ANALYSIS OF 3D FACE RECONSTRUCTION

M. Ramasubramanian[1], M.A. Dorai Rangaswamy[2]

[1]Research Scholar, Aarupadai Veedu Institute of Technology
[2]Research Supervisor , Vinayaka Missions University
Email: [1]mailtoraams@gmail.com, [2]drdorairs@yahoo.co.in

Abstract

The 3D shape reconstruction, from 2D images is an converse problem, and is therefore mathematically poorly posed. The solution to 3D shape reconstruction problem is to using a model based approach. Our paper presents an analysis by synthesis method for solving3D face reconstruction problems, using the anatomical landmarks and intensity from 2D frontal face images. To improve the quality of 3D shape reconstruction process we incorporate a number of steps in analysis by synthesis framework. First of all, we approached the 3D model construction problem by using rigid and non rigid surface registration. Secondly, we simplified the shape estimation by using multidimensional amoeba optimization to optimize shape parameters while mapping texture directly using 3D-2D alignment. Thirdly, we evaluate the quality of the 3D shape reconstruction in the context of 3D shape error as well as by visual analysis.

Keywords: Generic 3d models, morphing, animation, texture etc.

I.   INTRODUCTION

The three dimensional face reconstructions from the 2D face image have attracted a lot of recent attention due to its wide range of applications in biometrics etc. Among those, the outstanding applications are face recognition and expression analysis. Most of the literature surveys on face recognition deals with 2D intensity images rather than 3D face models. However, the 2D image of a human face changes for many reasons including the expressions, head pose, illumination, aging, makeup etc. Due to 3D nature of face, the pixel intensities in face images changes a lot more with changes in head pose and illumination than with a change in identity of the person. The advantage of using a reconstructed 3D face model for biometrics is that such a model can be reconstructed offline only once, and can then be used under varying conditions subsequently.

There are a various number of approaches for 3D face reconstruction including shape from shading [2], shape from silhouettes [3], shape from motion, and analysis by synthesis using morphable models. At present the approach which is most successful in reconstructing 3D face models is analysis by synthesis method. In this approach, after suitable alignment, and illumination estimation the parameters of a 3D statistical shape model are adjusted to optimize the match between the projected face shape and the targeted face image. The key source of information for reconstruction of a 3d shape from a 2D image is the intensity based features defining the texture and anatomical landmarks. The use of intensity information alone can be problematic in cases where input images have low contrast, illumination variation, shadows etc. Intensity based features such as edges can be unstable because edge detectors such as the Canny edge detector return spurious edges. Furthermore, the use of landmarks alone is problematic because landmarks on the face surface are sparse and difficult to locate as face images are mostly smooth [5].

There are some authors that use only anatomical landmarks for 3D reconstruction [6,7], and argue that landmarks are a more robust source of information for 3D shape estimation than intensity information. Some authors used an intensity based objective function [8], with stochastic Newton optimization which is computationally expensive. Other systems have been enhanced to incorporate additional information such as edges and specular reflection in the objective function, and used the Levenberg Marquardt optimisation for more robust fitting [9]. The main sources of error in reconstructing 3D shape from 2D images using an analysis by synthesis approach are caused 3D-2D alignment, illumination differences, shape differences and the quality of the dense correspondence between the 3D surfaces. There can be additional errors from 3D shape space due to the mesh resolution, the quality of the input 3D surfaces
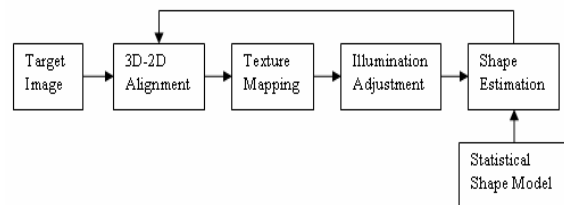

Fig1: Overview of the overall system

and the effectiveness of the optimization method. At present, most of the research on 3D face reconstruction from 2D images uses Cyberware laser scans as the quality of 3D surfaces is an important factor in 3D reconstruction. Current state of the art methods in 3D reconstruction from a single 2D image use 2D face recognition results based on synthetically rendered images from reconstructed models and visual inspection as measure of quality of reconstruction. However, it is clear that visual and the indirect measures are not accurate enough to measure quality of 3D shape reconstruction. 3D shape reconstruction techniques have been proven to

be effective for face recognition and expression analysis in an offline manner. However, texture can be deceptive when rendered on top of a 3D shape. We present results from samples where the corresponding 3D models of the subject are available to demonstrate the accuracy of 3D reconstruction. We proposed an approach to reconstruct 3D face models from frontal 2D face images using an analysis by synthesis approach. Our approach evaluates use of landmarks and intensity information for shape reconstruction. However, rather than estimating texture parameters[9,10] we directly apply texture from the target image onto the generic 3D model.

This helps in making the shape estimation function more robust as it avoids many of the errors associated with texture estimation. We use amoeba optimization as it is more robust and well suited to shape estimation when optimizing over the parameters of statistical shape model. Our system was tested using the Notre Dame 3D face database [11]. The entire cycle comprises establishing a dense 3D correspondence, 3D-2D alignment, illumination adjustment and shape estimation. The experimental results compare the quality of 3D reconstruction using three different schemes with different number of landmarks, intensity information.

II.   3D FACE RECONSTRUCTION

We now describe in detail the different steps in our approach which is illustrated in figure 1. It makes use of a dense 3D surface correspondence between objects, and involves alignment, illumination correction, texture mapping, and shape estimation using non linear optimisation.

A.   NOTRE DAME 3D FACE DATABASE

The main source of data we used in our experiments is the Notre Dame 3D face database, which comprises 278 different subjects, 166 of them having both gallery and

probe images [12]. The gallery and probe images were acquired with a four-week gap. The database consists of frontal scans with neutral expressions, having two scans per subject obtained within 11 to 13 weeks from each other. The dataset contains subjects from different origins including Caucasians, south Asian, east Asian, black with a fair mix of both the male and female population. The subjects were imaged using a Minolta Vivid 900 range scanner; this scanner used structured light reflected from the surface to infer 3D shape. The subjects were imaged approximately at 1.5 meters from the camera, against a plain gray background, with one front-above-centre spot light lighting their face. The scanner also captures a color image simultaneously with the range data which is registered automatically onto the corresponding surface.

B.   3D-3D CORRESPONDENCE

3D surfaces obtained using current technology present many challenges to statistical shape modeling (SSM). A raw surface may have holes, spikes etc. Raw surfaces also have different alignment and surface areas in addition to having a different numbers of points in the surface representation. The first step in any technique using SSM is to normalize the surface with respect to orientation and surface area. The second step is then to establish point to point correspondence between the surfaces in the input database. As a result of these steps extraneous portions of a surface were removed so that all the surfaces have similar area and are represented by the same number of corresponding vertices. The dense correspondence implies that each point on a surface represents the same anatomical position on all the other surfaces. Thus each vertex is effectively a landmark point once the correspondence is established.

Correspondence between 3D face models was established using a method proposed by

Papathoeodoros et al[13]. Initial alignment is found by using a sparse set of landmark points. These landmark points are first aligned using rigid registration. A template surface is then deformed using free-form deformation using each of the given surfaces as the target. Free-form deformation is preferred over thin plate splines because of its local nature, which results in an accurate and detailed representation of the target surfaces.

C.   3D STATISTICAL FACE MODEL

Once 3D-3D correspondence is established, every vertex is now effectively a landmarked point. This allows us to statistically encode the variation in the set of face shapes. Each face surface can be represented in the vector form as follows:

$$\big[\ \big]\big| = x_1, y_1, z_1, \ldots\ldots, x_M, y_M, z_M \text{———}(1)$$

A statistical shape model is then obtained by performing PCA on the normalized face space, and is given as:

$$\chi = \overline{\chi} + \sum_{i=0}^{t} p_i \sqrt{\lambda_i}\, \varphi_i \text{ ———}(2)$$

Where p is the parameter vector, and is expressed in units of standard deviations. The value of p for 3D face models normally varies within ±3 standard deviations. To synthesize new examples from the model, the parameter the vector p is varied (using first t modes) using a suitable optimization strategy. Each of the shape parameters corresponds to a mode of variation in the face space e.g face dimension, nose shape etc. Most of the variation in the 3D face model has been shown to come from the eyes, nose and lips[14]. At the same time, most of the face area is taken up by the cheek and forehead.

This has motivated the use of a segmented morph able model for shape estimation [15,26,27]. 90-95% of shape variation in our statistical shape model is explained by just 14-24

modes, and this allows us to reconstruct 3D faces using a small number of parameters. To explain 99% of the shape variation 74 modes are needed.

D. 3D-2D ALIGNMENT

A key step in reconstructing a 3D face from 2D face image is to determine the position, orientation and relative size of 3D face model in order to align its rendered image with the target image. This is again a very difficult problem and is referred to as head pose estimation [16] and 3D-2D registration [17]. Various approaches to solve this problem either use intensity or facial features (landmarks). We experimented with a number of different landmark configurations and after a series of experiments propose using up to 14 landmark points for aligning a mean face model with the target image. The landmarks being used for alignment need to be stable and easy to identify. We found that landmarks around the eyes (corners, centre), nose, lips and chin are the most suitable for aligning a 3D face model with a frontal or near frontal 2D face image.

The landmarks around face boundary are not being used for alignment. The mean 3D face in the shape model, and target image are landmarked as shown in figure 2. Once land marking has been completed it is possible to align the 3D image such that its projected landmarks match those of the 2D image. We do this by optimizing over translation, rotation and scaling parameters using amoeba optimization. We assume perspective projection throughout our work. Reconstruction accuracy is not necessarily dependent on accurate estimation of intrinsic camera parameters [6,15] when estimating shape for a specific class such as face. However it is still affected by alignment between projected 3D model and target 2D image especially if we are using intensity values for estimating shape. Let L3D be a landmark points on the mean surface, and Linp be the

corresponding landmark points on the input image. First, L3D will be projected using some initial camera parameters giving L2D. The Euclidean distance between the two is minimized by optimizing over the camera parameters as follows:

$$T(\beta) = \min EU(L_{2D}, L_{inp}) \text{ ———— (3)}$$

$$\text{where, } L_{2D} = P(L_{3D}, \beta) \text{ and } \beta = (T_x, T_y, T_z, R_x, R_y, R_z, S)$$

Here T is minimized by finding the optimal camera parameters using perspective projection P , where EU represents the Euclidean distance between the two sets of landmarks.
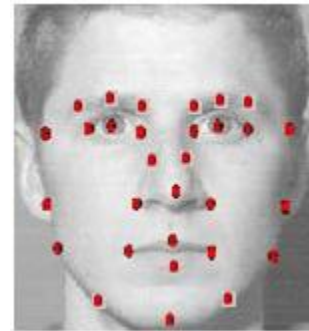

Fig 2:. Selection of Landmarks

E. TEXTURE MAPPING

After the alignment of the landmarks, texture from the target image is directly mapped onto the targeted 3D model using correspondence between projected vertices and pixels on the target face image. The accuracy of the texture mapping depends on both the spatial alignment, correspondence between shape. Both of these will be improved as we obtain an estimate of the target shape in an iterative scheme. An obvious advantage of texture mapping over texture synthesis, allows us to overcome difficulties that arise when a subject belongs to a different race having different skin colour and different facial features.

F. ADJUSTING ILLUMINATION

Then, shape is estimated by optimizing over the parameters of the statistical shape model.

However, shape estimation requires that we have spatial alignment as well as intensity alignment. Once spatial alignment is established, we adjust ambient, diffuse and specular properties of the 3D face model as follows:

$$T(\mid, \circledR) = EU(I_p, I_{inp}) \quad\text{———}\quad (4)$$

Here | represents the illumination (ambient, diffuse and specular properties) parameters and ®represents the camera parameters. The optimal transformation T is obtained by globally adjusting the model properties to align the intensities of the projected image (p I) with the target image (inp I). We have used phong illumination for photorealistic rendering.

G. SHAPE ESTIMATION

After performing spatial and intensity alignment, the major difference left between the projected image and the target image is the shape difference. This shape error is then minimized by optimizing over the shape parameters of statistical shape model. Shape reconstruction is a non-convex optimization problem involving many local minima. To reach the true minima, it is very important to initialize parameters close to true solution in addition to applying suitable constraints on the parameters. Rather than using a complex optimization scheme such as the stochastic Newton optimization or a different feature space (eg edges maps) we chose the simplex optimization. Downhill simplex is a multidimensional optimization algorithm described by Nelder and Mead[18]. In addition to being simple and robust this optimization method is geometric in nature and hence is more suitable for shape analysis. During each optimization step, downhill simplex can try a number of things such as reflection, expansion, contraction etc in one or more dimensions. These steps exploit the ability of the given statistical shape space to model the 3D shape from the target image by changing the 3D shape in different ways. Alignment and illumination estimates are iteratively updated for the current state of the shape model. The resulting camera, lighting and shape parameters are then used for shape estimation. Each of the iteration involves updating shape parameters using the distance between the projected face image and the target image as follows:

$$D(\alpha, \beta, \eta) = D_1(I_p, I_{inp}) + D_2(L_{2D}, L_{inp}) \quad\text{———}\quad (5)$$

D1 and D2 represent suitable distance measures between the pixel intensities and the corresponding landmark points on the two images respectively. The cost function (equation 5) uses the Euclidean distance between landmarks and the image intensity to optimize over the shape parameters (<) for the given camera and illumination parameters ((R),I). Here <is restricted within +/-3 standard deviations. The overall approach is as follows:

Step 1: Initial alignment between the mean surface and the target face image.

Step 2: Illumination adjustment

Step 3: Random initialization of the shape parameters

Step 4: Iterative update of alignment and illumination estimates

Step 5: Shape estimation using equation 5.Step 5 is terminated prematurely if optimization is not converging and restarted after repeating step 4.

III.   RESULTS & DISCUSSION

A set of 30 images and corresponding 3D models were chosen from the Notre Dame 3D face database. These images were landmarked with 30 points as shown in figure 2. The landmarked points on the inner parts of the face (21 in total) are used for alignment between 3D model and 2D image. Shape estimation is performed using two different configurations of landmark points. These are with or without the addition of the face boundary landmark points.

For shape estimation landmarks representing eye centre were not used. We have presented the results for shape estimation using three different approaches (1) using only the landmark points, (2) using both landmarks and intensity, (3) and using only the intensity. These experiments demonstrating the usefulness of landmark and intensity based information in 3D shape reconstruction. In figure 3, x-axis represents different subjects, where as y-axis shows the Euclidean distance between shape parameters. Average distance in the figure 3 represents distance of a subject in terms of shape parameters from all the other subjects in 3D face database. These results have been obtained using only first 21 modes of shape variation. We compare the results of four different schemes against the average. The landmarks only scheme uses 16 landmarks which includes four points on the face boundary. On the whole all schemes perform better than average for most subjects. However the landmarks only scheme gives worst performance among the four, we attribute this to lack of information. We found that we got the best performance when we used landmarks on inner part of face along with image intensity. This indicates that landmarks alone are not sufficient because they are sparse, whereas the accuracy of shape estimation is improved when image intensity was used with a selection of landmark points on the inner part of the face. These experiments were repeated at different mesh resolutions and it was found that system performs best when every fourth vertex is used for shape estimation. figure 4 gives a visual analysis of the performance of the system.

IV.   CONCLUSIONS & FUTURE WORK

In this paper, we presented a 3D face reconstruction technique for frontal face images. Our aim is to develop methods that can reconstruct 3D face models from more 2D images with certain degree of accuracy without the aid of specialized hardware and restrictive conditions. The technique presented in this paper is different from other similar methods to best of our knowledge in a number of ways:

1) 3D correspondence was established using rigid followed by non-rigid surface registration.

2) Alignment and shape estimation are interleaved in an iterative scheme.

3) Method uses both landmarks and intensity.

4) We rely on statistical model of 3D face shape only, whereas morphable models involves statistical model of both 3D face shape and texture and active appearance models[19] build statistical models of shape and texture for 2D images.

5) Shape estimation is based on the multidimensional amoeba optimizer. Furthermore, we provide a quantitative analysis of quality of 3D reconstruction. This has been discussed in [20-25]. By quantitatively assessing the 3D face reconstruction problem, we feel that the performance of the system can be further improved by a number of measures including:-
   - Using segmented models because most of the shape variation is in centre of the face.
   - Improving the quality of the initial 3D models and performance of the dense 3D correspondence.
   - Finding a cost function which is more suitable for shape estimation.
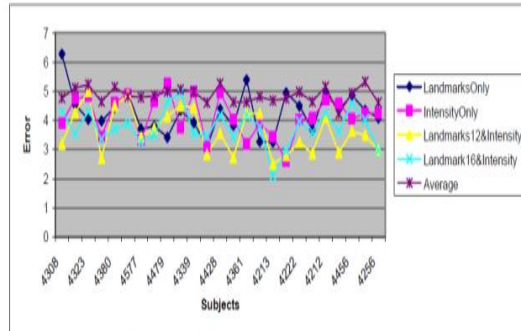   - Employing automatic landmarking

Figure 3. Shape estimation under different schemes.
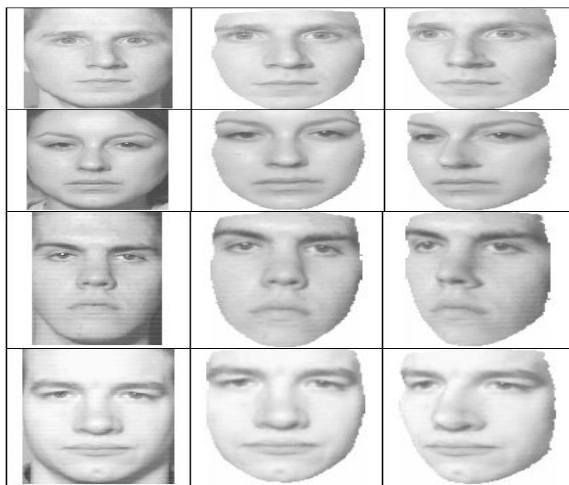
Fig 3: Shape estimation under different schemes.

Fig4: From left to right, first column shows the original 2D images, the 2nd and 3rd columns show renderings of the reconstructed 3D models.

REFERENCES

[1]   S. Hassan Amin, Duncan Gillies "Analysis of 3D Face Reconstruction" 14th International Conference on Image Analysis and Processing (ICIAP 2007)

[2]   R. R. Zhang, "Shape from shading: A survey," IEEE transactions on pattern analysis and machine intelligence, vol. 21, pp. 690-706, 1999.

[3]   B. Moghaddam, J. Lee, H. Pfister, and R. Machiraju, "Model-Based 3D Face Capture with Shape-from-Silhouettes," presented at Analysis and modeling of faces and gestures, Nice, France, 2003.

[4]   Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. Zhang,"Automatic 3D Reconstruction for Face Recognition,"presented at International conference on automatic face andgesture recognition, Seoul, 2004.

[5]   F. Pighin, J. Hecker, D. Lischinski, and R. Szeliski, "Synthesizing Realistic Facial Expressions From Photographs," presented at Computer graphics: SIGGRAPH 98; conference proceedings, Orlando; FL, 1998.

[6]   D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao, "Efficient 3D reconstruction for face recognition," Pattern Recognition, vol. 38, pp. 787-798, 2005.

[7]   B. Volker and V. Thomas, "Face Recognition Based on Fitting a 3D Morphable Model," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, pp. 1063-1074, 2003.

[8]   R. Sami and V. Thomas, "Estimating 3D Shape and Texture Using Pixel Intensity, Edges, Specular Highlights, Texture Constraints and a Prior," in Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2: IEEE Computer Society, 2005.

[9]   R. Sami, V. Blanz, and T. Vetter, "Face Identification by Fitting a 3D Morphable Model Using Linear Shape and Texture Error Functions," Lecture Notes in Computer Science, pp. 3-19, 2002.

[10] V. Thomas and B. Volker, "Estimating Coloured 3D Face Models from Single Images: An Example Based Approach," in Proceedings of the 5th European Conference on Computer Vision-Volume II Springer-Verlag, 1998.

[11] "Notredame 3D Face Database,http://www.nd.edu/~cvrl

[12] K. Chang, K. Bowyer, and P. Flynn, "Multi-Modal 2D and 3D Biometrics for Face Recognition," presented at Analysis and modeling of faces and gestures, Nice, France, 2003.

[13] T. Papatheodorou and D. Rueckert, "Evaluation of 3D Face Recognition Using Registration and PCA," presented at Audio- and video-based biometric person authentication: 5th International Conference, AVBPA, Hilton Rye Town, N.Y. USA, July 20-22 2005 : proceedings, Hilton Rye Town,N.Y., 2005.

[14] T. Papatheodorou, "3D Face Recognition Using Rigid and Non-Rigid Surface Registration," in Department of Computing: Imperial College London, 2006.

[15] B. Volker and V. Thomas, "A morphable model for the synthesis of 3D faces," in Proceedings of the 26th annual conference on Computer graphics and interactive techniques: ACM Press/Addison-Wesley Publishing Co., 1999.

[16] S. Srinivasan and K. Boyer, "Head Pose Estimation Using View Based Eigenspaces," presented at Pattern recognition, Quebec, Canada, 2002.

[17] M. J. Clarkson, D. Rueckert, D. L. G. Hill, and D. J. Hawkes, "Image Registration Using Photo-Consistency to Register 2D Optical Images of the Human Face to a 3D Surface Model," IEEE Transactions on Pattern Analysis and Machine Intelligence Pami, vol. 23, pp. 1266-1280, 2001.

[18] W. H. Press, Numerical recipes in C++ : the art of scientific computing, 2nd ed. Cambridge: Cambridge University Press, 2002.

[19] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—their training and application,"Comput. Vis. Image Underst., vol. 61, pp. 38-59, 1995.

[20] M. Ramasubramanian, Dr. M.A. Dorai Rangaswamy,"RECONSTRUCTION OF PRACTICAL 3D FACE MATCHING PART FROM 2D IMAGES – A CROSS BREED APPROACH", in the National Conference On Recent Approaches in Communication and Information Technology NCRACIT 2012, organized by Department Information Technology Madha Engineering College, Chennai on 20th March 2012.

[21] M. Ramasubramanian,Dr. M.A. Dorai Rangaswamy,"Reconstruction of sensible 3D Face Counterpart From 2D Images - A hybrid approach", in the International Journal of Research Journal of Computer Systems Engineering,ISSN: 2250-3005,Page No: 139-144,July,2012.

[22] M. Ramasubramanian, Dr.M.A.Dorai Rangaswamy,"EFFICIENT 3D OBJECT EXTRACTION FROM 2D IMAGES", in the National Conference On Emerging Trends In Computer Applications & Management NCETCAM 2013 organized by Department of Computer Application and Management of Aarupadai Veedu Insitute of Technology, Chennai on 17th, April 2013.

[23] M. Ramasubramanian, Dr.M.A.Dorai Rangaswamy,"3D OBJECT EXTRACTION FROM 2D OBJECT", in the National Conference on Architecture, Software systems and Green Computing NCASG-2013 , organized by Department of Computer Science and Engineering , Aarupadai Veedu Insitute of Technology, Chennai on 03rd April, 2013.

[24] M. Ramasubramanian, P.Shankar,Dr.M.A.Dorai Rangaswamy," 3D OBJECT CONVERSTION VIA 2D IMAGES A SURVEY REPORT", in the National Conference on Architecture, Software systems and Green Computing NCASG-2013 , organized by Department of Computer Science and Engineering , Aarupadai Veedu Insitute of Technology, Chennai on 03rd April, 2013

[25] M. Ramasubramanian, Dr. M.A. Dorai Rangaswamy,"3D OBJECT EXTRACTION FROM 2D OBJECT", in the International Journal on Emerging Trends & Technology in Computer Science, ISSN: ,Page No: ,June,2013.

[26] M.Ramasubramanian, P. Shankar,Dr.M.A.Dorai Rangaswamy," 3D OBJECT CONVERSTION VIA 2D IMAGES A SURVEY REPORT", in the International Journal on Emerging Trends & Technology in Computer Science, ISSN: Page No: ,June,2013.

[27] M. Ramasubramanian, Dr.M.A.Dorai Rangaswamy," EFFICIENT 3D OBJECT EXTRACTION FROM 2D IMAGES ", in the INTERNATIONAL CONFERENCE on Intelligent Interactive Systems and Assistive Technologies IISAT-2013, Coimbatore on August June,2013.

Mr. M. Ramasubramanian worked in the school of computer science and engineering CEG, Guindy, for 4+ years. He received his M.E. degree in the field of Computer Science and Engineering from Vinayaka Missions University in the year of 2009. He is presently working as a Senior Assistant Professor, in Aarupadai Veedu Institute of Technology,Vinayaka Missions University, India. He is doing his Research in the area of Image Processing in the same University, under the guidance of Dr.M.A. Dorai Rangaswamy.