Continuous Authentication for Mouse Gesture Recognition using Hidden Markov Model

[1]Chinmayee.KS, [2]Vanishree C
[1]Dept. of ECE, BCET, Bangalore, Karnataka, India
[2]Dept. of ECE, SCT Institute of Technology, Bangalore, Karnataka, India

Abstract

The mouse dynamics biometric is a behavioral biometric technology that extracts and analyzes the movement characteristics of the mouse input device when a computer user interacts with a graphical user interface for identification purposes. The existing mouse dynamics analyzes has continuous authentication or reauthentication for which exact results have been achieved. Static authentication using mouse gesture dynamics faces some challenges because of the limited amount of data that has captured. Authentication is the process of determining whether someone or something is, in fact, who or what it claim to be. A new category of biometrics that is gaining popularity is behaviometrics, where analysis focuses on the user s behavior while he interacts with computing systems for identification purposes. In this paper, a new mouse dynamics analysis framework uses mouse gesture dynamics for static authentication. The captured gestures are analyzed using a Hidden Markov Model. This results in improvement of both the accuracy and validation compared to the existing mouse dynamics approaches.

Keywords

Behavioral Biometric, Gesture, Mouse Dynamics Analysis Framework, Peirce's Criterion, Behaviometrics.

I. Introduction

Mouse dynamics deals with extracting the features related to the mouse movements and analyzing them to extract a signature, which is unique for every individual and can be used to discriminate different individuals. The aim of mouse dynamics biometric technology is its ability to continuously monitor the legal and illegal users based on their usage of a computer system. This is referred to as continuous authentication.

Continuous authentication is very useful for continuous monitoring applications such as intrusion detection. This paper first identifies the user movements or characteristics when the user interacts with the mouse, results in the generation of mouse gestures and checks every time when the user make session and provides authentication to the users. The mouse gesture is drawned in uni-stroke.

A mouse dynamics describe an individual behavior with a computer-based pointing device such as mouse. In context of authentication, biometrics have several advantages over traditional authentication techniques that verify identity based on something one knows (e.g. Password) or something one has (e.g. Hardware token). In particular no need of memorizing the gestures.

A mouse gesture results from the combination of computer mouse movements and clicks in a way that the software recognizes as a specific command. Biometrics refers to the identification of humans by their characteristics or traits. Biometric identifiers are characterized as physiological and behavioral characteristics. A physiological biometrics is related to voice, DNA, hand prints. A behavioral biometrics is related to the behavior of the persons.

A biometric system involves 2 phases, enrollment phase and verification phase. In the enrollment phase, user will draw a set of gestures several times on a computer monitor using mouse. The features are extracted from the captured data, analyze them and train the neural network that is later used for identification. In the verification phase, the user will be asked to replicate a subset of gestures drawn during the enrollment phase for authentication. The integration of a biometric subsystem in such a complex system environment requires taking into account concerns such as social impact, usability, interoperability, resilience, and scalability For instance, biometric systems carry private user information that could be misused if accessed by intruders, although such a threat could be mitigated using privacy-aware biometric cryptographic techniques.

Mouse gesture dynamics deals with biometric authentication. The mouse gesture dynamics uses a modular design of the neural network for classification. In the existing graphical password schemes, the user is not only accepted to memorize and remember the graphical passwords, the user has to hide the passwords during the login process to avoid surfing attacks. The mouse dynamics proposed schemes depends on the user biometric information and the user need not to memorize the gestures. There are only two mouse dynamics based biometric systems techniques available for static authentication due to the complexity or challenges.

One of the methods proposed by Syukri etal uses the signature drawn by the user as input during the static authentication process. The other method proposed by Revett etal uses mouse lock method for static authentication. In the proposed approach, the user draws the gesture at login time which are collected and analyzed for authentication purpose. Existing gestures based authentication systems uses other input devices such as stylus but in this paper, mouse is used as input device for capturing the gestures.

II. Mouse Gesture Detection

A. System Design

The approach to user authentication based on mouse gestures involves giving the user to draw one or several gestures and asking them to replicate the gestures a certain number of times. The produced replications are then compared against templates produced by the user during the enrollment phase.

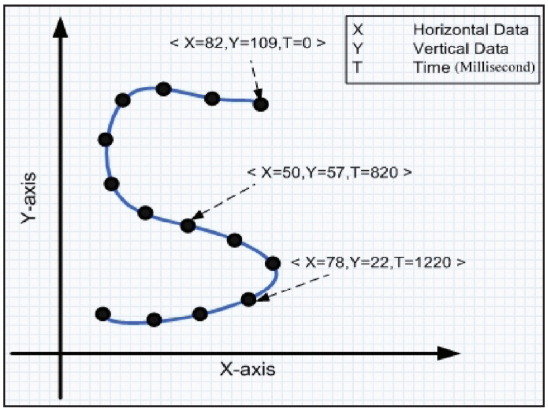


Fig. 1: Example of a Drawn Gesture Involving n=14 Data Points

The raw data collected from the drawing area consists of the horizontal coordinate (x-axis), vertical coordinate (y-axis) and the elapsed time in milliseconds at each pixel. Each gesture replication for a given gesture is defined as a sequence of data points. Each data point is represented by a triplet <x, y, t> indicates X-coordinate, Y-coordinate and elapsed time respectively. The jth replication of a gesture G is represented as a sequence $Gj = \{< x1j, y1j, t1j >, < x2j, y2j, t2j >, . . . < xnj, ynj, tnj >\}$, where n is the gesture size (GS) and each $< xij, yij, tij >$ where $(1 \leq i \leq n)$ is a data point. The main aim is to differentiate between individuals based on their behavioral biometrics while drawing mouse gestures.

The mouse dynamics analysis framework involves four modules.

1. Gesture creation module.
2. Data acquisition and preparation module.
3. Feature extraction module.
4. Classification module.

B. Gesture Creation Module

The gesture creation module is a simple drawing application where the user is asked to freely draw a set of gestures. Te main aim of this module is to make the user to draw the gestures in their own way to replicate them later.The gestures are not tied to any language and they do not necessarily have a meaning. The gestures should be in uni-stroke. For each gesture three parameters are collected namely horizontal coordinate, vertical coordinate and elapsed time in milliseconds. This is the first implementation module in the mouse gesture dynamics.

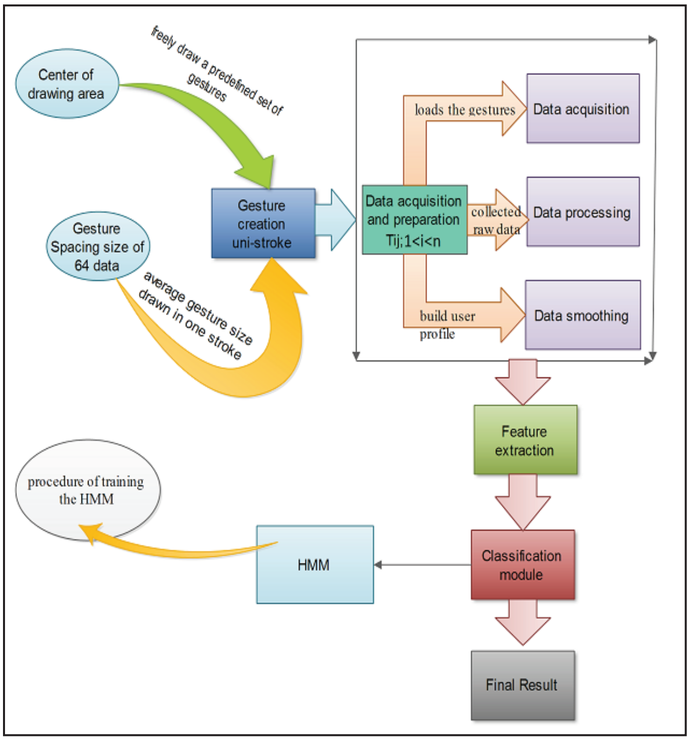

Center of drawing area
freely draw a predefined set of gestures
Gesture Spacing size of 64 data
average gesture size drawn in one stroke
Gesture creation uni-stroke
Data acquisition and preparation Tij;1<i<n
loads the gestures
Data acquisition
collected raw data
Data processing
build user profile
Data smoothing
Feature extraction
Classification module
HMM
procedure of training the HMM
Final Result

Fig. 2: Gesture Classification

C. Data Acquisition Module

The data acquisition component loads the gestures, created initially by the user using the gesture creation module and presents them to the user to replicate. The data acquisition module records the user interaction while drawing the gestures.

In Data preprocessing the data acquisition module preprocesses the collected raw data from the computer mouse in such a way that some noise patterns are ignored or dropped.

After preprocessing the raw data, the data acquisition module applies two types of normalizations for the input data. The first is normalization and the second is size normalization. The center normalization shifts the gesture to the center of the drawing area as implemented in the gesture creation module. Then normalize the size so that the final size of the gesture is equal to the size of the template gesture to compare the two gestures. The normalization can be applied by accepting gestures which is drawn by the users whose size is greater than or equal to the size of the template gesture.

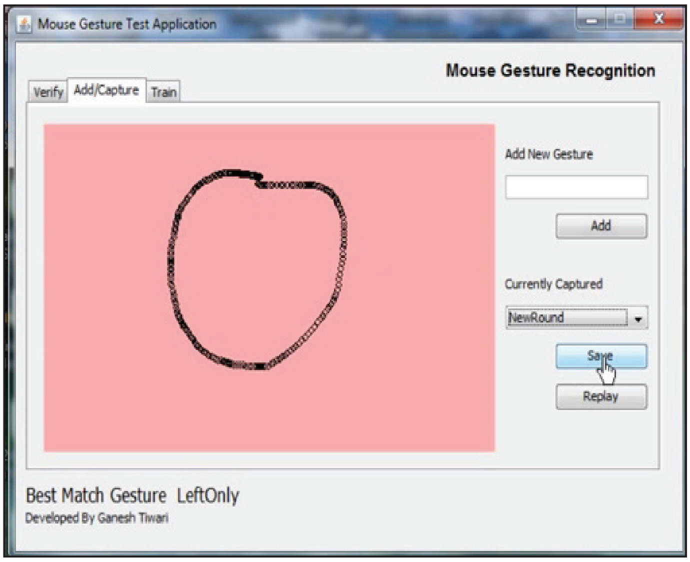

Mouse Gesture Test Application
Mouse Gesture Recognition
Verify
Add/Capture
Train
Add New Gesture
Add
Currently Captured
NewRound
Save
Replay
Best Match Gesture LeftOnly
Developed By Ganesh Tiwari

Fig. 3: Gesture Creation and Adding a New Gesture

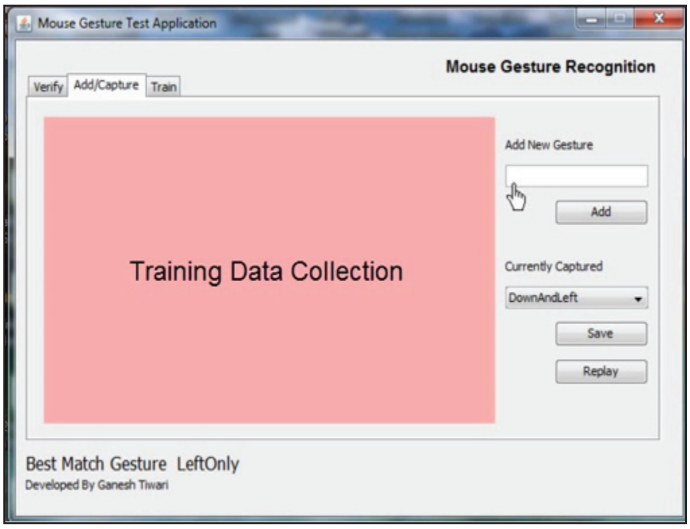

Mouse Gesture Test Application
Mouse Gesture Recognition
Verify
Add/Capture
Train
Training Data Collection
Add New Gesture
Add
Currently Captured
DownAndLeft
Save
Replay
Best Match Gesture LeftOnly
Developed By Ganesh Tiwari

Fig. 4: Training Data Collection

If the gesture size is bigger than the template size, then the k-means algorithm is used to cluster the data points into 64 clusters. The Euclidean distance is a distance measure between the data points in three dimensions <x, y, t> is used. Then the centroids of the 64 clusters are used to form the new gesture.

D. Outlier Removal and Data Smoothing

Data smoothing is used to eliminate noise and extract the real patterns from the data. The smoothing is used to smooth the data among the different replications obtained for each users. In general, human beings cannot draw the same gesture with the same exact details twice. This results in some variability in the replicas produced by the same user for the same gesture. Data smoothing will smooth such variability and minimize its effect on the learning process. The weighted least-squares regression (WLSR) method to smooth the data.

Algorithmic steps for k-means clustering:
Let X = {x_1,x_2,x_3,……..,x_n} be the set of data points and V = {v_1,v_2,…….,v_c} be the set of centers.

1. Randomly select ‘c’ cluster centers.
2. Calculate the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers..
4. Recalculate the new cluster center using:

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, ‘c_i’ represents the number of data points in i^{th} cluster.

5. Recalculate the distance between each data point and new obtained cluster centers.
6. If no data point was reassigned then stop, otherwise repeat from step 3).

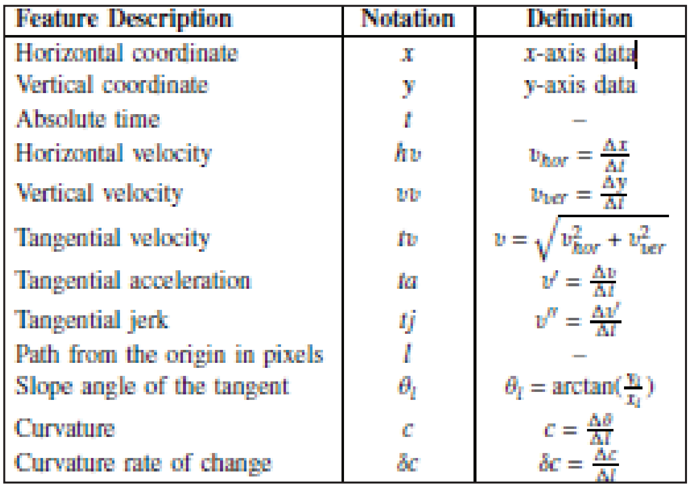
The Peirce’s criterion [21] is used to eliminate the outliers. Peirce’s criterion method is a robust statistical based method that does not make any assumptions about the data. Peirce’s criterion deals with the data that has several suspicious values.
Peirce’s criterion determines outliers by computing the maximum allowable deviation from the sample mean. Given m the sample size, n is the number of outliers and R is the ratio between the maximum allowable deviations to the standard deviation. The maximum allowable deviation is calculated by $dmax = R \times \sigma$, where σ is the standard deviation of the sample and xi is a data item which is considered as an outlier if $|x_i - x_m| > d_{max}$, where x_m is the sample mean. Using the Peirce’s criterion method and starting from n=1, the outliers are removed sequentially by incrementing the number of possible outliers, while keeping the original values of mean, standard deviation and sample size. The process is repeated until no other data item needs to be eliminated.
The outlier removal and data smoothing steps are applied only to horizontal coordinates and vertical coordinates data. The vector is constructed to aggregate the same occurrence of the first data point from each of the different replications. Then apply the Peirce’s criterion method and WLSR method to the data in the vector to produce clean and smoothed data. Repeat the process for each of the remaining data points of the gesture. The smoothing occurs only on the training samples and not on the test data.
The outlier removal and data smoothing module is implemented by using algorithm and assumes the following.

1. Let *m* be the number of replications.
2. Let *n* be the size of the gesture.
3. Let *pij* = (*xij, yij*) be a data point, where 1 ≤ *j* ≤ *m, 1* ≤ *i* ≤ *n.*
4. Given a gesture *G*, we denote by *Gj* the *jth* replica *Gj* = (*p1j, p2j, . . . , pnj*).
5. Let *Pi* denote a vector containing the *ith* data point from each of the different replications, where i = 1, 2, . . . , *n*: *Pi* = (*pi1, pi2, . . . , pim*).

E. Feature Extraction module

The feature extraction module extracts the features from the raw data. Features selection is made by analyzing sample data and identifying the features that exhibit strong reproducibility and discriminative capabilities. The feature extraction module extracts the features from the raw data. First the data are collected and then extracted.
Features that are obtained from the vector of data points that are intercepted between two mouse clicks can be used.The complete list of the extracted features is provided in Table 1.This module is mainly used for the extraction of features of the gesture drawned by the users. The features are x coordinate, y coordinate, absolute time, horizontal velocity, vertical velocity, tangential velocity, angle, tangent acceleration, curvature, curvature rate of change, path from the origin in the pixels, tangent jerk, slope angle of the tangent and so on.

Table 1: Feature Extracted From Raw Data



Feature Description	Notation	Definition
Horizontal coordinate	x	x-axis data
Vertical coordinate	y	y-axis data
Absolute time	t	–
Horizontal velocity	hv	$v_{hor} = \frac{\Delta x}{\Delta t}$
Vertical velocity	vv	$v_{ver} = \frac{\Delta y}{\Delta t}$
Tangential velocity	tv	$v = \sqrt{v_{hor}^2 + v_{ver}^2}$
Tangential acceleration	ta	$v' = \frac{\Delta v}{\Delta t}$
Tangential jerk	tj	$v'' = \frac{\Delta v'}{\Delta t}$
Path from the origin in pixels	l	–
Slope angle of the tangent	θ_t	$\theta_t = \arctan(\frac{y_t}{x_t})$
Curvature	c	$c = \frac{\Delta\theta}{\Delta t}$
Curvature rate of change	δc	$\delta c = \frac{\Delta c}{\Delta t}$

F. Classification Module

In this paper, first the technique known as principal component analysis is applied. Principal component analysis will reduce the correlation among two objects. It’s difficult to convert the principal component analysis to square matrix and to normalize to co variance. This results in low performance. The feed-forward back propagation multilayer network was tried. The training steps of this network were exhaustive and time consuming. The training process is stopped when it exceeds five hours (on a computer system with a 2GHz Core 2 Duo CPU and 2GB RAM) for only a population of two users.
The Hidden Markov Model was used. The Hidden markov model is mainly used for comparisons and recognition of gestures. The goal is to recognize the gestures. First the tools have been created for easily generating and modifying test suites of data. After the training has been generated and stored as a test suite, the HMM recognizer program instantiates a new codebook of specified size and a set of HMMs (one per gesture) with a specified number of states. The HMMs are immediately tested on the training data to verify the accuracy of training. The system can be worked for many hours. This gave good results.

III. Conclusion

In this paper, we highlighted the challenges faced by mouse dynamics biometric technology when it is applied to authentication and proposed a new mouse dynamics analysis framework to train the data that gave good results. The proposed framework uses hidden markov model for classification and uses Peirce’s criterion and weighted least -square s regression methods for outlier removal and data smoothing. Authentication is the process of determining whether someone or something is, in fact, who or what it claim to be. The ways in which someone may be authenticated fall into three categories: something the user knows, something the user has, and something the user is. The Hidden markov model is used as a classification module which yields in accurate results.

In the future work, we intended to enhance the accuracy using various techniques. Since the proposed system is entirely software based, integrating in a complex system environment such as e-commerce or e-learning portals should be straightforward from an implementation perspective. The verification time should be much faster in order to train the gestures. One of the challenges is the protection of systems against security attacks. Like other biometric techniques, mouse dynamics can be the target of reply attacks. Such threats can be mitigated by strengthening the protection of biometric templates using various techniques. Mouse dynamics can also be a target of generative attacks through forgeries. In our feature work, we planned to strengthen our system by investigating the impact of generative attacks against it.

References

[1] S. Patel, J. Pierce, G. Abowd,"A gesture-based authentication scheme for untrusted public terminals," In Proc. UIST, Oct. 2004.

[2] A. A. E. Ahmed, I. Traore,"A new biometric technology based on mouse dynamics", IEEE Transactions on Dependable and Secure Computing, 4(3), pp. 165-179, 2007.

[3] R. Plamondon, S. N. Srihari,"Online and off-line handwriting ecognition: A comprehensive survey," IEEE Trans. Pattern Anal. Mach. Intell., Vol. 22, No. 1, pp. 63–84, Jan. 2000.

[4] Ahmed Al-Khazzar, Nick Savage,"Graphical authentication based on user behaviour", In Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference. Poster: Preliminary investigation of gesture-based password: Integrating additional user behavioral features.

[5] H. Gamboa, A. Fred,"A behavioral biometric system based on human-comp. inter," In Proc. Conf. Biometric Tech. Human Identification, Vol. 5404, pp. 381–392, 2004.

[6] S. Cimato, M. Gamassi, V. Piuri, R. Sassi, F. Scotti, "Privacy-aware biometrics: Design and implementation of a multimodal verification system", In Proc. Annu. Comp. Sec. Apps. Conf., 2008, pp. 130–138.

[7] D. Lopresti, F. Monrose., L. Ballard,"Biometric authentication revisited: Understanding the impact of wolves in sheep's clothing", In Proc. 15th USENIX Sec. Symp., 2006.

[8] M. S. Obaidat, N. Boudriga,"Sec of e-Sys and Comp Networks", Cambridge, MA: Cambridge Univ. Press, 2007.

[9] M. Obaidat, B. Sadoun,"Verification of comp. users using keystroke dynamics," IEEE Trans. Syst., Man, Cybern., Vol. 27, No. 2, pp. 261–269, Apr. 1997.

[10] H. Gamboa, A. Fred,"A behavioral biometric system based on human-comp. inter", In Proc. Conf. Biometric Tech. Human Identification, Vol. 5404, pp. 381–392, 2004.

[11] A. A. E. Ahmed, I. Traor´e,"A new biometric tech. based on mouse dynamics," IEEE Trans. Dependable Secure Comput., Vol. 4, No. 3, pp. 165–179, Jul.–Sep. 2007.

[12] M. Pusara, C. E. Brodley,"User reauthentication via mouse movements", In Proc. ACM Workshop Visualization Data Mining Comp. Sec. (VizSEC/DMSEC), pp. 1–8, 2004.

[13] N. Zheng, A. Paloski, H. Wang,"An efficient user verification system via mouse movements", In Proc. 18th ACM Conf. Comp. Commun. Sec., 2011, pp. 139–150.

[14] K. Revett, H. Jahankhani, S. de Magalhaes, H. M. D. Santos,"A survey of user authentication based on mouse dynamics", In Proc. ICGeS, CCIS'12, pp. 210–219, 2008.

[15] P. Oel, P. Schmidt, A. Shmitt,"Time prediction of mouse-based cursor movements", In Proc. Joint AFIHM-BCS Conf. Human Comp. Inter., Vol. 2. Sep. 2001, pp. 37–40.

[16] A. A. E. Ahmed, I. Traor´e,"System and method for determining a comp. user profile from a motion-based input device," U.S.patent 10/555408, PCT/CA2004/000669, 2003.

[17] A. Nazar, I. Traor´e, A. Ahmed,"Inverse biometrics for mouse dynamics", Int. J. Artif. Intell. Pattern Recognit., Vol. 22, No. 3, pp. 461–495, May 2008.

[18] M. Gamassi, M. Lazzaroni, M. Misino, V. Piuri, D. Sana, F. Scotti,"Accuracy and performance of biometric systems", In Proc. Instrum. Meas. Tech. Conf., 2004, pp. 510–515.

[19] S. Bengio, J. Mariethoz,"A statistical significance test for person authentication", In Proc. Odyssey: Speaker Language Recognition Workshop, 2004.

[20] S. Lloyd,"Least squares quantization in PCM", IEEE Trans. Inform. Theory, Vol. 28, No. 2, pp. 129–137, Mar. 1982.

[21] S. Ross,"Peirce's criterion for the elimination of suspect experimental data", J. Eng. Tech., Vol. 20, No. 2, 2003.

[22] R. Biddle, S. Chiasson, P. C. V. Oorschot,"Graphical passwords: Learning from the first twelve years", School Comp. Sci., Carleton Univ., Ottawa, ON, Canada, Tech. Rep. TR-11-01, Jan. 2011.

[23] F. Azam,"Biologically inspired modular neural networks," Ph.D. dissertation, Virginia Polytechnic Instit. State Univ., Blacksburg, 2000.

Chinmayee KS, received her degree of post graduation in the year 2014, in VLSI Designs and Embedded systems from Visvesvaraya Technological University. Currently, working as an Assistant Professor in the Department of Electronics and Communication Engineering at Bangalore College of Engineering and Technology, Bangalore. Her interests includes digital signal processing, image processing, analog and digital communication systems.

Vanishree C, received her degree of post graduation in the year 2014, in VLSI Designs and Embedded systems from Visvesvaraya Technological University. Currently, working as an Assistant Professor in the department of Electronics and Communication Engineering at SCT Institute of Technology, Bangalore. Her interests includes Image Processing, wireless communication systems.