An Experimental Comparison of Face Detection Algorithms

Shivesh Bajpai, Amarjot Singh, and K. V. Karthik

Abstract—**Human face detection and tracking is an important research area having wide application in human machine interface, content-based image retrieval, video coding, gesture recognition, crowd surveillance and face recognition. Human face detection is extremely important and simultaneously a difficult problem in computer vision, mainly due to the dynamics and high degree of variability of the head. A large number of effective algorithms have been proposed for face detection in grey scale images ranging from simple edge-based methods to composite high-level approaches using modern and advanced pattern recognition approaches. The aim of the paper is to compare Gradient vector flow and silhouettes, two of the most widely used algorithms in the area of face detection. Both the algorithms were applied on a common database and the results were compared. This is the first paper which evaluates the runtime analysis of Gradient vector field methodology and compares with silhouettes segmentation technique. The paper also explains the factors affecting the performance and error incurred by both the algorithms. Finally, results are explained which proves the superiority of the silhouette segmentation method over Gradient vector flow method.**

Index Terms—**Face detection, gradient vector flow (GVF), active contour flow, silhoutte.**

I. Introduction

The current advancements in computer technologies have envisaged a sophisticated vision based mechanical world, amended by artificial intelligence. This trend motivated the development in machine intelligence especially in the field of computer vision. Computer vision focuses to replicate human vision. Computer vision systems focus on performing laborious and repetitive jobs such as surveillance, in assembly line recapitulation etc. The main application is driven towards more generalized application of vision related to a broader field of face recognition and video coding. Face detection is a brilliant and extremely efficient paradigm for a number of applications like object detection etc. It is in itself a fascinating problem to explore.

Human face can be easily detected by human beings without much effort but for machines, it is one of the most complicated tasks to perform, due to the complex 3-D shape of the face and the variation in appearance under different lighting environments. Many of the current technologies used for face detection work only on frontal faces of similar size. Another major challenge is the generality of detecting faces in grayscale image due to the lack of standard methods for determining illumination data or scene structure without performing preliminary extensive operations on the image. This encouraged the development of several successful techniques, capable of detecting faces in knavish experimental conditions. Vision systems have to perform a number of operations involving segmentation, extraction, and verification of faces and possibly facial features from an uncontrolled background in order to extract the face from an unknown scene. A large number of techniques have been proposed for the purpose.

In early 1970s, heuristic and anthropometric techniques [1] were utilized for face detection but due to large assumptions, a small change in the image would lead to the tuning or redesigning of the system. Despite several obstructions and the slow pace of growth in research, video coding and face recognition have started to become a reality. There have been several developments in the important aspects of face detection over the past decade. More robust segmentation techniques using color and motion have been developed. In addition to the segmentation techniques, there have been advancements in contour and templates which can track features more accurately. Feature based techniques used for detecting faces can be classified into three broad categories: low level analysis, feature analysis and active shape models. As the most primitive feature, edge representation was applied in face detection by Sakai et al. [1]. Besides edges, gray information can also be exploited to various parts on the face. A number of algorithms search for a local gray minima in the extracted face region [2]. Another convenient way of extracting the face from the scene is motion. Motion segmentation using silhouettes and Active contours can also be effectively used for face detection. Both these algorithms are the focus of this paper and hence, they have been discussed in detail below.

Motion segmentation is an extremely efficient approach capable of recognizing a moving foreground efficiently regardless of the background content. Motion segmentation is commonly achieved directly by frame difference using silhouettes. Moving silhouettes, including face and body parts are extracted on the basis of threshold value of accumulated frame difference. The method works on the principle of silhouettes generated by timed motion history image (tMHI). It is extremely smooth to extract face from the scene if video is available with the help silhouettes. Almost any silhouettes generation method including stereo disparity also known as stereo depth subtraction [3], infra-red back-lighting [4], frame differencing [5], color histogram back-projection [6], texture blob segmentation, range imagery foreground segmentation, etc can be used for generating silhouettes of the object of interest. We choose a simple background subtraction [7] method for the

Manuscript received June 13, 2012; revised August 10, 2012.

Amarjot Singh and K. V. Karthik are with National Institute of Technology, Warangal, 506004, India (e-mail: amarjotsingh@ieee.org).

Shivesh Bajpai is with Indian School of Mines Dhanbad, Jharkhand, India.

DOI: 10.7763/IJCTE.2013.V5.644

silhouettes generation. Another approach to spot the features is the active shape model. Unlike the previous models, this model focuses on the actual physical and hence higher-level appearance of features. The algorithm interacts with local image features (edges, brightness), gradually deforming into the shape of the facial feature. The first type uses a generic active contour called snakes introduced by Kass et al. in 1987 [8]. Further need for accuracy motivated the development of deformable templates [9] taking into account a priori of facial features. The third version [10] termed as smart snakes is a generic flexible model which provides the best efficient interpretation of the human face, possible by this method. The remaining paper can be classified into following sections. Next section explains the algorithms explained and compared in this paper in detail with formulation. Section III presents the results obtained by both the algorithms on various databases with the factors affecting their performance. Section IV describes the applications of both the algorithms to face detection systems. The last section V, concludes the results.

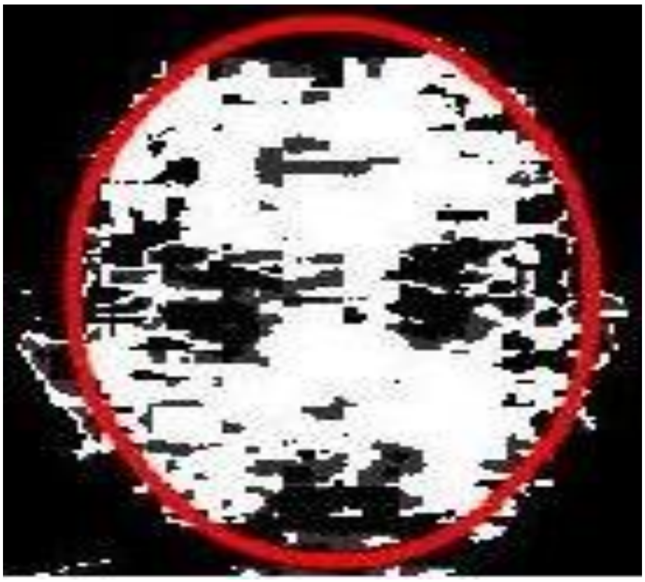


Fig. 1. Face extracted by silhouette segmentation technique

II. Algorithms

The algorithm used is described as below.

A. Active Contour Model

Unlike the previous described face detection techniques, this method aims to depict the actual high level appearance of features. They are commonly used to locate head boundary or edges [11]. The task is achieved by initializing the snake in the nearby proximity or region around the head. The snake gives the actual boundaries if released within approximate boundaries. The snake initialized converges onto the edges and subsequently assumes the shape of the head. The algorithm locks onto the features of interest which include mainly lines, edges or boundaries. The progress of traditional snake function $X(s)=[x(s),y(s)]$ for s ϵ $[0,\ 1]$ that moves through the spatial domain of an image $I(x,y)$ is obtained by minimizing the energy function given in equation 1,

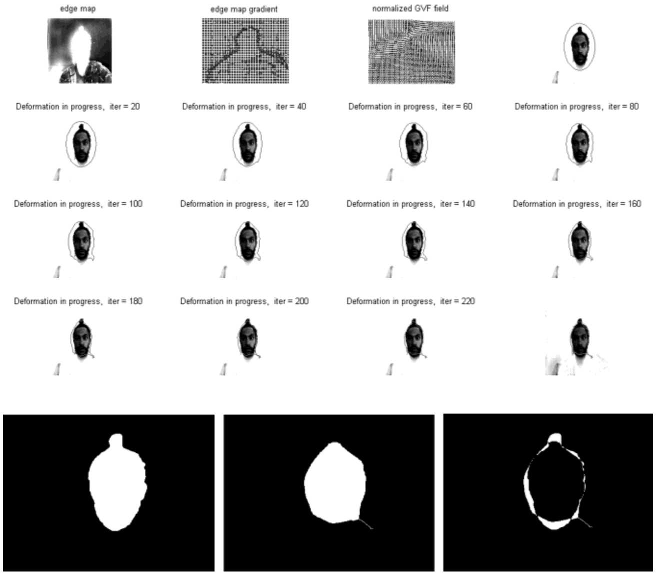

edge map
edge map gradient
normalized GVF field
Deformation in progress, iter = 20
Deformation in progress, iter = 40
Deformation in progress, iter = 60
Deformation in progress, iter = 80
Deformation in progress, iter = 100
Deformation in progress, iter = 120
Deformation in progress, iter = 140
Deformation in progress, iter = 160
Deformation in progress, iter = 180
Deformation in progress, iter = 200
Deformation in progress, iter = 220

Fig. 2. Face extracted by gradient vector field method

$$E_{energy}=\int_0^1 E_{energy}(X(s))ds \qquad (1)$$

where E_{in} and E_{ext} are the internal and external energies respectively. Internal energy E_{in} is a combination of elasticity and strain energy defined as

$$E_{in}=\frac{\alpha\left|\frac{dX(s)}{ds}\right|^2+\beta\left|\frac{dX(s)}{ds}\right|^2}{2} \qquad (2)$$

It is used to control the snake tension and rigidity. α represents elasticity while β stiffness in snake's internal energy.

The external energies which force the snake towards the edge are elaborated as

$$E_{ext}(x,y)=\left|-\nabla I(x,y)\right|^2 \qquad (3)$$

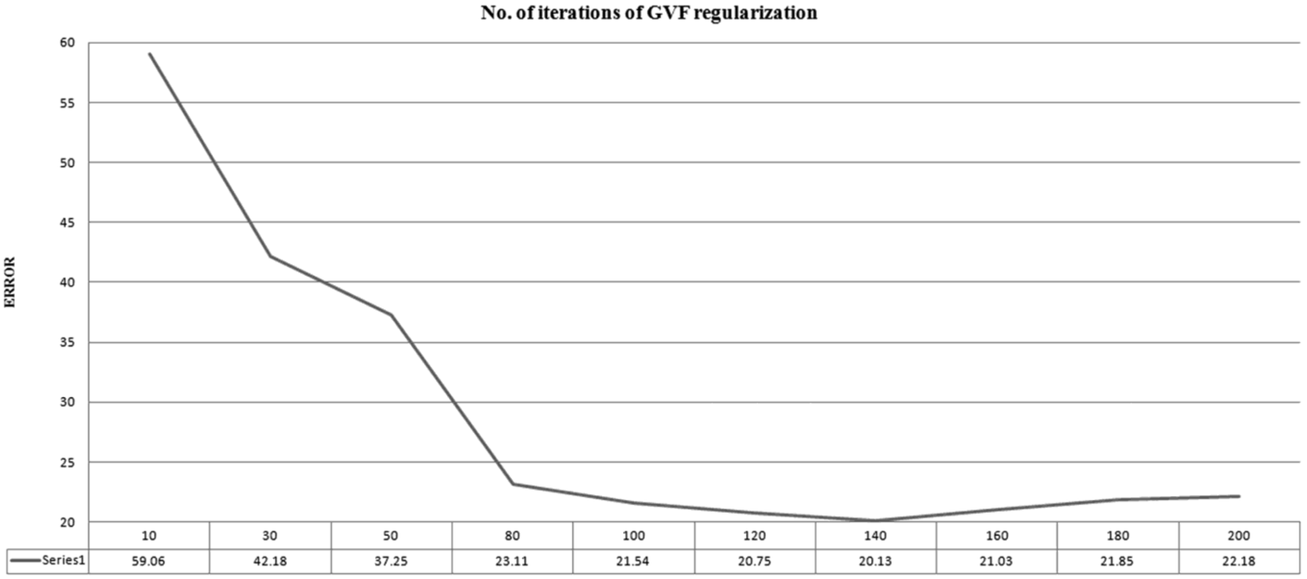
$$E_{ext}(x,y)=\left|-\nabla G_{\sigma}(x,y)*I(x,y)\right|^2 \qquad (4)$$

where $G_{\sigma}(x,y)$ is a two dimensional Gaussian function with standard deviation σ and ∇ as the gradient operator respectively. Minimization of energy provides the necessary force required for the snake to shrink to the succeeding position. This intimates that the internal energy is equal to the force by image gradient. It is expressed as

$$E_{energy}=\alpha X'(s)-\beta X^{iv}(s)-\nabla E_{ext}=0 \qquad (5)$$

$F_{in}=\alpha X'(s)-\beta X^{iv}(s)$ is defined as the internal force that tries to stop the stretching and bending while $F_{ext}=\nabla E_{ext}$ is the external force that tries to pull the snake towards the desired image edge. After the force balance

$$F_{in}+F_{ext}=0 \qquad (6)$$



	10	30	50	80	100	120	140	160	180	200
Series1	59.06	42.18	37.25	23.11	21.54	20.75	20.13	21.03	21.85	22.18

Fig. 3. Error variation with respect to the number of iterations in GVF method

Finally, the snake takes the shape of the object and the energy becomes minimum. The initialization of the snake is a difficult task as we have to acquire a suitable set of parameters as it is essential for the initialization process. It is difficult to automatically generate the set of parameters for the objects of interest. Hence these constants are decided by the user. Once the parameters are decided correctly and the snake is released in close proximity to the object, the face can be extracted successfully. It is an efficient method used in a number of applications requiring face detection.

B. Segmentation Using Silhouettes

In case of video, motion is the easiest way to extract moving objects from the scenes. Although there is recent work on more sophisticated techniques of background subtraction, we use a fast and simple method for the face extraction. In this method, the pixels which are a set number of standard deviations from the mean RGB background are labeled as foreground. In the next step, noise is removed by pixel dilation and region growing methods which finally leads to the extraction of the silhouette.

One of the simplest background model assumes that the brightness of every background pixel varies independently according to the normal distribution. The background features can be calculated by accumulating several dozens of frames, as well as their squares. The mean and standard deviation can be calculated as below

$$m(x,y)=\frac{t(x,y)}{N} \tag{7}$$

$$W(x,y)=\sqrt{\frac{t_q(x,y)}{N}-\left(\frac{t_q(x,y)}{N}\right)^2} \tag{8}$$

where N is the total number of frames collected. The pixels in positioned in certain frame are grouped to the moving object if

$$\left|m(x,y)-p(x,y)\right|>CW(x,y) \tag{9}$$

where C is equal to 3 according to the famous three sigma rule.

This algorithm is extremely efficient and extracts the pixels which vary or move with respect to a still background. As, here we are concerned only with the face, the remaining parts of the body are not of our interest. But, in case of human extraction from scenes, the connected portions to the moving object can be extracted by creating a mask to segment moving region.

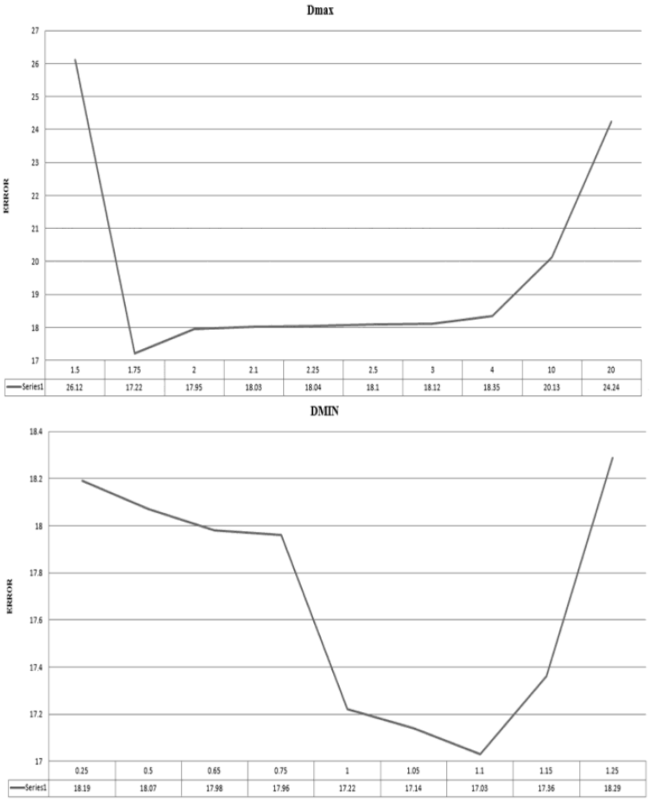
III. Results

The results obtained from the simulations enables us to investigate the capability of both the methodologies applied to synthetic and real time datasets. This section elaborates and compares the results obtained for face detection and extraction using active contour and silhouettes techniques tested on a pentium core 2 duo 1.83 ghz machine. the algorithms are compared on the basis of runtime analysis and the accuracy of face detected, compared across a common ground truth. this section further emphasizes on the factors affecting the performance of the algorithms, followed by their advantages as well as disadvantages. The results obtained for face detection using silhouette segmentation technique are shown in Fig. 1.

The tests are performed on the face of the author as shown in the figure mentioned above. A video sequence was given as input to the system. The method successfully detected and finally extracted the face of the author according to the algorithm described above. The whole process computes in less than 1 second. The algorithm works well for all kind of inputs. Silhouettes perform the task of recognition with high accuracy hence we don't compute the error for this methodology.

The simulation results for active contour method are shown in Fig. 2. The algorithm performed extremely well on A.R. database and ORL database with an accuracy of 97.88 % and 96.50% respectively. Fig. 2 shows the step by step evaluation of extraction on the input image. The method effectively extracts the author’s face in 230 iterations. The method takes 31 seconds for the whole process. The section also focuses on the other parameters responsible for the variations in the results on the basis of the error incurred by the specific parameter. As, the parameter have to adjusted manually, the accuracy mainly depends upon the selection of the parameters. Error is defined as the total number of pixels unclassified in the silhouette over the total number of pixels in the silhouette. Fig. 3 shows the variations on error with respect to normalization iterations. The minimum error is 20.13 for 140 iterations and it increases to 59.06 as the iterations drop to 10. Fig. 4 shows the variation in error with respect to the variations in alpha which is the elasticity parameter. The error varies from a maximum error of 17.61 for alpha equal to 0.06 to 17.03 for 0.05 alpha value. Another important parameter which affects the performance of snake is the initialization distance between the two points. The maximum error in the result is 20.24 for a maximum distance of 20 between initialization points specified for the snake initialization while an error of 18.29 for a minimum distance of 1.25. The error with respect to rigidity parameter beta varies from 17.03 to 17.71 for beta value equal to zero and 0.2 respectively. Similarly the maximum errors and minimum errors incurred by varying viscosity parameter (¥) and external force weight (ƙ) are 44.45 and 17.03 for viscosity parameter while 27.9 and 17.03 for external force weight respectively. The accuracy also depends on the number of iterations. The error variations with respect to the number of iterations are shown in Fig. 3. The error variations with respect to other parameters mentioned above are shown in Fig. 4.

The Parameters which give the best results are stated below. Alpha value should be set to 1.75, beta to 1.1 and gamma to 1. DMAX should be set to 1.75, DMIN to 1.1 and kappa to 0.6.



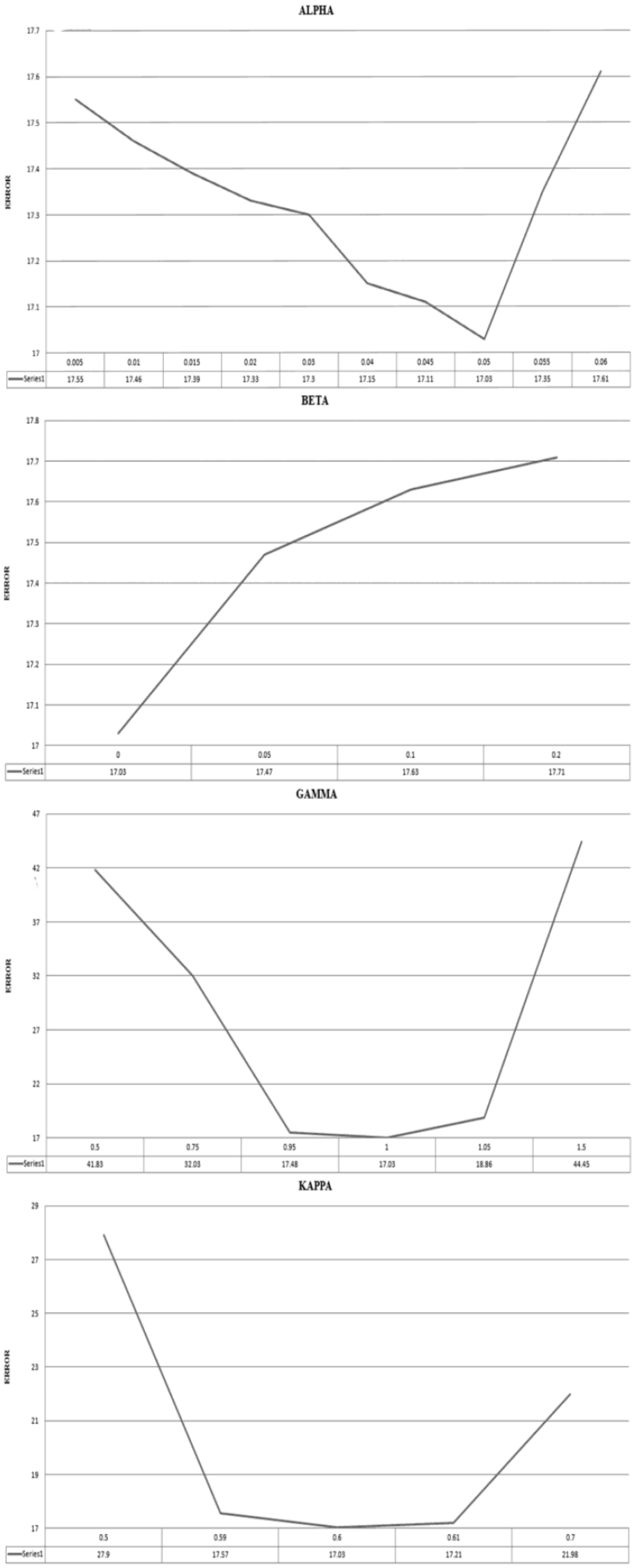


Fig. 4. a) Error variation with respect to the parameter Dmax b) Error variation with respect to the parameter Dmin c) Error variation with respect to the parameter α d) Error variation with respect to the parameter β e) Error variation with respect to the parameter γ f) Error variation with respect to the parameter κ

IV. Application to Face Detection

Face detection has huge potential as it is widely range of applications including biometric identification, video conferencing, video coding and indexing of images and video databases. With an increasing availability of images on the world wide web, face has become an important part for content has been used as a part of image search engine and has been used in webSeer [12] to mine the required content from based image retrieval. Neural network face detection system terabytes of video libraries. Face detection systems are also widely used in video conferencing systems where it is necessary to focus on the face of the speaker [13]. Human faces are mainly extracted using features such as skin, motion, contour geometry, facial analysis etc.

V. Conclusion

The section presents two most popular algorithms used for face detection, together with a brief presentation of some of the application areas. The results obtained from real time simulation by active contour method and silhouettes segmentation applied on real time data were presented in the paper. The following presents a concise summary with conclusions related to the performance of the algorithms.

Silhouettes perform much faster as compared to gradient vector field with higher accuracy. The silhouettes take less than 1 second to compute the result while on the other hand Gradient vector field takes 31 seconds to produce the results. The parameters of gvf has to be controlled manually which is time consuming process and may lead to error in the results, while silhouettes is totally automatic leading to higher accuracy. GVF gives better results when applied to images in controlled background. Another disadvantage of gvf is that it yields better results only on frontal image while any kind of image will give equally good results for silhouettes technique. Gradient vector field method fails to work on images which are very close to the image frame which again proves the superiority of silhouettes technique over gvf. On initialization of the gvf snake, it is released in close proximity to the face. As the face is very near to the frame, the snakes touches the frame of the image during expansion and compression, leading to an error. Silhouettes are much better in terms of accuracy as well as speed. Face detection is an important research area and the field has advanced into more complex and sophisticated areas which are still expensive to be implemented on a real platform, but it is likely to change with the upcoming advancement in computer hardware. Face recognition is still a very crucial area when it comes to face detection. Due to inexpensiveness of the technology, it is commonly used for CBIR systems, web search engines and digital video indexing. In spite of the advancements, accurate detection of facial feature such as corners of the eyes and mouth is still a hard problem to be solved.

References

[1] T. Sakai, M. Nagao, and T. Kanade, “Computer analysis and classification of photographs of human faces,” in *Proc. First USA Japan Computer Conference*, 1972, pp. 2-7.

[2] P. J. L. V. Beek, M. J. T. Reinders, B. Sankur, and J. C. A. V. D. Lubbe, “Semantic segmentation of videophone image sequences,” in *Proc. of SPIE Int. Conf. on Visual Communications and Image Processing.*

[3] D. Beymer and K. Konolige, “Real-time tracking of multiple people using stereo,” *IEEE FRAME-RATE Workshop*, 1999

[4] J. Davis and A. Bobick, “A robust human-silhouette extraction technique for interactive virtual environments,” in *Proceedings Modelling*, 1998

[5] G. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*. [Online]. Available: http://www.developer.intel.com/technology/itj/q21998/articles/art2.htm

[6] C. Bregler, “Learning and recognizing human dynamics in video Sequences,” in *Proceedings Computer Vision And Pattern Recognition*, June 1997, pp 568–574.

[7] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric model for background subtraction,” *IEEE FRAME-RATE Workshop*, 1999

[8] M. Kass, A. Witkin, and D. Trezopoulos, “Snakes: Active Contour Models,” *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321- 331,1987.

[9] L. Yuille, P. W. Hallinan, and D. S. Cohen “Features extraction from faces using deformable templates,” *International Journal of Computer Vision*, vol. 8, pp. 99-111, 1992.

[10] T. F. Cootes and C. J. Taylor, “Active shape models-smart snakes,” In *Proceedings of British Machine Vision Conference*, 1992, pp. 266-275.

[11] X. Chenyang and L. J. Prince, “Snakes, shapes and Gradient vector Flow,” *IEEE Transaction on Image Processing*, vol.7, no. 3 pp. 359-369, March 1998.

[12] C. Frankel, M. J. Swain, and V. Athitsos, “WebSeer: An Image Search Engine for the World Wide Web,” Technical Report, Computer Science Department, Univ. of Chicago, pp. 96-14, 1996.

[13] C. Wang, S. M. Griebel, and M. S. Brandstein, “Robust automatic video - conferencing with multiple cameras and microphones,” in *Proc. IEEE International Conference on Multimedia and Expo*, 2000.

Amarjot Singh is a Research Engineer with Tropical Marine Science Institute at National University of Singapore (NUS). He completed his Bachelors in Electrical and Electronics Engineering from National Institute of Technology Warangal. He is the recipient of Gold Medal for Excellence in research for Batch 2007-2011 of Electrical Department from National Institute of Technology Warangal. He has authored and co-authored 48 International Journal and Conference Publications. He holds the record in Asia Book of Records (India Book of Record Chapter) for having "Maximum Number (18) of International Research Publications by an Undergraduate Student". He has been awarded multiple prestigious fellowships over the years including the prestigious Gfar “Research Scholarship” for Excellence in Research from Gfar Research Germany and “Travel Fellowship” from Center for International Corporation in Science (CICS), India. He has also been recognized for his research at multiple international platforms and has been awarded 3rd position in IEEE Region 10 Paper Contest across Asia-Pacific Region and shortlisted as world finalist (Top 15) at IEEE President Change the World Competition. He is the founder and chairman of Illuminati, a potential research groups of students at National Institute of Technology Warangal (Well Known across a Number of Countries in Europe and Asia). He has worked with number of research organizations including INRIA-Sophia Antipolis (France), University of Bonn (Germany), Gfar Research (Germany), Twtbuck (India), Indian Institute of Technology Kanpur (India), Indian Institute of Science Bangalore (India) and Defense Research and Development Organization (DRDO), Hyderabad (India).