# HANDWRITING IN FORENSIC INVESTIGATIONS

Georgi Gluhchev

*Abstract: The process of automatic handwriting investigation in forensic science is described. The general scheme of a computer-based handwriting analysis system is used to point out at the basic problems of image enhancement and segmentation, feature extraction and decision-making. Factors that may compromise the accuracy of expert's conclusion are underlined and directions for future investigations are marked.*

*Keywords: Handwriting, Identification, Verification, Screening, Forensic Investigation, Image Enhancement, Feature Extraction, Decision-making.*

## 1. Introduction

The individual's authentication becomes a serious problem concerning different areas of the social and economic relations in the world. Its importance increases as a factor in the prevention of terrorist actions and illegal access to important information. The problem attracts the attention of researchers all over the world and during the last years a few international projects were launched aimed at the development of reliable systems for authentication using biometric information.

Writer authentication is one of the broadly investigated modality in this aspect. Until now it was mainly used in forensic investigations dealing with handwritten document analysis or signature verification. Despite of the long history in that respect handwriting investigation still remains a difficult, time-consuming and subjective process, where qualified experts evaluate the similarity between letters, strokes and writing styles on the basis of their experience. In all cases of writer identification the objectiveness of the analysis and reliability of conclusion are of great importance. However, the inevitable variation in writing under different conditions and psychological state or when a significant time gap exists between the incriminated and reference documents may mislead the expert. Also, extremely difficult are cases where the handwriting is deliberately changed. In such situations different experts may disagree as to who is the writer of a particular document and a wrong conclusion may be drawn.

Other problems in writer identification concern the expert's workload during the analysis and difficulties stemming from sometimes poor quality of handwritten materials.

To overcome the problems quantitative methods for objective handwriting analysis and adequate decision-making have to be developed and implemented.

To achieve this, serious scientific investigation is required in order to develop appropriate methods for feature measurement, selection of reliable sets of features, evaluation of the minimal number of handwriting elements which is necessary for the reliable decision-making, suggestion of robust classification algorithms dealing with mixtures of continuous and categorical variables. Major difficulties in this direction stem from the qualitative character of most of the handwriting parameters used by the experts.

Despite that the problem of writer identification is of great practical importance in forensic investigations a relatively small number of papers have been published until now [5,11,14,22]. The existing computer systems are aimed especially at the screening of similar handwritings from a large data base of handwritings. After that the identification problem is carried out manually by an expert on the basis of his own experience and subjective evaluation of the similarity between the handwritings under investigation. For this he compares visually strokes, letters or combinations of letters performing sometimes simple measurements.

The goal of the paper is to sketch the frame of a computer-based handwriting investigation system and discuss the problems of its major components (Fig. 1). It is organized in the following way: Section 2 concerns the improvement of image quality; Section 3 deals with feature extraction techniques; Section 4 describes decision-making and Section 5 points out at the unsolved problems.

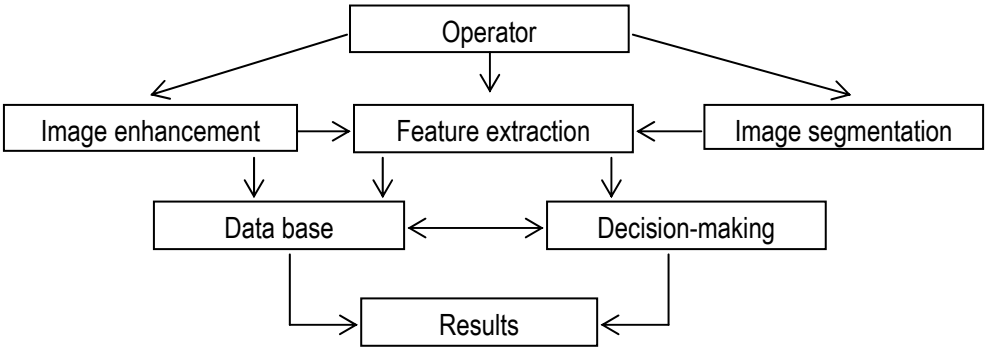


Figure 1. Block-diagram of a handwriting investigation system

## 2. Image Enhancement

The block-diagram in Fig. 1 shows that the development of a computerized handwriting investigation system follows the general methodology for the development of image processing and pattern recognition systems.

Since very often the handwritten materials are of poor quality, it is necessary to achieve some pre-processing. Its goal is two-fold: a) to improve image quality including contrast enhancement, random and structured noise reduction, and edge sharpening [2,3,9,17,19,21]. In such a way the image will offer better possibilities for the automatic analysis; b) to correct strokes and complex lines using morphological operations. This is especially important for the analysis of specific handwriting features, where the skeleton of the characters is used. Morphology allows to automatically connecting disrupted lines or cutting-off wrongly connected strokes.

## 3. Image Segmentation

Image segmentation is an essential step in the automatic document analysis. In handwriting analysis additional difficulties may arise due to the possibility of significant variation in rows', words' and letters' position.

The automatic segmentation includes background elimination as a first stage. Approaches based on histogram analysis (uniform background is supposed), locally adaptive binary trees and heuristic approaches (non-uniform background or presence of structured noise is assumed) are used.

The second stage concerns separation of rows. This operation is based on the analysis of histogram-like graphs obtained via horizontal projections of pixels, Hough transform or analysis of envelopes of continuous lines. While this operation could be easily achieved, special techniques are required for a proper detection of under-row and over-row elements of some characters. Also, the medial axis of the row may not be presented as a single straight line which requires piece-wise presentation.

The segmentation continues with the separation of the words. Different cases of concise writing or writing where letters are not connected between them are a challenge. For the word separation vertical projections of pixels from the corresponding row are analyzed, distances between envelopes of continuous lines are used, separation lines parallel to the predominant slope of the vertically oriented strokes could be applied, as well.

However, the most difficult problem concerns segmentation of letters and strokes. Except some special cases, e.g. child's writing, their delineation may be quite difficult even for a human being. A proper solution of this problem could be achieved via a man-machine dialog. The operator has to identify some specific points like end-points or vertexes of a polygon that encompasses a handwriting element. After that lines could be automatically investigated for the detection of points of intersection or bifurcation, local extremums and like [12,13,20].

## 4. Feature Extraction

Feature extraction is the crucial problem that should be solved. While during the last decades a common methodology for the handwriting analysis has been set up, many of the suggested features are of qualitative

character and are prone to different evaluation from different experts. Also, there are no strong recommendations as to what number of features is to be used for a reliable decision-making.

Different types of features may be investigated including graphometric, densitometric, categorical, model-based and topological invariants [2,5,6,7,8]. From the expert's point of view they are classified as general and specific features.

The general features are of categorical type and describe qualitative characteristics as: degree of connection between letters (usually three degrees are accepted: low, moderate and high), slope (right, left, upright), motion (rectilinear, curvilinear, angular or arched, loop-like, oval, wavy or spiral), elaboration (presence of ornaments), direction of movement (clock-wise or counter-clock-wise), quantity of movement (average number of strokes used to draw separate letters) and like [2,7,8,12,16]. They are difficult for automatic evaluation and are specified by the expert.

Specific features admit quantitative evaluation. They are known as graphometric and are aimed at the automatic or semi-automatic measurement of the following characteristics: distances between rows, height and width of letters, distances between letters, size of the above-row and under-row elements, distances between words, predominant slope, geometric parameters of handwriting elements like strokes, fragments and/or combination of characters [2,11,22].

Figure 2. a) Original image, b) Areas of different pressure

While many of the above-mentioned features can be easily imitated, there are features that are not clearly seen and, therefore, difficult to falsify. In that respect a special attention is paid to the distribution of pressure alongside the strokes (Fig. 2). It could be analyzed in different ways looking for a reliable description, e.g. evaluation of the geometric parameters of areas of different pressure and their mutual disposition at different writing elements, or pressure change alongside the skeleton of the elements. In that respect promising results have been reported in [15].

Except the described features, which are reasonable and intuitively clear for the experts, other features that do not express a particular property of the handwriting may be used as well. These include topologically invariant points associated with a particular character. According to this approach characters are divided into specific segments that can be transformed and compared piece-wise.

Another approach is based on the presentation of characters as elastic changes of an ideal model. Thus, a transformation between the model and the real character can be evaluated and its parameters used for classification.

Very important is the problem concerning the reliability evaluation of different set of features. A number of approaches may be used for this based on the information theory, statistics and classification power.

## 5. Decision-Making

The overall estimation of the similarity between two handwritings must be obtained as a combination of decision-making classifiers.

The decision-making for the specific features is based on the evaluation of the similarity between particular elements from the handwritings under investigation. Since the overall estimation will be based first on the estimations of separate elements and second on groups of elements, multi-level classifiers are to be used. The first level will concern the comparison of basic elements like strokes, letters and signs of punctuation. At the

output of these classifiers every element will be assigned a number that reflects the degree of similarity between the handwritings under investigation. Since a particular element may be detected in a few places in the text, an average similarity relative to this element will be calculated at the second stage. After the similarity is evaluated for all different elements, an overall evaluation is being obtained at the third level. One of the basic problems that has to be solved here concerns the weight factors of the elements, i.e. their classification power. Different types of decision rules could be used, including statistical, linear, heuristic, and NN-based [1,4,6,18].

The authenticity, where a forgery is expected, would be predominantly verified using stable features like pressure distribution. This is usually applied to small pieces of written text like particular words or signatures. Depending on how the pressure will be measured (areas of different pressure or a function alongside a skeleton line) different comparison techniques could be applied.

The categorical features are mainly used for the search of similar handwritings in a large database. Also, for the sake of one-to-one comparison mixed variables discriminant techniques could be used. A simple approach for the analysis of mixture of categorical and continuous data requires arbitrary scoring of all the categorical variables followed by the use of standard methods for multivariate continuous data, which in the case of classification means the use of techniques such as linear or quadratic discriminant analysis.

## 6. Discussion and Future Work

Different aspects and major problems that are to be solved for the development of a computer-based handwriting investigation system are described.

While the pre-processing stage of digital images is thoroughly investigated during the last decades, the well-known approaches may not work properly due to possible damages, background variation and/or poor image contrast. This requires locally adaptive methods to be developed, reflecting the specificity of the investigated images.

A big challenge is the selection of a reliable set of features. A computerized system must include as much as possible features that experts are accustomed to, but at the same time, special attention must be paid to the measurement of some parameters that are difficult for expert's evaluation, and therefore difficult for imitation, e.g. curvature at characteristic points, line smoothness or pressure distribution. Also, the expert has to have the possibility to interfere and suggest his one selection of features. This requires a friendly man-machine interface to be available.

The decision-making seems to be the most expert-independent part of the problem, since various objective measures of similarity (parametric, non-parametric, clustering) have been developed in pattern recognition theory. The different levels of similarity estimation will require the development of multi-level hierarchical classifiers.

Age-due variations in handwriting or changes due to different diseases must be investigated as well. This will increase the possibility for reliable writer identification when a significant time-gap between the handwritten materials exists or in case of a psychiatric disease [10].

A successful solution to all of the discussed problems will allow developing of a reliable and user-friendly computer system for handwriting analysis that could be implemented in police departments, bank and notary offices. It must be noted that such a system will help the expert to do an objective analysis, not to replace him.

The obtained solutions to specific handwriting analysis problems could be easily incorporated in a more complicated access-permit systems or person authentication systems at check points.

## Acknowledgements

This work is supported by the Ministry of Education and Sciences under contract # *И* 1302/2003.

## References

[1] Alexandre, Luís A., A. C. Campilho, M. Kamel. "Combining Unbiased and Independent Classifiers Using Geometric Mean", 11th Conf. of Portuguese Association on Pattern Recognition, 2000

[2] Angelov, A., D. Nestorov, S. Benchev, G. Gluhchev, D. Kamenov, P. Veleva. "A system for the automated handwriting investigation", In: Information Bulletine of RIFSC, Sofia, 1997, pp. 127-133

[3] Dimov, D. T., G. Y. Gluhchev. “A Locally Adaptive Binary-Tree Method for Binarization of Text Images”, In: Progress in Handwriting Recognition, (Edts A.C. Downton, S. Impedovo), World Scientific, London, 1996, pp. 575-580

[4] Gluhchev, G., V. Shapiro, D. Lalchev. "Probabilistic estimate of handwriting similarity", In: Theory and application of cybernetic systems, Bulgarian Academy of Sciences, Sofia,1991, 3,pp. 90-93

[5] Gluhchev, G., V. Shapiro, S. Ogorelkov. "Automatic investigation of handwritten characters", In: Forensic science and criminology, Sofia, 1989, 5,pp. 135-142

[6] Heutte, L., T. Paquet, J.V. Moreau, Y. Lecourtier and C. Olivier. "A Structural/Statistical Feature Based Vector for Handwritten Character Recognition", Pattern Recognition Letters, vol. 19, no. 7, pp. 629-641, 1998.

[7] Heutte, L., J.V. Moreau, B. Plessis, J.L. Plagnaud and Y. Lecourtier. "Handwritten numeral recognition based on multiple feature extractors", 2nd IAPR International Conference on Document Analysis and Recognition, IAPR-ICDAR'93, Tsukuba, Japan, IEEE Computer Society Press, pp. 167-170, 1993.

[8] Heutte, L., J.V. Moreau and Y. Lecourtier. "A new feature extraction strategy for handwritten pseudo-character recognition based on multiple feature extractors", 6th IGS International Conference on Handwriting and Drawing, IGS-ICOHD'93, Paris, France, pp. 186-188, 1993.

[9] Hummel R.A. “Image enhancement by histogram transformation”, Comput. Graphics and Image Processing, 6, 1976, pp. 184-197

[10] Kalcheva, E., G. Gluhchev. “Evaluation of handwriting changes in case of neurological diseases”, Proceedings of the CompSysTech ’2000, Sofia, pp.V.2-1–V.2-5

[11] Klement, V. “An Application System for the Computer Assisted Identification of Handwritings”, Proc. Int. Carnahan Conf. Security Technology, Zurich, 1982, pp. 75-79

[12] Lickforman-Sulem, L. “Extraction d’elements graphiques dans les images de manuscripts, CIFED’98, Quebec, Canada, 1998, pp. 198-207

[13] Lickforman-Sulem, L. and C. Faure. “A Hough Based Algorithm for Extracting Text Lines in Handwritten Documents”, ICDAR’95, Montreal, 1995, pp. 774-777

[14] Lima, Rui M., A. J. C. Campilho. "A System for Image Analysis of Documents", 7th Conf. of Portuguese Association on Pattern Recognition, pp. 1.5.1-1.5.6, 1995.

[15] Nestorov, D., V. Shapiro, P. Veleva, G. Gluhchev, A. Angelov, I. Stoyanov. "Towards objectivity of handwriting pressure analysis for static images", 6th Int. Conf. on Handwritting and Drawing ICOHD '93, 5-7 July, Paris, pp. 216-218

[16] Nosary, A., L. Heutte, T. Paquet and Y. Lecourtier. "Defining writer's invariants to adapt the recognition task", 5th International Conference on Document Analysis and Recognition, ICDAR'99, Bangalore, India, IEEE Computer Society Press, pp. 765-768, 1999.

[17] Pizer, S.M., E.P. Amburn, J.D. Austin et al. “Adaptive hiastgram equalization and its variations”,Comput. Vision, Graphics and Image Proc., 39, 1987, 355-368

[18] Plessis, B., A. Sicsu, L. Heutte, E. Menu, E. Lecolinet, O. Debon and J.V. Moreau. "A multi-classifier combination strategy for the recognition of handwritten cursive words", 2nd IAPR International Conference on Document Analysis and Recognition, IAPR-ICDAR'93, Tsukuba, Japan, IEEE Computer Society Press, pp. 642-645, 1993.

[19] Pratt W. K. Digital Image Processing, 2nd edn, John Wiley & Sons, 1991

[20] Shapiro, V., G. Gluhchev, V. Sgurev. "Handwritten document image segmentation and analysis", Pattern recognition letters, North Holland, 1993, 14,pp. 71-78

[21] Shapiro, V., G. Gluhchev, V. Sgurev. "Preprocessing for automatic examination of handwritten documents", 7th Scandinavian conf. on image analysis, Denmark, 1991,v.2, pp. 790-797

[22] Steinke, K. “Entwicklung von Mustererkennungsverfahren zur textunabhangigen Analyse von Handschriftenbildern”, Dissertation TH Aachen, 1981

## Author Information

Georgi Gluhchev – Institute of Information Technologies, BAS, Acad. G. Bonchev Str., bl.2, Sofia - 1113, Bulgaria; e-mail: gluhchev@iinf.bas.bg