Identification of Dynamic Objects in Video Streams for Multimedia Mining

Vibha L*, Chetana Hegde*, Prashanth S J*, P Deepa Shenoy*, Venugopal K R* and L M Patnaik**

Abstract: **Video Segmentation is one of the most challenging areas in Multimedia Mining. It deals with identifying an object of interest in a movie clip. It has wide application in the fields like Traffic surveillance, Security, Criminology etc. This paper proposes a technique for identifying a moving object in a video clip of stationary background for real time content based multimedia communication systems. First frames at certain distances are considered for observation and their differences with the current frame are computed to identify the areas that have no motion and these areas are then considered as the background. Next the background is eliminated by applying post processing techniques. The remaining region is then filled by the original pixels resulting in the identification of moving object. The method of least square is used to compare the accuracy of the output of this algorithm with the already existing algorithms.**

Key Words: **Background elimination, Frame Difference, Least square method, Object identification.**

I. Introduction

VIDEO mining can be defined as the unsupervised discovery of patterns in audio – visual content. The motivation for such discovery comes from the success of data mining techniques in discovering non-obvious patterns. In video mining we can discover interesting events in the video even without any prior knowledge about the events. The objective of video mining is to extract significant objects, characters and scenes in a video by determining their frequency of re-occurrence. Some of the basic requirements needed for extracting information in a video mining technique are i) It should be as unsupervised as possible. ii) It should have as few assumptions about the data as possible. iii) It should be computationally simple. iv) It should discover interesting events.

The success and significance of video mining depends on the content genre. Carefully scripted and staged videos like drama, news etc. posses extremely strong structure but has tremendous intra-genre variation in production styles that vary from country to country or content-creator. A Sports content falls somewhere in the middle as it is not scripted but is constrained by rules. Surveillance content is not scripted and often not constrained by any rules other than those imposed by the geometry of the setting and the laws of gravity. Each genre therefore poses a unique challenge to pattern discovery. Segmentation can be an extremely easy task if one has access to the production process that has created the discontinuities. For example, the generation of a synthetic image or of a synthetic video implies the modelling of the 3-D world and of its temporal evolution. However, if the segmentation intents to estimate what has been done during the production process, its task is extremely difficult and one has to recognize that the state of the art has still to be improved to lead to robust segmentation algorithms that are able to deal with generic images and video sequences. For multimedia services, segmentation may address different goals, which include object detection for coding (MPEG-4) or description (MPEG-7), estimation of partitions allowing an efficient encoding, temporal tracking or regions or shot-cut detection. The meaning of the word segmentation depends to a large extent on the application and the context in which it is used. The goal of any segmentation algorithm is to define a partition of the space. In the case of image and video, the space can be temporal (1-D), spatial (2-D) or spatio-temporal (3-D).

Video data consists of a set of frames, and a frame is considered as a still image. Video segmentation is one of the most challenging tasks in video mining as it requires a semantic understanding of the video to some extent. These images when displayed continuously one after another at fixed rates constitute the video stream. Segmentation subdivides a frame into its constituent regions or objects. The level to which the subdivision is carried depends on the problem being solved, i.e. segmentation should stop when the objects of interest in an application have been isolated.

This paper is organized as follows – Section 2 deals with the related work and Section 3 presents the architecture and model. Section 4 is about the problem definition. Section 5 presents the implementation of the proposed algorithm and the performance analysis. Section 6 contains the conclusion.

II. Related Work

A brief survey of the related work in the area of video segmentation is presented in this section. Video segmentation helps in the extraction of information about the shape of moving object in the video sequences. This concept is used for intelligent signal processing and content-based video coding presented in [1]. An image scene contains a number of video objects and the attempt is to encode the sequence that allows separate decoding and construction of objects. Nack et al., [2] and Salembier et al., [3] have discussed Multimedia

* Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bangalore, INDIA -560001. (e-mail: vibhal1@rediffmail.com, chetanahegde@yahoo.co.in)

** Microprocessor Applications Laboratory, Indian Institute of Science, Bangalore, INDIA-560012

content description related to the generation of region based representation with respect to MPEG-4 and MPEG-7.

Video segmentation algorithms can be broadly classified into two types based on their primary criteria for segmentation. Wang D. proposed a technique of Unsupervised video segmentation in [4], which consists of two phases i.e. initial segmentation and temporal tracking. The initial segmentation [5] is applied on the first frame of the video sequence, which performs spatial segmentation and partitions the first frame into homogeneous regions based on intensity, then the motion estimation is computed for determining the motion parameters for each region, and finally motion-based region merging is performed by grouping the regions to obtain the moving objects. After the initial segmentation, temporal tracking [6] is performed.

P. Salembier [7] found better results using spatial homogeneity as the primary criteria, which incorporates luminance and motion information simultaneously. The procedure includes the steps like joint marker extraction [8], [9], boundary decision and motion-based region fusion. Then the spatio-temporal boundaries are decided by the watershed algorithm. Choi et al., [10] used Joint similarity method for this purpose. Finally, motion-based region fusion is used for eliminating the redundant regions. The techniques used segmentation is as follows, first filters are used to simplify the image and then Watershed algorithm is applied for boundary detection [11]. Later the motion vector is computed using motion estimation and regions with similar motion are merged together to constitute the final object region. As watershed algorithm is being used they generate object boundaries which are more efficient and precise than any other methods. The change detection method is used as the primary segmentation criteria in [12] The major issue here is to guarantee robust detection in results in presence of noise, and many shortcomings are overcome by using Markov random field based on refining method. The position and shape of the moving object is determined using the frame difference concept, followed by a boundary fine-tuning process based on temporal information. Algorithms that deal with spatial domain processing first, without knowing much regarding the motion information will waste much of the computing power in segmenting the background.

Neri et al., [13] describes a solution to eliminate the uncovered background region by applying motion estimation on regions with significant frame difference. The object in the foreground is then identified when a good match is found between two frame differences. The remaining region is then discarded as unwanted areas. The shadow of the object in the background region may also affect the output in change detection based approach [14].

III. Architecture and Modeling

In many real-time applications like video conferencing, the camera is fixed. Some techniques proposed in paper [12], [18] use global motion estimation and comparison to compensate the change in background due to camera motion. In this algorithm, we will assume the stationary background for videos. The architecture and modelling of the proposed algorithm is shown in Figure 1.

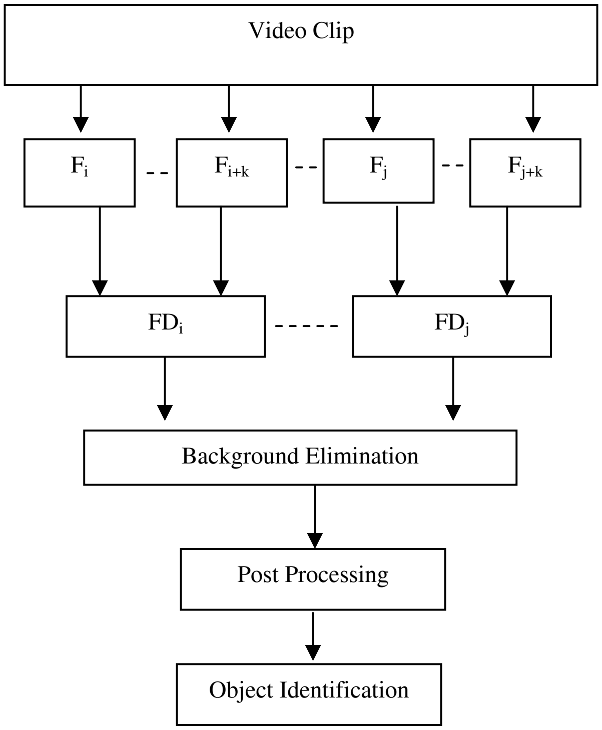

Fig. 1. Architecture for Object Identification

The flow of the algorithm is as follows: **A** video clip is read and it is decomposed into a number of frames**.** In the first stage difference between frames are computed i.e. F_i and F_{i+k}. In the next stage these differences are compared and the third stage involves eliminating pixels having the same values in the frame difference. The fourth phase is the post processing stage executed on the image obtained in third stage and the final phase is the object detection.

A. Frame Difference

Frame differences can be computed by finding the difference between consecutive frames but this will introduce computational complexity in case the video clips having slow-moving objects. Moreover this algorithm assumes a stationary background. Hence the difference between the frames at regular intervals (say, some integer k) is considered. If there are *n* frames, then we will get (n/k) frame differences (FD). The frame difference follows Gaussian distribution as indicated in equation (1)

$$p(FD) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{(FD-\mu)^2}{2\sigma^2} \right) \qquad (1)$$

Here, μ is the mean of FD and σ is the standard deviation of FD. The frame differences of some test sequences are as shown in Figure 2.

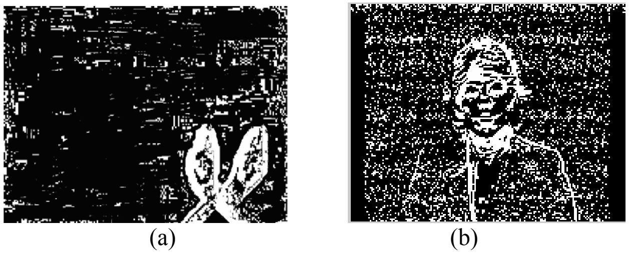

Fig. 2. Frame Difference
(a) Difference between 30th and 35th frames of Moving Hands
(b) Difference between 12th and 15th frames of Claire

B. Background Elimination

Once the frame differences are computed the pixels that belonging to the background region will have a value almost equal to zero, as the background is assumed stationary. Many a times because of camera noise, some of the pixels belonging to the background region may not tend to zero. These values are set to zero by comparing any two frame differences, say, FD_i and FD_j. Thus, the background region is eliminated and only the moving object region will contain non-zero pixel values. The images obtained after background elimination is as shown in the Figure 3.

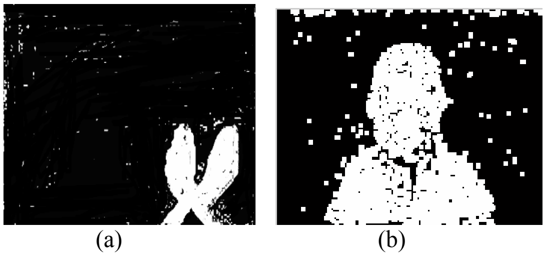

Fig. 3. After Background Elimination
(a) Image of Moving Hands (b) Image of Claire

C. Post Processing

Many a times due to camera noise and irregular object motion, there always exists some noise regions both in the object and background region. Moreover the object boundaries are also not very smooth; hence a post processing technique is required. Many post processing techniques are applied on the image that is obtained after background elimination. Initially, order-statistics filters are used, which are the spatial filters and whose response is based on ordering (ranking) the pixels contained in the image area encompassed by the filter. The response of the filter at any point is then determined by the ranking result. The current algorithm uses Median filter which is the best-known order-statistics filter. This filter replaces the value of a pixel by the median of the gray levels in the neighbourhood of that pixel. The formula used is –

$$\hat{f}(x, y) = median \ \{g(s, t)\} \qquad (2)$$

After applying the median filter, the resulting image is converted into a binary image. The morphological opening technique is applied on this binary image. The opening of *A* by *B* is simply erosion of *A* by *B* followed by dilation of the result by *B*. This can be given as –

$$A \circ B = (A \Theta B) \oplus B \qquad (3)$$

Here, *A* is the image and *B* is a structuring element. A flat structuring element containing 2*P+1 member is used in this algorithm. Here, P is any integer. To provide the number of rows and columns for structuring an element, a vector V of integers is used. In this situation, one structuring element member is located at the origin. The other members are located at 1*V, -1*V, 2*V, -2*V, . . ., P*V and -P*V. After applying the above explained pre-processing techniques, the new image obtained is as shown in the Figure 4.

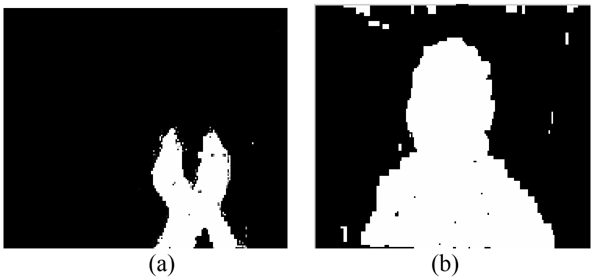

Fig. 4. After post-processing
(a) Image of Moving Hands (b) Image of Claire

D. Object Identification

The image obtained in the pre-processing step has less noise. And so, the background area is completely eliminated. Now, if the pixel values of this image are greater than certain threshold, then, those pixels are replaced by the pixels of the original frame. This process identifies the moving object as shown in Figure 5.

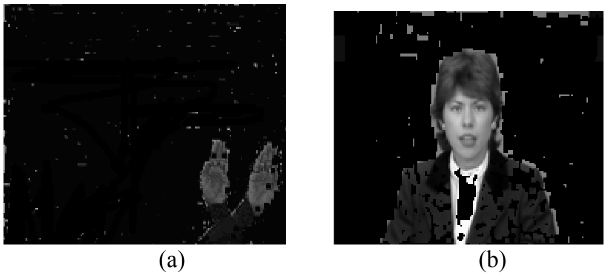

Fig. 5. Identification of Objects
(a) Moving Hands (b) Claire

IV. ALGORITHM

A. Problem Definition

Given a video clip in the format of QCIF (176 x 144), the objectives are:

i) To detect a moving object using the concept of background elimination technique.

ii) To improve the clarity of the moving object and compare it with the already existing algorithms.

B. Assumptions

i) The background of a video sequence is stationary.

ii) The motion of the object is slow.

C. Algorithm

Three major functions are involved in the proposed technique. The first function is to read a given video clip and convert it into frames as shown in Table I. Second function is to implement the major procedures like finding frame differences, eliminating the background, post-processing and then identifying the moving object as described in Table II. The last function used is to implement the Least Square Method (LSM) on the outputs obtained for performance comparison shown in Table III. The aim of the algorithm is to design an efficient moving object segmentation system. The algorithm has been implemented to achieve faster processing time as well as to improve quality of segmentation results. It was tested for QCIF (176 x 144) video sequences. The algorithms/pseudo codes for various steps involved are as shown below. Given a video clip, the initial problem is dividing it into number of frames. Each frame is then considered as an independent image. Every such image in RGB format is converted into Gray scale image.

In the next step, the difference between the frames at certain intervals is calculated. This interval can be decided based on the motion of moving object in a video sequence. If the object is moving quite fast, then the difference between every successive frame is considered. But, if the motion is slow, then the difference at the intervals of 3 or 5 frames is sufficient. The background region of the image is then eliminated. This is achieved by comparing any two frame differences, say, FD_i and FD_j. The matching pixels in FD_i and FD_j are considered to be a part of background and they are set to zero. All other pixels are unaltered. The image obtained by this procedure must undergo some post-processing techniques to remove the possible noise.

After post-processing, the image is compared with the one of the original frames (usually, the first frame). If the pixels are less than certain threshold, then they are ignored. Otherwise, they are replaced by the pixels of original image. This resulting image will be consisting of the moving object ignoring the background and hence satisfying our requirement.

TABLE I
To Read a Video Clip

```
Step 1:  Initialise an array  M_Array[] to an empty array.
Step 2: for i:= 1 to NumFrames in steps of 1
                Read each frame of video clip and store it
                 into M_Array[].
Step 3: Convert movie structures stored in M_Array[]
           into images.
Step 4: Convert the images obtained in Step 3 from
RGB
            to Gray format.
Step 5: Store all these gray images in an array viz.
            Im_Array[].
```

TABLE II
Object Identification

```
ALGORITHM IDOVS (Im_Array[],Rows, Cols, NumFrames)
//Input: An array of frames that are converted into images in
//         gray colour format viz. Im_Array[], Rows and Cols
//         indicating size of image and NumFrames indicating
//         total number of images in Im_Array[].
//Output: An image showing the moving object.
p:=1;
k:=5;      // any pre-defined value

// finding the frame differences
for i:=1 to Rows in steps of 1
    for j:=1 to Cols in steps of 1
        for m:=1 to NumFrames in steps of k
            FD[i,j,p]:=Im_Array[i,j,m+k]– Im_Array[i,j,m];
            p:= p+1;
        end for
    end for
end for

//Background Elimination
p:=2;
q:=4;      //any two pre-defined values

for i:=1 to Rows in steps of 1
    for j:=1 to Cols in steps of 1
        if (FD[i,j,p]==FD[i,j,q]) then
            BackElim[i, j]:= 0;
        else
            BackElim[i, j] := 255;
        end if
    end for
end for

//Post processing
K:= MedianFilter(BackElim);
G:= MorphologicalOpening(K);

//Object Identification
for i:=1 to Rows in steps of 1
    for j:=1 to Cols in steps of 1
        //TH is some observed threshold
        if G(i, j) >= TH then
            Object[i,j] = Im_Array[i, j, 1];
        end if
    end for
end for
```

V. IMPLEMENTATION AND PERFORMANCE ANALYSIS

Both the proposed algorithm and the background registration method are implemented using a Matlab 7. The performance analysis is done through the Least Square Method (LSM). The least square method is normally used to find the best-fit, given two sets of data. The method is as explained below.

TABLE III
Least Square Method

```
ALGORITHM LS (O[], BE[], Rows, Cols)
//Algorithm for finding least square value comparing
//original and identified images.
//Input: Original Image, O and identified image, BE
//     with Rows and Cols indicating size of images.
//Output: An integer value showing the least square
//           value.

LSValue:=0;
for i:=1 to Rows in steps of 1
      for j:=1 to Cols in steps of 1
            LSValue:= LSValue + {O(i,j) – BE(i, j)}^2
      end for
end for
```

Suppose that the data points are *(x_1-y_1), (x_2-y_2)… (x_n, y_n)* where x_1, x_2…. are the elements of first set and the *y1, y2,…* are the elements of second set. The deviation (error) d is calculated for each pair as

d_1=(x_1-y_1); d_2=(x_2-y_2) ….. d_n= (x_n - y_n)

According to the method of least squares, the best-fit must satisfy the following rule

$$\prod = d_1^{\,2} + d_2 + \ldots + d_n^{\,2} = \sum_{i=1}^{n} d_i^{\,2} = \min$$

This paper uses the least square method for comparing the outputs. Let O_{ij} be any frame of the input video clip, BE_{ij} be the result obtained through Background Elimination technique and BR_{ij} be the result obtained through Background Registration technique. Here, *i=1,2,…m* and *j=1,2,…n*. And *m* and *n* indicates rows and columns (i.e. size) of the image. The values are calculated using the formulae-

$$V1 = \sum_{i=1}^{m}\sum_{j=1}^{n}(O_y - BE_{ij})^2$$

And

$$V2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(O_y - BR_{ij})^2$$

It is observed through simulation that, *V1 < V2* for various test sequences. The actual values obtained for test sequences are given in Table IV. The outputs obtained through two different techniques are as shown in Figure 6. It is also observed that the clarity of the image obtained using our proposed algorithm is much clearer than the existing algorithm.

Simulation was carried out on standard QCIF sequences and on sequences captured in our laboratory. The results obtained from proposed algorithm are compared with those of background registration method. The Graph Showing Error Rates computed through Least Square Method is shown in Figure 6.

TABLE IV
Comparison of Error Rates

Video Sequence Name (i)	Value Obtained through the method of Least Squares for		Difference in Values (iv)=(ii)-(iii)
	Proposed Algorithm (Background Elimination Technique) (ii)	Existing Algorithm (Background Registration Technique) (iii)	
Moving Hands	13400352	14648832	-1248480
Claire	12442454	13141224	-698770
Salesman	13964820	15104228	-1139408
Grandma	12137846	13267234	-1129388

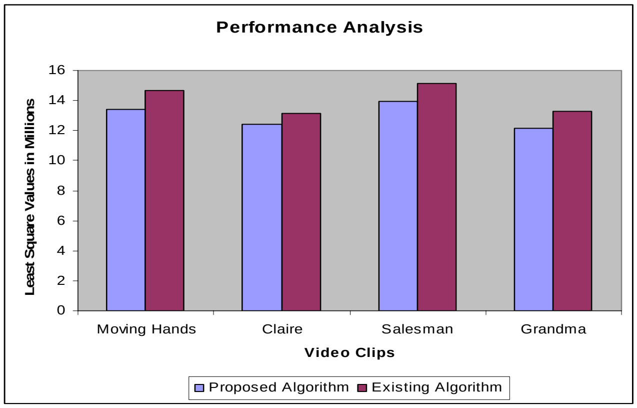

Fig. 6. Performance Analysis of Proposed and Existing Algorithm.

VI. CONCLUSION

In this paper, we propose an efficient algorithm for detecting a moving object using background elimination technique. Initially we compute the frame differences (FD) between frames F_i and F_{i+k}. The frame differences obtained are then compared with one another that help in identifying the stationary background image. The moving object is then isolated from the background. In the post processing step, the noise and shadow regions present in the moving object are eliminated using a morphological gradient operation that uses median filter without disturbing the object shape. This could be used in real time applications involving multimedia communication systems. The experimental results obtained indicate that the clarity of the image obtained using background elimination technique is much better than using background registration technique.

Good segmentation quality is achieved efficiently. In the future work we can consider the video sequences containing fast-moving objects and the sequences with more than one moving object.

VII. References

1. Sikora T., "The MPEG-4 Video Standard Verification Model", *IEEE Transactions, Circuits Systems, Video Technology,* vol. 7, Feb.1997, pp. 19-31.
2. Nack F. and Lindsay A. T., "Everything you Wanted to Know about MPEG-7: Part 2", *IEEE Multimedia*, vol.6, Dec. 1999, pp. 64-73.
3. Salembier P. and Marques F., "Region-based Representations of Image and Video: Segmentation Tools for Multimedia Services", *IEEE Transactions, Circuits Systems, Video Technology*, vol. 9, Dec. 1999, pp. 1147-1169.
4. Wang D., "Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking", *IEEE Transactions, Circuits Systems, Video Technology*, vol.8, Sept. 1998, pp. 539-546.
5. Y. Yokoyama, Y. Miyamoto and M. Ohta, "Very Low Bit Rate Video Coding using Arbitrarily Shaped Region-Based Motion Compensation", *IEEE Transactions, Circuits System. Video Technology,* vol. 5, pp. 500-507, Dec 1995
6. L. Wu, J. Benoise-Pineau, P. Delagnes and D. Barba, "Spatio-temporal Segmentation of Image Sequences for Object-Oriented Low Bit-Rate Image Coding", *Signal Processing: Image Communication.,* vol. 8, pp. 513-543, 1996.
7. P. Salembier, "Morphological Multiscale Segmentation for Image Coding", *Signal Processing,* vol. 38, pp. 359-386, 1994.
8. N. T. Watsuji, H. Katata and T. Aono, "Morphological Segmentation with Motion Based Feature Extraction", presented at *Int. Workshop on Coding Techniqus for Very Low Bit-Rate Video,* Tokyo, Nov. 8-10, 1995.
9. W. H. Hong, N.C. Kim and S.M. Lee, "Video Segmentation Using Spatial Proximity, Color and Motion Information for Region-Based Coding", in *Proceedings SPIE Visual Communications and Image Processing,* vol. 2308, pp. 1627-1633, 1994.
10. Choi J.C., Lee, S.W., and Kim, S.D., "Spatio-Temporal Video Segmentation Using a Joint Similarity Measure", *IEEE Transactions, Circuits Systems, Video Technology*, vol. 7, Apr. 1997, pp. 279-289.
11. Vibha L, Venugopal K. R. and L. M. Patnaik, "*A Study of Breast Cancer Using Data Mining Techniques*", Technical Report, University Visvesvaraya College of Engineering, Bangalore University, August 2003.
12. Aach T., Kaup A. and Mester R., "Statistical Model-Based Change Detection in Moving Video", *Signal Processing,* vol. 31, Mar.1993, pp. 165-180.
13. Neri,A., Colonnese S., Russo G. and Talone P., "Automatic Moving Object and Background Separation", *Signal Processing*, vol.66, Apr.1998, pp. 219-232.
14. Stauder J., Mech R. and Ostermann J., "Detection of Moving Cast Shadows for Object Segmentation", *IEEE Transaction Multimedia*, vol. 1, Mar. 1999, pp. 65-76.

VIII. Biographies

Vibha L. was born in India, on November 11, 1960. She graduated from U.V.C.E., completed her M.E. from U.V.C.E., has done her M.S. (Software Systems) from BITS., Pilani, and is currently a research scholar from MGR., University. She is presently employed as an Assistant Professor in department of CSE at Bangalore Institute of Technology.

Chetana Hegde was born in India, on July 22, 1978. She graduated from Karnatak University, Dharwad, completed her M.C.A. from Karnatak University, Dharwad, has done her M.Phil. (Comp. Sc.) from Madurai Kamaraj University. She is presently employed as Senior Lecturer in department of MCA at RNSIT, Bangalore.

Prashanth S. J. was born in India on July 13, 1984. He is an under-graduate student of the Department of Computer Science and Engineering at SJBIT, Bangalore.

P. Deepa Shenoy was born in India, on May 9 1961 She graduated from U.V.C.E., completed her M.E. from U.V.C.E., has done her M.S. (Systems and information) from BITS., Pilani, and has obiained her Ph.D in CSE from Bangalore University. She is presently employed as an Assistant Professor in department of CSE at U.V.C.E.

K.R. Venugopal graduated from U.V.C.E., obtained his masters degree in CSE from IISc Bangalore. He was awarded Ph.D in CSE from IIT madras, has authored 23 books, and has over 150 research papers to his credit.. His research interest includes Computer Networks, Digital Signal Processing and Data Mining. He is currently working as Principal and Dean of U.V.C.E.

L.M. Patnaik is currently working as a professor at CSE and Automation department at IISc. He has over 400 research publications in refereed International Journals and Conference Proceedings. He is a Fellow of all the four leading Science and Engineering Academies in India; Fellow of the IEEE and the Academy of Science for the Developing World.