FACE RECOGNITION USING MAXIMUM CONFIDENCE HIDDEN MARKOV MODEL

Swati Raut[1] & S. H. Patil[2]

[1]Student (M. Tech) Computer, Bharathi Vidyapeeth University, College of Engineering, Pune-21, Maharashtra, India
[2]Professor and Head of Department of Computer, Bharathi Vidyapeeth University, College of Engineering, Pune-21, Maharashtra, India

ABSTRACT

Different approaches have been proposed over the last few years for improving holistic methods for face recognition. Some of them include color processing, different face representations and image processing techniques to increase robustness against illumination changes. There has been also some research about the combination of different recognition methods, both at the feature and score levels. In this paper a new system for face recognition is proposed based on hidden makov models (HMMS). The performance of Face recognition by MC-HMM heavily depends on the choice of model parameters. In this paper, we use discriminative feature extraction method for face recognition . Experimental results illustrate that compared with the conventional HMM based face recognition algorithm the proposed method obtain better recognition accuracies and higher generalization ability.

KEYWORDS: *Hidden Markov Model, Face Recognition, Discriminative Feature Extraction, Pattern recognition Generalization ability.*

I. INTRODUCTION

Face recognition is an important research problem spanning numerous fields and disciplines. This technique was subsequently applied in partial pattern recognition application, more specifically in speech recognition problems [1].Face recognition is undoubtedly an interesting research area, growing in importance in recent years due to its applicability as biometric system in commercial and security application. Face recognition is a biometric approach that employs automated methods to verify or organize the identity of living person based on her/his physiological characteristics. Face recognition is a set of two tasks:-

1. Face identification: - Given a face image belong to person in database tell whose image is it.
2. Face Verification: - Given a face image that might not belong to database verify whether it is from the person it is claimed to be database.

As researcher interest in face recognition continued, many different algorithm were developed PCA (principal component analysis), LCA (linear discriminate analysis) [2].PCA commonly referred to as the use of Eigen faces, with PCA the probe and galley images must be the same size and must first be normalized to line up the eyes and mouth of the subjects within the images. The PCA approach typically requires the full frontal face to be presented each time, otherwise the image results in poor Performance. The primary advantage of this technique is that it can reduce data needed to identify the individual to 1/1000th of the data presented.LDA (linear discriminate analysis)is a statistical approach for classifying samples of unknown classes based on training samples with known classes. This technique aims to maximize between class (i.e. across users) variance and minimize within class (i.e. within user) variance. In this paper, we are concerned with two issues in establishing in HMMs for 2D object recognition [3]. The first issue involves a hybrid process of feature extraction and model estimation.

II. PROPOSED FACE RECOGNITION SYSTEM

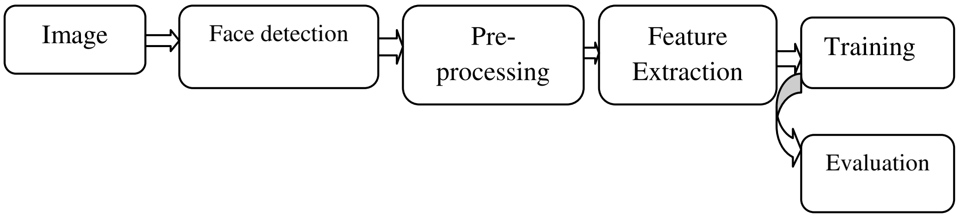

Fig1. Architecture of Face Recognition System

Fig 1. Shows general architecture of face recognition system [4].In this paper feature extraction plays most important part in HMM .we are using discriminative feature extraction algorithms [5].the presented face recognition scheme includes two key elements: preprocessing based on pixel averaging for dimension reduction and energy normalization to reduce the effect of image brightness feature extraction successively conducted by minimum classification error (MCE)[6][7].

2.1. Minimum Classification Error Training

Minimum classification error (MCE) training aims to estimate model distributions {p $(X|\Lambda_c)_{,c=1},\ldots,c$} in a fashion that classification errors of training. Data X={$X_{1},\ldots X_T$} are minimized. Target model Λ_c and competing model $\bar{\Lambda}_c$ can be alleviated. The trained models $\Lambda=\{\Lambda_{1,}\ldots\ldots\ldots, \Lambda_c\}$ should be general for the classification of test data, using log likelihood as discriminate function $g(X,\Lambda_c) = \log p(X|\Lambda_c)$, the first step was to calculate the misclassification measure.

$$d(X,\Lambda_c)= -g(X,\Lambda_c)+G(X,,\bar{\Lambda}_c)$$

$$= -\log p(X|\Lambda_c) + \left[\frac{1}{c-1}\sum_{\Lambda_{j \neq \Lambda c}} \rho \log p\left(X|\Lambda_j\right)\right]^{1/\rho} \quad (1)$$

Where ρ was a tuning parameter in antidiscriminant function G(X, $\bar{\Lambda}_c$) determined from C-1 non target models The third step was done through minimizing the expected loss or Bayes risk.

$$\Lambda_{MCE} = \arg\min_{\Lambda} E_X\left[l(X,\Lambda)\right]$$

$$= \arg\min_{\Lambda} E_X\left[\sum_{c=1}^{C} l(X,\Lambda_c)1(X\in\Lambda_c)\right] \quad (2)$$

1(.) is indicator function gradient descent algorithm used for optimization.

2.2. Feature Extraction Using Discriminative Training Algorithm

Discriminative training algorithm based on approximation of maximum mutual information (MMI) [8][9] Objective function and its maximization in a technique similar to the exception maximization (EM) algorithm[3].

LDA was known as discriminative feature extraction. In 2D pattern recognition transformation matrix W was estimated by maximizing fisher's ratio creation.

$$W_{LDA} = \arg_W \max \frac{\left|W^T S_b W\right|}{\left|W^T S_w W\right|} \tag{3}$$

Where s_b and s_w were between-class and within–class scatter matrices, respectively. $w^T{}_{LDA}$ used for discriminative features.LDA method was obtained by finding the eigenvectors of $s_w^{-1} s_b$ .MCE criterion was feasible to extract discriminative features through transformation matrix.

$$\beta = \alpha.(N_g - 1) \tag{4}$$

After discriminative feature extraction calculated by MCE optimization of objective function expressed as fisher's ratio creation. In discriminative feature extraction Gaussian parameters was considered.

2.3. Hidden Markov Model for Face recognition

A hidden markov model (HMM) is a statistical markov model in which the system being modeled is assumed to be a markov process with unobserved (hidden) states. An HMM can be considered as a simplest dynamic Bayesian network [10]. In regular markov model the state is directly visible to observer therefore the state transition probabilities are the only parameters. In a hidden markov model the state is not

Directly visible, but the output dependent on the state is visible. Each state has probability distribution over the possible output tokens. Therefore sequence of tokens generated by on HMM gives some information about the sequence of state.

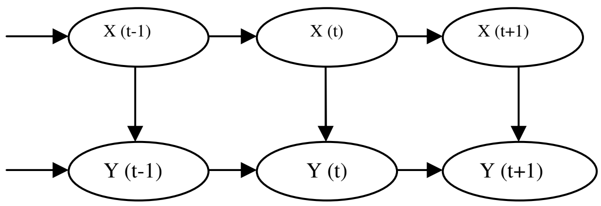

Fig 2. Architecture of Hidden Markov Model

The above fig2 shows the general architecture of an instantiated HMM. Each oval shape represents a random variable that can adopt any of number of values. The random variable x(t) is the hidden state at time t.(with x(t)={x1,x2,x3})the random variable y(t) is the observation at the t with (y(t)={y1,y2,y3}).

When we are using HMM for one dimensional gray scale face image. The significant facial regions (hair, forehead, eyes, nose, mouth) come in natural order from top to bottom [11].

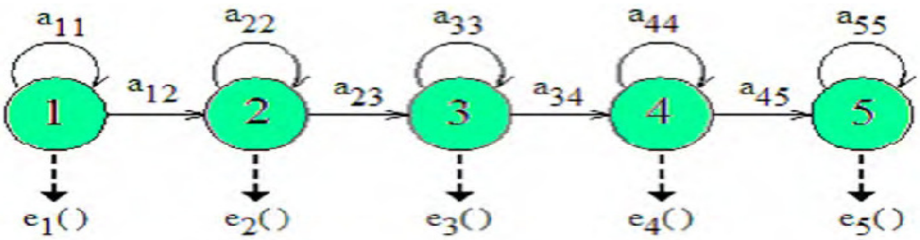

Fig3. Left to Right HMM for ace Recognition

In above fig 3. Each of these facial regions is assigned to a state in left to right 1D continues HMM. In 2D HMM we fallow the embedded HMMs for facial image representation and recognition[3] .facial image .X = $\{X_t\}=\{X_{tnm}\}$ of a person C are blocked with spatial indices (n,m) and modeled by structural HMM consisting of set of superstae.q_{nm} as shown I fig 4 .no skipping is allowed between super sates and embedded states.

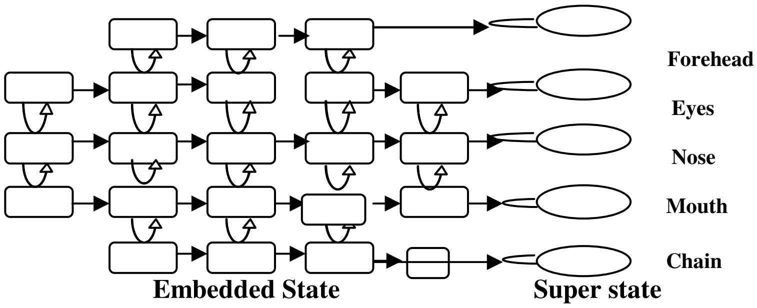

Fig 4 Facial images representation using HMM

There are five super sates characterizing the forehead, eyes, nose, mouth and chain in the vertical direction with initial state probabilities and state transition. Probabilities $\{\Pi_{\upsilon}, A_{\upsilon}\}=\{\pi_{sn}, a_{snSn+1}\}$.Each super state is itself a standard 1D HMM containing embedded states for modeling facial features in the horizontal direction. In total 21 embedded states are considered.

Fig 5. Input Face represented in embedded states and super states

$$b_{qnm}(x)=p(x|q_{qnm};B_h)=\sum_{l_{nm}}^{N_l}\omega_{qnm}l_{nm}N(x|\mu_{qnm}l_{nm},\Sigma q_{nm}l_{nm}) \tag{5}$$

Where $w_{qnm}l_{nm}$, $\mu q_{nm}l_{nm}$ are the mixture weight, mean vector and covariance matrix of embedded state q_{nm} And mixture component l_{nm} respectively [3].The trained parameter $\Omega=\{\Pi_v, A_v, \Pi_h, A_h, B_h\}$ the most likely class of a test image x_t is decided according to the like hood function given the optimal sequence $\{s_n, q_{nm}, l_{nm}\}$ of super states and embedded states.

III. RELATED WORKS

3.1. Hypothesis Test

In hypothesis test problem three steps considered:-
1 Define null hypothesis H_o and alternative hypothesis H_1 and choose significance level.
2. Calculate like hood function $p(X| H_o)$.
3. $p(X| H_1)$ are given the optimal solution is obtained by a like hood ratio test. [12].

$$LR=\frac{p(x|H_0)}{p(x|H_1)}\geq k \tag{6}$$

Given like hood function $p(x|H_0)$ and $p(x|H_1)$ used to obtain like hood ratio test where κ is decision threshold determined by significance level for distribution of LR. We perform facial modeling through optimizing the accumulated LR target models to competing models or equivalently maximizing the confidence of fitting X closer to target parameter Λ and farther from competing parameters $\bar{\Lambda}$.

3.2. Classification With an HMM

An HMM is a statistical model used to characterized the statistical properties of signal. The elements of HMM can formally defined by specifying the following parameters [3].

1. $X = \{X_t\}=\{X_{tnm}\}$:D-dimensional observation vector of frames $t = 1,\cdots,T$ in spatial indices $n= 1,\cdots,N$ and $m =1,\cdots,M$

2. $\{\Pi_h, A_h, B_h\}=\{\Pi_{qnm}, a_{\omega qnm}, {}_{\omega qnm,lnm}, \mu_{qnmlnm}, \sum_{qnmlnm}\}$:initial state probabilities, state transition probabilities and observation probabilities of embedded (horizontal) states $(q_{nm}q_{n,m+1}) =1,\cdots,N_q$ and mixture component $l_{nm}=1,\cdots,N_l$. observation probability $b_{qn}(x)$ consist of parameter of mixture weight ω_{qnm},l_{nm}, Gaussian mean vector.

3. $W = [W_d\ W_{D\text{-}d}]$: feature transformation matrix with reduced dimension.

4. $\delta_v(s_n),\delta_h(q_{nm})$:The beast confidence scores of super and embedded states at vertical and horizontal indices(n,m).

5. α ,β: discriminative factor with $\alpha = \beta(N_g -1)$ where N_g total number of Gaussians.

6. $\bar{\Lambda},q_{mn},l_{nm}$: competing model, embedded state and mixture component

7. .$\{s,q,l\} =\{s_n ,q_{nm}, l_{nm}\}$:sequence of super states, embedded states and mixture components.

8. $y_t(q_{nm,}l_{nm})$:posterior probability of x_{nm} staying in state q_{nm} and mixture component l_{nm} given the current parameter Λ and X.

$$Q(\Lambda'|\Lambda)=E_{s,q,l}[LLR(X,s,q,l|\Lambda')|X,\Lambda]= \sum_{s,q,l} P(s,q,l \mid X,\Lambda) \quad X \quad \sum_{t=1}^{T}$$

$$\Bigg[\log \pi'_{s1} \sum_{n=1}^{N} \Bigg[\log a'_{s_n s_{n+1}} + log\pi'_{qn1}$$

$$+\sum_{m=1}^{M}\Bigg[log\, a'_{qnmq_{n,m+1}}$$

$$+ \log N\left(W'^{T}_{d}x_{tnm}\middle|\mu'_{qnml_{nm}}\right) - \frac{a}{N_{g-1}} \sum_{(\bar{q}_{nm,}\bar{l}_{nm})\#(q_{nm},l_{nm})} \log N\left(W'^{T}_{d}x_{tnm}\middle|\mu'_{\bar{q}_{nm}\bar{l}_{nm}}, \textstyle\sum'_{\bar{q}_{nm}\bar{l}_{nm}}\right)\Bigg]\Bigg]\Bigg] \tag{7}$$

Where N_g is total number of Gaussians. Λ And $\bar{\Lambda}$ are the current estimate and new estimate respectively. Here we are concerned with finding solutions to the transformation matrix, mixture weights.HMM mean vectors and covariance matrices. The like hood function of competing models is determined from $N_g - 1$ Gaussians. In the maximization step, we maximize the auxiliary function $Q(\Lambda'|\Lambda)$ new estimate Λ' to find MC-HMM parameters.

$$\sum_{t=1,n=1,m=1}^{T.N.M.} \sum_{q_{nm}=1,l_{nm}=1}^{N_q N_1} \gamma_t(q_{nm}, l_{nm}) . \left\{ \begin{bmatrix} \log \omega'_{q_{nm} l_{nm}} + \log |W'| \\ + \frac{1}{2} \log \left| \Sigma^{l-1}_{q_{nm} l_{nm}} \right| \\ - \frac{1}{2} \left(W_d^{'T} X_{tnm} - \mu'_{q_{nm} l_{nm}} \right)^T \\ \Sigma^{l-1}_{q_{nm} l_{nm}} (W_d^{'T} X_{tnm} - \mu'_{q_{nm} l_{nm}}) \end{bmatrix} \right. \tag{8}$$

In (11) ,some constants corresponding to the last D-d dimensions are neglected. Interestingly , $\beta = \alpha.(N_g - 1)$ has the physical meaning of controlling the discriminative rate for model training. By using viterbi alignment for face image x_{tnm} ,we obtain $\lambda_t(q_{nm,} l_{nm}) = 1$ when x_{tnm} is not alien to state and mixture labels $(q_{nm,} l_{nm})$ and $\lambda_t(q_{nm,} l_{nm}) = 0$ when x_{tnm} is not aligned to $(q_{nm,} l_{nm})$. The transformation parameters w' get fix and maximize.

3.3. Implementation Procedure

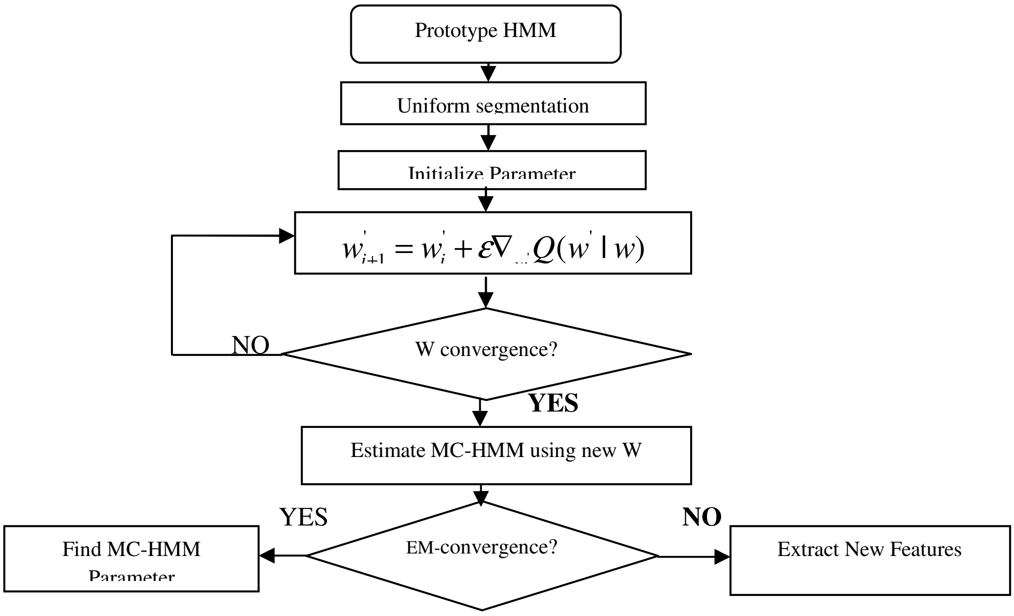

Fig6. Training Scheme

1. First the data is uniform ally segmented to obtain initial estimate of model parameter.
2. At the next iteration the uniform segmentation perform using the initial features of training image.
3. In next iteration HMM parameters are estimated using viterbi algorithm. Input image translated in to super states and embedded sates. Transform matrix W' Is then calculated. MC-HMM estimation converges after several EM iteration.

IV. RESULTS & DISCUSSION

In the experiments we use face database. (http://www.face.rec.org/database). Total 25 face images was selected and evaluated the effect of different numbers of classes from C=20 to C=153. We carried out ML-HMM and MCE-HMM [1][13] and MC-HMM[6] [7]for facial modeling .All images are manually cropped. Feature dimension d discriminative factor α in the MC Creation were varied for comparison.

Table1. Log-like hood ratio versus EM iteration for different discriminate factor α

Em-Iteration	Log like hood Ratio(0.001)	Log like hood Ratio(0.2)
1	-2.20	-2.18
5	-2.15	-2.21

In next comparison fig.7shows recognition accuracy. The MC-HMM with feature dimension reduction to d=16 is consider. In MC-HMM state alignment shall be matching facial features if HMM parameters are well trained. The MC-HMM with reduced feature dimension obtains quite good alignment in characterizing not only the vertical facial segments via super sates but also horizontal tiny texture via embedded states.

Table2. Recognition Rate (in percentage)

Method	Class Number(c)							
	20	40	60	80	100	120	140	160
MC-HMM (d=16)	94	93	92	90	89	87	86	85
MC-HMM (d=14)	84	83	82	80	79	78	77	72
MC-HMM (d=36)	88	89	87	86	85	84	83	80

In next comparison fig.8 shows comparison rate in percentage. MCE-HMM algorithm is also based on discriminative training algorithm. We compare MC-HMM with another and also with non-hmm methods using Eigen face and fisher faces. The cases of C=50 and C=100 are investigated. HMM methods are significantly better than non-hmm methods. MC-HMM achieves the best classification performance. The result shows the superiority of HMM modeling to Eigen face and fisher face.

Table3. Comparison Result in (percentage)

Methods	Recognition Rate In(%)	
	C=50	C=100
MC-HMM (d=36)	97	93
MC-HMM (d=16)	95	92
MCE-HMM (d=16)	92	90
ML-HMM (d=36)	90	89
Fisher face	87	83
Eigen face	83	78

Next Fig 9. Shows comparison of recognition time (in seconds) for different methods feature dimension. MC-HMM compared with other methods. MC-HMM having higher recognition time (sec).

Table4. Comparison of Training Time (In second)

Methods	Training Rate(In Seconds)
MC-HMM (d=36)	400
MC-HMM (d=16)	340
MCE-HMM (d=16)	250
ML-HMM (d=36)	289
Fisher face	255
Eigen face	89

V. CONCLUSION

This paper describes an MC-HMM approach for face recognition that use discriminative feature extraction, In which 2D-DCT co-efficient use as feature vector. The method is compared to the earlier HMM based face recognition system. By using Hypothesis test principle the new objective function was derived. The experimental result shows that this algorithm achieves higher recognition rate and stronger generalization ability. That is to say, this algorithm has a stronger ability to deal to deal with new data. The focus of this paper is on the use of MC-HMM techniques for face recognition. For this we have presented Discriminative Feature Extraction where a transformation matrix W was estimated by maximizing Fisher's ratio criterion. Although additional papers treating specific aspects of this field can be found in the literature, these are invariably based on one or another of the key techniques presented and reviewed here. Our goal has been to quickly enable the interested reader to review and understand the state-of-art for MC-HMM models applied to face recognition problems. It is clear that different techniques balance certain trade-offs between computational complexity, speed and accuracy of recognition and overall practicality and ease-of-use. Our hope is that this paper will make it easier for new researchers to understand and adopt MC-HMM for face analysis and recognition applications and continue to improve and refine the underlying techniques.

REFERENCES

[1]. A.V. Nefian and M.H. Hayes III, "An Embedded HMM- Based Approach for Face Detection and Recognition," Proc.IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, vol 6, pp. 3553-3556,1999
[2]. A.M. Martinez and A.C. Kak, "PCA versus LDA," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, pp. 228-233, 2001.
[3]. Jen-Tzung Chien, Senior Member, IEEE, and Chih-Pin Liao "Maximum Confidence Hidden Markov Modeling for Face Recognition" IEEE TRANSACTIONS ON PATTERN AND MACHINE INTELLIGENCE, VOL. 30, NO. 4, APRIL 2008
[4] Peter M. Corcoran and Claudia Iancu "Automatic Face Recognition System for Hidden Markov Model Techniques"
[5]. A. Ben-Yishai and D. Burshtein, "A Discriminative Training Algorithm for Hidden Markov Models," IEEE Trans. Speech and Audio Processing, vol. 12, no. 3, pp. 04-217, 2004
[6]. B.-H. Juang and S. Katagiri, "Discriminative Learning for Minimum Error Classification," IEEE Trans. Signal Processing, vol. 40, no. 12, pp. 3043-3054, 1992.
[7]. S. Katagiri, B.-H. Juang, and C.-H. Lee, "Pattern Recognition Using a Family of Design Algorithms Based upon the Generalized Probabilistic Descent Method," Proc. IEEE, vol. 86, no. 11,pp. 2345-2373, 1998.
[8]. L. Bahl, P. Brown, P. de Souza, and R. Mercer, "Maximum Mutual Information Estimation of Hidden Markov Model Parameters for Speech Recognition," Proc. IEEE Int'l Conf. Acoustic, Speech, and Signal Processing, vol. 11, pp. 49-52, 1986.
[9]. A. Ben-Yishai and D. Burshtein, "A Discriminative Training Algorithm for Hidden Markov Models," IEEE Trans. Speech and Audio Processing, vol. 12, no. 3, pp. 204- 217, 2004.
[10]. A.V. Nefian, "Embedded Bayesian Networks for Face Recognition,"Proc. IEEE Int'l Conf. Multimedia and Exp, vol. 2, pp. 133-136, 2002.
[11]. Lihaipeng, LiJingJiao"Implement of Face Recognition System Based on Hidden Markov Model" 978-1-4244- 5961-2/10/$26.00 ©2010 IEEE, 2010 Sixth International Conference on Natural Computation (ICNC 2010)
[12]. *Chih-Pin Liao and Jen-Tzung Chien"* Maximum Confidence Hidden Markov Modeling" 142440469X/ 06/$20.00 ©2006 IEEEICASSP 2006

Authors

Swati Raut is the student-M.tech (Computer) Bharati vidyapeeth university, College of engineering pune-21, Maharashtra, India.

S. H. Patil received the PhD degree in computer engineering. He is currently working as Professor and Head of Department of Computer engineering, Bharathi Vidyapeeth University, college of engineering Pune-21, Maharashtra, India