International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-2 Issue-2, May 2013

# Twelve Directional Feature Extraction for Handwritten English Character Recognition

**Sumedha B. Hallale, Geeta D. Salunke**

***Abstract*— Directional features have been successfully used for the recognition of both machine printed as well as handwritten characters. Selection of feature extraction method is probably the single most important factor in achieving high performance in pattern recognition. In this paper, twelve directional features are used for the recognition of handwritten English alphabets and numerals. The properties of similarity measure are analysed with directional pattern matching. Then the comparison is made between recognition rate of conventional and twelve directional feature extraction techniques. The experiment shows that directional feature extraction techniques are better than conventional one.***

***Index Terms*—Feature extraction, Pattern recognition, Directional pattern matching, Recognition rate.**

## I. INTRODUCTION

Nowadays, recognition systems are used in many fields that have different nature. The optical character recognition (OCR) was started from the recognition of machine printed digits and characters and then it was developed to the recognition of machine printed words. Gradually, handwritten digit, character and word recognition were introduced into this domain. Several research works have been focusing towards evolving the newer techniques that would reduce the preprocessing time and to provide higher recognition accuracy.

Recognition of handwritten character is one of the most interesting topics in pattern recognition. Applications of OCR is in different area especially digit recognition, which deals with postal mail sorting, bank check processing, form data entry, vehicle plate recognition, postal address block detection and recognition ,camera OCR etc..

The handwritten character recognition system is classified as on-line system and off-line system. The probability of recognizing handwriting recorded with digitizer as a time sequence of pen co-ordinates is known as on-line character recognition. In off-line system, the writing is usually captured optically by a scanner and is available as an image. It involves automatic conversion of text in an image into letter codes that are useable within computer.

The handwritten character recognition system has following stages.

**Revised Manuscript Received on 30 May 2013.**
* Correspondence Author
**Sumedha B. Hallale**, Electronics and Telecommunication Department, Pune University/ Genba Sopanrao Moze College of Engineering, Balewadi /Pune, India.
**Prof. Geeta Salunke,** Electronics and Telecommunication Department, Pune University/ Genba Sopanrao Moze College of Engineering, Balewadi /Pune, India.



- Image acquisition
- Pre-processing
- Feature extraction
- Classification

In this paper the features of directions of pixels of the characters with respect to their neighboring pixels are extracted with the help of gradient values. The direction has been divided into 12 regions with each region covering angle of 30 degree, hence direction value of any pixel may have only 12 values assigned from 1 to 12. This approach increases the information content and gives better recognition rate with reduced recognition time [1].

An overview of the paper is as follows: The section II covers the related work to this paper. In section III, a Character modeling, Skeletonization and Normalization are discussed. Section IV covers feature extraction techniques for handwritten character recognition. In Section V, Experimental results and analysis are provided. Finally, the conclusion and future scope have been offered in Section VI.

## II. LITERATURE REVIEW

Dayashankar Singh [1] presented a new feature extraction technique to calculate twelve directional features with the help of gradient values. Experimental result showed that twelve directional feature extraction techniques provide better results in terms of recognition accuracy, training time and classification time.

Cheng-Lin Liu [2] represented selected feature extraction methods in off-line handwritten Chinese character recognition. The experimental results showed, among the 8-directional, 12-directional and 16-directional feature extraction techniques, 12-directional method has better trade off between accuracy and complexity.

Bindu S. Moni, G. Raju [3] implemented a fixed meshing for the off-line recognition of handwritten isolated Malayalam characters. The 12-directional features are extracted to form a feature vector. Classification has been carried out by implementing the Quadratic discriminant Function and Modified Quadratic discriminant Function.

M. Amrouch, Y. es-saady et al [4] proposed a global approach to the recognition of handwritten Amazigh characters using directional features and Hidden Markov Model. Feature vector is extracted from an image of a character using sliding window technique based on Hough transform and translated into sequence of observations. Finally the class of the character is determined by Viterbi classifier. Hiromachi Fujisawa [5] presented the advances of directional features for Kanji character recognition and experimentally analyses similarity measure of directional pattern matching.


*Published By:*
*Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)*

The properties of similarity measure are also discussed and simulation experiments of directional pattern matching are conducted to validate the properties.

M. Ziaratban [6] presented an approach for character recognition termed as template matching. This technique extracts feature by searching the special templates in input images. For each template, the amount of matching is used as feature and position of the best matching in an image is found and saved..

## III. IMAGE PREPROCESSING

The raw data of handwritten characters, no matter how it is acquired, will be subjected to a number of preprocessing steps to make it useable. The preprocessing phase aims to extract the relevant textual parts and prepares them for segmentation and recognition. The main objectives of preprocessing are

- Noise removing
- Skeletonization
- Normalization

### *A. Noise removing*

Image noise is random variation of brightness or colour information in images, and is usually an aspect of electronic noise. It can be produced by the sensor and circuitry of a scanner or digital camera. Image noise is an undesirable by-product of image capture that adds spurious and extraneous information. The noise, introduced by the optical scanning device or the Writing instrument, causes disconnected line segments, bumps and gaps in lines, filled loops etc.. Prior to the character recognition, it is necessary to eliminate these imperfections. The noise reduction techniques can be categorized in two major groups as filtering, morphological operations.

### *B. Skeletonization*

The aim of the skeletonization is to extract a region-based shape feature representing the general form of an object. The skeleton of a shape is a thin version of that shape that is equidistant to its boundaries as in Fig. 1. The skeleton usually emphasizes geometrical and topological properties of the shape, such as its connectivity, topology, length, direction, and width.

**Fig.1. A shape and its skeleton**

### *C. Normalization*

In image processing, normalization is a process that changes the range of pixel intensity values. Normalization is sometimes called contrast stretching.It is the process of converting the random sized image into standard sized image. This size normalization avoids inter class variation among characters. Bilinear, Bi-cubic interpolation techniques are a few methods for size normalization [4].

## IV. FEATURE EXTRACTION

This is also called as data extraction & gives data from perspective areas. Features are a set of numbers that capture the salient characteristics of the segmented image.

The feature extraction methods for handwritten character recognition are based on two types of features: statistical and structural. The statistical features are derived from the statistical distributions of pixels, such as zoning, moments, projection histograms or direction histograms. Structural features are based on the topological and geometrical properties of the character, like strokes and their directions, end-points or intersection of segments and loops.

### *A. Conventional Feature Extraction*

In conventional method, if a line passes through a pixel (foreground), then corresponding pixel will be given value one (1), otherwise (background) it is taken as zero (0). In this method the numbers of inputs are greater which requires more time to classify the characters.

### B. *Directional Features Extraction*

Mathematically, the gradient of a two-variable function, here the image intensity function is at each image point, a 2D vector with the components given by the derivatives in the horizontal and vertical directions. At each image point, the gradient vector points in the direction of largest possible intensity increase, and the length of the gradient vector corresponds to the rate of change in that direction. This implies that the result of the Sobel operator at an image point which is in a region of constant image intensity is a zero vector and at a point on an edge is a vector which points across the edge, from darker to brighter values.

To extract gradient feature, 3×3 Sobel operators are used. It uses two templates to compute the gradient components in horizontal and vertical directions, respectively.

| | | |
|---|---|---|
| -1 | 0 | +1 |
| -2 | 0 | +2 |
| -1 | 0 | +1 |

Gx

| | | |
|---|---|---|
| +1 | +2 | +1 |
| 0 | 0 | 0 |
| -1 | -2 | -1 |

Gy

**Fig.2. Sobel operator templates**

The templates are shown in Fig. 2 and two gradient components at location (i, j) are calculated by:

$$g_v(i,j) = f(i-1,j+1) + 2f(i,j+1) + f(i+1,j+1) - f(i-1,j-1) - 2f(i,j-1) - f(i+1,j-i) \quad (1)$$

$$g_h(i,j) = f(i-1,j-1) + 2f(i-1,j) + f(i-1,j+1) - f(i+1,j-1) - 2f(i+1,j) - f(i+1,j+1) \quad (2)$$



Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

$$grad = \arctan[g_h(i,j)/g_v(i,j)] \quad (3)$$

A gradient vector of arbitrary direction is decomposed into two components coincident with the two neighboring standard directions by specifying a number of standard directions. The components are assigned to the corresponding direction planes and on decomposing all the gradient vectors, a number of feature values are extracted from each direction plane. Conventionally, the gradient is computed on each pixel of the normalized image [7].

After computing the gradient of each pixel of the character, the gradient values are mapped onto 12 direction values to the angle span of 30 degree between any two adjacent direction values. The orientations of these 12 directional values are shown in Fig. 3. [1].

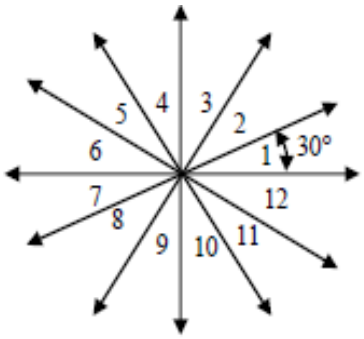


**Fig.3. Orientation of 12 directions**

To calculate directional distribution values of background pixels for each foreground pixel, we have used the masks for each direction. The pixel at center is foreground pixel under consideration to calculate directional distribution values of background. The weight for each direction is computed by using specific mask in particular direction [8]. The mapping of gradient values on 12 directional values can be given as in Fig. 4 below [1].

| Gradient values (g) | Direction |
|---|---|
| g = -1 | 0 |
| $0 \le g \le \Pi/6$ | 1 |
| $\Pi/6 < g \le \Pi/3$ | 2 |
| $\Pi/3 < g \le \Pi/2$ | 3 |
| $\Pi/2 < g \le 2\Pi/3$ | 4 |
| $2\Pi/3 < g \le 5\Pi/6$ | 5 |
| $5\Pi/6 < g \le \Pi$ | 6 |
| $\Pi < g \le 7\Pi/6$ | 7 |
| $7\Pi/6 < g \le 4\Pi/3$ | 8 |
| $4\Pi/3 < g \le 3\Pi/2$ | 9 |
| $3\Pi/2 < g \le 5\Pi/3$ | 10 |
| $5\Pi/3 < g \le 11\Pi/6$ | 11 |
| $11\Pi/6 < g \le 2\Pi$ | 12 |

**Fig.4. Mapping of gradient on 12 directions**

## V. EXPERIMENTAL RESULTS

In implementation, we have used two feature extraction techniques. For each technique and for each class we created feature matrix for 500 training images. Feature vector is obtained by taking a mean of feature matrix for each class. Then we took a testing image for which we extracted the features, say testing feature vector. The similarity between testing feature vector and feature vector of all the classes is calculated as below:

$$Simi(X_1, X_2) = \frac{< X_1, X_2 >}{\| X_1 \| . \| X_2 \|} \quad (4)$$

Representing two patterns in feature vectors $X_1$ and $X_2$, the similarity is calculated as per the correlation coefficient. The testing image belongs to the class which has the highest similarity. Similarly we carried this experiment for 200 testing images and 500 training images on each feature extraction technique to obtain recognition rate.

The following table I. shows the comparison between conventional and twelve directional feature extraction techniques with respect to their recognition rate.

**TABLE I. COMPARISON BETWEEN FEATURE EXTRACTION TECHNIQUES**

| Method | Conventional Method | 12 Directional Features |
|---|---|---|
| Class | | |
| A | 82.5 | 87.2 |
| B | 77.5 | 83.4 |
| C | 80.1 | 85.1 |
| D | 87.7 | 90.0 |
| E | 91.4 | 93.1 |
| F | 82.4 | 88.2 |
| G | 88.2 | 90.7 |
| H | 73.5 | 85.8 |
| I | 72.7 | 84.9 |
| J | 88.0 | 88.3 |
| K | 78.5 | 83.5 |
| L | 83.5 | 91.4 |
| M | 77.5 | 88.0 |
| N | 80.5 | 85.5 |
| O | 86.0 | 90.0 |
| P | 90.5 | 93.5 |
| Q | 85.0 | 88.9 |
| R | 81.0 | 86.4 |
| S | 85.0 | 90.8 |
| T | 77.5 | 86.7 |
| U | 81.0 | 85.7 |
| V | 86.5 | 89.6 |
| W | 90.5 | 91.4 |
| X | 86.1 | 93.4 |
| Y | 78.0 | 88.1 |
| Z | 82.0 | 87.3 |
| 0 | 75.4 | 87.5 |
| 1 | 77.0 | 85.3 |
| 2 | 85.0 | 88.7 |
| 3 | 85.7 | 89.4 |
| 4 | 87.0 | 88.7 |
| 5 | 87.4 | 91.5 |
| 6 | 86.0 | 90.2 |
| 7 | 82.0 | 88.7 |
| 8 | 81.5 | 83.5 |
| 9 | 83.2 | 88.0 |
| Average recognition rate (%) | 82.87 | 88.29 |



Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

## VI. CONCLUSION AND FUTURE SCOPE

In this paper an analysis has been carried out to compare the recognition accuracy of newly developed feature extraction technique with the conventional input pixel technique by using directional pattern matching. The recognition rates of conventional and 12 directional techniques are 82.87% and 88.29% respectively. Experimental result shows that the new approach provides better results for English alphabet characters and 0 to 9 digits as compared to other technique in terms of recognition accuracy, with respect to mean square values. In future, efforts can be made to improve the recognition accuracy of the characters by using more training samples and by making advancements in 12 directional feature extraction systems.

## REFERENCES


1. Dayashankar Singh, Sanjay Kr. Singh, Dr. (Mrs.) Maitreyee Dutta, "Hand written character recognition using twelve directional feature input and neural network", ©2010 International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 3.
2. Cheng-Lin Liu "Handwritten Chinese character recognition: effects of shape normalization and feature extraction", inria 00120408, Dec 2006.
3. Bindu S Moni,G Raju "Modified quadratic classifier and directional features for handwritten Malayalam character recognition", Computational Science-New Dimensions and Perspectives, NCCSE, 2011.
4. M. Amrouch, Y. es-saady, "Handwritten Amazigh character recognition system based on continuous HMMs and directional features", IJMER, vol. 2, issue 2, 2012, pp-436-411.
5. Hiromachi Fujisawa, Cheng-Lin Liu, "Directional pattern matching for character recognition revisited", proceedings of the ICDAR 2003, 0-7695-19601/03©2003 IEEE.
6. M. Ziaratban, K. Faez, F. Faradji, "Language-based feature extraction using template-matching in Farsi/Arabic handwritten numeral recognition", Ninth International Conference on Document Analysis and Recognition, pp. 297 - 301, 2007.
7. Cheng-Lin Liu, "Normalization-cooperated gradient feature extraction for handwritten character recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 8, August 2007.
8. Kartar Singh Siddharth, Mahesh Jangid, Renu Dhir, Rajneesh Rani, "Handwritten Gurmukhi character recognition using statistical and background directional distribution features", *International Journal on Computer Science and Engineering (0975-3397)*, Vol. 3 No. 6 June 2011.
9. Weipeng Zhang, Yuan Yan Tang, Yun Xue, "Handwritten character recognition using combined gradient and wavelet feature" ,1-4244-0605-6/06/$20.00 ©2006 IEEE
10. Anita Pal, Dayashankar Singh, "Handwritten English character recognition using neural network", International Journal of Computer Science & Communication Vol. 1, No. 2, July-December 2010, pp. 141-144


### AUTHOR PROFILE

**Sumedha Hallale** received the degree of B.E in Electronics and Telecommunication Engineering from Terna Engineering College, Osmanabad, Maharashtra in 2009 and persuing M.E. in Electronics & Telecommunication Engineering from Genba Sopanrao Moze College of Engineering, Balewadi, Pune.

**Geeta Salunke,** M. E. in Electronics and is assistant professor in the department of Electronics and Telecommunication Engineering at Genba Sopanrao Moze college of Engineering. With over 13 years of experience, she has published papers in national conference. Her main areas of interest include Image Processing and Communication System**.**



*Published By:*
*Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)*