Handwritten Recognition using Neural Networks Based on Multiple Feature Extraction Algorithms

Bassim AbdulBaqi*1, Ayad Ghazi Naser2, Maryam Khalaf Kadum1
1Department, of Computer Engineering and Information Technology, University of Technology, Iraq
2General Directorate of Vocational Education, Ministry of Education, Iraq

Abstract
Handwritten letters are vague in nature as there may not always be sharp perfectly straight lines, and curves not necessarily be clearly, unlike the printed letters. Furthermore, letter can be drawn in different sizes. Therefore, the proposed used of four methods based on two algorithms for feature extraction. The first algorithm discrete wavelet transform and second algorithm is Radon transform, by experiments was found that the feature extraction using radon transform is more precise than wavelet transform. The proposed technique computes Radon projections in different orientations and captures the directional features of letter images. The obtain recognition rate was 93% when used for train artificial neural networks. Comparative between four methods to choose the best has been done. Training method is based on Back-Propagation learning algorithm used by feed forward Neural Network.

Keywords: Feature extraction, Radon transform, Wavelet transform, Neural Networks.

Email:basimalani@yahoo.com
Introduction

Letter Recognition System (LRS) passes in steps: first step to image acquisition and then followed by second step preprocessing, third step is feature extraction finally training and classification is fourth step. Feature extraction is a vital step for pattern recognition, especially for handwritten letter recognition as there are varieties of handwritten styles depending on the subject’s age, gender and educational, as well as his/her mood while writing. Many feature extraction methods have been reported such as various moment features, gradient and distance-based features, geometric features, transform domain features, etc. Wavelet transform has been widely used in image processing and signal processing for image/signal enhancement, denoising, texture segmentation [1] based on its properties of short support, orthogonally, symmetry, higher order of vanishing moments, and more importantly, its multi-resolution decomposition analysis.

In the literature survey we have found that numbers of authors have attempted to recognize the handwritten letter using different techniques. [2] Used two types of wavelet features are proposed: Kirsch edge enhancement based 2D wavelets and 2D complex wavelets. Multilayer perceptron used as a classifier. [3], proposed a method for Character Recognition Using Radon Transformation and Principal Component Analysis in Postal Applications. [4] developed Automatic Handwritten Character Recognition System; the proposed approach consists of two stages: Extraction of Wavelet Relative Energy Distribution Feature based on wavelet transform and classification of different handwritten characters. Offline handwriting recognition-the transcription of images of handwritten text-is an interesting task, in that it combines computer vision with sequence learning. Letter Recognition System (LRS) designed for offline handwriting recognition. The proposed system for training and testing phases shown in figure 1.

![Figure 1- The proposed system for LRS.](image)

**Phases In LRS**

LRS consists of following stages: image acquisition, pre-processing, feature extraction, classification and recognition.

**Image Acquisition**

In this phase the input image is taken through camera or scanner.
Pre-processing

In preprocessing operations sample image are converted into gray scale. Then we applied these techniques, Gray scale image are converted into binary image using threshold value obtained by Otsu’s method[5], filtering operation, morphological operation, Binarization, and skeletonization of a digital image so that subsequent algorithms along the road to final classification can be made simple and more accurate. The corresponding objectives of Pre-processing methods are as follows:

Normalization

The input numeral image is normalized to equal size.

Noise removal

The major objective of noise removal is to remove any unwanted bit-patterns, which do not have any significance in the output. Wiener filter is lowpass-filters a grayscale image that has been degraded by constant power additive noise used for remove Gaussian noise as shows in figure 2. Median filtering is a nonlinear operation often used in image processing to reduce "salt and pepper" noise [6] as shows in figure 3.

Skeletonization

Skeletonization is the process which reduces the width of a line. This process can remove irregularities in letter. Skeletonization applied on letter stroke which have only one pixel width. It requires less memory to store information about input numeral. This method also requires less processing time. After preprocessing phase, a cleaned image is available that goes to the feature extraction phase. Figure 4 shows preprocessing stages of image [4].
**Feature Extraction**

Each character has its own differential features, which play an important role in pattern recognition. Feature extraction describes the relevant shape information contained in a pattern so that the task of classifying the pattern is made easy by a formal procedure.

The main goal of feature extraction is to obtain the most relevant information from the original data and represent that information in a lower dimensionality space. When the input data to an algorithm is too large and also may be redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector).

A term feature extraction is termed that transforms the input data into the set of features. This reduced representation instead of the full size input [7, 8].

In feature extraction stage each letter is represented as a feature vector, which becomes its identity. The major goal of feature extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements. Due to the nature of handwriting with its high degree of variability and imprecision obtaining these features, is a difficult task. Therefore used feature extraction methods are:

**Radon Transform (RT)**

The Radon transform is named after the Austrian mathematician Johann Karl August Radon. The main application of the Radon transform is CAT scans, where the inverse Radon transform is applied. The Radon transform can also be used for line detection.

The Radon transform computes projections of an image matrix along specified directions. A projection of a two dimensional function \( f(x,y) \) is a set of line integrals. The Radon function computes the line integrals from multiple sources along parallel paths, or beams, in a certain direction [9].

The beams are spaced 1 pixel unit apart. To represent an image, the radon function takes multiple, parallel-beam projections of the image from different angles by rotating the source around the center of the image. The figure 5 shows a single projection at a specified rotation angle.
The Radon transform is the projection of the image intensity along a radial line oriented at a specific angle. The radial coordinates are the values along the x'-axis, which is oriented at θ degrees counter clockwise from the x-axis. The origin of both axes is the center pixel of the image.

For example, the line integral of f(x,y) in the vertical direction is the projection of f(x,y) onto the x-axis; the line integral in the horizontal direction is the projection of f(x,y) onto the y-axis.

Projections can be computed along any angle θ, by use general equation of the Radon transformation

\[ R(x') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta - y \sin \theta - x') dx dy \quad \ldots (1) \]

where \( \delta(\cdot) \) is the delta function with value not equal zero only for argument equal 0, and:

\[ x' = x \cos \theta + y \sin \theta \]

where \( x' \) is the perpendicular distance of the beam from the origin and \( \theta \) is the angle of incidence of the beams.

The 2-D Radon transform is the mathematical relationship which maps the spatial domain (x,y) to the Radon domain (p,θ) where p is, θ is angle of radon transform. The Radon transform consists of taking a line integral along a line (ray) which passes through the object space. The \( \theta \) values are (0, 45, 90 and 135) to minimize of feature vector size. In figure 6 shown radon transform of image of letter (a), In figure 7 shown radon transform of image of letter (b).
Mean of Radon Transform (MRT)

The 2-D Radon transform with four angles $\theta$, the $\theta$ values are (0, 45, 90 and 135) is used to minimize of feature vector $f[v_i]$ size. Mean of the extracted four feature vector is taken. This vector computed accroding equation (3).

$$f[v_i] = \frac{1}{n} \sum_{i=1}^{n} R_{ij}$$  \hspace{1cm} ... (3)

Where $\theta = \{0, 45, 90, 135\} \ n=71$.

Wavelet Transform

The Wavelet Transform (WT) is a powerful technique for representing data at different scales and frequencies. In addition to being an efficient, highly intuitive framework for the representation and storage of multi-resolution images, the discrete wavelet transform (DWT) provides powerful insight into an image’s spatial and frequency characteristics. It provides the local information in both the space and the frequency domain. The ability to capture local information is essential because numerals are locally quite different [10].

Let $\phi(x)$ and $\psi(x)$ be the scaling and wavelet functions of variable $x$. The scaling and wavelet functions are calculated by the following equation (4) and (5) respectively,

$$\Phi_{jk}(x) = 2^{j/2} \phi(2^j x-k)$$  \hspace{1cm} ... (4)

$$\Psi_{jk}(x) = 2^{j/2} \psi(2^j x-k)$$  \hspace{1cm} ... (5)

Where $k$ denotes the position of the 1-d functions $\Phi(x)$ and $\psi(x)$ along the $x$ axis and $j$ denotes their width along the $x$ axis. $2^j$ denotes their height or amplitude.

In the 2D case, the Approximation, Horizontal, Vertical, Diagonal coefficients are calculated by the Eq. (6), (7), (8) and (9) respectively,

$$\Phi_{jk}^{\text{LL}}(x,y) = \Phi_{jk}(x)\Phi_{jk}(y)$$  \hspace{1cm} ... (6)

$$\Phi_{jk}^{\text{LH}}(x,y) = \Phi_{jk}(x)\Psi_{jk}(y)$$  \hspace{1cm} ... (7)

$$\Phi_{jk}^{\text{HL}}(x,y) = \Psi_{jk}(x)\Phi_{jk}(y)$$  \hspace{1cm} ... (8)

$$\Phi_{jk}^{\text{HH}}(x,y) = \Psi_{jk}(x)\Psi_{jk}(y)$$  \hspace{1cm} ... (9)

The character image is divided into ten sub-band images $\{[LL^3],[HL^3,LH^3,HH^3],[HL^2,LH^2,HH^2],[HL^2,LH^2,HH^2],[HL^1,LH^1,HH^1]\}$ after 3 scale haar wavelet decomposition, is shown in figure 8
Here, $LL^3$ sub-band image components are ignored as it represents the basic shape of the character image. LH gives the vertical high frequency component or horizontal details. HL gives the horizontal high frequency component or vertical details. HH gives the diagonal details. Then Wavelet Relative Energy Distribution Features are extracted from the Sub-band images $\{HL^2, LH^2, HH^2\}$, $\{HL^1, LH^1, HH^1\}$ with respect to the sub bands images $\{HL^3, LH^3, HH^3\}$. The extraction procedure is construction of sub-blocks from the sub-band images:

The wavelet transform sub bands ($HL^3, LH^3, HH^3$) as shown in figure 8, are three detail coefficients matrices at scale 3 and the size of each detail coefficient matrix is $(6 \times 6)$. The each detail coefficients matrix is divided into 36 sub-blocks each contain one pixel as shown in the figure 9.

The wavelet transform sub bands ($HL^2, LH^2, HH^2$) are three detail coefficients matrices at scale 2 and the size of each detail coefficient matrix is $(12 \times 12)$. Then each matrix is divided into 36 sub-blocks and the size of each sub-block is $(2 \times 2)$ pixels as shown in figure 10.

Figure 8-10 sub-band images of an image after 3 scale wavelet decompositions.

Figure 9- each detail coefficient matrix of size $(6 \times 6)$ at scale 3 is divided into $(1 \times 1)$ sub-blocks.
After wavelet transform, $HL_1$, $LH_1$, $HH_1$ are three detail coefficients matrices at scale 1 and the size of the each detail coefficient matrix is $(24 \times 24)$. The each matrix is divided into 36 sub-blocks and the size of the each subblock is $(4 \times 4)$ pixels as shown in the figure 11.

Extraction process of wavelet relative energy distribution feature:

As shown in the above figures (9), (10), (11), Horizontal, Vertical and Diagonal Sub-band images at scale1, scale 2and scale 3 are divided into 36 sub-blocks. The energy of the Sub-blocks of Horizontal, vertical, diagonal details at scale 3 are each divided by corresponding energy of the Sub-block of horizontal, vertical, diagonal details at scale2. So, the energy of the $1^{st}$, $2^{nd}$, $3^{rd}$, $4^{th}$ ... $36^{th}$ Sub-block of horizontal details at scale 3 are divided by the energy of the $1^{st}$, $2^{nd}$, $3^{rd}$, $4^{th}$ ... $36^{th}$ Sub-block of horizontal details at scale 2 respectively, the energy of the, $1^{st}$, $2^{nd}$, $3^{rd}$, $4^{th}$ ... $36^{th}$ Sub-block of vertical details at scale 3 are divided by the energy of the $1^{st}$, $2^{nd}$, $3^{rd}$, $4^{th}$ ... $36^{th}$ sub-block of vertical details at scale 2 respectively and the energy of the $1^{st}$, $2^{nd}$, $3^{rd}$, $4^{th}$ ... $36^{th}$ Sub-block of diagonal details at scale 3 are divided by the energy of the $1^{st}$, $2^{nd}$, $3^{rd}$, $4^{th}$ ... $36^{th}$ Sub-block of diagonal details at scale 2 respectively.

In the same way, the energy of the sub-blocks of horizontal, vertical and diagonal details at scale 3 are divided by the energy of the corresponding sub-block of horizontal, vertical and diagonal details at scale1.

So, one feature value can be extracted from each sub-block of the sub-band images at scale 1 and scale 2 comparing with the energy of corresponding sub-block at scale 3 following the above mentioned way. There are total six sub-band images at scale 1 and scale 2 and the six sub-band images are divided into $216(36 \times 6)$ sub-blocks.

The energy of the each Sub-block is calculated according to the equation (10).


\[ E_{\text{energy}_{\text{SubB}}(n)} = \sum_x \sum_y |\text{coef}(x, y)| \]  
\[ \text{where } x \in \text{SubB}(n), y \in \text{SubB}(n), \text{SubB}(n) \text{ is a sub-block in one sub-band image, where } n=1, 2, \ldots, 36. \text{ For each sub-block at scale 1 and scale 2, one Relative Energy Distribution Feature is extracted using the equation (11)} \]

\[ \text{Feature}_{\text{Sub}(n)} = \frac{\text{coef}_{\text{SubB}}(n)}{E_{\text{energy}_{\text{SubB}}(n)}+1} \]  
\[ \text{ coef }_{\text{SubB}}(n) \text{ is a coefficient value of one sub-block of each details Subband image at scale 3, Where } n=1, 2, \ldots, 36. 1 \text{ is added with the energy } \text{SubB}(n) \text{ because sometimes the value of energy } \text{SubB}(n) \text{ can be zero.} \]

So, from the 216 (36x6) Sub-blocks of 6 sub-band images at scale 1 and 2, 216 Relative Energy Distribution Features are extracted. So, the size of feature vector for each character image is 216. Three scales of letter (a) shown in figure 12.

![Wavelet Transform 3-scale](image)

**Wavelet Scalar Quantization Compression**

The Wavelet Scalar Quantization (WSQ) compression class of encoders involves a decomposition of the fingerprint image into a number of subbands, each of which represents information in a particular frequency band. The subband decomposition is achieved by a discrete wavelet transformation of the fingerprint image.

After the 2D wavelet decomposition, the original image can be represented by 64 subimages \([a, d_1, d_2, \ldots, d_{63}]\) where \(a\) is the zero-th subimage of the 64 subimages, which is a low resolution approximation of the original image, and \(d_j, j = 1, 2, \ldots, 63\) are the wavelet subimages containing the image details at different scales and orientations.

The ridge orientation as well as the ridge spatial frequency in different image regions represents the intrinsic nature of the fingerprint image. Both kind of information are well extracted into the wavelet coefficients subimages \(d_j, j = 1, 2, \ldots, 63\) of the 64 subbands, the zero-th band is equivalent of the DC component in the signal. Hence, this band contains no valuable information about the image features [11].

The normalized \(l_2\)-norm of each wavelet subimage \(d_j\) is computed in order to create a feature vector of length 63:

\[ [e_1, e_2, \ldots, e_{63}] \]

where

\[ e_j = \frac{\|d_j\|_2}{\sum_{i=1}^{63} \|d_i\|_2} \]  
\[ \text{for all } j = 1, 2, \ldots, 63. \text{ The } l_2\text{-norm } \|x\|_2 \text{ of a vector } x = [x_1, \ldots, x_N]^{T} \text{ is defined as} \]

\[ e(x) = \sqrt{\sum_{i=1}^{N} x_i^2} \]  
\[ \ldots (13) \]
In contrast with the wavelet transform which recursively decomposes subimages in the low frequency bands, the tree-structured wavelet transform (or wavelet packets) also decomposes the high frequency bands that contain significant information. So, feature vector contains more detailed information about the image energy distribution over different frequencies and orientations. It can thus provide more valuable properties for the discrimination of the fingerprint patterns. Figure (13) shows wavelet decomposition structure.

![Wavelet Decomposition Structure](image)

**Figure 13-** wavelet decomposition structure.

**Classification and recognition**

The Artificial neural networks (ANN) were used for the purpose of classification and recognition. For compute recognition rate we used the following equation

\[
\text{recognition rate} = \frac{C}{T} \times 100 \%
\]  \hspace{1cm} \text{... (14)}

Where \( C \) is a number of correct recognition and \( T \) is the total number of test sample. [12]

The used database for training and testing ANN is image collected 390 samples from different 7 people, 260 sample for training ANN and 130 samples for testing ANN. Multilayer networks have been used in optical character recognition systems for many years. Each node in a layer has full connections from the nodes in the previous layer and the proceeding layer. There are several layers of neurons: the input layer, hidden layers and the output layer. During the training phase, connection weights are learned. The output at a node is a function of the weighted sum of the connected nodes at the previous layer. The neural network structure in layers: input layer, hidden layer and output layer. Input layer depended on feature extraction method used. Different size of features vector effect on optimal number neurons in number of hidden layer hidden layer of neural network, therefore, Brute Force was used for choose best NN structure gives highest recognition rate. Figure (14) shows NN structure of Letter Recognition System using Radon Transform (LRS-RT) The NN of LRS-RT is trained and reached minimum error value (9.91 \( \times 10^{-6} \)) at 969 epochs and Figure (15) NN structure of Letter Recognition System using Mean Radon Transform (LRS-MRT) The NN of LRS-MRT is trained and reached minimum error value (9.95 \( \times 10^{-6} \)) at 12233 epochs and Figure 16 shows NN structure of Letter Recognition System using Discrete Wavelet Transform (LRS-DWT) The NN of LRS- DWT is trained and reached minimum error value (8.92\( \times 10^{-6} \)) at 3793 epochs. Figure 17 shows NN structure of Letter Recognition System using Wavelet Scalar quantization (LRS-WSQ) The NN of LRS- WSQ is trained and reached minimum error value (5.26 \( \times 10^{-6} \)) at 250 epochs.
Experimental results
In this paper four different methods were used, therefore length of feature vector is different for each method. Table 1 shows the length of feature vector for each used method:
After training NN the obtain result shown in table 2 that shows recognition rate of each method when used in LRS as feature extraction method. The Radon transform gives high recognition than wavelet Transform. Where \( L_i \) is number of neurons in the input layer, \( L_{H1} \) is the number of neurons in 1st hidden layer, \( L_{H2} \) is the number of neurons in 2nd hidden layer, \( L_o \) is the number of neurons in output layer.

### Table 1: Length of feature vector of method

<table>
<thead>
<tr>
<th>Method</th>
<th>Length of Feature vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>284</td>
</tr>
<tr>
<td>MRT</td>
<td>4</td>
</tr>
<tr>
<td>WT</td>
<td>216</td>
</tr>
<tr>
<td>WSQ</td>
<td>63</td>
</tr>
</tbody>
</table>

### Table 2: Recognition rate achieved by used method

<table>
<thead>
<tr>
<th>Method</th>
<th>NN-structure</th>
<th>Mean Square Error</th>
<th>NO. Iteration</th>
<th>Recognition using back propagation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( L_i )</td>
<td>( L_{H1} )</td>
<td>( L_{H2} )</td>
<td>( L_o )</td>
</tr>
<tr>
<td>RT</td>
<td>284</td>
<td>200</td>
<td>45</td>
<td>26</td>
</tr>
<tr>
<td>MRT</td>
<td>4</td>
<td>45</td>
<td>40</td>
<td>26</td>
</tr>
<tr>
<td>WT</td>
<td>216</td>
<td>300</td>
<td>104</td>
<td>26</td>
</tr>
<tr>
<td>WSQ</td>
<td>63</td>
<td>120</td>
<td>45</td>
<td>26</td>
</tr>
</tbody>
</table>

As shown MRT have 93% recognition rate which is the highest.

### Conclusions

According to the results of the performing, can conclude the following:
- In this paper, using neural networks for handwritten recognition is a field that attracting a lot of attention.
- Using the Mean of Radon transform feature extraction is more precise than wavelet transform, the feature vectors are fed to the Neural Networks.
- Using of additional features to improve classification results. It is known as increasing feature dimensionality

### References