Handwritten Digit Recognition using DCT and HMMs

Syed Salman Ali, Muhammad Usman Ghani

Lahore, Pakistan

Email:{syedsalmanali08, ch.usman.ghani}@gmail.com

Abstract—**Handwritten digits recognition has been an interesting area due to its applications in several fields. Recognition of bank account numbers and zip codes are a few examples. Handwritten digits recognition is not a trivial task due to presence of large variation in writing style in available data. In order to cope with this problem both features and classifier need to be efficient. In this research, transformation based features, Discrete Cosine Transform (2D-DCT), have been used. Hidden Markov models (HMMs) have been applied as classifier. The proposed algorithm has been trained and tested on Mixed National Institute of Standards and Technology (MNIST) handwritten digits database. The algorithm provides promising recognition results on MNIST database of handwritten digits.**

I. INTRODUCTION

Pattern recognition has become immense important due to ever demanding need of artificial intelligence in practical problems. Handwritten digits recognition is one such problem in which images of digits written by different authors are recognized by machines. In past, significant research has been done in the area of handwritten digits recognition.

Handwritten text recognition is an active area of research in pattern recognition. Handwritten digits recognition is a complex problem due to the fact that variation exists in writing style of different writers. The phenomenon that makes the problem more challenging is the inherent variation in writing style at different instances. Due to this reason, building a generic recognizer that is capable of recognizing infinitely large writers handwritten digits is almost impossible. It has applications in various fields, ranging from recognizing amounts from checks to recognizing postal codes on postal envelopes.

Similar to other pattern recognition problems, handwritten digit recognition requires efficient features along with rigorous classifier. Input images are highly correlated and thus cannot be used for classification as in its original form. Features are extracted to remove correlation and redundancy from input images. There are three different types of features used in document recognition: 1) Structural features, 2) Statistical features and 3) Transformation based features. The structural features extract structural information from image contents. The statistical features focus on the number or ratio of the black and white pixels in the certain region of images. Transformation based features convert the image signal from spatial domain to the desired domain.

The classification problem can also be divided into two categories 1) Supervised learning and 2) Unsupervised learning. In supervised learning labeled data is given to classifier for training. Whereas in unsupervised learning no labeling of data is required, classifier automatically defines the boundary among different classes of data. In proposed research transformation based features , i.e. Discrete Cosine Transform (DCT), has been used to extract features. For classification, supervised learning classifier, i.e. Hidden Markov Models (HMMs),have been used.

Rest of the paper is structured as follows: different researches reported in literature on subject matter are briefly described in section II, section III discusses methodology to conduct this research work including feature extraction and recognition tasks, results and discussion are given in section IV; and section V presents conclusion.

II. RELATED RESEARCH

Gorgevik et al. [1] developed support vector machines (SVM) based handwritten digits recognition system. They extracted four types of features from each digit image 1) projection histograms, 2) contour profiles, 3) ring-zones and 4) Kirsch features. They reported 97.27% recognition accuracy on National Institute of Standards and Technology (NIST) [2] handwritten digits database [3], when four types of features were used collectively. In [4] Chen et al. proposed max-min posterior pseudo-probabilities framework for handwritten digits recognition. They extracted 256-D directional features from the input image. Finally 256-D features were transformed into 128-D feature using Principal Component Analysis (PCA). They reported recognition accuracy of 98.76% on NIST database [3].

In [5] Labusch et al. proposed a sparse coding based feature extraction method with SVM as a classifier. They reported recognition accuracy of 99.41% on MNIST handwritten digits database [6]. In [7] Yang et al. proposed supervised matrix factorization method used directly as multi-class classifier. They reported recognition accuracy of 98.71% with supervised learning approach on MNIST handwritten digits database [6]. In [8] mixture of multi-class logistic regression model was proposed. They reported recognition accuracy of 98% on the Indian digit database provided by CENPARMI [9]. In [10] wavelet analysis based technique for feature extraction was reported. For classification, SVM and K-Nearest Neighbor (KNN) were used. An overall recognition accuracy of 97.04% was reported on MNIST digit database [6].

In [11] structural features have been used. The KNN classifier is employed for classification and recognition. They reported recognition accuracy of 96.94% on 5000 samples of MNIST digit database [6]. In [12] AlKhateeb et al. proposed Arabic handwritten digit recognition system using Dynamic
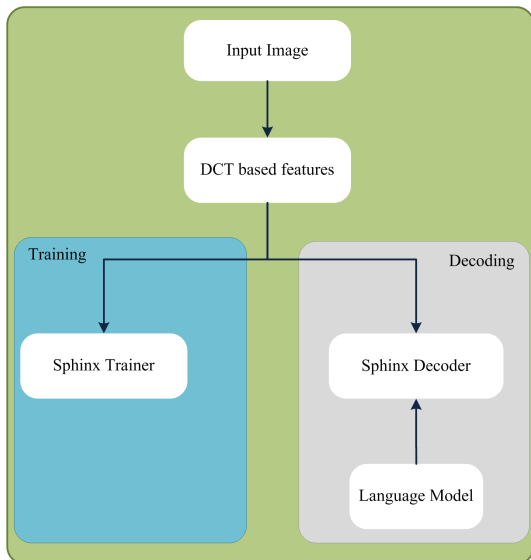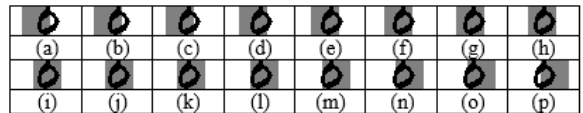
Fig. 1. System flow chart

Fig. 2. Illustration of window shifting for features extraction

Fig. 3. Selection of low frequency coefficients as features from 2D-DCT Matrix

Bayesian Network. They employed DCT coefficients based features for classification. They trained and tested the system on Arabic digits database (ADBase) which contains 70,000 Arabic digits [13]. They reported average recognition accuracy of 85.26% on 10,000 samples.

III. METHODOLOGY

In current research work application of 2D-DCT has been investigated for hand written digits recognition. HMMs have been used for classification and recognition. The overall flow chart for training and decoding is shown in Figure 1. The proposed system has two major modules: features extraction and classification. The detailed methodology is discussed in this section.

A. Feature Extraction

Features are of immense importance in every machine vision application. In current work global transformation based features, i.e. 2D-DCT [14], have been employed. Other transformation based techniques such as PCA are also used to extract features. However in contrast to PCA, DCT uses predefined basis images and is more favorable where extra computation is not desired. 2D-DCT has been widely used for data compression techniques in image processing [15]. Joint Photographic Experts Group (JPEG) standard [16] uses 2D-DCT to compress images. In image processing, 2D-DCT transforms image signal from spatial domain into frequency domain. It provides basis images and corresponding weight-age of each basis image. DCT has amazing property that it compresses the energy of input signal to low frequency components. Natural images do not contain much variation. This fact along with energy compression property of DCT enables to reconstruct the input image using low frequency components only. The coefficients of 2D-DCT have been used as features in current work.

In order to extract features a window of m x m is moved horizontally along the input image. Overlapping windows have
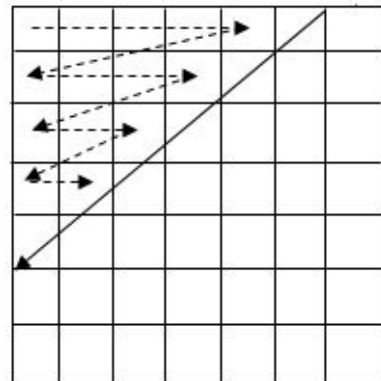
been used. The windows are shifted by one pixel to the right, as illustrated in Figure 2. The shaded portions in figure represent windowed image.

2D-DCT is applied on each windowed image. The resultant 2D-DCT matrix contains low as well as high frequency components. The coefficients of low frequency basis image are selected as features. The optimal number of coefficients to be selected as features have been found empirically. In order to select low frequency coefficients as features, an anti-diagonal is placed in the resultant matrix, as shown in Figure 3. The coefficients that lie above the diagonal are low frequency coefficients; and are selected as features in current work. The selection of coefficients is initiated from the top left corner of 2D-DCT matrix is shown in Figure 3. The selected coefficients are used to generate the feature vectors. These feature vectors are generated for each window. All feature vectors are combined and used for classifier training and recognition.

B. Classification and Recognition

HMMs [17] are probabilistic state machines in which transition from one state to another is hidden. HMM satisfies the Markov chain property in which next state is only dependent upon the current state. In current research work observation sequence is generated as windows slide across the image. Considering the Markov assumption that each observation sequence is only dependent upon the previous observation sequence, classification problem can be modeled as HMM. A simple HMM where state transitions occur in between neighboring states is used to model the current problem. The working HMM for five states is explained in Figure 4.

In order to develop HMM based classifier for current work, Sphinx[1][18] has been used. Sphinx is a tool kit that is designed

---

[1]Sphinx is speech recognition tool kit developed by Carnegie Mellon University
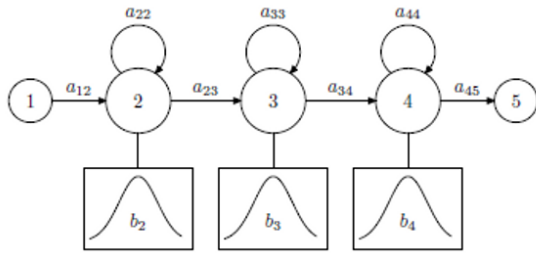
Fig. 4. Hidden Markov Model with five states

TABLE I. ACCURACY RESULTS FOR DIFFERENT FEATURE VECTOR LENGTHS

Feature Vector Length	Accuracy (%)
15	88.08
21	90.76
28	92.41
36	93.21
45	93.99
55	93.72
66	92.74

Fig. 5. Comparison of recognition accuracy for different vector lengths

TABLE II. ACCURACY RESULTS FOR DIFFERENT NUMBER OF GAUSSIAN MIXTURES

Number of mixture of Gaussian	Accuracy (%)
8	93.99
16	95.28
32	96.45
64	96.92
128	97.18
256	97.26

for speech recognition using HMMs. It provides a set of libraries to train and decode HMMs. A HMM with five states for each digit is generated. Baum welch algorithm [19] is used to train HMM parameters of each digit given the observation sequence. The observation sequence consists of 2D-DCT based features, as discussed in previous section. Once the system is trained, decoding is performed using Viterbi algorithm [20] to find optimal state sequence along with the language model.

IV. RESULTS AND DISCUSSION

MNIST handwritten database [6] has been used to train and test performance of proposed system. It contains 60,000 samples for training and 10,000 samples for testing. MNIST dataset is a subset of original NIST handwritten digits dataset [2]. All data samples have fixed size of 20x20 pixels. In current work HMMs have been trained using 60,000 training samples and testing have been performed over 10,000 samples. The first step in the process is to compute the features. The optimal number of 2D-DCT coefficients to be selected as features has been found by performing various experiments. In these experiments, window size has been remained constant i.e. 20. The experiments have been performed using first 15, 21, 28,36,45,55 and 66 coefficients of 2D-DCT progressively. In each experiment HMM with 5 states and 8 mixture of Gaussians have been used. The results of these experiments have been shown in Table I and are plotted in Figure 5. These results show that maximum accuracy is achieved using first 45 low frequency coefficients as features.

Once number of features have been optimized, the next parameter to be finalized is m, i.e. size of window. All image samples in MNIST database are 20x20 dimension. If m is chosen smaller than 20 then the resulting window will ignore the input signal that lies outside the window. On the other hand if m is chosen larger than 20 than it will add no additional information in the computation of features except redundancy. Hence optimal dimension of window to be selected is 20x20.
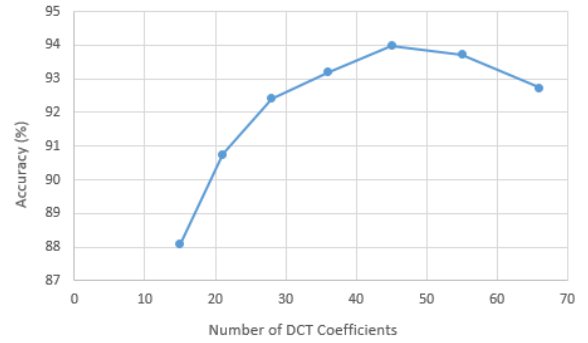
After the selection of number of features and window size, the next parameter to be optimized is number of mixture of Gaussians. This parameter has also been optimized empirically. The experiments have been performed using window size of 20x20. The number of features has been remained fixed as 45. Experiments have been repeated using 8, 16, 32, 64, 128 and 256 mixture of Gaussians. The recognition accuracy for these experiments has been mentioned in Table II and plotted in Figure 6. It is evident from the results that as number of mixture Gaussians increases, accuracy of the system increases.

Digits class-wise accuracy along with total and correct instances is presented in Table III.

V. CONCLUSION

In this paper application of 2D-DCT along with HMM as a classifier has been investigated. Sliding windows based approach and 2D-DCT is applied at each windowed image. The low frequency coefficients of 2D-DCT of the windowed image have been used as features. Considering the Markov chain property these features have been used to train HMM using
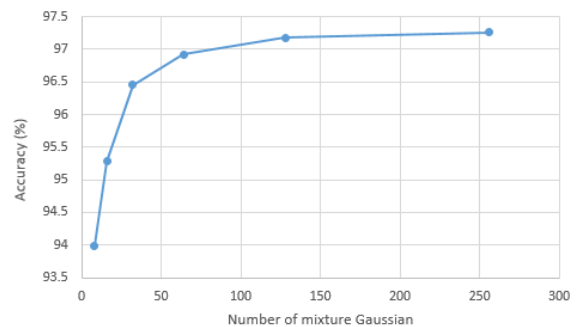

Fig. 6. Comparison of Accuracy for different number of mixture Gaussians

TABLE III.   DIGITS CLASS-WISE ACCURACY

Class	Total Instances	Correct	Accuracy (%)
0	980	971	99.08
1	1135	1127	99.30
2	1032	1014	98.26
3	1010	980	97.03
4	982	946	96.33
5	892	867	97.20
5	892	867	97.20
6	958	937	97.81
7	1028	986	95.91
8	974	943	96.82
9	1009	955	94.65

Sphinx. The number of features along with HMM parameter has been optimized empirically. The results of proposed system on MNIST database are comparable to state of the art.

REFERENCES

[1]   D. Gorgevik and D. Cakmakov, "Handwritten Digit Recognition by Combining SVM Classifiers," in *The International Conference on Computer as a Tool (EUROCON)*, 2005.

[2]   M. D. Garris, J. L. Blue and G. T. Candela, "NIST form-based handprint recognition system," NIST, 1997.

[3]   P. J. Grother, "NIST special database 19 handprinted forms and characters database," National Institute of Standards and Technology, 1995.

[4]   X. Chen, X. Liu and Y. Jia, "Learning Handwritten Digit Recognition by the Max-Min Posterior Pseudo-Probabilities Method," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, 2007.

[5]   K. Labusch, E. Barth and T. Martinetz, "Simple Method for High-Performance Digit Recognition Based on Sparse Coding," *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 19, no. 11, pp. 1985-1989, NOVEMBER 2008.

[6]   Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[7]   J. Yang, J. Wang and T. Huang, "Learning the Sparse Representation for Classification," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2011.

[8]   A. Gimenez, J. Andres-Ferrer, A. Juan and N. Serrano, "Discriminative Bernoulli Mixture Models for Handwritten Digit Recognition," in *International Conference on Document Analysis and Recognition (ICDAR)*, 2011.

[9]   Y. Al-Ohali, M. Cheriet and C. Suen, "Databases for recognition of handwritten Arabic cheques," *Pattern Recognition*, vol. 36, no. 1, pp. 111-121, 2003.

[10] M. S. Akhtar and H. A. Qureshi, "Handwritten Digit Recognition Through Wavelet Decomposition and Wavelet Packet Decomposition," in *Eighth International Conference on Digital Information Management (ICDIM)*, 2013.

[11] U. R. Babu, Y. Venkateswarlu and A. K. Chintha, "Handwritten Digit Recognition Using K-Nearest Neighbour Classifier," in *World Congress on Computing and Communication Technologies (WCCCT)*, 2014.

[12] J. H. AlKhateeb and M. Alseid, "DBN Based learning for Arabic Handwritten Digit Recognition Using DCT Features," in *6th International Conference on Computer Science and Information Technology (CSIT)*, 2014.

[13] S. Abdleazeem and E. El-Sherif, "Arabic handwritten digit recognition," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 11, no. 3, pp. 127-141, 2008.

[14] N. Ahmed, T. Natarajan and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, Vols. C-25, pp. 90-93, 1974.

[15] A. B. Watson, "Image compression using the discrete cosine transform," *Mathematica journal*, vol. 4, no. 1, 1994.

[16] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii-xxxiv, 1992.

[17] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

[18] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf and J. Woelfel, "Sphinx-4: A Flexible Open Source Framework for Speech Recognition," Sun Microsystems, Inc., Mountain View, CA, USA, 2004.

[19] L. R. Welch, "Hidden Markov models and the Baum-Welch algorithm," *IEEE Information Theory Society Newsletter*, vol. 53, no. 4, pp. 10-13, 2003.

[20] G. D. Forney Jr, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268-278, 1973.