# Feature Extraction based on DCT for Handwritten Digit Recognition

**Bouchra EL QACIMY[1], Mounir AIT KERROUM[2] and Ahmed HAMMOUCH[1]**

**[1] Laboratory LRGE, ENSET of Rabat, Mohamed V Souissi University
Rabat, Morocco**

**[2] Laboratory LARIT, Ibn Tufail University, Faculty of Science, ENCG of Kenitra
Kenitra, Morocco**

**Abstract**

Feature extraction is a crucial and challenging step in many pattern recognition problems and especially in handwritten digit recognition applications. However, the extraction of the most informative features with highly discriminatory ability to improve the classification accuracy and reduce complexity remains one of the most important problems for this task. This work investigates the effectiveness of four feature extraction approaches based on Discrete Cosine Transform (DCT) to capture discriminative features for handwritten Digits recognition. These approaches are: DCT upper left corner (ULC) coefficients, DCT zigzag coefficients, block based DCT ULC coefficients and block based DCT zigzag coefficients. The coefficients of each DCT variant are used as input data for Support Vector Machine (SVM) classifier to evaluate their performances. The objective of this work is to identify the optimal feature extraction approach that speed up the learning algorithms while maximizing the classification accuracy. We used the well known MNIST database in two variants: raw data and preprocessed data that dismisses the non-informative region of the observations in the dataset. The results have been analyzed and compared in terms of classification accuracy and reduction rate and the findings have demonstrated that the block based DCT zigzag feature extraction yields a superior performance than its counterparts.

***Keywords:*** *Feature extraction, Handwritten Digit Recognition, DCT, SVM Classification.*

## 1. Introduction

Feature extraction is one of the most crucial and challenging steps in many pattern recognition problems and especially in handwritten digit recognition applications such as postal mail sorting, bank check processing, form data entry, etc. For these applications, the accuracy and speed of digit recognition is crucial to the overall performance [1]. However, the extraction of the most informative features with highly discriminatory ability to improve the classification accuracy and reduce complexity remains one of the most important problems for this task.

In the literature, numerous works have contributed to feature extraction for handwritten digit recognition [2] [3] [4]. On the other hand the discrete cosine transform (DCT) has been widely used in pattern recognition problems. In his work on word based recognition system [5], Alkhateeb used DCT as feature extraction method and his results showed that DCT yielded to a higher recognition rate than its counterparts. In the same vein, Lawgali [6], in his work on Arabic isolated character recognition, compared DCT to discrete wavelet transformation (DWT) and the results showed the effectiveness of DCT features to lead to a better recognition rate. DCT is also used in face recognition [7], video text detection [8], car-plate recognition [9] [10] and Iris recognition [11]. However, little research had addressed the issue of what is the most effective method to retain DCT coefficients for handwritten digit classification.

In this work, we investigate the effectiveness of four feature extraction approaches based on Discrete Cosine Transform to capture discriminative features of handwritten Digits. These approaches are:

- DCT upper left corner (ULC) coefficients: in this approach, we retain the DCT coefficients in the upper left corner of the DCT matrix.
- DCT zigzag coefficients: the DCT coefficients are extracted in a zigzag fashion.
- Block based DCT ULC coefficients: the input image is partitioned in blocks, and the DCT upper left corner coefficients are retained for each block.
- Block based DCT zigzag coefficients: in this approach we apply DCT on each block and retain the N significant

coefficients in a zigzag fashion in order to form our feature vector.
To evaluate the performance of each approach in terms of classification accuracy and reduction rate (i.e. number of features retained), the DCT coefficients of each approach are used as input data for Support Vector Machine classifier and compared to the principal components extracted using PCA.

The objective of this work is to identify the optimal feature extraction approach that speed up the learning algorithms while maximizing the classification accuracy. The database retained for this work is the MNIST dataset [12] that we will describe in more details in the next section. A modified MNIST database has been also created using a preprocessing that eliminates the non-information-bearing areas in the original MNIST images in order to reduce dimensionality. The rest of this paper is organized as follows: in section 3 we give a brief overview of the DCT transformation and its different variants we used for dimension reduction. Next, in section 4 we present the SVM classifier retained for this work before discussing the results in section 5. Finally, the conclusion gives a brief summary and addresses some perspective for future work.

## 2. Data Description

For a fair comparison of performance of digit handwriting systems, several databases have been developed for research purposes. The handwritten digit database of CENPARMI [13] developed by Concordia University in Canada, contains 6000 digit images (600 images from each of 10 classes) collected from live mails of USPS, scanned at 166 DPI. In the database, 4000 images are specified for training and the remaining 2000 images are for testing. The Hitachi database contains 164158 samples of digit images collected from sampling sheets and real form (insurance application, bank transaction, etc.) images that were scanned at 200 DPI and is divided into several subsets of varying qualities [14]. Another large database which includes English characters (upper and lower cases) as well as number digits was provided by The National Institute of Standards and Technology (NIST) [15]. The MNIST database retained for this work was developed out of the original NIST [16]. MNIST is a reference database for handwritten digits [2] that have been widely used in handwritten digit research [16] [17] [18]. It was developed by Yann LeCun and Corinna Cortes [12]. It contains a training set of 60000 examples, and a test set of 10000 examples. The digits have been centered and size-normalized into 28x28 gray-scale images. Some sample images are shown in figure 1. The dataset is available at the homepage of LeCun [12].

Fig. 1 Sample images of MNIST dataset.

In order to reduce dimensionality, a preprocessing has been applied to the original MNIST database by eliminating the non-information-bearing areas. We will refer to this database with preprocessing as MNIST Reduced database. The preprocessing is described in more details in section 5.

## 3. Discrete Cosine Transform (DCT)

DCT initially used for image compression [19], have been of growing interest among the pattern recognition community [5] [20] [7]. DCT is a technique to convert data of the image into its elementary frequency components [21]. It clusters high value coefficients in the upper left corner and low value coefficients in the bottom right of the array(m,n) where [m n] is the image size. DCT coefficients f(u,v) of f(m,n) are computed by:

$$f(u,v) = \alpha(u)\alpha(v) \sum_{m=0}^{\mathrm{M}-1} \sum_{n=0}^{N-1} f(m,n) \cos\left(\frac{(2m+1)\pi u}{2M}\right) \cos\left(\frac{(2N+1)\pi v}{2M}\right)$$

where

$$\alpha(u) = \begin{cases} \frac{1}{\sqrt{M}}, & u = 0 \\ \sqrt{\frac{2}{M}}, & 1 \le u \le M-1 \end{cases}$$

and

$$\alpha(v) = \begin{cases} \frac{1}{\sqrt{N}}, & v = 0 \\ \sqrt{\frac{2}{N}}, & 1 \le v \le N-1 \end{cases}$$

One of the main characteristic of DCT is its ability to convert the energy of the image into a few coefficients [5]. Thus, the use of DCTs coefficients as features has become commonplace in the field of pattern recognition. In fact, DCT features has shown to be effective for several recognition problems. In his work on face recognition [22], Dabbaghchian et al. used DCT coefficients for feature extraction on the ORL and Yale database and reported the improvement of the results by using these features. Badrinath at al. [23] designed an efficient palmprint based recognition system using 1D-DCT features. The system has been tested on three databases, IITK [24], CASIA [25] and PolyU [26] databases and is found to be better than the well known palmprint systems. DCT features were also used for image watermarking [20], iris recognition [11] and speaker recognition [27].
Moreover, many works in character recognition using DCT features were reported in the literature. For Arabic handwritten isolated characters, Lawgali et al. [6] reports that DCT gives better accuracy than Discrete Wavelet transform (DWT). In the same vein, Khodadad et al. [28] used DCT features for online Arabic/Persian character recognition and asserted favorable results.

However, little research had investigated the effectiveness of the discrete cosine transform to capture discriminative features for handwritten digit classification. Furthermore, to our knowledge, no previous research has addressed the issue of what is the most effective method to retain DCT coefficients.

In this work we investigate the effectiveness of DCT features for handwritten digit recognition. Thus, we compare the performance of four variants of DCT coefficients namely:
- DCT upper left corner (ULC) coefficients;
- DCT zigzag coefficients;
- Block based DCT ULC coefficients;
- Block based DCT zigzag coefficients.

Our main goal is to identify the optimal feature extraction approach that reduces the dimension of MNIST data in order to speed up the learning algorithms while maximizing the classification accuracy.

Next, we will briefly present each method.

## 3.1 DCT ULC coefficients

This approach applies DCT on each image of the dataset and retains the N significant coefficients of the upper left corner from the transformed image as illustrated in figure 2. The retained coefficient matrix is then transformed into a vector to form the feature vector that will be fed to the classifier. The Number N of the coefficients to retain is chosen experimentally. The higher the number of retained coefficients the better the quality of the reconstructed image character.

| | | | | | |
|---|---|---|---|---|---|
| 1.1105 | -0.0008 | -0.5467 | -0.1147 | -0.4641 | 0.1671 |
| -0.0912 | -0.2609 | 0.1041 | 0.2192 | -0.0021 | 0.1706 |
| -0.7976 | -0.0444 | 0.2045 | 0.2014 | 0.6311 | -0.2251 |
| 0.0450 | 0.4270 | 0.0070 | -0.2411 | -0.1863 | -0.4503 |
| -0.2789 | 0.1106 | 0.4363 | -0.2564 | -0.3801 | 0.2052 |
| 0.1093 | -0.1092 | -0.2634 | -0.1891 | 0.3777 | 0.4663 |

Fig. 2 Example for retaining the 3x3 DCT ULC coefficients.

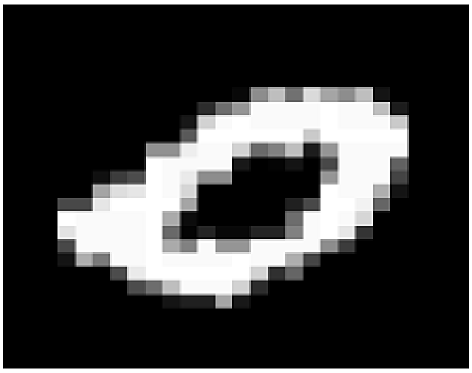
Figure 3 shows the original image character (a) with 28x28 pixels and the reconstructed image characters (b) with only 15x15 coefficients.

## 3.2 DCT zigzag coefficients

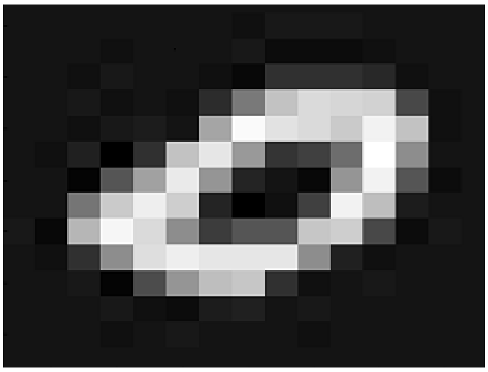
The same idea of the previous approach is applied: First, the DCT is applied to each image in the MNIST dataset, then the higher value DCT coefficients corresponding to low frequency are extracted in a zigzag order and stored in a vector sequence as illustrated in figure 4.

## 3.3 Block based DCT ULC coefficients

In the literature, for block based DCT, the image is divided into blocks of 8x8, then the DCT is applied to each block. In this work, we dispose of images normalized into 28x28 size image hence we used blocks of 7x7 then retained the most significant coefficients in the ULC of each block. The coefficients retained are concatenated to form a feature vector.



(a) Example of Initial Image size 28x28



(b) Image reconstructed with only 15x15 DCT ULC coefficients

Fig. 3 Example of image reconstruction using 15x15 DCT ULC Coefficients

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1.3389 | 0.1031 | -0.8184 | 0.0282 | -0.3595 | 0.2137 | 0.0836 | 0.1211 |
| -0.1247 | -0.2745 | 0.1307 | 0.3213 | -0.0154 | 0.0221 | -0.0581 | -0.0179 |
| -1.1597 | 0.1096 | 0.5865 | -0.0115 | 0.5230 | -0.2645 | -0.2246 | 0.1445 |
| 0.1313 | 0.5153 | -0.0810 | -0.5228 | -0.1390 | -0.4728 | 0.2067 | 0.0665 |
| -0.0678 | -0.0885 | 0.2793 | -0.2179 | -0.5709 | 0.2658 | 0.0897 | -0.1305 |
| 0.0962 | -0.2643 | -0.2542 | 0.1019 | 0.4105 | 0.3452 | -0.2991 | -0.0712 |
| 0.4661 | -0.0846 | -0.3691 | 0.2728 | 0.0486 | -0.2356 | -0.0526 | 0.1179 |
| -0.1097 | -0.0545 | 0.2691 | 0.2418 | -0.3686 | -0.2471 | 0.1926 | 0.0079 |

Fig. 4 Selecting DCT coefficients in a zigzag fashion

### 3.4 Block based DCT ULC coefficients

The same process used for the block based DCT ULC is applied here. For each 7x7 block, we apply the DCT and retain the N significant coefficients in a zigzag fashion in order to form our feature vector. This vector will be fed to an SVM classifier for classification. A brief overview of this classifier is presented in the next section.

## 4. Classification using SVM

Support Vector Machines (SVM) are a popular machine learning method for classification, regression, and other learning tasks [29]. The SVM algorithm was originally introduced by Vapnik [30] in his work on structural risk minimization and it was successfully evaluated on many pattern recognition problems [31]. SVM classifier uses kernels to give optimal decision boundary to separate between classes in higher dimensional feature spaces. For example, in two classes problem (triangles and circles sets of samples) as shown in figure 5, the basic form of linear SVM classifier tries to find an optimal hyperplane (or the maximal margin hyperplane) that separates the set of triangular samples from the set of circular samples.

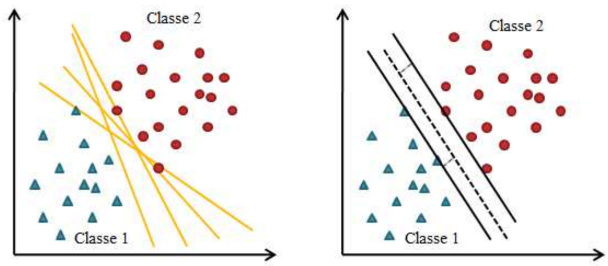



Fig. 5 Selecting optimal hyperplane for two linearly separable classes

In the case of two classes nonlinearly separable, SVM tries to separate between classes in a higher feature space using kernel functions as illustrated in figure 6.

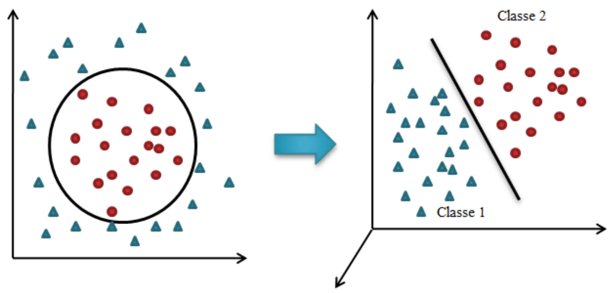


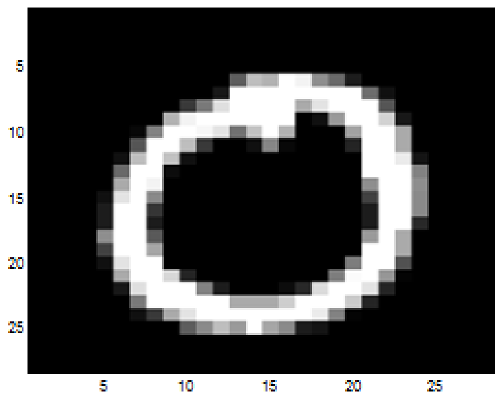

Fig. 6 Selecting optimal hyperplane for two nonlinearly separable classes

In this work, we use the LIBSVM package [29] that supports multi-class problem to classify the ten handwritten digit classes.

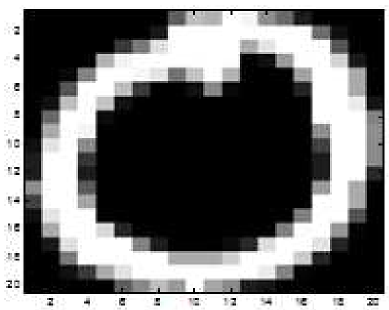

IJCSI
www.IJCSI.org

## 5. Experimental results

Experiments have been conducted in order to compare the classification performance of the four variants of DCT coefficients presented previously in section 3 and the principal components extracted using Principal Component Analysis (PCA). For this purpose, we used two datasets: the MNIST Dataset described in section 2 along with a reduced MNIST database. In the latter, a preprocessing has been applied to the original MNIST database by eliminating the non-information bearing areas. We proceeded as follows: First, for each image of size 28x28 we eliminated the margins since they don't hold any meaningful information (see figure 7). We ended up with image size ranging from 7x7 as a minimum to 20x20 as a maximum. Then, we size-normalized all the resulting images into 20x20. Hence, we reduced the dimensionality from 784 to just 400 feature pixels.



(a) Example of Initial Image size 28x28



(b) Isolated character size 20x20

Fig. 7 Example of dimension reduction of a sample image from size 28x28 to 20x20

The experiments were carried out in three steps:
- First we applied the feature extraction methods to raw data MNIST dataset as well as the MNIST reduced database in order to constitute feature vectors. For raw data MNIST dataset, we retained 15x15 features for DCT ULC, 100 for DCT zigzag, 256 for block based DCT ULC (i.e 4x4 in each block of 7x7 pixels), 160 for block based DCT zigzag (that is 10 coefficients for each block) and 80 principal components for PCA. As shown in figure 8, 80 principle components can interpret approximately 90% of total information, which suffices to be representative and informative.

Whereas, for the MNIST reduced database we retained 15x15 features for DCT ULC, 100 for DCT zigzag, 144 for block based DCT ULC (i.e. 3x3 in each block of 5x5 pixels), 128 for block based DCT zigzag (that is 8 coefficients for each block) and 80 principal components for PCA which can interpret approximately 90% of total information.

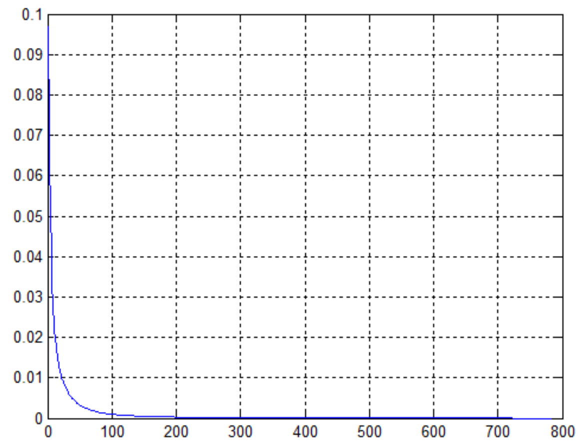


Fig. 8 Spectrum of singular values

- Second we performed a grid search with a view to find the optimal parameters c and gamma for each one of the 45 SVM classifiers corresponding to the couples of the MNIST Digit classes ranging from 0 to 9.
- Finally, we trained and tested the 45 SVM classifiers on, respectively, the 60000 training vectors and the 10000 test vectors of each dataset.

Table 1: Performances of retained feature extraction approaches in terms of dimension reduction and accuracy classification

| ***Feature Extraction Approach*** | ***Number of Features*** | Global Accuracy |
|---|---|---|
| Raw Data | 784 | 98.61 |
| PCA | 80 | 98.38 |
| DCT ULC | 225 | 98.66 |
| DCT zigzag | 100 | 98.71 |
| DCT Block ULC | 256 | 98.73 |
| DCT Block zigzag | 160 | 98.76 |

The results achieved for both databases MNIST and Reduced MNIST are summarized in Table 1 and Table 2 respectively.

Table 2: Performances of retained feature extraction approaches in terms of dimension reduction and accuracy classification

| ***Feature Extraction Approach*** | ***Number of Features*** | Global Accuracy |
|---|---|---|
| Reduced Database | 400 | 98.73 |
| PCA | 80 | 98.82 |
| DCT ULC | 225 | 98.71 |
| DCT zigzag | 100 | 98.78 |
| DCT Block ULC | 144 | 98.74 |
| DCT Block zigzag | 128 | 98.88 |

Figure 9 illustrates comparative performances of raw data features, PCA principal components and all retained DCT variants in terms of dimension reduction and global accuracy classification for both the MNIST Dataset and the Reduced MNIST Dataset.

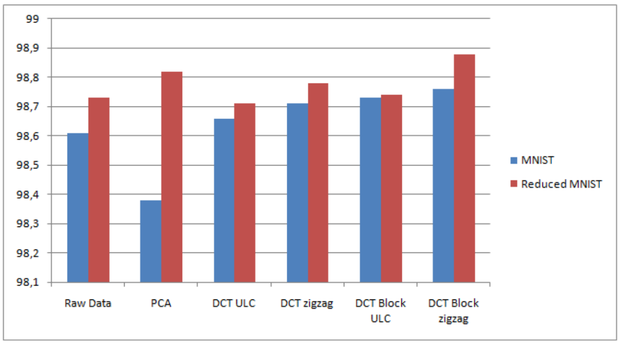



Fig. 9 Classification rate using the different Feature extraction techniques on both the MNIST Dataset and the Reduced MNIST Dataset

As we can see from figure 9, the classification rates of all the feature vectors of the reduced MNIST dataset is better than those of the MNIST Dataset. In the other hand, the classification of the four DCT variants outperforms the results of raw data classification and the principal components feature vector. Furthermore, the dimension of MNIST Dataset has been considerably reduced from 784 parameter pixels to just 225 for DCT ULC, 100 for DCT zigzag, 256 for block based DCT ULC and 160 parameters for block based DCT zigzag; leading to a better performance and improving the classification rate. Moreover, the two block based DCT features are more discriminative than the DCT applied to the whole image. This could be due to the fact that the block based DCT retains the most significant coefficients in different regions of the image which is more accurate. In sum, the DCT block based zigzag coefficients outperforms its counterparts for both datasets used in this work.

It is also worth of an investigation of the error distribution for each feature extraction method. As shown in figure 10, the histogram of errors rates of test data for six different feature sets have two peak patterns. We can conclude that classifying digit 9 and 7 may be more challenging than 0, 1 and 6.

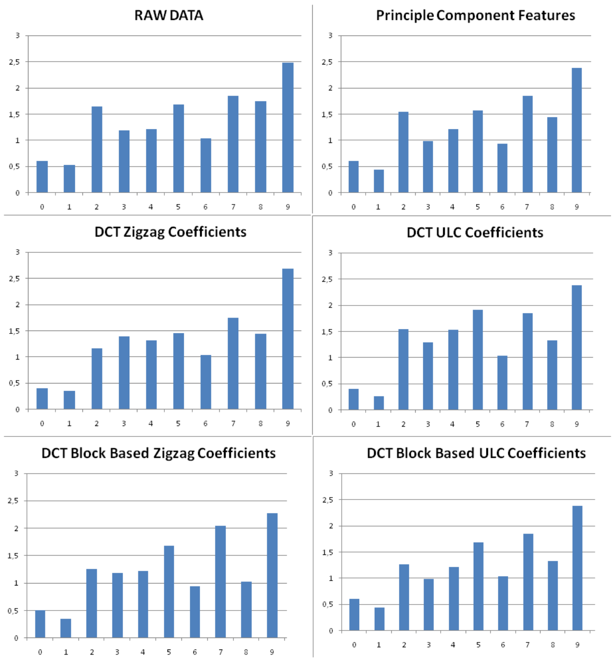



Fig. 10 Error distribution of test Data for six different feature sets using SVM Classifier

## 6. Conclusion

Feature extraction is an important step in handwritten digit recognition systems and for many other pattern recognition problems. In this work we have evaluated the effectiveness of four variants of the discrete cosine transform to capture discriminative features in order to achieve higher classification accuracy for handwritten digit recognition. These methods have been evaluated using the reference database MNIST and a modified MNIST database using a preprocessing step that eliminates the non-information baring areas. All the feature sets were fed to an SVM classifier to be evaluated in terms of accuracy and reduction rate. The results have demonstrated that the block based DCT zigzag coefficients lead to a higher accuracy compared to its counterparts. This performance is promising to integrate this method in our future work on developing an Arabic handwritten word recognition system. The performance of a recognition system doesn't only depend on feature extraction approach but also on the other steps of the recognition process such as the pre-processing stage and the classification algorithm. The combination of the latter along with feature extraction techniques is worth of an investigation in a future work.

## References

[1] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques," Pattern Recognition, vol. 36, no. 10, pp. 2271–2285, 2003.

[2] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger et al., "Comparison of learning algorithms for handwritten digit recognition," in International conference on artificial neural networks, vol. 60, 1995.

[3] C.-L. Liu, M. Koga, and H. Fujisawa, "Gabor feature extraction for character recognition: comparison with gradient feature," in Proceedings of the Eighth International Conference on Document Analysis and Recognition. IEEE, 2005, pp. 121–125.

[4] X.-T. Yuan and B.-G. Hu, "Robust feature extraction via information theoretic learning," in Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009, pp. 1193–1200.

[5] J. AlKhateeb, J. Ren, J. Jiang, S. S. Ipson, and H. El Abed, "Wordbased handwritten arabic scripts recognition using dct features and neural network classifier," in Proceedings of the 5th International MultiConference on Systems, Signals and Devices. IEEE, 2008, pp. 1–5.

[6] A. Lawgali, A. Bouridane, M. Angelova, and Z. Ghassemlooy, "Handwritten arabic character recognition: Which feature extraction method," International Journal of Advanced Science and Technology, vol. 34, pp. 1–8, 2011.

[7] V. V. Kohir and U. Desai, "Face recognition using a dct-hmm approach," in Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision (WACV'98). IEEE, 1998, pp. 226–231.

[8] C.-W. Ngo and C.-K. Chan, "Video text detection and segmentation for optical character recognition," Multimedia Systems, vol. 10, no. 3, pp. 261–272, 2005.

[9] M. Klima, P. Pata, K. Fliegel, and P. Hanzlik, "Image quality evaluation in security imaging systems," Aerospace and Electronic Systems Magazine, IEEE, vol. 22, no. 1, pp. 22–25, 2007.

[10] R. Parisi, E. Di Claudio, G. Lucarelli, and G. Orlandi, "Car plate recognition by neural networks and image processing," in Proceedings of the IEEE International Symposium on Circuits and Systems , vol. 3. IEEE, 1998, pp. 195–198.

[11] D. M. Monro, S. Rakshit, and D. Zhang, "Dct-based iris recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 586–595, 2007.

[12] Y. LeCun, "Mnist ocr data," http://www.research.att.com/yann/exdb/mnist/index.htm/.

[13] C. Y. Suen, C. Nadal, R. Legault, T. A. Mai, and L. Lam, "Computer recognition of unconstrained handwritten numerals," Proceedings of the IEEE, vol. 80, no. 7, pp. 1162–1180, 1992.

[14] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: investigation of normalization and feature extraction techniques," Pattern Recognition, vol. 37, no. 2, pp. 265–279, 2004.

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

[16] L. Deng, "The mnist database of handwritten digit images for machine learning research," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 141–142, 2012.

[17] B. El Kessab, C. Daoui, B. Bouikhalene, M. Fakir, and K. Moro, "Extraction method of handwritten digit recognition tested on the mnist database." International Journal of Advanced Science & Technology, vol. 50, 2013.

[18] Y. Qian, W. Xichang, Z. Huaying, S. Zhen, and L. Jiang, "Recognition method for handwritten digits based on improved chain code histogram feature," 3rd International Conference on Multimedia Technology. Atlantis Press, 2013.

[19] W. B. Pennebaker and J. L. Mitchell, JPEG: Still image data compression standard. Springer, 1993.

[20] M.-J. Tsai and H.-Y. Hung, "Dct and dwt-based image watermarking by using subsampling," in Proceedings of the 24th International Conference on Distributed Computing Systems Workshops. IEEE, 2004, pp. 184–189.

[21] A. Al-Haj, "Combined dwt-dct digital image watermarking." Journal of computer science, vol. 3, no. 9, 2007.

[22] S. Dabbaghchian, M. P. Ghaemmaghami, and A. Aghagolzadeh, "Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology," Pattern Recognition, vol. 43, no. 4, pp. 1431–1440, 2010.

[23] G. Badrinath, K. Tiwari, and P. Gupta, "An efficient palmprint based recognition system using 1d-dct features," in Intelligent Computing Technology. Springer, 2012, pp. 594–601.

[24] J. Kumar, A. Nigam, S. Prakash, and P. Gupta, "An efficient pose invariant face recognition system," in Proceedings of the International Conference on Soft Computing for Problem Solving, 2011. Springer, 2012, pp. 145–152.

[25] "Casia iris image database", http://www.sinobiometrics.com.

[26] "Polyu palmprint database", http://www.comp.polyu.edu.hk/biometrics/.

[27] M. McLaren, N. Scheffer, L. Ferrer, and Y. Lei, "Effective use of dcts for contextualizing features for speaker recognition," ICASSP, 2014.

[28] I. Khodadad, M. Sid-Ahmed, and E. Abdel-Raheem, "Online arabic/persian character recognition using neural network classifier and dct features," in Proceedings of the 54th IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), 2011, pp. 1–4.

[29] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no. 3, p. 27, 2011.

[30] V. N. Vapnik, "The nature of statistical learning theory. statistics for engineering and information science," Springer-Verlag, New York, 2000.

[31] V. Vapnik, Statistical learning theory. Wiley New York, 1998, vol. 2.