IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 6, Issue 9, April 2018)

Handwritten Character Recognition System using Gabor Filter and SVM Classifier

Sandli Bansal
M. Tech. Scholar
Shrinathji Institute of Technology and Engineering, Nathdwara (India)
sandlibansal@gmail.com

Ms. Komal Paliwal
Assistant Professor
Shrinathji Institute of Technology and Engineering, Nathdwara (India)
erkomalpaliwal.cs@gmail.com

Abstract – **The selection of features is an important step in any pattern recognition system. This selection of features is considered a combinatorial optimization problem and made the object of research in many disciplines. The main objective of the selection of features is to reduce the number of them by eliminating redundant and irrelevant features recognition system. The second objective of this feature selection is also to maintain and/or improve the performance of the classifier used by the recognition system. In this paper, support vector machine (SVM) based approach is proposed to solve this type of problem in the recognition of handwritten character. This work is capable of recognizing handwritten character with the help of morphological operation, edge detection, feature extraction using Gabor filter and support vector machine (SVM) based classifier.**

Keywords –**Edge Detection, Gabor Filter, Morphological Operations, SVM.**

I. INTRODUCTION

Writing in its various forms, printed and handwritten has always been an essential tool in human communication, and also omnipresent in the majority of sectors of its activities. It is used to preserve and archive knowledge. Because of this, man has always developed techniques aimed at its perpetuity throughout the generations. Indeed, with the emergence of new information technologies: electronics and computing, and the new increase in the power of machines, the automation of processing (reading, research and archiving) which are attached to it appears inescapable. It is of concern to researchers in the field of form recognition including automatic recognition of writing [1]. This is how research in this field began over the last few decades. The automatic recognition of writing is a complex computer process that aims to translate printed or handwritten text into a digitally encoded text, thus understandable by a machine, by transmitting to it the ability to read [2] [3]. The recognition of writing concerns more precisely all tasks related to the mass processing of paper documents. It therefore covers large repetitive applications with large databases such as automatic processing of administrative files, automatic sorting of postal mail, reading of amounts and bank checks, processing of postal addresses, Processing of forms, keyboardless interfaces, analysis of written gestures, reading of legacy documents, indexing of library archives and searching for information in databases [4].

The automation of one of these examples is an extremely difficult problem to implement, given the great variability associated with the writers' habits as well as the styles and forms of writing (manuscript, cursive or printed with many fonts) [5] [6]. Indeed, reading activity that is simple for a human is not an obvious task to be copied to a computer. Truly the accomplishment of this task requires that the machine acquires a prior knowledge base of domain and the use of a powerful mathematical formalism. The automatic reading of writing has seen considerable progress especially in the last decade. This is due, on the one hand, to the many works carried out leading to a variety of different approaches and, on the other hand, to the performance of computers and current acquisition systems coupled with modern statistical methods for example hidden markov models, support vector machines and neural networks. In addition, the availability of standard international databases for handwriting and printing enabled and allowed researchers to credibly report on the performance of their approaches in this field, with the possibility of comparing them with others approaches [7].

Despite the efforts and progress made in the field thanks to the many years of research on the subject, we are still far from the dream of a world without paper. Otherwise, there is still no reliable system able to handle natural writing as a whole. Indeed, the results published in the literature show that the

IJDACR
International Journal Of Digital Application & Contemporary Research
International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 6, Issue 9, April 2018)

recognition rates obtained are restricted to very limited fields of application (postal addresses, bank checks) or to very restricted categories of writing representing only one particular aspect of the current and spontaneous writing [8] [9] [10].

Because of this, automatic recognition of handwriting is still an active research topic. Recently, it was the subject of intense research. The number of jobs differs from one type of writing to another. For example, several scientific researches have been carried out on character recognition.

The research we have carried out focuses on the development of methods for the recognition of handwritten characters, taking into account the context by combining levels of analysis and morphological knowledge.

In the field of handwriting recognition, features can be described as a way to distinguish an object (vowels and consonants) of a class of another object (vowels and consonants) of another class. Therefore, it is necessary to define meaningful feature when developing a system of recognition. Features are generally defined by experience or intuition. Several features can be extracted. Representation features used is a vector representation. The vector size may be wide if a large number of features is extracted. It has been observed that some extracted features are irrelevant or redundant in recognition system. The selection of relevant features is against a complex issue and is the subject of much research.

We propose an optimization method in the context of this work for the selection and weighting of features of a system of recognition of isolated handwritten character. The classification method used is based on the theory of support vector machine (SVM). In this work, the optimization of the recognition system is advanced. The first criterion is to extract the features using Gabor filter, while the second criterion should improve or maintain the performance of the recognition system.

II. Proposed Method

The flow diagram of proposed character recognition system is shown by Figure 1. The processes are explained below:

A. Data Collection

Scanned image of alphabet data in .JPEG format is loaded to the system. The system we develop here is capable of identifying the character. So the user has to select the character of its own choice for recognition input. Once the input character is selected, several processes are applied to it.

B. Pre-Processing

The goal of the pre-processing is to facilitate the characterization of the form (vowels and consonants) or entity to recognize either by cleaning the picture of the form or reducing the amount of information to process only keep only the most relevant information. The image cleaning is basically to eliminate residual noise from the binarization. Reducing the amount of information to be processed can be obtained from operations to bring the line thickness to a single pixel or by monitoring feature or from extractors of upper contours, lower and / or interiors.

Note that some forms (vowels and consonants) are inclined or bent so it is necessary to normalize slope this form to segment the form (e.g. segmentation of a word in letters). This standardization is to correct the slope of a word or correct the inclination of the letters in a word to facilitate segmentation.

Segmentation is a phase of pre-processing. Its purpose is to locate and extract as precisely as possible the information to recognize. In this research work the segmentation is done using edge detection and morphological operations.

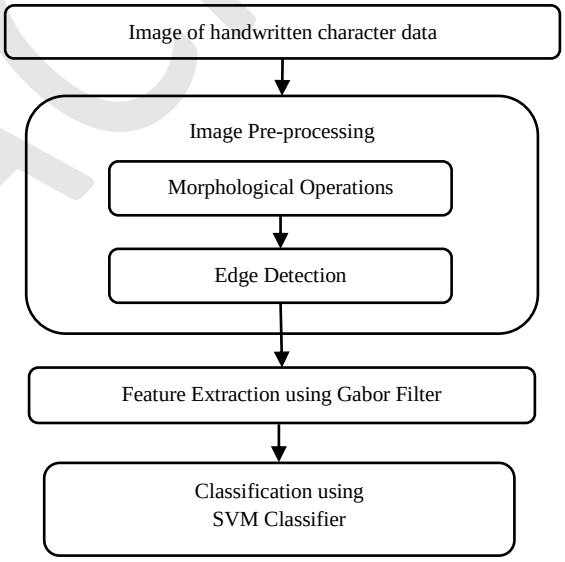

Figure 1: Flow diagram of proposed work

C. Feature Extraction using Gabor Filter

The purpose of feature extraction in the field of recognition is to express the feature in numerical or symbolic form called encoding. Depending on the case, the values of these features can be real, integer or binary. The vector composed of feature n represents a point in the new space of n dimensions. This stage of the recognition is to extract features to describe unequivocally the forms belonging to the same class of characters while differentiating them from other classes.

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 6, Issue 9, April 2018)

However, it is necessary to perform for each recognition problem experimental evaluation of some methods for extracting the most promising features. These experiments will make a wise choice features to extract as often, the use of a single method of feature extraction is not sufficient to achieve good performance of the recognition system.

An image represented in terms of $f(x,y)$ where x and y signifies the coordinates of pixels having size M×N is convoluted in frequency domain. The Fourier transform is applied as the solution for this step:

$$F\{f(x,y)\} = F(u,v) = \int_{-\infty}^{M-1}\int_{-\infty}^{N-1} f(x,y)e^{-j2\pi(ux/M+vy/N)}\,dx\,dy \quad (1)$$

In frequency domain the Gabor feature for an image $f(x,y)$ is the multiplication of convoluted image with Gabor filter bank $\Psi(x,y,\omega_m,\theta_n)$ given by:

$$O_{m,n}(x,y) = F(u,v) * \Psi(u,v,\varpi_m,\theta_n) \quad (2)$$

Where, $*$ is the convolution operator. The filter bank is created using m frequencies and n rotations $G(m \times n)$ that provides features points and is saved in form of vector. These feature vectors are classified using the support vector machine classifier.

D. Classification using Support Vector Machine

The classification is developing a decision rule that transforms attributes characterizing the forms in class membership (transition from code space to space-making) [11]. Before a decision model is integrated in a handwriting recognition system, you must have also previously two steps: the learning step and the test step.

As part of our project, support vector machine (SVM) is the method of classification of handwritten character recognition system which is described in the following heading.

Consider the training set $\{x_1, y_1\}, \ldots, \{x_\ell, y_\ell\}$, where $x \in X$ and $y \in \{-1, 1\}$, where ℓ is the number of observations and X is a distribution in space $\Re^n$. In the classification problem, the goal is to find an efficient method to construct the optimal separator hyperplane, i.e., with the greatest margin. To do this, one must find the vector w and the constant b, which minimize the norm $|w|^2 = w^T w$ (since it is inversely proportional to the margin), under the constraints:

$$w^T x_i + b \geq 1, \quad if\ y_i = 1 \quad (3)$$

$$w^T x_i + b \leq -1, \quad if\ y_i = -1 \quad (4)$$

Because one can accept some errors, one relaxes the constraints (3) & (4) and introduces an additional cost related to this relaxation, so that one arrives at the quadratic problem, QP, following:

$$\underset{w}{\text{Minimize}} \quad \frac{1}{2}(w^T w) + C\left[\sum_{i=1}^{\ell}\xi_i\right]$$

$$\text{Under the constraints} \quad y_i(w_i^T x + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad i = 1, \ldots, \ell \quad (5)$$

The problem (5) can be solved in the primal space (the space of parameters w and b). In fact, one solves the QP in the dual space, equation (6), (the Lagrange multiplier space) for two main reasons: 1) The constraints (4) and (5) are replaced by the associated Lagrange multipliers, and 2) We obtain a formulation of the problem where the training data appear as an internal product between vectors, which can then be replaced by kernel functions, then construct the hyperplane in the feature space and obtain functions Non-linear in the input space.

$$\underset{\alpha}{\text{Maximize}} \quad L_D(\alpha) = \sum_{i=1}^{\ell}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{\ell}\alpha_i\alpha_j y_i y_j (x_i^T x_j)$$

$$\text{Under the constraints} \quad \sum_{i=1}^{\ell} y_i\alpha_i = 0, \quad 0 \leq \alpha_i \leq C \quad i = 1, \ldots, \ell \quad (6)$$

Where, α_i is the Lagrange multiplier, associated with constraints. Parameter C controls the level of error in the classification.

The SVM evaluation function is defined as:

$$f(x) = \sum_{i=1}^{\ell}\alpha_i y_i k(x_i, x) + b \quad (7)$$

The examples x_i associated with the Lagrange multipliers α_i larger than zero correspond to the support vectors, and have a significant contribution to equation (7). Geometrically, these vectors reside in the margin defined by the separating hyperplane. The constant b represents the threshold of the hyperplane learned in the characteristic space. It can be calculated by the mean of the function (Equation 7), evaluated using the support vectors.

III. Simulation And Results

The performance of proposed algorithms has been studied by means of MATLAB simulation.

Figure 2: Input image

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 6, Issue 9, April 2018)

Figure 3: Gray image

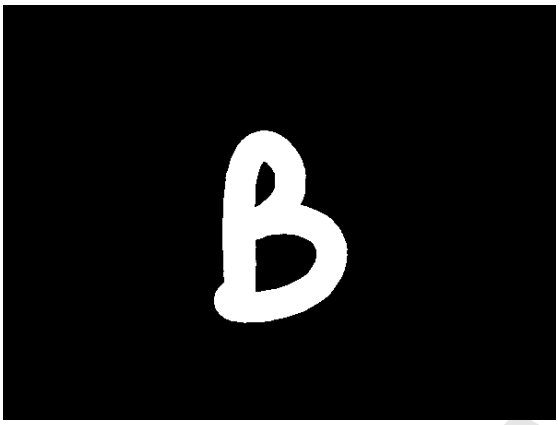

Figure 4: Binary image

Figure 5: Cropped binary image

Figure 6: Resized binary image

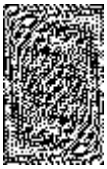

Figure 7: FFT of binary image

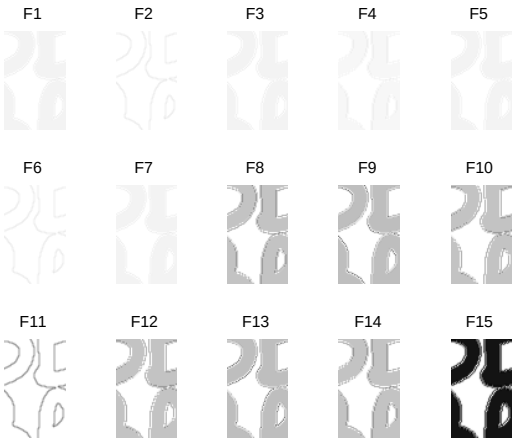

Figure 8: Gabor response

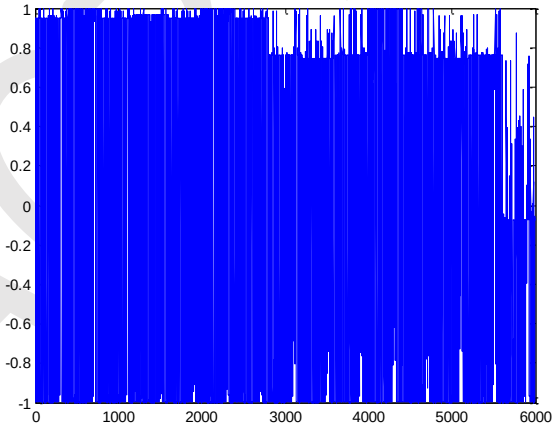

Figure 9: Feature vector of image

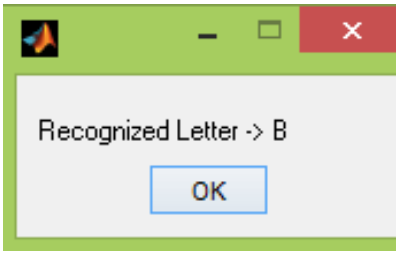

Figure 10: Recognized letter

IV. CONCLUSION

The work presented in this paper addresses the steps necessary to build a system of handwritten character recognition. For each of these stages are: pre-processing, extraction of features and classification, we have tried to propose an optimization method for the selection of features relevant recognition system. Thus, from the extraction step of the feature, the selection of relevant and non-redundant system is performed features. This selection is to reduce inputs of the classifier (SVM) while improving or maintaining the classification recognition rate. Main

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 6, Issue 9, April 2018)

contribution of this paper is the study of the implementation of the morphological operation and edge detection to select features extracted from Gabor filter. The results of this selection is satisfactory.

REFERENCE

[1] Gauri Katiyar, Shabana Mehfuz, "MLPNN Based Handwritten Character Recognition Using Combined Feature Extraction", IEEE, International Conference on Computing, Communication and Automation (ICCCA2015), pp. 1155-1159, July 2015.
[2] Duda, R.O., Hart, P.E. and Stork, D.G., 2012. Pattern classification. John Wiley & Sons.
[3] Jayadevan, R., Kolhe, S.R., Patil, P.M. and Pal, U., 2011. Offline recognition of Devanagari script: A survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41(6), pp.782-796.
[4] Kolman, E. and Margaliot, M., 2008. A new approach to knowledge-based design of recurrent neural networks. IEEE Transactions on Neural Networks, 19(8), pp.1389-1401.
[5] Gaur, A. and Yadav, S., 2015, January. Handwritten Hindi character recognition using k-means clustering and SVM. In Emerging Trends and Technologies in Libraries and Information Services (ETTLIS), 2015 4th International Symposium on (pp. 65-70). IEEE.
[6] Sameeksha Barve, "Artificial Neural Network Based On Optical Character Recognition", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 1, Issue 4, June 2012.
[7] Sameeksha Barve, "Optical Character Recognition Using Artificial Neural Network", International Journal of Advanced Technology & Engineering Research (IJATER), ISSN NO: 2250-3536 Volume 2, Issue 2, May 2012.
[8] Shabana Mehfuz, Gauri katiyar, "Intelligent Systems for Off-Line Handwritten Character Recognition: A Review", International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, Volume 2, Issue 4, April 2012.
[9] Dileep Kumar Patel, Tanmoy Som, Sushil Kumar Yadav, Manoj Kumar Singh, "Handwritten Character Recognition Using Multiresolution Technique and Euclidean Distance Metric", Journal of Signal and Information Processing, PP. 208-214., 2012.
[10] Rakesh Kumar Mandal, N. R. Manna, "Hand Written English Character Recognition using Column-wise Segmentation of Image Matrix (CSIM)", WSEAS Transactions On Computers, E-ISSN: 2224-2872, Issue 5, Volume 11, May 2012.
[11] Richard O. Duda, Peter E. Hart, & David G. Stork., "Pattern Classification", (Second edition). New York: Wiley-Interscience, 2001.