View-Based Word Recognition System

Marek Tabedzki [1], Khalid Saeed [2]
Faculty of Computer Science
Bialystok University of Technology
Wiejska 45A, 15-351 Bialystok, Poland.
[1] e-mail: tabedzki@ii.pb.bialystok.pl, [2] e-mail: aida@ii.pb.bialystok.pl
http://aragorn.pb.bialystok.pl/~zspinfo/

Abstract: *In this paper, a new method for word recognition and classification without segmentation is presented. The worked out algorithm is based on recognizing the whole word without separating it into letters. According to this algorithm, entire words are treated and analyzed as object images subject to classification. The method is based on the "view-based approach" presented in authors' previous works for hand and printed script recognition. The top and bottom views of each word are analyzed, and the characteristic points describing it are created. The procedure of the processing and recognition involves the application of neural networks and minimal eigenvalues of Toeplitz matrices. These methods are used to modify the "view-based" algorithm and improve the efficiency of the identification process. Printed words in Latin alphabet form the data base for the experiments introduced in this work. The obtained results are good and promising.*

***Keywords*:** Word Recognition without segmentation, Minimal Eigenvalues of Toeplitz matrices, Neural Networks.

1. INTRODUCTION

In this paper, we present a new approach for word classification and recognition. Contrary to most popular methods [1, 2] this method does not require stage of segmentation [3]. It is based on recognizing whole words, without dividing them into single letters. Every single word was treated as an image, then analyzed and classified in that manner. We used this method to recognize English names of animals, printed with various fonts in small caps.

Not only the stage of segmentation was omitted in this algorithm. Thinning [4] was also unnecessary, hence we analyzed the word without thinning, that is only the shape of the word is analyzed. The only necessary preprocessing was binarization (i.e. converting scanned image into a "black-and-white" format.)

The main engine of this method is strictly based on a hybrid view-based algorithm. Its essential ideas were presented in our previous works [5, 6]. Previously it was used for recognizing separated Latin letters (both machine-printed and handwritten) and bestows particularly commendable results.

For letter recognition four views of each image were analyzed. In case of words, only two of them (i.e. top and bottom) contain useful (and usable) information, so only these two views were examined. Next, fixed number of uniformly distributed characteristic points were taken of each view, and formed into vector describing tested word. This vector was the base for classification. Novelty introduced in this work, is attachment of values describing ratio (proportion) of analyzed image. This addition allowed to examine the length of the words and include this information in the recognition process.

Method was tested on database of 75 different words – English names of animals. Words were 2 to 12 characters long. Training Set (part of database containing knowledge of our system) was composed of these 75 words, printed with 6 different fonts each (small caps only.) Test Set (part of database used to verify the effectiveness of the method) contains the same 75 words, each printed with over 130 different fonts, both standard serif or sans serif fonts as well as fonts imitating handwriting. Some of them are presented in Fig. 1.

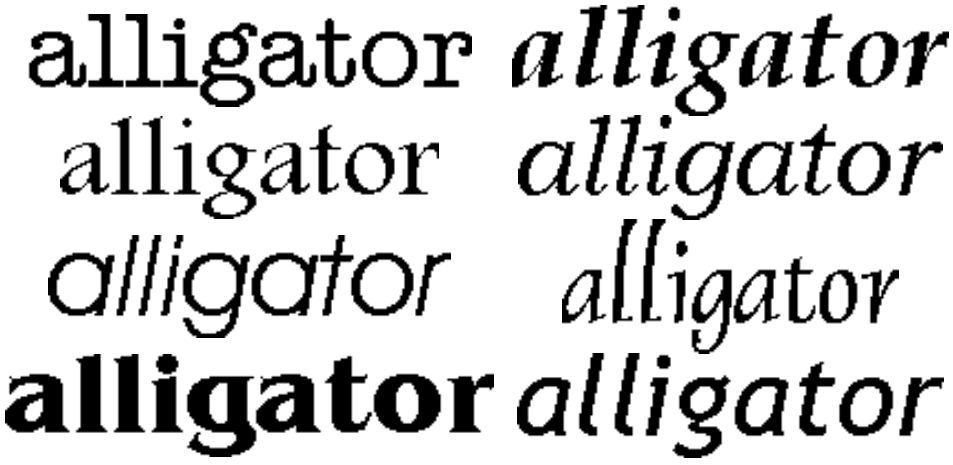

Fig.1 – Database sample

Words were classified with one of two methods. The simplest classification approach was to compare point-to-point characteristic vectors of each word and find in Base Set the most similar vector (the nearest neighbor.) The tested word was categorized to the class of his nearest neighbor. The second approach was classifier based on Artificial Neural Network [7]. In this case Base Set (called Training Set) was used to train neural network. Characteristic vector of the word in question was put on input of the neural network. On output we get class to which given word should be categorized.

2. VIEW-BASED APPROACH

This idea was first presented and fully described in our previous works [6, 8]. At first it was used for recognition of single characters. Here, and for the first time, it is applied to recognition of whole words. Hereafter, for your convenience, we repeat main propositions of previous works, as well as introduced modifications.

This method bases on fact, that for correct image recognition we usually need only partial information about its shape – its silhouette or contour.

We examine two "views" of each word, extracting from them a characteristic vector, which describes this given word. The view is a set of points that plot one of two projections of the object (top or bottom) – it consists of pixels belonging to the contour of a word and having extreme values of y coordinate – maximal for top, and minimal for bottom view (Fig. 2.)

Fig.2 – Two views of a sample word

Next, characteristic points are marked out on the surface of each view to describe the shape of that view. The method of choosing these points and quantity of them may vary. In our experiments 30 uniformly distributed characteristic points are taken for each view.

The next step is calculation of y coordinates for taken points. Thus we obtain two 30-element characteristic vectors describing given word. Novelty, we introduce addition of third vector describing aspect ratio of tested image. It consists of two values – width and height of the image. Then all three vectors are normalized, according to the formula of Eq. (1) so that their values are in the range <0, 1>.

$$x_j \leftarrow \frac{x_j}{\sqrt{x_1^2 + x_2^2 + \ldots + x_N^2}}, \qquad (1)$$

Next these three vectors are formed into one 62-element characteristic vector, which describes given word, and which is the base for further analysis and classification.

3. CLASSIFICATION WITH NEAREST NEIGHBOR METHOD

For the sake of classification three vectors describing views of tested word were combined into one 62-element vector.

When classifying with the method of simple comparison, vector describing tested word was compared with vectors describing words contained in Base Set (words from Base Set are already classified.) Then we search the nearest neighbor of tested word (i.e. such vector to which the distance is minimal.) We use 1-norm distance (Manhattan distance, Eq. 2) for that calculations.

$$d = \sum_{i=1}^{n} |x_i - y_i|, \qquad (2)$$

We assume that tested vector is of the same class as his nearest neighbor, i.e. image in question shows the same word, as image described by vector found in Base Set.

This method allowed us to achieve 88% correctly recognized words. Table 1 presents detailed results for some of the tested words and fonts.

Table 1. Results of recognition for selected words

Animal Name	Recognition Rate
dog	100%
mouse	96%
cockroach	93%
goat	89%
shark	82%
tiger	75%
ant	57%

As can be seen, results for different words differ. The best result was obtained for the word "dog" – 100% of samples were correctly recognized. The worst was "ant" with 57% recognition rate. All in all the results are promising – only for 7 words recognition rate is lower than 80%.

Table 2 presents details for some of used fonts. As can be seen, words printed with some of them, were recognized with 100% or nearly 100% effectiveness. In fact with more than one-half of used fonts obtained recognition rate was over 90%. Although some fonts lowers these results. These errors occurs in case of calligraphic fonts or fonts imitating handwriting. Probably it was because no such font was in the Base Set – it was composed of standard, "printed" fonts only. Therefore high results for some atypical fonts are all the more satisfactory.

Table 2. Results of recognition for selected fonts

Font Name	Recognition Rate
Lucida Bright Regular	100%
Georgia Bold	97%
Souvenir Demi	95%
Verdana Bold	84%
Comic Sans MS	65%
Bernhard Fashion	48%
Staccato	11%

4. CLASSIFICATION WITH NEURAL NETWORK

In some cases the method of Nearest Neighbors is quite inconvenient, because it needs to maintain big database of already classified vectors, and in order to recognize a word all the vectors from the database are compared with the vector describing the tested word. In case of huge database it may be time-consuming and hence inefficient. The use of Artificial Neural Networks allows us to avoid those costs.

In our research we use Multi-Layered Perceptron: classic feed-forward neural network, with one hidden layer (composed of 125 neurons), trained by the backpropagation method [9]. As a transfer function, we have taken the bipolar logistic sigmoid function (Eq. 3.)

$$F(x) = \frac{2}{1+e^{-x}} - 1, \tag{3}$$

Network has 62 inputs (because our input vector describing each word is also build of 62 elements) and 75 outputs (the number of classes in our database.) First network was trained with vectors from the Training Set. The training stage was performed, until recognition rate on Training Set climbs to 95% (it was about 500 epochs.) Next, fully trained network was tested with remaining words (words from the Test Set) – 62 element vector describing given word (obtained from view-based algorithm) was presented on input of a network. In the output we get information about class of input vector (i.e. what word it describes) – the number of output neuron on which we observe the biggest signal indicates the number of the class.

Classification with Neural Networks allowed us to obtain 80% correctly recognized words. As can be seen, the overall effectiveness is slightly lower than with the Nearest Neighbor method. It is mainly because of great number of classes and rather tiny Training Set (for comparison purposes we want to use exactly the same Training Set for both methods.) This result can be improved by expanding Training Set or adding some noise signals to it. Additionally some gain can be achieved by altering training method or net architecture. Those experiments were not included in this work, because it was not the main goal of this paper.

Table 3 presents detailed results for some chosen words. Although general result is lower, it can be seen, than some words were recognized with higher accuracy. Difference between the best and the worst recognition rates is reduced. This means, that neural networks have better adapted to given database.

Table 3. Results of recognition for selected words

Animal Name	Recognition Rate
dog	95%
mouse	86%
cockroach	69%
goat	88%
shark	94%
tiger	91%
ant	81%

Similarly in Table 4 some differences in recognition rates for various fonts can be observed – rates for some fonts are improved.

Table 4. Results of recognition for selected fonts

Font Name	Recognition Rate
Lucida Bright Regular	81%
Georgia Bold	83%
Souvenir Demi	91%
Verdana Bold	95%
Comic Sans MS	85%
Bernhard Fashion	40%
Staccato	19%

5. CLASSIFICATION WITH NEURAL NETWORK AND TOEPLITZ MATRICES

Experiments have also been done on using TM (Toeplitz Matrix minimal eigenvalues) along with the NN in a hybrid system to develop an algorithm for a less data feature vector without affecting the classification rate. TM was applied successfully with NN in classifying separate characters [11, 12]. This would be an important step in using TM and their minimal eigenvalues to large class groups of object images. The work in this direction is still under research and the current results are not publishing worthy.

6. CONCLUSIONS AND FUTURE WORK

In this paper a new method of word recognition was presented. It combines View-Based method (previously used for letter recognition) with well-known and popular Artificial Neural Networks. Results of experiments show that this method is successful for printed-word recognition. However, the efficiency for words printed with calligraphic fonts, or fonts imitating handwritings is not sohigh, anyhow generally the results are promising and encouraging for further work.

Moreover, this method has possibilities for further improvements and adjustments, and can be modified – especially in the classification stage, for example by altering neural net architecture, or using another training method, what can change final results. We also want to expand our database, by adding more both common and uncommon fonts.

Our next step will be to use this method for recognition of handwritten words. We are also planning to add two other methods, used in our previous works: dynamic time warping [10], which gave good results with signatures and the application of Toeplitz matrix minimal eigenvalues. The latter is under continuous studying and research.

7. REFERENCES

[1] D. Burr, *Experiments on Neural Net Recognition of Spoken and Written Text*, IEEE Trans. On Acoustic, Speech and Signal Proc., 36, no. 7, 1162-1168, 1988.

[2] K. Saeed. *Computer Graphics Analysis: A Method for Arbitrary Image Shape Description*. Volume 10, Number 2, 2001, pp. 185-194, MGV - International Journal on Machine Graphics and Vision, Institute of Computer Science. Polish Academy of Sciences. Warsaw 2001.

[3] R. C. Gonzalez and R. E Woods. *Digital Image Processing*. Addison-Wesley Publishing Company, 1992. pp. 7-9, 413-414.

[4] K. Saeed, M. Rybnik, M. Tabedzki, *Implementation and Advanced Results on the Non-interrupted Skeletonization Algorithm*, 9th CAIP Int. Conference on Computer Analysis of Images and Patterns, 5-7 Sept., Warsaw 2001. Proceedings published in: Lecture Notes in Computer Science - W. Skarbek (Ed.), Computer Analysis of Images and Patterns, pp. 601-609, Springer-Verlag Heidelberg: Berlin 2001.

[5] M. Rybnik, A. Chebira, K. Madani, K. Saeed, M. Tabedzki, M. Adamski, *A Hybrid Neural-Based Information-Processing Approach Combining a View-Based Feature Extractor and a Treelike Intelligent Classifier*, CISIM - Computer Information Systems and Industrial Management Applications, pp. 66-73, WSFiZ Press, Bialystok 2003.

[6] K. Saeed, M. Tabedzki, *A New Hybrid System for Recognition of Handwritten-Script*, COMPUTING - International Scientific Journal of Computing, Institute of Computer Information Technologies, Volume 3, Issue 1, pp. 50-57, Ternopil 2004.

[7] B. Widrow. R. Winter. R. Baxter. *Layered Neural Nets for Pattern Recognition.* IEEE Trans. ASSP, 1988, Vol. 36, pp. 1109-1118.

[8] K. Saeed, M. Tabedzki, M. Adamski, *A New View-Based Approach for object Recognition*, CONRADI, Research Review, Volume 2, Number 1, pp. 80-90, University of Vaasa, Turku 2003.

[9] S. I. Amari. *Mathematical theory of neural learning.* New Generation Computing. 1991. Vol. 8, pp. 281-294.

[10] K. Saeed, M. Adamski, *Offline Signature Classification with DTW application*, 14th Conference on Informatics Systems - KBIB'05 (in Polish), vol. I - pp. 455-460, Czestochowa 2005.

[11] K. Saeed. *A New Approach in Image Classification.* Proc. 5th International Conference on Digital Signal Processing and its Applications - DSPA'03. Moscow 2003. Vol. 1, pp. 49-52.

[12] K. Saeed, M. Tabedzki, *Cursive-Character Script Recognition using Toeplitz Model and Neural Networks*, pp.658-663, L. Rutkowski, J. Siekmann, R. Tadeusiewicz, L. Zadeh (Eds), Lecture Notes in Computer Science – LNCS 3070, Springer-Verlag, Heidelberg 2004.