Handwritten signature identification using basic concepts of graph theory

TOMISLAV FOTAK, MIROSLAV BAČA, PETRA KORUGA
Faculty of Organization and Informatics, Centre for biometrics
University of Zagreb
K.P. Krešimira IV, broj 15, 42000 Varaždin
CROATIA
tomislav.fotak@foi.hr, miroslav.baca@foi.hr, petra.koruga@foi.hr

Abstract: - Handwritten signature is being used in various applications on daily basis. The problem arises when someone decides to imitate our signature and steal our identity. Therefore, there is a need for adequate protection of signatures and a need for systems that can, with a great degree of certainty, identify who is the signatory. This paper presents previous work in the field of signature and writer identification to show the historical development of the idea and defines a new promising approach in handwritten signature identification based on some basic concepts of graph theory. This principle can be implemented on both on-line handwritten signature recognition systems and off-line handwritten signature recognition systems. Using graph norm for fast classification (filtration of potential users), followed by comparison of each signature graph concepts value against values stored in database, the system reports 94.25% identification accuracy.

Key-Words: - handwritten signature, signature recognition, identification, graph theory, biometrics, behavioral characteristics

1 Introduction

Handwritten signature can be defined as the name and surname of the person written by his or her own hand [1]. It is being used in various applications on daily basis. Whether one signs a contract, work documents, petition, or wants to approve a check payment, one will use personal signature to do all those things. We can say that personal signature is being used every day as a mean of giving our consent for an action or a set of actions that needs to be done. The problem arises when someone is trying to imitate our signature and steal our identity. If one does that good enough it could be used to make serious damage to us. Therefore, there is a need for adequate protection of our signature and it needs to be known who actually signed a document. In this case we are entering the field of personal identification. One way to identify people is to use biometric characteristics of each individual. In this paper we will focus on handwritten signature as a biometric characteristic. It belongs to behavioral biometrics and according to [2] it is widely acceptable and collectable biometric characteristic. On the other hand, because it is behavioral characteristic, it has greater entropy than other characteristics. We sign ourselves different every time, so it comes naturally to ask how it is possible identify someone with only his or her handwritten signature. While thinking about this, we notice that signature depends on almost everything. There are few key factors our signature depends on:

- *Physical and psychological state of the person* – includes illness, injuries, fears, heart rate, person's age, calmness, goodwill, etc.
- *Body position* – it is not the same if the person is standing or sitting while signing a document, where is person looking at a moment, what is the burden on signing hand, etc.
- *Writing surface and writing material (pen)* – signature will look different on the various types of paper. It will look different if taken with digitizing tablet or specialized pen. Writing with pen, pencil, stylus or feather also impacts person's signature.
- *Purpose of signing* – signature is usually significantly different if taken in formal environment then in informal.
- *Environmental factors* – environment and people that surround the signatory. This includes noise, luminance, temperature, humidity, etc.

It is easy to conclude that it would be impossible to take all these factors into consideration when developing identification system based on personal signature. Therefore, all handwritten signature authentication or identification systems are trying to

implement the best possible method which will try to summarize all these factors. One of those methods we find in the graph theory which will be described later in this paper.

This paper is organized as follows. In Section 2 basics and previous work, as well as the ideas that are based on previous work, and can be used to develop new methods and systems in the field of handwritten signature identification, are described. Section 3 presents architecture of the identification system and describes chosen concepts of graph theory. The second part of this section will describe identification procedure based on the concepts of graph theory. Finally Section 4 discusses given results and concludes the paper.

2 Signature identification

We already mentioned that handwritten signature is widely accepted and collectable biometric characteristic. This makes it suitable for further research and development of new authentication and identification methods. All those methods need to recognize signatures taken from the same person. From Fig.1 it is clear that this is not a trivial task, since it shows two signatures of the same person taken in the time interval of just a few seconds.

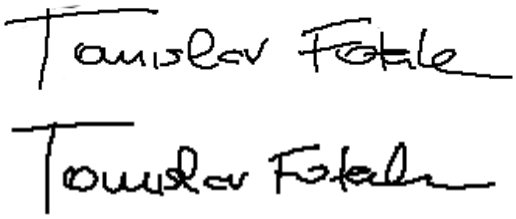


Fig.1. Two signatures of the same person taken within few seconds

While development of authentication methods based on this biometric characteristic is a common thing in the academic and research community, there are only few attempts of developing personal identification system based on handwritten signature.

2.1 Authentication vs. Identification

Both authentication (verification) and identification are important in biometrics. Understanding those terms gives us the basic knowledge needed to understand how the biometric systems actually work. They both rely on the database which contains records about users biometric characteristic features and comparing given biometric characteristic features with those stored in the database.

Biometric authentication is probably much simpler and more often used procedure than biometric identification. It answers the question: "Is the user really who he or she says to be?" User has to provide his or her username and biometric characteristic features to the system. Authentication is often referred as 'one-to-one' comparison.

On the other hand, biometric identification is known as 'one-to-many' comparison. In our context that means that user has to provide only his signature to the system. System will compare his signature against all signatures stored in the database and will calculate match results. The best result, if satisfying some other requirements will be suggested as the identified user. Therefore, biometric identification answers the question: "Who is the person?"

2.2 Previous work

As mentioned earlier, a lot of work has been done in the field of biometric signature verification, i.e. handwritten signature authentication. Some of those works also tried dealing with the signature identification, but because of the handwritten signature's great entropy it is hard to make a good identification system.

When dealing with signature identification, we can talk about off-line and on-line handwritten signature identification. The first one requires only signature image which has to be analyzed in some way. Person does not have to be physically present in the moment of the identification. On the other side, on-line handwritten signature identification requires physical presence of the person. It is usually done with the digitizing tablet or specialized pen which sends 'live data' to the biometric system. Since the approach presented later in this paper can be implemented as both off-line and on-line system we will cover previous work of the off-line and on-line handwritten identification systems. This will show the development of the idea of signature identification.

Handwritten signature is often equated with person's handwriting. In the context of this work this is not possible because we want to identify person's signature and was it made by the real person. Nevertheless, achievements in the field of

the handwriting recognition and writer identification can be very important for the handwritten signature identification because all the methods developed in this field can be implemented to identify signature. This is why we will mention some of the previous work in this field as a good idea that can be used in signature identification.

2.2.1 Off-line signature identification

Most of the work in the field of signature identification deals with off-line signature identification, i.e. off-line writer identification.

After the signature identification system was, according to American National Science and Technology Council, first developed in 1965, one of the first attempts to develop new approach in signature identification was found in [3]. They introduced the use of the revolving active deformable model as a way of capturing the unique characteristics of the overall structure of a signature because they believe that the overall structure of a signature uniquely determines the signature in the majority of cases. The computer-generated model interacted with the virtual gravity field created by the image gradient. Authors claimed that the experiments performed with a signature database showed that the proposed method is promising.

Another attempt to identify person's signature was presented in [4]. Authors used GSC (Gradient, Structural, and Concavity) approach to extract global, statistical, geometrical and topological features of the signature. The binary feature vector was associated with each signature sample and then the proximity of the sample to all other samples was calculated using the Correlation measure to express the similarity between two binary images. The k-nearest neighbor classification was used. The best result was for $k=3$ where they reported identification accuracy of 93.18%.

Not the signature identification algorithm, but the writer identification algorithm was presented by Said, Tan and Baker [5]. They took a global approach based on texture analysis, where each writer's handwriting is regarded as a different texture. They applied the multi-channel Gabor filtering technique followed by the weighted Euclidean distance for the recognition task and got result of 96% identification accuracy. The same principle was applied in [6]. It was proven that presented algorithm actually works with the approximately 95.7% identification accuracy. 2-D Gabor filter method has to convolute the whole image for the each orientation and each frequency. This is computational very costly.

Using wavelet-based GGD instead is presented in [7]. Authors summarize that compared with Gabor method GGD method achieves a higher accuracy and significantly reduces the computational time. Wavelets were also used in [8]. Authors proposed to use the rotated complex wavelet filters (RCWF) and dual tree complex wavelet transform (DTCWT) together to derive signature feature extraction, which captures information in twelve different directions. In identification phase, Canberra distance measure was used.

Different approach, not based on textures was presented in [9] where the distribution of the pixel gray levels within the line was considered. The curve associated with the gray levels in a stroke section was characterized by use of 4 shape parameters. Altogether, 22 parameters were extracted. Three different classifiers were used with and without genetic selection of the most significant parameters for the classifier. Then the classifiers were combined and the results show the gray level distribution within the writing.

Another direction in off-line writer identification process is using mathematical morphology. In [10] the feature vector was derived by means of morphologically processing the horizontal profiles (projection functions) of the words. The projections are derived and processed in segments in order to increase the discrimination efficiency of the feature vector. Both Bayesian classifiers and neural networks were employed to test the efficiency of the proposed feature. The achieved identification success using a long word exceeded 95%.

Use of neural networks is very popular in the handwritten signature identification process. Paper [11] combines image processing which consists in extracting significant parameters from the signature image and classification by a multi-layer perceptron which uses the previous parameters as input. The image processing step was described according to the intrinsic features of handwriting. Then, the proposed neural networks were compared with others classifiers as pseudo-inverse, k-nearest neighbors and k-means and the influence of preprocessing and bad segmentation was measured. For the identification task, they obtained an error rate of 2.8% when there is no rejection, and an error rate of 0.2% when 10% of the signatures were rejected. Another use of neural networks is

presented in [12]. They started with breaking the pixels into their RGB values and calculating their corresponding gray scale value which are used to train neural network. They implemented the basic algorithm of artificial neural network through back propagation algorithm and used three (Input, output and hidden) layers, six nodes (three in input layer, two in hidden layer and one in output layer). Artificial neural networks (ANN) were also used in [13] where authors presented an off-line signature recognition and verification system which is based on moment invariant method and ANN with back propagation algorithm used for network training. Two separate neural networks were designed; one for signature recognition, and another for verification (i.e. for detecting forgery). Both networks used a four-step process. Moment invariant vectors were obtained in the third step. They reported 100% signature identification accuracy on the small set of 30 signatures. Back propagation neural network and Radial Basis Function Network were used in [14]. The recognition rate of radial basis function was found to be better compared to that of back propagation network. The recognition rate in the proposed system lied between 90% and 100%. Another paper that uses neural networks and combines it with other methods for signature identification is [15]. Authors presented a parameterization system based on angles from signature edge (2D-shape) for off-line signature identification. They used three different classifiers, the nearest neighbor classifier (K-NN), neural networks (NN) and Hidden Markov Models (HMM) and got the best success rate of 84.64% using HMM.

Other approaches to off-line signature identification include use of Support Vector Machine. In [16] a new method for signature identification based on wavelet transform was proposed. This method uses Gabor Wavelet Transform (GWT) as feature extractor and Support Vector Machine (SVM) as classifier. Two experiments on two signature sets were done. The first is on a Persian signature set and other is on a Turkish signature set. Based on these experiments, identification rate have achieved 96% and more than 93% on Persian and Turkish signature set respectively. SVM has also been used in [17]. This work used Support Vector Machines to fuse multiple classifiers for an off-line signature system. From the signature images, global and local features were extracted and the signatures were verified with the help of Gaussian empirical rule, Euclidean and Mahalanobis distance based classifiers. SVM was used to fuse matching scores of these matchers. Finally, recognition of query signatures was done by comparing it with all signatures of the database.

There are other identification methods, but there are only one or two papers that deal with those methods. These include use of Contourlet transform as mentioned in [18]. After preprocessing stage, by applying a special type of Contourlet transform on signature image, related Contourlet coefficients were computed and feature vector was created. Euclidean distance was used as classifier.

Besides that, use of fractals is mentioned in [19]. Advantage was taken from the autosimilarity properties that are present in one's handwriting. In order to do that, some invariant patterns characterizing the writing were extracted. During the training step these invariant patterns appeared along a fractal compression process and then they were organized in a reference base that can be associated with the writer. A pattern matching process was performed using all the reference bases successively. The results of this analyze were estimated through the signal to noise ratio.

One could notice that neural networks are main approach in the off-line signature identification. This is possible because signature identification can be considered as the pattern recognition problem, where neural networks play important role. Their implementation has always been of great interest of the researchers.

2.2.2 On-line signature identification

On-line handwritten signature identification is harder to find in the literature than off-line handwritten signature identification. While on-line signature verification is common subject among biometric community, there are only a few papers on on-line handwritten signature identification.

Hidden Markov Models (HMM) are frequently used during authentication process. Therefore, it would be reasonable to apply this approach to handwritten identification. HMM are usually used for handwritten word recognition, thus it can be applied to on-line signature recognition. Hidden Markov Models are part of the statistical word recognition approach. We cannot find works related to signature identification, but there are few works that deal with handwriting recognition and can be applied in signature identification. Such work is presented in [20] which described a Hidden Markov

Model based writer independent handwriting recognition system. A combination of point oriented and stroke oriented features yields improved accuracy. The general recognition framework is composed of Hidden Markov Models, representing strokes and characters, embedded in a grammar network representing the vocabulary. The main characteristic of the system is that segmentation and recognition of handwritten words are carried out simultaneously in an integrated process.

Another idea in this field is based on Gaussian Mixture Models (GMMs). In [21] the task of writer identification of on-line handwriting captured from a whiteboard was addressed. The system is based on Gaussian mixture models. The training data of all writers are used to train a universal background model (UBM) from which a client specific model is obtained by adaptation. The system is tested using text from 200 different writers. A writer identification rate of 98.56% on the paragraph and of 88.96% on the text line level was achieved.

Those were just ideas what could be implemented to develop on-line handwritten signature identification system. What has been really achieved in this field slightly differs from given ideas. In 2008, a paper that covers the area of signature slant identification was presented [22]. Signatures were captured using a tablet and saved in a digitized format of x and y values. Then it was filtered and calculated for its angle and degree. In the end the signature was classified to its slant category. A slant algorithm was created and coded into a functional system. An experiment consisting of 50 signatures were tested and the finding showed the angle and degree of the slant in every signature. The result was then tested for its accuracy with an available 10 sample of created proofed signatures. The results showed an accuracy of 80% correct slant identification. Authors claim that this algorithm would be able to give some degree of contribution in the area of signature recognition.

Signature direction, slant, baseline, pressure, speed and numbers of pen ups and downs were recognized as the main on-line signature features [23]. While [22] deals with signature slant, paper [23] discussed baseline extraction algorithm for online signature recognition based on vector rules. Signatures were taken from twenty randomly selected individuals with different background. In order to validate the algorithm, the capture image of each signature was used as a sample for a developed questionnaire to be given to human expert. These questionnaires were all about identifying the baseline of the signatures. Both results from automatic baseline detector and the questionnaire were compared, and it showed that the algorithm was 90% accurate. Authors concluded that the algorithm proposed is acceptable to represent extraction of signature features based on baseline.

Presented features can be considered as the global features of given signature since they were applied on the whole signature and not just on some parts of signature. Using some more global feature was presented in [24]. The information was extracted as time functions of various dynamic properties of the signatures. Thirty-one features were identified and extracted from each signature. Different feature reduction approaches and classifiers were used to assess their suitability for this application. Rough set approach has resulted in a reduced set of nine features that were found to capture the essential characteristics required for signature identification. Rough set classifier has achieved 100% correct classification rate using naïve Bayes classifier and Rough set, which demonstrates its suitability and effectiveness for online signature identification.

The last approach that will be mentioned here includes artificial immune theory. Paper [25] presented an approach for online signature recognition which extracts the most commonly used signature features and utilizes the self-learning and self-adaptation of artificial immune theory to obtain new models with higher distinguishability when the training samples are limited. Experiments showed that this approach performs well in sample training and results in satisfactory verification rate and identification rate.

After we presented some of the previous work in the field of handwritten signature identification, it is clear that this problem is still very unexplored and suitable for further development. In the rest of our work we will show a new approach to handwritten signature identification based on graph theory.

3 Signature identification system

If one wants to implement signature identification system to gain more security in the company, one would probably use on-line identification system. System architecture of an ordinary on-line handwritten identification system consists of one main module. It is called identification module and

it is responsible for all the identification logic. This module contains some of the previously described approaches in signature identification or a completely new approach.

System interacts with user by user interface. User is asked to place his or her signature on some kind of specialized gadget. System records signature main data and derives some new data. This data is then passed to the identification module which also requires data from data template storage. Identification module compares signatory data against all templates in the database, thus finding the best match. Person is identified if best match template satisfies certain predefined rules of identification. Simplified signature identification system's architecture is given in Fig. 2.

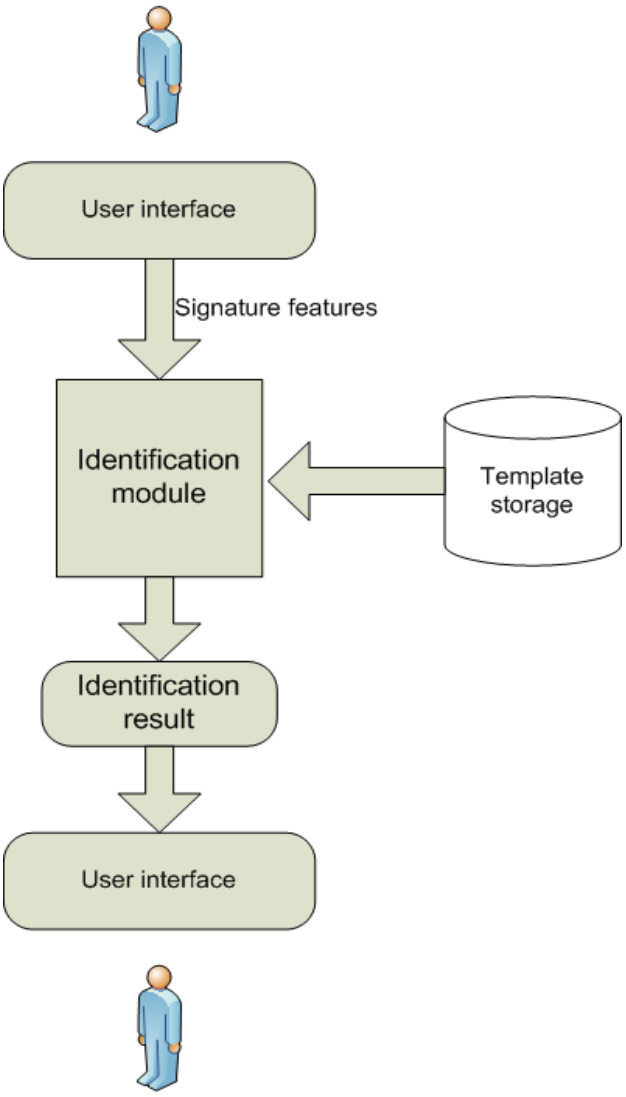


Fig.2. Simplified architecture of the on-line handwritten signature identification system

In our system, user is asked to sign on digitizing tablet. This device sends signal containing three components to the computer. Let *S* be that signal. It is actually three-dimensional vector with the following representation:

$$S = [x, y, z] \quad (1)$$

Where

x – Pen position value on x-axis in (x, y)-coordinate system

y – Pen Position value on y-axis in (x, y)-coordinate system

z – Value of pen pressure (not used in this work)

In this phase of the work we will need just *x* and *y* values. They will be used to construct signature's graph and to calculate some graph features. Coordinate *z* will be used in our future work. If we denote that vector with *S** then we have the following vector representation that will be used in our system:

$$S^* = [x, y] \quad (2)$$

The fact that only those values will be used makes this method suitable even for off-line signature identification, i.e. method can be easily implemented on the handwritten signature image.

Now we can use system architecture shown in Fig.2. It will be divided into three modules:

1. Data acquisition
2. Feature extraction
3. Identification process

3.1 Data acquisition

As mentioned before, our system uses digitizing tablet for data acquisition. Tablet sends packets containing pen position values as well as the pen pressure value. The last one will be disregarded, so we will collect only pen position values. Those will be presented in the tablet specific coordinate system with the origin, i.e. (0, 0) coordinates in the upper left corner of the device. All packets will be stored in the list and can be retrieved when needed.

3.2 Feature extraction

In the data acquisition process some packets are specially monitored. What we need to find in order to continue identification process is the minimum

value of the x-coordinate *Min(X)* and the maximum value of the x-coordinate *Max(X)*. From those values we can calculate signature width (*SW*) which will play important role later in the work. Signature width can be expressed as in (3):

$$SW = Max(X) - Min(X) \quad (3)$$

The second thing that we need to know is the number of strokes in the signature. Stroke is defined as the line drawn from the time since the user puts down the pen to the contact surface until it is filled, i.e. until pen-up occurs. We will denote number of strokes with $N_{strokes}$.

3.2.1 Used concepts of graph theory

With the acquired data we can construct mathematical graph of data. Graph is the main concept of the graph theory. In our case, undirected weighted graphs will be used.

An undirected mathematical graph *G* is an ordered pair *(V, E)* in which *V* represents a set of vertices (nodes) and $E \subseteq V \times V$ is a set of unordered pairs form *V* called set of edges of the graph *G* [26]. In other words, mathematical graph is a set of vertices that are connected by links called edges. Every edge connects only two vertices, and every two vertices can be called adjacent only if they are connected with an edge. To an every edge *E* we can assign some non-negative number *w* which is then referred as the weight of the edge *E*. If all the edges of the graph *G* have weight assigned to them then the graph *G* is called *weighted graph*. How do we assign weights to graph edges will be shown later in the paper.

The first feature that we need to extract from the graph is the information whether the graph is connected or not. Graph is connected if we choose one vertex and traveling along the edges of the graph manage to reach all other vertices of the same graph. In the computer science this can be achieved by implementing 'light versions' of Depth-First Search or Breadth-First Search algorithms which will tell us if they have searched through all the graph vertices. This will be the main factor for the best match scoring in the identification process.

The connectivity of a given graph will be very important in determining next graph features. We want to know is the graph Eulerian or has the graph Euler path (Eulerian trail). Rare mathematical graphs are Eulerian graphs, thus if someone's signature produces that kind of graph it would certainly be of the great influence to the identification process. Graph is Eulerian if and only if all the vertices have an even degree (degree of the vertex is the number of edges connected with the vertex).

Eulerian trail is a trail in a graph which visits every edge exactly once. Graph will have Eulerian trail if and only if it has at most two vertices with an odd degree. Later, from the signature graph will be shown that this is also very rare feature.

Next feature taken in consideration will be the possibility that the graph is Hamiltonian. The problem with Hamiltonian graphs is that there is no simple characterization of Hamiltonian graphs. Because of that we will use one of the characterizations called Dirac theorem. According to this theorem, a simple graph with *n* vertices ($n \geq 3$) is Hamiltonian if each vertex has degree *n/2* or greater. This feature of the signature graph is also very rare, so it could have great impact on the identification result if it is present.

All graph features described so far can be implemented on both non-weighted and weighted graphs. Next in the line are features and concepts that can be implemented on weighted graphs only. First of all we need to calculate overall weight of the graph. This can be achieved if we sum all the weights in the graph. We will denote this feature with W_{Graph}. From each graph, if it is possible, a minimum spanning tree will be calculated. A tree is a connected graph that does not contain cycles. That means that there is a unique path between every two vertices in the graph. Spanning tree *ST* of the graph *G* contains all the vertices of the graph *G* and only some edges of the same graph. To get the minimum cost spanning tree we will use well known Kruskal algorithm. Cost of the spanning tree will be denoted with MST_{Cost}. Since signatures of one person can deviate in width and height, the normalized minimum spanning tree cost should be calculated. This is expressed in (4):

$$N\_MST_{Cost} = \frac{MST_{Cost}}{W_{Graph}} \quad (4)$$

Beside the cost of the minimum spanning tree, it is important to know which edges are in the minimum spanning tree. Our empirical research showed that trained signature should have the same minimum spanning tree edges almost every time when it is consisting from the same stroke number.

Described graph features will be used in the handwritten signature identification process. But, before we can continue with that, we have to know how signature graph is created.

3.2.2 Signature's graph creation

Signature graph is based on the number of the strokes in the signature. Each stroke is a line and has two characteristic points (pen-down point and pen-up point). Therefore, each signature must have at least one stroke. Example of a stroke and its characteristic points is shown in Fig. 3.

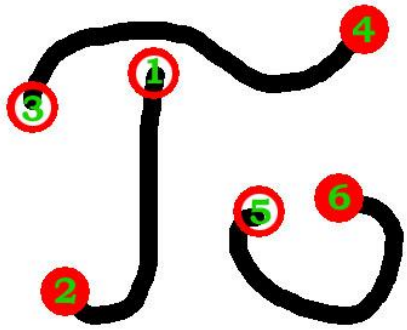


Fig.3. Three strokes from the beginning of the signature

Fig.3. shows three strokes from the beginning of person's signature, along with characteristic points of strokes and the number of the point in respect to time of occurrence. As we mentioned earlier, a stroke is a line from pen-down occurrence (open circle in Fig.3) to pen-up occurrence (closed circle). Number of point is an integer number and it will be very important in the process of creating a signature's graph.

Characteristic points of each stroke will represent graph vertices and will be connected with and edge. But, before we can create a graph it is necessary to divide signature in three segments according to signature's width expressed in (3). Then, the following procedure will be applied in each segment separately. Considering one segment of signature, two vertices will be adjacent if their label divided by 2 produces the same reminder or two vertices are characteristic points of the same stroke. Therefore, in our example we are creating undirected graph and have set of edges $E = \{(1,2), (1,3), (1,5), (2,4), (2,6), (3,4), (3,5), (4,6), (5,6)\}$. Example of graph created form Fig.3. is shown in Fig.4. As we can see, all points with an odd number label are connected between themselves, as well as all points with an even number label and all points that originate from the same stroke.

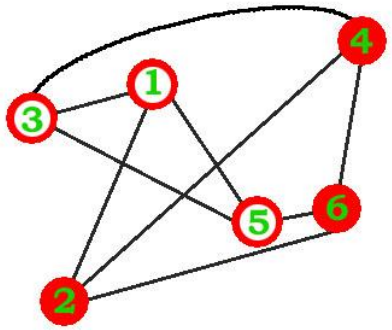


Fig.4. Graph created from strokes in Fig.3

It is possible that in one or more segments there will be no adjacent vertices. This will produce an unconnected graph. To prevent every graph from being unconnected the vertex form the first segment which is the closest to the second segment is connected with the vertex in the second segment which is the nearest to the first segment. Similar to that, second and third segments are connected. If we assume that graph from Fig.4 is just the first segment graph then we must connect it with the second segment if it is possible. This will not be possible only if there are no vertices in the second segment. The connection between the first and the second segment is visible in Fig.5. Even though points with labels 4 and 7 do not have same reminder when divided by 2, they are adjacent because they are in two adjacent segments of person's signature and they are the closest to the segments border (vertical line in the Fig.5).

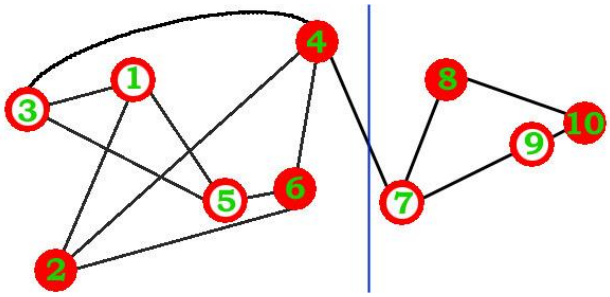
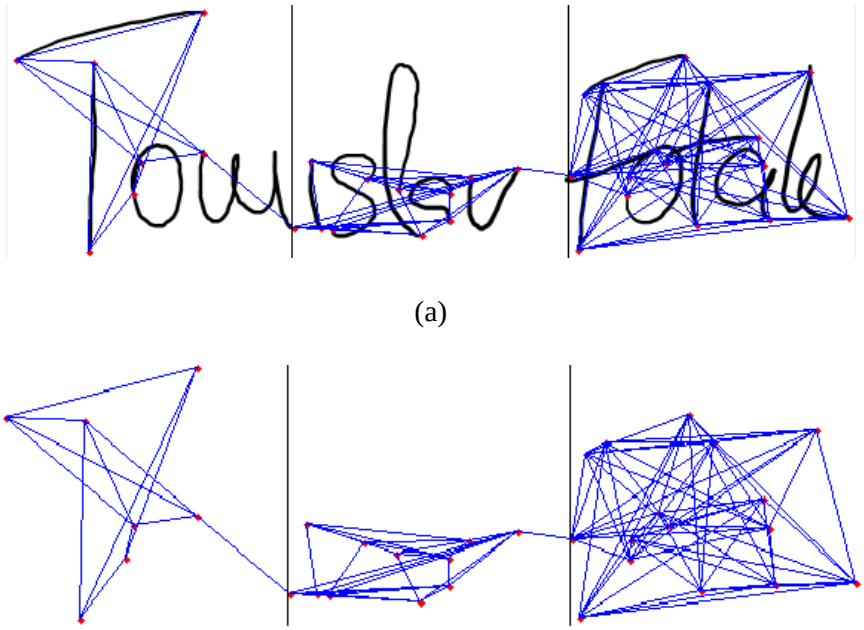


Fig.5. Connection between first and second segment of person's signature

Example of real signature graph in practice is shown in Fig.6. Vertex labels are not shown but it is important to know that they start from zero. This information will be valuable later when determining edges that are in the minimum spanning tree of the given graph.



(a)

(b)

Fig.6. Example of signature graph: (a) with signature in the background; (b) without the signature in the background

Described graph can be referred as the global signature graph. But, in the identification process the local graphs will be used as well. Those are graphs extracted from each of three signature segments. Vertices in each segment get internal label. Vertices from the same stroke, that are in given segment will be adjacent, as well as the vertices whose internal label divided by 2 produces same reminder. Signature graph with its internal labels for every segment is shown in Fig.7 (a). Derived local graphs for every segment are shown in Fig. 7 (b).

From each of these graphs minimum spanning tree will be calculated using Kruskal algorithm. To be able to calculate minimum spanning tree we have to assign weights to all edges of the given graph. There are various ways to determine edge weights. We decided to use digitizing tablet specific coordinate system and combine it with Euclidean distances. This is where our expression (2) comes to a play. Tablet sends packets containing x and y values of the current pen position. Packets that represent pen-down and pen-up events are remembered because they are the first and the last point of a stroke. All those packets are stored in a list of packets and this is where we give them global and internal labels. When determining adjacent vertices of the given graph we use procedure described above. For now, we have set of edges without its weights. To assign weight to an edge Euclidean distance between two adjacent vertices is used.

If a graph has two adjacent vertices v_i and v_j with their coordinates (x_i, y_i) and (x_j, y_j) respectively, weight assigned to edge connecting these two vertices is expressed as follows:

$$w_{i,j} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{5}$$

For example, let us take a look at the vertices in Fig.3. Represented in the tablet coordinate system, vertex one (v_1) has coordinates (105, 65), vertex two (v_2) has coordinates (40, 190) and vertex three (v_3) has coordinates (20, 85). We can see that v_1 and v_2 are from the same stroke. Therefore, they will be connected with an edge. Weight of an edge connecting them will be calculated as in (5): $w_{1,2} = \sqrt{(40-105)^2 + (190-65)^2} = 125$. We can do the same for the vertices v_1 and v_3: $w_{1,3} = \sqrt{(20-105)^2 + (85-65)^2} = 87{,}3212$. We cannot calculate the weight between vertices v_2 and v_3 because they are not connected with an edge.

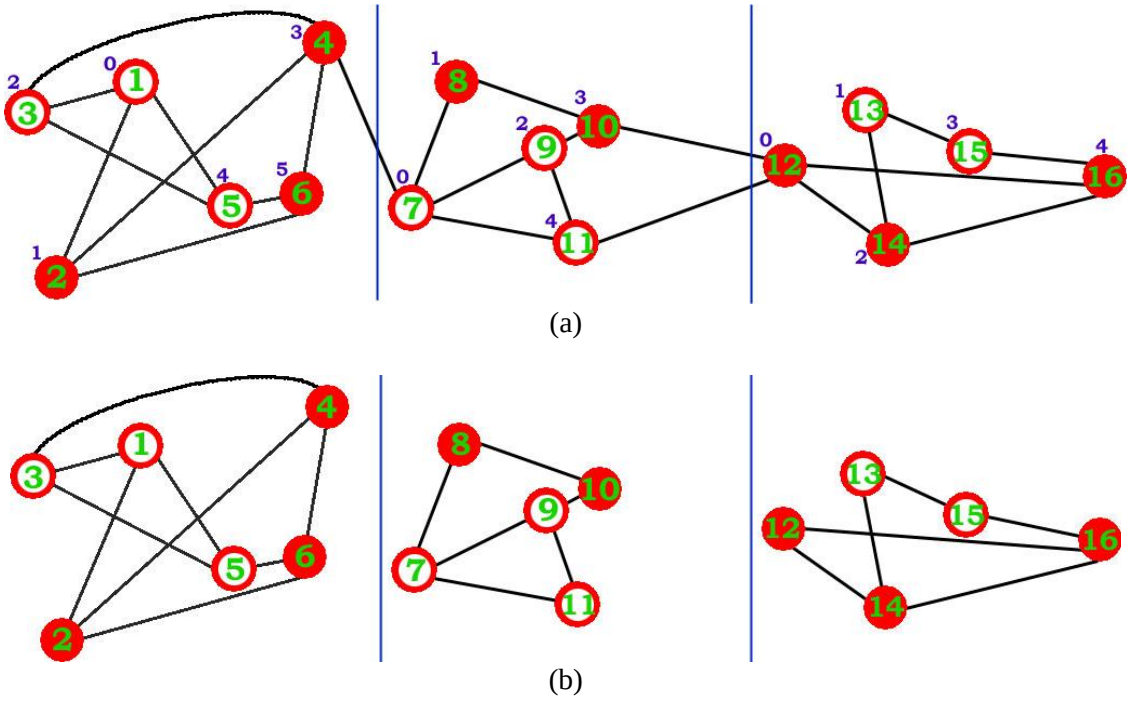


Fig.7. All graphs of our example: (a) Global signature's graph with internal labels of every segment vertices; (b) Signature's graphs of each segment

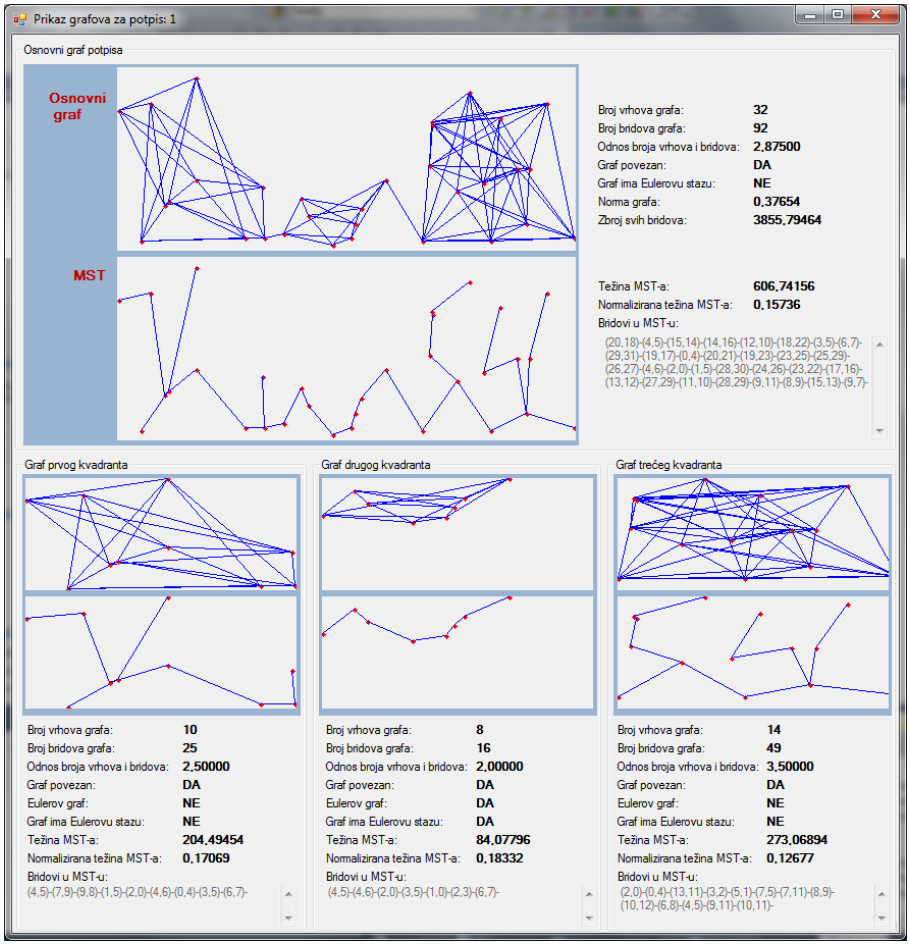


Fig.8. All graphs and theirs features of one signature

To calculate an overall graph weight (W_{Graph}) we have to sum all graph weights. This feature is important in our work.

All graphs and theirs features of one signature are shown in Fig.8. This figure is a screenshot of the software window developed for the purpose of this work. This window is shown upon user registration in the system. Since it is in Croatian, we will now explain what it shows. Global graph of a signature is referred as the 'Osnovni graf'. This is the main graph we constructed following the method described above. Below this graph we can find minimum spanning tree of given graph ('MST'). On the right side of these graphs are graph features shown. They are described in the Section 3.2.1. We show number of vertices (32 vertices), number of edges (92 edges), edges to vertices ratio (2,875), is the graph connected (DA, this means YES), has the graph Eulerian trial (NE, this means NO), graph norm (0,3765; will be described later in the paper), overall graph weight (3855,79) as well as the minimum spanning tree specific features, such as minimum spanning tree overall cost (606,74), normalized minimum spanning tree cost (0,15736) according to (4) and finally all edges in minimum spanning tree (referred as 'Bridovi u MST-u). Below these data and global graph window is divided into 3 parts. Every part represents local graph and corresponding minimum spanning tree along with the graph features of one graph segment. First, second and third segments are referred as 'Graf prvog kvadranta', 'Graf drugog kvadranta' and 'Graf trećeg kvadranta' respectively.

Upon user registration, it is recommended that user signs itself 15 times, but no less than 10 times. Each signature has its own graphs, so it is possible to store in database up to 60 graphs for one user (up to 15 global and 45 local graphs). In the identification process, all stored graphs will be taken in consideration when finding the best match.

3.3 Identification process

In this process, calculated graph features of given signature have to be compared against all signature templates in database to find the best match. This can be very time consuming. To prevent comparing against all templates and to speed-up the system, it is possible to make fast classifier (filter) that would extract just those users that satisfy certain rule needed to continue identification.

3.3.1 Graph norm

Graph norm (*N(G)*) is a single real number that represents the whole signature graph. Since graph depends on the number of the strokes in the signature and the width of the signature, these two characteristics will make a base for graph normalization. We will normalize the overall weight of the graph. This is the feature that varies depending on the stroke number and the signature width. Using (3), the graph norm can be represented as following:

$$N(G) = \frac{W_{Graph}}{N_{strokes} \cdot SW} \tag{6}$$

The person with the trained handwritten signature will always have almost the same value of graph norm. This is what makes this feature so valuable in the identification process.

To be able to use this feature, during the registration process, mean, standard deviation, minimum and maximum value of this feature should be recorded.

3.3.2 Identification procedure

After the signatory placed his or her signature on the digitizing tablet, the identification procedure can start. Given signature is being classified according to the graph norm (mean ± 1 standard deviation) and stroke number. If the system matched given features with just one user, it suggests that user as a potential signatory and the identification process is over. But this would be very rare if we have database that contains more than ten users. If more than one user were recognized as potential signatories, the identification procedure should continue with comparing user graph against all the graphs of potential users. This is done separately for each potential user. Each graph of the registered signature of the potential user is being compared against the given signature graph. Comparison is done separately for global and local graphs.

For the identification purposes of finding the best match, the score system was developed. Maximum points that potential signature from database can score against given signature is predefined and it consists from points for global graph and points for local graphs. Each graph feature has also predefined percentage in overall points. If the graph is not connected, features such as Eulerian graph, Eulerian trial, Hamiltonian graph, MST value, Normalized MST value and MST edges will not be taken in

consideration. Therefore, other features will have greater percentage in the overall points. If the graph is connected then all the features are taken in consideration. How much points will each feature gain depends on deviation of the stored signature feature from the given signature feature.

At the end of identification procedure for one potential user, statistical indicators of all his or her results are obtained. The best match is found in the user whose results have the biggest arithmetic mean of all the results obtained.

3.4 System results

The proposed identification system was tested on the small base of 27 users. Each user has to provide 15 signatures upon registration in the system. Overall of 400 identification tests on live users were made. In 377 cases method correctly identified the user, in 17 cases it identified the wrong user but the right user was suggested as potential after the fast classification phase and in 6 cases no user was suggested in the fast classification phase. This yields 94.25% identification accuracy. This is promising result for the further development of the proposed identification system.

4 Conclusion

In this paper we presented a new approach in on-line handwritten signature identification. It is based on graph theory which is already proven as a good tool for some biometric systems that use other biometric characteristics (e.g. face recognition). The fact that it is using graphs as main concept makes it suitable for implementation even in the off-line handwritten identification systems. The results of the proposed identification method are very promising and show the great potential of graph theory in the field of handwritten signature identification. Our method is not perfect and we are aware of that. So far, it implements only some basic concepts of graph theory. Since this field is very wide, our future work will be directed to find more suitable graph features for handwritten signature identification. We will also try to increase the security of the proposed system by combining it with the basic on-line signature features such as presented in [27]. In this way the security will be increased because the identified person should also pass the verification test, i.e. his or her on-line handwriting features must match those in the database. We will also try to include the coordinate *z* in our calculations of graph features. This will enable us more accurate system but it will no longer be available for easy use in off-line signature recognition systems. It would be needed to develop a method that can represent image pixels as the pen pressure. For now, we have given a good base for our future work.

5 Acknowledgments

Shown results come out from the scientific project Methodology of biometrics characteristics evaluation (016-01611992-1721) and technological project Multiple biometric authentication using smart card (2008-043) financed by the Ministry of Science, Education and Sport, Republic of Croatia.

References:

[1] V. Anić et al., *Hrvatski enciklopedijski rječnik*, in Croatian, Novi Liber, 2002.

[2] A.K. Jain, A. Ross, S. Prabhakar, An Introduction to Biometric Recognition, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol.14, No.1, 2004, pp. 4-20.

[3] I. Pavlidis, N.P. Papanikolopoulos, R. Mavuduru, Signature identification through the use of deformable structures, *Signal Processing*, Vol.71, No.2, 1998, pp. 187-201.

[4] M.K. Kalera, S. Srihari, A. Xu, Offline Signature Verification and Identification Using Distance Statistics. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol.18, No.7, 2004, pp. 1339-1360.

[5] H.E.S. Said, T.N. Tan, K.D. Baker, Personal identification based on handwriting, *Pattern Recognition*, Vol.33, 2000, pp. 149-160.

[6] Y. Zhu, T. Tan, Biometric personal identification based on handwriting, *Pattern Recognition*, Proc 2, 2000, pp. 797-800.

[7] Z. He, X. You, Y.Y. Tang, B. Fang and J. Du, Handwriting-based personal identification. In Proceedings of IJPRAI. 2006, pp. 209-225.

[8] M.S. Shirdhonkar, M. Kokare, Off-line Handwritten Signature Identification Using Rotated Complex Wavelets Filter, *International Journal of Computer Science Issues*, Vol.8, Issue 1, 2011, pp. 478-482.

[9] M. Virotius, A. Seropian, N. Vincent, Writer Identification From Gray Level Distribution, *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, 2003, pp. 1168-1172.

[10] E.N. Zois, V. Anastassopoulos, Morphological waveform coding for writer identification, *Pattern Recongition*, Vol.33, 2000, pp. 385-398.

[11] I. Pottier, G. Burel, Identification and Authentication of Handwritten Signatures with a Connectionist Approach, *1994 IEEE International Conference on Neural Networks*, Vol.5, 1994, pp. 2948-2951.

[12] D. Bhattacharyya, T. Kim, Design of Artificial Neural Network for Handwritten Signature Recognition, *International Journal Of Computers and Communications*, Vol.4, No.3, 2010, pp. 59-66.

[13] C. OZ, F. Ercal, Z. Demir, Signature Recognition and Verification with ANN, available at Elektrik Mühendisleri Odası web http://www.emo.org.tr/ekler/8b7dc6e8b36bcaa_ek.pdf, Accessed: 27th March 2011.

[14] J. Ashok, E.G. Rajan, Writer Identification and Recognition Using Radial Basis Function, *International Journal of Computer Science and Information Technologies*, Vol.1, No.2, pp. 51-57.

[15] J.C. Briceño, C.M. Travieso, M.A. Ferrer, J.B. Alonso, F. Vargas, Angular contour parameterization for signature identification, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, pp. 358-365.

[16] M.H. Sigari, M.R. Pourshahabi, H.R. Pourreza, Offline Handwritten Signature Identification using Grid Gabor Features and Support Vector Machine, *16th Iranian Conference on Electrical Engineering*, 2008, pp. 281-286.

[17] D.R. Kisku, P. Gupta, J.K. Sing, Off-line Signature Identification by Fusion of Multiple Classifiers using Statistical Learning Theory, *International Journal of Security and Its Applications*, 2010, pp. 35-45.

[18] M.R. Pourshahabi, M.H. Sigari, H.R. Pourreza, Offline Handwritten Signature Identification and Verification Using Contourlet Transform, *2009 International Conference of Soft Computing and Pattern Recognition*, 2009, pp. 670-673.

[19] A. Seropian, M. Grimaldi, N.Vincent, Writer Identification based on the fractal construction of a reference base, *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, 2003.

[20] J. Hu, S.G. Lim, M.K. Brown, Writer independent on-line handwriting recognition using an HMM approach, *Pattern Recognition*, Vol.33, 2000, pp. 133-147.

[21] A. Schlapbach, M. Liwicki, H. Bunke, A writer identification system for on-line whiteboard data, *Pattern Recognition*, Vol.41, 2008, pp. 2381-2397.

[22] M.R. Shamsuddin and A. Mohamed, Online signature slant feature identification algorithm. *WSEAS Transactions on Computer Research*, Vol.3, No.3 2008, pp. 121-130.

[23] A. Mohamed, R. Yusof, S.A. Rahman, S. Mutalib, Baseline extraction algorithm for online signature recognition, *WSEAS Transactions on Systems*, Vol.8, No.4, 2009, pp. 491-500.

[24] W. Al-Mayyan, H.S. Own, H. Zedan, Rough set approach to online signature identification, *Digital Signal Processing*, Vol.21, No.3, 2011, pp. 477-485.

[25] Z. Lisha, S. Zhengxing, Online Signature Recognition Using Artificial Immune Model, *Journal of Computer Aided Design and Computer Graphics*, Vol.19, No.3, 2007, pp. 311-317.

[26] B. Divjak, A. Lovrenčić, *Diskretna matematika s teorijom grafova*, in Croatian, TIVA-FOI, 2005.

[27] M. Bača, P. Koruga, T. Fotak, Basic on-line handwritten signature features for personal biometric authentication, *2011 Proceedings of the 34th International Convention MIPRO*, 2011, pp. 116-120.