Automatic Extraction of Facial Regions using CLM for the Application of Face Recognition

M. Annalakshmi, S.Md.Mansoor Roomi, V.Karthik

Abstract: ***Facial Image processing is an active research area which involves various applications such as face detection, face recognition, person identification and also demographic information collection that is age, gender and race from the face. In general, all these applications fall under either holistic approach or local approach. In holistic approach, the whole face image is used for further processing. In local approach, the face is divided into various blocks which may contains the facial regions like eyes, eye brows, mouth, nose, cheek and chin regions. In order to develop a robust algorithm against scale, illumination, pose and expression issues, a component based or part based approaches were developed. For component based approach, the extraction of various face parts needs an automatic method to crop the facial regions closest to the manual cropping. The method of automatically extracting various facial parts is addressed in this paper. The Constrained Local Model (CLM) approach is used to identify the facial landmarks which in turn used to segregate the different facial parts. The performance of proposed approach is evaluated by ground truth of the respective facial components. The correspondence between the automatic extracted facial regions and ground truth is evaluated by SIFT descriptor. Experimental results on matching manually cropped facial regions against automatically extracted regions show that the CLM approach achieves promising performance. The extraction of various facial region will be very useful in Face recognition.***

Keywords: ***Constrained Local Model, facial land mark localization, face region extraction, SIFT descriptor***

I. INTRODUCTION

A Human face conveys variety of information which includes person identification, age, gender and race. The existing approaches either represent faces with a holistic representation or use local features, e.g. eyebrows, eyes, nose and mouth for face related applications. An automatic face detection and analysis is a challenging problem in computer vision, and has been actively researched for applications such as face verification, face tracking, person identification, gender recognition and age estimation. The automatic facial parts extraction is necessary to perform various automatic face image processing. It is hard to acquire facial landmark locations, head pose estimation from face images suffering from extreme poses, illumination and resolution variations.

There are many automatic facial land mark localization methods such as constrained local model (CLM)-based, active appearance model (AAM)-based, Active Shape Model (ASM)-based methods. CLM-based methods utilized a shape model and patch model to detect a facial feature point.

AAM-based methods[4,5] fit a shape model to an image by minimizing texture synthesis errors. The Active Shape Model (ASM) [18,9] is an effective way to locate facial features, to model both shape and texture, and also to find correlation between them from an observed training set. Facial feature points are mainly located around facial components such as eyes, mouth, nose and chin. The proposed method is very useful for facial region extraction.

The key contributions of this work includes are

- Automatic extraction of various facial regions.
- Matching of automatic extracted facial regions against manually cropped facial regions are analyzed.

The remainder of this paper is organized as follows. In Section II the literature review is discussed. The facial parts extraction of the proposed system is discussed in Section III. In Section IV, experiments are performed to verify the effectiveness of the proposed system by matching automatic extracted face regions to manually extracted face regions. Finally, Section V includes conclusion and directions of future work.

II. LITERATURE REVIEW

Facial parts extraction was started by face detectors Viola and Jones [1], Yang et al[2] which returns a rectangular bounding box, implies the face location. Different numbers of facial feature points are labeled for various application scenarios as, a 17-point model, 29-point model or 68-point model. These points cover several frequently-used areas: eyes, eyebrows, nose, and mouth. These areas carry the important information for both generative and discriminative purposes. The accurate detection of facial feature points helps to locate the different facial components effectively.

Constrained local model (CLM)-based methods, Cristinacce and Cootes [3] consider the appearance variation around each facial feature point with the assistance of a corresponding local expert. Facial feature points are then predicted from these response maps refined by a shape prior which is generally learned from training shapes. This method outperforms the state-of-the-art RLMS fitting method and the tree-based method.

Revised Manuscript Received on July 30, 2020.

* Correspondence Author

M.Annalakshmi*, is currently working as Associate Professor in Electronics and Communication Engineering, Sethu Institute of Technology, Kariapatti,Tamil Nadu, India. E-mail: annam.baluss@gmail.com

S.Md.Mansoor Roomi, is currently working as Associate Professor in Electronics and Communication Engineering,Thiagarajar College of Engineering,Madurai, Tamil Nadu, India. E-mail: smmroomi@tce.edu

V.Karthik, is currently working as Assistand Professor in Electronics and Communication Engineering, Sethu Institute of Technology, Kariapatti,Tamil Nadu, India. E-mail: karthik2055@gmail.com

Retrieval Number: I7087079920/2020©BEIESP
DOI: 10.35940/ijitee.I7087.079920
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

Lee and Kim [4] explored the shape-normalized and fitted shape appearance of the tensor-based active appearance model (AAM) , proposed by Cootes et al [5], in which the input image is transformed into a normalized image to conduct variation robust face recognition.

AAM is applied by Stegmann et al [6] to medical image analysis. For robust shape tracking, a fusion strategy is proposed by Zhou et al [7] to incorporate subspace model constraints.

In order to separate the shape from the texture to favour the sketch generation process, Chen et al [8] applied active shape model (ASM) proposed by Cootes and Taylor [9]. In face hallucination by Wang et al [10] and facial swapping by Bitouk et al [11] facial feature point detection is an essential pre-processing step.

To control the variation of facial appearance, Weise et al [12] proposed facial animation which generally detects facial feature points. To robustly describe the variation of facial expression across different poses, the combination of 2D and 3D view based AAM is proposed by Sung and Kim, [13]. In 3D face modeling the correspondence of facial feature points plays an important role Blanz and Vetter [14]. AAM is applied by Anderson et al [15] to track robustly over a very large set of facial data with expressions and to synthesize video realistic renderings in the visual text-to-speech system.

Akshay Asthana et al [16] proposed a novelistic approach of discriminative regression, which is based on the CLM frame-work. This method is known as Discriminative Response Map Fitting (DRMF) which outperforms the state-of-the-art RLMS fitting method and the tree-based method. Ying Tai [17] proposed an Orthogonal Procrustes Regression based approach, which shows high efficiency for the misaligned test images with pose variations. Yet this method takes into considerations only the horizontal pose variation.

From the literature, the Constrained Local Model approach is more accurate and robust than the other methods like ASM and AAM approach. In order to localize the individual facial regions, the proposed method utilizes CLM method to detect facial landmarks.

The extractions of the localized facial regions are useful for many applications like demographic information including gender, age and race classification.

The success rate of the proposed system is analyzed by manual cropped facial parts using image matching algorithm. The detailed explanation of the proposed system is given in section III.

III. PROPOSED METHOD

The proposed method consists of three different modules such as face detection, facial landmark localization and facial regions extraction which are shown in Fig. 1. The input image is applied to a Constrained Local Model.

The CLM model first detects the face using Viola Jones[1] face detector. The model gives 68 feature points for the detected face.

These points include nose, eyes, eye brows, mouth and outer boundary of the face. For the ease of analysis, the face regions are cropped. The feature points obtained are then used for segregating the various parts of the face.

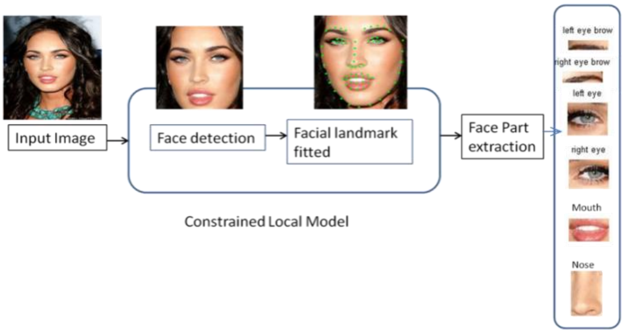


Fig. 1. Block diagram of the proposed method

A. Constrained Local Model (CLM)

Constrained Local Models is the category of part based models which exploit the local image patches taken around the landmark points or feature points. CLM model consists of two parts such as Patch Model and Shape Model. The Patch Model uses the local image around the feature points and the Shape Model describes the shape variations of feature points. A joint shape and texture model is employed in Patch Model. From the training samples, the set of grey scale vectors and normalized shape co-ordinates are obtained to create linear models as follows.

$$X = \overline{X} + O_s b_s \qquad G = \overline{G} + O_g b_g \qquad (1)$$

Where $\overline{X}$ is the mean shape, a set of orthogonal modes of variation is denoted by O_s and a set of shape parameters is represented by b_s. Similarly the mean normalised grey-level vector is defined by $\overline{G}$, a set of orthogonal modes of variation is given by O_g and b_g is a set of grey-level parameters. The shape and template texture models are combined using a PCA to produce one joint model.

$$b = O_c C \qquad (2)$$

Where $O_c = \begin{pmatrix} O_{cs} \\ O_{cg} \end{pmatrix}$ and $b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix}$

Where b is the concatenated shape and texture parameter vector, Ws is the weighting parameter, C is a set of joint appearance parameters, Oc is the orthogonal matrix computed using PCA, which divided into two different matrices Ocs and Ocg which jointly computes the shape and texture parameters given a joint parameter vector C. For each feature point, a response image is obtained in Shape Model. If high value is obtained in the response image, then the match score is high. This response image is used to determine the position of each feature point. The best position for each point from the response image is obtained, taking into account the allowed shape variation.

B. Facial Landmark Localization

The facial feature points are obtained by CLM based approach. CLM model consists of two parts; one part describes the shape variations of feature points called Shape Model, and the other part describes each patch of image around the feature points, called Patch Model.

The sample result of CLM approach is shown in Fig. 1. The feature points are numbered as

- 1 to 17 represents the outer boundary
- 18 to 27 is represents the eyebrow
- 28 to 36 represents the nose
- 31 is represents the nose tip
- 37 to 48 represents the eyes
- 49 to 68 represents the mouth

C. Facial Part Localization

The CLM approach produces the 68 fiducial points on the face region. The spatial co- ordinates of the all these 68 points are extracted.

Eyebrow extraction:

The eyebrow of the detected face is indicated by 10 fiducial points and this facial component is extracted by using their spatial coordinates. The left and right eyebrow is indicated by 5 points each. Let the bounding box of the eyebrows are

$$\left[\alpha_i^{eb} \quad \beta_i^{eb} \quad \omega_i^{eb} \quad H_i^{eb}\right]$$

$$X_i^{eb} = \left\{x_{(i-1)*5+1}^{eb}, \ldots x_{i*5}^{eb}\right\} \quad \forall i = 1,2 \text{ -- (3)}$$

$$Y_i^{eb} = \left\{y_{(i-1)*5+1}^{eb}, \ldots y_{i*5}^{eb}\right\} \quad \forall i = 1,2 \text{ -- (4)}$$

$$\left[\alpha_i^{eb} \quad \beta_i^{eb}\right] = \left[\min\left(X_i^{eb}\right) \quad \min\left(Y_i^{eb}\right)\right] \forall\, i = 1,2 \text{ - (5)}$$

$$\left[\omega_i^{eb} \quad H_i^{eb}\right] = \left[\left(\max\left(X_i^{eb}\right) - \alpha_i^{eb}\right) \quad \left(\max\left(Y_i^{eb}\right) - \beta_i^{eb}\right)\right] \quad \forall\, i = 1,2 \text{ - (6)}$$

Where i=1, 2 represents the left and right eyebrow region

Eye Extraction:

The eye region of the detected face is marked by 12 points. Each eye region is covered by 6 fiducial points. The spatial coordinates are given by,

$$X_i^{e} = \left\{x_{(i-1)*6+1}^{e}, \ldots x_{i*6}^{e}\right\} \quad \forall i = 1,2 \text{ -- (7)}$$

$$Y_i^{e} = \left\{y_{(i-1)*6+1}^{e}, \ldots y_{i*6}^{e}\right\} \quad \forall i = 1,2 \text{ -- (8)}$$

$$\left[\alpha_i^{e} \quad \beta_i^{e}\right] = \left[x_{(i-1)*6+1}^{e} \quad \min\left(y_{(i-1)*6+j}^{e}\right)\right] \quad \forall\, i = 1,2\ \& j = 2,3 \text{ -- (9)}$$

$$\left[\omega_i^{e} \quad H_i^{e}\right] = \left[\left(x_{(i-1)*6+4}^{e} - \alpha_i^{e}\right) \quad \left(\max\left(y_{(i-1)*6+k}^{e}\right) - \beta_i^{e}\right)\right] \quad \forall\, i = 1,2\ \& k = 5,6 \text{ -(10)}$$

Nose Extraction:

The nose region from the face is extracted by using spatial coordinates of 9 fiducial points which is given by,

$$X^{N} = \left\{x_i^{N}\right\} \quad \forall i = 1,2,\ldots.9 \text{ ---- (11)}$$

$$Y^{N} = \left\{y_i^{N}\right\} \quad \forall i = 1,2,\ldots.9 \text{ ---- (12)}$$

$$\left[\alpha^{N} \quad \beta^{N}\right] = \left[\frac{\left(x_5^{N} - x_5^{e}\right)}{2} + x_5^{e} \quad y_1^{N}\right] \text{ ---- (13)}$$

$$\left[\omega^{N} \quad H^{N}\right] = \left[\left(x_7^{e} - x_4^{e}\right) \quad \max\left(x_{i+4}^{N}\right) - y_1^{N}\right] \forall i = 1,2,..5 \text{ --- (14)}$$

Mouth Extraction:

The mouth region of the detected face is extracted by using spatial co ordinates of 20 fiducial points are given by,

$$X^{M} = \left\{x_i\right\} \quad \forall i = 1,2,\ldots.20 \text{ ---- (15)}$$

$$Y^{M} = \left\{y_i\right\} \quad \forall i = 1,2,\ldots.20 \text{ ---- (16)}$$

$$\left[\alpha^{M} \quad \beta^{M}\right] = \left[\min\left(X^{M}\right) \quad \min\left(Y^{M}\right)\right] \text{ ---- (17)}$$

$$\left[\omega^{M} \quad H^{M}\right] = \left[\left(\max\left(X^{M}\right) - \alpha^{M}\right) \quad \left(\max\left(Y^{M}\right) - \beta^{M}\right)\right] \text{ -- (18)}$$

From the equations (3) to (18), α and β are the starting point of the facial parts to be cropped, ω is the Width of the facial region, H is the height of the facial parts and (x,y) is the spatial co ordinates of the each facial regions which is marked in the subscript. The various facial parts are indicated by the rectangular bounding box as shown in Fig.2. The same region is extracted using given formulae.

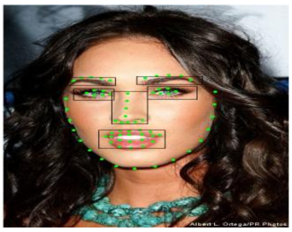


Fig. 2. Facial Part Localization

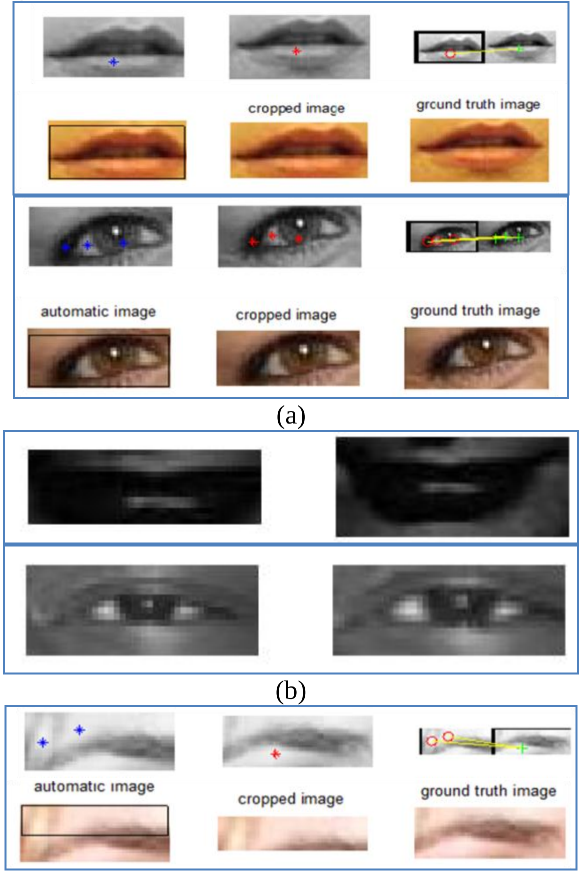
D. Face Parts Verification:

The performance of the automatic extracted facial region against manually cropped facial regions is performed by image matching technique using SIFT feature descriptor. It determines the correspondence between two image features. To understand the success rate of the automatic face parts extraction, the manual cropping face parts are used instead of state of the art method. This approach gives the best comparison result. For analyzing the effectiveness of extracted face parts, the SIFT feature descriptor is obtained from the both the automatic and manual cropped regions. Then matching points between the two images are obtained. The minimum Euclidean distance between the corresponding matching points is identified and considered as a reference point for analyzing whether the automatically extracted facial region consist of required facial part in terms of area of the region.

IV. RESULTS AND DISCUSSION

For experimentation, the images from the “Labeled Face Parts in the Wild (LFPW) Dataset” are used. This database includes the issues like scale, illumination, pose and expression variation. The extracted facial regions are verified with manually cropped facial parts whether the extracted region contains the required facial region or not. To accomplish this task, SIFT descriptor is obtained from both automatic and manual cropped image. The SIFT descriptor produces the matched points between the images. The correspondences between the images are identified from the matched points. The matched point having minimum distance is selected as best reference point for verifying the automatic extracted face region. By using matched points, the exact facial region is cropped from the extracted facial region. The rectangular bounding box has drawn inside the automatic extracted facial region shows the required facial region which is shown in Fig.5(a). The area of the automatic extracted facial region (A1) and ground truth image (manually cropped facial region) (A2) is computed.

The difference between these two areas determines the automatic extracted facial region as positive or negative. The results of image matching by SIFT descriptor is shown in Fig.5(a).



(a)

(b)

(c)

Fig. 4. (a) perfect match (b) not matched due illumination and occlusion (c) not matched due to improper feature matching point

The Fig.4(a) shows that the many of the automatically cropped face parts are closely equivalent to that of manually cropped face parts. But very few parts do not matched with the manual cropped parts even though they are visually same due to illumination and occlusion as shown in Fig.4 (b) and also due to improper matching point between the images as shown in Fig.4(c). From the 72 sample images, the performance of the SIFT feature descriptor and Human observations are obtained and the result is shown in Table 1. It is observed that the overall accuracy produced by the proposed system is above 92.59% and for different face parts are above 95% except eyebrows.

Table-I. Matching result of automatic extracted images against manual cropped images

Facial Regions	No. of Images	SIFT matching		Human observation		Accuracy (%)
		Matched	Not Matched	Matched	Not Matched	
Mouth	72	69	3	70	2	95.83
Left Eye Brow	72	59	13	70	2	81.94
Right Eye Brow	72	65	7	70	2	90.28
Left Eye	72	68	4	71	1	94.44
Right Eye	72	69	3	72	0	95.83
Nose	72	70	2	72	0	97.22

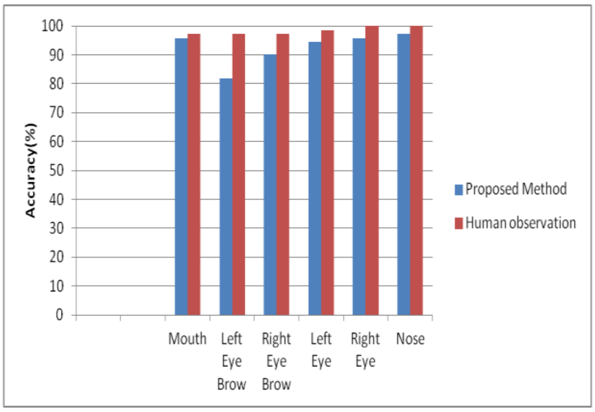


Fig. 6. Accuracy of the proposed system Vs. Human observation

V. CONCLUSION

The detailed experiment was conducted in LFPW face database to analyze the success of the proposed system. Instead of using whole face region for various applications like face recognition and demographic data collection that includes age, gender and race of a person, facial components can be effectively used. The proposed method is deployed in the face recognition application even if the whole face is not visible to the system. The part based facial recognition is achieved through the proposed method. The proposed method is used to extract the facial components automatically once the face region is localized. In order to ensure that the automatic cropped region is a required region, SIFT descriptor is applied on the automatically extracted facial regions as well as the manually cropped facial regions. The efficiency of the proposed system is compared with human observation. The proposed system produced the accuracy more than 95% for all facial parts except the eyebrows. In future, the variation like pose is sorted out by proper alignment technique prior to facial part extraction.

REFERENCES

1. Viola P and Jones M , "Robust real-time face Detection", International Journal of ComputerVision , 57(2) ,2004, 137-154
2. 2. Yang M, Kriegman D and Ahuja N, "Detecting faces in images: a survey", IEEE Transactions on Pattern Analysis and Machine Intelligence 24(1) , 2002, 34-58
3. Cristinacce D and Cootes T , "Feature detection and tracking with constrained local models", In: Proceedings of British Machine Vision Conference, 2006b , pp 929-938
4. Lee H andKim D, "Tensor-based AAM with continuous variation estimation: application to variation Robust face recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence 31(6), 2009, 1102-1116
5. Cootes T, Edwards G and Taylor C, "A comparative evaluation of active appearance model algorithms", In: Proceedings of British Machine Vision Conference, 1998b, pp 680-689

6. Stegmann M, Ersboll B and Larsen R, "FAME-a flexible appearance modeling environment", IEEE Transactions on Medical Imaging 22(10), 2003,1319-1331
7. Zhou X, Comaniciu D and Gupta A , "An information fusion framework for robust shape tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence 27(1) , 2005, 115-129
8. Chen H, Xu Y, Shum H, Zhu S and Zheng N, "Example-based facial sketch generation with non- parametric sampling", In: Proceedings of IEEE International Conference on Computer Vision, , 2001, pp 433-438
9. Cootes T and Taylor C, "Active shape models- 'smart snakes"', In: Proceedings of British Machine Vision Conference, 1992, pp 266-275
10. 10.Wang N, Tao D, Gao X, Li X and Li J, "A comprehensive survey to face hallucination, International Journal of Computer Vision106(1), 2014, 9-30,
11. 11.Bitouk D, Kumar N, Dhillon S, Belhumeur P and Nayar S, "Face swapping: automatically replacing faces in photographs", In: Proceedings of SIGGRAPH, 2008, pp 39.1-39.8
12. 12.Weise T, Bouaziz S, Li H and Pauly M, "Realtime performance-based facial animation", In: Proceedings of SIGGRAPH, 2011, pp 77.1-77.9
13. Sung J and Kim D, "Pose-robust facial expression recognition using view-based 2d+3d AAM", IEEE Transactions on Systems, Man andCybernetics, Part A: Systems and Humans38(4), 2008, 852-866
14. Blanz V, Vetter T "A morphable model for the synthesis of 3d faces",
15. In: Proceedings of SIGGRAPH, 1999, pp 187-194
16. Anderson R, Stenger B, Cipolla R and Wan V, "Expressive visual text-to-speech using active appearance models", In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp 3382-3389
17. Akshay Asthana, Stefanous Zafeiriou, Shiyang Cheng and Maja Pantic, "Robust Discriminative Response Map Fitting with Constrained Local Model", CVPR, 2013.
18. Ying Tai, Jian Yang, Yigong Zhang,Lei Luo,Jianjun Qian, and Yu Chen, "Face Recognition with pose variation and misalignment via Orthogonal Procrustes Regression", IEEE transactions on image processing, vol. 25, no. 6, 2016
19. T. F. Cootes, C. J. Taylor, D. H. Cooper and J.Graham, "Active shape models-their training and application," Comp. Vis. And Image Understand, vol. 61, no. 1, 1995, pp. 38-59

AUTHORS PROFILE

M. Annalakshmi received her B.E. degree from Thiagarajar College of Engineering in 1998, her M.E. degree in Optical Communication from Alagappa Chettiar College of Engineering and Technology in 2005 and her Ph.D. in image processing from Anna University, Chennai in 2019. Her research interest includes Compute Vision and Pattern Recognition.

.

S. Mohamed Mansoor Roomi received his B.E. degree from Madurai Kamaraj University, in 1990, his M.E. degree in Power Systems and Communication Systems from Thiagarajar College of Engineering in 1992 and 1997 and his Ph.D. in Image Analysis from Madurai Kamaraj University in 2009. He has authored and co-authored more than 250 papers in various journals and conference proceedings and numerous technical and industrial project reports.

V. KARTHIK received the B.E. degree in electronics and communication engineering from National Engineering College, Kovilpatti, Anna University, Chennai, and received the M.Tech degree in VLSI Design from SATHYABAMA UNIVERSITY, India, in the year of 2009 and 2011 respectively. He is presently working as Assistant Professor, Department of Electronics and Communication Engineering at Sethu Institute of Technology India. He has published 15 research papers in the National & International Conferences.