserman[5] . The problem is to reconstruct surfaces which are continuous almost everywhere or, in other words , continuous in patches. To reach a satisfactory formalization of this principle, they have developed a membrane model: Imagine an elastic membrane which we are trying to fit to a surface, the edge will appear as tears in the membrane. Depending on how elastic is the membrane there may be more or less edges. The membrane is described by an energy function (the elastic energy of the membrane) which has to be minimized in order to find an equilibrium state. The energy has three components:

D: A measure of faithfulness to the data:

$$D = \int (u - d)^2 dA$$

where $u(x, y)$ represents the membrane and $d(x, y)$ represents the data.

S: A measure of how the function $u(x, y)$ is deformed

$$S = \lambda^2 \int (\nabla u)^2 dA$$

P: The sum of penalties α levied for each break in the membrane

$$p = \alpha Z$$

where Z is a measure of the set of contours along which $u(x, y)$ is discontinuous The elastic energy of the membrane is then given by

$$E = D + S + P = \int (u - d)^2 dA + \lambda^2 \int (\nabla u)^2 dA + \alpha Z$$

There is Strong relation between the *weak membrane* model and MRF models. An elastic system can also be considered from a probabilistic view point. The link between the elastic energy and probability P is

$$P \propto exp(\frac{-E}{T})$$

that is the Gibbs distribution. however the *weak membrane model* operates with mechanical analogies, representing *a priori* knowledge from a mechanical point of view, while MRF modelization is purely probabilistic.

Reward Punishment(RP) model

The auto logistic model can be generalized to multi level logistic(MLL) model, also called Strauss process and generalized Ising model. There are $M(> 2)$ discrete label set, $L = 1, 2, ..., M$. In this type of models, a clique potential depends on the type c (related to size, shape and possibly orientation) of the clique and local configuration $f_c \cong f_i | i \epsilon c$. For cliques containing more thanone site $(c > 1)$, the MLL clique potentials are defined by

$$V_c(f) = \begin{cases} +\alpha_c & \text{if all sites on c have the same label} \\ -\alpha_c & \text{otherwise} \end{cases} \tag{2.10}$$

where α_c is the potential for type-c cliques.

We have chosen a simple case of Ising model, In our case we have studied the behavior of reward and punishment given by the model, depending on the homogeneity of the class. If the adjacent pixel is same as that of the center pixel then a reward is assigned to the energy function, otherwise punishment and the amount of reward and punishment is dependent on the homogeneity of the given image. So the clique potential of the model is given by:

$$V_c(z) = \begin{cases} +\delta_c & \text{if } |z_i - z_j| = 0 \\ -\delta_c & \text{if } |z_i - z_j| \neq 0 \end{cases} \tag{2.11}$$

where δ_c is selected on ad hoc manner in our case.

2.1.4 Gibbs Random Field

A set of random variables Z is said to be a *Gibbs random field (GRF)* on S with respect to N if and only if its configuration obey a *Gibbs distribution*. A Gibbs distribution takes the following form.

$$P(Z = z) = \frac{1}{Z'} \times e^{-\frac{U(z)}{T}} \tag{2.12}$$

where

$$Z' = \sum_{z \in \mathbf{Z}} e^{-\frac{U(z)}{T}} \tag{2.13}$$

Z is a normalizing constant called the partition function, T is a constant called the temperature which shall be assumed to be 1 unless otherwise stated, and $U(Z)$ is

the energy function. The energy

$$U(Z) = \sum_{c \in C} V_c(z) \tag{2.14}$$

is a sum of clique potentials $V_c(z)$ over all possible cliques C. The value of $V_c(z)$ depends on the local configuration on the clique C[16][35].

A GRF is said to be homogeneous if $V_c(z)$ is independent of the relative position of the clique c in S. It is said to be isotropic if V_c is independent of the orientation of c. It is considerably simpler to specify a GRF distribution if it is homogeneous or isotropic than one without such properties. The homogeneity is assumed in most MRF vision modes for mathematical and computational convenience. The isotropy is a property of direction-independent blob-like regions[35].

To calculate a Gibbs distribution, it is necessary to evaluate the partition function Z' which is the sum over all possible configurations in **Z**. $P(Z = z)$ measures the probability of the occurrence of a particular configuration, or pattern, z. The more probable configuration are those with lower energies. The temperature T controls the sharpness of the distribution. When the temperature is high, all configurations tend to be equally distributed. Near the zero temperature, the distribution concentrates around the global energy minima.

For discrete labeling problem, a clique potential $V_c(z)$ can be specified by a number of parameters. For example, letting $z_c = (z_i, z_{i'}, z_{i''})$ be the local configuration on a triple clique $c = \{i, i', i''\}$, z_c takes finite number of states and therefore $V_c(z)$ takes a finite number of values. Sometimes, it may be convenient to express the energy of a Gibb's distribution as the sum of several terms, each ascribed to cliques of a certain size, that is,

$$U(z) = \sum_{\{i\} \in C_1} V_1(z_i) + \sum_{\{i,i'\} \in C_2} V_2(z_i, z_{i'}) + \sum_{\{i,i',i''\} \in C_3} V_3(z_i, z_{i'}, z_{i''}) \tag{2.15}$$

The above implies a homogeneous Gibbs distribution because $V_1,\ V_2,\ V_3$ are independent of the locations of $i,\ i',\ i"$. For nonhomogeneous Gibbs distributions, the clique functions should be written as $V_1(i, z_i),\ V_2(i, i'', z_i)$, and so on[35].

2.1.5 Markov-Gibbs Equivalence

An MRF is characterized by its local property whereas a GRF is characterized by its global property. The Hammersley-Clifford theorem establishes the equivalence of these two types of properties. The theorem states that *Z is an MRF on S with respect to N if and only if Z is a GRF on S with respect to N.*

The practical value of the theorem is that it provides a simple way of specifying the joint probability. One can specify the joint probability $P(Z = z)$ by specifying the clique potential functions $V_c(z)$ and choosing appropriate potential functions for desired system behavior. How to choose the forms and parameters of the potential functions for proper encoding of the constraints is a major issue in MRF modeling. The forms of the potential functions determine the forms of the Gibbs distribution. When all the parameters involved in the potential functions are specified, the Gibbs distribution is completely defined.

To calculate the joint probability of an MRF, which is a Gibbs distribution, it is necessary to evaluate the partition function (2.65). Because it is the sum over a combinatorial number of configurations, the computation is usually intractable. The explicit evaluation can be avoided in maximum probability based MRF vision models when $U(z)$ contains no unknown parameters. But this is not true when the parameter estimation is also a part of the problem. In the latter case, the energy function $U(z) = U(z/\theta)$ is also a function of parameters θ and so is the partition function $Z' = Z'(\theta)$. The evaluation of $Z'(\theta)$ is required. To circumvent

the formidable difficulty therein, the joint probability is often approximated in practice[35][16].

2.2 LINE PROCESS

Smoothness is a generic assumption in MRF models which characterizes the spatial coherence and homogeneity of image lattice. However improper imposition of it can lead to undesirable, over-smoothed solutions. It is necessary to take care of discontinuities when using smoothness prior. To avoid the problem of over-smoothing Geman and Geman proposed the underlying MRF (label process) with an additional line process. The line process is neither a data nor the target of estimation. Rather, it is an auxiliary process which is coupled to the label process in such a manner that the joint probability distribution of intensity function is locally smooth with line process for discontinuities. The prior on the line process is often selected to emphasize continuous line and to reject spurious edge elements. Such a model has the desirable property of promoting structure within the image without causing over-smoothing. A couple of MRFs are defined on the image lattice, one is for intensity or label field, other is the dual lattice for the edge field or "line field" . A line process comprises a lattice $S^{'}$ of random variable $f \in F$, whose sites $i^{'} \in S^{'}$ corresponded with vertical and horizontal boundaries between adjacent pixels of the image lattice. It takes the values from $0,\ 1$ which signifies the absence or occurrence of edges. $z_{i'} = 1$ of the line process variable indicates that a discontinuity is detected between the neighboring pixels j and i, i.e. $V_{i,j}(z_i, z_j)$ is taken same before.

Another neighborhood N is defined over the dual lattice $S^{'}$ for line sites. Each pixel has four line site neighbors. Image lattice can be represented as $S \cup S^{'}$. The

(2.62) can be represented with the incorporation of the line fiels as

$$P(Z = z,\ F = f)\ =\ \frac{1}{Z'}e^{-\frac{U(z,f)}{T}} \tag{2.16}$$

The resulting MAP estimation can therefore defined using a Gibbs posterior distribution whose prior energy function is

$$U(z, f)\ =\ U(z|f)\ +\ U(f) \tag{2.17}$$

Assignment of line field is preferred as it results in smaller energy and better estimation[16][35].

Chapter 3

OBJECT DETECTION USING TEMPORAL SEGMENTATION

Segmentation is a process that subdivides an image into its constituent regions or objects.The level to which the subdivision is carried depends on the problem being solved. That is, segmentation should stop when the objects of interest in an application have been isolated. Segmentation of nontrivial images is one of the most difficult tasks in image processing. Motion is a powerful cue used by humans and animals to extract objects of interest from a background of irrelevant detail. Video segmentation refers to the identification of regions in a frame of video that are homogeneous in some sense. Most real image sequences contain multiple moving objects or multiple motions. Motion segmentation refers to labeling pixels that are associated with each independently moving 3-D object in a sequence featuring multiple motions. A closely related problem is optical flow segmentation , which refers to grouping together those optical flow vectors that are associated with the same 3-D motion and/or structure. These two problems are identical when we have a dense optical flow field with an optical flow vector for every pixel. It should not come as a surprise that motion-based segmentation is

an integral part of many image sequence analysis problems, including: improved optical flow estimation, 3-D motion and structure estimation in the presence of multiple moving objects,and higher-level description of the temporal variations and/or the content of video imagery. In the first case, the segmentation labels help to identify optical flow boundaries and occlusion regions where the smoothness constraint should be turned off. Segmentation is required in the second case, because a distinct parameter set is needed to model the flow vectors associated with each independently moving 3-D object. Finally, in the third case, segmentation information may be considered as a high-level (object-level) description of the frame-to-frame motion information as opposed to the low-level (pixel-level) motion information provided by the individual flow vectors. As with any segmentation problem, proper feature selection facilitates effective motion segmentation. In general, application of standard image segmentation methods directly to optical flow data may not yield meaningful results, since an object moving in 3-D usually generates a spatially varying optical flow field. For example in the case of a single rotating object, there is no flow at the center of rotation, and the magnitude of the flow vectors grows as we move away from the center of rotation. The mapping parameters depend on the 3-D motion parameters, the rotation matrix R and the translation vector T, and the model of the object surface, such as the orientation of the plane in the case of a piecewise planar model. Since each independently moving object and/or different surface structure will best fit a different parametric mapping, parameters of a suitably selected mapping will be used as features to distinguish between different 3-D motions and surface structures. Direct methods, which utilize spatio-temporal image gradients may be considered as extension of the case of multiple motion. A suitable parametric motion model has subsequently been used for optical flow segmentation using clustering or maximum a posteriori (MAP) estimation. The accuracy of segmentation results clearly

depends on the accuracy of the estimated optical flow field. As mentioned earlier, optical flow estimates are usually not reliable around moving object boundaries due to occlusion and use of smoothness constraints. Thus, optical flow estimation and segmentation are mutually interrelated, and should be addressed simultaneously for best results. We consider direct methods for segmentations of images into independently moving regions based on spatio-temporal image intensity and gradient information. This is in contrast to first estimating the optical flow field between two frames and then segmenting the image based on the estimated optical flow field. We start with a simple thresholding method that segments images into "changed" and "unchanged regions". Thresholding is often used to segment a video frame into "changed" versus "unchanged" regions with respect to the previous frame. The unchanged regions denote the stationary background, while the changed regions denote the moving and occlusion areas. We define the frame difference $FD_{k,k-1}(x_1, x_2)$ between the frames k and k-1 as

$$FD_{k,k-1}\left(x_1, x_2\right) = s\left(x_1, x_2, k\right) - s\left(x_1, x_2, k-1\right) \tag{3.1}$$

which is the pixel-by-pixeldifference between the two frames. Assuming that the illumination remains more or less constant from frame to frame, the pixel locations where $FD_{k,k-1}(x_1, x_2)$ differ from zero indicate "changed" regions. However, the frame difference hardly ever becomes exactly zero, because of the presence of observation noise. In order to distinguish the nonzero differences that are due to noise from those that are due road scene change , segmentation can be achieved by thresholding the difference image as

$$X = \begin{cases} 1 & \text{if} \quad \mid FD_{k,k-1}(x_1, x_2) \mid > T \\ o & \qquad Otherwise. \end{cases} \tag{3.2}$$

where T is an appropriate threshold.

3.1 IMAGE SEGMENTATION

Segmentation is an important process in automated image analysis. It is during segmentation that regions of interest are extracted from an image for subsequent processing such as surface description and object recognition. It is the low level operation concerned with partitioning images by determining disjoint and homogeneous regions, or, equivalently, by finding edges or boundaries. The homogeneous regions, or the edges are supposed to correspond to actual objects or parts of them within the images. Thus, in a large number of applications in image processing and computer vision, segmentation plays a fundamental role as the first step before applying to images for higher level operations such as recognition, semantic interpretation and representation. Segmentation can be defined as follows:

Let I denote an image and H define a certain homogeneity predicate, then the segmentation of I is a partition P of I into a set of N regions $R_n,\ n = 1, 2, ...N$ such that:

1. $\bigcup_{n=1}^{N} R_n = I$ with $R_n \cap R_m \neq 0, n \neq m$
2. $H(R_n) = TRUE \quad \forall n$
3. $H(R_n \cup R_m) = FALSE \quad \forall R_n$ and R_m adjacent

Condition 1) states that partition has to cover the whole image; condition 2) states that each region has to be homogeneous with respect to predicate H; condition 3) states that no two adjacent region cannot be merged into a single region that satisfies the predicate H. Regions of image segmentation should be uniform and homogeneous with respect to some characteristics such as gray tone, texture or color. Region interiors should be simple and without many small holes. Adjacent

regions of segmentation should have significantly different values with respect to the characteristic on which they are uniform. Boundaries of each segment should be simple, not ragged and must be spatially accurate.

3.2 VIDEO SEGMENTATION

Video segmentation refers to the identification of regions in a frame of video that are homogeneous in some sense. Different features and homogeneity criteria generally leads to different segmentation of same data; for example, color segmentation, texture segmentation, and motion segmentation usually result in segmentation maps. Furthermore, there is no guarantees that any of the resulting segmentation will semantically meaningful, since semantically meaningful region may have multiple colors, multiple textures, or multiple motions. Generally motion segmentation is closely related to two other problems, motion (change) detection and motion estimation. Change detection is a special case of motion segmentation with only two regions, namely changed and unchanged regions(in the case of static cameras) or global and local motion regions(in the case of moving cameras). An important distinction between the change detection and motion segmentation is that the former can achieved without motion estimation if the scene is recorded with a static camera. Change detection in the case of a moving camera and general motion segmentation, in contrast, require some sort of global or local motion estimation, either explicitly or implicitly. It should not come as a surprise that motion/object segmentation is an integral part of many video analysis problems, including (i) improved motion (optical flow) estimation, (ii) three- dimensional (3-D) motion and structure estimation in the presence of multiple moving objects, and (iii) description of the temporal variation or content of video. In the former case,the segmentation labels help to identify optical flow boundaries(motion

edges) and occlusion regions where the smoothness constraint should be turned off. Segmentation is required in the second case, because distinct 3-D motion and structure parameters are needed to model the flow vectors associated with each independently moving objects. Finally in third case segmentation information may be employed in an object level description of frame to frame motion as opposed to a pixel level description provided by individual flow vectors.

Video segmentation has applications in the field of face and gait -based human recognition, event detection, activity recognition, activity based human recognition,detection of the position of the object, detection of the behaviors of the insects, fault diagnosis in rolling plants, visual recognition, detect and model the abnormal behavior of the insects, anomaly detection, tracking, robotics applications, autonomous navigations, dynamic scene analysis, target tracking and path detection etc.

3.3 TEMPORAL SEGMENTATION

Motion is a powerful cue used by humans and many animals to extract objects of interest from a background of irrelevant detail. In imaging applications, motion arises from a relative displacement between the sensing system and the scene being viewed, such as in robotic applications, autonomous navigation and dynamic scene analysis.

3.3.1 Spatial Techniques

Basic approach

One of the simplest approaches for detecting changes between two image frames $f(x, y, t_i)$ and $f(x, y, t_j)$ taken at times t_i and t_j, respectively, is to compare the

two images pixel by pixel. One procedure for doing this is to form a difference image. Suppose that we have a reference image containing only stationary components. Comparing this image against a subsequent image of the same scene, but including a moving object, results in the difference of the two images canceling the stationary elements, leaving only nonzero entries that correspond to the nonstationary image components.

A difference image between two images taken at times t_i and t_j may be defined as

$$d_{i,j}(x,y) = \begin{cases} 1 & \text{if} \quad \mid f(x,y,t_i) - f(x,y,t_j) \mid > T \\ o & \qquad\qquad Otherwise. \end{cases} \tag{3.3}$$

where T is a specified threshold. Note that $d_{i,j}(x,y)$ has a value of 1 at spatial coordinates (x,y) only if the gray-level difference between the two images is appreciably different at those coordinates, as determined by the specified threshold T. It is assumed that all images are of the same size. Finally, we note that the values of the coordinates (x,y) in (3.3) span the dimensions of these images, so that the difference image $d_{i,j}(x,y)$ also is of same size as the images in the sequence.

In dynamic image processing, all pixels in $d_{i,j}(x,y)$ with value 1 are considered the result of object motion. This approach is applicable only if the two images are registered spatially and if the illumination is relatively constant within the bounds established by T. In practice, 1-valued entries in $d_{i,j}(x,y)$ often arise as a result of noise. Typically, these entries are isolated points in the difference image, and a simple approach to their removal is to form 4- or 8-connected regions of 1's in $d_{i,j}(x,y)$ and then ignore any region that has less than a predetermined number of entries. Although it may result in ignoring small and/or slow-moving objects, this approach improves the chances that the remaining entries in the difference

image actually are the result of motion.

Accumulative differences

Isolated entries resulting from noise is not an insignificant problem when trying to extract motion components from a sequence of images. Although the number of these entries can be reduced by a thresholded connectivity analysis, this filtering process can also remove small or slow-moving objects as noted in the previous section. One way to address this problem is by considering changes at a pixel location over several frames, thus introducing a ”memory” into the process. The idea is to ignore changes that occur only sporadically over a frame sequence and can therefore be attributed to random noise.

Consider a sequence of image frames $f(x, y, t_1), f(x, y, t_2).....f(x, y, t_n)$ and let $f(x, y, t_1)$ be the reference image. An *accumulative difference image* (ADI) is formed by comparing this reference image with every subsequent image in the sequence. A counter for each pixel location in the accumulative image is incremented every time a difference occurs at that pixel location between the reference and an image in the sequence. Thus when the kth frame is being compared with the reference, the entry in a given pixel of the accumulative image gives the number of times the gray level at that position was different from the corresponding pixel value in the reference image. Often useful is consideration of three types of accumulative difference images: absolute, positive, and negative ADIs. Assuming that the gray-level values of the moving objects are larger than the background, these three types of ADIs are defined as follows. Let $R(x, y)$ denote the reference image and, to simplify the notation, let k denote t_k, so that $f(x, y, k) = f(x, y, t_k)$. We assume that $R(x, y) = f(x, y, 1)$. Then, for any $k > 1$, and keeping in mind that the values of the ADIs are counts, we define the

following for all relevant values of (x, y):

$$A_k(x,y) = \begin{cases} A_{k-1}(x,y) + 1 & \text{if} \quad \mid R(x,y) - f(x,y,k) \mid > T \\ A_{k-1}(x,y) & Otherwise. \end{cases} \tag{3.4}$$

$$P_k(x,y) = \begin{cases} P_{k-1}(x,y) + 1 & \text{if} \quad \mid R(x,y) - f(x,y,k) \mid > T \\ P_{k-1}(x,y) & Otherwise. \end{cases} \tag{3.5}$$

and

$$N_k(x,y) = \begin{cases} N_{k-1}(x,y) + 1 & \text{if} \quad [R(x,y) - f(x,y,k)] < -T \\ N_{k-1}(x,y) & Otherwise. \end{cases} \tag{3.6}$$

where $A_k(x, y)$, $P_k(x, y)$ and $N_k(x, y)$ are the absolute, positive, and negative ADIs, respectively, after the kth image in the sequence is encountered. It is understood that these ADIs start out with all zero values (counts). The images in the sequence are all assumed to be of the same size. The order of the inequalities and signs of the thresholds in (3.5) and (3.6) are reversed if the gray-level values of the background pixels are greater than the levels of the moving objects.

Establishing a Reference Image

A key to the success of the techniques discussed in the preceding two sections is having a reference image against which subsequent comparisons can be made. As indicated, the difference between two images in a dynamic imaging problem has the tendency to cancel all stationary components, leaving only image elements that correspond to noise and to the moving objects. The noise problem can be handled by the filtering approach mentioned earlier or by forming an accumulative difference image, as discussed in the preceding section.

In practice, obtaining a reference image with only stationary elements is not always possible, and building a reference from a set of images containing one or more moving objects becomes necessary. This necessity applies particularly to situations describing busy scenes or in cases where frequent updating is required. One procedure for generating a reference image is as follows. Consider the first image in a sequence to be the reference image. When a nonstationary component has moved completely out of its position in the reference frame, the corresponding background in the present frame can be duplicated in the location originally occupied by the object in the reference frame. When all moving objects have moved completely out of their original positions, a reference image containing only stationary components will have been created. Object displacement can be established by monitoring the changes in the positive ADI.

3.4 ALGORITHM FOR TEMPORAL SEGMENTATION

The salient steps of the Hybrid Algorithm are as follows

1. Initially two frames are taken one as a reference frame and another frame in which object is present and identification of object is performed on that frame.

2. A Change Detection Mask (CDM) is obtained by taking the difference between the considered frame and the reference frame.

3. The difference between the frame is thresholded by global thresholding approach, which gives a binary image with two regions that is object and background.

4. In final stage the intersection of object region and original image frame is taken to find out the Moving Objects.

3.5 RESULTS AND DISCUSSION

In simulation, two types of situations are considered. The first one is when reference frame is available, while the second one is in the absence of reference frames. Fig. 3.1 shows the Hall monitoring video sequence. The original Hall monitoring sequence which is considered as the reference frame is shown in Fig. 3.1(a). The movement in the hall is shown in different video sequences as shown in Fig. 3.1(b). The change detection masks are shown in Fig. 3.1(c). It is observed from the CDMs that there are many other objects i.e parts of the background presence in the CDMs. Temporal Segmentation is carried out and the corresponding VOPs of different frames are shown in Fig. 3.1(d). It can be observed from Fig. 3.1(d) that the video objects could be detected but there are few other background patches. However, ignoring the minor background patches in the VOPs it can be concluded that with the availability of reference frames, the objects could be detected accurately.

The second example considered is Bowing video sequence as shown in Fig. 3.2. In this case, the reference frame is shown in Fig. 3.2(a). With the activity in the video, frames 57, 58, 59, 60 are shown in Fig. 3.2(b) where the moving object is thee human activity. The CDMs obtained with the use of reference frames contains lots of background information besides foreground information. VOPs are generated using Temporal Segmentation and it is observed that the moving object could be detected with less error. Hence in this case also with reference frame, temporal segmentation could produce better results. The third example is the Hall

monitoring sequence with a different type of activity. The corresponding CDMs are also shown in Fig. 3.3(a) and the CDMs are with more background information. The corresponding VOPs are shown in Fig. 3.3(d) where it can be observed that the object could be detected with vary background patches.

The second case considered is when no reference frame is available. The first example considered is the Akiyo Video sequence as shown in Fig. 3.4(a). The VOPs generated are shown in Fig. 3.4(c) where it can be observed that some parts of the moving object could be detected but in a dithered way. Hence, it can be concluded that without availability of reference frames temporal segmentation method fails to detect the objects. This observation is also corroborated with the second example considered as shown in Fig. 3.5. This is a Grandma video sequence, where reference frame is not available and hence the VOPs are very much distorted as shown in Fig. 3.5(c). It is observed that only some effect of the silhouette is present in the sequence. Thus it can be concluded that temporal segmentation is not suitable for object detection when reference frame is not available.

The limitation of the existing temporal segmentation methods are as follows

1. It does not give good result in presence of noise and illumination variation
2. It can not able to give good result with poor resolution
3. case will be more critical in absence of reference frame
4. It may not give any result if there is slow movements in the sequences.
5. Substantial amount of object movement is required in order to generate reference frame.
6. If Object size is large it may also fails to generate reference frame.

(a) Original Hall Monitoring Video Sequence Frame No.6

(b) Original Hall Monitoring Video Sequence Frame No.49,50,51,52

(c) CDM of Frame No.49,50,51,52 using Frame No. 6 as reference

(d) VOP of Frame No.49,50,51,52

Figure 3.1: VOP Generation of Hall Monitoring Sequence using Temporal Segmentation

(a) Original Bowing Video Sequence Frame No.1

(b) Original Bowing Video Sequence Frame No.57,58,59,60

(c) CDM of Frame No.57,58,59,60 using Frame No. 1 as reference

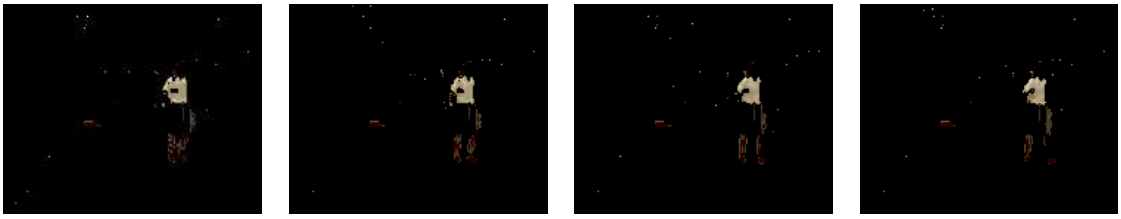
(d) VOP of Frame No.57,58,59,60

Figure 3.2: VOP Generation for Bowing Video Sequence using Temporal Segmentation

(a) Original Hall Monitoring Video Sequence Frame No.6

(b) Original Hall Monitoring Video Sequence Frame No.292, 293, 294, 295

(c) CDM of Frame No.292, 293, 294, 2955 using Frame No. 6 as reference

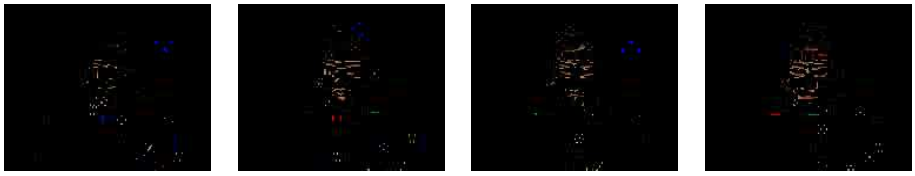


(d) VOP of Frame No.292,293,294,295

Figure 3.3: VOP Generation for Hall Video Sequence using Temporal Segmentation

(a) Original Akiyo Video Sequence Frame No.75

(b) Original Akiyo Video Sequence Frame No.76,77,78,79



(c) VOP of Frame No.76,77,78,79

Figure 3.4: VOP Generation for Akiyo Sequence using Temporal Segmentation

(a) Original Grandma Video Sequence Frame No.11

(b) Original Grandma Video Sequence Frame No.12,13,14,15

(c) VOP of Frame No.12,13,14,15

Figure 3.5: VOP Generation for Grandma Sequence using Temporal Segmentation

Chapter 4

OBJECT DETECTION USING COMPOUND MRF MODEL BASED SPATIO-TEMPORAL SEGMENTATION

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],[2],[21],[16]. The video segmentation methodologies have extensively used stochastic image models, particularly Markov Random Field (MRF) model, as the model for video sequences [19],[25],[26]. MRF model has proved to be an effective stochastic model for image segmentation [35],[17],[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 [19] 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 [20] with different filtering techniques. Extraction of moving object and tracking of the same has been achieved in spatio-

temporal framework [24] with Genetic algorithm serving as the optimization tool for image segmentation. Recently, MRF model has been used to model spatial entities in each frame [24] and Distributed Genetic algorithm (DGA) has been used to obtain segmentation. Modified version of DGA has been proposed [25] to obtain segmentation of video sequences in spatio-temporal framework. Besides, video segmentation and foreground subtraction has been achieved using the spatio-temporal notion [27],[28] 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 [29]. 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. The problem of extraction of moving target from the background has been investigated [22] where adaptive thresholding based scheme has been employed to segment the images.

In this Chapter we propose a scheme to detect moving object in a video sequence. There could be substantial movement in the moving objects from frame to frame of a video sequence or the movement could be slow enough to be missed by temporal segmentation. In order to take care of both the situation, we obtain spatial segmentation of the given frame and in the sequence use the same results to obtain temporal segmentation. The accuracy of temporal segmentation greatly depends upon the accuracy of spatial segmentation. The results of the temporal segmentation is used to obtain the video object plane and hence moving object detection. The spatial segmentation problem is formulated in spatio-temporal framework. A compound MRF model is proposed to model the spatial as well as temporal pixels of the video sequence. The compound MRF model consists of

three MRF, one to model the spatial entities of the given frame; the second MRF model take care of attributes in the temporal direction and the third MRF model is used to take care of edge features in the temporal direction. Thus a compound MRF model is used to model the video. The problem is formulated as a pixel labeling problem and the pixel label estimates are the maximum a posteriori (MAP) estimates of the given problem. By and large the Simulated Annealing (SA) algorithm [16] is used to obtain the MAP estimates, instead we have proposed a hybrid algorithm based on local global attributes to obtain the MAP estimates and hence segmentation. The proposed scheme has been tested for a wide verity of sequences and it is observed that with the proposed edge based compound MRF model yields better segmentation results than that of edgeless model. The ground truth image is constructed manually and the percentage of misclassification is obtained based on the ground truth images. The proposed method is compared with JSEG [14] method and it is found that the proposed method outperformed JSEG in terms of misclassification error.

The pixels labels thus obtained for each frames are being used for temporal segmentation. Temporal segmentation is obtained using the change detection masks and the history of the labels of different frames. Thereafter, Video Object Plane is constructed using the temporal segmentation and original frames. it has been observed that this scheme could detect moving object more efficiently than that of using only temporal segmentation. The edges of the moving objects could be preserved and this could be preserved and this could be attributed to the edge preserving property of the proposed model.The VOP constructed using the edge-based model and it is observed that the video segmentation results has two class, one moving object and the other background. The scheme was tested for different video sequence and even slow movements in the video could be detected.

4.1 MOVING OBJECT DETECTION

Usually, CDM is the difference of two consecutive frames. The gray value of pixels on CDM could be either high due to changes such as motion or significant illumination changes or low due to noise and variation in illumination. These low value changes cause improper generation of VOP. We have proposed a method of obtaining the CDM, where inspite of taking the gray level difference of two consecutive frame, the difference between the label of two consecutive frames, are obtained followed by thresholding. The CDM obtained with that of label difference produces better result than that of using CDM with difference in gray level.

4.2 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 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.

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

where $\eta_{s,t}$ 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.

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

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

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

4.2.1 Segmentation in MAP frame work

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 denote 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.

$$\hat{x} = \arg \max_{x} P(X = x | Y = y) \tag{4.1}$$

Where $\hat{x}$ denotes the estimated labels. Since, x is unknown it is very difficult to evaluate (4.1), hence, using Baye's theorem (4.1) can be written as

$$\hat{x} = \arg \max_{x} \frac{P(Y = y | X = x) P(X = x)}{P(Y = y)} \tag{4.2}$$

Since y is known, the prior probability $P(Y = y)$ is constant. hence (4.2) reduces to

$$\hat{x} = \arg \max_{x} P(Y = y | X = x, \theta) P(X = x, \theta) \tag{4.3}$$

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

$$P(X = x) = e^{-U(x,\theta)} = e^{\left[-\sum_{c\epsilon C}[V_{sc}(x)+V_{tec}(x)+V_{teec}(x)]\right]} \tag{4.4}$$

In (4.4) $V_{sc}(x)$ 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.(4.4) is called the edge-based 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

$$P(N = y - x|X, \theta) = \frac{1}{\sqrt{(2\pi)^n det\,[k]}} e^{-\frac{1}{2}(y-x)^T K^{-1}(y-x)} \tag{4.5}$$

Where k is the covariance matrix. Assuming decorrelation of the three RGB planes and the variance to be same among each plane, (4.5) can be expressed as

$$P(N = y - x|X, \theta) = \frac{1}{\sqrt{(2\pi)^3\sigma^3}} e^{-\frac{1}{2\sigma^2}(y-x)^2} \tag{4.6}$$

In (4.6) Variance σ^2 corresponds to the Gaussian degradation. Hence (4.3) can be expressed as

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

The a priori model having the three components is attributed as edge based model.

$$\hat{x} = \arg\ \max_{x}\ e^{-\left[\frac{\|y-x\|^2}{2\sigma^2}+\sum_{c\epsilon C} V_{sc}(x)+V_{tec}(x)+V_{teec}(x)\right]} \tag{4.7}$$

Maximizing (4.7) is tantamount to minimizing the

$$\hat{x} = \arg\ \min_{x}\ \left\{\left[\frac{\| y-x \|^2}{2\sigma^2}\right] + \left[\sum_{c\epsilon C} V_{sc}(x) + V_{tec}(x) + V_{teec}(x)\right]\right\} \tag{4.8}$$

$\hat{x}$ in (4.8) 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

4.2.2 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) [17] 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 $(U < threshold)$.

4.3 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

$$H = \{h_s | 0 \leq s \leq (M_1 - 1)(M_2 - 1)\} \tag{4.9}$$

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.

4.4 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.

4.4.1 Modification in CDM

By and large CDM is the difference of two consecutive frames. The gray value of pixels on CDM could be either high due to changes such as motion or significant illumination changes or low due to noise and variation in illumination. These low value changes cause improper generation of VOP. We have proposed a method of obtaining the CDM, where inspite of taking the gray level difference of two consecutive frame, the difference between the label of two consecutive frames, are obtained followed by thresholding. The CDM obtained with that of label difference produces better result than that of using CDM with difference in gray level.

4.5 CENTROID CALCULATION ALGORITHM

Using a optimal threshold value and the VOP available for previous frame the temporal segmentation of the current frame is obtained. The cluster of the object region is transformed to a gray level image, where the object region is differed from the background region by two gray level either 0 or 255. Which can be given as,

$$x = \begin{cases} 255 & \text{if} \quad it\ is\ in\ object \\ o & \quad Otherwise. \end{cases}$$

where the centroid $(\hat{x}_{n_c}, \hat{y}_{n_c})$ of the binary temporal segmented image is given as,

$$\hat{x}_{n_c} = \frac{\sum_{i\epsilon T} x_{n_i} c(i)}{\sum_{i\epsilon T} c(i)} \tag{4.10}$$

$$\hat{y}_{n_c} = \frac{\sum_{i\epsilon T} y_{n_i} c(i)}{\sum_{i\epsilon T} c(i)} \tag{4.11}$$

4.6 SIMULATION AND RESULT DISCUSSION

The two video models edge less and edgebased model have been tested with three different video sequences, namely Suzie, Akiyo and Motherbaby video sequences. For these two models, the two different strategies are adopted while obtaining the CDMs. The first one is when the original frame is considered and the second one is when the estimated label frames are considered. In all the cases we have considered RGB color model.

Fig. 4.2 shows the results of the Suzie video sequence. The original sequence is shown in Fig. 4.2(a). It can be seen from the original sequence that there is slow movements of the object in different frames such as 5, 8 and 11. Besides, the reference frame is not available. Hence temporal segmentation method would fail in this case.

Hence, the spatio-temporal segmentation together with the temporal segmentation is used to detect the video objects. The ground truth images for spatial

segmentation are shown in Fig. 4.2(b). For spatial segmentation it has been assumed to be Gaussian. The standard deviation σ for this process is 3.34 while the MRF model parameters are $\alpha = 0.01$, $\beta = 0.007$ and $\gamma = 0.001$. These parameters are considered for different video images are tabulated in Table. 4.2. The spatio-temporal based segmentation using edgeless and edgebased model are shown in Fig. 4.2(d) and 4.2(e) the sharpness in lips of the face has been observed while the lips are smoothed in case of edgeless model. The corresponding JSEG result is shown in Fig. 4.2(c). It can be observed that the part of the face is merged with the hair part and similarly there are more misclassified labels in this. This is also reflected in the percentage of misclassification error provided in Table. 4.1. As seen from the Table. 4.1. the error for JSEG is 4.5 percentage which is quite high as compared to 0.4 and 0.3 for edgeless and edgebased model. Even though the misclassification errors are close in both the cases, the sharpness of the features has been preserved. However in other cases there is appreciable amount of difference in error between edge less and edgebased approaches. The temporal segmentation as obtained using the original video sequence are shown in Fig. 4.2(f) and the corresponding VOPs are shown in Fig. 4.2(g) where it can be observed that the object could be separated from the background. Even for slow movement of the objects in frames, this method could detect the objects.

Similar observation are also made for other two video sequences are shown in Fig. 4.3 and Fig. 4.4 shows the Akiyo news reading video sequence where there are slow movements of the different parts of the body. It is observed that the JSEG groped the whole faces one class while the edgebased model preserved the edges. The misclassification is again low in case of edgebased model. The temporal segmentation and VOPs are shown in Fig. 4.3(f) and (g). For slow movements, the object could be detected.

The third video sequence, as shown in Fig. 4.4(a) has movements that can be viewed as moderate and even in this case also the edgebased model proved to be better as corroborated from the misclassification error. Temporal segmentation and VOPs are shown in Fig. 4.4(f) and Fig. 4.4(g). As observed in the VOPs there are some background pixels present in the edges of the head and also at the top of the head. All the above results used original frames to compute the CDMs and hence the VOPs. The model parameters for the sequence are given in Table. 4.2

In the subsequent part we consider the label frames to obtain the CDMs as opposed to the original frames. It is found that the objects detected using the label frames and edge based model is more precise than using the original frames. The movements again are slow as well as moderately fast and the object could be detected. The Grandma video sequence is shown in Fig. 4.5, the objects detected using original frames are shown in Fig. 4.5(f). It can be observed that near the shoulder some background part has reflected and hence this does not belong to the object part. Fig. 4.5(i) shows the objects detected using the label frames and it is seen that the object could be detected properly without any background part. The model parameters and the misclassification error is given in Table. 4.2. and.Table. 4.3. The $'+'$ in Fig. 4.5(i) indicates the centroid of the object detected. In this case also edge based model proved to be better for slow moving objects. Similarly Fig. 4.8(d) observation can be made for the Akiyo video sequence shown in Fig. 4.6. The edgebased model with label frames for temporal segmentation detected objects more efficiently than edgeless model using original sequences. Thus can be seen from the results shown in case of Fig. 4.6 and Fig. 4.8. For Traffic video sequence as shown in Fig. 4.9, 4.10 and 4.11, with a moderately fast movements, the proposed scheme could detect the vehicles without any missing parts. Fig. 4.8

shows the result for a single object i.e car was seen from Fig. 4.10(i), with the use of label frames it could be detected properly. Fig. 4.11 shows multiple objects in the scenes and the moving object could be detected as seen in Fig. 4.11. The third traffic sequence also corroborate the above findings.

Thus, for slow as well as moderately fast movements the edge based model with label frames proved to be better than edge less model.

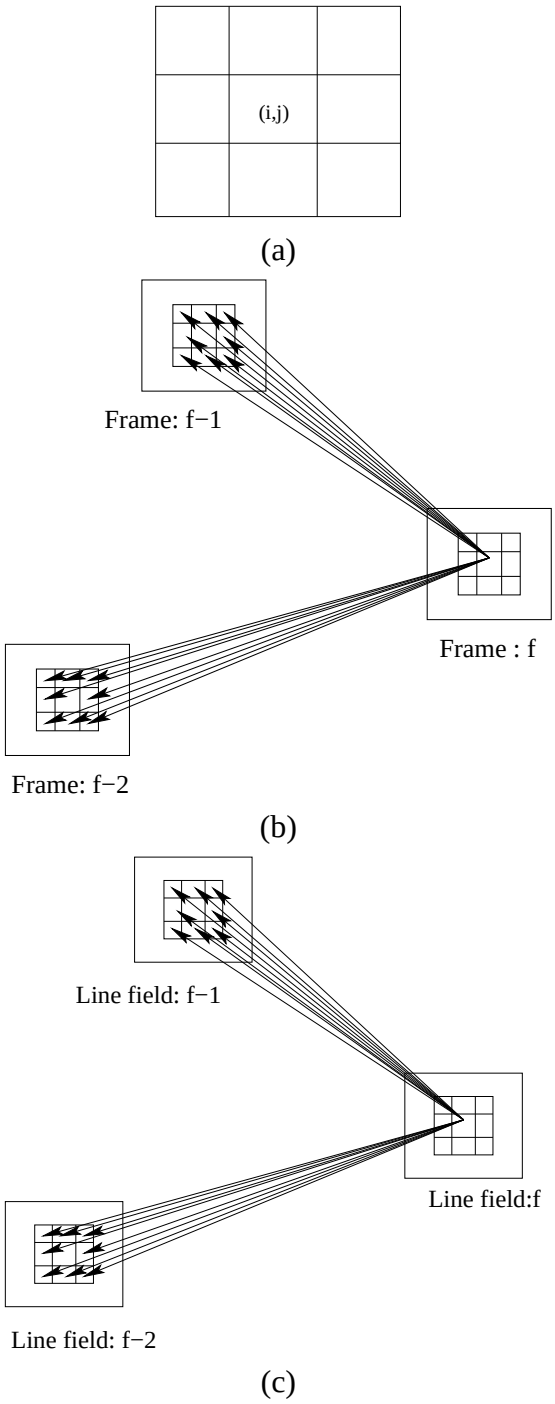


Figure 4.1: (a) MRF modeling in the spatial direction (b) MRF modeling taking two previous frames in the temporal direction (c) MRF with two additional frames with line fields to take care of edge features

(a) Original Suzie Video Sequence Frame No.5,8,11

(b) Ground Truth of Suzie Video Sequence Frame No.5,8,11

(c) JSEG Results for Suzie Video Sequence Frame No.5,8,11

(d) Edgeless Result of Suzie Video Sequence Frame No.5,8,11

(e) Edgebased Result of Suzie Video Sequence Frame No.5,8,11

(f) Temporal Segmentation Result of Suzie Video Sequence Frame No.5,8,11

(g) Extracted VOP of Suzie Video Sequence Frame No.5,8,11

Figure 4.2: Detection of Moving Object in Suzie Video Sequence

(a) Original Akiyo Video Sequence Frame No.75,88,101

(b) Ground Truth of Akiyo Frame No.75,88,101

(c) JSEG Results for Akiyo Frame No.75,88,101

(d) Edgeless Result of Akiyo Frame No.75,88,101

(e) Edgebased Result of Akiyo Frame No.75,88,101

(f) Temporal Segmentation Result of Akiyo Frame No.75,88,101

(g) Extracted VOP of Akiyo Frame No.75,88,101

Figure 4.3: Detection of Moving Object in Akiyo Video Sequence

(a) Original Mother Baby Video Sequence Frame No.65,74,83

(b) Ground Truth of Mother baby Frame No.65,74,83

(c) JSEG Results for Mother Baby Frame No.65,74,83

(d) Edgeless Result of Mother Baby Frame No.65,74,83

(e) Edgebased Result of Mother Baby Frame No.65,74,83

(f) Temporal Segmentation Result of Mother Baby Frame No.65,74,83

(g) Extracted VOP of Mother Baby Frame No.65,74,83

Figure 4.4: Detection of Moving Object in Mother Baby Video Sequence

$Video$	$FrameNo.$	$Edgeless$	$Edgebased$	$JSEG$
$Suzie$	5	0.40	0.30	4.50
	8	0.35	0.10	5.50
	11	0.45	0.1	7.50
$Akiyo$	75	0.80	0.60	2.00
	88	1.20	0.90	2.50
	101	2.70	0.90	2.50
$MotherBaby$	65	1.10	0.20	4.70
	74	2.80	1.00	9.90
	83	1.70	0.10	7.10

Table 4.1: Percentage of Misclassification Error

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

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

(c) Segmentation of Frame No.12,13,14,15 with Edge based Compound MRF Model,

(d) Segmentation result with JSEG Scheme

(e) Temporal Segementation result of Frame No. 12,13,14,15 using CDM of Original Frames

(f) Detected Moving Object of Frame No.12,13,14,15 using results(e)

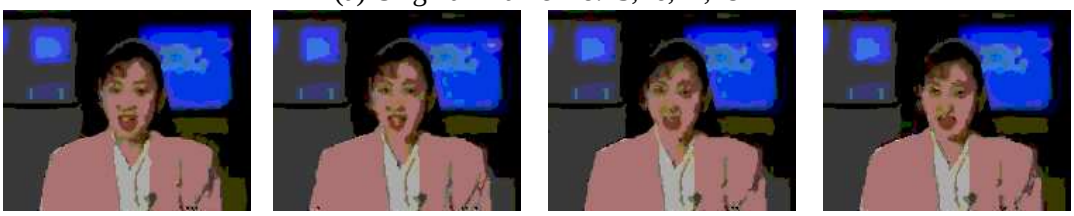
(g) Temporal Segementation result of Frame No. 12,13,14,15 using CDM of Label Frames

(h) Detected Moving Object of Frame No.12,13,14,15 using result (g)

(i) Tracked Object of Frame No.12,13,14,15 using result (h)

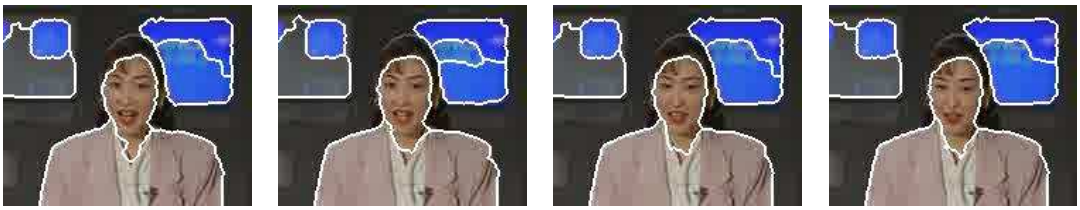
Figure 4.5: VOP Generation of Grandma video sequences

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



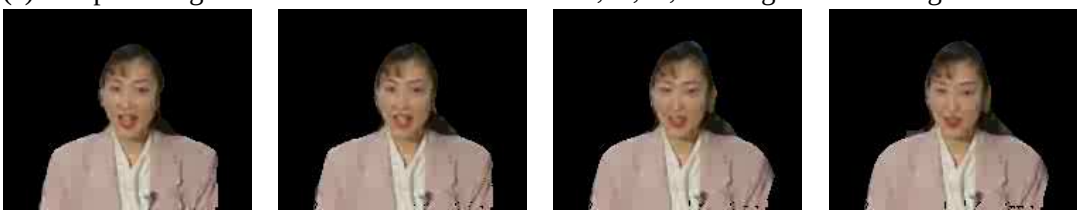
(b) Ground truth of Frame No.75,76,77,78

(c) Segmentation of Frame No.75,76,77,78 with Edge based Compound MRF Model



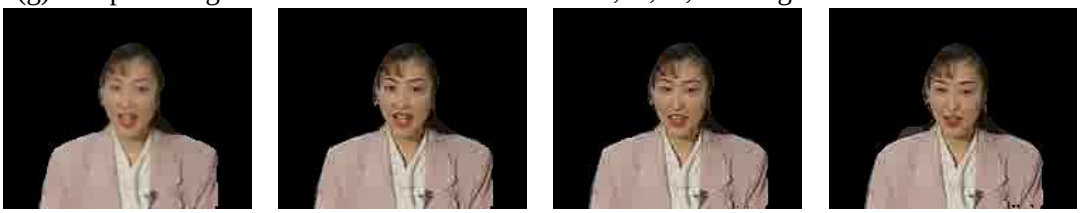
(d) Segmentation result with JSEG Scheme

(e) Temporal Segementation result of Frame No.75,76,77,78 using CDM of Original Frames

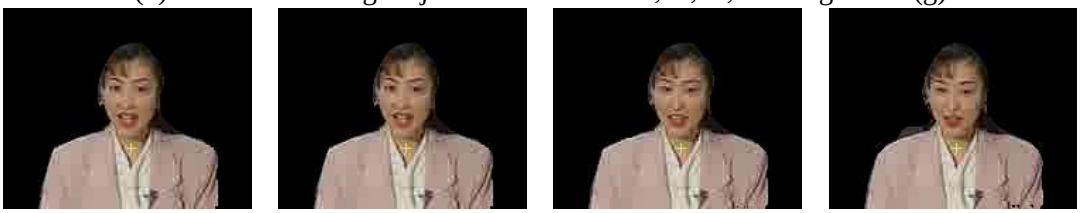


(f) Detected Moving Object of Frame No.75,76,77,78 using results(e)

(g) Temporal Segementation result of Frame No.75,76,77,78 using CDM of Label Frames



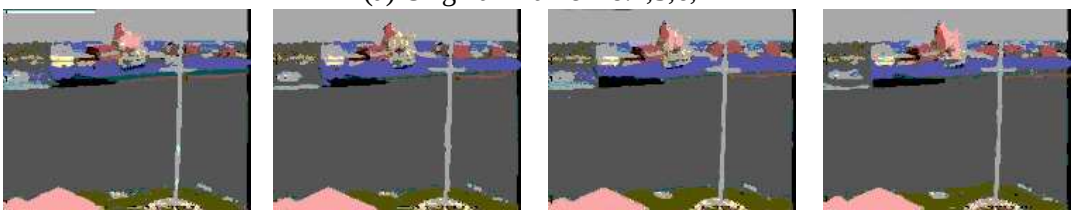
(h) Detected Moving Object of Frame No.75,76,77,78 using result (g)



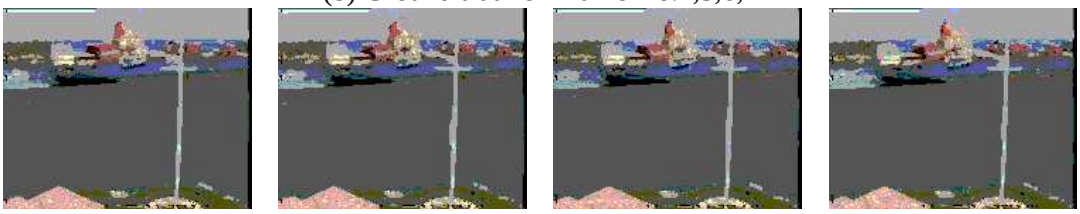
(i) Tracked Object of Frame No.75,76,77,78 using result (h)

Figure 4.6: VOP Generation of Akiyo video sequences

(a) Original Frame No.4,5,6,7



(b) Ground truth of Frame No.4,5,6,7



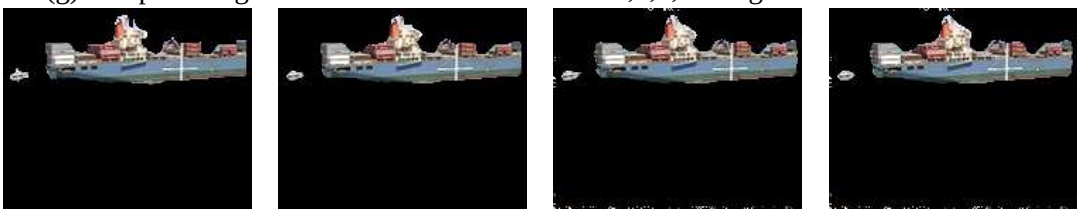
(c) Segmentation of Frame No.4,5,6,7 with Edge based Compound MRF Model

(d) Segmentation result with JSEG Scheme

(e) Temporal Segementation result of Frame No.4,5,6,7 using CDM of Original Frames

(f) Detected Moving Object of Frame No.4,5,6,7 using results(e)

(g) Temporal Segementation result of Frame No.4,5,6,7 using CDM of Label Frames



(h) Detected Moving Object of Frame No.4,5,6,7 using result (g)

Figure 4.7: VOP Generation of Container video sequences

(a) Original Frame No.5,6,7,8

(b) Ground truth of Frame No.5,6,7,8

(c) Segmentation of Frame No.5,6,7,8 with Edge based Compound MRF Model

(d) Segmentation result with JSEG Scheme

(e) Temporal Segementation result of Frame No.5,6,7,8 using CDM of Original Frames

(f) Detected Moving Object of Frame No.5,6,7,8 using results(e)

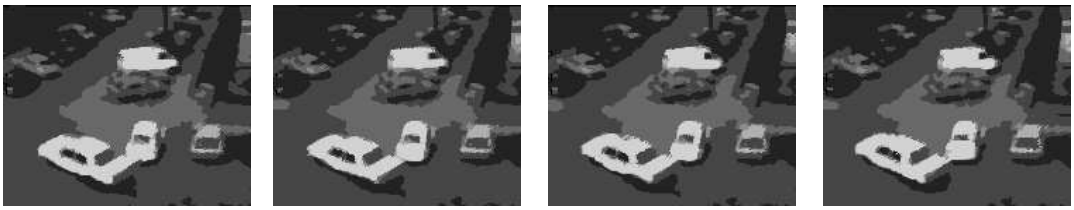
(g) Temporal Segementation result of Frame No.5,6,7,8 using CDM of Label Frames

(h) Detected Moving Object of Frame No.5,6,7,8 using result (g)

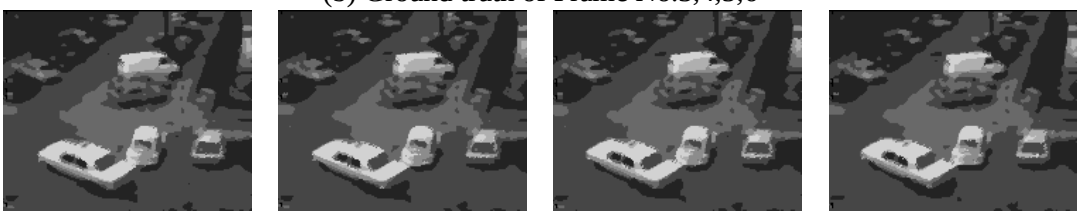
(i) Tracked Object of Frame No.5,6,7,8 using result (h)

Figure 4.8: VOP Generation of Suzie video sequences

(a) Original Frame No.3,4,5,6



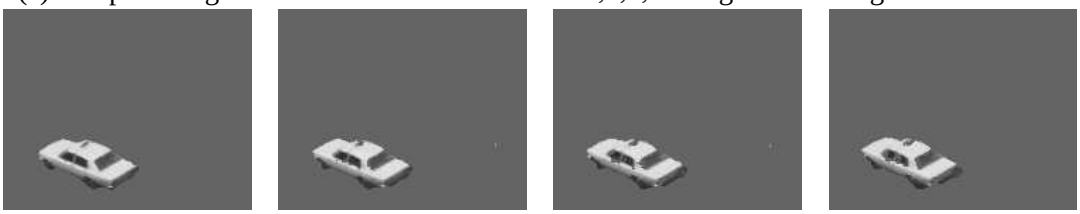
(b) Ground truth of Frame No.3,4,5,6



(c) Segmentation of Frame No.3,4,5,6 with Edge based Compound MRF Model

(d) Segmentation result with JSEG Scheme

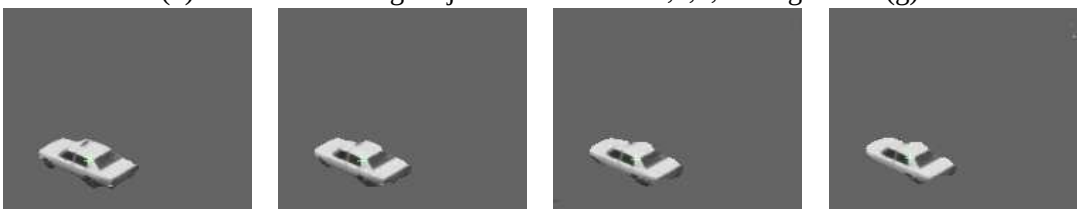
(e) Temporal Segementation result of Frame No.3,4,5,6 using CDM of segmented Frames



(f) Detected Moving Object of Frame No.3,4,5,6 using results(e)

(g) Temporal Segementation result of Frame No.3,4,5,6 using CDM of segmented Frames

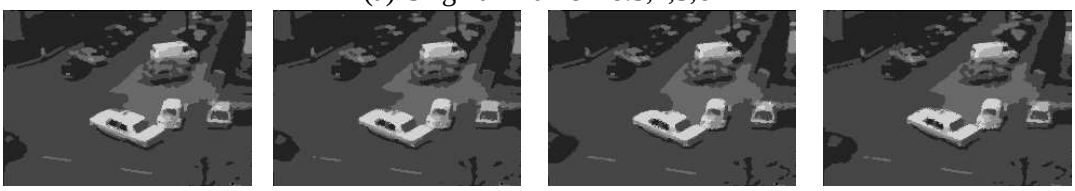
(h) Detected Moving Object of Frame No.3,4,5,6 using results(g)



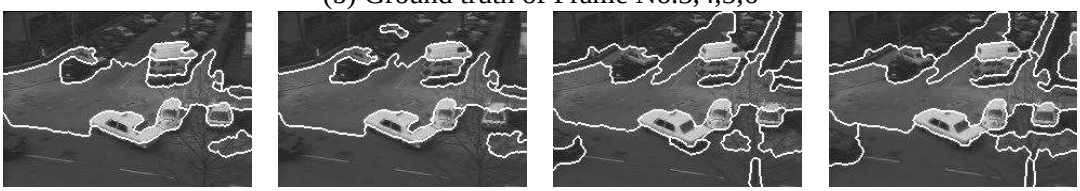
(i) Tracked Moving Object of Frame No.3,4,5,6 using results(h)

Figure 4.9: VOP Generation of Traffic video Car sequences

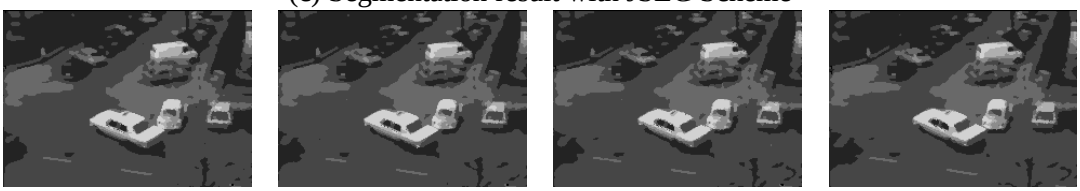
(a) Original Frame No.3,4,5,6



(b) Ground truth of Frame No.3,4,5,6

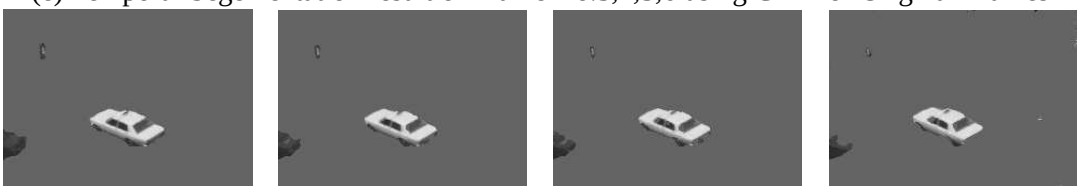


(c) Segmentation result with JSEG Scheme



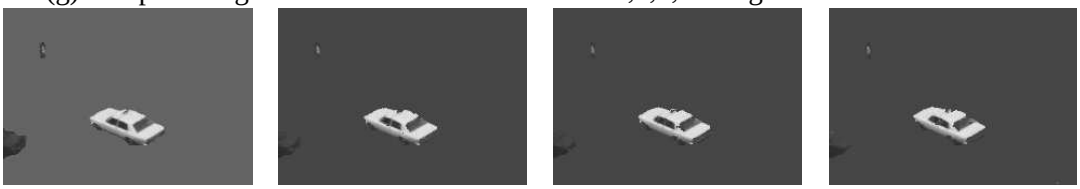
(d) Segmentation of Frame No.3,4,5,6 with Edge based Compound MRF Model

(e) Temporal Segementation result of Frame No.3,4,5,6 using CDM of Original Frames



(f) Detected Moving Object of Frame No.3,4,5,6 using results(e)

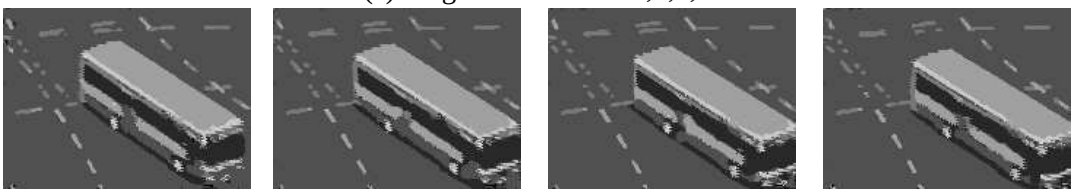
(g) Temporal Segementation result of Frame No.3,4,5,6 using CDM of Label Frames



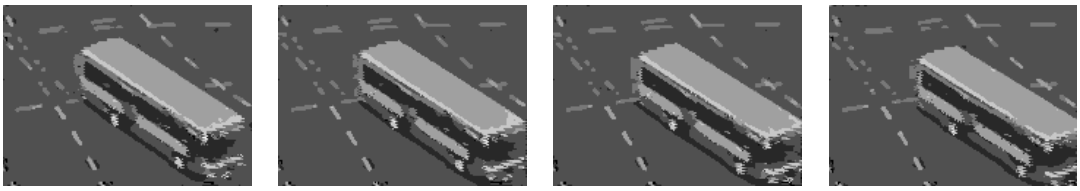
(h) Detected Moving Object of Frame No.3,4,5,6 using result (g)

Figure 4.10: VOP Generation of Canada Traffic video sequences

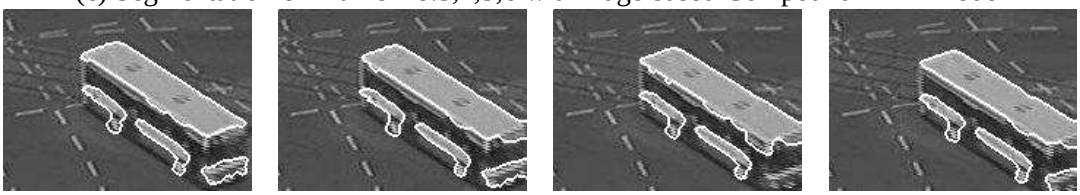
(a) Original Frame No.3,4,5,6



(b) Ground truth of Frame No.3,4,5,6



(c) Segmentation of Frame No.3,4,5,6 with Edge based Compound MRF Model



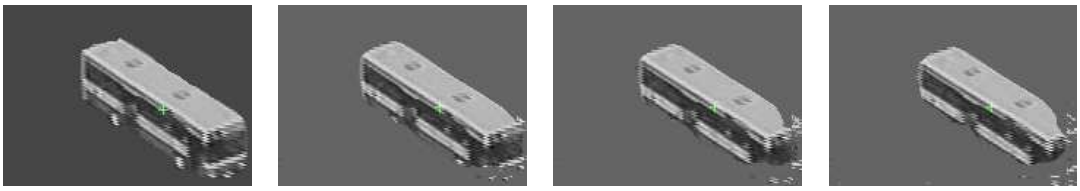
(d) Segmentation result with JSEG Scheme

(e) Temporal Segementation result of Frame No.3,4,5,6 using CDM of segmented Frames

(f) Detected Moving Object of Frame No.3,4,5,6 using results(e)

(g) Temporal Segementation result of Frame No.3,4,5,6 using CDM of segmented Frames

(h) Detected Moving Object of Frame No.3,4,5,6 using results(g)

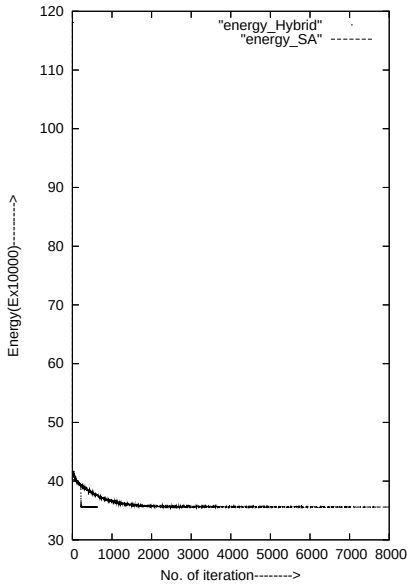


(i) Tracked Moving Object of Frame No.3,4,5,6 using results(g)

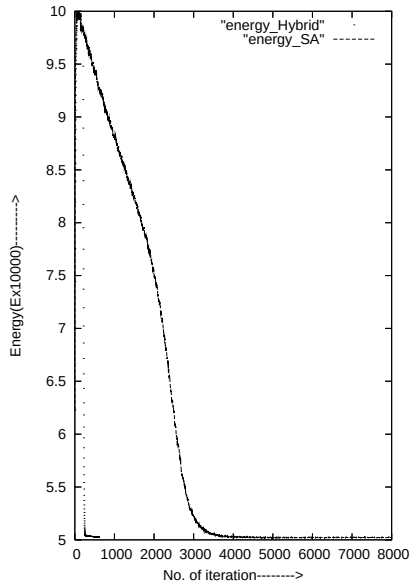
Figure 4.11: VOP Generation of Bus video sequences

VIDEO	α	β	γ	σ
Suzie	0.01	0.007	0.001	3.34
Akiyo	0.009	0.008	0.007	2.0
Mother & Daughter	0.01	0.007	0.005	5.5
Grandma	0.05	0.009	0.007	5.19
Container	0.01	0.009	0.001	2.44
Traffic Car	0.01	0.009	0.007	3.0
Traffic Cannada	0.01	0.009	0.007	3.0
Traffic Bus	0.01	0.009	0.007	3.0

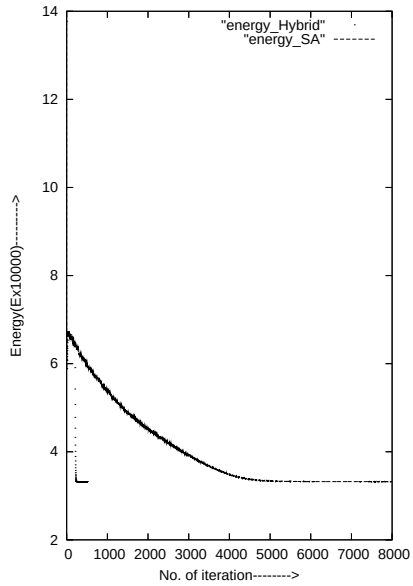
Table 4.2: Compond MRF Model Parameters for diffrent videos



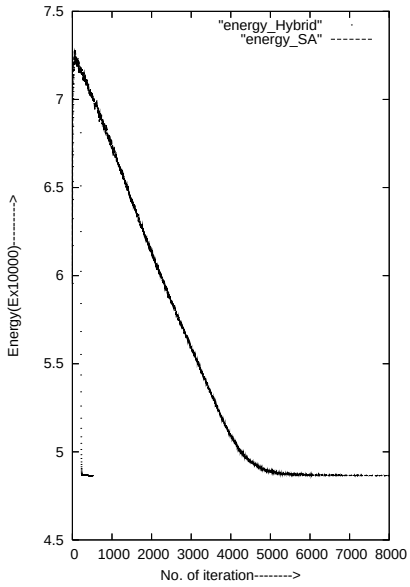
(a) For Akiyo Frame No.75



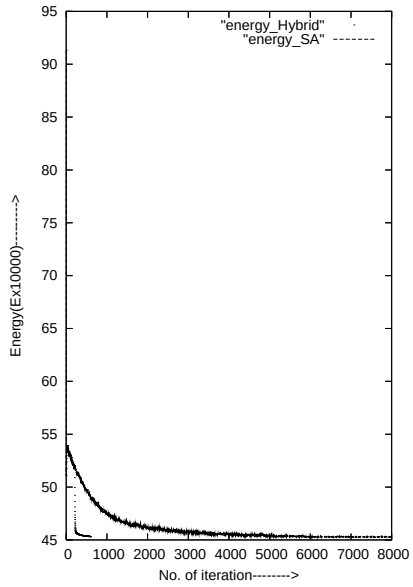
(b) For Grandma Frame No.12



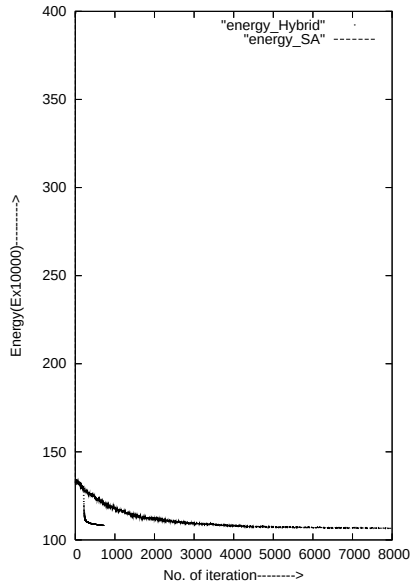
(c) For Container Frame No. 4



(d) For Suzie Frame No. 5



(e) For Traffic car Frame No.3



(f) For Cannada Traffic Frame No. 3

Figure 4.12: Energy plot of different Video Sequences

$Video$	$FrameNo.$	$JSEG$	$Edgebased$
$Grandma$	12	6.82	0.24
	13	3.77	0.30
	14	4.29	0.25
	15	4.20	0.29
$Akiyo$	75	6.20	0.12
	76	4.90	0.12
	77	5.50	0.10
	78	5.40	0.10
$Container$	4	2.55	0.10
	5	6.44	0.14
	6	4.61	0.17
	7	4.46	0.15
$Suzie$	4	2.55	0.10
	5	6.44	0.14
	6	4.61	0.17
	7	4.46	0.15
$TrafficCar$	3	9.56	0.75
	4	10.44	0.41
	5	7.56	0.65
	6	22.05	0.61
$TrafficBus$	3	6.10	0.18
	4	5.27	0.40
	5	4.97	0.44
	6	5.10	0.39
$CanadaTraffic$	3	5.95	0.1
	4	8.23	0.41
	5	16.65	0.52
	6	7.1	0.46

Table 4.3: Percentage of Misclassification Error

Chapter 5

DETECTION OF SLOW MOVING VIDEO OBJECTS USING COMPOUND MARKOV RANDOM FIELD MODEL

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 Chapter, we propose a Markov random Field (MRF) model based scheme to detect slow movements in a video sequence. In order to enhance the efficacy of the earlier schemes, a new MRF model for video sequences is proposed. In the frame sequences, there are changes from frame to frame because of the object in the video. We assume these changes not to be abrupt ones and hence are expected to have a temporal neighborhood dependency. These changes in the consecutive frames are modeled as MRF [26]. Therefore the proposed a priori MRF model of the video sequence takes in to account these changes of the frames together with the edges in temporal direction.

In this piece of work, 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 [25]. 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, edgemaps in temporal direction. The edge maps helps in preserving the edges of the moving objects. in order to detect slow moments 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 next use it as the initials one for the next frame. ICM (Iterative Conditional Mode) algorithm [17] is used to obtain the 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 [14] method and it is found that the proposed method outperformed JSEG in terms of misclassification error.

5.1 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. The a priori distribution takes care of the spatial model of X_t, the temporal modeling taking care ofX_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} and respectively. The MRF model taking care of changes in temporal direction of the 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

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

where $\eta_{s,t}$ 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.

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

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

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

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

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

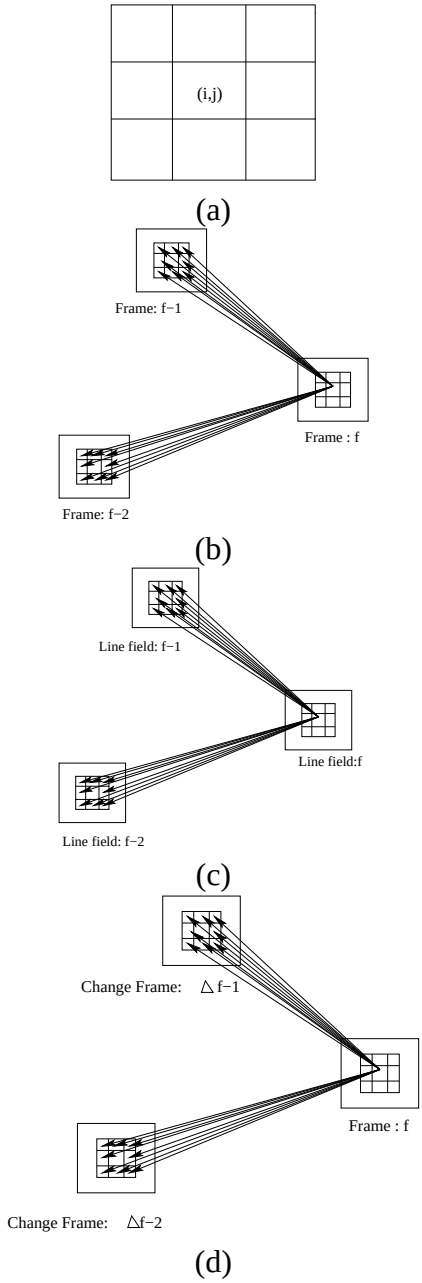


Figure 5.1: (a) MRF modeling in the spatial direction (b) MRF modeling taking two previous frames in the temporal direction (c) MRF with two additional frames with line fields to take care of edge features (d) MRF with two change frame to incorporate changes

5.1.1 Spatio-temporal Segmentation in MAP frame work

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 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 ΔX_{t-1} and ΔX_{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 can be estimated from the observed random field Y_t . The label field is estimated by maximizing the following posterior probability distributions.

$$\hat{x} = \arg \max_{x} P(X = x|Y = y,) \tag{5.1}$$

Where $\hat{x}$ denotes the estimated labels. Since, x is unknown it is very difficult to evaluate (5.1), hence, using Baye's theorem (5.1) can be written as

$$\hat{x} = \arg\ \max_{x}\ \frac{P(Y = y|X = x)P(X = x)}{P(Y = y)} \tag{5.2}$$

Since y is known, the prior probability $P(Y = y)$ is constant. hence (5.2) reduces to

$$\hat{x} = \arg\ \max_{x}\ P(Y = y|X = x, \theta)P(X = x, \theta) \tag{5.3}$$

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

$$P(X = x) = e^{-U(x,\theta)} = e^{\left[-\sum_{c\epsilon C}[V_{sc}(x)+V_{tec}(x)+V_{teec}(x)+V_{tch}(x)]\right]} \tag{5.4}$$

In (5.4) $V_{sc}(x)$ 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 and $V_{tch}(x)$denotes clique potential incorporating change feature. We have proposed this additional feature in the temporal direction.(5.4) is called the edgebased 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

$$P(N = y - x|X, \theta) = \frac{1}{\sqrt{(2\pi)^n det\,[k]}} e^{-\frac{1}{2}(y-x)^T K^{-1}(y-x)} \tag{5.5}$$

Where k is the covariance matrix. Assuming decorrelation of the three RGB planes and the variance to be same among each plane, (5.5) can be expressed as

$$P(N = y - x|X, \theta) = \frac{1}{\sqrt{(2\pi)^3 \sigma^3}} e^{-\frac{1}{2\sigma^2}(y-x)^2} \tag{5.6}$$

In (5.6) Variance σ^2 corresponds to the Gaussian degradation. Hence (5.3) can be expressed as

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

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.

$$\hat{x} = \arg\ \max_x\ \left[e^{-\frac{\left[-\|y-x\|^2\right]}{2\sigma^2} + \sum_{c\epsilon C} V_{sc}(x)+V_{tec}(x)+V_{teec}(x)+V_{tch}(x)} \right] \tag{5.7}$$

Maximizing (5.7) is tantamount to minimizing the

$$\hat{x} = \arg\ \min_x\ \left\{ \left[\frac{\| y - x \|^2}{2\sigma^2} \right] + \left[\sum_{c\epsilon C} V_{sc}(x) + V_{tec}(x) + V_{teec}(x) + V_{tch}(x) \right] \right\} \tag{5.8}$$

$\hat{x}$ in (5.8) 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.

5.2 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.

5.3 RESULTS AND DISCUSSION

Five different video sequences have been considered to validate the change based MRF model. The a priori MRF distributions of the change based model have additional model parameters besides edge based model. In this case, the model parameters μ_1 and μ_2 have also been selected on a trial and error basis. Fig. 5.2 shows the Grandma video sequences. Fig. 5.2(d) shows the spatial segmentation of edge based model and Fig. 5.2(i) shows results due to change model. The corresponding tracked objects are shown in Fig. 5.2(h) and Fig. 5.2(l). The model parameters selected are given in Table. 5.2, and the misclassification error is in Table. 5.1. From Fig. 5.2(i) there are some misclassified pixels in the shoulder of the Grandma where as seen from Fig. 5.2(l) the $'+'$ symbol indicates the centroid of the object. The change based MRF model with the label frame could better detect the object than the edge based model. As observed in the previous section the JSEG method yields segmentation result having more misclassification error than edgebased model.

Similar observations are also made with the Akiyo video sequence shown in Fig. 5.3. In this case, comparing 5.3(f) and 5.3(l), it is observed that there are some background information because of use of original frames where as with the use of label frames the result has improved and the results are further improved by use of the change based model.

Fig. 5.4 shows another video sequence with slow moving objects. As observed the spatial segmentation accuracy in case of change model is comparable with that of the edgebased model. The detected objects in case of change based model are comparable with that of edgebased model with some of the background noise being eliminated. Analogous observation are also made for the Container video sequence shown in Fig. 5.5. A flag which could not be detected properly could be detected in case of change based model. Fig. 5.6 shows the results obtained for the traffic sequence which has multiple objects in the scene. As observed from Fig. 5.6(k) and the change based model could detect the moving object while other objects have been static and hence considered background. Edgebased also produced similar result with some dots as the background noise. Hence even in multiple scene the proposed method could track the objects. As seen from Table. 5.1 the misclassification error for traffic sequence is lowest as compared to JSEG and edgebased model.

Thus, the change based MRF model exhibited improved accuracy as compared to the edgebased model. The moving objects in this sequence could be detected for slow as well as moderately fast moving sequences

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

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

(c) Segmentation result with JSEG Scheme

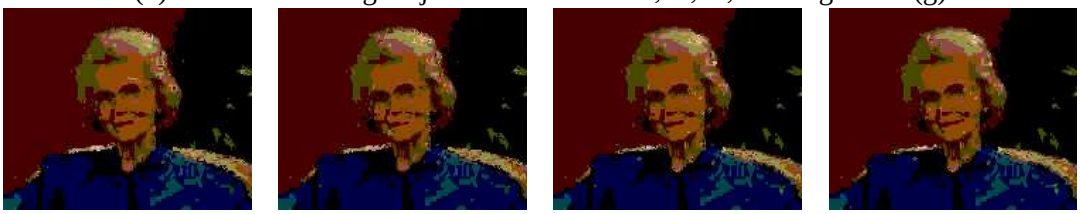
(d) Segmentation of Frame No.12,13,14,15 with Edge based Compound MRF Model,

(e) Temporal Segementation result of Frame No. 12,13,14,15 using CDM of Original Frames

(f) Detected Moving Object of Frame No.12,13,14,15 using results(e)

(g) Temporal Segementation result of Frame No. 12,13,14,15 using CDM of Label Frames

(h) Detected Moving Object of Frame No.12,13,14,15 using result (g)



(i) Segmentation of Frame No.12,13,14,15 with Change Model

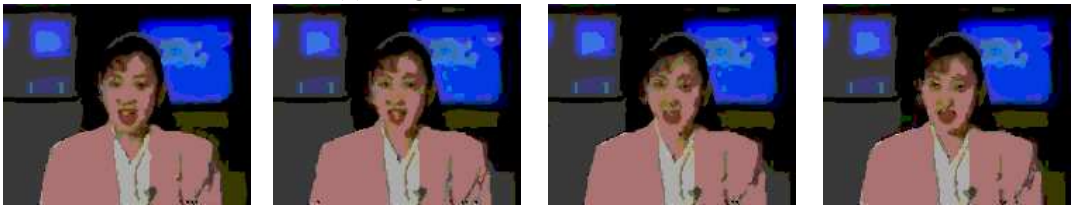
(j) Temporal Segementation result of Frame No. 12,13,14,15 using CDM of Label Frames

(k) Detected Moving Object of Frame No.12,13,14,15 using result (j)

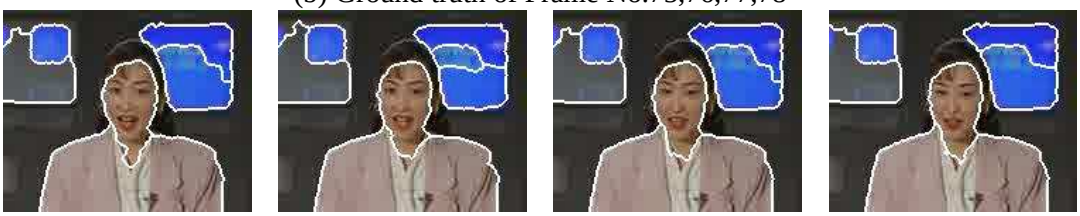
(l) Tracked Object of Frame No.12,13,14,15 using result (k)

Figure 5.2: VOP Generation of Grandma video sequences

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



(b) Ground truth of Frame No.75,76,77,78



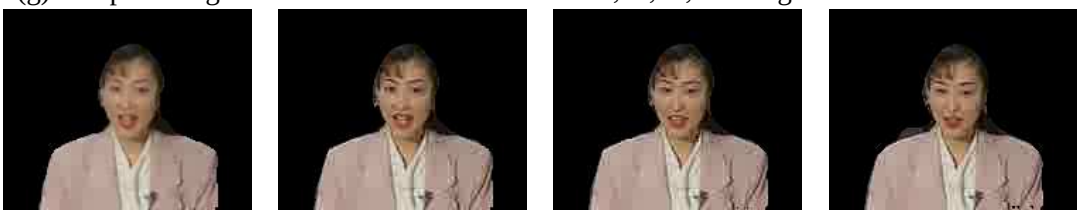
(c) Segmentation result with JSEG Scheme

(d) Segmentation of Frame No.75,76,77,78 with Edge based Compound MRF Model

(e) Temporal Segementation result of Frame No.75,76,77,78 using CDM of Original Frames

(f) Detected Moving Object of Frame No.75,76,77,78 using results(e)

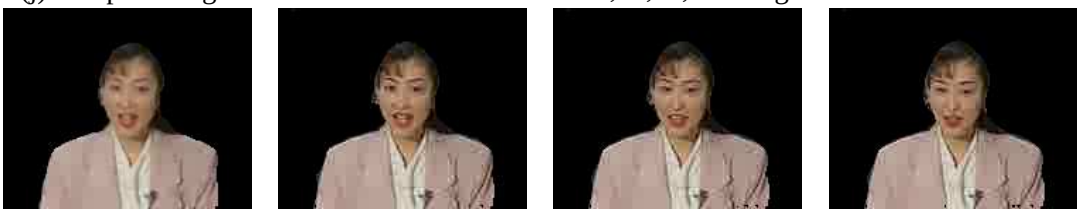
(g) Temporal Segementation result of Frame No.75,76,77,78 using CDM of Label Frames



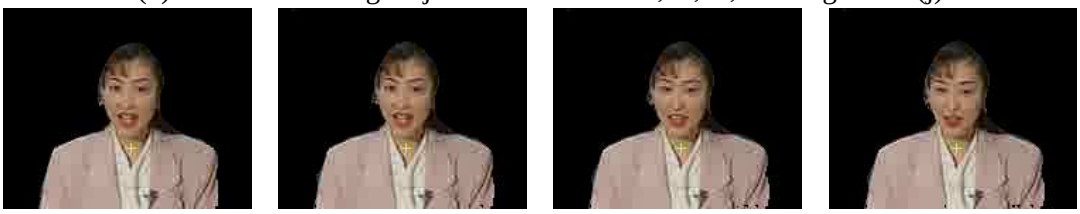
(h) Detected Moving Object of Frame No.75,76,77,78 using result (g)

(i) Segmentation of Frame No.75,76,77,78 with Change Model

(j) Temporal Segementation result of Frame No.75,76,77,78 using CDM of Label Frames



(k) Detected Moving Object of Frame No.75,76,77,78 using result (j)



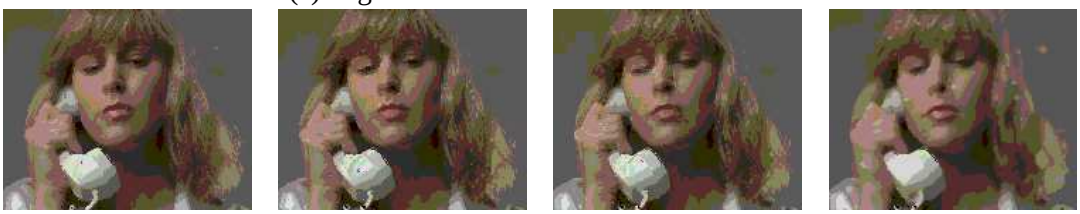
(l) Tracked Object of Frame No.75,76,77,78 using result (k)

Figure 5.3: VOP Generation of Akiyo video sequences

(a) Original Frame No.5,6,7,8

(b) Ground truth of Frame No.5,6,7,8

(c) Segmentation result with JSEG Scheme



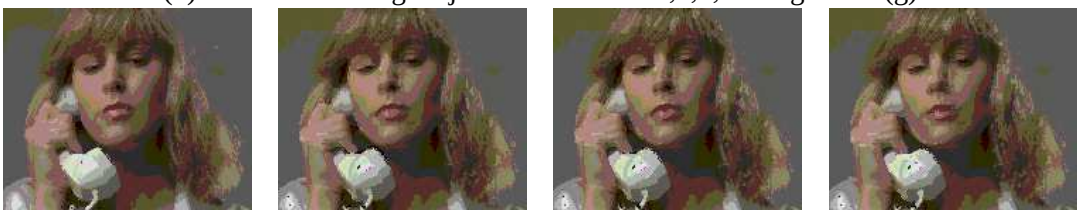
(d) Segmentation of Frame No.5,6,7,8 with Edge based Compound MRF Model

(e) Temporal Segementation result of Frame No.5,6,7,8 using CDM of Original Frames

(f) Detected Moving Object of Frame No.5,6,7,8 using results(e)

(g) Temporal Segementation result of Frame No.5,6,7,8 using CDM of Label Frames

(h) Detected Moving Object of Frame No.5,6,7,8 using result (g)



(i) Segmentation of Frame No.5,6,7,8 with Change Model

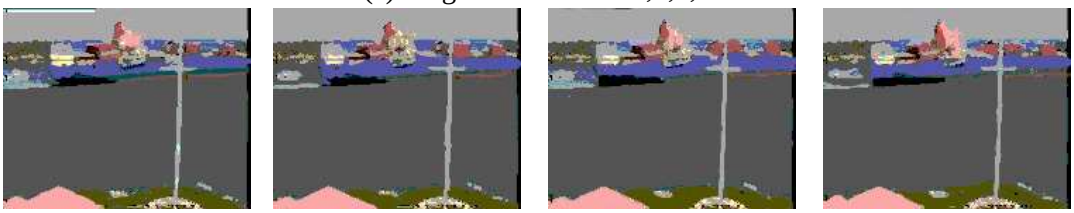
(j) Temporal Segementation result of Frame No.5,6,7,8 using CDM of Label Frames

(k) Detected Moving Object of Frame No.5,6,7,8 using result (j)

(l) Detected Moving Object of Frame No.5,6,7,8 using result (k)

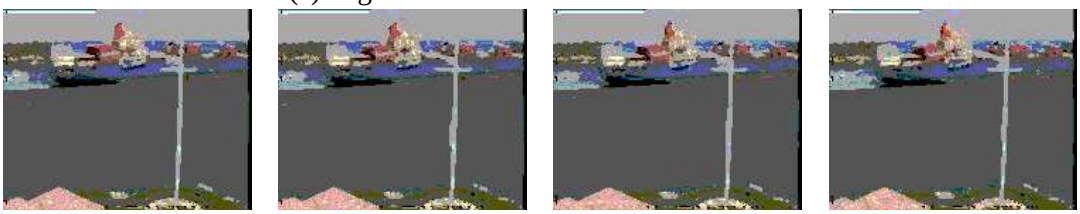
Figure 5.4: VOP Generation of Suzie video sequences

(a) Original Frame No.3,4,5,6



(b) Ground truth of Frame No.3,4,5,6

(c) Segmentation result with JSEG Scheme

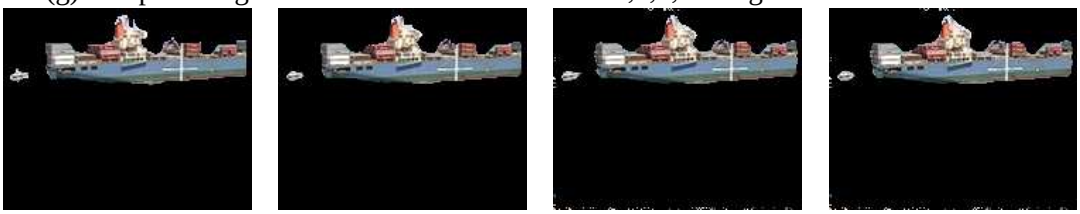


(d) Segmentation of Frame No.3,4,5,6 with Edge based Compound MRF Model

(e) Temporal Segementation result of Frame No.3,4,5,6 using CDM of Original Frames

(f) Detected Moving Object of Frame No.3,4,5,6 using results(e)

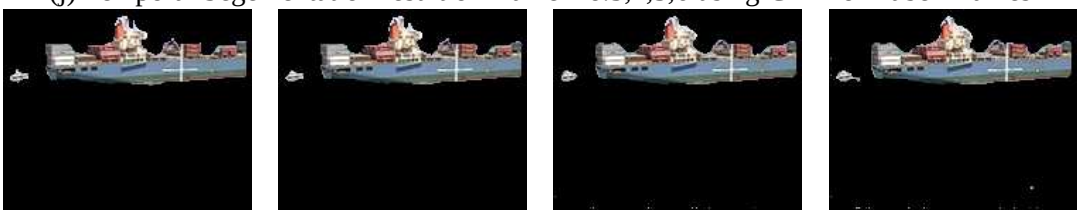
(g) Temporal Segementation result of Frame No.3,4,5,6 using CDM of Label Frames



(h) Detected Moving Object of Frame No.3,4,5,6 using result (g)

(i) Segmentation of Frame No.3,4,5,6 with Change Model

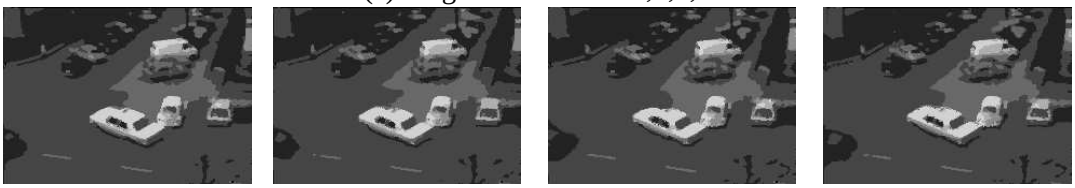
(j) Temporal Segementation result of Frame No.3,4,5,6 using CDM of Label Frames



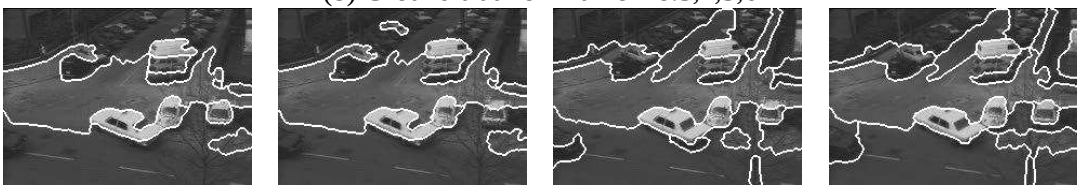
(k) Detected Moving Object of Frame No.3,4,5,6 using result (j)

Figure 5.5: VOP Generation of Container video sequences

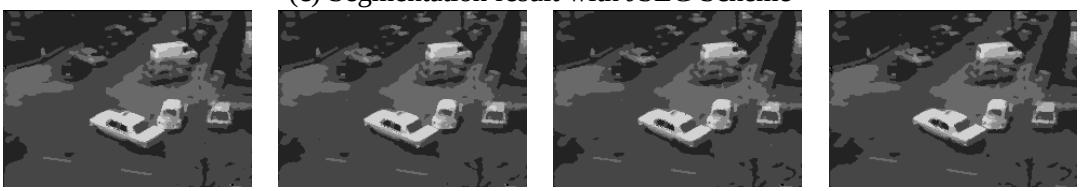
(a) Original Frame No.3,4,5,6



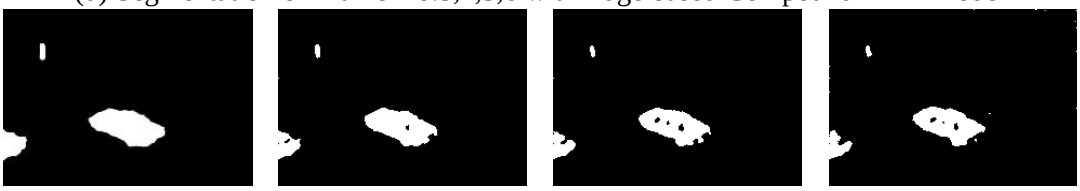
(b) Ground truth of Frame No.3,4,5,6



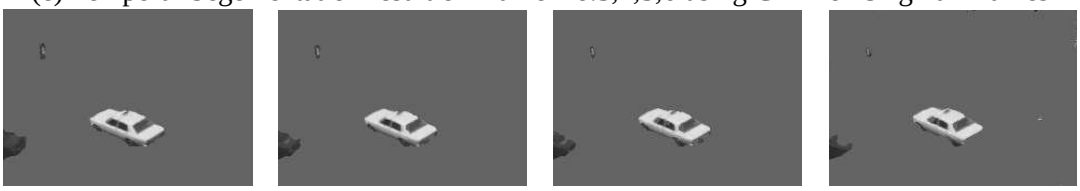
(c) Segmentation result with JSEG Scheme



(d) Segmentation of Frame No.3,4,5,6 with Edge based Compound MRF Model

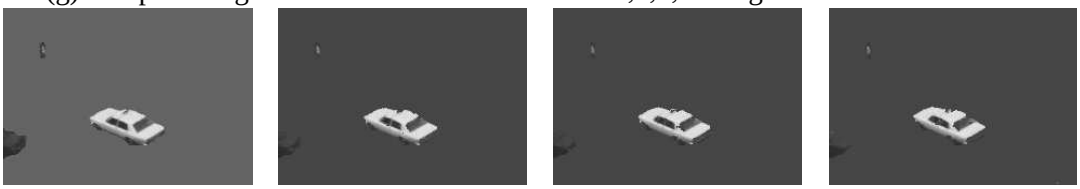


(e) Temporal Segementation result of Frame No.3,4,5,6 using CDM of Original Frames

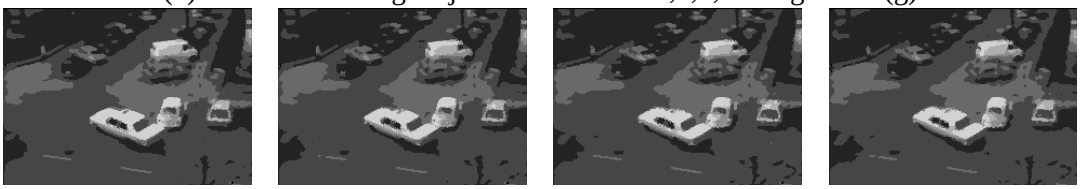


(f) Detected Moving Object of Frame No.3,4,5,6 using results(e)

(g) Temporal Segementation result of Frame No.3,4,5,6 using CDM of Label Frames



(h) Detected Moving Object of Frame No.3,4,5,6 using result (g)



(i) Segmentation of Frame No.3,4,5,6 with Change Model

(j) Temporal Segementation result of Frame No.3,4,5,6 using CDM of Label Frames

(k) Detected Moving Object of Frame No.3,4,5,6 using result (j)

Figure 5.6: VOP Generation of Cannada Traffic video sequences

$Video$	$FrameNo.$	$JSEG$	$Edgebased$	$ChangeModel$
$Grandma$	12	6.82	0.24	0.40
	13	3.77	0.30	0.40
	14	4.29	0.25	0.39
	15	4.20	0.29	0.12
$Akiyo$	75	6.20	0.12	0.12
	76	4.90	0.12	0.18
	77	5.50	0.10	0.19
	78	5.40	0.10	0.22
$Container$	4	2.55	0.10	0.15
	5	6.44	0.14	0.15
	6	4.61	0.17	0.23
	7	4.46	0.15	0.24
$Suzie$	4	2.55	0.10	0.24
	5	6.44	0.14	0.24
	6	4.61	0.17	0.24
	7	4.46	0.15	0.24
$CanadaTraffic$	3	5.95	0.1	0.19
	4	8.23	0.41	0.16
	5	16.65	0.52	0.13
	6	7.1	0.46	0.21

Table 5.1: Percentage of Misclassification Error

VIDEO	α	β	γ	σ	$\mu 1$	$\mu 2$
Grandma	0.05	0.009	0.007	5.19	0.1	0.01
Akiyo	0.009	0.008	0.007	2.0	0.1	0.01
Suzie	0.01	0.007	0.001	3.34	0.1	0.01
Container	0.01	0.009	0.001	2.44	0.1	0.01
Traffic Video	0.01	0.009	0.007	3.0	0.01	0.01

Table 5.2: Compond MRF Model Parameters for diffrent videos

Chapter 6

AN EVOLUTIONARY BASED SLOW AND FAST MOVING VIDEO OBJECTS DETECTION SCHEME USING COMPOUND MRF MODEL

It has been observed in the previous proposed scheme that spatial segmentation of each frame has to be obtained to find out temporal segmentation. Spatial segmentation of every frames is a time consuming procedure and hence the object detection scheme takes appreciable amount of time. This forbids the feasibility of real time implementation. In order to reduce the computational burden, we compute the spatial segmentation of a given frame using the proposed spatio-temporal approach. The spatial segmentation of subsequent frames are obtained starting from the segmentation of given frame with adaptation strategy. Detection of video object at any frame is obtained using the frame together with the temporal segmentation. Spatial segmentation only one frame is obtained using spatio-temporal formulation of previous section [25].

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 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 [35]. 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 [14] method and it is observed that the proposed method is found to be better than former one in the context of misclassification error.

6.1 PROPOSED APPROACH OF OBJECT DETECTION

In this approach, we obtain the spatial segmentation of a frame known as the initial frame, The spatial segmentation is formulated in spatio-temporal framework using edge based MRF model as in Section. 4. Hybrid algorithm is used to obtain the MAP estimates of the pixel labels. Thereafter, segmentation of successive frames are obtained by evolving the labels of the initial frames with the proposed evolution strategy. Thus, estimation of labels of other frames are not necessary.
In order to construct the VOP, temporal segmentation is obtained with the labels of different frames as opposed to the original frames. The history of the labels are

being used to obtain the temporal segmentation (Global thresholding is used for to obtain the CDM). The VOP are constructed using the temporal segmentation and the original frame sequence. Thus scheme avoids the estimation of labels of each frame. This reduces the computational burden and makes it feasible for real time implementation.

6.2 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.

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,

$$x_{(t+d)i} = x_t - \mid y_{t+d} - y_t \mid + y_{t+d(y_{t+d}-y_t)} \tag{6.1}$$

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.

6.3 ITERATED CONDITIONAL MODE ALGORITHM

Since it is difficult to maximize the joint probability of an MRF, Besag proposed a deterministic algorithm called Iterated Conditional Modes (ICM) which maximizes local conditional probabilities sequentially. The ICM algorithm uses the greedy strategy in the iterative local maximization. Given the data x and the other labels $z_{S-i}^{(k)}$, the algorithm sequentially updates each $z_i^{(k)}$ into $z_i^{(k+1)}$ by maximizing $P(z_i \mid x, z_{S-i})$, the conditional probability, with respect to z_i. Two assumptions are made in calculating $P(z_i \mid x, z_{S-I})$:

1. The observation components $x_1,\ x_2,\ x_3 ... x_m$ are conditionally independent given z and each x_i has the same known conditional density function $p(x_i \mid z_i$ dependent only on z_i. Thus

$$p(x \mid z) = \prod_i p(x_i \mid z_i) \tag{6.2}$$

2. The second assumption is that z depends on the labels in the local neighborhood, which is the Markovianity.

From the two assumptions and the Bayes theorem, it follows that

$$P(z_i \mid x, z_{S-i}) \propto p(x_i \mid z_i) P(z_i \mid z_{N_i}) \tag{6.3}$$

Obviously, $P(z_i \mid x_i,\ z_{N_i}^k)$ is much easier to maximize than $P(z \mid x)$, which is the point of ICM. Maximizing (4.18) is equivalent to minimizing the corresponding posterior energy using the following rule.

$$z_i^{k+1} \longleftarrow \underset{z_i}{arg\,max}\ U(z_i \mid x_i,\ f_{N_i}^{(k)}) \tag{6.4}$$

The result obtained by ICM depends very much on the initial estimator $z^{(0)}$ and the ICM is locally convergent[35].

6.4 SIMULATION AND RESULT DISCUSSION

We have considered four types of video sequences as shown in Fig. 6.1, Fig. 6.2, Fig. 6.3 and Fig. 6.4. Fig. 6.1 and Fig. 6.4 corresponds to slow movements of the sequence where as Fig. 6.2 and Fig. 6.3 corresponds to video sequences with fast moving objects. Fig. 6.1(a) shows Grandma image of 12th, 37th, 62nd and 87th frame. It is observed from these frames that there are slow changes. The corresponding ground truth image constructed manually are shown in Fig. 6.1(b). Fig. 6.1(c) 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.6.1(c) evolves to produce the initial segmentation results corresponding to 18, 24, 30 and 37th frame as shown in Fig. 6.1(d). Using 37 th frame crude result in Fig. 6.1(d) as the crude segmentation, ICM is run to obtain the segmentation of 37th frame as shown in Fig. 6.1(e). Analogously for the 62nd frame segmentation result of 37th frame evolves to obtain crude segmentation of 62nd frame as shown in Fig. 6.1(f). ICM is run starting in 62nd frame crude result of Fig. 6.1(f) and the segmented results obtained for 62nd frame is shown in Fig. 6.1(g). Similarly result is obtained for 87th frame from evolving crude result of 87th frame. The temporal segmentation result obtained using the segmented result instead of original frames are shown in Fig. 6.1(k) and the corresponding VOPs are shown in Fig. 6.1(l). 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 6.1(o). 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. 6.1(j). The $\%$age of misclassification error is given in

Table. 6.2 and it can be observed that the proposed method has less misclassification error as compared to JSEG method.

The time required for execution of the edgebased scheme is 104sec while the time required for the evolving scheme is 9sec. Thus there is a time saving of order 10. The scheme implemented in a $Pentium4(D)$, $3GHz$, $L2\ cache\ 4MB$, $1GBRAM$, $667FSB$ PC. The execution time required for other sequences also are much less as compared to the edgebased model.

The second video considered is the Akiyo video sequence as shown in Fig. 6.2. Fig. 6.2(c) show the spatial segmentation of 75th frame using spatio-temporal formulation and hybrid algorithm. The MRF model parameters are tabulated in Table. 6.1. The evolutionary strategy is applied to 75th frame to obtain segmentation of 79, 83, 87 and 95th frame as shown in Fig. 6.2(d). ICM is run on 95th frame crude result and the final segmentation is obtained is shown in Fig. 6.2(e) . Other segmented result obtained using the evolving procedure is shown in Fig. 6.2(g) and (i) .Segmentation of JSEG is shown in Fig. 6.2(j). The video objects could be detected properly. The third example considered is the container video sequence as shown in Fig. 6.3. Fig. 6.3(n) shows the detection of video objects with the evolving scheme and it is observed that the object could be tracked without any background effect.The time taken by the proposed scheme is also 10 times less that of obtaining spatial segmentation of each frame.The 4th example considered is the claire video sequence shown in Fig. 6.4. Similar observations are also made in this case. Fig. 6.4(l) shows the tracked object using the label maps in CDMs. Where as using original frames the results are shown in Fig. 6.4(o). The object detected using the label frame based CDMs are sometimes better than that of using original frames. Thus the evolutionary approach based scheme has much less computational burden and hence is viable from real time complementation

VIDEO	α	β	γ	σ
Grandma	0.05	0.009	0.007	5.19
Akiyo	0.009	0.008	0.007	2.0
Container	0.01	0.009	0.001	2.44
Claire	0.009	0.008	0.007	1.00

Table 6.1: Parameters for diffrent videos of the given videos

strategy.

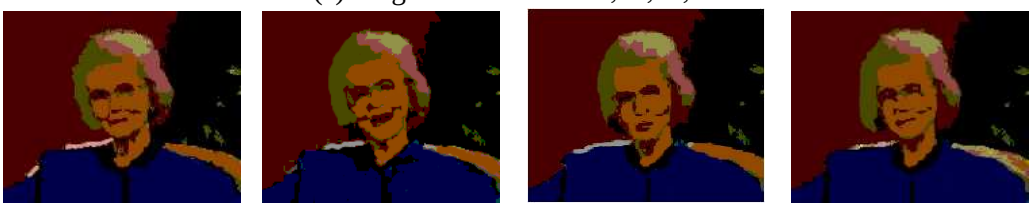
$Video$	$FrameNo.$	$Evolving$	$JSEG$
$Grandma$	12	0.24	6.82
	37	0.15	4.65
	62	0.15	4.5
	87	0.12	3.88
$Akiyo$	75	0.12	6.20
	95	0.10	1.62
	115	0.15	1.65
	135	0.15	1.80
$Container$	4	0.10	2.55
	12	0.11	2.55
	20	0.13	1.51
	24	0.13	2.08
$Claire$	3	0.41	2.95
	7	0.39	2.47
	11	0.76	2.91
	15	0.76	2.91

Table 6.2: Percentage of Misclassification Error

$Video$	$FrameNo.$	$EdgeBased$	$Evolving$
$Grandma$	37	104	9
	62	104	9
	87	104	9
$Akiyo$	95	82	8
	115	82	8
	135	82	8
$Container$	12	112	12
	20	112	12
	28	112	12
$Claire$	7	94	8
	11	94	8
	15	94	8

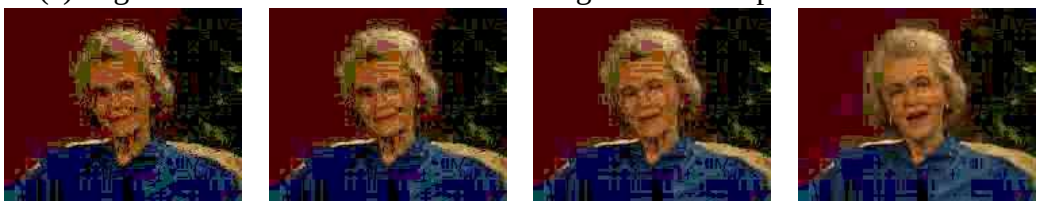
Table 6.3: Time required for execution of the programme in Second

(a) Original Frame No.12,37,62,87



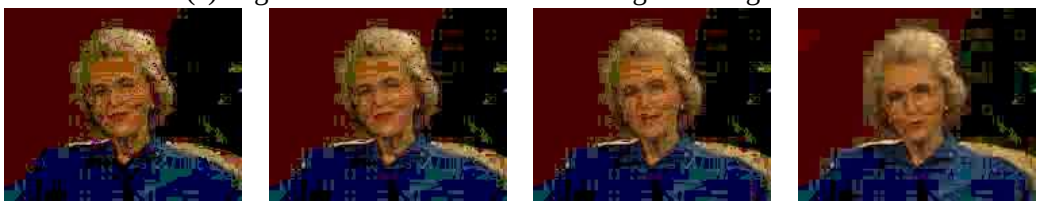
(b) Ground truth of Frame No.12,37,62,87

(c) Segmentation of Frame No.12 with Edge based Compound MRF Model



(d) Evolving Crude result of Frame No. 18,24,30,37

(e) Segmentation of Frame No.37 using Evolving scheme



(f) Evolving Crude Result of Frame No. 41,47,53,62

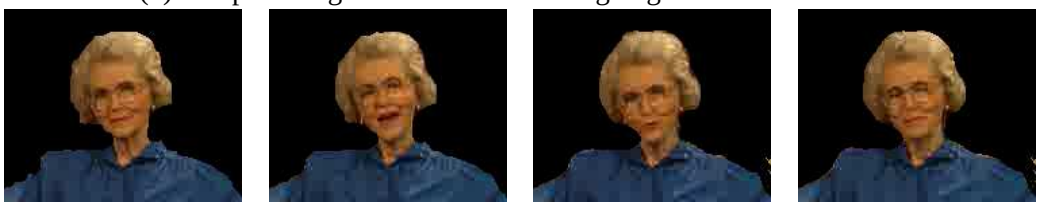
(g) Segmentation of Frame No.62 using Evolving Scheme

(h) Evolving Crude Result of Frame No. 68,74,80,87

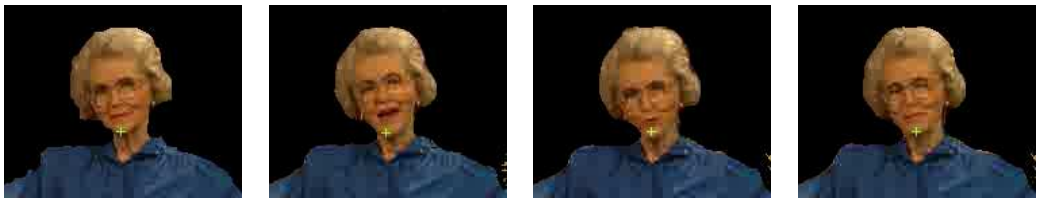
(i) Segmentation of Frame No.87 using Evolving scheme

(j) Segmentation Result using JSEG Scheme

(k) Temporal Segmentation Result using Segmented Result CDM



(l) VOP Extracted using Temporal Segmentation Result (i)



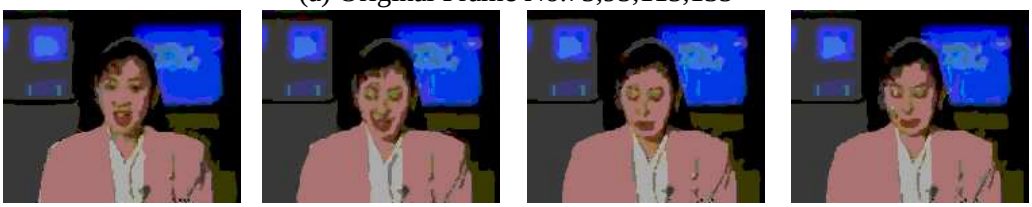
(m) Tracked Moving Object

(n) Temporal Segmentation Result using Original Frame CDM

(o) VOP Extracted using Temporal Segmentation Result (k)

Figure 6.1: VOP Generation for Grandma Video using Evolving Scheme

(a) Original Frame No.75,95,115,135



(b) Ground truth of Frame No.75,95,115,135

(c) Segmentation of Frame No.75 with Edge based Compound MRF Model

(d) Evolving Crude Result of Frame No. 79,83,87,95

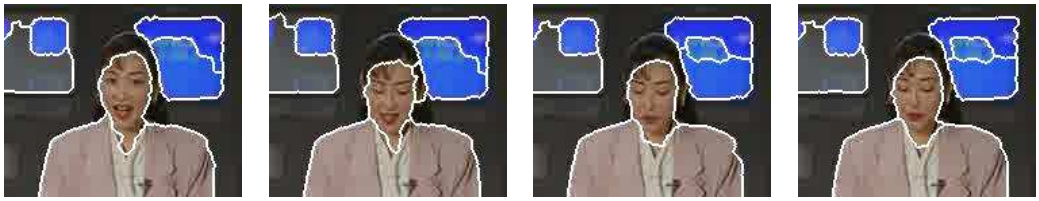
(e) Segmentation of Akiyo video Frame No.95 using Evolving scheme

(f) Evolving Crude Result of Frame No. 100,105,110,115

(g) Segmentation of Akiyo video Frame No.115 using Evolving Scheme

(h) Evolving Crude Result of Frame No. 120,125,130,135

(i) Segmentation of Akiyo video Frame No.115 using Evolving Scheme

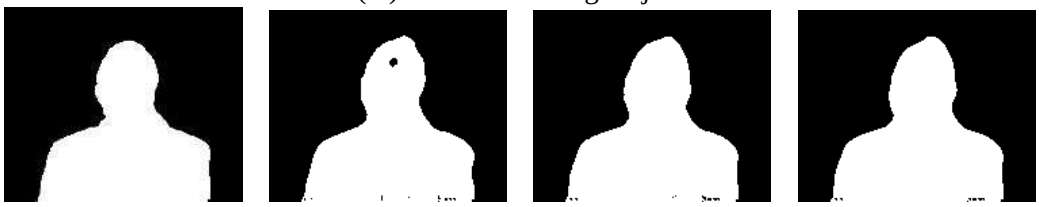


(j) Segmentation Resul using JSEG Scheme

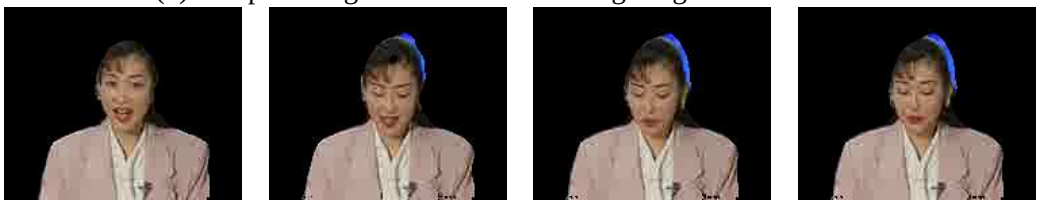
(k) Temporal Segmentation Result using Segmented Result CDM

(l) VOP Extracted by Evolving Scheme using Temporal Segmentation Result (k)

(m) Tracked Moving Object



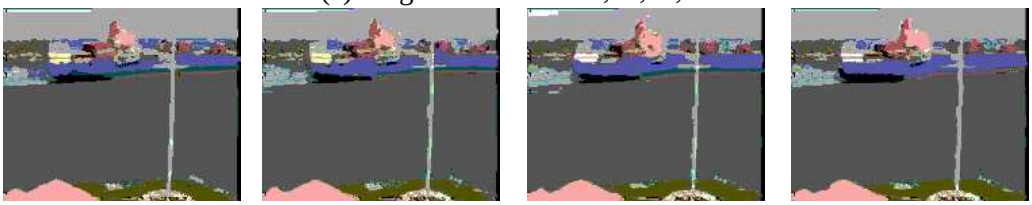
(n) Temporal Segmentation Result using Original Frame CDM



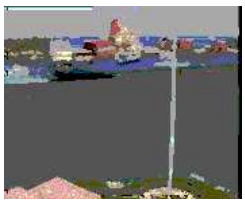
(o) VOP Extracted using Temporal Segmentation Result (m)

Figure 6.2: VOP Generation for Akiyo Video

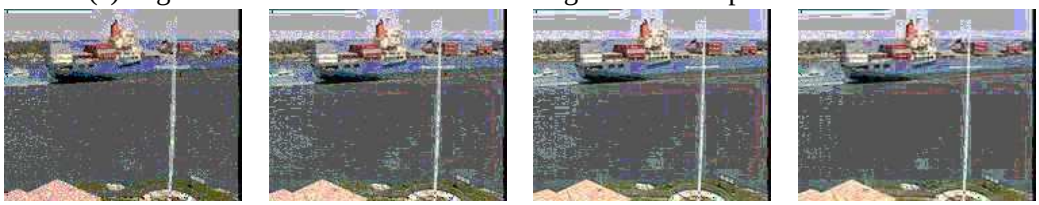
(a) Original Frame No.4,12,20,28



(b) Ground truth of Frame No.4,12,20,28



(c) Segmentation of Frame No.4 with Edge based Compound MRF Model



(d) Evolving Crude Result of Frame No. 6,8,10,12

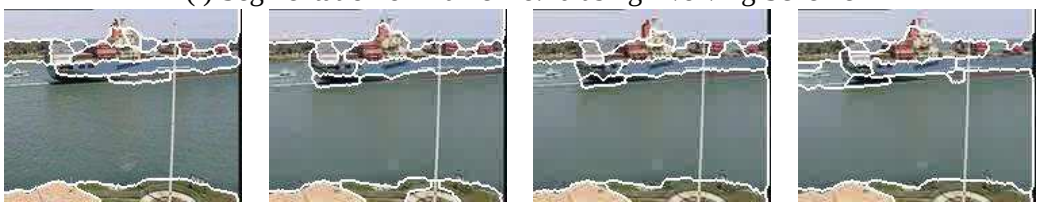
(e) Segmentation of Frame No.12 using Evolving scheme

(f) Evolving Crude Result of Frame No. 14,16,18,20

(g) Segmentation of Frame No.20 using Evolving Scheme

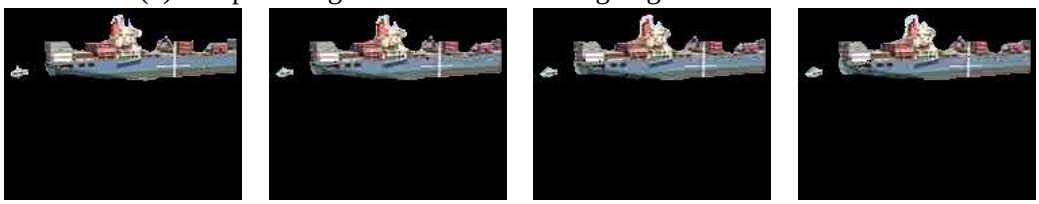
(h) Evolving Crude Result of Frame No. 22,24,26,28

(i) Segmentation of Frame No.20 using Evolving Scheme



(j) Segmentation Result using JSEG Scheme

(k) Temporal Segmentation Result using Segmented Result CDM



(l) VOP Extracted by Evolving Scheme using Temporal Segmentation Result (i)

(m) Temporal Segmentation Result using Original Frame CDM

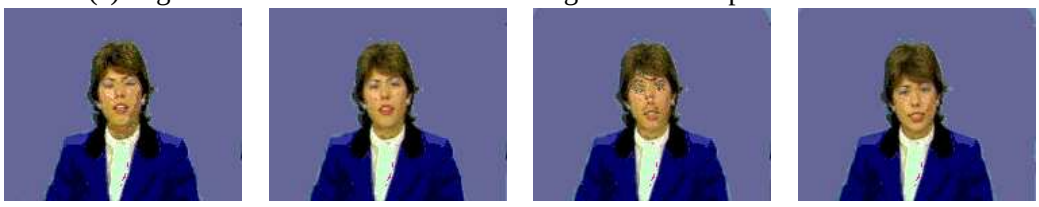
(n) VOP Extracted using Temporal Segmentation Result (k)

Figure 6.3: VOP Generated using Container Video

(a) Original Frame No.3,7,11,15

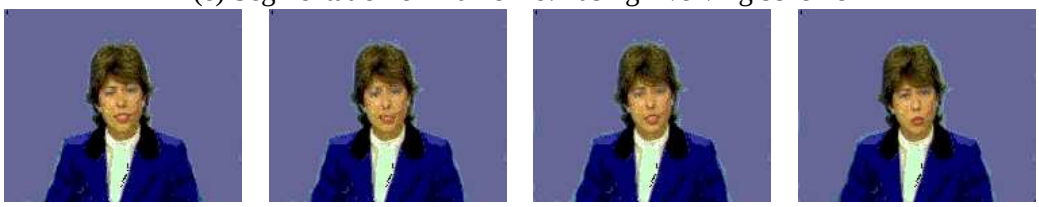
(b) Ground truth of Frame No.3,7,11,15

(c) Segmentation of Frame No.3 with Edge based Compound MRF Model



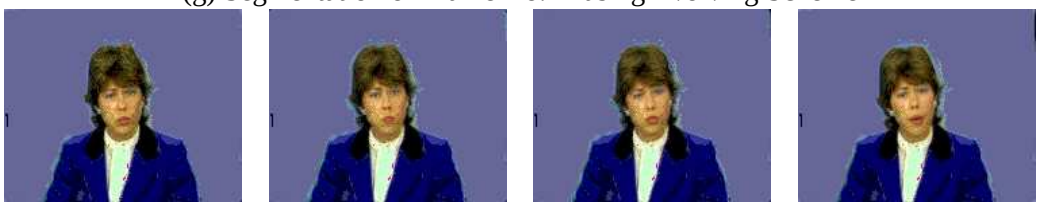
(d) Evolving Crude Result of Frame No. 4,5,6,7

(e) Segmentation of Frame No.7 using Evolving scheme



(f) Evolving Crude Result of Frame No. 8,9,10,11

(g) Segmentation of Frame No.11 using Evolving Scheme

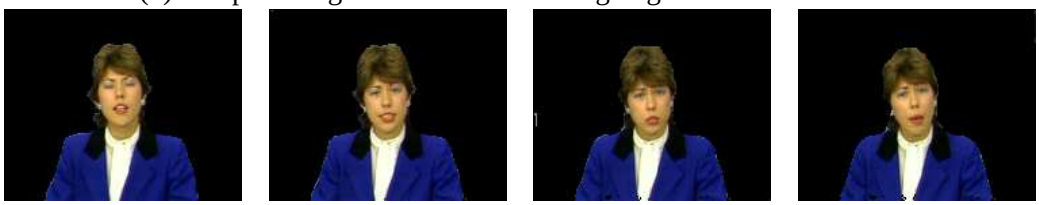


(h) Evolving Crude Result of Frame No. 12,13,14,15

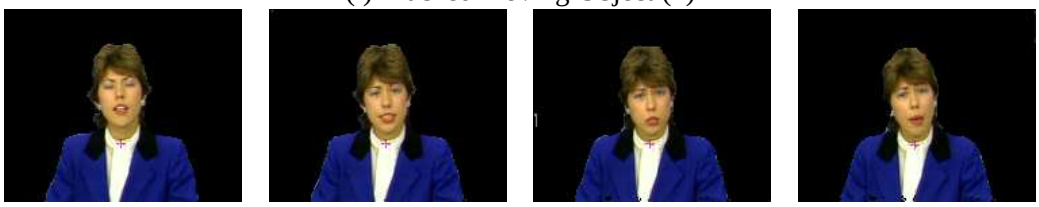
(i) Segmentation of Frame No.15 using Evolving Scheme

(j) Segmentation Result using JSEG Scheme

(k) Temporal Segmentation Result using Segmented Result CDM

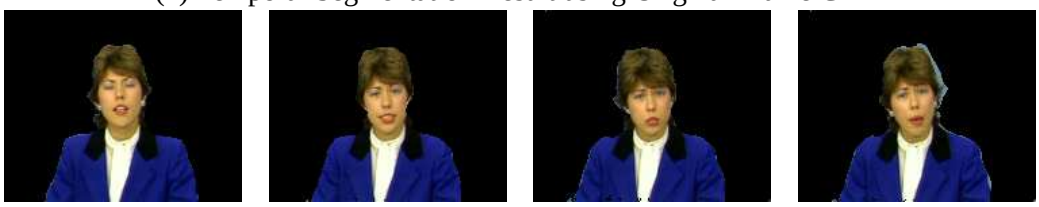


(l) Tracked moving Object (k)



(m) Tracked Moving Object

(n) Temporal Segmentation Result using Original Frame CDM



(o) VOP Extracted using Temporal Segmentation Result (k)

Figure 6.4: VOP Generated using Claire Video

Chapter 7

VIDEO OBJECT DETECTION USING MRF MODEL AND ADAPTIVE THRESHOLDING

Detecting regions of changes in video frames is of widespread interest due to a large number of applications in diverse disciplines. Change detection is widely used in video processing and analysis. Change detection researchers employ many common processing steps and core algorithms [3]. The change mask may result from a combination of underlying factors, including appearance or disappearance of objects, motion of objects relative to the background, or shape changes of objects. In addition, stationary objects can undergo changes in brightness or color. A key issue is that the change mask should not contain unimportant or nuisance forms of change, such as those induced by camera motion, sensor noise, illumination variation, nonuniform attenuation, or atmospheric absorption. The notions of significantly different and unimportant vary by application, which sometimes makes it difficult to directly compare algorithms. Estimating the change mask is often a first step toward the more ambitious goal of change understanding: segmenting and classifying changes by semantic type, which usually requires tools tailored to a particular application. Image differencing followed by thresholding is a popular method for change detection [9]. Thresholding plays a pivotal role in

such change detection methods. Many thresholding methods have been proposed in literatures, however, few of them are specific to change detection. Thresholding methods can be classified into gray-level distribution based [34] and spatial properties based [24].

In this chapter we have obtained the temporal segmentation scheme using the notion of adaptive thresholding and feature entropy. In existing temporal segmentation CDM, we have obtained from the original frames and global thresholding. The performance detoriates when the frames are noisy or variation in conditions of illumination is there. Hence, the notion of adaptive thresholding has been applied. The scheme also failed to give a good performance if there is a object present in a background of multiple class. So a modified CDM is proposed to obtain the change detection between the frames. we have proposed entropy based adaptive thresholding to obtain appropriate CDMs and hence the moving object parts of the video sequence. However, the spatio-temporal segmentation in MRF-MAP framework, as mentioned in the previous section is used to obtain the spatial segmentation. This spatial segmentation is combined with adaptive thresholding based temporal segmentation to construct the VOPs and thus moving object detection. The results obtained using adaptive thresholding is found to be superior to that of using global thresholding method.

7.1 ADAPTIVE THRESHOLDING

The problem that arises when illumination is not sufficiently uniform may be tackled by permitting the threshold to vary *adaptively* (or *dynamically*) over the whole image. In principle, there are several way of achieving this. One involves modeling the background the background within the image. Another is to work out

a local threshold value for each pixel by examining the range of intensities in its neighborhood. A third approach is to split the image into subimages and deal with them independently. Though this last method will clearly run into problems at the boundaries between subimages, and by the time these problems have been solved it will look more like one of the other two methods. Ultimately, all such methods must operate on identical principles. The differences arise in the rigor with which the threshold is calculated at different locations and in the amount of computation required in each case. In real-time applications the problem amounts to finding how to estimate a set of thresholds with a minimum amount of computation. The problem can sometimes be solved rather neatly in the following way. On some occasions-such as in automated assembly applications-it is possible to obtain an image of the background in the absence of any objects. This appears to solve the problem of adaptive thresholding in rigorous manner, since the tedious task of modeling the background has already been carried out. However, some caution is needed within this approach. Objects bring with them not only shadows (which can in some sense be regarded as part of the objects), but also an additional effect due to the reflections they cast over the background and other objects. This additional effect is nonlinear, in the sense that it is necessary to add not only the difference between the object and the background intensity in each case but also an intensity that depends on the products of the reflectance of pairs of objects.

Since the threshold used for each pixel depends on the location of the pixel in terms of the subimages, this type of thresholding is adaptive. Let us consider an example. All the subimages that didn't contain a boundary between object and background had variances of less than 75. All subimages containing boundaries had variances in excess of 100. Each subimage with variance greater than 100 was segmented with a threshold computed for that subimage using any of the global thresholding algorithm. There may be three approaches for finding the threshold

Figure 7.1: Image having size $M \times N$ is divided into 12 non-overlapping subimages, each of size $a \times b$, is thresholded by different thresholds $T1, T2, ..., T12$.

in adaptive thresholding:

1. The Chow and Kaneko Approach
2. Local Thresholding Approach
3. Adaptive Window Approach

7.2 CHOW AND KANEKO APPROACH

Chow and Kaneko [36] proposed a method in 1972 which is widely recognized as the standard technique for adaptive thresholding. It performs a thoroughgoing analysis of the background intensity variation, making few compromises to save computation. In this method , the image is divided into a regular array of overlapping subimages, and individual intensity histograms are constructed for each one. Those that are unimodal are ignored since they are assumed not to provide any useful information that can help in modeling the background intensity variation. However, the bimodal distributions are well suited to this task. These are individually fitted to pairs of Gaussian distributions of adjustable height and width, and the threshold values are located. Thresholds are then found, by interpolation, for the unimodal distributions. Finally, a second stage of interpolation is necessary to find the correct thresholding value at each pixel.

One problem with this approach is that if the individual subimages are made very small in an effort to model the background illumination more exactly, the statistics of the individual distributions become worse, their minima become less well defined, and the thresholds deduced from them are no longer significant i,e. it does not pay to make subimages too small and that ultimately only a certain level of accuracy can be achieved in modeling the background in this way. The situation is highly data dependent, but little can be expected to be gained by reducing the subimage size below 32×32 pixels. Chow and Kaneko employed 256×256 pixel

images and divided these into a 7×7 array of 64×64 pixel subimages with 50% overlap.

Overall, this approach involves considerable computation, and in real-time applications it may well not be viable for this reason. However, this type of approach is of considerable value in certain medical, remote sensing, and space applications.

7.3 LOCAL THRESHOLDING APPROACH

In real-time applications the alternative approach mentioned earlier is often more useful for finding local thresholds. It involves analyzing intensities in the neighborhood of each pixel to determine the optimum local thresholding level. Ideally, the above histogram technique would be repeated at each pixel, but this would significantly increase the computational load of this already computationally intensive technique. Thus, it is necessary to obtain the vital information by an efficient sampling procedure. One simple means of achieving this is to take a suitably computed function of nearby intensity values as the threshold. Often the mean of the local intensity distribution is taken, since this is a simple statistics and gives good results in some cases.For example, in astronomical images stars have been thresholded in this way.

Another frequently used statistic is the mean of the maximum and minimum values in the local intensity distribution. Whatever the sizes of the two main peaks of the distribution, this statistic often gives a reasonable estimate of the position of the histogram minimum. The theory presented earlier shows that this method will only be accurate if

1. the intensity profiles of object edges are symmetrical,

2. noise acts uniformly everywhere in the image so that the widths of two peaks of the distribution are similar, and

3. the heights of the two distributions do not differ markedly.

Sometimes these assumptions are definitely invalid-for example, when looking for(dark) cracks in eggs or other products. In such cases the mean and maximum of the local intensity distribution can be found, and a thresholdcan be deduced using the statistic

$$T = mean - (maximum - mean) \tag{7.1}$$

where the strategy is to estimate the lowest intensity in the bright background assuming the distribution of noise to be symmetrical. Use of the mean here is realistic only if the crack is narrow and does not affect the value of the mean significantly. If it does, then the statistic can be adjusted by use of an ad-hoc parameter

$$T = mean - k(maximum - mean) \tag{7.2}$$

where k may be as low as 0.5.

This method is computationally less intensive but they are somewhat unreliable because of the effects of noise. All these methods work well only if the size of the neighborhood selected for estimating the required threshold is sufficiently large to span a significant amount of object and background. In many practical cases, this is not possible and the method then adjusts itself erroneously, for example, so that it finds darker spots within dark objects as well as segmenting the dark objects themselves.

7.4 ADAPTIVE WINDOW BASED APPROACH

One of the primary disturbances sources id from uneven lighting, which often exists in the capturing of an image, especially during field operation. The main causes for uneven lighting are:

1. the light may not be always stable
2. the object is so large such that it creates an uneven distribution of the light, and
3. the background is unable to be optically isolated from shadows of other objects.

One possible solution to this problem is to partition the whole image into certain small windows, and then use those existing methods to threshold each small window. This process is called thresholding in partitioned windows. The smaller the window size is, the better the result will be. However, when the window size becomes too small, it can produce the problem of homogeneous windows, i,e., windows contain only background or object pixels. As a consequence, black areas called ghost objects will occur after thresholding. Therefore, there is a need to develop a new technique for automatically selecting window size in order to obtain optimal result i.e., adaptive window selection. This technique is based on the pyramid data structure manipulation, and the window size is adaptively selected according to Lorentz information measure.

7.5 ENTROPY

The entropy of a system as defined by Shannon [37] gives a measure of our ignorance about its actual structure. In the context of information theory, shannon's function is based on the concept that information gain from an event is inversely related to its probability of occurrence. The logarithmic behavior of entropy is considered to incorporate the additive property of information.

7.5.1 Shannon's entropy

Shanon defined the entropy of an $n-$state system as

$$H = -\sum_i p_i \ \log_2 p_i \tag{7.3}$$

where p_i is the probability of occurrence of the event i and

$$\sum_i p_i = 1, \quad 0 \leq p_i \leq 1 \tag{7.4}$$

In case of a binary system, the entropy becomes

$$H = p \ \log_2 p - (1-p) \ \log_2(1-p) \tag{7.5}$$

The entropy H is claimed to express a measure of ignorance about the actual structure of the system. In order to explain why such an expression is taken as a measure of ignorance, let us critically examine the philosophy behind shanon's entropic measure with an example given below.

Suppose a six-faced die, covered with a box, is placed on a table and someone is asked to guess the number on the top most face of the die. Since the exact state of the die is not known, he/she can describe the state of the die by the probability distribution of occurrences of different faces on the top. In other world, the state of the die can be expressed by specifying $p_i, \ i = 1, 2, ..., 6$; where p_i is the probability that the ith face is the topmost face. Obviously,

$$0 \leq p_i \leq 1 \quad and \quad \sum_{i=1}^{6} p_i = 1$$

When the box is opened, the state of the die becomes known to us and we gain some information. A very natural question arises, How much information did we gain ?

Let $p_k = \max_i p_i$:the most probable event and $p_m = \min_i p_i$:the least probable event. Now, if the kth face appears on the top, the gain in information would

be minimum, whereas the occurrence of the mth face on the top would result in the maximum gain.

Thus we see that the gain in information from an event is inversely related to its probability of occurrence.This, of course, intuitively seems all right. For example, if somebody says, ”The sun rises in the east”, the information content of the statement is practically nil. On the other hand if one says, ”He is ten feet in height”, the information content of the statement is very high, as it is an unlikely event. A commonly used measure of such a gain is

$$\triangle I = \log_2 (1/p_1) = -\log_2 (p_i) \tag{7.6}$$

In order to justify the logarithmic function, the following points can be stated:

1. It gives additive property of information. To make it more clear, suppose two independent events m and n with probabilities of occurrence p_m and p_n have occurred jointly, then the additive property says

   $$\triangle I(p_m \cdot p_n) = \triangle I(p_m) + \triangle I(p_n) \tag{7.7}$$

   where $(p_m \cdot p_n)$ is the probability of the joint occurrence of the events m and n. Thus the additive property can be stated as follows. The information gain from the joint occurrence of more than one event is equal to the sum of information gain from their individual occurrence.

2. The gain in information from an absolutely certain event is zero, i.e., $\triangle I(p_i = 1) = 0$.

3. As p_i increases, $\triangle I(p_i)$ decreases. Gain in information from the experiment can be written as

   $$H = E(\triangle I) = -\sum_i p_i \ \log_2 p_i \tag{7.8}$$

   The value of H denotes the entropy (shanon’s entropy) of the system.

7.5.2 Entropic measures for image processing

Based on the concept of Shanon's entropy, different authors have defined entropy for an image and its extension to fuzzy sets. Let us consider those measures and the associated problems when applied to image processing and recognition problems.

Let $X = [X(m,n)]_{P\times Q}$ be an image of size $P \times Q$, where $X(m,n)$ is the gray value at (m,n); $X(m,n) \in G_L = 0, 1, ..., L-1$, is the set of gray levels. Let N_i be the frequency of the gray level i. Then

$$\sum_{i=0}^{L-1} N_i = P \times Q = N(say) \tag{7.9}$$

Lets consider the gray level histogram of X an $L-$ symbol source, independently from the underlying image. In addition to this, they also assumed that these symbols are statistically independent.

Following Shanon's definition of entropy from (7.3), the entropy of image (histogram) is defined as

$$H = -\sum_{i=0}^{L-1} p_i \ \log_2 p_i \tag{7.10}$$

7.6 PROPOSED METHODS

The proposed method of section 3.4 based on the Lorentz information measure and is greatly dependent on the proper choice of initial window size. In order to ameliorate this situation, we propose a method of window growing instead of window merging.

7.6.1 Window growing based on feature entropy

The basic notion of window growing is to fix the window size primarily focussing on the information measure of the image at different scale. In other words, fixing

the size of the window not only depends on the entropy of the chosen window but also the feature entropy of the window. The edges of the window are considered as the features and the feature entropy is computed. Since, the edge map represents the image information at a different scale, the entropy at this scale also plays a pivotal role for image segmentation. Thus, the basic notion is to capture the information at a different scale. It is known that entropy can be

$$H_w = \sum_{i=1}^{G} p_i \ \log_e \left(\frac{1}{p_i} \right) \tag{7.11}$$

where p_i is the probability distribution of the ith gray value, H_w denotes entropy of the window, G denotes the total number of gray values. Over a given window, the edge map is computed and the entropy of the edge map is

$$H_{wf} = \sum_{i=1}^{G} p_{f_i} \ \log_e \left(\frac{1}{p_{f_i}} \right) \tag{7.12}$$

where H_{wf} denotes the entropy of the edge map of the window. The following are the two proposed window growing criterion.

case I: WG-I

The window is fixed if the following is satisfied

$$H_w > Th \tag{7.13}$$

where Th is selected based on the entropy of the total image.

case II: WG-II

The following criterion is considered for window fixing after the grow of the window.

$$H_w > Th$$

$$subject\ to\ the\ constraint,\ H_{wf} > Th_f \tag{7.14}$$

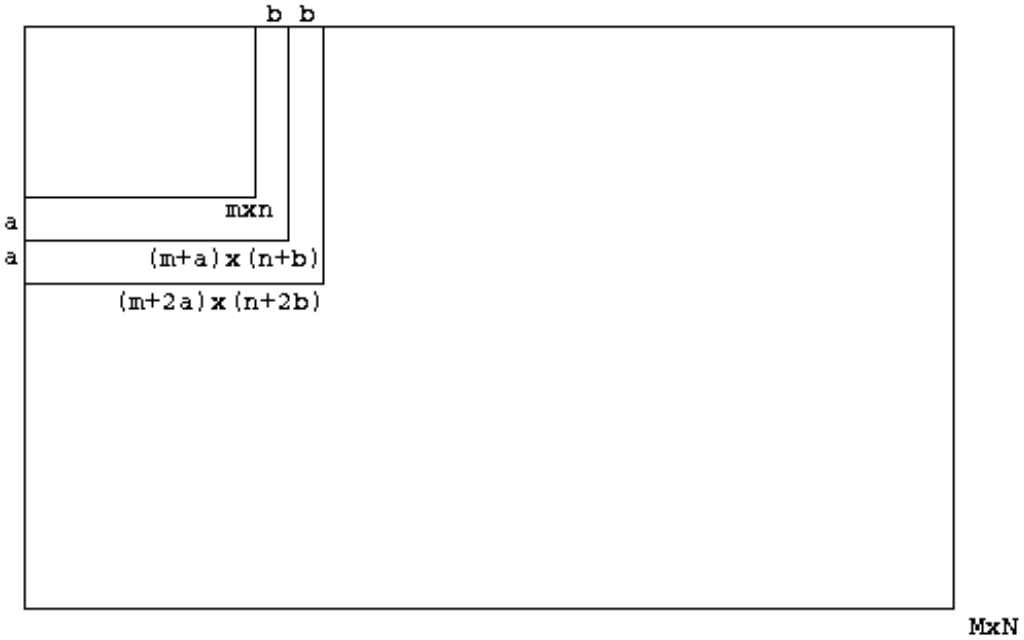


Figure 7.2: Illustration of Window growing method

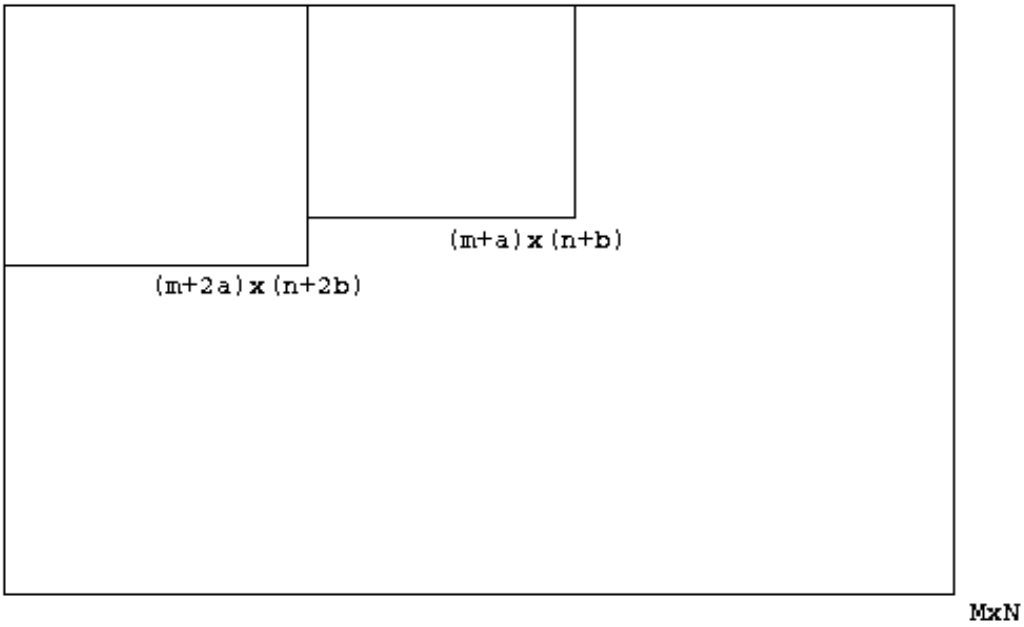


Figure 7.3: Illustration of Window growing method

The thresholds Th and Th_f for the above inequality are chosen based on the total entropy of the image and that of edge map respectively. Empirically, it is found that the thresholds are closer to the entropy of the whole image and whole edge map.

Figure 4.1 shows the illustration of window growing method. First a window of size $m \times n$ is chosen and it is merged with size $a \times b$ to make a window of size $(m + a) \times (n + b)$ and so on till condition is not satisfied. Figure 4.2 shows that, after fixing of one window, another of size $m \times n$ started from adjacent side. The followings are the salient steps of the algorithm.

Algorithm

1. Choose a window of size w.
2. Determine the entropy from the gray value distribution of the considered window.
3. Compute the edge map and determine the entropy of the edge map of the window.
4. Choose two thresholds Th and Th_f and test the conditions of the (7.13) and (7.14).
5. If the window is fixed, then start from the next window. If not fixed, then increase the window size by $10 - 25$.
6. Repeat steps 2-5 till he whole image is exhausted.

7.7 PROPOSED CDMs

Usually, CDM is obtained by taking the difference of the original frames, which fails to give satisfactory result in case of a object present with multiple class back-

ground, situation where there is less variation in the gray level of the object. Hence, a modified CDM is proposed. Initially we take the absolute difference of the estimated labels of consecutive frames i.e if y_t and y_{t-1} are two consecutive frames, and x_t and x_{t-1} are the estimated labels a new sequence is obtained using

$$x_{d-t} = |x_t - x_{t-1}| \tag{7.15}$$

a new sequence of x_{d-t} is created and the CDM is the difference of this consecutive frames of this sequence. Thus, the modified CDM is

$$CDM_m = |x_{(d-t)_t} - x_{(d-t)_{t-1}}| \tag{7.16}$$

This CDM_m is subjected to adaptive thresholding to obtain temporal segmentation.

7.8 OBJECT DETECTION USING ADAPTIVE THRESHOLDING

In this scheme also the spatial segmentation of each frame is obtained. The spatio-temporal framework as given in Section. 4 is used to obtain the labels of a given frame. The video sequence is modeled as Compound edgebased MRF and the pixel labeling problem is formulated using MAP estimation criterion. The MAP estimates are obtained using the hybrid Algorithm. Thus, the spatial segmentation of individual frames are obtained.

The proposed adaptive thresholding is used to obtain the temporal segmentation. Initially the CDMs are obtained using difference of the estimated labels and because of noise where the CDM becomes noisy. This noisy CDM, when used for the temporal segmentation yields noises and hence wrong object detection. The

proposed entropy based adaptive thresholding scheme is used in CDMs to obtain the moving objects while eliminating noises. The window growing based adaptive scheme is used to obtain a accurate CDMs. The temporal segmentation is obtained with the CDMs while using the history of the pixel labels. Thereafter the VOPs are constructed and hence the object is detected. In this scheme, the noises in CDMs that otherwise have falsely detected moving objects are avoided.

7.9 SIMULATION AND RESULTS DISCUSSION

In this chapter the object detection is based on spatial segmentation and temporal segmentation. The edgebased MRF model is used and the MAP estimates are obtained by Hybrid Algorithm. In simulation five different examples have been considered. The first example is the Grandma video sequence as shown in Fig. 7.4. The edge based MRF model is used and the spatial segmentation is obtained as shown in Fig. 7.4(c). The temporal segmentation is obtained using the CDM and global thresholding. These are shown in Fig. 7.4(e). The object detected is shown in Fig. 7.4(f) where there are some background pixels reflected in the foreground. Some noisy pixels are still present with the foreground. Temporal segmentation obtained using adaptive thresholding is shown in Fig. 7.4(g) where it can be observed that noisy pixels are absent and hence the detected objects using this are shown in Fig. 7.4(h). It can be seen that the background pixels earlier reflected is absent and the objects are detected correctly and the tracking is done accordingly. The MRF model parameters are tabulated in Table. 7.1.

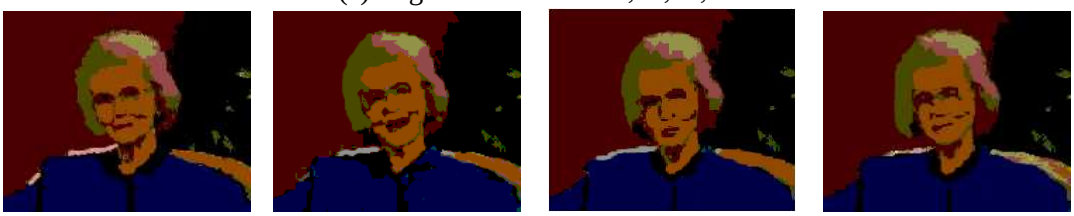
The second example considered is the Claire Video sequence shown in Fig. 7.5. As observed from Fig. 7.5(e) in the temporal segmentation using global thresholding, some portion of the object such as portions from the head is miss-

ing. In case of the temporal segmentation using adaptive thresholding, the above missing portion appear in the object. Hence, in the detected part of the moving object is the complete object itself. Thus, adaptive thresholding could eliminate the background noises and hence the moving objects could be tracked.

The next three sequences are traffic sequences with single moving objects or multiple moving objects. Fig. 7.6 shows the results for traffic sequence having three moving objects. The global thresholding approach detected two objects(Car and the man) as shown in Fig. 7.6(f). As observed from Fig. 7.6(h), using adaptive thresholding approach two object could be detected. On the second traffic sequence as shown in Fig. 7.7, the single object is detected in case of adaptive thresholding where as in global thresholding the back portion of the car is missing as seen from Fig. 7.7(e). Fig. 7.8 shows the case of multiple moving objects and global thresholding approach produce results with many missing parts of the moving object as seen from Fig. 7.8(e) and Fig. 7.8(f). Fig. 7.8(h) shows the results obtained using adaptive thresholding approach, where it can be seen that all the parts of the moving object has been detected.

The last example is shown in Fig. 7.9 where the original frames are blurred ones. Use of global thresholding in temporal segmentation has lots of missing parts of the moving objects as seen from Fig. 7.9(e). Adaptive thresholding could detect the objects fully as shown in Fig. 7.9(h). Thus the proposed adaptive thresholding could take care of blurred situation.

(a)Original Frame No.12,37,62,87

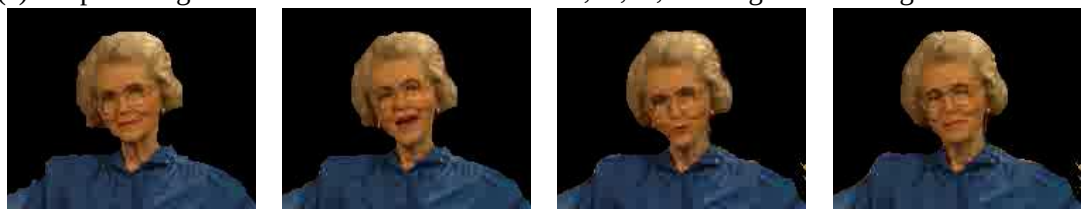


(b)Ground truth of Frame No.12,37,62,87

(c) Segmentation of Frame No.12,37,62,87 with Edge based Compound MRF Model

(d)Segmentation result with JSEG Scheme

(e)Temporal Segementation result of Frame No.12,37,62,87 using CDM of segmented Frames



(f)Detected Moving Object of Frame No.12,37,62,87 using results(e)

(g)Temporal Segementation result of Frame No.12,37,62,87 using proposed CDM and adaptive thresholding

(h)Detected Moving Object of Frame No.12,37,62,87 using results(g)

(i)Tracked Moving Object of Frame No.12,37,62,87 using results(g)

Figure 7.4: VOP Generation of Grandma video sequences

(a)Original Frame No.3,7,11,15

(b)Ground truth of Frame No.3,7,11,15

(c) Segmentation of Frame No.3,7,11,15 with Edge based Compound MRF Model

(d)Segmentation result with JSEG Scheme

(e)Temporal Segementation result of Frame No.3,7,11,15 using CDM of segmented Frames

(f)Detected Moving Object of Frame No.3,7,11,15 using results(e)

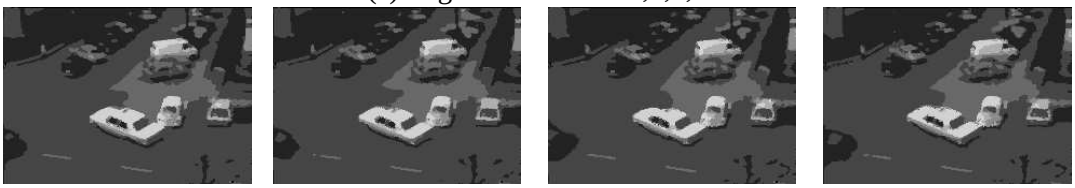
(g)Temporal Segementation result of Frame No.3,7,11,15 using proposed CDM and adaptive thresholding

(h)Detected Moving Object of Frame No.3,7,11,15 using results(g)

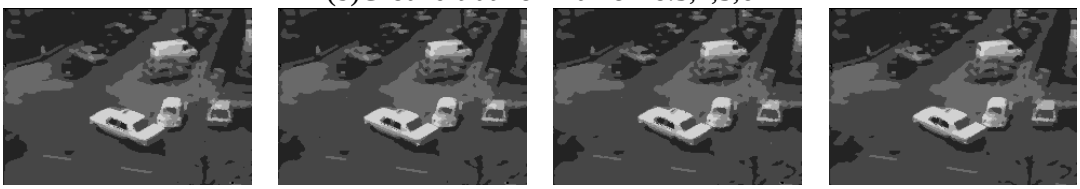
(i)Tracked Moving Object of Frame No.3,7,11,15 using results(g)

Figure 7.5: VOP Generation of Claire sequences

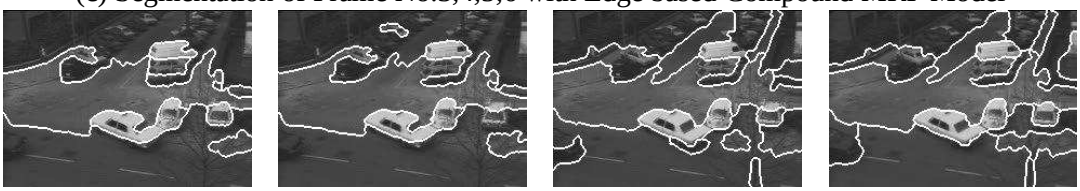
(a)Original Frame No.3,4,5,6



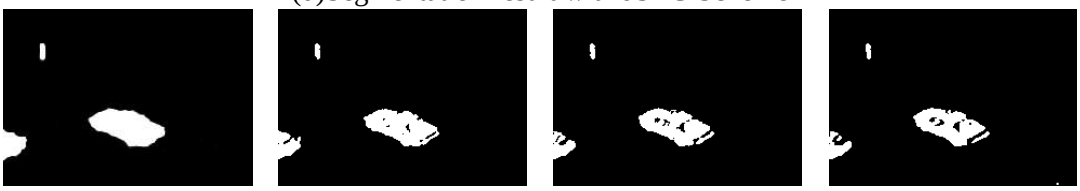
(b)Ground truth of Frame No.3,4,5,6



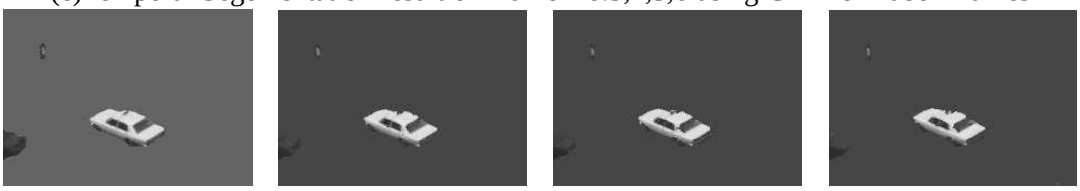
(c) Segmentation of Frame No.3,4,5,6 with Edge based Compound MRF Model



(d)Segmentation result with JSEG Scheme

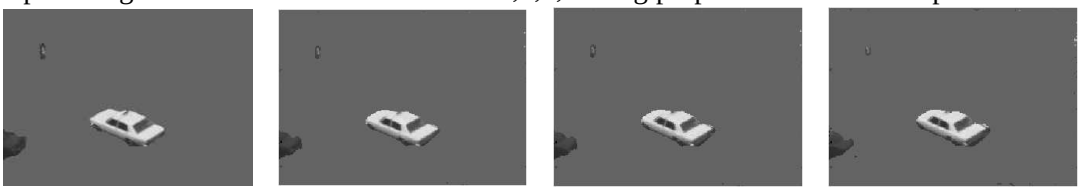


(e)Temporal Segementation result of Frame No.3,4,5,6 using CDM of Label Frames



(f)Detected Moving Object of Frame No.3,4,5,6 using result (g)

(g)Temporal Segementation result of Frame No.3,4,5,6 using proposed CDM and adaptive thresholding

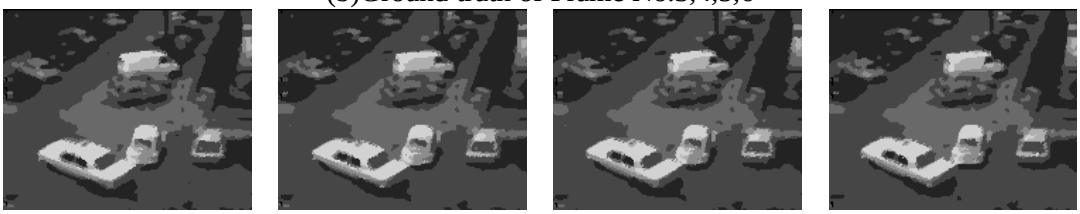


(h)Detected Moving Object of Frame No.3,4,5,6 using results(g)

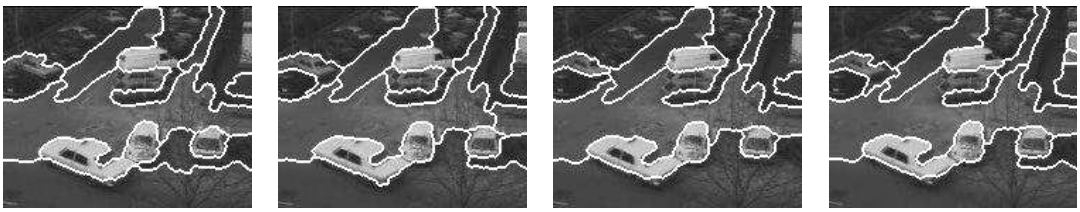
Figure 7.6: VOP Generation of Canada Traffic Video sequences

(a)Original Frame No.3,4,5,6

(b)Ground truth of Frame No.3,4,5,6



(c) Segmentation of Frame No.3,4,5,6 with Edge based Compound MRF Model

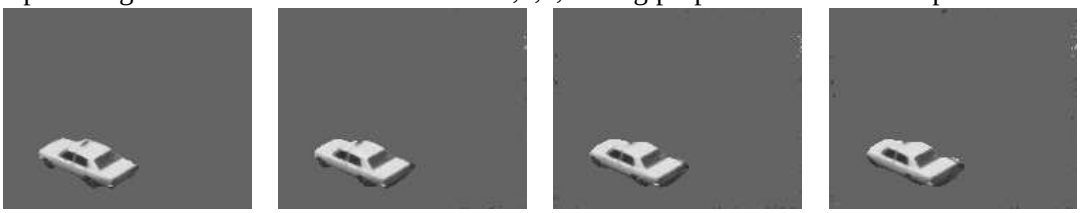


(d)Segmentation result with JSEG Scheme

(e)Temporal Segementation result of Frame No.3,4,5,6 using CDM of segmented Frames

(f)Detected Moving Object of Frame No.3,4,5,6 using results(e)

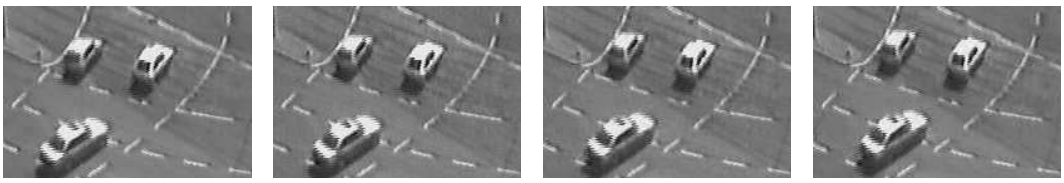
(g)Temporal Segementation result of Frame No.3,4,5,6 using proposed CDM and adaptive thresholding



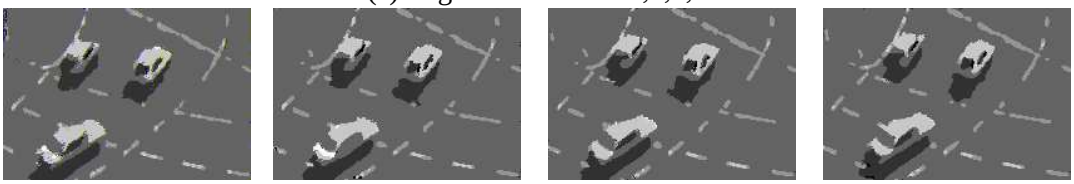
(h)Detected Moving Object of Frame No.3,4,5,6 using results(g)

(i)Tracked Moving Object of Frame No.3,4,5,6 using results(g)

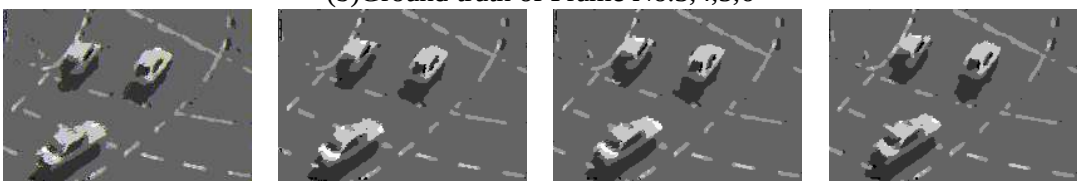
Figure 7.7: VOP Generation of Traffic video sequences



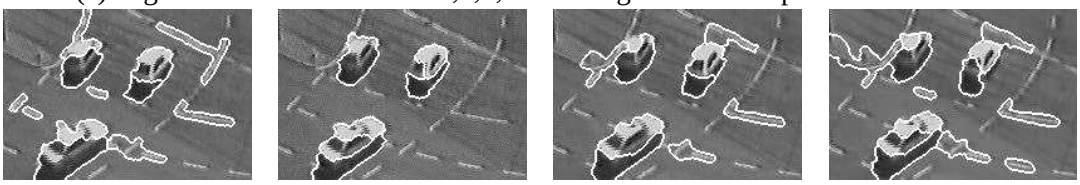
(a)Original Frame No.3,4,5,6



(b)Ground truth of Frame No.3,4,5,6



(c) Segmentation of Frame No.3,4,5,6 with Edge based Compound MRF Model

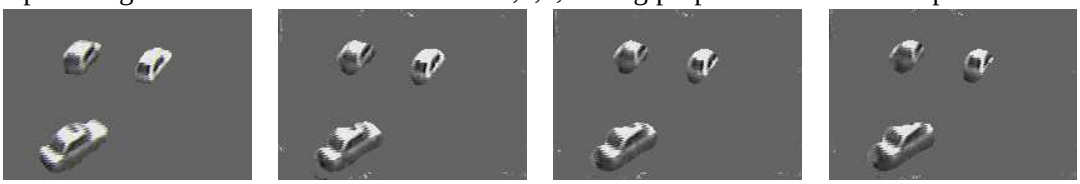


(d)Segmentation result with JSEG Scheme

(e)Temporal Segementation result of Frame No.3,4,5,6 using CDM of segmented Frames

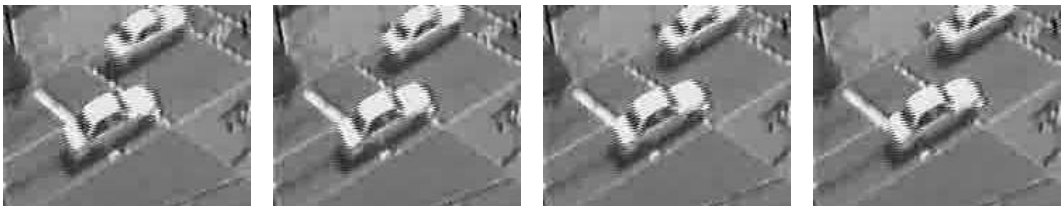
(f)Detected Moving Object of Frame No.3,4,5,6 using results(e)

(g)Temporal Segementation result of Frame No.3,4,5,6 using proposed CDM and adaptive thresholding

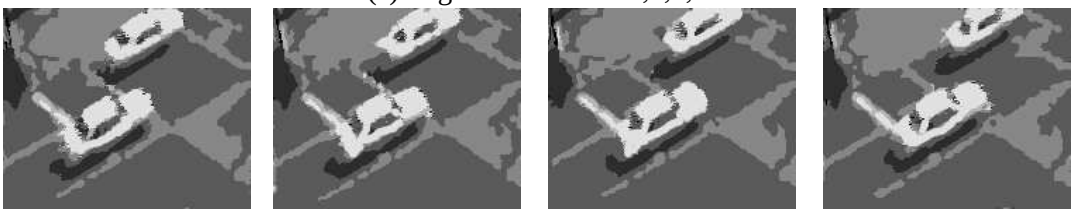


(h)Detected Moving Object of Frame No.3,4,5,6 using results(g)

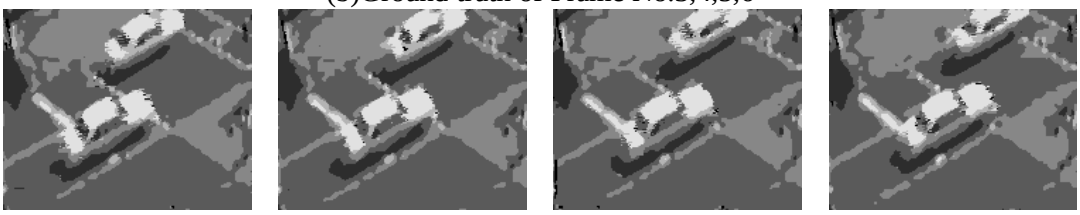
Figure 7.8: VOP Generation of Traffic-2 video sequences



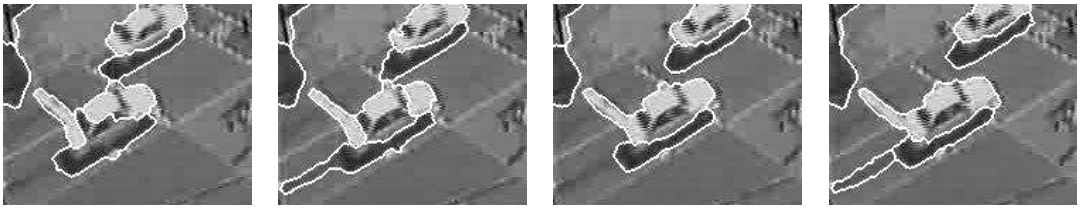
(a)Original Frame No.3,4,5,6



(b)Ground truth of Frame No.3,4,5,6

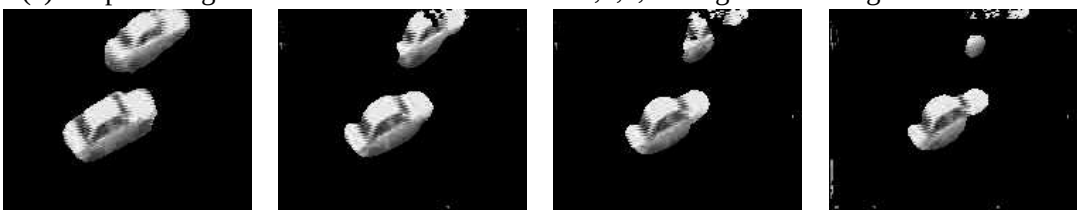


(c) Segmentation of Frame No.3,4,5,6 with Edge based Compound MRF Model



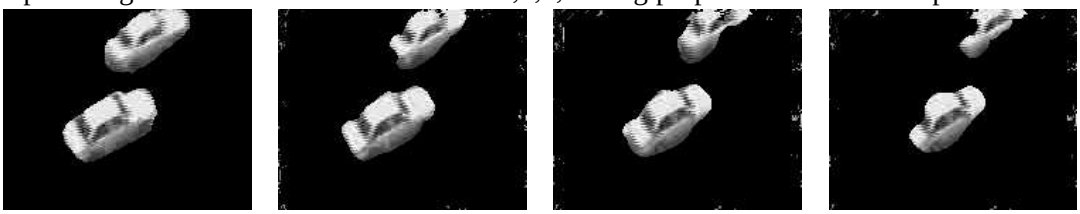
(d)Segmentation result with JSEG Scheme

(e)Temporal Segementation result of Frame No.3,4,5,6 using CDM of segmented Frames



(f)Detected Moving Object of Frame No.3,4,5,6 using results(e)

(g)Temporal Segementation result of Frame No.3,4,5,6 using proposed CDM and adaptive thresholding



(h)Detected Moving Object of Frame No.3,4,5,6 using results(g)

Figure 7.9: VOP Generation of Sequence video sequences

VIDEO	α	β	γ	σ
Grandma	0.05	0.009	0.007	5.19
Claire	0.009	0.008	0.007	1.00
Traffic Cannada	0.01	0.009	0.007	3.0
Traffic Car	0.01	0.009	0.007	3.0
Traffic Sequence	0.009	0.008	0.007	3.0
Traffic Car-2	0.01	0.008	0.007	4.0
Traffic Bus	0.01	0.009	0.007	3.0

Table 7.1: Parameters for diffrent videos of the given videos

$Video$	$FrameNo.$	$Evolving$	$JSEG$
$Grandma$	12	0.24	6.82
	37	0.15	4.65
	62	0.15	4.5
	87	0.12	3.88
$Claire$	3	0.41	2.95
	7	0.39	2.47
	11	0.76	2.91
	15	0.76	2.91
$CanadaTraffic$	3	0.1	5.95
	4	0.41	8.23
	5	0.52	16.65
	6	0.46	7.1
$TrafficCar$	3	0.75	9.56
	4	0.41	10.44
	5	0.65	7.56
	6	0.61	22.05
$TrafficSequence$	3	1.41	15.03
	4	1.25	11.50
	5	1.33	15.19
	6	0.84	17.73
$TrafficCar-2$	3	1.53	7.02
	4	1.54	7.82
	5	2.37	6.49
	6	1.37	5.83
$TrafficBus$	3	6.10	0.18
	4	5.27	0.40
	5	4.97	0.44
	6	5.10	0.39

Table 7.2: Percentage of Misclassification Error

Chapter 8

CONCLUSION

In this dissertation, the problem of slow as well as fast moving video objects detection is addressed. Initially the existing temporal segmentation with CDMs could detect fast moving video objects but failed with slow moving video objects. In the scheme it has been assumed to have the reference frame. Often in practice reference frame may not be available. Therefore the problem is formulated using spatio-temporal framework.

The spatial segmentation problem is considered in supervising mode where the model parameters are assumed to be known a priori. A compound MRF model is proposed to model the video sequence. In the first case, the a priori distribution of the model takes care of the pixel distribution of a frame spatially and also the pixel distribution in the temporal directions. This is called edge less MRF model. The edge features in the temporal directions are extracted and the MRF a priori distribution is modified to take edge features in the temporal directions besides the edge features in the spatial domain. This model has been named as the edge based model. The spatial segmentation problem is formulated as a pixel labeling problem and the pixel labels are estimated using MAP estimation criterion. Simulated Annealing algorithm used to obtain the MAP estimates. It has been observed

that SA is computationally involved and hence takes appreciable amount of time to converge to the solution. Hence, a Hybrid Algorithm exploiting the globally convergent features of SA and the local convergent features of ICM is proposed to obtain the MAP estimates. The proposed algorithm is found to be much faster than that of SA. It is approximately 10 or more times faster than that of SA. The only bottleneck was to fix the epochs for global convergence on a trial and error basis. The results obtained by the Hybrid Algorithm are found to be comparable with that of SA. Thus a substantial saving in computational time could be achieved. As far as the proposed MRF model is concerns, it is proved to be an efficient model for modeling the video sequences. The only bottleneck of the scheme is that the model parameters are selected on a trial and error basis.

The performance of the scheme could further be improved by changing the model. The changes of the frames were obtained by taking the difference is CDMs were obtained. The CDMs for difference are assumed to have some temporal dependence and continuity and hence the changes are also modeled as MRFs. Hence, the a priori distribution of MRF distribution not only took care of the edges of the temporal sequences but also took into account the changes in the temporal direction. This model is known as the Change based MRF model. This model proved to be more efficient than the other models. Temporal segmentation in this case is obtained using CDMs together with the history of the labels. Here the CDMs are obtained by taking the difference of the estimated labels of each frame rather than the original frames. This scheme with this model proved to be best among edge based and edgeless approaches. The above scheme is quite computationally involved because in spatial segmentation of every image frame has to be obtained. This prohibited from the idea of running the scheme for real time sequences.

In order to make a viable scheme from the practical stand point, an evolutionary approach based segmentation scheme is proposed. In this scheme, segmentation of only one frame has to be obtained by spatio-temporal frame with MAP estimation principle. The rest of the label maps of different frames are obtained by evolving the label map of the initial frame with the proposed evolution strategy. This reduced the computational burden appreciably thus, leading a stepping stone for real time implementation. Here also the temporal segmentation uses the labels of different frames as opposed to the original frames. It has been observed that there are some errors in the object detection and tracking. It was due to the presence of noise in the CDMs reflected from original frames or variation of illumination in the original frames.

Hence, to take care of such situations an entropy based adaptive threshold strategy is proposed to eliminate the noises in CDMs. The temporal segmentation and the VOPs thus constructed are found to be better than all other methods. All the proposed scheme are supervised in native because the parameters are selected on trial and error basis. the schemes can be made unsupervised with estimation of model parameters together with the labels. Model parameter estimation is worth pursuing. Fusing label fields to obtain improve results is also worth pursuing.

Bibliography

[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] R .F. Gonzalez, and R. E. Wood,*Digital Image Processing*, Singapore, Pearson Education,2001.

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

[5] A. Blake, and A. Zisserman,"Visual Reconstruction".London: MIT press, 1987.

[6] M. Kim, J. choi, D. Kim and H. Lee, "A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences based on Spatio-Temporal information,"*IEEE Transaction on circuits and Systems for Video Technology*, Vol. 9, No. 8, pp. 1216-1226, Dec 1999.

[7] A. Cavallaro and T. Ebrahimi, "Change Detection based on Color Edges,"*IEEE International Symposium on Circuits and Systems 2001*ISCAS2001, Vol. 2, pp. 141-144, 2001.

[8] L. shi, Z. Zhang and P. An, "Automatic segmentation of Video Object Plane based on Object tracking and Matching,"*proceedings of 2001 International*

Symposium on Intelligent multimedia, Video and Speech Processing, pp. 510-513, May 2001.

[9] H. Chen, F. Qi and S. Zhang, ”Supervised Video Object Segmentation using a Small Number of Interaction, ”*Proceedings. of International Conference on Acoustics, speech,and signal processing,*, pp.383-386, 6-10th April 2003.

[10] D. Butler, S. Sridharan and V. M. Bove,”Real Time Adaptive Background segmentation, ”*Proceedings. of International Conference on Acoustics, speech,and signal processing,*, pp.341-344, 6-10th April 2003.

[11] Ç. Erdem, B. Sankur and A. M. Teklap, ”Performance Measure for Video Object segmentation and Tracking, ”*IEEE Transaction on Image Processing,*, vol. 13, No. 7, pp.937-951, July 2004.

[12] J. Zhang, L. Zhang and H. M. Tai ”Efficient Video Object Segmentation Using Adaptive Background Registration and Edge based Change detection Techniques ,” *Proceedings of IEEE International Conf. on multimedia and Expo*,”ICME 04, Southwest Jiaotong Univ., Chengdu, China, Vol.2, pp.1467-1470, April 2004

[13] S. G. Sun, D. M. kwak, W. B. Jang and D. J. Kim ”Small target detection using Center-surround Difference with Locally Adaptive threshold ,” *Proceedings of 4th International Symposium on Image and Signal processing and Analysis*,” ISPA05 pp.402-407, 2005.

[14] Y.Deng,B.S.Manjunath, ”Unsupervised Segmentation of Color-Texture Regions in Images and Video,“ *IEEE Transactions on Pattern Analysis And Machine Intelligence* vol. 23,pp. 800-810, 2001.

[15] A. Papoulis and S. U. pillai,”Probability, Random Variables and Stochastic Process”.Mc Graw Hill,2002.

[16] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,”*IEEE Trans. on Pattern Analysis and Machine Intelligence*, volume 6, no. 6,pp. 721-741, Nov. 1984.

[17] 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.

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

[19] 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.

[20] 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.

[21] E. Y. Kim, S. H. Park and H. J. Kim, “ A Genetic Algorithm-based segmentation of Markov Random Field Modeled images, ”*IEEE Signal Processing letters*, vol.7, No.11, pp.301-303, 2000.

[22] B. G. Kim, D. J. Kim and D. J. Park, “Novel precision target detection with adaptive thresholding for dynamic image segmentation, ”*Machine Vision and Applications, Springer-Verlag*, vol.12, pp.259-270, 2001.

[23] E. Y. Kim, S. W. Hwang, S. H. Park and H. J. Kim ”Spatiotemporal segmentation using genetic algorithms, ”*Pattern Recognition*, vol.34, no.10, pp.2063-2066,2001

[24] S. W. Hwang, E. Y. Kim, S. H. Park and H. J. Kim ”Object Extraction And Tracking using Genetic Algorithms, ”*Proceedings. of International Conference on Image Processing ,Thessaloniki, Greece*, pp.383-386, 7-10th oct 2001

[25] E. Y. Kim and K. Jung, “Genetic Algorithms for Video Segmentation, ”*Pattern Recognition*, vol.38, No.1, pp.59-73, 2005.

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

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

[28] 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

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

[30] P. M. Jodoin, M. Mignotte and C. Rosenberger, “Segmentation Framework Based on Label Field Fusion ,” *IEEE Transactions on image Processing*, vol.16, No. 10, pp.2535-2550, Oct. 2007.

[31] Q. shi, L. Wang, L. Cheng and A. Smola, “Discriminative Human Segmentation and Recognition using Semi-Markov Model, ”*Proc. of IEEE Computer Society conference on Computer Vision and Pattern Recognition*CVPR-08, 22-28 Jan 2008.

[32] H. Zhou, Y. Yuan, Y. Zhang and C. Shi, “Non-rigid Object Tracking in complex Scenes, ”*Pattern Recognition Letters* 2008.

[33] X. Pan, and Y. Wu, “GSM-MRF based classification approach for real time moving object detection, ”*Journal of Zhejiang University SCIENCE A* vol. 9, No. 2, pp. 250-255, 2008

[34] C. Su, and A. Amer, “A Real Time Adaptive Thresholding for Video Change Detection, ”*Proc. IEEE International Conference in Image Processing 2006* , pp. 157-160, 2006

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

[36] E. R. Davies,*Machine Vision Theory, Algorithm, Practices* 3rd Edition, Elsevier.

[37] C .E. Shanon,*A Mathematical Theory of Communication* Bell System Technology, pp. 379-423, 1948.

[38] A. N. Rajagopalan, P. Burlina and R. Chellappa, ”Detection of people in images,” *Proc. of the IEEE International Joint Conference on Neural Networks* (IJCNN’99), vol.4, pp. 2747-2752, July 1999.

[39] R. Chellappa and A. N. Rajagopalan, ”Tracking humans in video,” *Proc. of the International Conference on Multimedia Processing and Systems* (ICMPS’00), vol.1, pp. 111-114, August 2000.

[40] M. Kamath, U.B. Desai and S. Chaudhuri,”Direct Parametric Object Detection in Tomographic Images,” *Image and Vision Computing*, vol-16, no-10, pp 669-676, 1998.

[41] S. Pal, P. K. Biswas and A. Abraham, ”Face Recognition Using Interpolated Bezier Curve Based Representation,” *Proc. of the International Conference*

on Information Technology: Coding and Computing (ITCC'04),vol. 1, pp. 45-49, 2004.

[42] M. Kokare, B. N. Chatterji, P. K. Biswas, "Cosine-modulated wavelet based texture features for content-based image retrieval,"*Pattern Recognition Letters* vol. 25, No. 4, 391-398, 2004.