SIGN LANGUAGE VIDEO SEGMENTATION WITH LEVEL SETS FUSING COLOR, TEXTURE, BOUNDARY AND SHAPE FEATURES

P.V.V.Kishore[1] and P.Rajesh Kumar[2]

[1,2]Department of Electronics and Communications Engineering, Andhra University,
Visakhapatnam City, Andhra Pradesh, INDIA
pvvkishore@gmail.com
rajeshauce@gmail.com

ABSTRACT

This paper presents a new and improved concept for segmenting gestures of sign language. The algorithm presented extracts signs from video sequences under various non static backgrounds. The signs are segmented which are normally hands and head of the signing person by minimizing the energy function of the level set fused by various image characteristics such as colour, texture, boundary and shape information. From RGB color video three color planes are extracted and one color plane is used based on the contrasting environments presented by the video background. Texture edge map provides spatial information which makes the color features more distinctive for video segmentation. The boundary features are extracted by forming image edge map form the existing color and texture features. The shape of the sign is calculated dynamically and is made adaptive to each video frame for segmentation of occlude objects. The energy minimization is achieved using level sets. Experiments show that our approach provides excellent segmentation on signer videos for different signs under robust environments such as diverse backgrounds, sundry illumination and different signers.

KEYWORDS

Sign Language, Video Segmentation, Color/Texture extraction, Boundary Information, Shape Extraction, Level Sets

1. INTRODUCTION

Sign language is the basic mode of expression and communication for deaf people. Sign language involves hand shapes, hand tracking, hand orientation with respect head and other body parts, along with head movements and facial expressions. The pointed aim of sign language recognition system is to correctly reproduce verbal or text in accordance with the given sign. To realize such a system in real time, various attributes of the signer has to be considered such as hands and head movement, facial expressions and body pose. Stokoe et.al. [1] showed that signs are made of basic articulacy units referred to as phonemes in similarity to that of words in spoken language.

The primary challenge faced by any sign language recognition system is the ability to track the signer in the video of the signer with a variety of background clutter. The backgrounds used so far by many researchers are simple. We also had our system developed for simple backgrounds [2]. But to make the system work under complex backgrounds, different lightening conditions, indoors, outdoors and signer independent researchers are looking for new models and algorithms to solve these problems.

DOI : 10.5121/sipij.2012.3307

This research paper addresses the problem of segmenting hands and head along with facial features required to understand a particular sign for the video of the signer under various backgrounds. This problem is attacked by many researchers from around the globe by various algorithms such as background subtraction [3], skin color based segmentation [4] and motion based segmentation [6].

In background subtraction [3] various algorithms have been proposed such as frame differencing, adaptive mean filtering, adaptive median filtering and Gaussian Mixture Models (GMM). Of all the proposed algorithms GMM gives good segmentation levels for all kinds of complex video backgrounds. But the main factor in GMM is the speed of operation. GMM is slow compared to other methods providing a good segmentation result. GMM estimates the probabilities of foreground objects and background in to two different classes on each pixel and based on the maximum likelihood estimation it extracts the object class. Qing-song Zhu et.al. proposed for video object segmentation by dynamically constructing background models of Gaussian mixtures and segmenting foreground pixels through background subtraction [3] . The algorithm works well for surveillance videos but cannot show promising results for sign videos.

Wen-kai Tsai et.al showed the human skin color can be adaptively used for segmenting humans from a series of cluttered backgrounds [4]. The proposed method is based on gesture skin color that is obtained from least squares approximation solutions and Gaussian distribution model. This model has improvement in the segmentation of complex backgrounds, shadow elimination and light-reflection effects. Qiuyu Zhang et.al. proposed dynamic background updation algorithm which is based on difference background image between consecutive video frames using ‘3σ Principle’ of normal distribution of hand gestures detection to cope with the problem [5]. The algorithm showed better results for complex background.

In [6,7,8] temporal segmentation is performed on continuous video sequences using motion trajectories obtained by fitting respective points of motion region to the curve. This method assumes that these points submit to a certain distribution limiting the application of this algorithm to long video sequences. The algorithms discussed are not robust when applied to sign language videos. To effectively segment sign language videos and extract hand and head gestures we propose active contours based video segmentation.

Active contours or popularly known in the research community as ‘snakes’ is a active research area with applications to image and video segmentation predominantly to locate object boundaries. They are also used for video object tracking applications. Active contours come under the category of model based segmentation methods giving good results in the last few years [9,10,11]. The active contours was first introduced by Terzopoulos [12,13]. The basic idea behind active contours is to start with a curve anywhere in the image and move the curve in such way that it sticks to the boundaries of the objects in the image, thus separating the background of the image from its objects. The original snakes algorithm was prone to topological disturbances and is exceedingly susceptible to initial conditions. However with invention of level sets [14] topological changes in the objects are automatically handled. Nevertheless all active contours are depending on gradient of the image for end the growth of the curve. Chan and Vese (CV Model) [15,16,17] proposed a new level sets method based on Mumford-Shah distance for image segmentation. CV Model for level sets does not necessarily consider gradient for stopping the curve evolution.

In general the object segmentation using active contours based level sets face a few challenges considering the videos they are applied on. Firstly the contours get easily distracted if the background in the video sequence contains clutters which the case of sign language videos under real time environments. Second problem is when working with intensity images it becomes difficult to locate true boundaries of objects under varied lighting and the objects mostly blends with the background of the image. This problem can be solved to an extent if texture of the object is included as a feature for the contour to fit the boundaries of the object. Finally video segmentation gets seriously affected due to occlusions. The foreground objects are partially or completely lost during occlusions. These three challenges hamper the performance of a good sign language recognition system during segmentation phase.

This paper addresses the above discussed challenges and solves them to achieve excellent video object segmentation for different sign language videos. The active contours method used effectively manages segmentation of multiple moving objects on static and non static cluttered backgrounds along with intra-object occlusions. Our proposed method brings together multiple characteristics of video image to segment the objects. These characteristics comprise color and texture information of objects, image boundary edge map and the shape of the object form the prior to segmentation of the current frame from the previous frame.

The segmentation is devised by minimizing a force function which is a combination of color, texture, boundary and prior shape information of the objects to find their boundaries in all the video frames. The color and texture information is formulated by separating out foreground objects from background by minimizing the distance between them. The boundary information is calculated by using a gradient operator, which enables the contour to align itself to the edges of objects in the image. The prior shape information is obtained from the properties of level set contour, which has the ability to identify objects under occlusion. We have successfully applied the method to sign language video sequences under different situations for segmenting hands and head portions.

The rest of the paper is organized as follows: sect. 2, we present introduction on active contours, sect. 3, proposed method for multiple object segmentation, sect. 4, we discuss the justification of our proposed approach to sign language segmentation under various test conditions, sect.5 we provide a brief conclusion and discuss the future prospects of our proposed method.

2. Active Contours -Theoretical Background

The active contours are elastic models of continuous, flexible curve that is imposed upon and matched to the image by varying the elastic parameters. The fundamental idea is to make the curve or snake to fit tightly to the boundary of a particular object in the image. The design of evolution equation is such that the snake can easily embrace the object of importance, to be able to develop a similarity. The first snake model was proposed by kass[18]. The minimization energy function in order to achieve equilibrium is

$$\mathfrak{E}^{Snake} = \int_0^1 \left\{ \mathfrak{E}_{int}(\boldsymbol{v}(\mathcal{s})) + \mathfrak{E}_{image}(\boldsymbol{v}(\mathcal{s})) \right\} d\mathcal{s} \qquad [1]$$

Where the location of the snake on the image is represented parametrically by a planer curve

$$\boldsymbol{v}(\mathcal{s}) = (\mathcal{x}(\mathcal{s}), \mathcal{y}(\mathcal{s})) \qquad [2]$$

And $\mathfrak{E}_{int}$ represents the internal energy of the curve due to bending and $\mathfrak{E}_{image}$ represents the image forces that push the snake towards the desired object.

The internal energy model was defined as

$$\mathfrak{E}_{int} = \frac{(\alpha(s)|\boldsymbol{v}_{\mathcal{s}}(\mathcal{s})|^2 + \beta(\mathcal{s})|\boldsymbol{v}_{\mathcal{s}\mathcal{s}}(\mathcal{s})|^2)}{2}, \quad \mathcal{s} \in [0,1] \qquad [3]$$

Where $\boldsymbol{v}_{\mathscr{s}}(\mathscr{s})$ First derivative of $\boldsymbol{v}(\mathscr{s})$ and $\boldsymbol{v}_{\mathscr{s}\mathscr{s}}(\mathscr{s})$ is Second order derivative o f $\boldsymbol{v}(\mathscr{s})$ with respect to $\mathscr{s}$. The model of image energy is defined as

$$\mathfrak{E}_{image} = -|\nabla I(x,y)|^2 \qquad [4]$$

The first derivative of$v(\boldsymbol{\mathscr{s}})$ with respect to'$\mathscr{s}$' gives us rate of change of length of the curve. The coefficient $\alpha(\boldsymbol{\mathscr{s}})$ allows the curve to have smaller or larger degree of contraction of the curve and therefore makes the snake act like an elastic string. The second derivative of $v(\boldsymbol{s})$ with respect to'$\mathscr{s}$' gives us rate of change of curvature. The coefficient $\beta(\boldsymbol{\mathscr{s}})$ regulates the rate of the change of the curve in the direction normal to the boundary, preserving the smoothness of the curve. By adjusting these two coefficients, the curve gets an appropriate elasticity and is able to embrace the object of interest.

2.1. Global Region Based Segmentation- The Chan-Vese Model

In Chan-Vese [19] Image segmentation model, consider a gray scale image as real valued function of space $\boldsymbol{\mathcal{J}}: \boldsymbol{\mathcal{S}} \to \mathbb{R}$ defined on image space $\boldsymbol{\mathcal{S}} \to \mathbb{R}^2$, where $\mathbb{R}$ is a set of real numbers. A point in the image $(x, y) \in \boldsymbol{\mathcal{S}}$ is termed as pixel and the function value $\mathcal{J}(x) = \mathcal{J}$ as pixel value. The basic idea of chan-vese active contour model is to find a contour $\mho: \boldsymbol{\mathcal{S}} \to \mathbb{R}^2$, that optimally approximate the image $\boldsymbol{\mathcal{J}}$ to a single real gray value $\Phi_{internal}$ on the inside of the contour $\mho$**,** and the outside of the contour $\mho$**,** by another gray value $\Phi_{external}$. The solution for the above problem comes in the form of finding optimal contour $\widetilde{\mho}$ and a pair of optimal gray scale values $\widetilde{\Phi} = (\widetilde{\Phi}_{internal}, \widetilde{\Phi}_{external})$, which is formulated mathematically as

$$\boldsymbol{\mathfrak{E}^{CV}(\widetilde{\mho}, \widetilde{\Phi}) = \min_{\mho, \Phi} \mathfrak{E}^{CV}(\mho, \Phi)} \quad [5]$$

where $\mathfrak{E}^{CV}$ is the energy function defined by chan-vese model as an analogous to piece wise linear Mumford-Shah [16] model which approximates the gray scale image $\mathcal{J}(x)$ by a piecewise smooth function $\mho$ as a solution to the minimization problem

$$\boldsymbol{\mathfrak{E}^{CV} = \lambda_1 \int_{\mho} d\mathscr{s} + \lambda_2 \left[\frac{1}{2}\int_{int(\mho)} (\mathcal{J}(x) - \Phi_{internal})^2\, dx + \frac{1}{2}\int_{ext(\mho)} (\mathcal{J}(x) - \Phi_{external})^2\, dx\right]} \quad [6]$$

The first term in the eq.6 indicates arc length $arg\min_{\mho,\Phi} \lambda_1 \times Length(\mho)$ which guarantee evenness of $\mho$**.** The second term has two integrals. The first integral function pushes the contour $\mho$ towards the image $\mathcal{J}$ while the second integral function ensures the differentiability on the contour $\mho$**.** The Mumford-Shah considers the edge map of the image as the boundary. The weight parameters $\lambda_1\, and\, \lambda_2 > 0$. Solution for eq.6 is a complicated one. Hence a more simpler piecewise constant formulation of Mumford-Shah distance function is

$$\boldsymbol{\mathfrak{E}^{CV} = \lambda_1 \int_{\mho} d\mathscr{s} + \lambda_2 \frac{1}{2}\int_{int(\mho)} (\mathcal{J}(x) - \mho(x))^2\, dxdy} \qquad [7]$$

Compared to Mumford-Shah model, Chen-Vese Model consists of an additional term imprisoning the area enclosed and a further simplification , Φ is allowed to have two values corresponding to the mean values of the pixels inside and outside $\mho$,

$$\mho(\boldsymbol{x}) = \begin{cases} \Phi_{internal}, Where\ x\ is\ inside\ \mho = \frac{\mathbf{1}}{|\boldsymbol{int}(\mho)|}\int_{\mathbf{int}(\mho)} \boldsymbol{\mathcal{J}}(\boldsymbol{x})\boldsymbol{dx} \\ \Phi_{external}, Where\ x\ is\ outside\ \mho\ = \frac{\mathbf{1}}{|\boldsymbol{ext}(\mho)|}\int_{\mathbf{ext}(\mho)} \boldsymbol{\mathcal{J}}(\boldsymbol{x})\boldsymbol{dx} \end{cases} \qquad [8]$$

CV model calculates and finds the values of $\mho$ **,** that is best fits the image $\boldsymbol{\mathcal{J}}$ using the energy term in eq. 8.

$$\mathfrak{E}^{\mathbf{CV}} = \boldsymbol{\lambda}_{\mathbf{1}}\int_{\mho} \mathbf{d}\boldsymbol{s} + \boldsymbol{\upsilon}\int_{\mathbf{int}(\mho)} \mho(\mathbf{x})\mathbf{dx} + \boldsymbol{\lambda}_{\mathbf{2}}\left[\int_{\mathbf{int}(\mho)} |\mathcal{J}(\mathbf{x}) - \boldsymbol{\Phi}_{\mathbf{internal}}|^{\mathbf{2}}\,\mathbf{dx} + \int_{\mathbf{ext}(\mho)} |\mathcal{J}(\mathbf{x}) - \boldsymbol{\Phi}_{\mathbf{external}}|^{\mathbf{2}}\,\mathbf{dx}\right] \qquad [9]$$

The first two terms are regularizing parameters for contours length and its area to control the size of the contour. The third and fourth terms make the model $\mho(x)$ adapt to the objects in the image $\boldsymbol{\mathcal{J}}(\mathbf{x})$**.** Image segmentation deals with finding the global minimum to the problem defined in eq.[9].

2.2. The Level Set Model

James A. Sethian and Stanley Osher [14] represented boundaries of $\mho(x)$ implicitly and model their propagation using appropriate partial differential equations. The boundary is given by level sets of a function $\phi(x)$. In level sets method, the interface boundary is characterize by a zero level set function $\phi(x) = 0, where\ \phi: \mathbb{R}^{\mathbf{2}} \to \mathbb{R}$. $\mho$ is defined for all values of x,

$$\mho = \{\phi(x) = 0, x \in \mathbb{R}^2\} \qquad [10]$$

The sign of $\phi(x)$ defines whether x is inside the contour $\mho$ or external to it. The sets $\mho^{Int} = \{x, \phi(x) \le 0\}$ and $\mho^{Ext} = \{x, \phi(x) > 0\}$. The level set evolves based on the curvature κ of the image objects and assuming the curve moves towards the outward normal $\vec{\boldsymbol{n}}$ defined in terms of parameter ϕ as

$$\boldsymbol{\kappa} = \nabla.\left[\frac{\nabla\phi}{|\nabla\phi|}\right]\ \ and\ \ \vec{\boldsymbol{n}} = \frac{\nabla\phi}{|\nabla\phi|} \qquad [11]$$

Usually the curve $\mho$ evolution is a time dependent process and the time dependent level set function $\phi: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ as $\mho(\mathrm{t}) = \{\phi(x,t) = 0, x \in \mathbb{R}^2\}$**.** One way to solve is to approximate spatial derivatives of motion and update the position of the curve over time. This method of solving the level sets is prone to unsteadiness due to erroneousness detection of position of the curve.

A different approach was proposed from the theory of level sets in [11]. Start with a zero level set $\phi(x) = 0$ of higher dimension function and entrench the object curvature. Initializing the level set function ϕ at $t = 0$, we have

$$\phi(x, t = 0) = \pm d \qquad [12]$$

Where $'d'$ is signed distance function (sdf) from x to the curvature of the image object. If d is a positive value x is outside the object boundary and if d is a negative value x is inside the object boundary. The goal is to construct an equation for evolution of $\phi(x,t)$ to embrace the object boundaries from zero level set $\phi(x) = 0$.

We can propagate the zero level set $\phi(x) = 0$, by solving a convection equation containing the velocity field v, which propagates all the level sets as

$$\phi_t + v.\nabla\phi = 0 \qquad [13]$$

The motion is normal velocity of the curve which is given by eq.11 as $v = F\vec{\boldsymbol{n}} = F\frac{\nabla\phi}{|\nabla\phi|}$. Inserting in eq.13 we have a level set equation of the form

$$\phi_t + F|\nabla\phi| = 0 \qquad [14]$$

Eq.14 is a type of Hamilton-Jacobi equation. The speed term F is dependent of object curvature $\boldsymbol{\kappa} = \nabla.\left[\frac{\nabla\phi}{|\nabla\phi|}\right]$, which can be formulated as

$$F(\boldsymbol{\kappa}) = F_0 + F_1(F) \qquad [15]$$

Eq.15 drives the contour to level out with the high curvature regions together with a diffusion term.

3. Sign Video Segmentation Module

This section presents the video image sequence segmentation proposed to extract hands and head segments of the signer form a variety of video backgrounds under different lighting conditions with diverse signers. A video sequence is defined as a sequence of image frames $\boldsymbol{\mathcal{I}(x,y,t)}: \mathfrak{D} \rightarrow \mathbb{R}$, where the images change over time. Alternatively a succession of image frames can be represented as $\boldsymbol{\mathcal{I}^{(n)}}$ where $0 \leq n \leq \infty$. The basic principle behind our proposed segmentation technique is to localize the segmentation of one or more moving objects of the n^{th} frame from the cues available from previous segmented frames $\boldsymbol{\mathcal{I}^{(1)}, \mathcal{I}^{(2)}} \ldots\ldots\ldots.. \boldsymbol{\mathcal{I}^{(N)}}$ such that subsequent contours $\boldsymbol{\mho^1, \mho^2} \ldots\ldots\ldots\ldots \boldsymbol{\mho^N}$ are available. The sign videos are composed of many moving objects along with the hands and head of the signer. We considered signers hands and hand as image foreground denoted by $\boldsymbol{\mathcal{I}_f^{(n)}}$.and rest of the objects as the image background $\boldsymbol{\mathcal{I}_b^{(n)}}$ for the image $\boldsymbol{\mathcal{I}^{(n)}}$ in the video sequence. We may denote foreground contour of the hands and head by $\boldsymbol{\mho_f^{(n)}}$. Our proposed video segmentation algorithm segments hands and head of signers using color, texture, boundary and shape information about the signer given precedent understanding of hand and head shapes from $\boldsymbol{\mathcal{I}_f^{(n-1)}}$ and $\boldsymbol{\mathcal{I}_b^{(n-1)}}$. The outline of the algorithm in the form of a block diagram is shown in figure1.

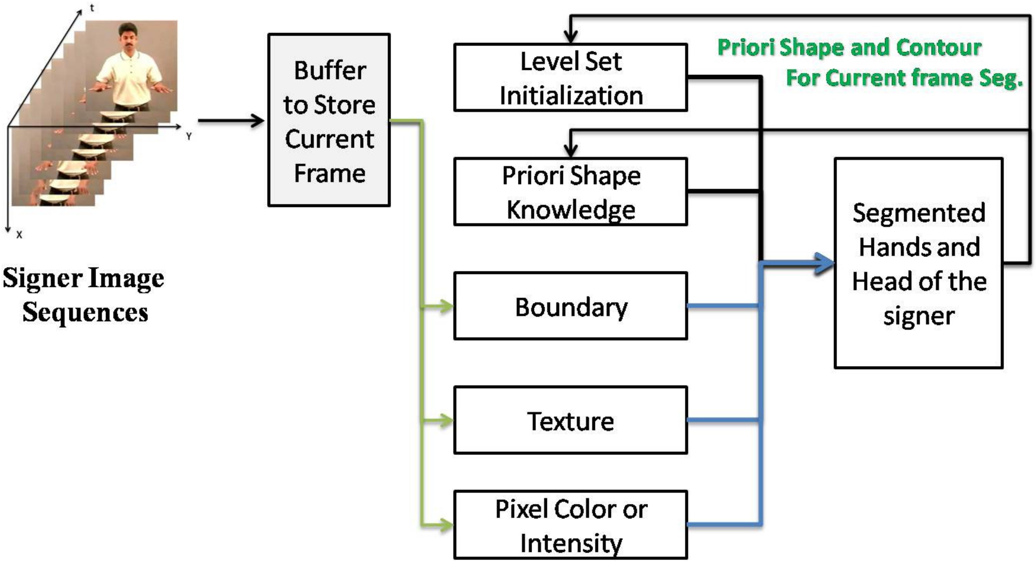
t
Y
x
Signer Image Sequences
Buffer to Store Current Frame
Level Set Initialization
Priori Shape and Contour For Current frame Seg.
Priori Shape Knowledge
Segmented Hands and Head of the signer
Boundary
Texture
Pixel Color or Intensity

Figure 1: Process flow block representation of the segmentation algorithm

3.1. Color and Texture Features Module

Color plays a vital role in segmentation of complex images easily. There are various color models, but the RGB color model is most common in video acquisition. Processing on a RGB color frame increases the size of feature vector and thereby making the segmentation process sluggish. Instead of working with gray scale images which store intensity information about each pixel, we used each of the three color planes separately to extract each color feature vector. This allows us to work with only one plane at a time depending on the background color level. We choose manually the color plane which highlights the human object from a background of clutter. We also experimented with all three color planes which will be discussed in the results section. Once a color plane is identified, texture features are calculated using coorelogram of pixel neighbourhood [20, 21]. Texture is an irregular distribution of pixel intensities in an image. Allam.et.al [22] established that co-occurrence matrix (CM's) produce better texture classification results than other methods. Gray Co-occurrence matrix (GLCM) presented by Haralick.et.al [23] is most effectively used algorithm for texture feature extraction for image segmentation.

Let us consider a color plane of our original RGB video. The R color plane is now considered as a $M \times N$ R coded 2D image. The element of co-occurrence matrix $C_{d,\theta}$ defines the joint probability of a pixel x_i of R color intensity $\boldsymbol{r_i}$ at a distance d and orientation θ to another pixel x_j at R color intensity $\boldsymbol{r_j}$.

$$C_{d,\theta}(r_i, r_j) = Pr\{I(z_1) = r_i \wedge I(z_2) = r_j : |z_1 - z_2|_\theta = d\} \qquad [16]$$

where $|z_1 - z_2|_\theta$ gives the distance between pixels. For each co-occurrence matrix, we calculate four statistical properties: contrast (C), correlation (CO), energy (EN) and homogeneity (H) defined as follows

$$C = \sum_{i,j} |i - j|^2 \, C_{d,\theta}(r_i, r_j) \qquad [17]$$

$$CO = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} C_{d,\theta}(r_i, r_j) \qquad [18]$$

$$EN = \sum_{i,j} \left(C_{d,\theta}(r_i, r_j) \right)^2 \qquad [19]$$

$$H = \sum_{i,j} \frac{C_{d,\theta}(r_i, r_j)}{1 + |i-j|} \qquad [20]$$

The sense in which the above parameters are used as texture feature is described as follows. The contrast represents inertia and variance of the texture. The correlation term gives correlation between different elements of GLCM. CO is high for more complex textures. From eq.18 $\mu_i \; and \; \mu_j$ mean values along $i \; and \; j$ directions.$\sigma_i \; and \; \sigma_j$ represents variances. Energy term describes the uniformity of texture. Homogeneity is taken as a measure of coarsenesses of the texture. We used four different orientations $\theta = \{0,45,90,135\}$ and two distance measures $d = (1, -1)$ for calculation of GLCM.

Finally a feature vector $f^{vect}(\boldsymbol{x})$ is produced which is a combination of any one or all of the color planes and texture vector. Thus $f^{\boldsymbol{vect}}(\boldsymbol{x}) = \{f_1(\boldsymbol{x}), f_2(\boldsymbol{x}) \ldots \ldots \ldots . f_n(\boldsymbol{x})\}$ the feature vector contains color and texture values of each pixel in the image. This is a five dimension feature vector containing the first vector for any of the three color planes and the next four vectors for texture. We can also use all three color planes to represent color, and then the feature vector becomes a seven dimension feature vector.

Most of the image sequences contain many classes of color and texture. Hence we classify them as background and foreground pixels using K-Means clustering algorithm. Given n −dimensional feature vector $f^{vect}(\boldsymbol{x})$ and the K-Means algorithm classify this n −vector into k −catogiries. The centroids $\boldsymbol{S_c}$ of each group are used to identify each of the k −clusters where $c = 1,2 \ldots N, for \; all \; N > 1$. For every new classification the difference between the new vector and all the centroids is computed. The centroid corresponding to smallest distance is judged as the vector that belongs to the group.

$$d = \min_c |f^{vect}(\boldsymbol{x}) - \boldsymbol{S_c}| \qquad [21]$$

d is the distance of every new $f^{vect}(\boldsymbol{x})$ of each frame to the previously computed centroids. All the pixels are classified and average of all pixel values in each cluster is calculated. The centroids are replaced by new average and pixels are classified again until all cluster centers become rigid. In the first frame $\boldsymbol{\mathcal{J}}^{(\mathbf{0})}$ all objects and background clusters are created. The object region contains three clusters of foreground $\boldsymbol{C}_{\boldsymbol{f_i}}^{(\mathbf{0})}, where \; i = 1 \; to \; 3$ and background region $\boldsymbol{C}_{\boldsymbol{b_j}}^{(\mathbf{0})}, where \; j = 1,2$ into two clusters**.** We will assume at this point is that objects in the video sequence pretty much remain same compared to background that varies due to camera movement or changes in background scenes. This can be taken care by periodically updating the background clusters with some threshold if the changes in consecutive background frame cross the specified threshold.

To move the contour on to the objects of interest we minimize the following energy function$\mathfrak{E}^{\mathbf{CT}}$ from color and texture according to the initial object contour $\mathfrak{U}^{\mathbf{obj}}$

$$\mathfrak{E}^{\mathbf{CT}}\left(\mho^{\mathbf{obj}}\right) = \sum_{\mathbf{i=1}}^{\mathbf{3}}\int_{\mathbf{obj}} \mathbf{D}\left(\boldsymbol{C}_{\boldsymbol{f}_i}^{(\boldsymbol{n}-\mathbf{1})}, \boldsymbol{C}_{\boldsymbol{f}_i}^{(\boldsymbol{n})}\right)\boldsymbol{dx} + \sum_{\mathbf{j=1}}^{\mathbf{2}}\int_{\mathbf{back}} \mathbf{D}\left(\boldsymbol{C}_{\boldsymbol{b}_j}^{(\boldsymbol{n}-\mathbf{1})}, \boldsymbol{C}_{\boldsymbol{b}_j}^{(\boldsymbol{n})}\right)\boldsymbol{dx} \qquad [22]$$

where $\boldsymbol{C}_{\boldsymbol{f}_i}^{(\boldsymbol{n}-\mathbf{1})}$ and $\boldsymbol{C}_{\boldsymbol{b}_j}^{(\boldsymbol{n}-\mathbf{1})}$ are object and background centroids of previous frame. $\boldsymbol{C}_{\boldsymbol{f}_i}^{(\boldsymbol{n})}$ and $\boldsymbol{C}_{\boldsymbol{b}_j}^{(\boldsymbol{n})}$ are object and background clusters from current frame. The $(n-1)$ frame cluster centroids will become the n^{th} frame initial centroid and the object contour is moved by minimizing the Euclidean distance between the two centroids. We can implement this by assigning pixel $\boldsymbol{x}$ to object in the current frame when $\mathbf{D}\left(\boldsymbol{C}_{\boldsymbol{f}_i}^{(\boldsymbol{n}-\mathbf{1})}, \boldsymbol{C}_{\boldsymbol{f}_i}^{(\boldsymbol{n})}\right) > \mathbf{D}\left(\boldsymbol{C}_{\boldsymbol{b}_j}^{(\boldsymbol{n}-\mathbf{1})}, \boldsymbol{C}_{\boldsymbol{b}_j}^{(\boldsymbol{n})}\right)$ and to the background otherwise.

3.2. Object Boundary Module

In the earlier module the focus was on extracting region information with the objective of minimizing the object contour energy function, which then segments the objects of interest in the video sequence. But poor lightning can impact image region information in a big way. Hence we use boundary edge map of the image objects which only depends on image derivatives. The way out would be to couple the region information in the form of color and texture features to boundary edge map to create a good segmentation of image objects.

We define the boundary $\boldsymbol{\mathcal{B}}^{\boldsymbol{obj}}(\boldsymbol{x})$ as pixels that are present in edge map of the object. The boundary pixels can be calculated by using gradient operator on the image. To align the initial contour $\mho^{\mathbf{obj}}$ from previous frame to the objects in the current frame to pixels on the boundary we propose to minimize the following energy function

$$\mathfrak{E}^{\mathbf{B}}\left(\mho^{\mathbf{obj}}\right) = \int_{\mathbf{arc}\left(\mathbf{L}^{\mathbf{obj}}\right)} \mathbf{g}\left(\boldsymbol{\mathcal{B}}^{\boldsymbol{obj}}(\boldsymbol{x})\right)\boldsymbol{dx} \qquad [23]$$

where $\mathbf{arc}\left(\mathbf{L}^{\mathbf{obj}}\right)$is the length of the object boundary. The function g is an edge detection function. The boundary energy reaches to a minimum when the initial contour aligns itself with the boundary of the objects in the image. The minimization of energy in eq.23 also results in a smooth curve during evolution of the contour [24].

3.3. Shape Influence Module

Even with color, texture and boundary values of pixels in the image, the greatest challenge comes when object pixels and background pixels share the same color and texture information. This happens because we are trying to segment non rigid objects that are hands of the signer along with finger positions and orientations which change frequently in sign video. The problem will influence the propagation of contour and results in meagre segmentation of video sequences. The contour can be influenced by giving information regarding the shape of the object computed from the previous frames.

The following method in [25, 26] is used to construct the influence of shape of non-rigid objects in the image sequence. As for the first fame $\boldsymbol{\mathcal{J}}^{(\mathbf{0})}$ where prior shape information is not available we just use the region and boundary information for segmentation. For $\boldsymbol{\mathcal{J}}^{(\boldsymbol{n})}\ \forall\, \boldsymbol{n} \geq \mathbf{1}$, the segmentation of $\boldsymbol{\mathcal{J}}^{(\boldsymbol{n})}$ is given by the level set contour $\mho^{\boldsymbol{n}}$ which minimizes the energy function

$$\mathfrak{E}^{\mathbf{S}} = \min_{\mathbf{T}} \int_{\mathbf{int}(\mho)} \boldsymbol{\phi}_{\mathbf{0}}\left(\boldsymbol{T}^{-\mathbf{1}}\boldsymbol{x}\right)\boldsymbol{dx} \qquad [24]$$

where the minimum is calculated over Euclidian Similarity Transformations $T: \mathbb{R}^2 \to \mathbb{R}^2$ which is a combination of translational and rotational parameters. Minimizing over groups of transformations to achieve rigid object interactions was proposed by chan and zhu [25]. We propose to use from [27] a non-rigid shape influence term in this paper. Now let us recollect $\mho$ indicate the active contour and $\mho^{\mathbf{0}}$ be the active contour for the shape from the first frame. Let $\phi: \mathbb{R}^2 \to \mathbb{R} = \phi(x)$ be a level set distance function associated with contour $\mho$ and $\phi_0: \mathbb{R}^2 \to \mathbb{R} = \phi_0(x)$ is a level set function with contour $\mho^{\mathbf{0}}$ from first image frame $\mathcal{J}^{(\mathbf{0})}$. let x be a pixel in the image space $\mathbb{R}$ fixed, $\phi(x) = \phi(\mho; x)$ is actually a function of contour$\mho$**.** The initial contour $\mho$ aligns itself with the object contour $\mho^{\mathbf{0}}$ in the first frame that is the initial contour for the next frame in the video sequence coming from the contour in the previous frame. Hence the shape interaction term proposed in this paper has the from

$$\mathfrak{E}^{\mathrm{S}}\left(\mho^{\mathbf{obj}}\right) = \int_{\boldsymbol{int}(\mho)} \boldsymbol{\phi_0(x)dx} \qquad [25]$$

Thus by applying shape energy to the level set we can effectively segment sign video and we could differentiate between object contour modifications due to motion and shape changes.

3.4. Integrated Energy Functional for Video Segmentation

By integrating the energy functions from color, texture, boundary and shape modules we formulate the following energy functional of the active contour as

$$\mathfrak{E}^{\mathrm{T}}\left(\mho^{\mathbf{obj}}\right) = \boldsymbol{\zeta}\mathfrak{E}^{\mathrm{CT}}\left(\mho^{\mathbf{obj}}\right) + \boldsymbol{\eta}\mathfrak{E}^{\mathrm{B}}\left(\mho^{\mathbf{obj}}\right) + \boldsymbol{\chi}\mathfrak{E}^{\mathrm{S}}\left(\mho^{\mathbf{obj}}\right) \qquad [26]$$

where $\boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\chi}$ are weighting parameters that provide stability to contribution from different energy terms. All terms are positive real numbers. The minimization of the energy function is done with the help of Euler-Lagrange equations and realized using level set functions. The resultant level set formulation is

$$\frac{d\boldsymbol{\phi}^{\boldsymbol{n}}(\boldsymbol{x},\boldsymbol{t})}{dt} = \Bigg(-\boldsymbol{\zeta}\left(\mathbf{D}\left(\boldsymbol{C}_{\boldsymbol{f}_i}^{(\boldsymbol{n}-\mathbf{1})}, \boldsymbol{C}_{\boldsymbol{f}_i}^{(\boldsymbol{n})}\right) + \mathbf{D}\left(\boldsymbol{C}_{\boldsymbol{b}_j}^{(\boldsymbol{n}-\mathbf{1})}, \boldsymbol{C}_{\boldsymbol{b}_j}^{(\boldsymbol{n})}\right)\right) \\ + \boldsymbol{\eta}\left(\mathbf{g}\left(\boldsymbol{\mathcal{B}}^{obj}(\boldsymbol{x})\right) + \nabla.\left(\mathbf{g}\left(\boldsymbol{\mathcal{B}}^{obj}(\boldsymbol{x})\right)\left[\frac{\nabla\phi}{|\nabla\phi|}\right]\right)\right) - \boldsymbol{\chi}\boldsymbol{\phi}_{\mathbf{0}}(\boldsymbol{x})\Bigg) \|\nabla\phi^n\| \qquad [27]$$

The numerical implementation of above equation eq.27 is computed using narrowband algorithm [28]. The algorithm approximates all the derivatives in eq.27 using finite differences. The level set function is reinitialized when the zero level set clutches the boundary of the object in the image frame.

4. EXPERIMENTAL RESULTS

In this section we show the results obtained on sign videos acquired under diverse surroundings. To validate the proposed method we use a measure to compute the correctness of spatial location of segmented objects. Suppose $\boldsymbol{\mathcal{A}}_{(\boldsymbol{obj})}^{(\boldsymbol{n})}$ is the area of segmented object in the nth frame from the proposed method and $\boldsymbol{\mathcal{G}}_{(\boldsymbol{obj})}^{(\boldsymbol{n})}$ is the ground truth area of the object attained by hand segmenting the same object in the same frame. $\boldsymbol{\mathcal{A}}_{(\boldsymbol{back})}^{(\boldsymbol{n})}$ area of the background in the current frame and $\boldsymbol{\mathcal{G}}_{(\boldsymbol{back})}^{(\boldsymbol{n})}$ is the ground truth background area. $\boldsymbol{\mathcal{A}}_{(\boldsymbol{img})}^{(\boldsymbol{n})}$ is the total area of the image under test. The object segmentation error is calculated from the equation

$$\varepsilon = \left| \frac{\left(\boldsymbol{\mathcal{A}}_{(\boldsymbol{obj})}^{(\boldsymbol{n})} - \boldsymbol{\mathcal{G}}_{(\boldsymbol{obj})}^{(\boldsymbol{n})}\right) + \left(\boldsymbol{\mathcal{G}}_{(\boldsymbol{back})}^{(\boldsymbol{n})} - \boldsymbol{\mathcal{A}}_{(\boldsymbol{back})}^{(\boldsymbol{n})}\right)}{\boldsymbol{\mathcal{A}}_{(\boldsymbol{img})}^{(\boldsymbol{n})}} \right| \qquad [28]$$

The error $\varepsilon \in [0,1]$ gives area intersections of segmented object to their background by total area of the image frame. In a sense this error tells us the percentage of misclassified pixels in each frame of the video sequence.

The first frame $\boldsymbol{\mathcal{J}}^{(\mathbf{0})}$ is segmented by calculating the feature vector consisting of single color plane (R or G or B) and texture information along with boundary edge map. The proposed method is then applied to the remaining frames of the video sequence. The segmentation result of the previous frames is used as a mask or initial contour for the current frame and so on. The energy minimization function in eq.26 is employed to process the level sets and to produce an optimal contour for segmentation of each frame. In all video sequences initial contour is a circle of radius 25 pixels that can be placed near to the object of interest.

In the first experiment we started with a video sequence which is produced in a controlled environment. The video is shot in a lab using web cam with dark background and with an additional constraint that signer should also wear a dark shirt. This video sequence is part of the database we have created for sign language recognition project. Our Sign language database consists of 310 signs with 8 different signers. The frame size of the 320×480. Figure 2 shows where we run our segmentation algorithm with values of $\zeta = 0.3, \eta = 0.5$ and $\chi = 0.2$. The object and background clusters are made of three and two clusters. All the results are compared with results produced from implementing the CV algorithm in [17].

The experiments are performed in R color plane. As such we can do it any color plane or with color videos. The problem with full color sequences pertaining to sign language is that sign language videos contain large sequence of frames with lot of information to be extracted. Figure 2(a) shows four frames from a video sequence of the signer performing a sign related to English alphabet ‘X’. This simple sign contain 181 frames. Figure 2(b) shows the results obtained from our proposed method. The inclusion of prior shape term in the level sets segments that the right finger shape in spite of being blocked by the finger from left hand. This segmentation result will help in good sign recognition.

Figure 3(a) and 3(b) shows the effectiveness of the proposed algorithm against the CV model. Segmentation Error is calculated for the sequence in figure 2(a) against the ground truth result obtained manually. The error is plotted in figure 4 for both proposed method and CV method.

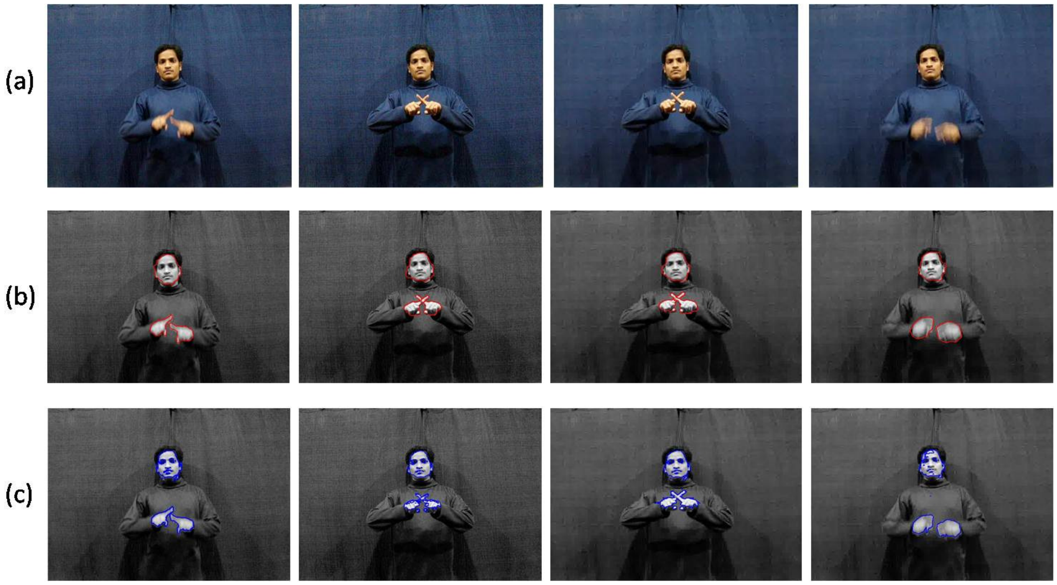

Figure 2. Experiment one showing our proposed segmentation algorithm on sign language videos under laboratory conditions. Frames 10, 24, 59,80 are shown. Row (a) shows four original frames. Row (b) shows the results from proposed algorithm and Row (c) Shows results of CV algorithm in [17].

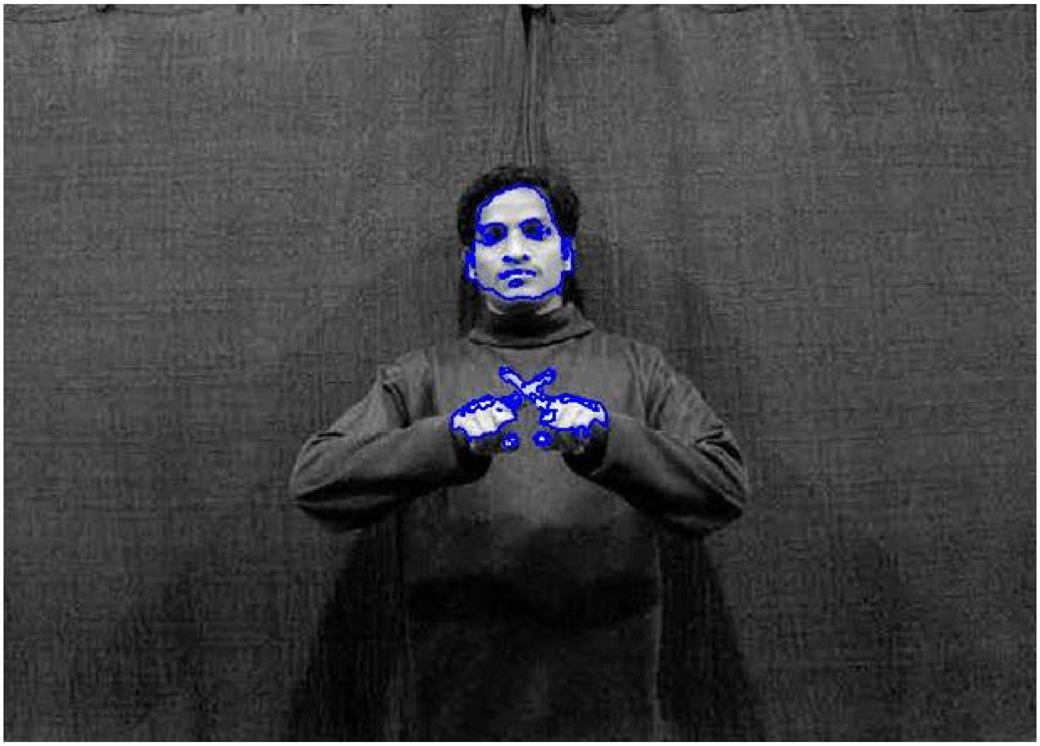

Figure 3(a). Showing the enlarged result with CV method [17] of previous Experiment

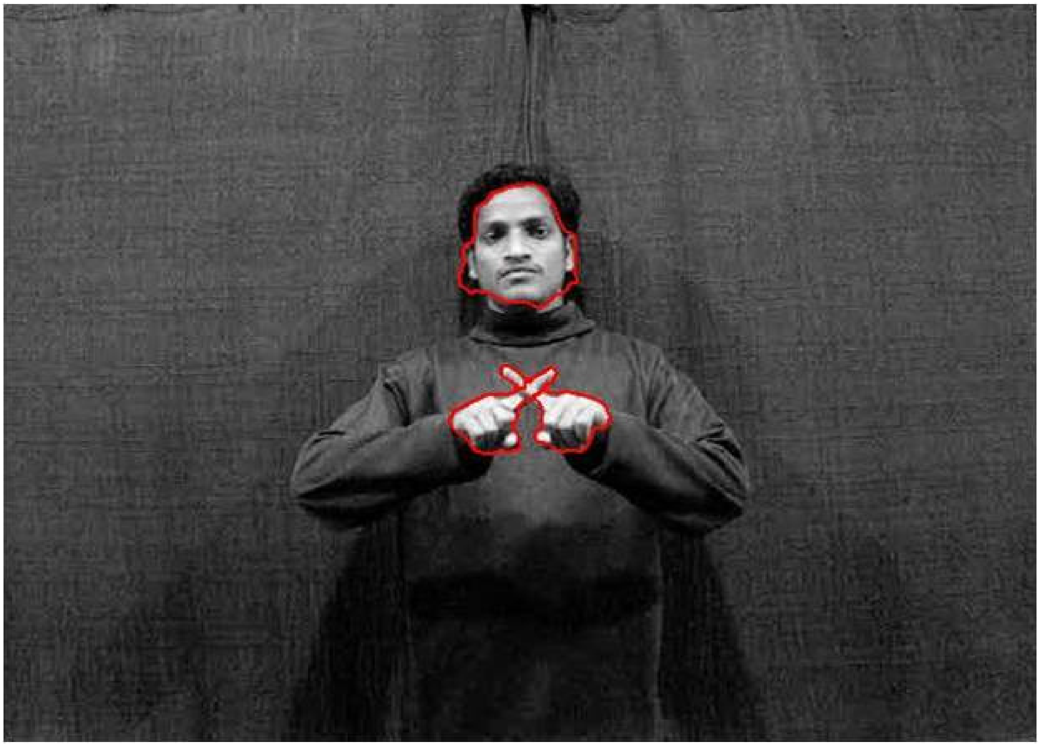

Figure 3(b). Showing the enlarged result with Proposed Method that is with Color, Texture, Boundary and prior shape info of previous Experiment

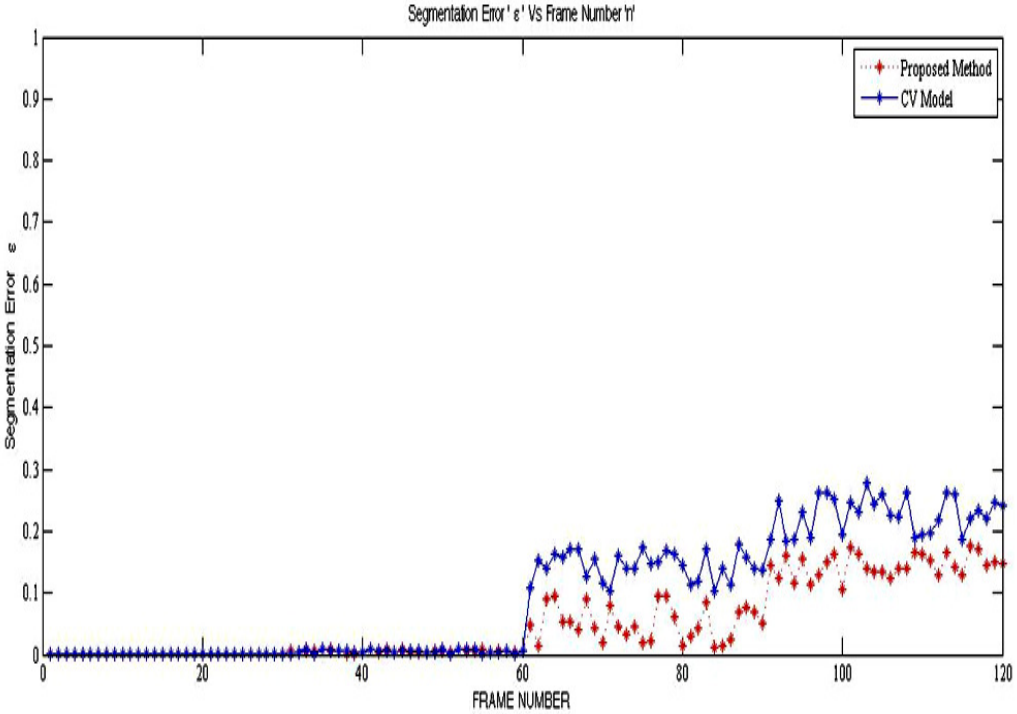

Figure 4. Segmentation Error for the sequence in figure 2(a)

We also experimented with noise added to the video sequence to test our proposed method. For this purpose we added to our video sequence in the previous experiment a white Gaussian noise of zero mean and standard deviation $\sigma = 2$. The results are shown in figure 5. As we observe the level set in CV model deviated drastically from the set perimeter for segmentation. The advantage of our method is clearly visible as it is able to segment the hands and head portions without much difficulty. This is due to the additional prior shape from the previous frame. Here for this video sequence we have increase the shape weighing term to $\chi = 0.43$ to influence the contour to shape information. The segmentation error is plotted in figure 6 which indicated there is a larger deviation from ground truth segmentation in case of CV model when compared to level sets with prior shape knowledge.

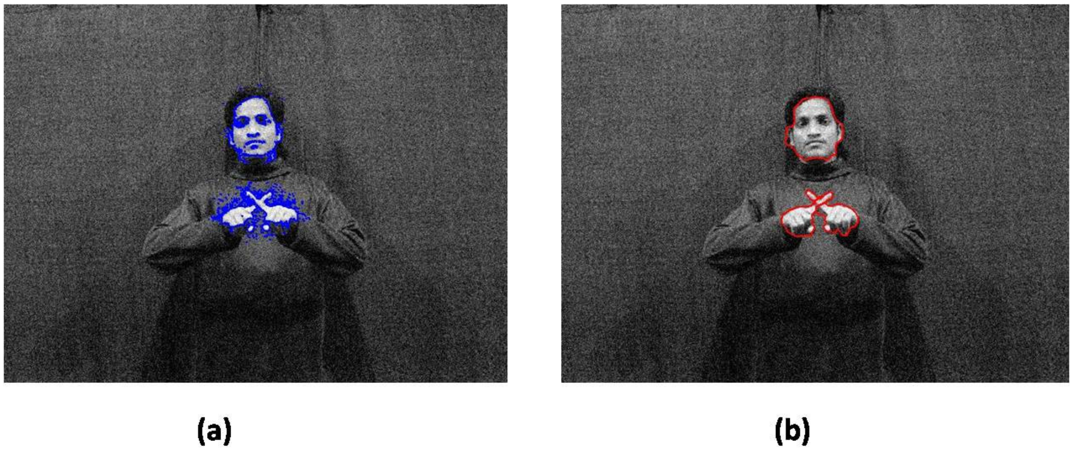

Figure 5. Example showing robustness of segmentation to noise. (a) is CV model and (b) Proposed method both with white Gaussian noise of $\sigma = 2$ and $\mu = 0$

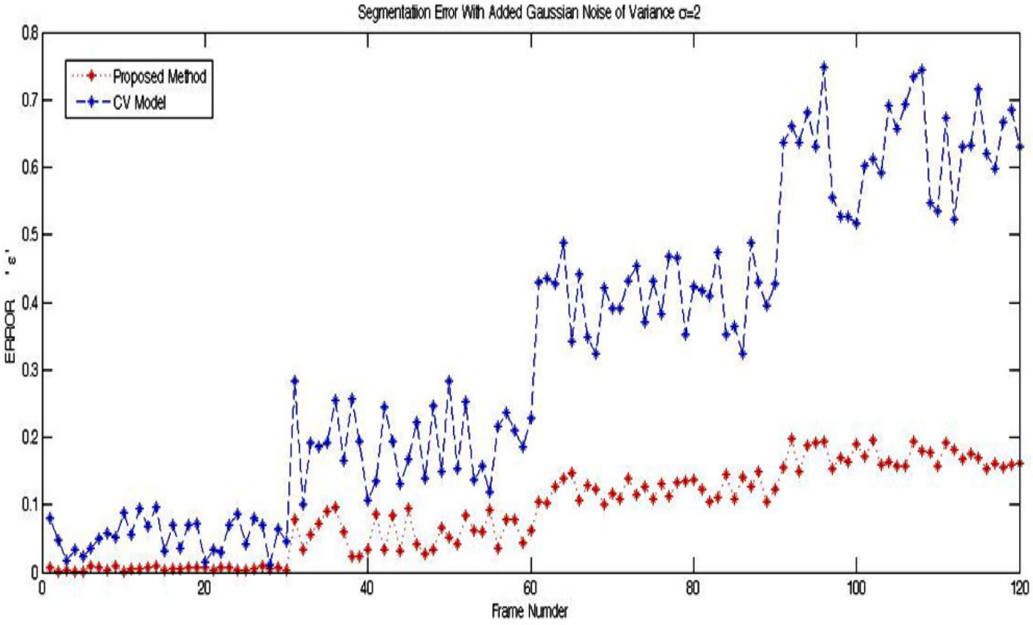

Figure 6. Segmentation Error Plot with noise added to the image sequence in figure 2(a)

We also experimented with more real time scenarios so that sign language recognition system can be implemented under real time. The video sequence that is considered is taken in a restaurant where it is difficult to identify the signer and the hands of the signer with multitude of background clutter. For this image sequence we have manually extracted the signer's hands and head portions from the first frame $\boldsymbol{\mathcal{I}}^{(\mathbf{0})}$ which is used to initialize the proposed level set. The weighing parameters in eq.26 $\zeta = 0.24, \eta = 0.21$ and $\chi = 0.63$. We observed that segmentation is good if shape term in eq.26 weight is increased. Because in this real time video the color and texture information does not reveal much of information. Similarly the boundary information also provides insufficient data under the influence of such a background clutter.

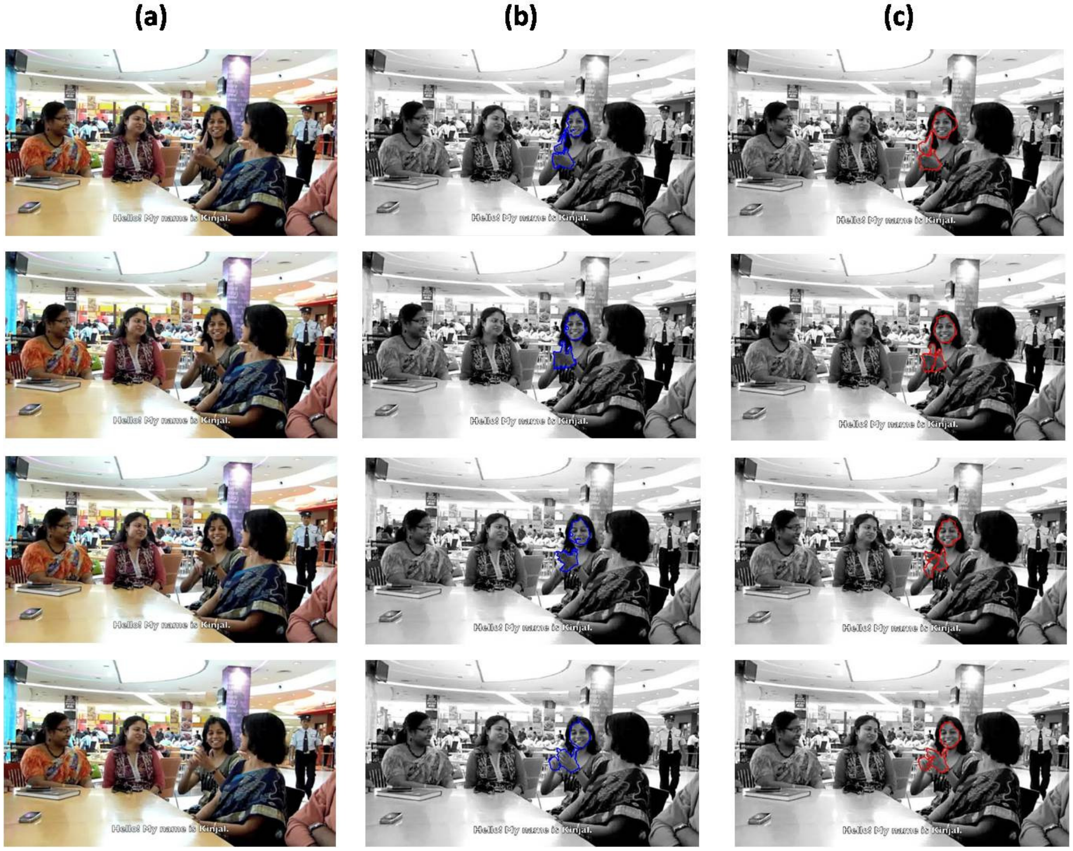

Figure 7. Comparison of Segmentation results for real time video sequence where frames 1110, 1123, 1190 and 1221 are shown. Column (a) original video sequence, column (b) results from CV model and Column (c) results from proposed method.

We observe occlusions of hands and head very frequently in sign language videos. Most sign language recognition systems insist that the signer should face the camera directly to avoid occlusions of hands largely. This problem is solved using our level set method. The figure 8 shows only this. We initialized contour for only right hand of the signer in the first frame. With the left hand coming in the path of right hand as can be observed from the original sequence in figure 8(a), it's difficult to segment the shape of right hand. But the results in figure 8(b) show the segmentation of right hand only with occlusions from left hand. This breakthrough will help is designing sign language systems with utmost robustness.

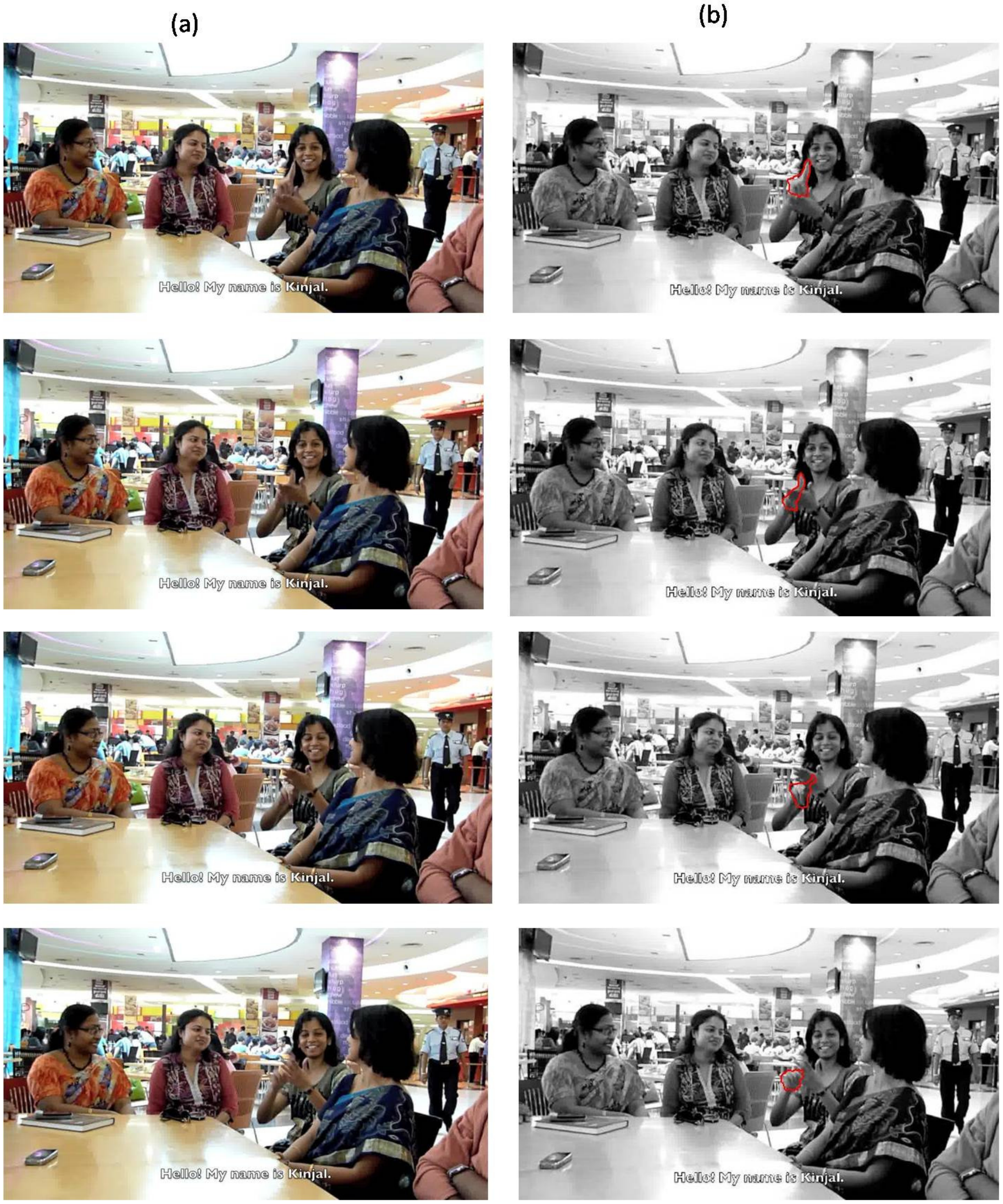

Figure 8. Showing the influence of prior shape knowledge. Here only the occluded right hand is segmented. Column (a) showing original image sequence and column (b) the segmentation result.

The only problem with our model, as we can observe the segmentation shape is not exact to that of right hand. This is due to the imperfect segmentation in the previous frame which is taken as a mask to current frame. This problem can be fixed by reinitializing the active contour whenever occlusion period is longer.

The final experiment shows the supremacy of our proposed technique when the video sequence contains fast moving objects in contrast to hand and head movements. The video sequence is shot on an Indian road and in the natural environment. Figure 9 shows the original sequence in column (a) along with the results of CV model in column (b) and our method in column (c).

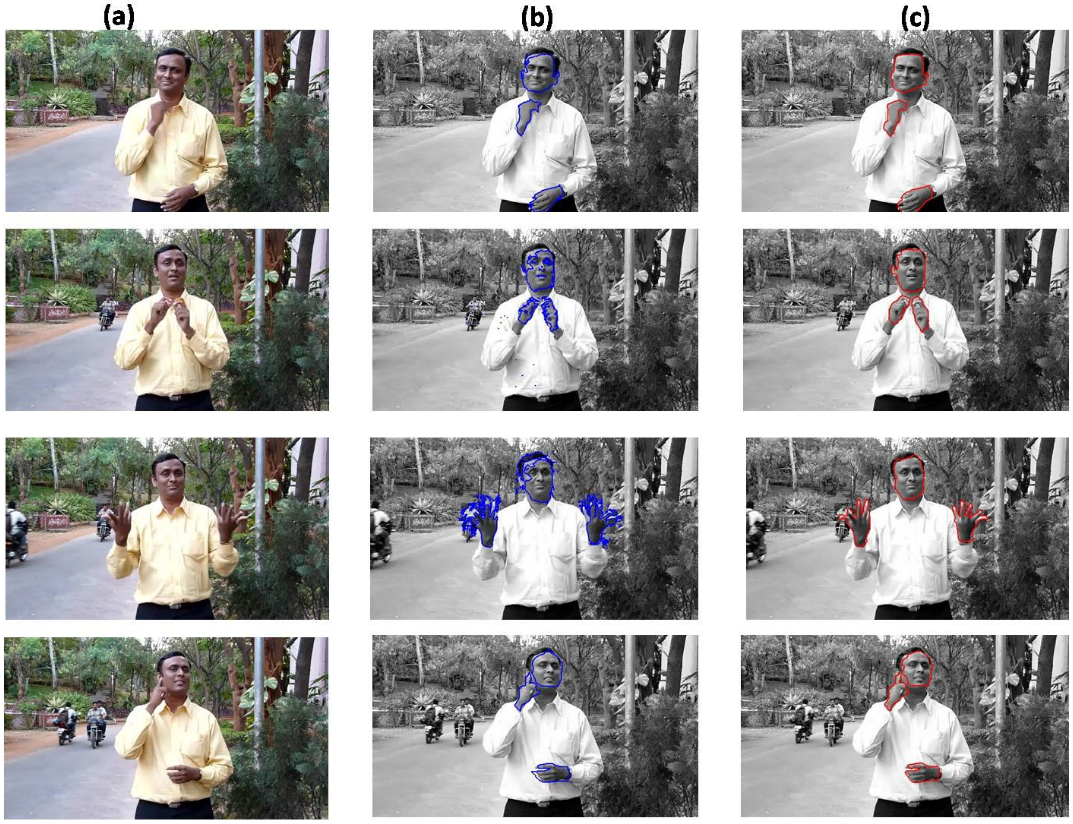

Figure 9. Frames of a sign video sequence on a Indian road and under natural environment. Column (a) is original sequence of frames 39, 54,79 and 99. Column (b) CV method and Column (c) our proposed method.

Observation of third row expose the disadvantage associated with CV method of segmentation. In this video the background object suddenly appears in the frame to which the CV technique provides much resistance and the final segmentation result include the object. But providing prior shape information along with object color, texture and boundary edge map proves the strength of our method. Also we get the unwanted segments in the form of leaves of trees in the background for left hand of the signer in row three for CV method which is not an issue with our method.

The plot between segmentation errors calculated using eq.27 and frame number, i.e. error per frame shows the error increases as the bike enters the frame for CV method. As pointed earlier error in our method also increasing due to re-initialization problem of the level set to the current frame. Error can be minimized by employing a re-initialization algorithm to initialize the initial contour whenever there is large change in the object of interest otherwise if there is change in the background. For the moment our method has shown that prior knowledge of shape along with other cues such as color, texture and boundary information can provide good segmentation results for segmenting hand gestures of sign language.

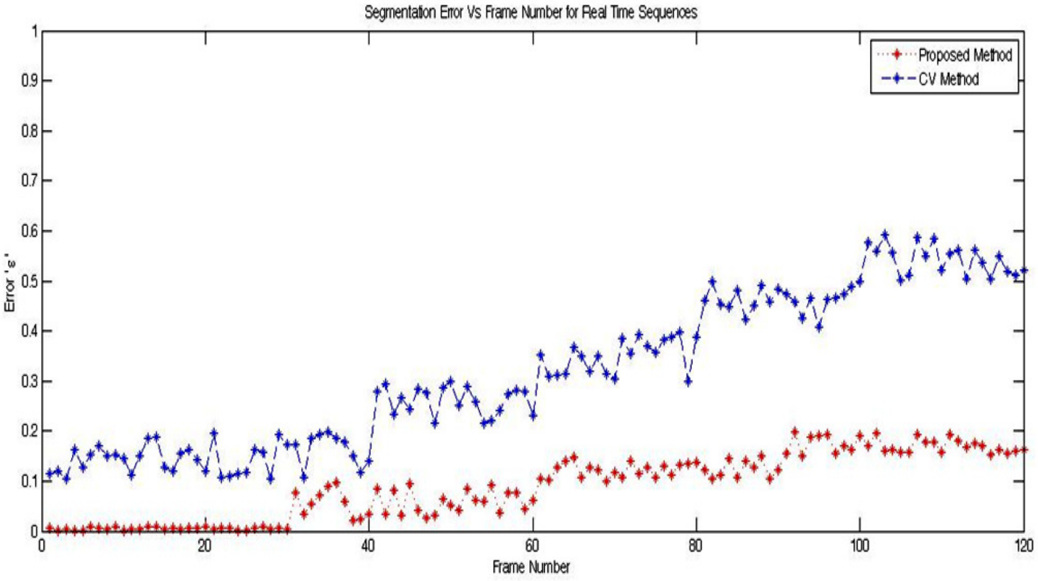

Figure 10. Segmentation Error versus frame number plot for results in 9(b) and 9(c).

5. Conclusions and Discussions

This paper brings us a little closer in making sign language recognition system a reality. The method proposed combines effectively the color, texture, boundary and prior shape information to produce an effective video segmentation of sign language videos under various harsh environments such as cluttered backgrounds, poor lighting, fast moving objects and occlusions. The color and texture information is extracted statistically by creating a feature vector and classifying each pixel in the frame to object and background pixel. Boundary information is provided by divergence operator along with the curvature of the object under consideration. Including Shape information from the previous frame did a whole lot of difference to the level set minimization to segment effectively the occulted hand from other hand and also head some times. We have effectively demonstrated by experimentation of the proposed method by applying it to video sequences under various conditions. Nevertheless there are challenges which are to addressed to apply this method to continuous sign language recognition systems to carry out the segmentation in real time.

The segmentation experiments are run on an Intel core i3 2.5GHz processor with 3GB RAM using MATLAB software. The average running time was 6 frames per second with around 40 level set iterations per frame. The resolution of the video is kept to minimum along with color information which in this paper was restricted to only one color plane. This is to reduce the computation time which otherwise increases for higher resolutions and full color processing. The speed of the algorithm can be increased by using fast numerical schemes for determining the level sets as given in [29].

References

[1] W. Stokoe, D. Casterline, and C. Croneberg,(1965) "*A Dictionary of American Sign Language on Linguistic Principles*." Gallaudet College Press, Washington D.C., USA.

[2] P.V.V.Kishore, P.Rajesh Kumar, E.Kiran Kumar & S.R.C.KIshore (2011). *"Video Audio Interface for Recognizing Gestures of Indian Sign Language"* International Journal of Image Processing(IJIP), CSC Journals, Vol. 5, Issue(4), pp479-503.

[3] Qing-song Zhu, Yao-qin Xie, Lei Wang (2010) *Video Object Segmentation by Fusion of Spatio-Temporal Information Based on Gaussian Mixture Model*, Bulletin of advanced technology research, vol. 5, No. 10, pp38-43.

[4] Wen-kai Tsai, Chung-chi Lin, Shyue-wen Yang, Ming-hwa Sheu, Ching-lung Su (2008), *"Adaptive Motion Gesture Segmentation"*. The 2008 International Conference on Embedded Software and Systems Symposia (ICESS2008), pp386-391.

[5] Qiuyu Zhang, Fan Chen, Xinwen Liu,(2008), "Hand Gesture Detection and Segmentation Based on Difference Background Image with Complex Background" The 2008 International conference on Embedded Software and Systems Symposia (ICESS2008), pp338-343.

[6] N. Peyrard and P. Bouthemy (2002) "*Content-based video segmentation using statistical motion models"*. In *Proc. BMVC*, Cardiff, pages 527–536.

[7] S. V. Porter, M. Mirmehdi, and B. T. Thomas (2003), "*Temporal video segmentation and classification of edit effects"*. Image and Vision Computing, Vol. 21 No.13-14,pp1097–1106.

[8] Qiulei Dong, Yihong Wu and Zhanyi Hu,(2006), "*Gesture Segmentation from a Video Sequence Using Greedy Similarity Measure",* Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), pp233-236.

[9] Collins, R.T., Liu, Y., Leordeanu, M. (2005) "*Online selection of discriminative tracking features"*. IEEE Trans. Pattern Anal. Mach. Intell. Vol.27 No.10 , pp1631–1643.

[10] Ginmo Chung and Luminita Vese. (2003) *"Image segmentation using a multilayer level set approach.Technical Report*" 03-53, UCLA.

[11] R. Malladi, J.A. Sethian, and B.C. Vemuri. (1995) *"Shape modeling with front propagation: A level set approach*". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.17 No. 2,pp158-175.

[12] D. Terzopoulos and K. Fleischer (1988) "*Deformable models"*. The Visual Computer, Vol.4 No.6, pp306-331.

[13] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer.(1987) "*Elastically deformable models*" In Comp.Graphics Proc., pages 205-214. ACM Press/ ACM SIGGRAPH.

[14] Osher, S., Sethian, J. (1988) Fronts propagating with curvature-dependant speed: algorithms based on Hammilton-Jacobi formulations.J. Comput. Phys. Vol.79 No.1, pp12–49

[15] Luminita Vese and Tony Chan.(2002) "*A multiphase level set framework for image segmentation using the mumford and shah model"*. International Journal of Computer Vision, Vol.50,No. 3 pp271-293.

[16] D. Mumford and J. Shah.(1989) *"Optimal approximation by piecewise smooth functions and associated variational problems"* Comm. Pure Appl. Math, Vol.42,pp577-685.

[17] T. Chan and L. Vese(2001) "*Active contours without edges*". IEEE Transations on Image Processing,Vol. 10 No.2,pp266-277.

[18] M.Kass, A Witkin, D Terzopoulos (1987),*"Snakes: Active Contour Models"*, Int. J. of Computer Vision*,* pp 321-331.

[19] G. Chung, L.A. Vese, (2005) *"Energy Minimization Based Segmentation and Denoising Using a Multilayer Level Set Approach,"* Energy Minimization Methods in Computer Vision and Pattern Recognition, vol. 3757/2005, pp. 439–455.

[20] Tuceryan, M., Jain, A. (1998) : Texture analysis. In: Chen, C.H., Pau, L.F., Wang, P.S.P. (eds) The Handbook of Pattern Recognition and Computer Vision, 2nd edn. chap. 2.1, 207–248.World Scientific Publishing, Singapore.

[21] Allili, M.S., Ziou, D.,(2006), *“Automatic color-texture image segmentation by using active contours”*. In: Proceedings of 1st IEEE International Workshop on Intelligent Computing in Pattern Analysis/Synthesis, Xi’an, China, 26–27, LNCS 4153, pp. 495–504.

[22] S. Allam, M. Adel, P. Refregier, (1997) *“Fast algorithm for texture discrimination by use of a separable orthonormal decomposition of the co-occurrence matrix,”* Applied Optics, vol.36, pp.8313–8321.

[23] R. M. Haralick, K. Shangmugam, I. Dinstein, (1973) *“Textural Feature for Image Classification,”* IEEE Trans on Systems, Man, Cybernetics, 3(6), pp.610—621.

[24] Goldenberg, R., Kimmel, R., Rivlin, M. and Rudzsky, M (2002), *“Fast Geodesic Active Contour,”* IEEE Trans. Pattern Anal. Machine Intell, vol.24, pp603-619.

[25] Chan, T., & Zhu, W. (2005). *Level set based prior segmentation.* ProceedingCVPR, vol.(2), pp1164–1170.

[26] Cremers, D., Soatto, S.: A pseudo-distance for shape priors in level set segmentation. In: Proceedings of 2nd IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision, Nice, France, 13–16 October 2003, pp169–176.

[27] Ketut Fundana · Niels C. Overgaard · Anders Heyden,(2008), *“Variational Segmentation of Image Sequences Using Region-Based Active Contours and Deformable Shape Priors”* Int J Comput Vis vol.80: pp289–299.

[28] Adalsteinsson, D., Sethian, J,(1995) “*A fast level set method for propagating surfaces”*. J. Comput. Phys. 118 vol (2), pp269–277.

[29] Weickert, J., K¨uhne, G.: Fast methods for implicit active contour models. In: Osher, S., Paragios, N.: (eds) Geometric Level Set Methods in Imaging, Vision and Graphics, chap. 3, pp. 44–57. Springer, Heidelberg.

[30] www.deafsigns.org

Authors

P.V.V.Kishore (SMIEEE’07) received his M.Tech degree in electronics from Cochin University of science and technology in the year 2003, and currently pursuing PhD at Andhra University College of engineering in Department of ECE from 2008. He is working as research scholar at the Andhra university ECE department. He received B.Tech degree in electronics and communications engineering from JNTU, Hyd. in 2000. His research interests are digital signal and image processing, computational intelligence, human computer interaction, human object interactions. He is currently a student member of IEEE.

Dr. P.Rajesh Kumar (MIEEE’00, FIETE’09) received his Ph.D degree from Andhra University College of Engineering for his thesis on Radar Signal Processing in 2007. He is currently working as associate professor at Dept. of ECE, Andhra University College of engineering, Visakhapatnam, Andhra Pradesh. He is also Assistant Principal of Andhra University college of Engineering, Visakhapatnam, Andhra Pradesh. He as produced numerous research papers in national and international journals and conferences. He has guided various research projects. His research interests are digital signal and image processing, computational intelligence, human computer interaction, radar signal processing.