A robust system for printed and handwritten character recognition of images obtained by camera phone

H. El Bahi, Z. Mahani, A. Zatni and S. Saoud

Université Ibn Zohr, ESTA, Laboratoire Matériaux, Systèmes et Technologies de l'information
B.P: 33/S, 8000 Agadir - Maroc
elbahihassan@gmail.com; z.mahani@uiz.ac.ma,
http://www.esta.ac.ma/

Abstract: - In Recent years, character recognition has gained more importance in the area of pattern recognition owning to its application in various domains. The biggest challenge is to build an efficient optical character recognition system (OCR) able to recognize documents, also to allow overcoming the problems of blurred and noisy image. Many OCRs systems are been applied, but less interest have been given to document images obtained by camera phone. In this paper, we will present a complete offline handwritten and machine-printed character recognition system for isolated character acquired via camera-mobile. Our system includes five stages namely: preprocessing, segmentation, feature extraction and classification. We investigated various techniques in the preprocessing stage in order to select the best. In feature extraction and classification stages, we examined several features methods with three different types of classifiers The Support vector machines (SVM), The Naïve Bayes (NB) and the Multilayer Perceptron (MLP). We performed the experiments with two databases of handwritten and machine-printed character images. The results indicate that the proposed system is very effective and yields good recognition rate for character images obtained by camera phone.

Key-Words: - Preprocessing, Feature Extraction, Classification, OCR, SVM and MLP.

1 Introduction

The field of pattern recognition has become one of the broad areas where more and more researchers have worked. The goal of researchers in this field is to find algorithms that can solve on a computer the problems of pattern recognition, which are intuitively resolved by humans. Optical character recognition (OCR) is one of the fields in pattern recognition; its purpose is to transform an image of handwritten, typewritten or printed text into an understandable representation that a computer can easily recognize. Consequently, the OCR system is applied in several applications in various domains such as: bank check processing, postal code recognition, mail sorting, digital libraries, security system, etc. However, OCR system development is a non trivial task because the word have an infinite number of representations due to that each person writes with his own way, which is different from the others, and also in view of the fact that there are many font to print with many styles (bold, italic, underlined, etc), with different complex layouts. Depending on the type of writing that a system should recognize (manuscript, cursive or print), operations to be performed and the results can vary significantly. The OCR system is widely divided into two types: On-line recognition and Off-line recognition. In the On-line type, the process of writing recognition is performed at the same time when the user is writing, while the off-line type is static recognition in which the writing recognition is carried out after completion of writing.

Several OCR systems have been proposed by researchers, but fewer attentions have been given to document image recognition acquired via camera phone. Therefore In this paper, our objective is mainly interested in the development of an off-line English handwriting and machine-printed characters recognition system, in which the images are obtained by camera phone.

Habitually, the phases form the structures of handwriting recognition system are: Pre-processing, Segmentation, Feature extraction, Classification and Post-processing. Indeed, the use of a very good method in Pre-processing stage will allow facilitating the other phases. However, Extraction of good features and employing of a perfect classification method are the main keys for improving recognition accuracy rate.

The remainder of the paper is organized as follows: In the section 2 we present a review of some work of text enhancement methods and thresholding techniques. Section 3 gives an overview of our proposed OCR system and also

gives descriptions of the methods that we used throughout the OCR process, which includes the following stages: Binarization, Noise removing, Segmentation, Skeletonization, Normalization, Feature extraction and Classification. Finally, some experimental results and a comparison between different classifiers with using different combinations of features extraction methods which illustrate more this work are shown in Section 4.

2 PREPROCESSING

After document image acquisition, preprocessing is the first step in most image processing systems and pattern recognition, in this phase one of the basic operation preformed is: Text enhancement. The text enhancement has become an indispensable aspect of the document analysis process; its goal is restoration of the distortions appeared during the acquisition and making the image more meaningful through the preservation of textual information [1]. In recent years, this method has gained more importance and become an area of constant research. It is a very important phase in the document analysis which can improve and facilitate the tasks that come next, such as: segmentation, feature extraction and the classification.

Traditional text enhancement algorithms aims at separating document image into two layers namely, the foreground text (or regions of interest) and the document background. In the literature, there are various techniques of text enhancement. Generally these techniques can be classified into the following several groups: the first one is based on global thresholding [11.28], the second is based on local (adaptive) thresholding [14.15.29.13]. The global thresholding methods try to find a single threshold value calculated from an overall measure for the whole image, and then each pixel is classified into text (foreground) or background based on its grayscale value. The advantage of global methods is very fast. Moreover, they work well for typical scanned document image. One of the drawbacks of these methods is that they can't give a good result in the case of presence of noise or low quality document image. To overcome these problems, local (adaptive) thresholding methods have been proposed. These methods calculate a single threshold value for each pixel, and this value is determined by grayscale information from local neighborhood of the pixel. The main disadvantage of these techniques is the fact that its effectiveness depends completely on the window's size and the character stroke width.

Otsu [11] is the most popular method among the global methods, the algorithm selects the global threshold value that minimizes the variance among two classes of pixels; the text and background pixels. When the gray level of the pixel is below the global threshold so it is defined as a background, otherwise it is defined as a text region.

Kitter and Illingworth [13] proposed one of the most important minimum error techniques; the main purpose is to minimize the average pixel classification error directly, employing either exhaustive search or an iterative algorithm [13].

One of the good performing local thresholding techniques was suggested by Niblack [14]. The basic principle of this method is for each pixel (x, y) a threshold T value is computed based on the mean m and standard deviation s in a local neighborhood (x, y) by the following form:

$$T(x, y) = m(x, y) + k.s(x, y) \tag{1}$$

Where k is a negative constant equal to -0.2. To conserve local information in an image the size of the neighborhood of the pixel windows should be small, which consequently gives rise to a lot of noise in the background image regions.

Sauvola's technique [15] finds a solution to this problem by adding a hypothesis on the gray value of text (foreground) and non text (background) pixels; text pixels have gray value close to 0 and non text pixels have gray value near 255. Then the local thresholding value will be calculated by the following equation:

$$T(x, y) = m(x, y).\left(1 - k.\left(1 - \frac{s(x, y)}{R}\right)\right) \tag{2}$$

Where k is a constant equal to 0.5, and R denotes the dynamic range of the standard deviation s (defined as $R = 128$ for a grayscale documents).

The thresholding are usually very limited in the case of restoration of degraded documents image. As a solution recently, another category of document image enhancement approach has been proposed in the literature based on partial differential equation (PDE). This approach is usually able to remove noise and restoration of degraded images, without losing the essential information for readability (see [2.3.4.5.16]).

Many methods based on partial differential equation (PDE) have appeared, particularly those using nonlinear diffusion. The most known of these methods is Perona-Malik [9] which is aiming at smoothing an image and reducing the noise while

simultaneously preserving and enhancing the image features such as the edges.

The Perona and Malik equation is defined as:

$$\begin{cases} \dfrac{\partial u}{\partial t} = div\big(g(|\nabla u|).\nabla u \big) \\ u(x,y,0) = u_0(x,y) \end{cases} \qquad (3)$$

Where $|\nabla u|$ is the gradient modulus of the image and $g(|\nabla u|)$ is an edge stopping function which is chosen to satisfy $g(0)=1$ and $g(x) \to 0$ where $x \to \infty$.

Recently F. Darira [2] suggested an approach for enhancing degraded textual document image and reducing the noise. This approach based on combination of the diffusion model of Perona-Malik and the tensor-driven diffusion of Weickert. R. Farrahi Moghaddam and M. Cheriet [12] presented an enhancement technique of low quality document based on a physical model of documents degradation.

Another method is proposed by [10] for preservation of meaningful textual information in document image, this method allows overcoming the problems usually encountered in image obtained by camera phone, it's based on solving a partial differential equation (PDE). Initially, they used a model that combines reflectance and the non uniform illumination of the image , and then they estimated the non uniform illumination by relying on the solution of a partial differential equation method (5), after that they calculated the reflectance depending on the image and the non uniform illumination already estimated (for more details see [10,17]).

$$\begin{cases} w_t = d \max(0, d\Delta^A w), \\ \dfrac{\partial w}{\partial t} = 0 \\ w(t=0) = u, \end{cases} \qquad (4)$$

Where w the log of the non uniform illumination, and d the grayscale of the text equals 1 or -1 according to the image background.

In figure 1, we have examined the results of several methods of thresholding and text enhancement in order to select the best method for our OCR system, we have tested: Otsu [11], Kitter and Illingworth [13], Niblack [14], Sauvola [15] and Perona-Malik [9]. The testing image example is obtained from a camera phone (SAMSUNG Galaxy S III).

Otsu [11], Niblack [14] and Kitter and Illingworth [13], can preserve the text information, but they generate a certain amount of noise in the non text regions (background) due to the low contrast and the variation of the illumination in the document image. The diffusion model of Perona-Malik [9] removes the noises in non-text regions, but it is unable to removing the image background. The Sauvola method [15] smooth efficiently the noise, it also is excellent in terms of readability and the preservation of the meaningful textual information.
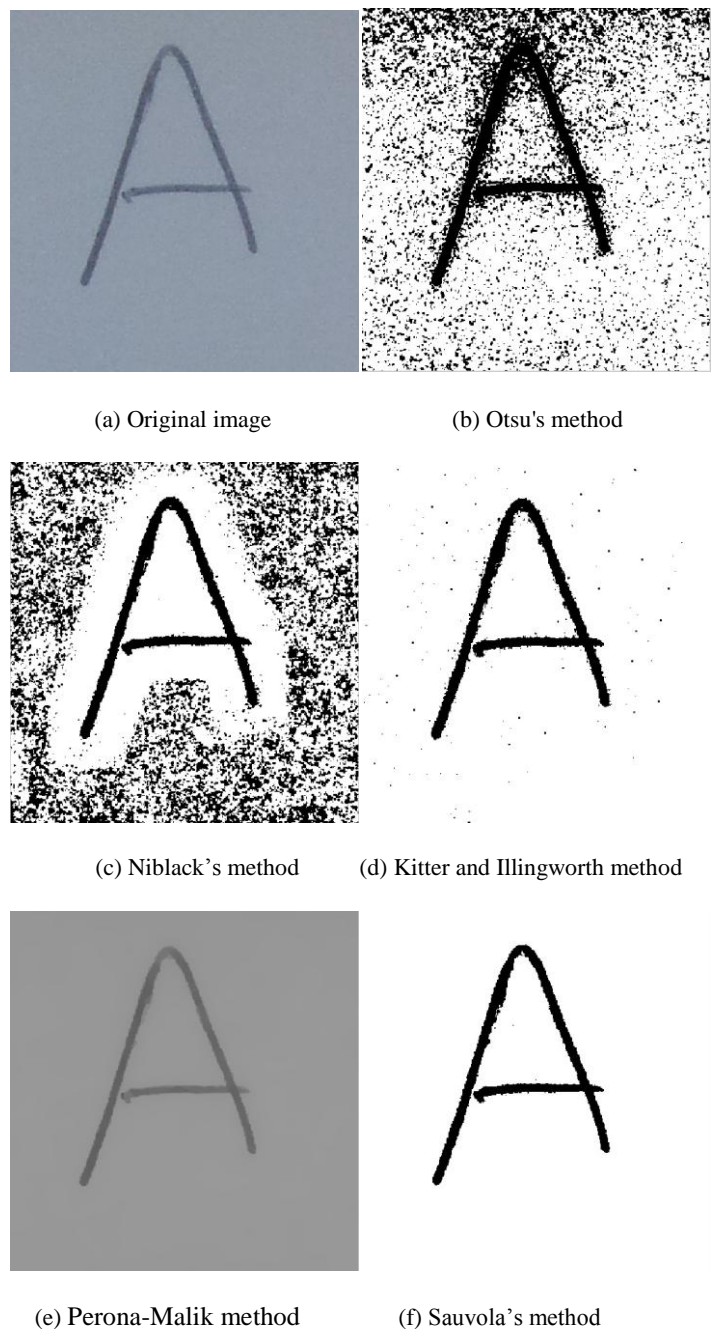

(a) Original image   (b) Otsu's method

(c) Niblack's method   (d) Kitter and Illingworth method

(e) Perona-Malik method   (f) Sauvola's method

Fig. 1 Original image obtained from an SAMSUNG Galaxy S III

3 The PROPOSED RECOGNITION SYSTEM

In this section a system of off-line handwritten optical character recognition (OCR) is proposed. Generally, an OCR system is a mechanism that includes several stages for translating an image of printed or handwritten text into a form that the machine can manipulate. These stages are called: preprocessing, segmentation, feature extraction and classification. The schematic diagram of these phases is shown in figure 2.
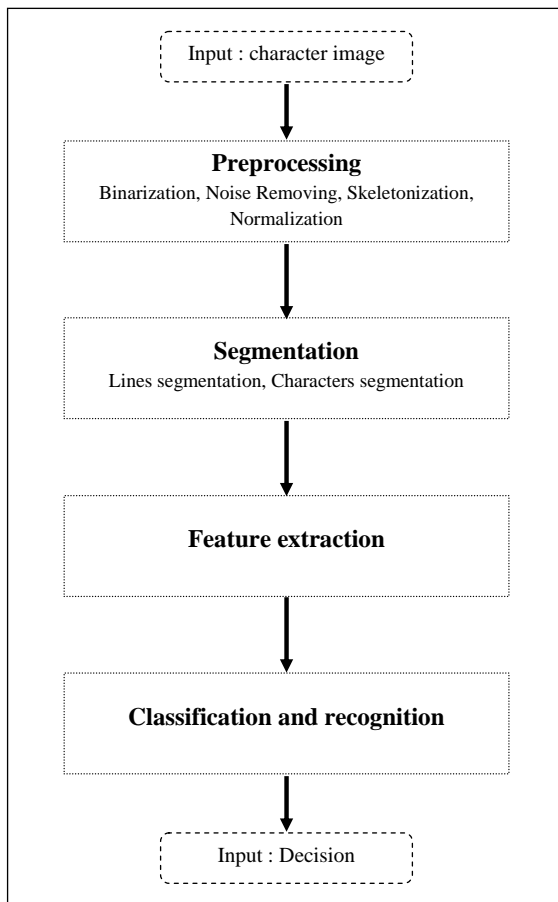

Fig. 2 Schematic diagram of the proposed OCR system

3.1 *Image Acquisition*

The image acquisition is the first step in all image processing systems and pattern recognition. In this work the images of characters are obtained form a camera phone (Samsung S3). A sample of characters image is shown in figure 3.

3.2 *RGB to Gray Image:*

The purpose of this step is to convert an RGB image (combination of Red, Green and Blue colors) to a grayscale format. Therefore we calculated the gray level for each pixel by using the following formula:

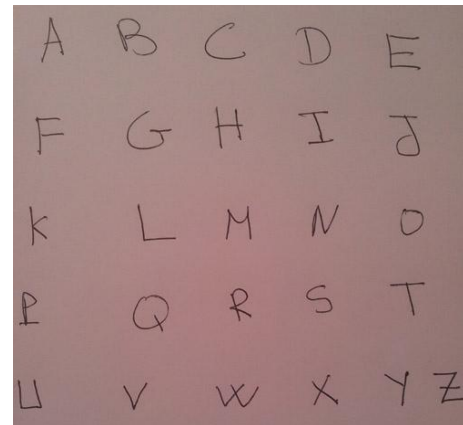$$Y = 0.298*R + 0.587*G + 0.114*B \quad (5)$$

Fig. 3 Sample of characters image

3.3 *Binarization:*

According to the result of comparison between the different algorithms of binarization that we saw in section 2, we have chosen to use Sauvola [15] method. It achieves better results than the other methods; it also leads to the preservation of the meaningful textual information and removes the noise in non-text regions (background). Figure 4 below shows the results obtained with the binarization Sauvola method.
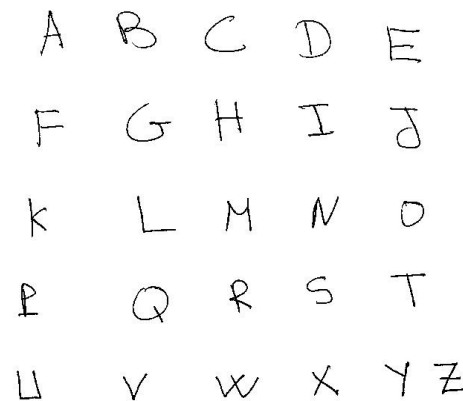

Fig. 4 Binarization with the Sauvola method

3.4 *Noise removing:*

Noise which is in the images is one of the big difficulties in optical character recognition process. The aim of this part is to remove and eliminate this obstacle; there are several methods that allow us to overcome this problem. In this work we decided to use the morphology operations to detect and delete small areas of less than 40 pixels.

3.5 *Segmentation:*

Segmentation is one of the principal stages of OCR process. With the use of a very good segmentation method, the recognition accuracy will be increased. In this phase, an input image that contains a sequence of characters will be subdivided into sub images of isolated characters. In our proposed system the segmentation is carried out with line segmentation and character segmentation. In line segmentation, we determine the horizontal projection histogram of pixels for each row (figure 5) in order to distinguish between regions that have high density (lines) and regions that have low density (interspace among the lines). Similarly for character segmentation, once we get the lines we use the vertical projection histogram of pixels for each column (figure 6) to obtain the individual characters.
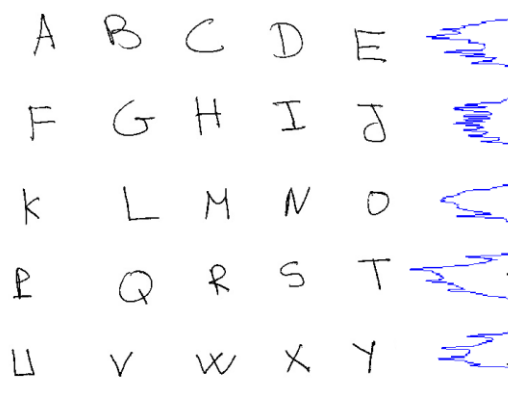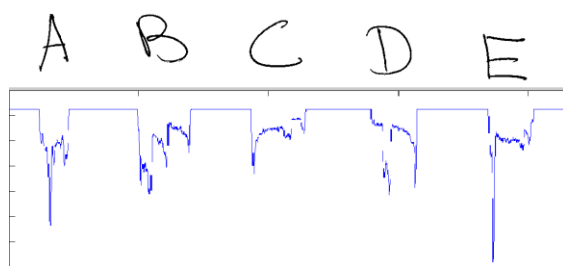

Fig. 5 Lines segmentation

Fig. 6 Characters segmentation

3.6 *Skeletonization:*

The skeletonization or thinning is an important preprocessing operation performed to simplify the representation of an image and convert it into another image easier to treat. The basic idea of skeletonization is to reduce the thickness character image to one-pixel while preserving its connectivity and its topological properties. A number of thinning algorithms have been proposed and applied to OCR system. In this work we selected the algorithm of Zang and Suen [18] owing to its strength and its

rapidity. The result obtained before and after applying the thinning algorithm is given in the figure 7.
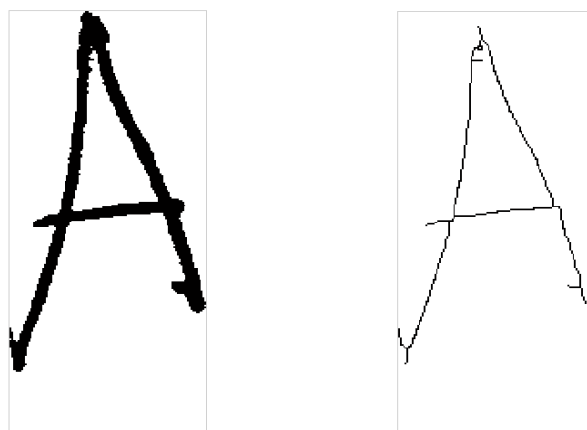
(a) Before thinning   (b) After thinning

Fig. 7 The result of the Zang and Suen algorithm

3.7 *Normalization:*

Normalization is an important task allowing converting an image with arbitrary size to an image with a fixed size. It tends to reduce or eliminate as much as possible the variability related to the difference in sizes and styles. In order to facilitate the feature extraction phase, likewise for improving the classification accuracy rate. In this paper, every character image has 60 x 50 pixels.

3.8 *Feature extraction:*

Feature extraction is a necessary and important stage for any OCR system. Its role is to represent the input data (character image) in a vector of fixed dimension. This vector contains the characteristic which are most relevant of image. However, a right feature method is the main key to improving recognition accuracy rate. In literature, there are several features extraction methods which are categorized into three groups: Structural features, Statistical features and global transformation features. In this paper, we have tested seven methods: Histograms of oriented Gradients (HOG), Grey Level Co-occurrence Matrix (GLCM), Zernike Moments, Gabor Filters, Zoning, Projection Histogram, and Distance Profile. In the following we will give the principle of each method, thereafter to constitute the features vectors we have divided all the methods into 11 various feature vectors include single features and combinations features, the Table 1 shows these features vectors as well the sizes of each one of them.

3.8.1 *Histograms of oriented Gradients:*

Histograms of oriented Gradients (HOG) features were introduced by Dalal and Triggs [26] based on Scale-Invariant Features transform (SIFT) descriptors [27]. HOG features is widely used in several applications in the fields of computer vision and image processing with intention of doing the localization and detection of objects. In this work we attempt to extract the features from characters images by using HOG technique. The main idea of this technique is that local object appearance and shape can frequently be expressed enough well by distribution of local intensity gradients or edges directions. The computation of HOG features can be obtained by the following steps: Gradient computation, histogram generation and histogram normalization.

In the first step the gradient magnitude was calculated by filtering the image with two masks Gx and Gy in horizontal and vertical directions respectively. In this work we have used the following masks:

$$Gx = (-1, 0, 1) \text{ and } Gy = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \quad (6)$$

After that, for each pixel the magnitude of the gradient $m(x, y)$ has been obtained by combining between horizontal and vertical approximations by the following form:

$$m(x, y) = \sqrt{Dx^2 + Dy^2} \quad (7)$$

With Dx and Dy are the approximations of horizontal and vertical gradient obtained by convolving the masks with the original image.

The orientation $\theta(x, y)$ of the gradient in each pixel was also calculated (Fig. 8-b) as follows:

$$\theta(x, y) = \tan^{-1}\left(\frac{Gx}{Gy}\right) \quad (8)$$

The second stage starts with division of the image into small regions called cells (Fig. 8-a). In this paper all gradient histogram are computed along rectangular cells (RHOG) by the using of unsigned gradient. For each cell we have used 9 bins $i = 1, 2, ..., 3$ and all the pixels participate in the vote. Then we accumulate the vote into the appropriate bin for each orientation. The vote has been done by

the magnitude of the gradient. After that we can construct the feature vector of an image by regrouping all histograms for each cell together in order to form one vector (Fig. 8-c).

The last step is the normalization, in this step in order to avoid the variability in the image and improve the accuracy, the histogram values are normalized with a L1-norm.
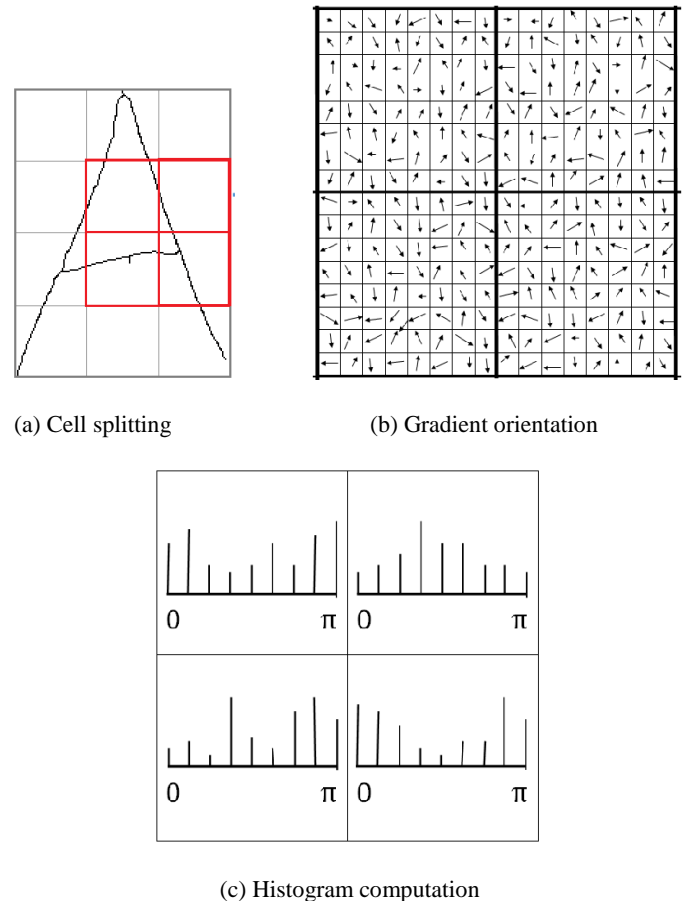

(a) Cell splitting   (b) Gradient orientation

(c) Histogram computation

Fig. 8 Calculate Histograms of Oriented Gradients

3.8.2 *Gray Level Co-occurrence Matrix:*

Gray Level Co-occurrence Matrix (GLCM) technique is an approach for extracting statistical texture features that have been proposed by Haralick [19]. the main principle of GLCM is to counts the number of times various combinations of pixel gray levels occur in a given image. Haralick defines 14 statistical features measured from the GLCM. In this work, five important features are used namely energy, contrast, correlation, entropy and homogeneity.

3.8.3 *Zernike moments:*

Zernike moments were introduced by Teague based on the orthogonal Zernike polynomials. They have

been used widely in different patterns recognitions applications, due the fact they are invariant to rotation, robustness to noise and can be readily constructed to an arbitrary order.

The Zernike moments [21] of order n and repetition m are defined as follows of an image $I(x, y)$:

$$Z_{mn} = \frac{m+1}{\pi} \int_x \int_y I(x, y) \left[V_{mn}(x, y) \right] dx dy \qquad (9)$$

Where $V_{mn}(x, y)$ is represented in polar coordinates as follows:

$$V_{mn}(r, \theta) = R_{mn}(r) e^{-jn\theta} \qquad (10)$$

Where $R_{mn}(r)$ is the orthogonal radial polynomial given as!

$$R_{mn}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} (-1)^s \frac{(m-s)!}{s! \left(\frac{(m+|n|)}{2} - s \right)! \left(\frac{(m-|n|)}{2} - s \right)!} r^{m-2s}$$

3.8.4 *Gabor filters:*
As a powerful feature, the Gabor filters [22] have been successfully applied in numerous pattern recognitions including face recognition fingerprint recognition …, as well as optical characters recognition. The Gabor filters are defined by a complex sinusoidal modulated by a Gaussian envelope described as follows:

$$G(x, y, \theta, f) = e^{-\frac{1}{2} \left(\frac{R_1^2}{\sigma_x^2} + \frac{R_2^2}{\sigma_y^2} \right)} \cos(2\pi f x_\theta) \qquad (11)$$

Where:

$$R_1 = x \cos(\theta) + y \sin(\theta)$$
$$R_2 = y \cos(\theta) - x \sin(\theta)$$

f represents the frequency of the Sinusoidal plane wave along the direction θ, and (σ_x, σ_y) explain the standard deviations of the Gaussian envelope along x and y directions.

3.8.5 *Zoning:*
The zoning technique [20] is a statistical region-based feature extraction, it aim is to get the local characteristics in lieu of global characteristic. Therefore, according to the size normalized character image (60 x 50 pixels), we divided it into

30 (6 x 5) zones of 10 x 10 pixels size, then we calculated the densities of pixels in each zone, finally we are getting 30 features.

3.8.6 *Projection histogram:*
Projection histogram descriptor is a statistical feature, According to this feature we have we have used two direction of projection horizontal and vertical traversing. The horizontal histogram of the character is generated by counting the number of black pixels in each row, similarly the vertical histogram of character computed by counting the number of black pixels in each column. At the last we will have 60, 50 feature depending on the direction projection.

3.8.7 *Distance profile:*
In distance profile feature [23] the distance (number of pixels) between the bounding box of image and the first pixel of foreground will be calculated. We have employed four type of profiles sides left, right, top and bottom. Concerning left and right profiles, they are extracted by counting the distance from the left bounding box and the right bounding box respectively to the nearest foreground pixels in each row. Then as well, top and bottom profiles, they are extracted by counting the distance from the top bounding box and the bottom bounding box respectively to the nearest foreground pixels in each column.

Feature Method	Contained feature	Size
FM1	Zernike Moments	32
FM2	GLCM	5
FM3	Gabor filters	32
FM4	Zoning	30
FM5	Projection Histogram Horizontal	60
FM6	Projection Histogram Vertical	50
FM7	Horizontal + Vertical Histogram	110
FM8	Distance Profile (Left + Top)	110
FM9	Distance Profile (Right + Bottom)	110
FM10	Distance Profile (R + T + R + B)	220
FM11	Histograms of oriented Gradients	81

Table 1 Combination of the different feature vectors

3.9   *Classification:*

The classification stage involves finding the most suitable model to the input character image. In general, the classification is composed of two steps: Learning and decision. During the learning step the system learn about the relevant properties of the models classes by using a training set of samples. Afterward, decision step, we seek to predict the model closer that belongs to him the character image. In litterateur, there are many types of classifiers that have been implemented in off-line handwritten optical character recognition problems. Among them, in this paper we have used three classifiers: The Support vector machines (SVM), The Naïve Bayes (NB) and the multi-layer Perceptron (MLP) artificial neural network.

3.9.1   *Support vector machines:*

The Support vector machines (SVM) are a category that belongs to the supervised machine learning models that can be used to classifications or regressions problems, they proposed by Vapnik [24]. SVM modeling was originally used to optimize the linear hyperplane which separate two classes, That is to say, the empty region around the decision boundary determined by the distance to the nearest training pattern [24].

We consider the problem of classification the group of training data $(x_i, y_i)_{i=1,...,l}$ into two classes, where $y_i \in \{-1, +1\}$ is the class (label) and x_i is a feature vector. In the classification with SVM model a label y_i will be assigned to a feature vector x_i by evaluating:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b \right) \quad (12)$$

Where α_i corresponding to the weighs and b was the bias, these two variables are called SVM parameters and adopted into training by maximizing:

$$L_D = \sum_i x_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j S_i S_j K(P_i, P_j) \quad (13)$$

Within the constraints:

$$\sum_i x_i S_i = 0 \quad \text{And} \quad 0 \le \alpha_i \le c \quad (14)$$

Where C is a positive constant, and $K(P_i, P_j)$ is named the Kernel function of the SVM model. In this paper, we use the polynomial Kernel; it's given by the following form:

$$K(P_i, P_j) = \left(P_i * P_j + 1 \right)^d \quad (15)$$

In the case where the classes are not linearly separable, it necessary to combine several SVMs models to solve a multi-class classification problem. The strategy is the "one against all" which involves building a SVM per class, so each classifier trained to distinguish the data of his class from those of all other classes. Another classic strategy is to use "one against one" which building a SVM classifier for each pair of classes. In this work, we have selected "one against one" strategy for multi-class classification.

3.9.2   *Naive Bayes classifier:*

The Naïve Bayes classifier is one of the simplest methods in supervised machine learning models based on applying Bayes' theorem. It assumes that values of input features $X_1, X_2, ..., X_n$ are all conditionally independent of any other given Y. Taking an example, $X = \langle X_1, X_2 \rangle$ are two features, therefore we get:

$$P(X \mid Y) = P(X_1, X_2 \mid Y) \quad (16)$$
$$= P(X_1 \mid X_2, Y) P(X_2 \mid Y)$$
$$= P(X_1 \mid Y) P(X_2 \mid Y)$$

Generally speaking, when a set of features $X = \langle X_1, X_2, ..., X_d \rangle$ of dimension d are conditionally independent of any other given Y, as a result we have

$$P(X_1...X_d \mid Y) = \prod_{i=1}^{d} P(X_i \mid Y) \quad (17)$$
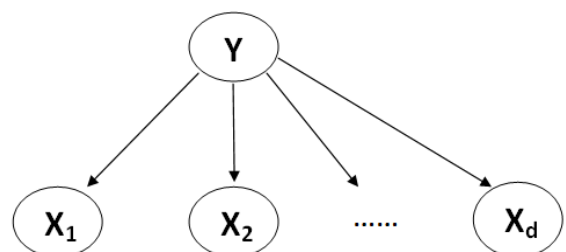

Fig. 9 Bayesian network for the Naive Bayes classifier.

3.9.3 *Artificial neural networks:*

Artificial Neural network approach has been extensively used for pattern recognition problems; it's a set of connected neurons running in parallel that allow learning and recognition. There are many types of neural networks; in our work we have selected the Multilayer Perceptron (MLP) architecture especially since it provides a simple implementation with satisfactory capacity for character recognition.

MLP is a feed forward artificial neural network organized in layers, including at least three layers of neurons, in particular the first is the input layer, the last is the output layer, and at least a hidden layer (figure 10), in a way that each neuron is connected to all the previous and next layers. Indeed, neurons of the first layer are connected to the external word and receive the input data, the number of these neurons varies depending on the size of the feature vector. The sum of the neurons in the output layer is determined by the total number of classes' recognition. Therefore, the output includes 26 neurons, which correspond to number of character. Hidden layers are used in order to solve some problems such as non-linearly separable problem, generally, the size of this layer is between the number of neurons in the input and output layers [25,26].
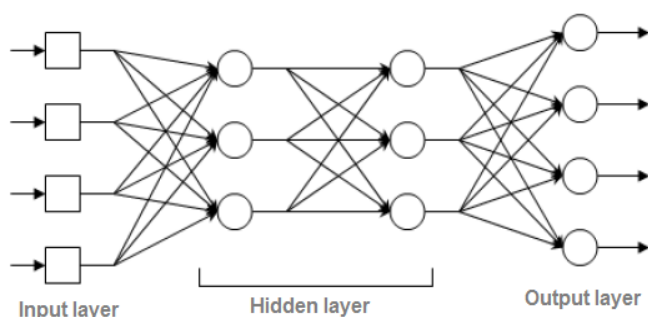

Fig. 10: Multilayer Perceptron model.

In this paper, the back propagation learning algorithm is used for the recognition and classification stages, the purpose of this algorithm is to modify the weight and bias for each neurons in order to be able to give results very to the correct values of input. For the parameters of this algorithm, we have chosen 1000 epoch to training the network with 0.3 for learning rate and 0.2 for momentum rate. As for activation function we worked with the sigmoid function defined by the following:

$$f(x) = \frac{1}{1 + e^{-x}} \qquad (18)$$

4 Numerical results

Due to the absence of a standard database of handwritten or machine-printed characters acquired by camera phone, we have constructed two databases of upper-case English character (A to Z) images obtained via camera phone. The handwritten database contains 130 samples of 26 characters classes, collected from 10 different writers presented in figure (13). As a result the database consists of 3380 samples. The samples of the machine-printed database are generated using 15 different fonts as shown in figure (14), for each fonts we have used three size (viz. 20 points, 26 points and 36 points), as well different styles are used namely : Normal, Bold and Italic. Therefore this database contains 130 samples of 26 characters classes, accordingly a total of 3380 samples. For both databases the samples are divided randomly into two set, one for training stage (2990 samples) and the other for testing stage (390 samples).

We have employed the camera phone of SAMSUNG Galaxy S III, the camera oh this phone has 8 megapixels. By default, image returned from this mobile has a high resolution of 3264 x 2448 pixels. All the experiments are carried out in MATLAB 7.9 environment with using a PC with windows 7 as an operating system equipped with Intel (R) Core (TM) i7-3337U processor 1.80 GHz and 4 Go RAM.

The results obtained from our training and Testing of both databases are presented in Table 2 and 3. For classification stage we have used three classifiers: the Naïve Bayes (NB), the Support Vector Machine (SVM), the Multilayer Perceptron (MLP), and for each classifier we employed a set of different features extraction methods that are indicated in Table 1.

In table 2, the experiments were carried out on handwritten characters database. We found that the Histograms of oriented Gradients (FM 11) provides higher recognition and learning rate, with the achievement of a rate of 87.43%, 97.85% and 97.94% as recognition accuracy, respectively for NB, SVM and MLP. Zoning (FM 4) achieves a very good recognition rate with 86.86%, 95.20% and 97.43% respectively for NB, SVM and MLP.

The result presented in table 3, show the learning and recognition rates of the machine-printed characters database obtained using three different types of classifiers with various features methods. It is observed the same as the results of handwritten database, the HOG (FM 11) and Zoning (FM 4)

Classifier Feature vector	Naive Bayes (NB)		Support Vector Machine (SVM)		Multilayer Perceptron (MLP)	
	Learning R.	Recognition R.	Learning R.	Recognition R.	Learning R.	Recognition R.
FM 1	50.10 %	45.38 %	61.87 %	53.58 %	70.20 %	48.71 %
FM 2	57.25 %	54.87 %	58.66 %	55.64 %	66.38 %	63.07 %
FM 3	59.48 %	58.32 %	78.71%	76.92 %	90.80 %	81.28 %
FM 4	87.69 %	86.86 %	95.38 %	95.20 %	97.72 %	97.43 %
FM 5	74.05 %	73.58 %	71.27 %	69.48 %	76.12 %	73.07 %
FM 6	57.01 %	52.05 %	52.78 %	52.30 %	62.56 %	61.00 %
FM 7	87.43 %	87.18 %	88.01 %	85.89 %	93.58 %	93.55 %
FM 8	66.75 %	62.82 %	86.18 %	80.25 %	94.94 %	82.30 %
FM 9	80.93 %	77.94 %	94.71 %	87.43 %	98.02 %	88.71 %
FM 10	89.29 %	86.41 %	99.09 %	94.35 %	99.29 %	95.12 %
FM 11	89,76 %	87,43 %	98,20 %	97,85 %	99,16 %	97,94 %

Table 2: Handwritten characters recognition results using different features vectors with NB, SVM and MLP classifiers.

Classifier Feature vector	Naive Bayes (NB)		Support Vector Machine (SVM)		Multilayer Perceptron (MLP)	
	Learning R.	Recognition R.	Learning R.	Recognition R.	Learning R.	Recognition R.
FM 1	63.11 %	66.66 %	79.53 %	74.35 %	87.25 %	74.10 %
FM 2	37.84 %	40 %	42.36 %	45.89 %	62.42 %	63.33 %
FM 3	53.51 %	59.74 %	72.40 %	79.74 %	90.96 %	85.64 %
FM 4	91.33 %	91.28 %	95.11 %	94.61 %	99.56 %	97.17 %
FM 5	49.23 %	48.46 %	70.16 %	64.35 %	78.56 %	63.33 %
FM 6	49.89 %	46.66 %	51.23 %	47.69 %	61.97 %	52.30 %
FM 7	70.66 %	70.76 %	88.76 %	82.56 %	93.57 %	84.87 %
FM 8	70.06 %	71.02 %	88.46 %	76.66 %	92.47 %	81.28 %
FM 9	78.22 %	76.92 %	98.22 %	87.94 %	97.35 %	89.48 %
FM 10	89.43 %	87.69 %	99.83 %	96.15 %	98.36 %	92.05 %
FM 11	86.92 %	83,53 %	97,18 %	94,41 %	99,63 %	96,66 %

Table 3: Machine-printed characters recognition results using different features vectors with NB, SVM and MLP classifiers.

have attained the highest recognition accuracy. HOG achieves 83.53%, 94.41% and 96.66% accuracy respectively for NB, SVM and MLP. Zoning achieves 91.28%, 94.61% and 97.17% accuracy respectively for NB, SVM and MLP.

Also FM 7 and FM 10 give some encouraging results in the both databases. However the recognition rate improves when we combine vertical and horizontal projection histogram method (FM 7), same thing when we combine the four types (left, right, top and bottom) of distance profile feature method (FM 10), but the problem which arises from these two lasts methods (FM 7 and FM 10) is that they take more time for training phase especially for MLP classifier, due to the size of its vector (such as 220 features for FM 7 and 110 features for FM 10).

All the results obtained using the three classifiers are compared in figures 11 and 12. According to these results, it can be noticed that the best results always are obtained with the Multilayer Artificial Perceptron (MLP). One important key to success of MLP is the size of the hidden layer, in this work we have tested the different sizes and as a result we opted to use the following size: the size of feature vector plus number of classes (26 characters) divided by two [25, 26, 35]. In SVM classifier, we obtained the highest result with the use of polynomial Kernel and the value 1 for cost parameter C. Regarding the subject of using the Naïve Bayes (NB) classifier, the test shown that its performance is mediocre at out recognition problem.
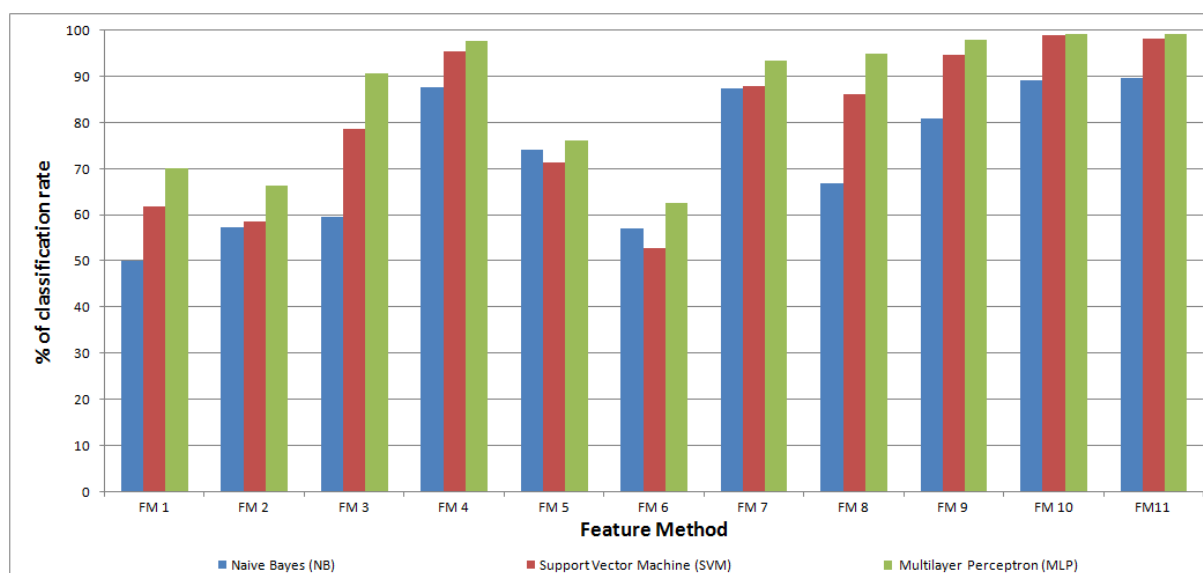
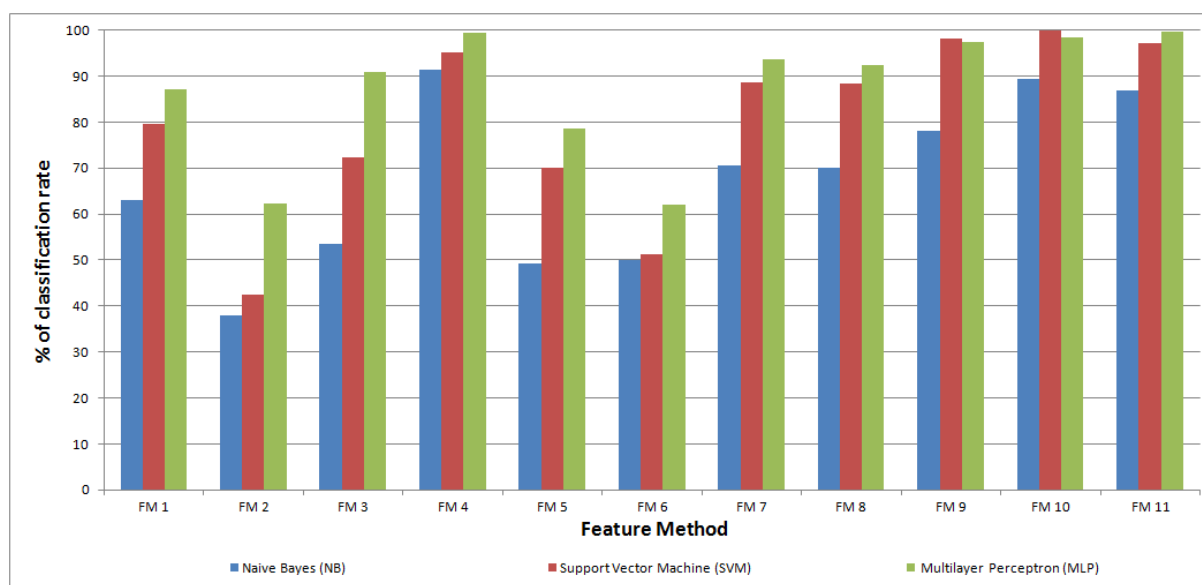Fig. 11: A comparison of numerical results obtained on handwritten characters database.

Fig. 12: A comparison of numerical results obtained on machine-printed characters database

In order to further assess the capabilities and the robustness of our system the combine HOG feature and MLP classifier, we conducted other experiments with the use of MNIST handwritten digits dataset. MNIST contains 60000 training samples of handwritten digits and 10000 testing samples of handwritten digits, this dataset is available for download from (http://yann.lecun.com/exdb/mnist/). The images in the MNIST are centered and normalized to 20x20 pixels; figure 15 shows some samples of the digits of this dataset. Table 4 presents the result obtained with our system compared with other results published on the MNIST dataset [30, 31, 32, 33, 34]. Our system provides an encouraging result with an accuracy rate of 99.47% and outperforms some results already published on the MNIST dataset.

Method	Correct (%)	Error (%)
Dan et al. [30]	91.20%	8.80%
Zhao et al. [31]	91.24%	8.76%
Mayraz et al. [32]	98.30%	1.70%
Yu et al. [33]	99.10%	0.90%
Our method	99.48%	0.52%
Ciresan et al. [34]	99.77%	0.23%

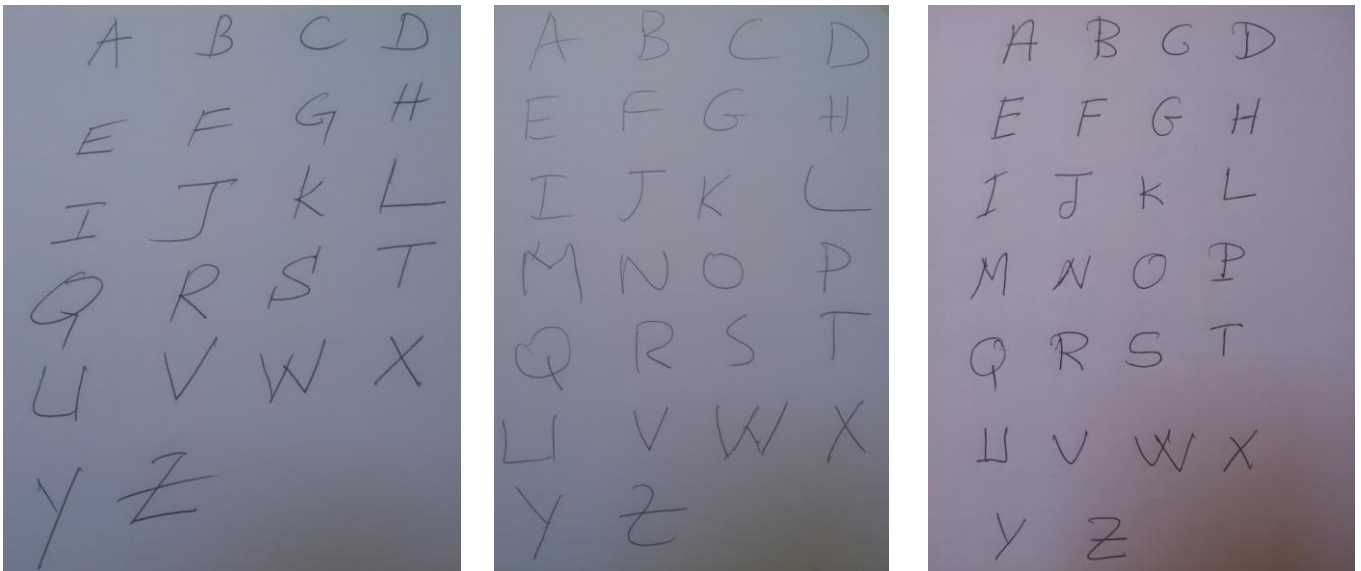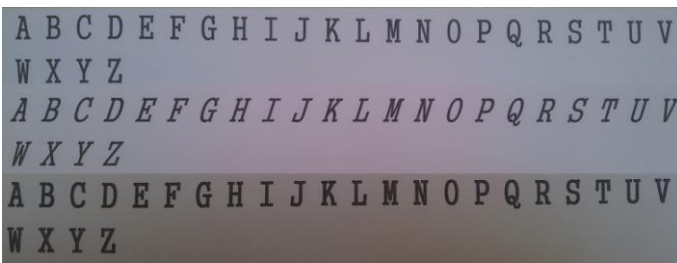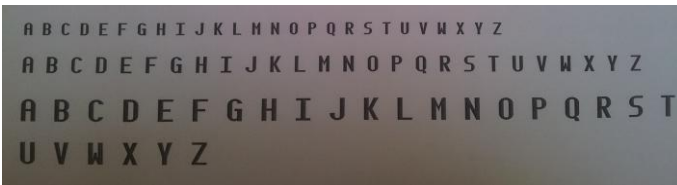Table 4. The recognition rate results on MNIST dataset

Fig. 13: Samples of three writers

(a) GungsuhChe font (Normal, Italic, Bold)

(b) Envy Code R font (20 points, 26 points and 36 points)

Fig. 14: Samples of two fonts with different sizes and styles

Fig. 15: Samples images in MNIST database.

In term of performance and efficiency, we can conclude that MLP classifier with Histograms of oriented Gradients feature are powerful tools for solving the problem of handwriting and printed characters image recognition.

4 Conclusion

An offline English handwritten and machine-printed characters recognition system for isolated image characters acquired via camera phone was introduced in this paper. Several thresholding methods have been analyzed and compared; as a result we've chosen Sauvola [15] method duo its ability to remove the noise and its conservation of the textual information. The experim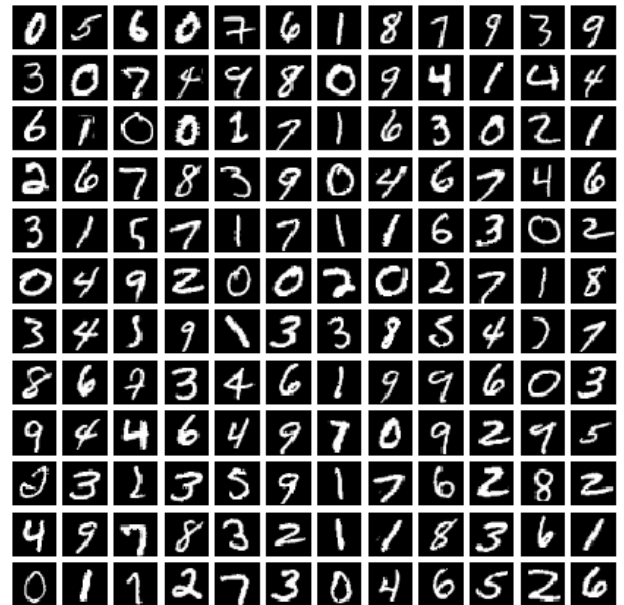ents were performed with two databases of handwritten and machine-printed character images. We examined several classifiers and for each classifier we have tested a set of features methods. The results obtained in this paper that has been compared and analyzed have shown that Multilayer Perceptron (MLP) with Histograms of oriented Gradients (HOG) feature are the best in terms of recognition accuracy rate.

In the future, we'll try to improve the results by using or adding other feature methods, also we will attempt to optimize the code to implement it in the mobile phone. In addition, we will extend our system to the recognition of handwriting or printed words.

References:

[1] Henry S. Baird, "The State of the Art of Document Image Degradation Modeling," In *Proc. of 4 th IAPR International Workshop on Document Analysis Systems*. Rio de Janeiro, 2000, pp. 1–16.

[2] F. Drira, F. Le Bourgeois and H. Emptoz, "Document images restoration by a new tensor based diffusion process: Application to the recognition of old printed documents," *10th International Conference on Document Analysis and Recognition (ICDAR09)*. Barcelone, 2009, pp. 321–325.

[3] B. Smith, "RSLDI: Restoration of single-sided low-quality document images," *Pattern Recognition, Special Issue on Handwriting Recognition*. no. 42, pp. 3355-3364 (2009).

[4] I. Nwogu, Z. Shi, and V. Govindaraju, "Pde-based enhancement of low quality documents," In *The (ICDAR07)*, page 541-545, Vol.01, 2007.

[5] S. Saoud, Z. Mahani, M. El-Rhabi and A. Hakim, "Document scanning in a tough environment: application to camera phone," *International Journal of Imaging and Robotics (IJIR), Special issue on Practical Perspective of Digital Imaging for Computational Applications,* 9(1):1-16, 2013.

[6] J. Kim and H. Lee, "Joint nonuniform illumination estimation and deblurring for bar code signals," *Optic Express.*, vol. 15, issue 22, pp. 14817–14837, 2007.

[7] H.G. Barrow and J.M. Tenenbaum," Recovering intrinsic scene characteristics from images," In *CVS78,* pages 326, 1978.

[8] Szirmay-Kalos Laszlo, "Monte-Carlo Global Illumination Methods - State of the Art and New Developments," *SCCG'99,* Invited talk, 1999.

[9] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.* vol. 12, no. 7, pp. 629–639, 1990.

[10] Z. Mahani, J. Zahid, S. Saoud, M. El Rhabi and A. Hakim, "Text enhancement by pde's based methods," *Lecture Notes in Computer Science, Image and signal Processing*, 7340:65-76, 2012.

[11] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics SMC-9 (1)*, pp. 62–66, 1979.

[12] R. Moghaddam and M. Cheriet, "Low quality document image modeling and enhancement," *International Journal of Document Analysis and Recognition*, vol. 11, no. 4, pp. 183–201, 2009.

[13] Kittler, J., Illingworth, J, "On threshold selection using clustering criteria," *IEEE transactions on systems, man, and cybernetics*, 15:652–655, 1985.

[14] W. Niblack, "An introduction to digital image processing," *Prentice-Hall, Englewood Cliffs, New Jersey*, pp. 115–116, 1986.

[15] J. Sauvola and M. Pietikainen, "Adaptive Document Image Binarization," *Pattern Recognition 33(2)*, pp. 225–236, 2000.

[16] M. El Rhabi and G. Rochefort, Realeyes3D SA, patent. Available : http://patentscope.wipo.int/search/en/WO20091 12710

[17] H El Bahi, Z Mahani, A Zatni, "An enhancement text method for image acquired via digital cameras by PDE's stable model," *In Proc. of Proceedings of the 18th International Conference on Circuits, Santorini Island, Greece*, pp. 309-313, 2014.

[18] Zhang, T.Y., and Suen, C.Y. "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM, vol.27 (3),* pp.236–240, 1984.

[19] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics,* vol. 3, pp. 610-621, 1973.

[20] A. B. S. Hussain, G. T. Toussaint and R. W. Donaldson, "Results obtained using a simple character recognition procedure on Munson's hand printed data", *IEEE Transactions on Computers*, pp. 201-205, 1972.

[21] S. K. Hwang, W. Y .Kim, "A novel approach to the fast computation of Zernike moments", *Pattern Recognition* (36), pp. 2065–2076, 2006.

[22] J.G. Daugman, " Two-dimensional spectral analysis of cortical receptive field profile," *Vision Research,* 20, pp. 847-856, 1980.

[23] Siddharth, K.S., Jangid, M., Dhir, R., Rani, R., "Handwritten Gurmukhi Character Recognition Using Statistical and Background Directional Distribution Features", *International Journal on Computer Science and Engineering (IJCSE),* Vol. 3 No. 6, pp. 2332–2345, 2011

[24] Vapnik. V. 1995. The Nature of Statistical Learning Theory. Springer, N.Y. ISBN 0-387-94559-8.

[25] K.W. Wong, C.S. Leung & S.J. Chang, "Handwritten digit recognition using multi-layer feedforward neural networks with periodic and

monotonic activation functions", *ICPR*, vol. 3, 2002, pp. 106–109.

[26]   D. Navneet, and T. Bill, "Histograms of Oriented Gradients for Human Detection" *IEEE Computer Vision and Pattern Recognition*, 886–893, 2005.

[27]   D.Lowe, "Distinctive Image Features from Scale-Invariant Keypoints" *International Journal of Computer Vision*, 60(2), 91-110, 2004.

[28] A Kefai, T Sari, H Bahi, *Text/ Background separation in the degraded document images by combining several thresholding techniques,* Wseas Transactions on Signal Processing, ., vol. 10, 2014, pp.436-443.

[29] A. Tigora, "An Image Binarization Algorithm Using Watershed-Based Local Thresholding*, in Proc. 1st WSEAS International Conference on Image Processing and Pattern Recognition (IPPR '13), Budapest, Hungary*, December 10-12, 2013, pp.154-160.

[30] Z. Dan and C. Xu, "The recognition of handwritten digits based on bp neural network and the implementation on android," *Third International Conference on Intelligent System Design and Engineering Applications*, pp. 1498–1501, 2013.

[31] Z. Zhao, C.L. Liu, and M. Zhao. "Handwriting representation and recognition through a sparse projection and low-rank recovery framework," *In International Joint Conference on Neural Networks (IJCNN),* pp. 1-8, 2013.

[32] G. Mayraz and G. E. Hinton, "Recognizing handwritten digits using hierarchical products of experts" *EEE Transactions on Pattern Analysis and Machine Intelligence,* vol 24, pp. 189–197, 2002.

[33] N. Yu and P. Jiao ,"Handwritten Digits Recognition Approach Research based on Distance &Kernel PCA" *IEEE fifth International Conference on Advanced Computational Intelligence(ICACI)* pp. 689-693, 2012.

[34] D. Ciresan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification" *In CVPR*, pp. 3642-3649, 2012.

[35] J Pradeep, E Srinivasan, S Himavathi, "An Investigation on the Performance of Hybrid Features for Feed Forward Neural Network Based English Handwritten Character Recognition System" *WSEAS Transactions on Signal Processing*, pp. 21-29, 2014.