

AUTOMATIC FAST VIDEO OBJECT DETECTION AND TRACKING ON VIDEO SURVEILLANCE SYSTEM

V. Arunachalam¹, I. Sorimuthu², V. Rajagopal³ and B. Sankaragomathi⁴

^{1,2}Department of Electronics and Communication Engineering, Einstein College of Engineering, India
E-mail: ¹arunachal004@gmail.com and ²i.sorimuthu1989@gmail.com

³Department of Electronics and Communication Engineering, Sri Vidya College of Engineering and Technology, India
E-mail: vrajgo@gmail.com

⁴Department of Electronics and Instrumentation Engineering, National Engineering College, India
E-mail: sankaragomathi.paramasivan@gmail.com

Abstract

This paper describes the advance techniques for object detection and tracking in video. Most visual surveillance systems start with motion detection. Motion detection methods attempt to locate connected regions of pixels that represent the moving objects within the scene; different approaches include frame-to-frame difference, background subtraction and motion analysis. The motion detection can be achieved by Principle Component Analysis (PCA) and then separate an objects from background using background subtraction. The detected object can be segmented. Segmentation consists of two schemes: one for spatial segmentation and the other for temporal segmentation. Tracking approach can be done in each frame of detected Object. Pixel label problem can be alleviated by the MAP (Maximum a Posteriori) technique.

Keywords:

Background Subtraction, Object Tracking, Principle Component Analysis, Spatio-Temporal Segmentation, MAP

1. INTRODUCTION

Visual surveillance in dynamic scene with multiple cameras, attempts to detect, recognize and track certain objects from image sequences, and more importantly to understand and describe object behaviors. The main goal of visual surveillance is to develop intelligent visual surveillance to replace the traditional passive video surveillance that is proving ineffective as the number of cameras exceed the capability of human operators to monitor them. The goal of visual surveillance is not only to put cameras in the place of human eyes, but also to accomplish the entire surveillance task as automatically as possible. The capability of being able to analyze human movements and their activities from image sequences is crucial for visual surveillance.

The independent component analysis (ICA) particularly aims at indoor surveillance for possible applications in home-care and health-care monitoring, where moving and motionless persons must be reliably detected. An ICA model [3] that directly measures the statistical independency based on the estimations of joint and marginal probability density functions from relative frequency distributions is used. An ICA model can well separate two highly-correlated images. In the detection stage, the trained de-mixing vector is used to separate the foreground in a scene image with respect to the reference background image. Hopfield neural network (HNN) for solving the image change detection problem between two images. A difference image is obtained by subtracting pixel by pixel both images. The network topology is

built so that each pixel in the difference image is a node in the network. The HNN model [1] allows each node to take on analog state values, i.e., to determine the strength of the change. HNN model overcomes this drawback and for each pixel allows one to achieve a tradeoff between the influence of its neighborhood and its own criterion. This is mapped under the energy function to be minimized.

For analog Hopfield networks, the total input into a node is converted into an output value by a sigmoid monotonic activation function instead of the thresholding [2] operation for discrete Hopfield networks. The analog Hopfield's model allows each node to take on analog state values. Unlike most widely used approaches, where binary labels (changed/unchanged) are assigned to each pixel, the analog property provides the strength of the change. The discrete cosine transforms (DCT) coefficients (including ac coefficients) at block level to represent background, and adapt the background by updating DCT coefficients [5]. The segmentation approach can extract foreground objects with pixel accuracy through a two-stage process. First a new background subtraction technique in the DCT domain is exploited to identify the block regions fully or partially occupied by foreground objects, and then pixels from these foreground blocks are further classified in the spatial domain. The mixture of Gaussians (MoG) based approach has obtained tremendous popularity due to its capability to model multimodal backgrounds. Most approaches only exploit dc coefficients of the discrete cosine transformation (DCT) to identify moving regions.

Pfinder ("person finder") that substantially solves the problem for arbitrarily complex but single-erosion, fixed-camera situations. Pfinder has been used as a real-time interface device for information, and performance spaces, video games, and a distributed virtual reality populated by artificial life. It has also been used as a preprocessor for gesture recognition systems, including one that can recognize a 40-word subset of American Sign Language with near perfect accuracy. Pfinder [4] adopts a Maximum a Posteriori Probability (MAP) approach to detection and tracking of the human body using simple 2D models. It incorporates a priori knowledge about people primarily to bootstrap itself and to recover from errors. Pfinder expects the scene to be significantly less dynamic than the user. Although Pfinder has the ability to compensate for small, or gradual changes in the scene or the lighting, it cannot compensate for large, sudden changes in the scene. If such changes occur, they are likely to be mistakenly considered part of the foreground region, and an attempt will be made to explain them in the user model.

A.Cavallaro and T. Ebrahimi [6] have 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, supervised video object segmentation has been [7]. 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 Interpolated Bezier Curve Based Representation scheme [16] is also proposed to recognize the face. An object detection scheme using direct parametric approach in the tomographic images [17] are also proposed Stochastic model [8] particularly Markov Random Field Models, have been extensively used [9]-[10] 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 [15] 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, multi-resolution approach is adhered. Gibbs Markov Random Field Model has been used to obtain 3-D spatiotemporal segmentation [13]. The region growing approach is used to obtain segmentation. E. Y. Kim et al. [12] have used MRF to model each frame sequence and the observed sequence is assumed to be degraded by independent identically distributed (IID) 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). Recently MRF modeling has been used to model the video sequences but the segmentation problem has been formulated using Spatio-temporal framework [11]. 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. [14] 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 [18] 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.

Hence our motion detection gives the better performance and reduced noise than the existing system. The combination of spatio-temporal segmentation also gives the good resolution than the edge, edgeless and JSEG segmentation.

2. METHODOLOGIES

This section deals with the procedure of motion detection from the input video based on PCA algorithm. The detected moving object separated using background subtraction. Moreover combination spatio-temporal segmentation introduced

a color based RGB. Object detection and tracking performed by MAP-MRF combination.

2.1 MOTION DETECTION

Detection of moving objects is usually the first stage of video processing chain and its results are used by further processing modules. Most video segmentation algorithms usually employ spatial and/or temporal information in order to generate binary masks of objects However; simple time-averaging of video frames is insufficient for a surveillance system because of limited adapting capabilities. The solution implemented in the framework utilizes spatial segmentation for detection of moving objects in video sequences, using background subtraction algorithm. This approach is based on modeling pixels as mixtures of Gaussians and using an on-line approximation to update the model. This method proved to be useful in many applications, as it is able to cope with illumination changes and to adapt to the background model accordingly to the changes in the scene, e.g. when motionless foreground objects eventually become a part of the background. Background modeling is used to model current background of the scene and to differentiate foreground pixels of moving objects from the background. Each pixel in the image is modeled with a mixture of K Gaussian distributions for this purpose. The probability that a pixel has the value xt at the time t is given as,

$$p(x_t) = \sum_{i=1}^K w_i^t \eta \left(x_t, \mu_i^t, \Sigma_i^t \right) \quad (1)$$

where, w_i^t denotes the weight and μ_i^t and Σ_i^t are the mean vector and the covariance matrix of i -th distribution at the time t , and d is the normal probability density function,

$$\eta(x_t, \mu, \Sigma) = \frac{1}{(2\pi)^{0.5D} \sqrt{|\Sigma|}} e^{-0.5(x_t - \mu)^T \Sigma^{-1} (x_t - \mu)} \quad (2)$$

where, D is the number of elements describing pixel color; for the RGB color space D is equal to 3. It is assumed that each Gaussian distribution represents a different background color of a pixel. With every new video frame, the parameters w and t of distributions for each pixel is updated according to the on-line K-means Approximation algorithm. In the first step, distributions are ordered based on the value of the r coefficient given as,

$$r = \frac{\omega}{\sqrt{|\Sigma|}}$$

where, $|\Sigma|$ is the determinant of the covariance matrix r . A particular color of the scene background is usually more often present in the observation data than any color of foreground objects and as such is characterized by the low variance. Thus a distribution with a higher r value represents the background color more accurately. Every new pixel value xt is checked against existing distributions, starting from the distribution with the highest value of the r coefficient, until the first match is found. If there is a matching distribution, its mean and variance values are tuned according to the current value of the pixel; the speed of converging is determined by the learning rate α . Only the first D distributions of pixel x in time t ordered by the

decreasing r coefficient value are used as the background model where D is defined as,

$$D'_x = \arg \min_d \left(\sum_{i=1}^d w_i^t > T \right) \quad (3)$$

If T is small, then the background model is usually unimodal. If T is higher, the background color distribution may be multimodal, which could result in more than one color being included in the background model. This make possible to model periodic changes in the background, properly. If the current pixel value does not match any of the first D distributions, it is considered as a part of a foreground object.

2.2 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) is a mathematical procedure that performs a dimensionality reduction by extracting the principal components of the multi-dimensional data. The first principal component is the linear combination of the original dimensions that has the highest variability. The n -th principal component is the linear combination with the maximum variability, being orthogonal to the $n - 1$ first principal component. A principal component can be defined as a linear combination of optimally-weighted observed variables. In order to understand the meaning of this definition, it is necessary to first describe how subject scores on a principal component are computed. In the course of performing a principal component analysis, it is possible to calculate a score for each subject on a given principal component.

It is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Depending on the field of application, it is also named the discrete Karhunen–Loève transform (KLT), the Hotelling transform or proper orthogonal decomposition (POD).

PCA Algorithm

Step 1: Calculate the mean

$$u[m] = \frac{1}{N} \sum_{n=1}^N X[m, n]$$

Step 2: Calculate the deviation

$$\mathbf{B} = \mathbf{X} - \mathbf{u}\mathbf{h}$$

Step 3: Find the Co-variance matrix

$$\mathbf{C} = E[\mathbf{B} \otimes \mathbf{B}] = E[\mathbf{B}\mathbf{B}^*] = \frac{1}{N} \sum \mathbf{B}\mathbf{B}^*$$

Step 4: Find the Eigen values and Eigen vectors of the Co-variance matrix

$$\mathbf{V}^T \mathbf{C} \mathbf{V} = \mathbf{D}$$

Step 5: Dimensionality reduction

2.3 BACKGROUND SUBTRACTION

Background subtraction is the process of separating out foreground objects from the background in a sequence of video frames. Background subtraction is used in many emerging video applications, such as video surveillance, traffic monitoring, and gesture recognition for human-machine interfaces and etc.

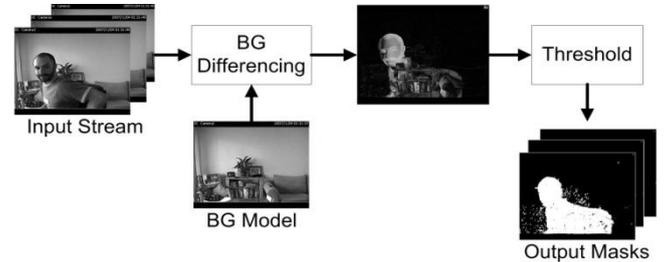


Fig.1. General Framework for Background subtraction

$$I = I_b - I_f$$

where, I_b is background image and I_f is foreground image. Processing a video stream to segment foreground objects from the background is a critical first step in many computer vision applications. The general framework for background subtraction is shown in Fig.1. Background subtraction (BGS) is a commonly used technique for achieving this segmentation. The popularity of BGS largely comes from its computational efficiency, which allows applications such as human-computer interaction, video surveillance, and traffic monitoring to meet their real-time goals.

2.4 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 lead to different segmentation of same data; for example, color segmentation, texture segmentation, and motion segmentation usually result in segmentation maps. Furthermore, there is no guarantee that any of the resulting segmentation will be semantically meaningful, since a 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 [7] 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 be 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 object. 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.

2.5 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. The block diagram for object detection can be illustrated in Fig.2.

2.6 SPATIAL SEGMENTATION

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[13].

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

$$d_{i,j}(x, y) = \begin{cases} 1 & \text{if } |f(x, y, t_i) - f(x, y, t_j)| > T \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

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 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.

The combined scheme applies in a first step the general blocks of camera motion estimation and compensation and scene cut detection which can be seen as a kind of pre-processing in order to eliminate the influence of a moving camera. In a second step, either temporal or combined Spatio-temporal segmentation of each image are carried out, depending on the requirements. The reason for this is that in general only performing temporal segmentation requires less computational complexity.

On the other hand, taking into account also spatial segmentation leads to more accurate segmentation results, but increases the computational complexity of the segmentation. For temporal segmentation, still two possible algorithms are under consideration, both having been verified by extensive cross-checking. It will be one main task for the work on segmentation in the second phase of the MPEG-4 development to decide which of these algorithms performs better. For spatial segmentation, only one algorithm is considered, which however has not been cross checked.

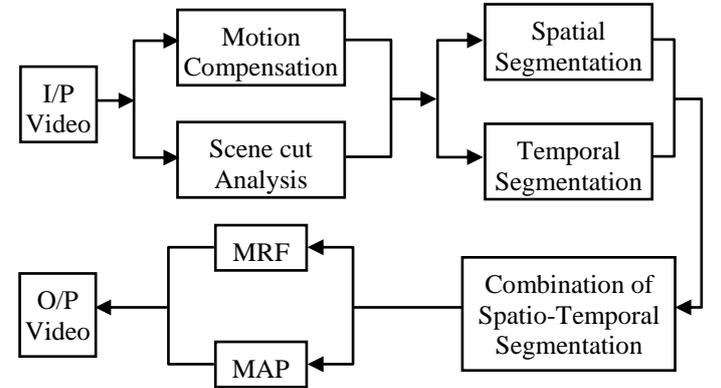


Fig.2. Spatial Temporal Segmentation Description

Finally, if temporal and spatial segmentation is performed, both temporal and spatial segmentation results are combined. Up to now, several approaches have been investigated for this task, however no final algorithm has been decided upon yet. Thus, it will be the second main task of the group to work out an appropriate algorithm for combining the temporal and spatial segmentation results.

2.7 MAXIMUM A POSTERIORI PROBABILITY (MAP)

In Bayesian statistics, a maximum a posteriori probability (MAP) estimate is a mode of the posterior distribution. The MAP can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data. It is closely related to Fisher's method of maximum likelihood (ML), but employs an augmented optimization objective which incorporates a prior distribution over the quantity one wants to estimate. MAP equation can be denoted as,

$$\hat{x} = \arg \min_x \left\{ \left[\frac{\|y - x\|^2}{2\sigma^2} \right] + \left[\sum_{c \in C} [V_{sc}(x) + V_{tec}(x) + V_{tecc}(x)] \right] \right\} \quad (5)$$

where, x is the segmentation of y , y is the observed video sequence, $V_{sc}(xt)$ is the Clique potential function in the spatial domain, $V_{tec}(xt)$ is the Clique potential in the temporal domain, $V_{teec}(xt)$ is the Clique potential in the temporal domain incorporating edge features, σ^2 is the Variance of Gaussian process.

2.8 MARKOV RANDOM FIELD (MRF)

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.

Let $Z = \{Z_1, Z_2, \dots, Z_m\}$ 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* [12]. We use the notion $Z_i = z_i$ to denote the event that Z_i takes the value z_i and the notion $(Z_1 = z_1, Z_2 = z_2, \dots, Z_m = z_m)$ to denote the joint event. For simplicity a joint event is abbreviated as $Z = z$ where, $z = \{z_1, z_2, \dots\}$ 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 $P(Z = z) = P(Z_1 = z_1, Z_2 = z_2, \dots, Z_m = z_m)$ 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,

$$P(Z = z) > 0, \forall z \in Z \text{ (Positivity)} \quad (6)$$

$$P(z_i | z_{S-i}) = P(z_i | z_{N_i}) \text{ (Markovianity)} \quad (7)$$

where, $S-i$ is the set difference, z_{S-i} denotes the set of labels at the sites in $S-i$ and $z_{N_i} = \{z_i | i \in N_i\}$ 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 characteristic of Z . In MRF, only neighboring labels have direct interactions with each other.

Color Based Approach: Unlike many other image features (e.g. shape) color is relatively constant under viewpoint changes and it is easy to be acquired. Although color is not always appropriate as the sole means of detecting and tracking objects, but the low computational cost of the algorithms proposed makes color a desirable feature to exploit when appropriate. An algorithm to detect and track vehicles or pedestrians in real-time using color histogram based technique. They created a Gaussian Mixture Model to describe the color distribution within the sequence of images and to segment the image into background and objects. Object occlusion was handled using an occlusion buffer. Tracking of multiple faces in real time at full frame size and rate using color cues. This simple tracking method is based on tracking regions of similar normalized color from frame to frame. These regions are defined within the extent of the object to be tracked with fixed size and relative positions. Each region is characterized by a color vector computed by sub-sampling the pixels within the region, which represents the averaged color of pixels within this region. They even achieved some degree of robustness to occlusion by explicitly modeling the occlusion process.

3. RESULTS AND DISCUSSION

Motion Detection: Motion detection is a process of confirming a change in position of an object relative to its surroundings or the change in the surroundings relative to an object. Results of moving object detection in the continuous frames are shown in Fig.3.

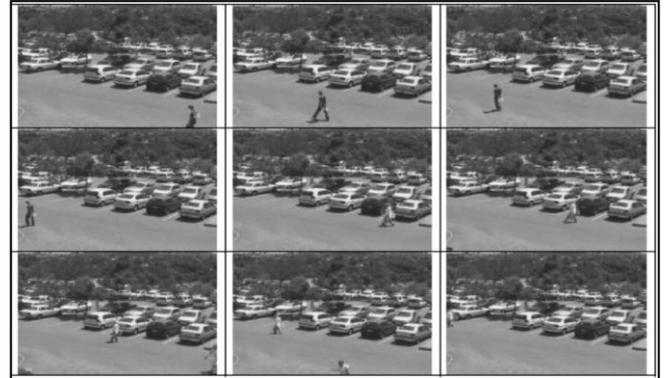


Fig.3. Results of moving object detection in the continuous frames

Background Subtraction: Background subtraction is the process of separating out foreground objects from the background in a sequence of video frames. The results of the background subtraction are shown in Fig.4.



Fig.4(a)

Fig.4(b)

Fig.4(c)

Fig.4(a) Background image (b) Foreground image
(c) Background Subtracted image

The first one is when the original frame (Fig.5(a)) is considered and the second one is when the estimated label frames (Fig.5(b)) are considered. In all the cases we have considered RGB color model. The Spatio-Temporal segmentation together with the temporal segmentation is used to detect the video objects. Fig.6(a), 6(b) and 6(c) represent the tracking frames at time T_1 , T_2 and T_3 respectively.



Fig.5(a) Input Sequence

Fig.5(b) Segmented Frame



Fig.6(a)

Fig.6(b)

Fig.6(c)

4. CONCLUSION

Principal Component Analysis (PCA) is suitable for any type of video while the existing methods only suitable for compressed or ordinary videos. The noise and execution time is less compared to other methods. The spatio-temporal spatial segmentation result of the initial frame is obtained by edge based MRF modeling and a MAP estimation algorithm. The segmentation result of the initial frame together with some change information from other frames is used to generate an initialization for segmentation of other frames. It is found that the proposed approach produces better segmentation results compared to edgeless and JSEG segmentation schemes and comparable results with edge based approach. The scheme also gives better accuracy and is used for live stream purpose to the considered MRF based segmentation schemes for a number of video sequences.

ACKNOWLEDGEMENT

My deepest thanks to Prof. V. Rajagopal, Dr. Sankara Gomathi and Ms. P.R. Gayathri Devi for guiding and correcting various documents of mine with attention and care.

REFERENCES

- [1] R.J. Radke, S. Andra, O. Al-Kofahi and B. Roysam, "Image change detection algorithms: A systematic survey", *IEEE Transactions on Image Processing*, Vol. 14, No. 3, pp. 294–307, 2005.
- [2] T. Ridler and S. Calvard, "Picture thresholding using an iterative selection method", *IEEE Transactions on Systems, Man, Cybernetics*, Vol. 8, No. 8, pp. 630–632, 1978.
- [3] Du-Ming Tsai and Shia-Chih Lai, "Independent Component Analysis-Based Background Subtraction for Indoor Surveillance", *IEEE Transactions on Image Processing*, Vol. 18, No. 1, pp. 158–167, 2009.
- [4] C.R. Wren, A. Azarbayehani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 780–785, 1997.
- [5] N.M. Oliver, B. Rosario, and A.P. Pentland, "A Bayesian computer vision system for modeling human interactions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 831–843, 2000.
- [6] A. Cavallaro and T. Ebrahimi, "Change Detection based on Color Edges", *IEEE International Symposium on Circuits and Systems*, Vol. 2, pp. 141–144, 2001.
- [7] 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, 2003.
- [8] A. Papoulis and S. U. Pillai, "Probability, Random Variables and Stochastic Process", Tata McGraw Hill Education, 2002.
- [9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, pp. 721–741, 1984.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] G.K. Wu and T.R. Reed, "Image sequence processing using spatiotemporal segmentation", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 9, No. 5, pp. 798–807, 1999.
- [14] 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*, Vol. 2, pp. 383–386, 2001.
- [15] R.O. Hinds, and T.N. Pappas, "An Adaptive Clustering algorithm for Segmentation of Video Sequences", *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, Vol. 4, pp. 2427–2430, 1995.
- [16] S. Pal, P.K. Biswas and A. Abraham, "Face Recognition Using Interpolated Bezier Curve Based Representation", *Proceedings of the International Conference on Information Technology: Coding and Computing*, Vol. 1, pp. 45–49, 2004.
- [17] 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.
- [18] E.Y. Kim and S.H. Park, "Automatic video Segmentation using genetic algorithms", *Pattern Recognition Letters*, Vol. 27, No. 11, pp. 1252–1265, 2006.