Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 93 (2016) 39645-39649

ISSN: 2229-712X

On Inclusion of hidden View for improved handwritten character recognition

Vijay T. Patil[1] and S. A. Patekar[2]

[1]Department of Computer Engineering, Vidyalankar Polytechnic, Wadala, Mumbai-37.

[2] Department of Computer Engineering, Vidyalankar Institute of Technology, Wadala, Mumbai-37.

ARTICLE INFO

Article history:

Received: 8 July 2013;

Received in revised form:

14 April 2016;

Accepted: 20 April 2016;

Keywords

2D view,

Extra View,

SVM,

NN.

ABSTRACT

The paper proposes Handwritten Character Recognition method using 2D view and Support Vector Machine (SVM). In this all the character images are Pre-processed (includes Normalization and Noise Removal), which are further used for feature extraction using two dimensional (2D) views. From each character four different views (Top, Bottom, Left, and Right) are obtained called as basic views. All basic views are not able to collect the complete information of character image. The hidden information is capture separately called as extra views. From each view 16 features are extracted and combined to obtain 80 features. These features are used to train SVM to separate different classes of characters. Handwritten Character database is used for training and testing of SVM classifier. Support Vector Machine provides a good recognition result for lower case characters and upper case characters are 82%.

1. Introduction

Today, many researches have been done to recognize Handwritten characteristics have got million application because in terms of vital link between machine and then communication. Several algorithms have been proposed by many researchers but finding the efficiency of these algorithms is still in the quest. In this paper our approach is not to refine the algorithm but select the best of the algorithm. In this paper the focus is to find out the complete information of the handwritten character. Many feature selection methods are available but very few talk about the structure of the image and none of them talk about the curved character image. The new 2D view feature extraction method is used to gather data from the character image. From this method we get top, bottom, left and right view of the image. All the views formed are not sufficient to reconstruct the image because data from inner part of image is not collected. So the missing data has to be identified and this is done by removing all the views from the original image, the remaining pixels are considered as extra view of the image. The outcome will be the remaining pixels which are not identified by any one of the basic views.

When these features are used to train SVM, which enhanced good results for lower and upper case characters.

2. Implementation Details

Inputs to the system are the character images which have been processed by different stages like preprocessing, feature extraction and classification. In pre-processing, the character image is normalized to a standard size. Feature extraction process gets the two dimensional (2D) views of each image. These feature data is applied to train SVM in training phase. SVM creates structure with this training data can be called as model. Then test character images can be classified in particular classes, depending on their features, by SVM.

The block diagram of our HCR system is shown in Fig. 1.

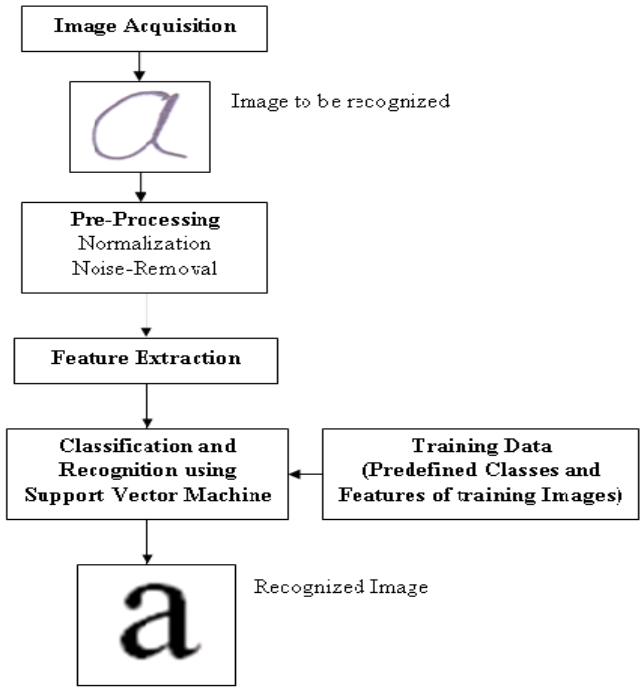


Figure 1: Block diagram of Handwritten Character Recognition System

2.1 Image Acquisition:

We will acquire an image to our system as an input. This image should have a specific format, for example, bmp, jpg format. This image can be acquired through the scanner or, digital camera or other digital input devices.

2.2 Preprocessing:

After acquiring the image, it will be processed through sequence of preprocessing steps to be ready for the next step. In HCR system, typical preprocessing operations include normalization, smoothing and noise reduction of a digital image so that, algorithms and classification can be made simple and more accurate.

Tele:

E-mail address: vijay.patil@vpt.edu.in

Normalization is to regulate the size, position, and shape of character images, so as to reduce the shape variation between the images of same class. In normalization process all images are converted into fixed size format of 64x64 pixels. Then convert the grayscale image into a binary image. The output image replaces all pixels in the input image with luminance greater than level with the value 1 (white) and replaces all other pixels with the value 0 (black).

Median filtering is applied on normalized images for removal of noise from the image. The objective of noise removal is to remove any unwanted bit patterns, which do not have any significance in the output.

Noise removal reducing noise in an image. For on-line there is no noise to eliminate so no need for the noise removal. In off-line mode, the noise may come from the writing style or from the optical device captures the image.

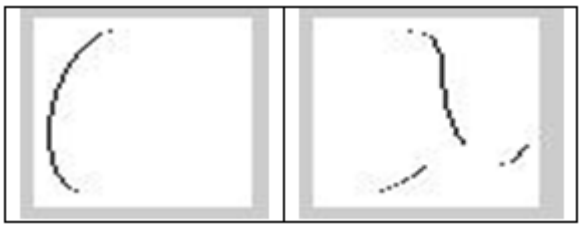
2.3 Feature Extraction:

Feature extraction is an important data gathering step, in which important data of character images are collected. The feature should include all the data with which the original image can be identified. Many available feature selection methods are studies but very few talk about the structure of the image and none of them talk about the curved character image. The new 2D view feature extraction method is applied for HECR system. This method is to get the top, bottom, left, right view of the image. As shown in Figure 2 the views of the image can be found by selecting the pixel from the original image. For selecting the left view start scanning image from top to bottom and each row from left to right. Count all the white pixels until get the first black pixel, then go to next row. Then for selecting the top view start scanning image from left to right and each column from top to bottom. Right view can be achieved by scanning image from top to bottom and each row from right to left. And bottom view achieved by scanning image from left to right, and each column from bottom to top image.

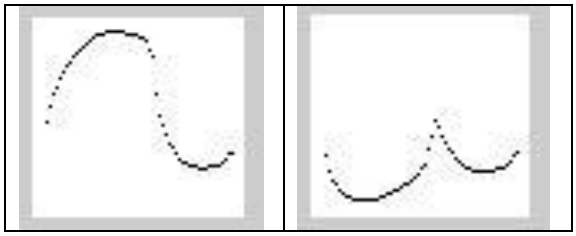
2.3.1 Incorporation of Extra View:

All the views (Top, Bottom, Left and Right view) formed are not sufficient to reconstruct the image because data from inner part of image is not collected. The data which is not identified or not collected in the basic views can be obtained by removing the data of basic views from the preprocessed image. This data is considered as extra view of the image. Incorporation of extra view makes handwritten character recognition more effective and accurate.

Figure 2: (a) 2D direction of character image



(b) (c)

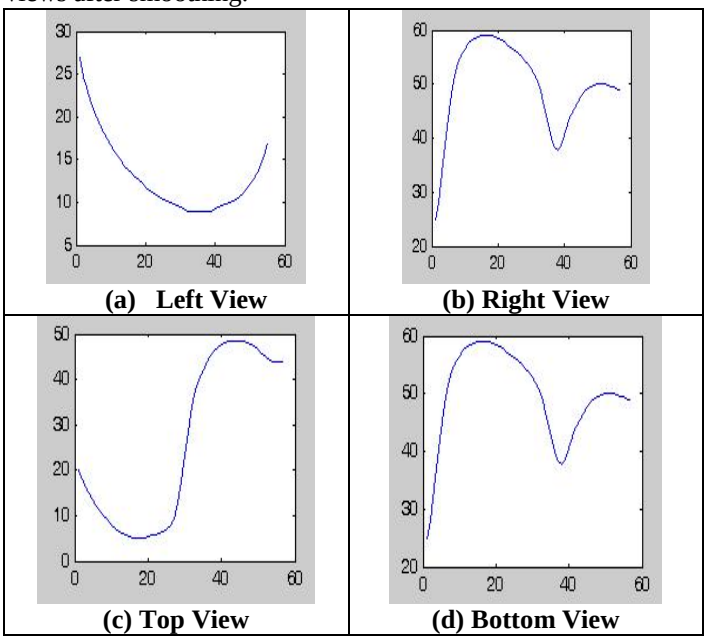


(d) (e)

Figure 2: (b) Left view (c) Right View (d) Top View (e) Bottom View

2.4 Smoothing

We can use a low pass averaging filter to eliminate the noise. The aim of smoothing is to give a general idea of relatively slow changes of value with little attention paid to the close matching of data values, while curve fitting concentrates on achieving as close a match as possible. Fig. 3 shows the views after smoothing.



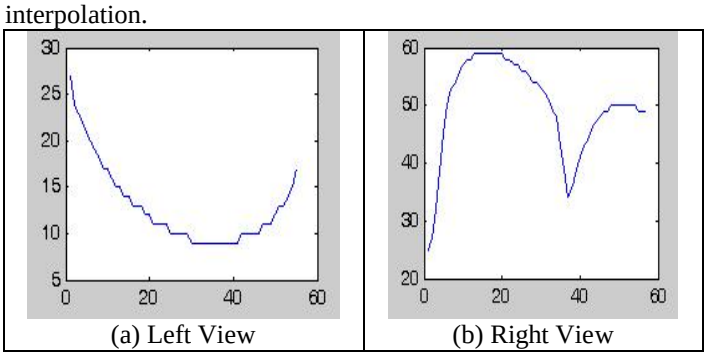
(a) Left View **(b) Right View**

(c) Top View **(d) Bottom View**

Figure 3: Smoothing of views (a) Left (b) Right (c) Top (d) Bottom

2.5 Interpolation:

Next step in feature extraction is Interpolation to find the curve which fit the views perfectly. Interpolation is a method of constructing new data points within the range of a discrete set of known data points. For finding the curve fitting to the points in the views collected from the image the cubic interpolation method is used. This method uses four points to find the fine curve at each point so that the curve will be more close to fit all the points in the image views. Fig. 4 shows the views after interpolation.



(a) Left View (b) Right View

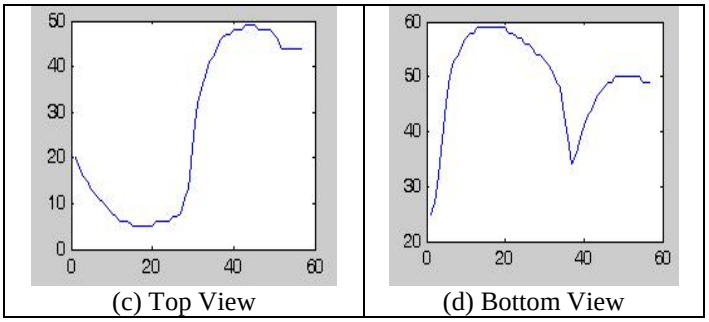


Figure 4: Interpolation of views (a) Left (b) Right (c) Top (d) Bottom

2.6 Sampling:

Sampling is an important aspect of data collection. Figure 5 shows sampling effect on all views of character image. All the views are interpolated and made ready for sampling. This process will collect sample data from the curve so that the curve representation will not change and sampled data can represent the same curve as original one. After interpolation we will get 64 points which are now a curve or can be represented as polynomial. This can be represented as the feature set but further it can be reduced by sampling the data with sampling rate 1/4. After sampling the feature data size of every view is 16 points. Finally each character image will have 80 points. These features are easy to interpret, compute and they have good information about the structure of the character.

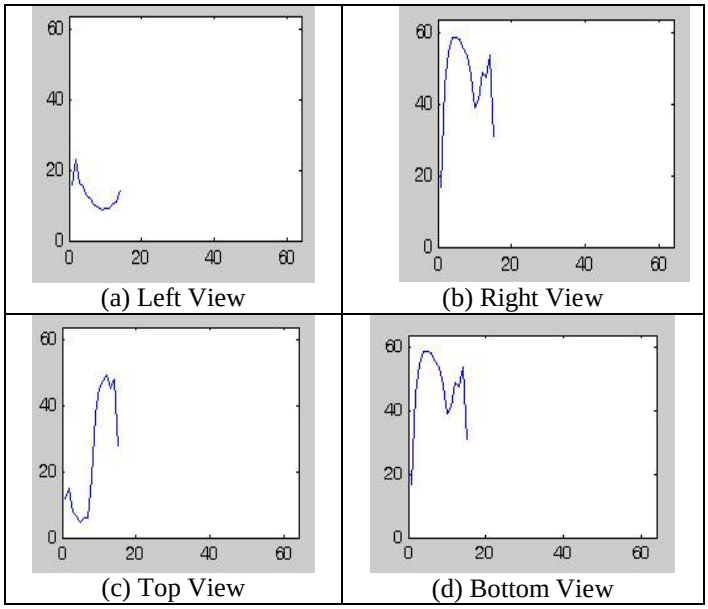


Figure 5: Sampling of views (a) Left (b) Right (c) Top (d) Bottom

2.7 Classification and recognition using Support Vector Machine:

The objective of any machine capable of learning is to achieve good generalization performance, given a finite amount of training data, by striking a balance between the goodness of fit attained on a given training dataset and the ability of the machine to achieve error-free recognition on other datasets.

The Support Vector Machine (SVM) can be characterized as a supervised learning algorithm capable of solving linear and non linear classification problems. The principle of an SVM is to map the input data onto a higher dimensional feature space nonlinearly related to the input space and determine a separating hyper plane with maximum margin between the two classes in the feature space. This results in a nonlinear boundary in the input space. The optimal separating hyper plane can be determined without any computations in the higher dimensional feature space by using kernel functions in the input space. Commonly used kernels include:

Kernel	Function
Linear	$K(x, y) = (x.y)$
Gaussian (Radial basis function)	$K(x, y) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2})$
Polynomial	$K(x, y) = (x.y)^p$
Tangent Hyperbolic	$K(x, y) = \tanh(x.y - \theta)$

3. Experimental Result

After preprocessing and feature extraction the datasets for SVM are created which include 80 feature points of each character and the respective class label. The training dataset is provided to the SVM to create the structure which can be used for testing. Radial Bases Function (RBF) Kernel is used in the experiments for SVM. The overall performance of recognition rate on the test set for the SVM classifier is shown in Table 1 and Table No. 2. Figure 6 shows the character recognition output of single character.

Table 1: SVM classification accuracy for Lower Case Characters (a-z)

	Training Data set	Testing Data set	Accuracy	Classification
Classes	5	5	23/50	46%
Samples for each class	10	10		
Classes	26	26	34/52	65.38%
Samples for each class	8	2		
Classes	26	26	670/832	80%
Samples for each class	60	32		
Classes	26	26	860/1040	82.69%
Samples for each class	70	40		

Table 2: SVM classification accuracy for Upper Case Characters (A-Z)

	Training Data set	Testing Data set	Accuracy	Classification With EV
Classes	5	5	33/50	66%
Samples for each class	15	10		
Classes	10	10	147/260	56.53%
Samples for each class	25	25		
Classes	26	26	319/520	61.34%
Samples for each class	30	20		
Classes	26	26	532/650	81.78%
Samples for each class	40	25		

Table 3: SVM classification accuracy with EV and normal SVM for special character Q and G

	Training Data set	Testing Data set	Classification with normal SVM	Classification With EV
Classes	1	1	58%	66%
Samples for class Q	40	25		
Classes	1	1	68%	82%
Samples for class G	40	25		

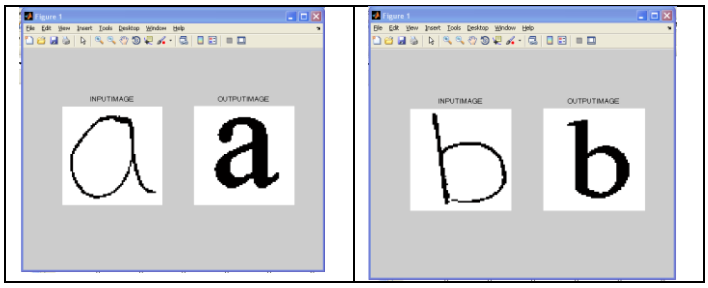


Figure 6: Single Character Recognition result (Lower Case- 'a' and 'b')

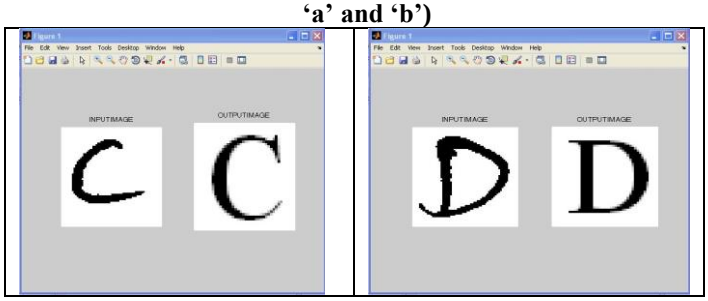


Figure 7: Single Character Recognition result (Upper Case- 'C' and 'D')

4. Conclusion

The focus of this paper is on to find hidden information of character which we called as Extra View. Experiments show that these features have good discrimination ability. For training and test sets, the basic views provide 64 features for each character image, which are not sufficient for representing full information of character image completely. Then E-View added to feature extraction to get all most all the features from each character image, which increased the features of each character image to 80 and accuracy of SVM is increased. It can be concluded that features extracted from the character images are well represent the complete input information of character image; the accuracy of character recognition will increased definitely specially for hidden information of characters like Q and G. SVM have been demonstrated superior classification accuracies to neural classifiers in many experiments. Here the paper represent procedure to extract complete information of handwritten character, together with the SVM improves handwritten character recognition. Experiments shows that the by adding extra view improves the accuracy about 20% than the result of normal views.

5. References

[1] Miguel Po-Hsien Wu, "*Handwritten Character Recognition*", The school of Information Technology and Electrical Engineering, University of Queensland, Oct 29, 2003

[2]Nafiz Arica, "*An Offline Character Recognition system for Free Style Handwriting*", thesis of Department of Computer Engg. The Middle East Technical University, Sep-1998

[3] R. Plamondon and S. N. Srihari, "*Online and Offline handwritten recognition: a comprehensive survey*", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22. No.1, January 2000.

[4] Wikipedia online, http://en.wikipedia.org/wiki

[5] The odoros Evgenuiu and Massimilliano pontil, "*Statistical Learning Theory: a primer*", 1998

[6] Dewi Nasien, Habibollah Haron, "*Support Vector Machine for English Handwritten Character Recognition*", 2010 Second International Conference on Computing Engineering and Applications

[7] Tutorial Slides by Mingyue Tan, "*Support Vector Machine and its Applications*", The university of British Columbia Nov 26,2004

[8]Vikramaditya Jakkula, "*Tutorial on Support Vector Machine*", School of EECS, Washington State University, Pullman 99164.

[9] Burges C.J., "*Tutorial on support vector Machines for pattern Recognition*", Data mining and Knowledge Discovery, 2,1998, kluwer Academic Piblishers, Boston.

[10]Chin-Wei Hsu, Chih-Chung Chang and Chih-Jen Lin, "*A Practical Guide to Support Vector Classification*", Department of Computer Science, National Taiwan University, Taipei 106, Taiwan, May 19, 2009

[11]Steve R. Gunn, "*Support Vector Machines for Classification and Regression*", Technical Report Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, 10 May 1998.

[12]Abdul Rahim Ahmad, Christain Viard-Faudin, "*Online Handwriting Recognition using Support Vector Machine*", Proceedings of Second International Conference on Artificial Intelligence in Engineering and Technology, August 3-5, 2004, Kota Kinabalu, Malaysia

[13]Seong Wook Joo, "*Linear Discriminant Function and SVM*"

[14] Max Welling, "*Support Vector Machines*", Department of Computer Science, University of Toronto, Canada

[15] Henok Girma, "*A Tutorial on Support Vector Machine*", Center of Experimental Mechanics, University of Ljubljana, 2009.

[16] Abdul Rahim Ahmad, Marzuki Khalid, Rubiyah Yusof, "*Machine Learning Using Support Vector Machines*"

[17] http://www.google.com search Kernel Trick, Learning from examples

[18]Sami Gazzah and Najoua Ben Amara, "*Neural Networks and Support Vector Machines Classifiers for Writer Identification using Arabic Script*", The International Arab Journal of Information Technology, Vol. 5, No. 1, January 2008.

[19] H. Swethalakshmi, Anitha Jayaraman, V. Srinivasa Chakravarthy, "*Online Handwritten Character Recognition of Devnagari and Telugu Characters using Support Vector Machines*"

[20] Cheng Lin Liu and Hiromichi Fujisawa, "*Classification and Learning Methods for Character Recognition: Advances and Remaining Problems*", National Laboratory of Pattern Recognition, Instt of Automation, Beijing, China

[21] Dhruv Batra, Girish Singhal and Shantanu Choudhary, " *Gabor Filter based Fingerprint Classification using Support Vector Machines*", IEEE India Annual Conference 2004, INDICON 2004.

[22] Nor Amizam Jusoh, Dr. Jasni Md Zain, "*Comparision between Techniques in Feature Extraction*", Proceeding of International conference on Electrical Engineering and

Informatics, Instt Teknologi bandung, Indonesia June 17-19, 2007.
[23]Oivind Due Trier, Anil Jain, "*Feature Extraction Methods for Character Recognition –A survey*", Pattern Recognition Society, vol 29, No. 4, 1996
[24]Cheng Lin Liu, "*Handwritten Chinese Character Recognition: Effects of Shape Normalization and Feature Extraction*", National Laboratory of Pattern Recognition (NLPR) Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P. R. China.
[25] Dr. P.Santhi Thilagam, Balaji Bombade, "*Handwritten Kannada Character Recognition system using Support Vector Machine*", International conference on Computer Science and Engineering, Karnataka, 2009.