WRITER IDENTIFICATION SYSTEM FOR ETHIOPIC HANDWRITING

Daniel Demoze and Eneyew Adugna
Department of Electrical & Computer Engineering
Addis Ababa University

ABSTRACT

Writer identification is a popular and ongoing research area having a wide variety of applications in banking, criminal justice system, access control, determining the authenticity of handwritten mails, etc.

In this paper, an off-line text independent Ethiopic writer identification system has been proposed. The system uses 50 *handwritten text blocks collected from* 25 *volunteers* (*each person was made to write on two A*-4 *size pages in Ethiopic handwriting*)*. These text blocks are scanned and stored for further processing by the identification system.*

Two approaches have been employed for feature extraction from the handwritten images: texture level using multi-channel Gabor Energy Features and the character-shape (*allograph*) *level using codebook of connected component contour.*

Experimental results demonstrate that 93% *correct identification, in a hit list of size* 3, *and* 76%, *in a hit list of size* 1, *using Gabor energy features and* 96% *correct identification, in hit list of* 3, *and* 92%, *in hit list* 1, *using codebook of connected component contours, are acquired.*

Keywords: *writer identification, writer verification, image pre-processing, multi-channel Gabor filtering, Gabor Energy Feature, Connected Component Contour and Codebook.*

INTRODUCTION

Biometric modalities are classified into two broad categories: physiological biometrics that perform person identification based on measuring a physical property of the human body (e.g. fingerprint, face, iris, retinal blood vessels, hand geometry, DNA) and behavioral biometrics that use individual traits of a person's behavior for identification (e.g. voice, gait, keystroke dynamics, signature, handwriting).

Identifying the writer of a handwritten sample using automatic image-based methods is an interesting pattern recognition problem with a wide variety of applications including banking, forensics, access control, determining the authenticity of handwritten mails, and historic document analysis fields.

A writer identification system performs a one-to-many search in a large database with handwriting samples of known authorship and returns a likely list of candidates. This list is further scrutinized by the forensic expert who makes the final decision regarding the identity of the author of the questioned sample.

Writer identification method fall into two broad categories [1]: text-dependent vs text-independent methods. The text-dependent methods [2, 3, 4] are very similar to signature verification techniques and use the comparison between individual characters or words of known text (ASCII) content. These methods therefore require the prior segmentation by hand of the relevant information. The text-independent methods [5, 6, 7, 8] use statistical features extracted from the entire image of a text block.

A scientific validation of individuality of handwriting is performed by Srihari et al. [9] who proposed a large number of features divided into two categories: Macro-features and Micro-features. Micro-features are better than macro-features in identification tests with a performance exceeding 80%. A multilayer perceptron or parametric distributions are used for writer verification with an accuracy of about 96% using Latin script.

Bensefia et al. [5] used connected components generated by a handwriting segmentation method to encode the individual characteristics of handwriting independent of the text content. Connected component clustering is used to define a feature space common for all documents in the dataset. Experimental results are reported on three datasets containing 88 writers, 39 writers (historical documents) and 150 writers, with 2 samples (text blocks) per writer. Writer identification rates around 90% are reported on the different test datasets.

Said et al. [6] proposed a text-independent approach and derived writer-specific texture features using multichannel Gabor filtering and gray-scale co-occurrence matrices. Two sets of 20 writers, 25 samples per writer are used in the evaluation. Nearest-centroid classification using weighted Euclidean distance and Gabor features achieved 96% writer identification accuracy.

Schomaker et al. [8] present a new approach, using connected component contours codebook and its probability density function. Also, combining connected-component contours with an independent edge-based orientation and curvature PDF yields very high correct identification rates.

Bulacu et al. [10] evaluated the performance of edge-based directional probability distributions as features in comparison to a number of non-angular features. It is noted that the joint probability distribution of the angle combination of two hinged edge fragments outperforms all other individual features.

Although this research area is necessary to combat crime and terrorist threats, there is no research that has ever been conducted on Ethiopic handwriting. Thus, this area has not been even touched and this is a pioneer work, which shall open the way for other researchers to involve in the area.

After closely reviewing the reported works related to handwriting identification, in this thesis the proposed systems are modeled by implementing preprocessing and two feature extraction techniques (Gabor Energy and Connected Component Contour). Gabor energy is used for feature extraction in the first approach and connected component segmentation technique is used for preprocessing in the second approach.

The rest of this paper is organized as follows: *i*) an overview of Ethiopic characters and characteristics of Ethiopic handwriting is presented; *ii*) writer identification system implementations using Gabor energy and connected component feature extracting techniques are discussed, respectively; *iii*) experimental results of the two proposed writer identification systems is evaluated using the x^2 distance measure; and *iv*) finally, discussion on the implementation and results is provided followed by conclusion and recommendations on possible outlooks for future work is provided.

THE ETHIOPIC WRITING

As the main concern of this work is the Ethiopic handwriting, this section discusses the Ethiopic character set as a whole and characteristics of Ethiopic handwriting.

The Ge'ez script, which dates back to AD 300, is one of the native African writing systems [11]. The original Ge'ez script has only 26 consonants and has no vowel indications until around 350 A.D. [12]. But, later, vowels were incorporated by adding six non-basic characters for each consonant symbol. These were created by making structural changes to the existing consonant symbols and this made the total number of characters in Ge'ez script 182 [12]. Eventually, the Ge'ez script gave birth to the Amharic writing system, which is now used to write Amharic and other Semitic languages like Tigre, Harari and Gurage in Ethiopia [13].

In addition to the 26 consonants and their six non-basic forms that exist in Ge'ez system, the Amharic writing system later added 7 other consonants whose shape were derived from some of the existing Ge'ez characters [12]. This increases the total number of characters to 231 (i.e., 33×7) in the newly born Amharic character set.

The total number of characters in the Amharic character set is 306 if the Ethiopian numerals, punctuation symbols, and the extended character group (□, □ etc) are counted with the alphabetic characters. From these, 231+7 of them are alphabetic letter. Basically, the number of basic characters is 33 and the total number of alphabetic characters will be 231. But, the character □ is included as a special character to represent the sound 'v' from Latin-based languages and it has also six non-basic forms.

Amharic writing system is often called syllabary rather than an alphabet because the seven orders of Amharic characters indicated above represent syllable combination consisting of consonant and a following vowel. The non-basic forms (vocalization) are derived from the basic forms (consonants) by attaching small appendages (diacritic marks) to the right, left, top, or bottom in more or less regular modification. Some are formed by adding strokes, others by adding loops or other forms of differentiation to each core character.

For character based writer identification character segmentation is a major issue. In most handwritings since the characters are written in connected way, i.e in cursive way, segmentation is a difficult task. But in Ethiopic handwriting characters are written separated and it is much easier to segment a character or part of a character.

SYSTEM IMPLEMENTATION USING GABOR ENERGY FEATURE EXTRACTION TECHNIQUE

Based on the idea that has been presented in Said et al. [6], we can assume handwriting as texture image and writer identification as texture classification. For this purpose, in the first step the image of a document is preprocessed and then a set of Gabor filters with eight equidistant orientations and three spatial frequencies were applied, resulting in 24 filtered images. Then, features are extracted from these filtered images based on Gabor-energy method.

Preprocessing

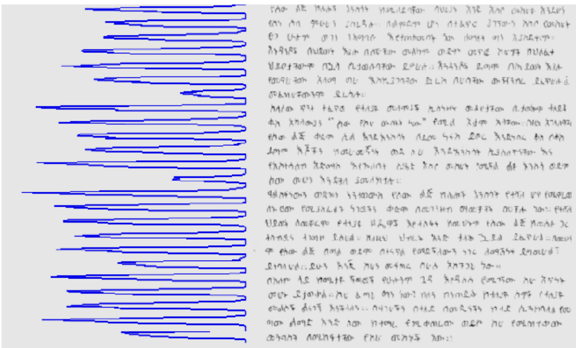
The original scanned image was gray scale. To change the image into binary Otsu algorithm [14] is used. It may contain characters of different sizes and spaces between text lines. For the purpose of texture feature extraction, the input documents need to be normalized to create a uniform block of text. Figure 1b shows an original handwriting image and this procedure can be accomplished in the following four steps.

Locate text lines

Line segmentation allows the ascenders and descenders of consecutive lines to be separated. In the manuscripts it is observed that lines consist of a series of horizontal components from left to right. Projection profile technique has been widely used in line and word segmentation for machine printed documents [15]. In this technique a one dimensional function (1D) of the pixel values is obtained by projecting binary image on to the horizontal axis (vertical axis is *x*-axis and horizontal axis is *y*-axis). Let $f(x, y)$ be digital value of pixel (x, y) in digital image, where zero value correspond to black pixel and one correspond to white pixel. The horizontal projection profile (HPP) is defined as

$$p(x) = \sum_{y=0}^{w} f(x, y) \qquad (1)$$

where W is the width of the image. Figure 1 shows sample handwritten image in (*b*) and its HPP in (*a*). The horizontal projection profile (HPP) of the document is computed. The valley between peaks corresponds to text lines. The distance between two valleys corresponds to the space between each text line as shown in Fig. 1. To segment text lines the peak points are used.



(*a*) (*b*)

Figure 1 Extraction of lines from handwritten text.
(*a*) Horizontal projection profile
(*b*) Handwritten image

Normalize height of text lines

Since the size of each text line may differ greatly in the handwriting documents, it is necessary to resize each character to a similar size. Given that the height of each text line can be obtained, as indicated above, text line height can easily be normalized. The average size of the text line is used for normalization.

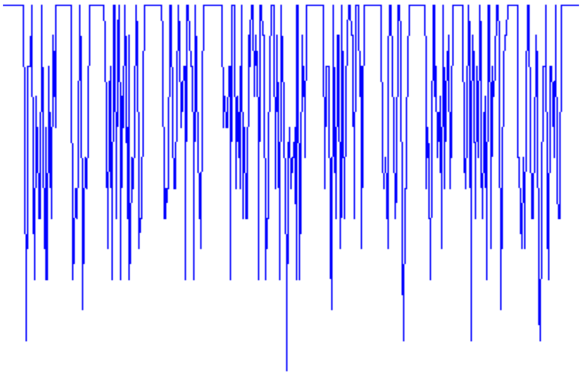
Normalize spacing

The handwriting image may contain different spaces between characters, words and text lines. Different spacing may influence the texture of the images, so the normalization of spaces is necessary. After text line localization, the vertical projection profile (VPP) is computed to determine spacing between characters or words. Let $f(x, y)$ be digital value of pixel (x, y) in digital image, where zero value correspond to black pixel and one correspond to white pixel. The vertical projection profile (VPP) is defined as:

$$p(y) = \sum_{x=0}^{h} f(x, y) \qquad (2)$$

where h is the height of the image. Figure 2 shows text line (a) and its vertical projection profile (*b*).

(*a*)



(*b*)

Figure 2 Normalizing word spacing. (*a*) Text line (*b*) Vertical projection profile.

Form a uniform texture image by text padding

The input image may contain incomplete or partially justified text lines. The blank spaces are filled up by means of text padding. Padding may also be applied if the handwriting document contains only a small number of characters. In this work, the text is padded to create a block of a predefined size. The result of this step has been shown in Fig. 3.

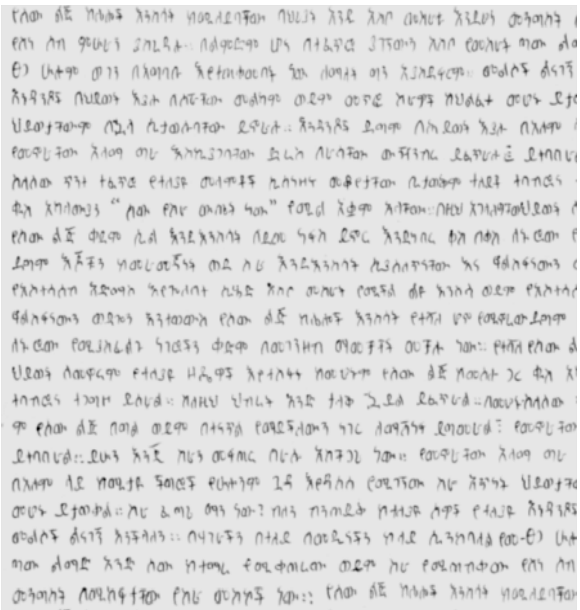


Figure 3 The final preprocessed image

Gabor Filter

In image processing, a Gabor filter, named after Dennis Gabor, is a linear filter used for feature extraction. Frequency and orientation representations of Gabor filter are similar to those of human visual system, and it has been found to be particularly appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. The Gabor filters are self-similar – all filters can be generated from one mother wavelet by dilation and rotation.

Its impulse response is defined by a harmonic function multiplied by a Gaussian function. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function.

The following family of two-dimensional Gabor function is employed to model the properties of simple cells (Eq. (3)).

$$g_{\varepsilon,\eta,\lambda,\theta,\varphi}(x,y)=\exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right) \quad (3)$$

where

$$x'=(x-\varepsilon)\cos\theta-(y-\eta)\sin\theta$$

$$y'=(x-\varepsilon)\sin\theta+(y-\eta)\cos\theta$$

and the pair (ε, η) determines the center of a receptive field in image coordinates. The standard deviation σ of the Gaussian envelope specifies the size of the receptive field. The parameter γ, called the spatial aspect ratio, determines the ellipticity of the receptive field. The parameter λ is the wavelength and 1/ λ the spatial frequency of the channel that is modeled by Gabor functions. Based on experiments, the frequency bandwidth of simple cells in the visual cortex is about one octave [16]. Therefore, the ratio σ/λ that determines the spatial frequency bandwidth is fixed to 0.56, which corresponds to a bandwidth one octave at half-response:

$$\frac{\sigma}{\lambda}=\frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}\cdot\frac{2^b+1}{2^b-1} \quad (4)$$

where the parameter b is bandwidth (in octaves).

The parameter θ specifies the orientation of the normal to the parallel excitatory and inhibitory stripe zones — this normal is the axis x' in Eq. (3) — which can be observed in the receptive fields of simple cells.

Finally, the parameter φ, which is a phase offset in the argument of the harmonic factor $s(2\pi x'/\lambda+\varphi)$, determines the symmetry of the function $g_{\varepsilon,\eta,\lambda,\theta,\varphi}(x,y)$: for $\varphi=0$ and $\varphi=180$ it is symmetric with respect to the center (ε,η) of the receptive field; for $\varphi=-\frac{1}{2}\pi$ and $\varphi=\frac{1}{2}\pi$, the function is anti symmetric and all other cases are asymmetric mixtures of these two.

Using the above parameterization, one can compute the response $\gamma_{\varepsilon,\eta,\lambda,\theta,\varphi}(x,y)$ of a simple cell modeled by a receptive field function $g_{\varepsilon,\eta,\lambda,\theta,\varphi}(x,y)$ to an input image $f(x,y)$ with gray level distribution as follows:

$$\gamma_{\varepsilon,\eta,\lambda,\theta,\varphi}(x,y)=f(x,y)*g_{\varepsilon,\eta,\lambda,\theta,\varphi}(x,y) \quad (5)$$

The Gabor energy is related to a model of complex cells which combines the responses of a pair of simple cells with a phase difference of $\pi/2$. The results of a pair of symmetric and anti-symmetric filters are combined into the Gabor energy as follows:

$$E_{\varepsilon,\eta,\lambda,\theta}=\sqrt{\gamma^2{}_{\varepsilon,\eta,\lambda,\theta,o}+\gamma^2{}_{\varepsilon,\eta,\lambda,\theta,(-\pi/2)}} \quad (6)$$

Indeed in multi-channel Gabor filtering, each channel is modeled by a pair of Gabor filter which are symmetric and anti-symmetric. For feature extraction, different methods: Linear Gabor features, Thresholded Gabor features, Gabor-energy features and complex moments features can be applied to the output of Gabor filters.

Filter Design

Hubel and Wiesel [17] deduced that simple cells are sensitive to specific orientations with approximate bandwidths of 30^o. Therefore, we use a bank of Gabor filters with three frequencies (λ = 5.4, 8.2, and 10.8) and eight equidistant orientations $\theta=k(\pi/8), k=1,2,...8)$. The frequencies and orientations are selected such that appropriate coverage of the spatial-frequency domain is achieved as shown in Fig. 4.

In fact, by this method the input image is decomposed into a number of filtered images, each of which contains intensity variation over a narrow band of frequency and orientation. Therefore, a total of 24 filtered images are obtained.

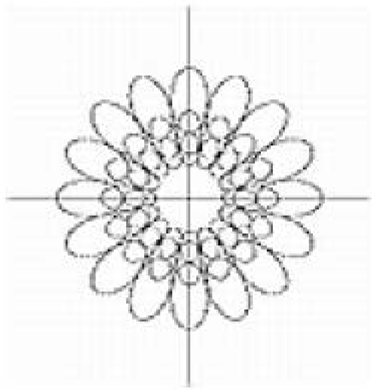


Figure 4 Coverage of spatial frequency domain by using Gabor filters [18].

Gabor Energy Features

To classify textures features can be extracted from the output of the filter using different types of Gabor features. One of these feature types is Gabor energy, which is good in texture classification [19]. Thus, in this thesis Gabor energy is used for feature extraction.

The quantity of Gabor energy (Eq. (6)) is computed based on the previous bank of Gabor filters and a set of 24 filtered images are acquired. Then, features are extracted from these images based on the mean and standard deviation values.

SYSTEM IMPLEMENTATION USING CONNECTED-COMPONENT CONTOUR FEATURE EXTRACTION TECHNIQUE

This section introduces an allograph or character shape based method for writer identification. The fundamental underpinning of this method is the idea of assuming that each writer acts as a stochastic generator of ink-traces, or connected components. The probability distribution of these simple shapes in a given handwriting sample is characteristic for the writer and is computed using a common codebook of connected components obtained by clustering. Originally proposed in [8], the theoretical model that supports this approach is also provided here in its essential aspects.

Theoretical Model

The allographic shape variations reflecting the character forms engrained in the motor memory of the writer allows for very effective writer identification and verification. Schomaker proposed a theoretical model and provided an experimental evaluation for this allograph-level approach to writer identification [8]. The main aspects of this model are presented below.

Assume there exists a finite list S of allographs for a given alphabet L. Each allograph S_{li} is considered to be the i^{th} allowable shape (style) variation of a letter $l \in \mathrm{L}$ which should in principle be legible at the receiving end of the writer-reader communication line [13]. The source of allographic variation may be located in teaching methods and individual preferences. The human writer is thus considered to be a pattern generator, stochastically selecting each allograph shape S_{li} when a letter l is about to be written [20]. It is assumed that the probability density function $p_w(S)$, i.e., the probability of allographs being emitted by writer w, will be informative in the identification of writer w if it holds that

$$w \neq v \Rightarrow p_w(S) \neq p_v(S) \qquad (7)$$

where w and v denote writers, S is a common allograph codebook and $p(.)$ represents the discrete PDF for allograph emission. This (i.e. Eq. (7)) will be realizable if for handwritten samples u emitted by w and characterized by

$$\bar{x}_{wu} = p_{wu}(S) \qquad (8)$$

and assuming that the sample u is representative

$$\bar{x}_{wu} \approx p_w(S) \qquad (9)$$

it holds that $\forall a, b, c, w, v \neq w$:

$$\Delta(\vec{x}_{wa}, \vec{x}_{wb}) < \Delta(\vec{x}_{wa}, \vec{x}_{vb}) \qquad (10)$$

where Δ is an appropriate distance function on PDFs $\bar{x}$; v and w denote writers, as before, and a, b, c are handwriting-sample identifiers. Equation (10) states that, in feature space, the distance between any two samples of the same writer is smaller than the distance between any two samples by different writers. In ideal circumstances, this relation would always hold, leading to perfect writer identification. Note that in this model (Eq. (7)), the implication is unidirectional: in case of forged handwriting, $p_w(S)$ does not equal $p_v(S)$ but writer w poses as v ($w = v$).

A problem at this point is that an exhaustive list S of allographs for a particular script and alphabet is difficult to obtain in order to implement this stochastic allograph emission model. Clustering of character shapes with a known letter label is possible and has been realized [21]. However, the amount of handwritten image data for which no character ground truth exists vastly exceeds the size of commercial and academic training sets which are labeled at the level of individual characters. At this point in time, a commonly accepted list of handwritten allographs (and their commonly accepted names, e.g., in Latin, such as in the classification of species in the field of biology) does not exist, as yet. In this respect, it is noteworthy that for machine-print fonts, with their minute shape differences in comparison to handwriting variation, named font categories exist (e.g., Times-Roman, Helvetica, etc.), whereas we do not use generally agreed names for handwritten character families.

Therefore, it would be conducive to use an approach which avoids expensive character labeling at both training and operational stages. Contrary to character segmentation in handwriting, connected components can be detected reliably and in a non-parametric manner. The question then, is whether such sub-allographic or supra-allographic text fragments might be usable for writer identification.

If each allograph S_{li} is composed of a non-empty set of connected components c_j, i.e., $S_{li} = \{c_1, c_2, \ldots, c_m\}$, then let us assume that a finite set or codebook C of connected components for all possible allographs can be estimated. We will use the empirical distribution of connected component occurrence as an approximation for the writer-specific allograph-emission probability. We will assume a finite set or codebook C of connected component and we will estimate and use $p_w(C)$ as the writer descriptor in the identification tests. In the next sections, we will describe the construction of the connected component codebook C, the computation of an estimate of the writer-specific pattern-emission PDF $p_w(C)$, and an appropriate distance function for these PDFs.

A potential concern is the phenomenon of touching characters, but this phenomenon is hardly found in Ethiopic handwriting as compared to western handwritings.

Connected-Component Segmentation and Contour Tracing

In English and Arabic handwriting where free-style cursive handwriting is common, connected-components may encompass several characters or syllables. A segmentation method that isolates individual characters remains an elusive goal for handwriting research. But in Ethiopic handwriting the characters are written separately.

The allographs are extracted as connected components, followed by size normalization to 30×30 pixel bitmaps, preserving the aspect ratio of the original pattern.

The sample images were processed in order to extract the connected components representing the handwritten ink. For each connected component using MATLAB function 'bwlabel'. The contour of the allograph was computed using Moore's algorithm, starting at the left-most pixel in a counter-clockwise fashion. The resulting contour coordinate sequence was resampled to contain 200 (X, Y) coordinates pairs. The resulting fixed dimensional ($N = 400$) vector will be dubbed Connected-Component Contour (CO3).

Connected-Component Codebook Generation

As indicated before, hand written samples have been collected from 25 writers. The connected-component have been extracted from each of these 50 samples using MATLAB function 'bwlabel' yielding a training set containing a total of 32182 patterns (normalized bitmaps).

In this work k-means clustering method is used to generate connected component codebook for Ethiopic characters. The size of the codebook (the number of clusters used) yielding optimal performance is an important parameter in this method. In the experiments, a codebook of size 500 clusters has been used.

Allograph-usage PDF Computation

One bin is allocated to every connected component in the codebook and a shape occurrence histogram is computed for every handwritten sample. For every connected component extracted from a sample after segmentation, the nearest codebook connected component g is found using Euclidean distance and this occurrence is counted into the corresponding histogram bin. The histogram is normalized to a PDF $p(g)$ that acts as the writer descriptor used for identification and verification.

EXPRIMENTAL RESULTS AND DISCUSSIONS

For evaluation of the proposed methods, 25 participants were selected randomly and each was asked to copy out a desired text in his/her natural handwriting on two A4 size pages. Handwritten documents are digitally scanned at 200 dpi resolution and processed using the two writer identification methods.

Since the number of writers in a realistic problem is very large, use of techniques such as the support-vector machine (SVM) or multilayer perceptron (MLP) is not trivial in the writer identification problem. In this thesis work a simple classifier with low computational cost: χ^2 distance measure is used. The features that are extracted for unknown input text are compared with the features of known writers. The writer that his/her features have minimum distance from features of unknown input text is considered as identity of unknown input text. The χ^2 distance measure is defined as follow:

$$x_{ij}^{\;2} = \sum_{k-1}^{n} \frac{(f_{ki} - m_{kj})^2}{(f_{ki} + m_{kj})} \tag{11}$$

where f_{ki} is the k^{th} feature of the unknown input text i and m_{kj} is the mean value of the k^{th} feature of writer j, that are computed from training blocks of writer j. The advantage of using the χ^2 distance measure is that differences for small features are weighed more importantly than weighted Euclidean distance.

Results obtained using Gabor Energy Feature

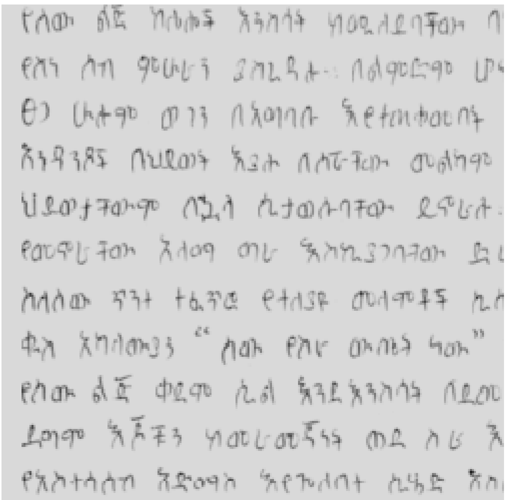
Prior to texture analysis, handwriting documents need to be preprocessed. For this purpose, handwriting images are normalized with respect to different word spacing, line spacing, etc and finally each image is made to be 512 x 512 pixels. In order to decrease computational cost each preprocessed image is divided into four non-overlapping blocks each with size 256 x 256 pixels, and four blocks from the first preprocessed image is considered for training and four blocks from the second preprocessed image is used for the testing from each writer.

The performance of the described methods for feature extraction was investigated and results are given in the following.

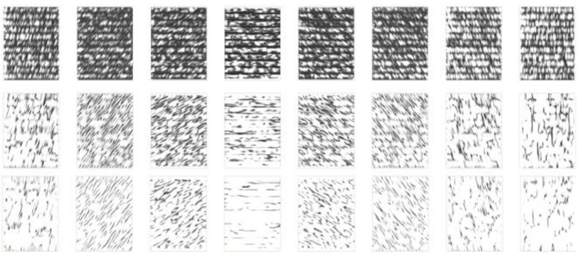
As sample, preprocessed block image and its filtered images using eight orientation and three special frequencies are show in Fig. 5. The parameters used in the experiment are shown in Table 1.

Table 1: Parameters used in the experiment

Parameter	symbol	Values
Orientation	θ	$\left\{\frac{\pi}{8},\frac{2\pi}{8},\frac{3\pi}{8},\frac{4\pi}{8},\frac{5\pi}{8},\frac{6\pi}{8},\frac{7\pi}{8},\right\}$
Wavelength	λ	{5.4,8.2,10.8}
Aspect Ratio	γ	1



(*a*)



(*b*)

Figure 5 Sample block image (*a*) and its Gabor filtered images (*b*)

After the block image is filtered using Gabor filter, Gabor energy operation is performed on it. Features are extracted calculating mean and standard deviation of the black pixel distribution. Thus, 48 feature values are obtained for each block image.

Feature classification is done by using χ^2 distance measure discussed above and an identification rate of 93% in hit list of size 3 and 76% in hit list of size 1 is achieved using all the 48 features values.

Results Obtained using Connected Component Contour

Before feature is extracted, connected components are segmented using connected component labeling function ‘bwlabel’, which is a built in MATLAB tool. After connected components are segmented, their contour is traced using Moore neighbor algorithm.

The Code Book is generated by clustering connected component extracted from 50 samples collected from the 25 writers, which are a total of 32182. To start the clustering work the first 500 centriods or connected components is used from the 32182 extracted connected components.

It is also possible to vary or increase the cluster size but due to limitation on the pc used for experimentation (Intel Pentium - processor speed 1.70 GHz, RAM size 2 GB and RAM speed of 593 MHz) cluster size of 500 is used. On this pc increasing the cluster size more than this makes the clustering work slow.

Feature is extracted using PDF as discussed in the previous section and each handwritten image will have 500 feature values.

As experimental result shows, an identification rate of 96% in hit list 3 and 92% in hit list 1 is achieved.

CONCLUSIONS AND FUTURE WORKS

In this thesis work, hand writing identification system for Ethiopic handwriting has been developed. The technique to identify writers was χ^2 distance measure. The handwriting samples were taken under normal environment and the writers were made to write at their normal writing speed.

For feature extraction, two approaches are followed: Gabor Energy and Connected Component Contours. For the first approach image-processing techniques are implemented in MATLAB to make texture analysis. After extracting the features, χ^2 distance measure method is used for identification in both approaches.

Experimental results demonstrate that, 93% correct identification in a hit list with size 3 and 76% in a hit list with size 1 by using Gabor energy features. While using connected component contours, 96% correct identification in hit list 3 and 92% in hit list 1 are acquired.

Although no research has been reported on Ethiopic handwriting for comparison, the experimental results showed that acceptable identification systems are achieved. Further improvements can be investigated according to the following suggestions:

1. It is possible to achieve better performance with improvement in preprocessing stages.
2. It can be possible to get a better result by fusing the two approaches seen in this paper.

REFERENCE

[1] Plamondon, R. and Lorette, G. "*Automatic Signature Verification and Writer Identification - the state of the art*". Pattern Recognition, Vol. 22, No. 2, pp. 107-131, 1989.

[2] Srihari, S. N., Arora, H., Cha, S. H. and Lee, S. "*Individuality of Handwriting,*" Journal of Forensic Sciences, Vol. 47, No. 4, pp. 1-17, 2002.

[3] Zhang, B., Srihari, S. and Lee, S. "*Individuality of Handwritten Characters*", Seventh International Conference on Document Analysis and Recognition (ICDAR'03), Vol. 2, pp. 1086-1090, 2003.

[4] Zois, E. N. and Anastassopoulos, V. "*Morphological Waveform Coding for Writer Identification*", Pattern Recognition, Vol. 33, No. 3, pp. 385-398, 2000.

[5] Bensefia, A., Paquet, T. and Heutte, L. "*A Writer Identification and Verification System*", Pattern Recognition Letters, Vol. 26, No. 13, pp. 2080-2092, 2005.

[6] Said, H., Tan, T. and Baker, K. "*Personal Identification Based on Handwriting*", Pattern Recognition, 14th International Conference on Pattern Recognition (ICPR'98), Vol. 2, pp. 1761, 1998.

[7] Schlapbach, A. and Bunke, H. "*Using HMM-based Recognizers for Writer Identification and Verification*", In Proceedings of 9th IWFHR, pp. 167-172, 2004.

[8] Schomaker, L. and Bulacu, M. "*Automatic Writer Identification Using Connected Component Contours and Edge-based Features of Uppercase Western Script*", IEEE Transactions on PAMI, Vol. 26, No. 6, pp. 787-798, 2004.

[9] Srihari, S. N., Arora, H., Cha, S. H. and Lee, S. "*Individuality of Handwriting: a Validation Study*," Proceedings of 6th IEEE Conference on Document Analysis and Recognition, pp. 106-109, 2001.

[10] Bulacu, M., Schomaker, L. and Vuurpijl, L. "*Writer Identification Using Edge-based Directional Features*," Seventh International Conference on Document Analysis and Recognition, pp. 937-941, 2003.

[11] Alwiya, S. Omar, "*African Languages*", Microsoft Encarta Reference Library 2004, Microsoft Corporation.

[12] Messay-Hailemariam: "*Handwritten Amharic Character Recognition: The Case of Postal Addresses*", (Masters Thesis). Addis Ababa University, School of Information Studies for Africa, Addis Ababa University, 2003.

[13] Rao, A. R. and Lohse, G. L. "Identifying High Level Features of Texture Perception," *CVGIP: Graphical Models and Image Processing*, Vol. 55, Issue 3, pp. 5218-5233, 1993.

[14] Otsu, N. "*A Threshold Selection Method for Gray-level Histogram*," IEEE Trans. On Systems, Man and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.

[15] Ha, J., Haralik, R.M. and Philips, I. T. "*Document Image Decomposition by the Bounding-box Projection Technique*", In ICDR, pp. 1119-1122, 1995.

[16] Pollen, D. A. and Ronner, S. F. "*Visual Cortical Neurons as Localized Spatial Frequency Filters*", IEEE Transactions on Systems, Man and Cybernetics, Vol. 13, No. 5, pp. 907- 916, 1983.

[17] Hubel, D. H. and Wiesel, T. N. "*Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas* (18 and 19) of the Cat," J. of Neurophysiology, Vol. 28, pp. 229-289, 1965.

[18] Kruizinga, P. and Petkov, N. "*Nonlinear Operator for Oriented Texture*," IEEE Transaction on Image Processing, Vol. 8, No. 10, pp. 1395 – 1407, 1999.

[19] Kruizinga, P., Petkov, N. and Grigorescu, S. E." *Comparison of Texture Features Based on Gabor Filters*", Institute of Mathematics and Computing Science, University of Groningen, The Netherlands, 1999.

[20] Shannon, C. E. "*A Mathematical Theory of Communication*," Bell System Technical Journal, Vol. 27, No. 3, pp. 379-423, 1948.

[21] Vuurpijl, L. G. and Schomaker, L. R. B. "*Coarse Writing-style Clustering Based on Simple Stroke-related Features*", In A.C. Downton, & S. Impedovo (Eds.), Progress in Handwriting Recognition (pp. 29-34). London: World Scientific, 1997.