

Canonical Correlated Kernel PCA Method for Face Recognition

¹N.NagaMounica and ²Ch.Ganapathy Reddy

¹Master Scholar, DECE, GNITS.

²Head of the Dept., ECE, GNITS.

Jawaharlal Nehru Technological University
Hyderabad, India

Abstract—Practical face recognition systems are sometimes confronted with low-resolution face images. To address this problem, a super-resolution method that uses nonlinear mappings to infer coherent features that favor higher recognition of the nearest neighbor (NN) classifiers for recognition of single LR face image is presented. Canonical correlation analysis is applied to establish the coherent subspaces between the principal component analyses (PCA) based features of high-resolution (HR) and LR face images. The obtained features from PCA are not good enough for dimensionality reduction and computational complexity when large set of databases are taken into consideration. To overcome that problem Kernel PCA is introduced. Then, a nonlinear mapping between HR/LR features can be built by radial basis functions (RBFs) with lower regression errors in the coherent feature space than in the KPCA feature space. Thus, we can compute super-resolved coherent features corresponding to an input LR image according to the trained RBF model efficiently and accurately. And, face identity can be obtained by feeding these super-resolved features to a simple NN classifier. Extensive experiments on the Yale database show that the proposed method outperforms the state-of-the-art face recognition algorithms for single LR image in terms of both recognition rate and robustness to facial variations of pose and expression.

Keywords: Face Recognition, Kernel PCA, Canonical Correlation Analyses.

I Introduction

The face is our primary focus of attention in social intercourse, playing a major role in conveying identity and emotion. Although the ability to infer intelligence or character from facial appearance is suspect, the human ability to recognize faces is remarkable. We can recognize thousands of faces learned throughout our lifetime and identify familiar faces at a glance even after years of separation. This skill is quite robust, despite large changes in the visual stimulus due to viewing conditions, expression, aging, and distractions such as glasses, beards or changes in hair style.

The performance of a real-world face recognition system usually declines when the input face images are degraded seriously, such as low-resolution (LR) with size of only 12×12 pixels. This is a critical problem for surveillance circumstances. Compared with high-resolution (HR) images, these LR images lose some discriminative details across different persons. These low-resolution (LR) images are common in practice, usually caused by the limited accuracy of available hardware and capturing device. Thus, enhancement of recognition performance under LR conditions is desirable in various applications. Typical scenarios include security surveillance, where subjects are far away from the camera and their faces are quite small in the field of view. Another application of face recognition for LR images is to automatically organize group photos in digital family albums or social networking service. In this paper, we focus on improving the recognition performance in the case where only a single face “snapshot” of LR is available.

Researchers in the machine learning community strive for devising sophisticated classifiers (recognizers) in order to increase the recognition rate on inputs of low quality. Jun Liu proposed that, the point of notice that the facts that every image matrix can always have the well-known Singular Value Decomposition (SVD) and can be regarded as a composition of a set of base images generated by SVD, and can be further point out that the leading base images (those corresponding to large singular values) on one hand are sensitive to the aforementioned facial variations and on the other hand dominate the composition of the face image [1]. Kwak&Pedrycz presented a technique concerned with an enhanced independent component analysis (ICA) and its application to face recognition. Typically, face representations obtained by ICA involve unsupervised learning and high-order statistics [2].

A new method of face recognition based on fuzzy clustering and parallel NNs is proposed by Jianming Lu [3]. The face patterns are divided into several small-scale neural networks based on fuzzy clustering and they are combined to obtain the recognition result. The most widely used Fuzzy Clustering Algorithm is the FCM algorithm, is a data clustering algorithm in which each data point is associated with a cluster through a

membership degree. This technique divides a collection of data points into fuzzy groups and finds a cluster center in each group such that a cost function of a dissimilarity measure is minimized. Independent component analysis (ICA), which finds linear transformation of the data that maximize the statistical independence, appeared in the last two decade as a new data analysis tool proposed by Sezel [4].

To super-resolve faces, adapt the robust method in which models the image formation process and does not rely on a facial image prior, thus avoiding hallucination. As is typically done for super-resolution methods will describe the algorithm using standard notation from linear algebra, assuming each image has all of its pixel values in a vector. In the actual implementation, the solution process is carried out with more practical operations on 2D pixel arrays [17]. The system presented here is based on a face recognition system described in. In the proposed system, individual faces were represented by a rectangular graph, each node labeled with a set of complex Gabor wavelet coefficients, called a jet. Only the magnitudes of the coefficients were used for matching and recognition [18].

The main disadvantages of conventional methods are pose, illumination and facial expressions and also low resolution images. To overcome the drawback of conventional method a new algorithm is designed. Specifically the holistic PCA features of training HR and their corresponding LR face images in the training phase are calculated. The PCA features give a set of components in feature space which may not be easily interpretable in terms of the input space. To overcome that problem Kernel PCA is presented and KPCA features of HR and LR are calculated. Subsequently, CCA is applied to extract coherent features that have maximal correlation between the training HR and LR features. In order to directly connect the LR features to their HR counterparts, RBFs are employed to construct the nonlinear mappings between the features in the coherent subspaces. Given an input LR face image, the coherent SR feature is obtained by mapping the LR feature via the learnt RBFs in the coherent subspace for recognition. Higher recognition rates can be achieved by a simple NN classifier.

The rest of this paper is organized as follows. In Section II, review related works on SR for face recognition and applications of CCA and RBF model are briefed. In Section III, the framework of proposed method is introduced. Section IV gives the details of proposed method, and is followed by extensive experiments in Section V. Section VI concludes this paper.

II. Related Works

The work applies KPCA, CCA and RBF to feature domain SR for the recognition of LR face images. The most relevant works are briefly reviewed.

Kernel principal component analysis (kernel PCA) is an extension of principal component analysis (PCA) using techniques of kernel methods. Using a kernel, the originally linear operations of PCA are done in a reproducing kernel Hilbert space with a non-linear mapping [8]. The K-PCA based Face recognition reduces the no. of features to be compared in the processing. To understand the utility of kernel PCA, particularly for clustering, observe that, while N points cannot in general be linearly separated in $d < N$ dimensions, they can almost always be linearly separated in $d \geq N$ dimensions. That is, given N points, x_i , if we map them to an N -dimension space with $\Phi(x_i) = \delta_{ij}$ where $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^N$ and Φ_{ij} is the Kronecker delta. Kernel PCA has been demonstrated to be useful for novelty detection and image de-noising.

SR techniques are central to a variety of applications ranging from digital photography to publishing. Furthermore, face image SR methods are often applied to enhance the face recognition rate of LR image sequences. Lin *et al.* [5] applied optical flow SR algorithm as a preprocessing stage to improve the face recognition performance of LR face images. A sequence of video frames of a subject is also applied for creating a SR image of the face with increased resolution and reduced blur for face recognition. Given a single LR face image, Jia *et al.* generated the SR identity parameter vector for recognition by incorporating the tensor structure that models multiple factors into the similar Bayesian framework as that in [6]. These studies strive for an effective approach to combining information from multiple images/sources into recognition. Instead, we aim at extracting SR features from a single LR image which is suitable for performance improvement on NN classifiers.

CCA was first developed by Hotelling to find bases for two sets of random vectors such that the correlation between the projections of the vectors onto the bases is maximized. Classical CCA has been generalized in various ways, such as kernel CCA to maximize nonlinear correlation, and tensor CCA for multiple sets of variables [9]. CCA and its extensions can be used whenever there is a need to establish a relationship between two sets of variables. The main difference between CCA and KPCA is that CCA is closely related to mutual information. CCA can also be used to measure the similarity between two image sets for object and action recognition [9]. In this paper, we apply CCA to establish the coherent subspaces for HR and LR face images.

RBF was introduced by Broom head and Lowe for the purpose of exact function interpolation. Algorithms based on RBFs are commonly applied for statistical learning, geometric data analysis, and pattern recognition. RBFs can also be applied for the problem of reconstructing a surface from scattered points sampled

on a physical shape. As pointed out, RBF neural networks are best suited for learning continuous or piecewise continuous [10] approximation and the RBF neural classifier was applied for face recognition to cope with small training sets of a high-dimensional problem efficiently [11]. The SR face image corresponding to the input LR face feature was obtained by RBF mappings. Here, RBF-based mapping to build the regression model between the features of LR and HR face images are applied.

III. Formulation of Problem and Algorithm Overview

The holistic features, specifically the KPCA features, are applied for the recognition of LR face images, since local features are not applicable any more for the images of very low resolution in our applications. The problem of feature-domain SR for LR face recognition turns out to be the inference of SR features c_h from an input LR image I_l given a training set consisting of HR images I_h and their corresponding LR versions I_l .

The high-dimensional structure formed by face images in the high-dimensional pixel space is homeomorphism with a geometric structure in lower dimensional pixel space, and the down sampling process preserves the intrinsic structures in the high-dimensional image manifold. This means that the features of HR and LR face images share a common topological structure, and thus they are coherent through the structure. As stated, in the recognition of LR face images, the KPCA features are generally applied. However, this coherence does not always hold in practice in the KPCA space. We need to find a feature subspace where the coherence between the topological structure of HR and LR face images is established and the HR feature can be estimated more accurately.

CCA is a statistical method to study the linear relationships between two sets of variables with two or more variables in each set and to determine the particular variables which attribute to these relationships. This method can be seen as the problem to select the linear functions of the two sets of variables such that the correlation between the two linear functions is maximized. We apply CCA transformation to the KPCA feature sets of LR and HR face images in order to find the coherent feature subspaces, in which the correlation of the topological structures of LR and HR is maximal. The topological structures are more coherent and it is easier to establish the mapping relationship after CCA transformation. We apply the RBF-based mapping to build the regression model between the features of HR and LR face images in the coherent subspace by taking advantages of the salient features of RBF regression such as fast learning and generalization ability.

Figure 1 provides the flow chart of the proposed method. Our approach is divided into training, testing and classification phases. The corresponding HR and LR face image sets are used for training to obtain the base vectors of CCA transformation and the parameters of RBF regression. In the testing stage, we calculate the KPCA coefficients of a given LR image and project the KPCA features into the coherent subspace using the learnt base vectors. Hence, the SR coherent feature corresponding to the given input LR face image can be obtained by simply applying the learnt RBF mappings. And, an NN classification is performed on these super-resolved features for face recognition.

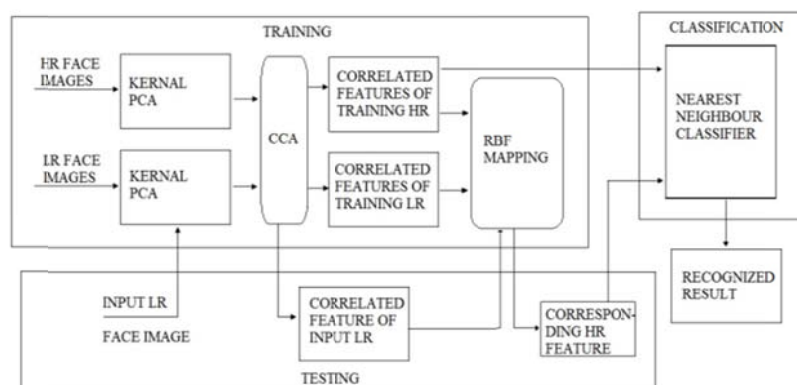


Figure 1 Flow chart of proposed method

IV. Proposed Algorithm

A. Kernel PCA

Kernel principal component analysis (kernel PCA) is an extension of principal component analysis (PCA) using techniques of kernel methods. It is easy to construct a hyper plane that divides the points into arbitrary clusters. Of course, this Φ creates linearly independent vectors, so there is no covariance on which to perform eigen decomposition explicitly as we would in linear PCA. Instead, in kernel PCA, a non-trivial, arbitrary Φ function is 'chosen' that is never calculated explicitly, allowing the possibility to use very high dimensional Φ 's if we never have to actually evaluate the data in that space. Since we generally try to avoid working in the Φ -space, which we will call the 'feature space', we can create the N-by-N kernel by

$$K = k(x, y) = (\Phi(x), \Phi(y)) = \Phi(x)^T \Phi(y) \quad (1)$$

which represents the inner product space (see Gramian matrix) of the otherwise intractable feature space. The dual form that arises in the creation of a kernel allows us to mathematically formulate a version of PCA in which we never actually solve the eigenvectors and Eigen values of the covariance matrix in the $\Phi(x)$ -space (see Kernel trick). The N-elements in each column of K represent the dot product of one point of the transformed data with respect to all the transformed points (N points). To evaluate the projection from a point in the feature space $\Phi(x)$ onto the kth principal component V (where exponent k means the component k, not powers of k) is

$$V^{kT} \Phi(x) = (\sum_{i=1}^N a_i^k \Phi(x_i))^T \Phi(x) \quad (2)$$

The Gaussian kernel is defined as

$$k(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} \quad (3)$$

B. Feature Vector Extraction

In this section, the detailed procedure of proposed algorithm is present. Different from the mixture models, kernel PCA just works with a single PCA. It is an extension of PCA to non-linear distributions. Instead of directly doing a PCA, the n data points x_i are mapped into a higher-dimensional (possibly infinite-dimensional) feature space [12]. As stated, the problem of SR of feature domain for face recognition is formulated as the inference of the HR domain feature c_h from an input LR image I_l , given the training sets of HR and LR face images, $I^H = \{I_i^H\}_{i=1}^m$ and $I^L = \{I_i^L\}_{i=1}^m$ where m denotes the size of the training sets. The dimension of the image data, which is much larger than the number of training images, leads to huge computational costs. So, the holistic features of face images are obtained by KPCA, which represents a given face image by a weighted combination of eigenfaces. We define

$$x_i^H = (B^H)^T (I_i^H - \mu^H) \quad (4)$$

where μ^H is the corresponding mean face of HR training face images and x_i^H is the feature vector of face image I_i^H . B^H is the feature extraction matrix obtained by the HR training face images and is made up of orthogonal eigenvectors of $(\hat{I}^H)^T \times \hat{I}^H$ corresponding to the Eigen values being ordered in descending order. Similarly, the feature of LR face image is represented as

$$x_i^L = (B^L)^T (I_i^L - \mu^L) \quad (5)$$

Where B^L and μ^L are the feature extraction matrix and the mean face obtained by LR training face images, respectively. Then, we have the PCA feature vectors of HR and LR training sets. The following process of our algorithm is based on these KPCA feature vectors.

C. Canonical Correlation Analysis

Canonical correlation analysis has been used to study the correlations between two sets of variables. In our study of feature-domain SR for LR face recognition, the relationship between HR and LR feature vectors should be learned by the training sets. Thus, given an input LR face features, the corresponding SR features can be obtained for recognition. In the existing methods, this relationship is directly obtained by the KPCA features of LR and HR face images. Corresponding HR and LR images of the same face have differences only in resolution, thus, they are coherent through their intrinsic structures. In order to learn the relationship between HR and LR feature vectors more exactly, we apply CCA to incorporate the intrinsic topological structure as the prior constraint. In the coherent subspace obtained by CCA transformation, the solution space of HR feature corresponding to a given LR image is reduced. Then, the more exact coherent SR features can be obtained for recognition in the coherent subspace.

Specifically, from the PCA feature training sets X^H and X^L , we first subtract their mean values X^H and X^L , respectively, which yields the centralized data sets. CCA finds two base vectors

V^H and V^L for datasets X^H and X^L in order to maximize the correlation coefficient between vectors C^H and C^L . The correlation coefficient is defined as

$$\rho = \frac{E[C^H C^L]}{\sqrt{E[(C^H)^2] E[(C^L)^2]}} \quad (6)$$

Where $E[C^H C^L]$ denotes mathematical expectation. To find the base vectors V^H and V^L , we define $c_{11} = [x^{\sim H} (x^{\sim H})^T]$ and $c_{22} = [x^{\sim L} (x^{\sim L})^T]$ as the within-set covariance matrices of X^H and X^L , respectively, while $c_{12} = [x^{\sim H} (x^{\sim L})^T]$ and $c_{21} = [x^{\sim L} (x^{\sim H})^T]$ as their between-set covariance matrices. Then, we compute

$$R_1 = C_{11}^{-1} C_{12} C_{22}^{-1} C_{21} \quad (7)$$

$$R_2 = C_{22}^{-1} C_{21} C_{11}^{-1} C_{12} \quad (8)$$

V^H is made up of the eigenvectors of R_1 when the Eigen values of R_1 are ordered in descending order. Similarly, the eigenvectors of R_2 compose V^L . We obtain the corresponding projected coefficient sets C^H and C^L of the KPCA feature sets X^H and X^L projected into the coherent sub spaces using the following base vectors:

$$C_i^H = (V^H)^T X_i^H \quad (9)$$

$$C_i^L = (V^L)^T X_i^L \quad (10)$$

As there exists a coherent intrinsic structure embedded in the HR and LR feature sets X^H and X^L , the correlation between the two sets C^H and C^L is increased and their topological structures are more coherent after the transformation. Then, the relationship between HR and LR features is more exactly established in the coherent subspace.

D. RBF Mapping

As the coherent subspace is obtained, the nonlinear mapping relationship between the coherent features of HR and LR will be learned by the training features. This problem can be formulated as finding an approximate function to establish the mapping between the coherent features of HR and LR face images. RBFs are typically used to build up function approximations. So, we apply RBF to construct the mapping relationship. RBF uses radial symmetry function to transform the multivariate data approximation problem into the unary approximation problem, and can interpolate no uniform distribution of high-dimensional data smoothly.

The form of RBFs used to build up function approximations is

$$f_i(\cdot) = \sum_{j=1}^m w_j \phi(\|t_i - t_j\|) \quad (11)$$

where the approximating function $f_i(\cdot)$ is represented as a sum of m RBFs $\phi(\cdot)$, each associated with a different center t_j , and w_j is the weighting coefficient. The form has been particularly used in nonlinear systems. In our implementation, we apply multi quadric basis function

$$\phi(\cdot) = \sqrt{\|t_i - t_j\|^2 + 1} \quad (12)$$

In order to apply RBFs, first, we train the weighting coefficients by training coherent features of HR and LR face images. The approximate value we want to obtain is the coherent HR features, while the input value is the coherent features of LR face images. So, in the training stage, we substitute the coherent features of LR face images C_i^L and C_j^L for t_i and t_j , and the coherent HR feature C_i^H corresponding to C_i^L for f_i . The aim of RBFs is to establish the nonlinear mappings between C_i^H and C_i^L .

E. Super Resolution for Recognition

We feed the coherent features super-resolved from the features of LR faces to an NN classifier to achieve the face recognition. In the testing phase, given an LR face image l , the KPCA feature vector x_l of the input face image is computed as

$$x_i = (B^L)^T (I_l - \mu^L) \quad (13)$$

In our algorithm, we execute our recognition process in the coherent subspaces. So, the PCA feature vector X_l is transformed to the coherent subspace using

$$c_i = (V^L)^T (x_i - x^{-L}) \quad (14)$$

The coherent SR feature c_h is obtained by feeding the coherent feature of the LR face image c_l to the trained RBF mapping. We will get

$$c_h = w \cdot [\varphi(c_1^L, c_i), \dots \dots \varphi(c_m^L, c_i)]^T \quad (15)$$

Finally, we apply the coherent feature c_h and $C^H = \{c_i^H\}_{i=1}^m$ for recognition based on the NN classification with L2 norm

$$g_k(c_h) = \min(\|C_h - C_{ik}^H\|_2) \quad i=1,2,\dots,m \quad (16)$$

Where C_{ik}^H represents the i th sample in the k th class in C^H .

V. Experimental Results And Discussion

The experiments are performed on the YALE face database. In order to demonstrate the effectiveness of designed algorithm, comparison is done for the face recognition rate with other methods.



(a) Mean image of HR

(b) Mean image of LR



(c) Training of HR images



(d) Training of LR images



(e) Testing input LR image (f) Recognized image

Figure 2.face images in Yale Database

The proposed algorithm undergoes 3 phases i.e training, testing and classification. The figures (a)-(d) are training images, the figure(e) is the testing stage input image and the figure (f) is the recognized result.

TABLE 1 CUMULATIVE RECOGNITION RESULTS FOR YALE DATABASE

Rank	1	2	3	4	5
KPCA	0.950	0.965	0.980	0.985	0.990
PCA	0.930	0.945	0.970	0.9775	0.985
CLPM	0.915	0.930	0.955	0.960	0.965
WRANG	0.910	0.930	0.945	0.950	0.975

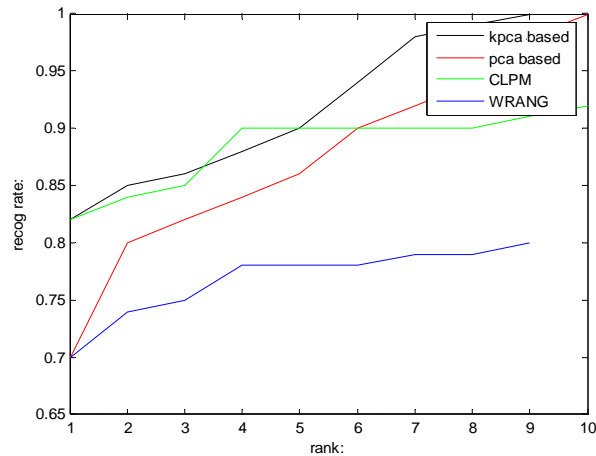


Figure 3. Comparison with other methods

The figure 3 shows the comparison of recognition rate of proposed method with other methods. From the results the proposed method achieves higher recognition rate.

The table 1 shows that cumulative results for Yale database. From the table it is clear that the recognition rate for KPCA method is higher than the other methods. So compared to other methods, KPCA method achieves the good recognition result.

VI. Conclusions

For the problem of LR face images resulting in lower recognition rate, a KPCA method in the feature domain for face recognition was proposed in this paper. The obtained KPCA features are efficient in dimensionality reduction. CCA was applied to obtain the coherent subspaces between the holistic features of HR and LR face images, and RBF model was used to construct the nonlinear mapping relationship between the coherent features. Then, the SR feature in the HR space of the single-input LR face image was obtained for recognition. Experiments show that even the simple NN classifier can implement high recognition rates in the coherent subspaces. Compared to other feature-domain SR methods, proposed method is more robust under the variations of expression, pose, lighting, and down sampling rate and has a higher recognition rate.

REFERENCES

- [1]. S.-W Lee. and J. Park, "Low resolution face recognition based on support vector data description," *Pattern Recognit.*, vol. 39, no. 9, pp. 1809–1812, Sep. 2006.
- [2]. O. Sezer, Y. Altunbasak, and A. Ercil, "Face recognition with independent component-based super-resolution," in *Proc. SPIE Visual Commun. Image Process.*, vol. 6077. San Francisco, CA, 2006, pp. 52–66.
- [3]. J. Lu, X. Yuan, and T. Yahagi, "A method of face recognition based on fuzzy c-means clustering and associated sub-NNs," *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 150–160, Jan. 2007.
- [4]. K.-C. Kwak and W. Pedrycz, "Face recognition using an enhanced independent component analysis approach," *IEEE Trans. Neural Netw.*, vol. 18, no. 2, pp. 530–541, Mar. 2007.
- [5]. F. Lin, J. Cook, V. Chandran, and S. Sridharan, "Face recognition from super-resolved images," in *Proc. 8th Int. Symp. Signal Process. Appl.*, vol. 2. Sydney, NSW, Australia, Aug. 2005, pp. 667–670.
- [6]. S. Zhou, V. Krueger, and R. Chellappa, "Probabilistic recognition of human faces from video," *Comput. Vis. Image Understanding*, vol. 91, nos. 1–2, pp. 214–245, Jul.–Aug. 2003.
- [7]. H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, 1936.
- [8]. http://en.wikipedia.org/wiki/Kernel_principal_component_analysis
- [9]. T.-K. Kim, S.-F. Wong, and R. Cipolla, "Tensor canonical correlation analysis for action classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, Jun. 2007, pp. 1–8.
- [10]. J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Comput.*, vol. 1, no. 2, pp. 281–294, 1989.
- [11]. M. Er, S. Wu, J. Lu, and H. Toh, "Face recognition with radial basis function (RBF) neural networks," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 697–710, May 2002.
- [12]. <http://www.heikohoffmann.de/htmlthesis/node36.html>

Author's Profile



N.NagaMounica received B.Tech degree in Electronics and Communication Engineering from Bhoj Reddy Engineering College for Women, JNTUH, Hyderabad. She is presently pursuing M.Tech in Digital Electronics and Communication Engineering specialization in Electronics and Communication Department from G. Narayanamma Institute of Technology and Sciences, JNTUH, Hyderabad. Her Research interests are in image processing and Electronics and Communication.