applied sciences

MDPI

Article

Short Words for Writer Identification Using Neural Networks

Georgia Koukiou

Electronics Laboratory, Physics Department, University of Patras, 26504 Patras, Greece; gkoukiou@upatras.gr; Tel.: +30-2610-996-147

Abstract: In biometrics, it is desirable to distinguish a person using only a short sample of his handwriting. This problem is treated in the present work using only a short word with three letters. It is shown that short words can contribute to high-performance writer identification if line characteristics are extracted using morphological directional transformations. Thus, directional morphological structuring elements are used as a tool for extracting this kind of information with the morphological opening operation. The line characteristics are organized based on Markov chains so that the elements of the transition matrix are used as feature vectors for identification. The Markov chains describe the alternation in the directional line features along the word. The analysis of the feature space is carried out using the Fisher linear discriminant method. The identification performance is assessed using neural networks, where the simplest neural structures are sought. The capabilities of these simple neural structures are investigated theoretically concerning the achieved separability into the feature space. The identification capabilities of the neural networks are further assessed using the leave-one-out method. It is proved that the neural methods achieve identification performance that approaches 100%. The significance of the proposed method is that it is the only one in the literature that presents high identification performance using only one short word. Furthermore, the features used as well as the classifiers are simple and robust. The method is independent of the language used regardless of the direction of writing. The NIST database is used for extracting short-length words having only three letters each.

Keywords: writer verification; morphological line features; time-series modeling; clustering analysis; language independence; Markov chains; neural networks

check for updates

Citation: Koukiou, G. Short Words for Writer Identification Using Neural Networks. *Appl. Sci.* **2023**, *13*, 6841. https://doi.org/10.3390/app13116841

Academic Editor: Krzysztof Koszela

Received: 16 December 2022
Revised: 28 May 2023
Accepted: 3 June 2023
Published: 5 June 2023

Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Automated handwriting recognition has been developed for decades, and one of its main purposes is to be used in legal and forensic cases to identify a writer. Optical character recognition (OCR) is another broad area in handwriting recognition, which handles formatting, performs correct segmentation into characters, and finds the most justifiable words. For writer recognition, both handwritten text and signatures have been used. When the text is used for author identification, the used characteristics may be static, which can be extracted from the lines drawn by the writer, or dynamic characteristics recorded during writing using a pen-based computer screen surface, sensing the pen-tip pressure, speed, and acceleration. The dynamic features add improved reliability in writer recognition [1].

In the field of writer recognition from text, and without considering the dynamic characteristics, there is a variety of methods for extracting characteristics from the lines of the text with significant results and reliability in recognition. An extensive analysis of various pattern recognition methods for handwriting recognition in historical documents is presented in [2]. In [3], a review is presented on the state of the art in offline text-independent writer identification methods for three major languages, namely English, Chinese, and Arabic. In [4], the authors described how oriented Basic Image Feature Columns (oBIF columns) can be used for writer identification.

A novel method for writer identification called graphemes or fraglets is presented in [5]. Graphemes extracted from a document are represented in terms of Fourier and wavelet descriptors. In [6], a feature vector is derived using morphological processing of the horizontal profiles (projection functions) of the words. In [7], an end-to-end framework for online text-independent writer identification using a recurrent neural network (RNN) is presented. A novel offline text-independent writer identification method is proposed in [8] based on scale-invariant feature transform (SIFT), composed of training, enrollment, and identification stages.

Special emphasis has been given recently to the use of neural networks and deep neural networks for handwriting identification. Some of the most representative publications follow. In [9], the application of neuroevolution to the automatic design of convolutional neural network (CNN) topologies is explored, developing two novel solutions based on genetic algorithms and grammatical evolution. The MNIST dataset for handwritten digit recognition is used, achieving a result that is highly competitive with the state-of-the-art without any kind of data augmentation or preprocessing. A new neural network architecture is proposed in [10], which combines a deep CNN with an encoder–decoder, called sequence-to-sequence, to solve the problem of recognizing isolated handwritten words. The proposed architecture aims to identify the characters and contextualize them with their neighbors to recognize any given word. A survey of the top state-of-the-art contributions reported on the MNIST dataset for handwritten digit recognition can be found in [11]. In addition, a new dataset, EMNIST, which was introduced in 2017 and involves both digits and letters, with a larger amount of data acquired from a database different than MNIST is explored, and some results on EMNIST are surveyed. The research in [12] focuses on the recognition of Farsi handwriting digits and illustrates applications in biomedical science. The paper aims at identifying handwritten Farsi digits written in different handwritten styles. Digits are classified using several traditional methods, including K-nearest neighbor, artificial neural network, and support vector machine (SVM) classifiers. New features of digits, namely, geometric and correlation-based features, have been demonstrated to achieve better recognition performance.

Furthermore, a new dataset of Arabic letters written exclusively by children aged 7–12, which is called Hijja, is presented in [13]. The dataset contains 47,434 characters written by 591 participants. An automatic handwriting recognition model based on CNNs is proposed. Some sets of simple-shaped geometric features are used for achieving the offline verification of signatures in [14]. These features include baseline slant angle (BSA), aspect ratio (AR), normalized area (NA), the center of gravity, as well as the line's slope that joins the center of gravities of the signature's image two splits. The similarity measure within the feature space between the two signatures is determined using the Euclidian distance. Several modern techniques for neural network training and recognition enhancement are based on their structures' symmetry [15]. Such approaches demonstrate impressive results, both for recognition practice and for understanding data transformation processes in various feature spaces. This survey examines symmetrical neural network architectures—Siamese and triplet. Offline handwritten signature verification is one of the most prevalent and prominent biometric methods in many application fields [16]. This study proposes a two-stage Siamese neural network model for accurate offline handwritten signature verification with two main ideas: (a) adopting a two-stage Siamese neural network to verify original and enhanced handwritten signatures simultaneously and (b) utilizing the focal loss to deal with the extreme imbalance between positive and negative offline signatures. From the perspective of handwriting analysis, the authors of [17] exemplified and systematized general methods for an informal falsification of artificial neural networks applied to the verification of offline handwritten document authorship. The work in [18] proposes an approach to analyze the images of simple graphemes using CNNs.

Recently, a few papers were published that propose methods for offline writer identification, which are different from the method proposed in the present work. Specifically, an offline system for text-independent writer identification of handwritten documents was

presented in [19]. The system extracts a handwritten connected component contour, which in turn is divided into segments of a specific length. The system uses the proposed features to train a k-means clustering algorithm to construct a codebook of size K. The method uses occurrence histograms of the extracted features in the codebook to create a final feature vector for each handwritten document. The work in [20] exploits the textural information in handwriting to characterize writers from historical documents. The authors use oBIF (oriented Basic Image Features) and hinge features and introduce a novel moment-based matching method to compare the feature vectors extracted from writing samples. Classification is based on the minimization of a similarity criterion using the proposed moment distance. Finally, a comparative study in [21] evaluates several deep learning architectures, such as VGG16, ResNet50, MobileNet, Xception, and EfficientNet, end-to-end to examine their advantages over offline handwriting for writer identification problems with IAM and CVL databases.

In the present work, writer verification is carried out based on short-length words and neural networks. We were inspired by the need that exists several times to recognize/distinguish a person only from one of his words that he may have left in a small note. The features used are the elements in the transition matrix of the Markov chains, which describe the alternation in the directional morphological line features as moving from left to right along the image of the word. The separability of the clusters formed on the high dimensional space is assessed using the simplest possible neural structures. A theoretical insight into the decision region formation is given based on the capabilities of the perceptrons in the input layer to draw linear (N-1)-D subspaces in the N-D feature space. The database used [22] is the web-released NIST (National Institute of Standards and Technology) Special Database 19 (SD19). It was possible to extract a small number of specific short words such as "the" or "end" for testing our features for writer identification. This technique can be applied as well in offline signature recognition making use of the directional information that is contained in its lines.

The contents of this paper are as follows. The basic steps of the proposed writer identification method are given in Section 2. The NIST database used in the experiments is described in detail in Section 3. In Section 4, preprocessing of each word using the line morphological transformations is carried out in order to extract the directionality of the line features of the word. In Section 5, the formation of the feature vector is explained, which is based on the Markov chains formed as the morphological structuring elements (SEs) fit along the word. The Fisher linear discriminant (FLD) method is used to give an insight into the feature space. In Section 6, a basic theoretical background is provided regarding the separability of the feature space using very simple neural structures. The experimental results, where simple neural networks classify successfully the writers using small words, are provided in Section 7. The conclusions are drawn in Section 8.

2. Methodology—Outline of the Verification Method

Writer verification is carried out based on short words. The NIST database is used to extract these words from a total of ten writers. A word written by an unknown person must be classified as belonging to a specific person among those which have been recorded. The approach is independent of the language used.

The independence of the method from the language used is based on the fact that the features to be extracted reveal the orientation of the lines along the word that make up the letters of the word. We suggest that this orientation be determined using simple oriented morphological SEs. Thus, with the morphological operation of opening, these elements reveal the line orientation for each position on the word. The alternating orientation of the lines, as we move from left to right along the word, is captured using a discrete first-order Markov process (Markov chain). The method is independent of the language in which the word is written, even if the direction of writing is from right to left, since the Markov chain format is independent of the direction of writing. The eigenvalues of the transition matrix for this process are the features used to identify the writer.

The steps followed in the proposed verification procedure are:

a. Preprocessing an image of the word from an unknown person. The preprocessing stage involves thresholding so that the word is isolated from the background as well as isolated pixel (noise) removal and thinning the lines of the word.
b. Extract orientation characteristics using morphological transformations. The morphological opening is used with directional SEs. The SE that fits better in each position on the word characterizes the orientation at this position and is recorded.
c. Form the feature for identification by describing the alternation in orientation characteristics using Markov chains. The transition matrix is used as the feature vector.
d. Demonstrate the feature space in lower dimensions using the FLD approach.
e. Experimental procedure testing the cluster separability capabilities of the simplest possible neural structures in the original high-dimensional feature space. Test performance when morphological line features of different lengths are used.
f. Characterize the feature space and the cluster separability according to the obtained neural structures.

This procedure is schematically shown in Figure 1.

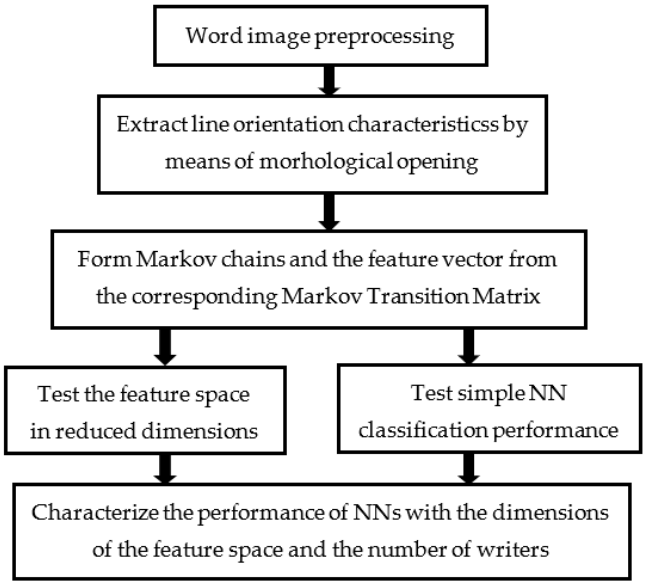


Figure 1. Schematic representation showing the proposed writer identification procedure.

3. The NIST Database 19

The NIST database contains thousands of writers who have written the same single text only once. The updated web-released NIST Special Database 19 (SD19) was used in our experiments. The original Special Database 19 contains the full-page binary images of 3669 handwriting sample forms (HSFs—Figure 2) and 814.255 segmented handprinted digit and alphabetic characters from those forms [22].

Using the HSFs from the NIST Special Database 19, we had the opportunity to test the discrimination between writers when the length of the word is short and the number of the same words coming from a single writer is very small. Since only one text sample from each person was available, we used those words that appear many times in the text so that a cluster can be formed for each writer and word. Specifically, as shown in Figure 2, we used the word 'the', which is repeated six times in the free text, and the word 'and', which is repeated three times. A large number of writers was available so that writer verification using short words and simple neural structures could be assessed.

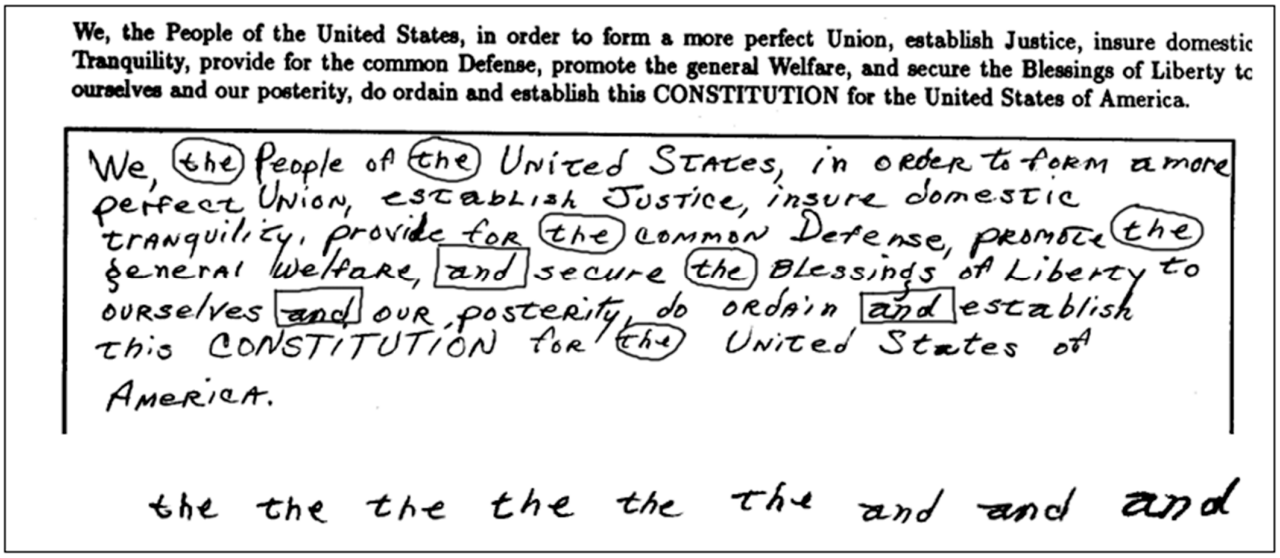


Figure 2. One of the full-page binary images of the 3669 HSFs from the original NIST Special Database 19 (f0000_14.png). Six replicas for the word "the" and three for the word "and" were extracted to uniquely represent the specific writer.

4. Preprocessing and Directional Morphological Transformations

Preprocessing requires the isolation of the word from its background, keeping all its directional characteristics and simultaneously transforming it into a thin line so that only the necessary information is kept. Preprocessing algorithms are applied to the short words selected from the NIST database as shown in Figure 3. Firstly, the negative of the word image is obtained to work on a black background, and after that, thinning is applied using morphological algorithms in order to maintain the strong line features of the word.



Figure 3. A thresholded word image after negation and line-thinning using morphological transforms.

The thinned image in Figure 3 is after that processed with directional morphological structuring elements (SE) to extract, in each location of the image, the type of the dominating directionality. Directional SEs of length three are shown in Figure 4. When an SE fits well on a part of the line image, this means that the lines in this part of the image are oriented mainly towards the direction of the SE. The degree of fitting for an SE on a specific part of a line image can be estimated using the morphological operation called opening [6]. The more an SE with a specific directionality fits the line image at a specific position of the word, the more the lines in the original image towards this direction remain unaffected when opening is performed on the image. Figure 5 demonstrates the effect of a vertical line SE of length 3 on a 3-pixel-wide strip containing the letter "h".

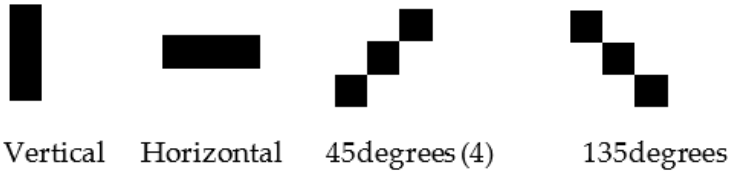


Figure 4. Directional structuring elements with a length of 3 pixels.

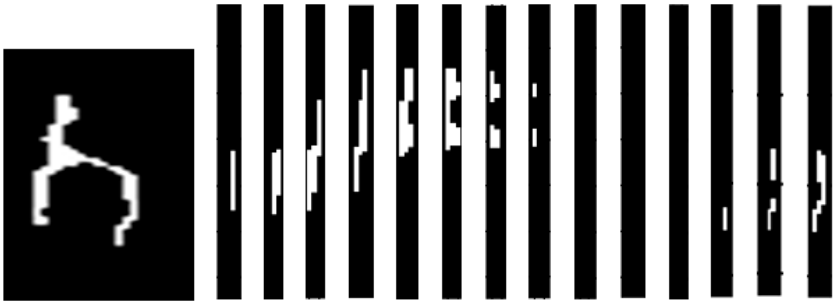


Figure 5. Processing of the letter 'h' in the word 'the' with morphological opening using a vertical directional structuring element of length 3. The vertical lines in the letter remain unchanged. The vertical line content of the letter is obvious from the strips that contain white information.

In more detail, the directionality of the lines in each location of a specific word, namely, the word "the", is recorded using morphological openings along the word. In the case of SEs with a length of 3 pixels, the word is scanned pixel by pixel from left to right in strips at a width of 3 pixels (Figure 5). This means that 66% of the area of each strip is the same as its neighboring strips. Each strip is morphologically opened separately with each one of the four directional SEs depicted in Figure 4. To accommodate the opening operation at the margins of the strip, one line of pixels was added around the strip. It is assumed that in each strip, the best fitting SE is the one which leaves most of the pixels in the specific strip unaffected after opening. This SE represents the specific strip. The alternation in the best fitting SEs can be considered as the states of a Markov model, which change stochastically according to the model. The modeling of the alternation in the SEs as a Markov states sequence is examined in the next section. The Markov models constitute the features for writer identification.

The morphological processing of the word 'the' with the four directional SEs, given in Figure 4, was carried out, and the sequence of the best fitting SEs was derived. Figure 6 depicts a sample for the word "the" with its corresponding sequence of the best fitting directional SEs.

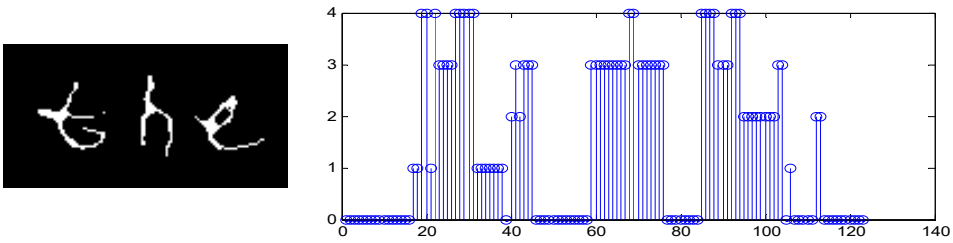


Figure 6. The word 'the' in image f0000_1_1.tif from the NIST Special Database 19 and the corresponding sequence of the orientation SEs from Figure 3 that fit along the word. Each number corresponds to one of the SEs. Zero corresponds to blank space.

5. Feature Extraction and Feature Space

The feature vector for writer identification is formed using first-order Markov chain models (FMCs), which describe the alternation in the different orientation SEs that fit in each specific location of the word as the opening is performed along the word. A Markov chain is composed of a sequence of states (representing the different oriented SEs) with the strict property that the probability to have a specific state—SE in a specific strip—depends solely on the state of the previous strip, i.e.,

$$\Pr(r_i[k]|r_j[k-1], r_m[k-2], \ldots, r_n[1]) = \Pr(r_i[k]|r_j[k-1]) = t_{i/j} \quad (1)$$

where r_i, r_j, ... r_n stands for the different states (i.e., the SEs with specific orientation); in this work, there are five states, taking values from 0 to 4. k stands for the spatial variable

that corresponds to the movement along the word. State 0 corresponds to a blank space or no SE present in the sequence. The absence of SEs along the word is a significant feature for writer identification since it reveals blanks between letters. A matrix T that contains all possible transition probabilities $t_{j/i}$ is a square matrix called the transition matrix. Row i in the transition matrix T designates the previous state, while column j designates the next state, and each of the rows in a Markov transition matrix adds up to one.

For a specific word (of sufficient length), the transition matrix T possesses specific values and uniquely characterizes the writer. Accordingly, we used the transition matrix T as the feature for identifying a writer among others. Eventually, the matrix T contains second-order statistics regarding the transitions among the oriented SEs. Experimentally, the feature vector used for writer identification is the transition matrix given in a reshaped 1-D vector v of size 25×1:

$$T = \begin{pmatrix} t_{1/1} & \cdots & t_{5/1} \\ \vdots & \ddots & \vdots \\ t_{1/5} & \cdots & t_{5/5} \end{pmatrix} \overset{yields}{\rightarrow} \begin{pmatrix} t_{1/1} \\ t_{1/2} \\ \vdots \\ t_{1/5} \\ t_{2/1} \\ \vdots \\ \vdots \\ t_{5/5} \end{pmatrix} \overset{yields}{\rightarrow} \vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ \vdots \\ \vdots \\ v_{25} \end{pmatrix} \tag{2}$$

Dimensionality reduction from 25 dimensions into 2 or 3 dimensions is necessary to perceptually understand the separability of the clusters. It is achieved using the FLD method, i.e., the projection using a suitable linear transformation W in a new space. The vectors w_i of W are the new directions, where each 25-D vector v_i derived from the Markov chain models will be projected. This linear transformation is:

$$y_i = W^t v_i \tag{3}$$

where y_i is the transformed v_i. One of the important criteria used in the literature [23] for best separating the clusters is given by maximizing the function $J(w)$ in the transformed space:

$$J(w) = \frac{w^t S_B w}{w^t S_W w} \tag{4}$$

The transformation vectors w_i of W are obtained from the solution of the generalized eigenvalue problem:

$$S_B w_i = \lambda_i S_W w_i \tag{5}$$

where S_W is called the within-scatter-matrix and S_B is the between-scatter-matrix. The solution of (5) gives the eigenvalues that correspond to each of the above eigenvectors w_i. From this solution, we will use the most important directions corresponding to the largest eigenvalues. Accordingly, the sum of the two largest eigenvalues over the sum of all eigenvalues gives the quality of cluster separability in the reduced 2-D feature space. The clustering capabilities of the directional SEs of lengths 3 and 5 were examined and demonstrated in the case of the short-length words 'and' and 'the' obtained from the NIST Special Database 19. When both SEs are combined giving a 50-D feature vector, the largest eigenvalues obtained for the word "the" are those given in Table 1.

It is well known that the separability information in each direction is relevant to the square root of the corresponding eigenvalue. Accordingly, almost 65% of the information for cluster separability is held in the three directions corresponding to the first three largest eigenvalues. However, significant information for cluster separability is held in the other directions as well. Figure 7a,b presents the separation of the clusters corresponding to 10 writers. Each cluster has six members resulting from the word 'the' written six times by

each writer. Obviously, the clusters are separable, although we are working in the three dimensions which correspond to the three largest eigenvalues, while the use of the other dimensions can support better separability. For demonstration purposes, the ten clusters are depicted in the three most important directions, as can be seen in two different views with different aspects in Figure 7a,b.

Table 1. Ten largest eigenvalues corresponding to the most prominent directions for cluster separability in the 50-D feature space, for the word "the".

λ:	363	121	47	32	14	11	7	5	4	1
$\sqrt{\lambda}$:	19	11	6.8	5.6	3.7	3.3	2.6	2.2	2	1

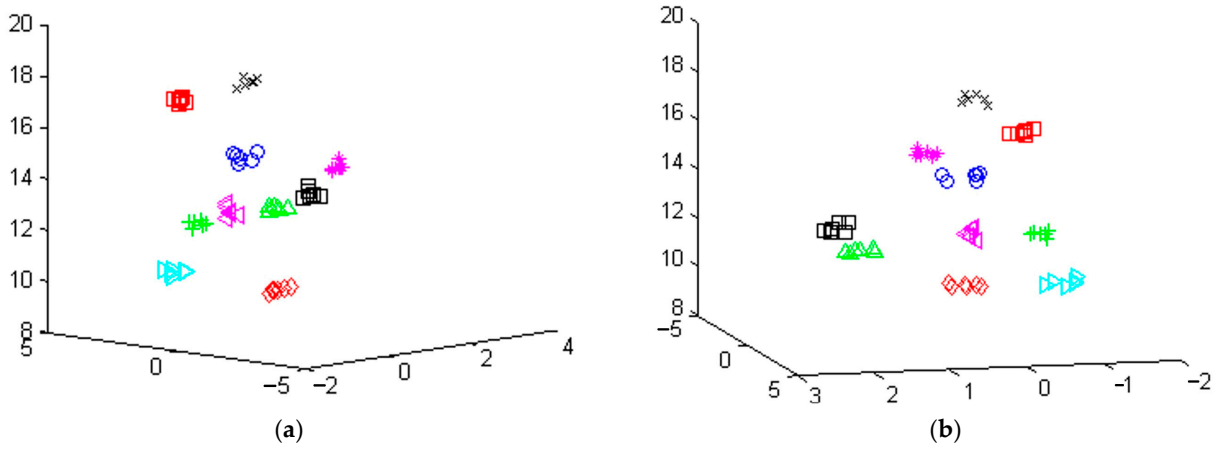


Figure 7. Writer discrimination results. The clusters correspond to the same word "the", written 6 times by 10 different writers. The FLD is used to project the clusters onto three dimensions with the maximum separability. The measurements were obtained using morphological structuring element of length 3 and 5; consequently, dimensionality reduction from 50 to 3 dimensions was performed. Two different aspects of the same 3-D feature space are given in (**a**,**b**).

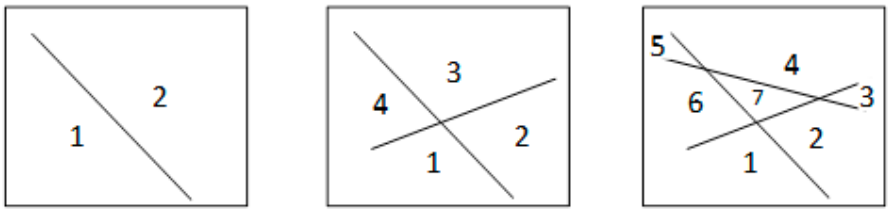
In practice, using neural networks with all available inputs, the separability information distributed in all feature space dimensions is taken into consideration. This gives the neural structures the superiority of providing higher writer separability performance in the case where short words are used, as described in the next section.

6. Discussion of Space Separability

The capabilities of the conventional multilayer neural networks were described in an early tutorial paper by Lipmann [24]. In that work, it was clearly stated that:

a. The first layer of perceptrons draws in the N-D feature space linear subspaces of (N−1) D dimensionality.
b. The second (and the third, if necessary) layers of perceptrons combine the above linear (N−1) D subspaces to create any type of convex and non-convex decision regions in the original feature space.

When the cluster populations are well separated, such as those shown in Figure 7a,b, it is expected that the neural structure that will be suitable for separating the feature space will be the simplest possible. For example, if two populations are to be separated, only one linear subspace is needed, as shown in Figure 8a, while with two subspaces in a general position, we can divide the space into four regions (Figure 8b), and with three subspaces in a general position, we can divide the 2-D feature space into seven regions (Figure 8c).



(**a**) (**b**) (**c**)

Figure 8. The maximum number of separable regions in a 2-D feature space. (**a**) Two regions are defined with one linear subspace. (**b**) Four regions are defined with two linear subspaces. (**c**) Seven regions are defined with three subspaces in general position.

According to the previous material, the question which arises is:

"How many separable regions can be obtained in a n-dimensional feature space when m hyperplanes of dimensions n−1 are employed is a general position?"

The answer is obtained from the following proposition [25]:

Let A be an n-dimensional arrangement of m hyperplanes with n-1 dimensions in a general position. The number r(A) of well-separated regions determined in the n-dimensional space for the general case is

$$r(A) = 1 + m + \binom{m}{2} + \cdots \binom{m}{n} \tag{6}$$

For example, if n = 50, as in the case of features extracted in Sections 4 and 5, then for m = 3 subspaces, we obtain a number of well separable regions as

$$r(A) = 1 + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 8 \tag{7}$$

while for m = 4 subspaces, the number r(A) of well separated regions is

$$r(A) = 1 + \binom{4}{1} + \binom{4}{2} + \binom{4}{3} + \binom{4}{4} = 1 + 4 + 6 + 4 + 1 = 16 \tag{8}$$

According to Equations (7) and (8), the number of (n−1)-D subspaces required to separate the ten populations shown in Figure 7a,b should be more than three to obtain ten well-separated regions. However, this inference is not valid since the neural structure combines the (n−1)-D linear subspaces drawn with the first layer of perceptrons in order to form more complicated separable regions. This is depicted in the following using the populations of the writers shown in Figure 7a,b, on the 2-D space, on which the clusters have their best separability according to the FLD (5). Figure 9 shows the position of the clusters in the two highest separability dimensions, while Figure 10 depicts the way these populations have been separated using a neural structure of two layers with three and four perceptrons, respectively. Obviously, the linear subspaces that separate the populations are curved according to the cluster's shape.

As it is depicted in Figure 10, a very simple neural structure can separate the ten clusters with well-formed decision regions. For further exploitation of the capability of a simple neural network to separate the same ten clusters, we retrained the same network in order to test if the 2-D feature space is again separated in a similar way. The results are depicted in Figure 11. It is obvious that retraining the same sized neural network (3–4) converges again to the same minimum, which means to almost the same weights, and thus to the same decision regions as in Figure 10. The whole procedure reveals stability in the convergence of the neural structure with two layers of three and four perceptrons, respectively.

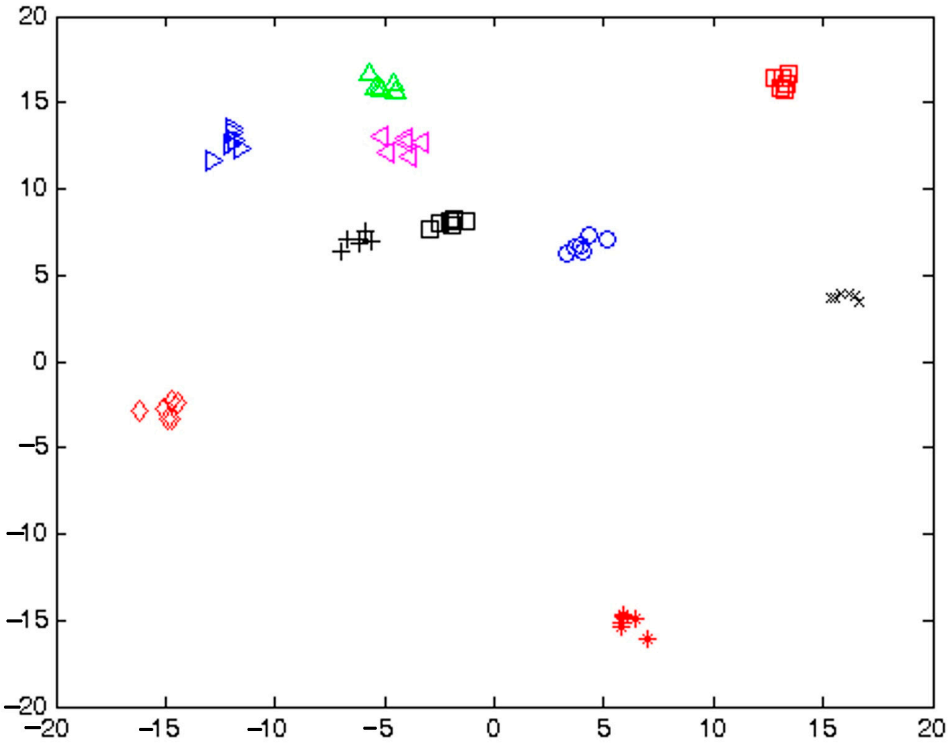


Figure 9. The 2-D space with maximum separability obtained using the FLD from the 50-D space. The writers are well separable.

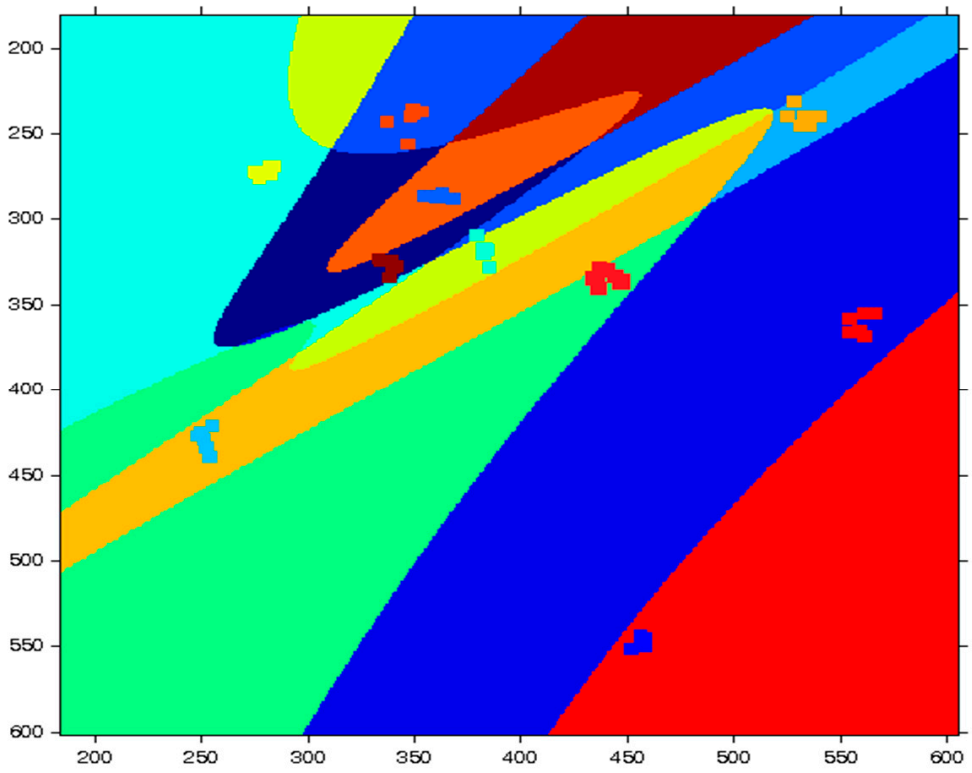


Figure 10. The separation of the feature space shown in Figure 8 using a neural structure of two layers with three and four perceptrons, respectively. Colored regions represent the decision regions belonging to each separate cluster.

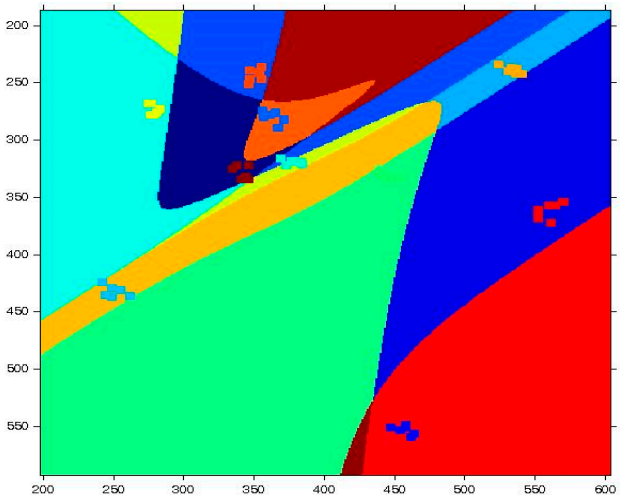


Figure 11. In the case of retraining the neural structure of two layers with three and four perceptrons, respectively, with the same clusters, the achieved separation of the feature space is almost the same as the one shown in Figure 10 (convergence stability). Colored regions represent the decision regions belonging to each separate cluster.

7. Classification Results with Simple Neural Networks

For experimentation with neural networks and assessing their capabilities to discriminate among writers based on short words, the data used were as follows:

a. Ten writers from the NIST database. From each writer, we have six samples of the word "the" and three samples of the word "and".
b. For each word, a 25-D feature vector is created using the directional morphological SEs of length 3, and its classification capabilities are tested.
c. For each word, a 25-D feature vector is created using the directional morphological SEs of length 5, and its classification capabilities are tested as well.
d. We merge the above 25-D vectors and perform writer identification with vectors of 50-D. The clarification capabilities of the simple neural networks are tested using the 50-D vectors.
e. We test the generalization of the neural structures to classify the writers using the leave-one-out method.

Based on the material provided in the following paragraphs, it is evident that simple neural networks with two layers and a few perceptrons achieve 100% classification performance.

7.1. Classification Based on the Word "the" Using an SE with Length 3

In this case, 10 clusters are formed in the 25-D feature space. Each cluster contains only six members since the word "the" appears only six times in the available text from the NIST database. We started the training of the neural network with complicated neural structures. Gradually, we proceeded with simpler structures, and we reached a two-layer structure with three neurons in the first layer and four neurons (3–4 neural structure) in the output layer, which performed a total separation of the ten clusters.

Moreover, the classification performance of a two-layer, 3–4 neural structure was tested using the leave-one-out method. Accordingly, one member from each of the ten clusters was extracted from the training of the neural network. So, five-member clusters were used for the training. Again, a two-layer, 3–4 neuron neural structure was found suitable for separating the ten clusters. With this network, the excluded members of the clusters are correctly classified. This happens irrespectively of the specific member excluded each time from the cluster.

7.2. Classification Based on the Word "and" Using an SE of Length 3

The same results were obtained when the word "and" was used for writer discrimination. In this case, each of the ten clusters has three members given that the word "and" appears in the text three times. The size of the neural structure that achieves this is again a two-layer neural structure with three and four neurons in each layer, respectively.

In this case, the leave-one-out method was used as well and performed successfully, irrelevantly of the excluded member from each cluster. All excluded members were successfully classified with the neural structure trained with the other ones.

7.3. Classification Based on the Word "the" Using SEs of Length 3 and 5—Formation of a 50-D Feature Vector

In this case, as expected, it was found that the 10 clusters formed with 6 members each (number of the words "the") are moving away from each other since new information is inserted by the features based on the directional SEs of length 5. Accordingly, a similar two-layer neural structure with three and four neurons performed successfully in separating the ten writers. The leave-one-out method performs successfully in this case as well.

7.4. Classification Performance Comparison

The performance comparison between different methods for handwriting discrimination (handwriting texts or signatures) is not an easy task due to the different databases used and mainly the different methods/algorithms used. Due to the fact that the choice of the algorithm implemented for a method belongs to other researchers, performance comparisons are practically impossible and can only be carried out qualitatively based on the numerical results given in each manuscript.

The handwriting discrimination method proposed in this paper presents higher classification success compared to the methods given in [14,16], as well as almost all the other methods presented in the literature given in the Introduction.

The high classification success of the proposed method is due to the fact that the feature space is not significantly reduced in dimensions, as shown using the FLD method and obvious from the values of the eigenvalues in Table 1. This means that there exists significant information content along more than ten dimensions in the feature space.

8. Conclusions

The writer verification problem faced in this work is based only on the information contained in one short word. Specifically, if a word must be attributed to a specific person, the most suitable procedure is to ask the person to write down the specific word and then classify the word as belonging to one of the persons whom a neural network has been trained with.

The features used correspond to the elements of the Markov chain model formed with the alternation in the directional information on the lines along the word. The directional information is captured with fitting directional SEs on the lines of the word.

The verification problem is analyzed using the feature spaces formed with the elements in the transition matrices of the Markov models. Since the features used are of high dimensionality, the FLD method is applied to reduce the dimensionality of the space, thus giving an insight into the clusters' separability. The writers can be seen to be separated very well in the reduced 2-dimentional feature space. If we take into consideration that a significant amount of information (almost 40%) regarding the separability of the clusters is lost due to dimensionality reduction, the classification success of the neural structure is expected superior since these structures take advantage of the whole dimensionality of the feature space. We consider the achieved separability satisfactory since all writer clusters are quite apart. Thus, writer discrimination using neural networks is achieved by means of the simplest ones. For ten writers, this corresponds to a two-layer network with three perceptrons in the input layer and four in the output one. The whole procedure

reveals stability in the convergence of the neural structure with two layers with three and four perceptrons.

The proposed method has significant advantages over existing techniques in retrieving handwriting information that is irrelevant to the language used. This is due to the fact that the word is analyzed along the writing axis as a time signal containing the orientation of the lines. The method is independent of the language in which the word is written even if the direction of writing is from right to left since the Markov chain is formatted independent of the direction of writing. As explained in this manuscript, this time-like signal has first-order Markov properties. This fact gives rise to the creation of the features. However, the procedure is very sensitive to the preprocessing stages, i.e., thresholding and thinning. Missing significant directional information in these preprocessing stages would significantly deteriorate the final feature vector.

Our future work will concentrate on combining decisions (fusion) for writer identification. Using decisions coming from more than two words, the final inference for writer identification can be radically improved, especially in cases where the number of writers is high.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available upon request.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Chaabouni, A.; Boubaker, H.; Kherallah, M.; Alimi, A.M.; El Abed, H. Combining of Off-line and On-line Feature Extraction Approaches for Writer Identification. In Proceedings of the IEEE International Conference on Document Analysis and Recognition, Beijing, China, 18–21 September 2011.
2. Fischer, A. Handwriting Recognition in Historical Documents. Ph.D. Thesis, Institute for Computer Science and Applied Mathematics, University at Bern, Bern, Switzerland, 2012.
3. Tan, G.J.; Sulong, G.; MohdRahim, M.S. Writer identification: A comparative study across three world major languages. *Forensic Sci. Int.* **2017**, *279*, 41–52. [CrossRef] [PubMed]
4. Newell, A.J.; Griffin, L.D. Writer identification using oriented Basic Image Features and the Delta encoding. *Pattern Recognit.* **2014**, *47*, 2255–2265. [CrossRef]
5. Kumar, R.; Chanda, B.; Sharma, J.D. A novel sparse model based forensic writer identification. *Pattern Recognit. Lett.* **2014**, *35*, 105–112. [CrossRef]
6. Zois, E.N.; Anastassopoulos, V. Morphological waveform coding for writer identification. *Pattern Recognit.* **2000**, *33*, 385–398. [CrossRef]
7. Zhang, X.Y.; Xie, G.S.; Liu, C.L.; Bengio, Y. End-to-End Online Writer Identification with Recurrent Neural Network. *IEEE Trans. Hum. Mach. Syst.* **2016**, *47*, 285–292. [CrossRef]
8. Wu, X.; Tang, Y.Y.; Bu, W. Offline Text-Independent Writer Identification Based on Scale Invariant Feature Transform. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 526–536. [CrossRef]
9. Baldominos, A.; Saez, Y.; Isasi, P. Evolutionary convolutional neural networks: An application to handwriting recognition. *Neurocomputing* **2018**, *283*, 38–52. [CrossRef]
10. Sueiras, J.; Ruiz, V.; Sanchez, A.; Velez, J.F. Offline continuous handwriting recognition using sequence to sequence neural networks. *Neurocomputing* **2018**, *289*, 119–128. [CrossRef]
11. Baldominos, A.; Saez, Y.; Isasi, P. A Survey of Handwritten Character Recognition with MNIST and EMNIST. *Appl. Sci.* **2019**, *9*, 3169. [CrossRef]
12. Nanehkaran, Y.A.; Zhang, D.; Salimi, S.; Chen, J.; Tian, Y.; Al-Nabhan, N. Analysis and comparison of machine learning classifiers and deep neural networks techniques for recognition of Farsi handwritten digits. *J. Supercomput.* **2021**, *77*, 3193–3222. [CrossRef]
13. Altwaijry, N.; Al-Turaiki, I. Arabic handwriting recognition system using convolutional neural network. *Neural Comput. Appl.* **2021**, *33*, 2249–2261. [CrossRef]
14. Tahir, N.M.; Ausat, A.N.; Bature, U.I.; Abubakar, K.A.; Gambo, I. Off-line Handwritten Signature Verification System: Artificial Neural Network Approach. *Int. J. Intell. Syst. Appl.* **2021**, *1*, 45–57. [CrossRef]
15. Ilina, O.; Ziyadinov, V.; Klenov, N.; Tereshonok, M. A Survey on Symmetrical Neural Network Architectures and Applications. *Symmetry* **2022**, *14*, 1391. [CrossRef]

16. Xiao, W.; Ding, Y. A Two-Stage Siamese Network Model for Offline Handwritten Signature Verification. *Symmetry* **2022**, *14*, 1216. [CrossRef]
17. Marcinowski, M. Evaluation of neural networks applied in forensics; handwriting verification example. *Aust. J. Forensic Sci.* **2022**, 1–10. [CrossRef]
18. Mora, M.; Naranjo-Torres, J.; Aubin, V. Convolutional Neural Networks for Off-Line Writer Identification Based on Simple Graphemes. *Appl. Sci.* **2020**, *10*, 7999. [CrossRef]
19. Ahmed, B.Q.; Hassan, Y.F.; Elsayed, A.S. Offline text-independent writer identification using a codebook with structural features. *PLoS ONE* **2023**, *18*, e0284680. [CrossRef] [PubMed]
20. Gattal, A.; Djeddi, C.; Abbas, F.; Siddiqi, I.; Bouderah, B. A new method for writer identification based on historical documents. *J. Intell. Syst.* **2023**, *32*, 20220244. [CrossRef]
21. Suteddy, W.; Agustini, D.; Adiwilaga, A.; Atmanto, D. End-To-End Evaluation of Deep Learning Architectures for Off-Line Handwriting Writer Identification: A Comparative Study. *JOIV Int. J. Inform. Vis.* **2023**, *7*, 178–185. [CrossRef]
22. Grother, P.; Hanaoka, K. *NIST Special Database 19 (SD19), Handprinted Forms and Characters*, 2nd ed.; Information Access Division, Information Technology Laboratory, NIST, National Institute of Standards and Technology, U.S. Department of Commerce: Gaithersburg, MD, USA, 2019.
23. Duda, R.; Hart, P.; Stork, D. *Pattern Classification*, 2nd ed.; Wiley & Sons: New York, NY, USA, 2001.
24. Lippmann, R. An Introduction to computing with neural nets. *IEEE ASSP Magazine* **1987**, *4*, 4–22. [CrossRef]
25. Stanley, R.P. An Introduction to Hyperplane Arrangements. In *Geometric Combinatorics*, 1st ed.; Miller, E., Reiner, V., Sturmfels, B., Eds.; IAS/Park City Mathematics Series: Park City, UT, USA, 2007; Volume 13, pp. 13–30.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.