

Automatic Cheque Processing System

Mohit Mehta, *Member, IACSIT*, Rupesh Sancheti and Ajay Marchya

Abstract— Automatic Cheque Processing is one of the most widely researched areas in document analysis and biometric. Various methodologies have been proposed in this area for Automatic Cheque Processing and forgery detection. An account holder gives cheques to another person as account payee or self-cheque. It is been observed that a number of forgery cases have been registered as cheque forgery, where some person has forged the signature of another person and provided a self-cheque to himself. In this paper we propose a mechanism for recognition of cheque fields, like name, amount and also verify the signature and it's authenticity. We propose a unique two stage model of Automatic Cheque processing with detecting skilled forgery in the signature by combining two feature types namely Sum graph and HMM and classify them with knowledge based classifier and probability neural network. We proposed a unique technique of using HMM as feature rather than a classifier as being widely proposed by most of the authors in signature recognition. Results show a higher false rejection than false acceptance rate. Character segmentation accuracy is found to be 95%, character recognition efficiency 83%, Digit recognition efficiency is 91%.and system detects forgeries with an accuracy of 80% and can detect the signatures with 91% accuracy.

Index Terms— Nearest Mean Classifier, PNN, Biometric, Sum Graph, HMM, Forgery Detection

I. INTRODUCTION

The widespread use of bank cheques in daily life makes the development of cheque processing systems of fundamental relevance to banks and other financial institutions. Bank transactions involving cheques are still increasing throughout the world in spite of the overall rapid emergence of electronic payments by credit cards. However, fraud committed in cheques is also growing at an equally alarming rate with consequent losses. Automatic bank cheque processing systems are hence needed not only to counter the growing cheque fraud menace but also to improve productivity and allow for advanced customer services. The automatic processing of a bank cheque involves extraction and recognition of handwritten or user entered information from different data fields on the cheque such as courtesy amount, legal amount, date, payee and signature. A system that is able to read cheque automatically would be very helpful, especially if it is fast and accurate. Even if misclassification occurs, the mistake could potentially be detected during the recognition process; however it is more desirable that the system rejects a cheque in case of doubt so that it can be directed to manual processing from the beginning. In order to

produce a successful cheque processing system, many sub-problems have to be solved such as background and noise removal, recognition of the immense styles of handwriting and signatures, touching and overlapping data in various fields of information and errors in the recognition techniques [1].

In the proposed system we will consider only 1) Courtesy amount recognition 2) Payee Recognition 3) Signature

The study of Automatic Cheque Processing with Skilled Forgery detection involves following independent Studies

- 1) Image Preprocessing (Filtering and Noise Removal)
- 2) Segmentation of Bank Cheque
- 3) Feature Extraction
- 4) Classification
- 5) Statistical analysis

II. RELATED WORK

Describes a novel method for automatically segmenting and recognizing the various information fields present on a bank cheque. The uniqueness of the approach lies in the fact that it doesn't necessitate any prior information and requires minimum human intervention. The extraction of segmented fields is accomplished by means of a connectivity-based approach. For the recognition part, authors have proposed four innovative features, namely; entropy, energy, and aspect ratio and average fuzzy membership values [1].

Elaborates the signature verification for banking applications. Most of the banks issue the cheque books to their customers having different background colors where as the original specimen signature is always taken on the white paper at the time of the opening of the account. When these signatures are verified digitally the results are not accurate. In this paper, we have proposed a mechanism that automates the offline signature verification for bank cheques even with different background colors. Work present in [2] is focused, to examine whether an input signature of colored bank cheque is a genuine signature or a forged. This task is performed by comparing the collected signature samples (white background) with input signatures (colored background). The Signature Verification of Colored Cheques (SVCC) system of verifying the signatures having different background colors in spite of white paper specimens through variation of color intensity is discussed and presented [2].

A hidden Markov model (HMM) based approach is proposed for recognition of offline handwritten Devanagari words. The histogram of chain-code directions in the image-strips, scanned from left to right by a sliding window, is used as the feature vector. A continuous density HMM is proposed to recognize a word image. In [3] the states of the HMM are not determined a priori, but are determined automatically based on a database of handwritten word images. A handwritten word image is assumed to be a string

Mohit Mehta is with MIT Mandasaur, India as lecturer, Tel:0091-9981393677, email: mohit.mehta@mitmandasaur.info

Rupesh Sancheti is with MIT Mandasaur, India as lecturer, Tel:0091-9827213991, email: rupesh.sancheti@mitmandasaur.info

Ajay Marchya is with MIT Mandasaur, India as lecturer, email: ajay.marchya@mitmandasaur.info

of several image frame primitives. These are in fact the states of the proposed HMM and are found using a certain mixture distribution. One HMM is constructed for each word. To classify an unknown word image, its class conditional probability for each [3].

Reports the contribution to signature verification considering different forgery types in an HMM framework. It also elaborates that the signature verification is a function of HMM state transition [4].

III. PROPOSED WORK

We propose ROI based segmentation for cheque segmentation. First we separate Payee Field, Courtesy Field and Signature. After the Separation of the Field Character segmentation is performed only for courtesy digit and payee field and is not applied for the signature field. We use probabilistic neural network for recognition of for courtesy digit and payee field.

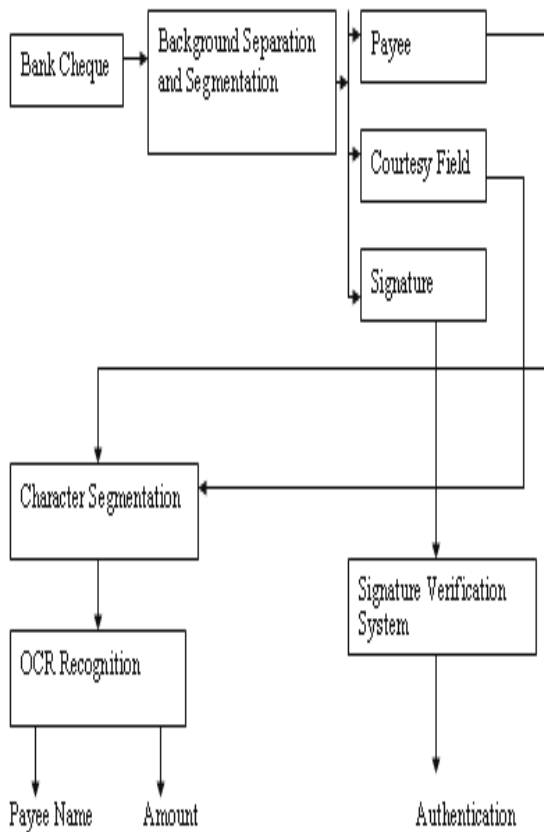


Figure 1: Overall System Block Diagram

We propose a method for signature verification, which is based on Perception and Probability. It means first the system roughly determines to which class a signature belongs to and then it finally decides whether the signature can be accepted or not.

Perception presents the class, which a signature “possibly” belongs and the pattern classification based on state transition determines if at all it belongs to that class. Beside it defines an accurate closeness function. We propose such a system where we combine the spatial features of sum graph and HMM and classify them separately by a PNN and

Knowledge based classifier.

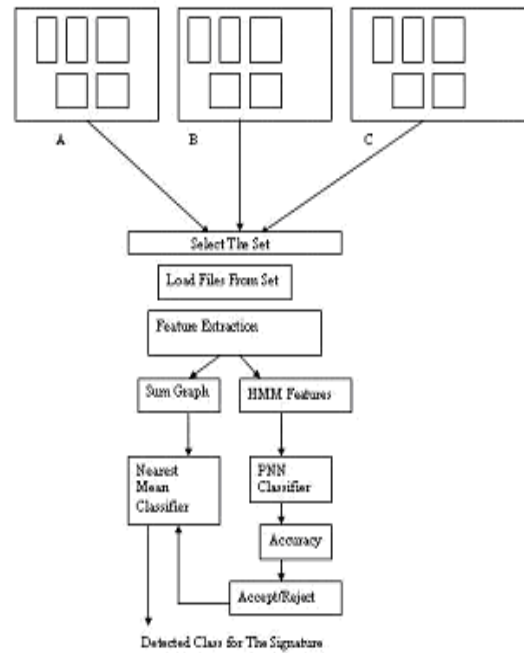


Figure 2: Signature Verification system

A. Preprocessing

Cheque needs to be preprocessed before using it for the verification process.

1) Noise Removal

A simple 3x3 kernel median filtering is applied followed by morphological smoothing to clean the spatial noise characterized by small components. It will eliminate the image abnormalities such as scanning error and ink variation. Resizing the area normalizes the filtered images.

We have not implemented rotation invariance. We have considered that the image follows a same pattern. In some of the works the signatures are rotated to a single monotonous plane. This may not be quite a suitable technique because every person signs at a different slant. Hence slants are also important preserving features

2) Binarization

All gray scale images are converted into binary image. In this process the mean gray level value of the images are extracted and values greater than the mean are mapped to white where as the values lesser than that are marked as black.

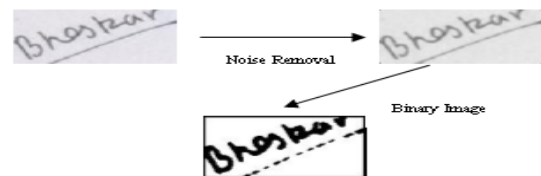


Figure 3: Noise Removal and Binarization

B. Segmentation and Extraction of Fields

For cheque segmentation technique we have considered region based segmentation. The scanned cheque is first resized to 540x540 size. Now the fields are segmented by feeding the crop points. This is done by first applying manual segmentation of the field. Then averaging the points. The

algorithm is as follows.

- 1) First manually segment the payee, signature and courtesy amount field by using imcrop.
- 2) Extract the area for each part.
- 3) Repeat step 1, five times.
- 4) Get the average starting and ending point for the area.
- 5) Select the average points as the seed points.
- 6) Automatically segment the cheque based on the values as determined by step 4.

Once the fields are segmented, they are fed to character segmentation system. Handwritten character segmentation system of ours works with sum graph approach. Here we first convert the character image to binary image. Then we apply 3x3 kernel median filtering, followed by morphological smoothing. Then we convert the image into negative image. It gives white characters on a black background. The sum graph of the image is taken. The image is segmented based on the occurrence of minima in the graph. Once characters are segmented, they are saved as separate image in corresponding directory. Character segmentation is performed only for courtesy digit and payee field and is not applied for the signature field.

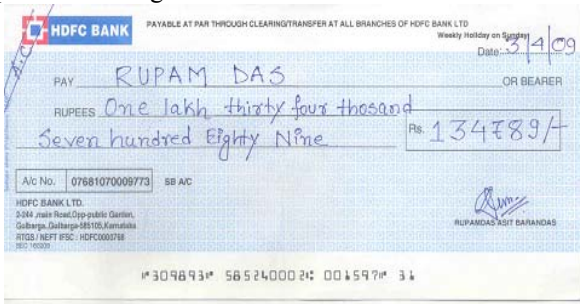


Figure 4: Cheque Image

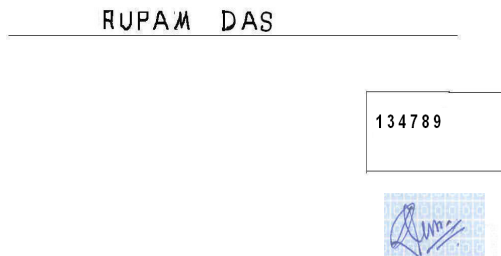


Figure 5: Image after separating the fields

C. Feature Extraction

We consider two different types of features viz: 1) Sum graph and 2) HMM features.

1) Sum Graph

Let us consider that our Input image be $I(i, j)$;

Then the sum graph represents the column wise summation of the color values.

$$S(j) = \sum_{i=1}^N I(i, j) \quad (1)$$

Where N =number of columns.

Sum graph is an important feature in detection of characters or numerals in document analysis, specially handwritten character recognition. The area under each peak represents a distinct feature.

The minima from the beginning and the end are removed

to get only the sun graph corresponding to handwritten part only. We have 1×128 matrix to representing the sum graph feature. Transpose is taken of this matrix to convert it to row vector. This is passed as a feature to knowledge-based classifier (nearest mean).

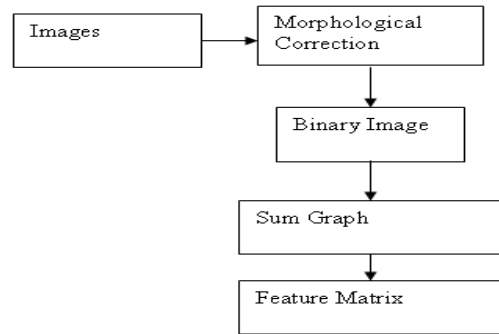


Figure 6: Simple block diagram depicting Sum Graph Feature Extraction

2) Hidden Markov Model

First we determine the probability matrix of appearance of 1 w.r.t 0 and 1 and probability of occurrence of 0 w.r.t 1 and 0. This pattern is unique for all the signatures. As the signatures are considered random in nature, we consider a hidden markov model that can generate such a pattern. This matrix is use to generate the states of hidden markov model. Mean and Standard deviation of this state are extracted and are used as feature vector for PNN.

The Hidden Markov Model is a finite set of states, each of which is associated with a (generally multidimensional) probability distribution. Transitions among the states are governed by a set of probabilities called transition probabilities. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are "hidden" to the outside; hence the name Hidden Markov Model

HMM is quantitatively described by Equation 2

$$\lambda = (A, B, \pi) \quad (2)$$

Where, A is the state transition matrix and gives the probability of transiting from one state to another, B is the observation probability matrix and gives the probability that a particular Observation symbol occurs in a particular state and π is the initial probability distribution for the states.

Here we have obtained the co-occurrence matrix of the binary signature, which gives four probabilities. Hence we construct a set of probabilities, $S = \{ \{1,1\}, \{1,0\}, \{0,1\}, \{0,0\} \}$. Initial probability π is considered as 1. State generated by giving this set as input to the HMM model represents the variability in signature pattern.

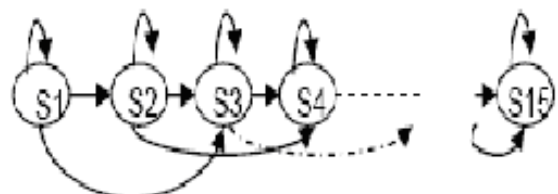


Figure 7: State diagram of HMM model

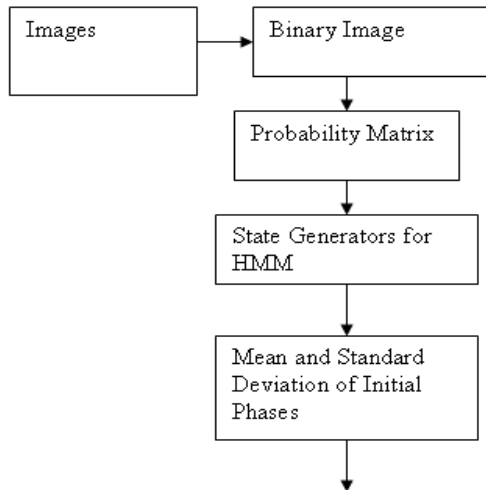


Figure 8: Feature Vector for Neural Network

IV. RECOGNITION OF BANK CHEQUE FIELDS

We adopt character recognition system based on projection histogram technique. Here first the digit and alpha database of handwritten characters and digits are trained. For training, we extract the projection histogram of each feature. These features are fed to a neural network. We use separate network for digit and alphabet recognition. Then the segmented characters are extracted one by one from the directory where they were saved and recognized by the neural network.

Each state now represents a matrix, which can define a probability matrix as calculated prior to HMM. The statistical mean and standard deviation of each value is taken as the training vectors for a Probabilistic Neural Network. We use Probabilistic Neural Network for classifying a signature as one of the signature of the database. Here there are total N input layer, which represent each signature in the database. The mapping function is weight matrix with standard deviation of the first feature value

Neural network is a trained based on input statistical feature. Output vector is the class representing the signature person. The result of the neural network is further mapped with the results of the knowledge-based classifier. A knowledge-based classifier (Nearest mean) is designed to classify the signature based on sum graph features.

Differential of the sum graph points from each sample w.r.t the input vector is calculated. Then the smallest of them is calculated. The smallest set represents the detected class. The accuracy of classification is measured. If the accuracy is less than 80% than the class is rejected.

We use nearest mean classifier as the knowledge based classifier. The nearest mean classifier considers the linear set of values as the training sets. These can be directly the image matrix or the frequency transformed data or some other information. Training process groups these values in a set and labels them as the classes. The input image features would be extracted and their closeness with all the training classes would be compared. Whichever class values are closer to the features (maximum match of minimum deviation from the respective class values) is detected as the matching class for the signature, as shown in figure 10.

For calculating the distance between the training class vectors and test vectors we use following distance measure.

$$d_k = \sum (x_s - k_s) \quad (3)$$

Where d_k represents the distance of test sample from k th class. x_s is the sum graph of test sample and k_s is the sum graph of k th class.

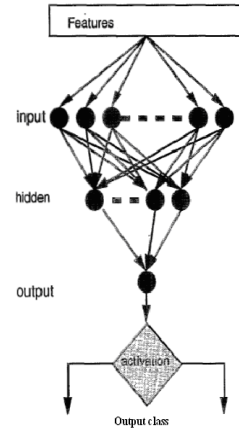


Figure 9: Neural Network Classifier

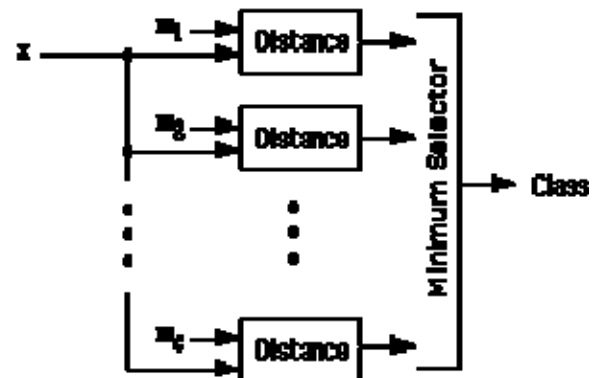


Figure 10: Knowledge based Minimum Distance classifier, m_1, m_2, \dots represents the training features, x represents the test sample

V. RESULTS AND ANALYSIS

The results show that the system maintains a good correct classification rate for any number of classes. The lowest rate of correct classification is 86% and the highest is 92%. We define correct classification as follows.

$$\text{Rate of Correct Classification} = \frac{\text{Total Number of Input Classes} - (\text{False Accepted Test} + \text{False Rejected Test})}{\text{Total Number of Input Classes}} \times 100 \quad (4)$$

Where, (False Accepted Test+ False Rejected Test) presents the total false classification.

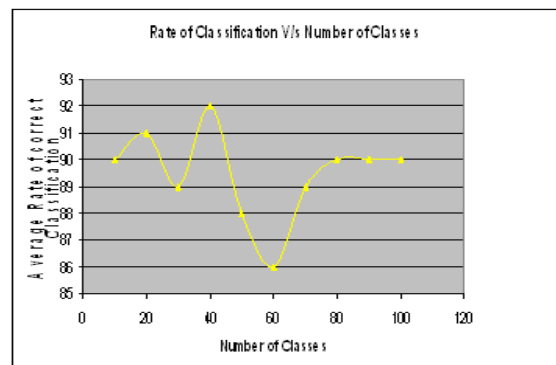


Figure 11(Result1): Number of classes V/s Rate of Classification

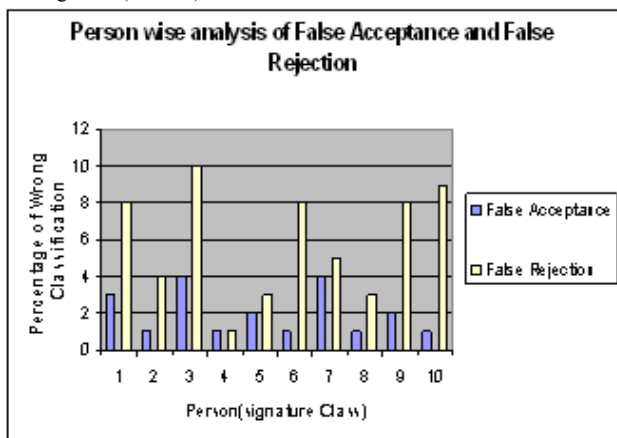


Figure 12(Result 2): Person wise analysis

Figure 12 Shows the Graph of independent false acceptance and rejection rate for 10 different classes. The graph shows that for most of the user the false rejection rate is higher than the false acceptance rate. It is the most desired condition in biometric authentication and signature techniques. Therefore we can say that our system performance is satisfactory.

VI. CONCLUSION AND FUTURE WORK

We have developed Automatic cheque processing with main emphasis in recognition of skilled forgery. The work can process cheques of a particular bank only. Further techniques can be developed to segment the cheque fields based on visual information rather than ROI based segmentation. The accuracy can be further improved by integrating more statistical moments like variance, skewness etc. Here we have not analyzed the time variant analysis of the persons. i.e. samples of the persons were collected at the same time. By nature of signatures, the signature of same person shows slight variations. The algorithm can be further developed to detect such variability by combining more features. The system can be more accurate if more efficient handwritten character recognition is implemented.

REFERENCES

- [1] Vamsi Krishna Madasu, Brian Charles Lovell, "Automatic Segmentation and Recognition of Bank Cheque Fields", Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA 2005), 2005, ISBN: 0-7695-2467-2, pp. 33
- [2] Javed Ahmed Mahar, Mohammad Khalid Khan, Prof. Dr. Mumtaz Hussain Mahar, "Off-Line Signature Verifications of Bank Cheque Having Different Background Colors", IEEE/ACS International Conference on Computer Systems and Applications 2007, 738-745(2007)
- [3] Bikash Shaw, Swapan Kumar Parui, Malayappan Shridhar, "Offline Handwritten Devanagari Word Recognition: A holistic approach based on directional chain code feature and HMM", Proceedings of the 2008 International Conference on Information Technology (ICIT), 2008, ISBN: 978-0-7695-3513-5, pp. 203-208
- [4] Edson Justino, Edson J.R. Justino, Flavio Bortolozzi, Robert Sabourin, "Off-Line Signature Verification Using HMM for Random, Simple and Skilled Forgeries", Sixth International Conference on Document Analysis and Recognition, 2001, ISBN: 0-7695-1263-1, pp.1031-1034
- [5] Debasish Jena, Banshidhar Majhi, Saroj Kumar Panigrahy, Sanjay Kumar Jena, "Improved Offline Signature Verification Scheme Using Feature Point Extraction Method", 7th IEEE International Conference

- on Cognitive Informatics (ICCI'08), 2008, ISBN: 978-1-4244-2538-9, pp. 475-480
- [6] Abhay Bansal, Divye Garg, Anand Gupta, "A Pattern Matching Classifier for Offline Signature Verification", First International Conference on Emerging Trends in Engineering and Technology, 2008, ISBN: 978-0-7695-3267-7, pp.1160-1163
- [7] Filipe Coelho, Luis Batista, Luis F. Teixeira, Jaime S. Cardoso, "Automatic System for the Recognition of Amounts in Handwritten Cheques", Proceedings of the International Conference on Signal Processing and Multimedia Applications (SIGMAP 2008), 2008, ISBN: 978-989-81111-60-9, pp. 320-324
- [8] Jing Wen, Bin Fang, YuanYan Tang, TaiPing Zhang, "Offline signature verification: A new rotation invariant approach", IEEE International Conference on Systems, Man and Cybernetics (ICSMC), 2007, ISBN: 978-1-4244-0991-4, pp. 3583-3587
- [9] Drouhard J-P, Sabourin R, Godbout M, "Evaluation of a training method and of various rejection criteria for a neural network classifier used for off-line signature verification", IEEE International Conference on Neural Networks (ICNN), 1994, ISBN: 0-7803-1901-X, Volume: 7, pp. 4294-4299



Mohit Mehta, received B.E. degree in Information Technology from MIT, Mandsaur, Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal (M.P.), India in 2005 and received M.Tech in Information Technology from Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal (M.P.), India in 2010. He has Four & Half years of teaching experience in the field of Information Technology. Currently, he is working as Lecturer in Information Technology Department with

MIT, Mandsaur (M.P.), India. His research interests are in Pattern recognition. He has published one International and three national research papers. He is a member of Indian Society for Technical Education (ISTE) and International Association of Computer Science & Information Technology (IACSIT).



Rupesh Sanchati, received B.E. degree in Electronics and communication from MIT, Mandsaur, Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal (M.P.), India in 2004. He has Six years of teaching experience in the field of Electronics and Computer. Currently, he is working as Lecturer in Computer Science Department with MIT, Mandsaur (M.P.), India. His research interests are in Pattern recognition. He has published one International and one national research papers.



Ajay Marchya, received his B.E. degree in Information Technology from MIT, Mandsaur, Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal (M.P.), India in 2005. He has three years of industrial experience in the software company Patni Computers Systems LTD. (Mumbai) and one and half year of teaching experience in MIT Mandsaur (M.P.). Currently, he is working as Lecturer in Information Technology Department with MIT, Mandsaur (M.P.), India. His research interests are in Pattern recognition.