2nd International Workshop on Automated Forensic Handwriting Analysis (AFHA) 2013

22-23 August 2013, Washington DC, USA

AFHA 2013

ONLINE PROCEEDINGS

PREFACE

Handwriting is considered as a representative of human behavior and characteristics for centuries. With the evolution of modern computing technologies, researchers have moved towards the automated analysis of handwriting. This shift has been reinforced by the interest various industries have in this field. One of the most important applications of automated handwriting analysis systems is in forensic environments. Until now, most of the forensic handwriting analysis cases are solved without actual application of automated systems. This is because there is an ever increasing gap between the demands of Forensic Handwriting Experts (FHEs) and the computer science community. Actually the underlying issue is the incapability of most of the state-of-the-art automatic handwriting examination systems to be directly applicable to the forensic cases. This is because the computer science community in general has moved by considering the cases which are either trivial w.r.t. forensic situations or not considered the needs of FHEs. Thus, there is a great demand to bring the forensic experts and the computer science experts under one roof. The 2nd International Workshop and Tutorial on Automated Forensic Handwriting Analysis (AFHA) 2013, like its predecessor AFHA 2011, serves this purpose.

The AFHA 2013 takes place on 22-23 August 2013, in Washington DC, USA, and is organized as a two-day combined workshop and tutorial covering a diverse range of topics influencing handwriting analysis in the forensic science and in computer science (particularly, in pattern recognition).

On the first day, an introductory tutorial on forensic handwriting examination and automatic handwriting/signature analysis is given. This includes a description of the forensics point of view and examples of real casework as well as a summary of important approaches in the area of automated handwriting examination. The major topics include: how forensic experts make comparisons (similarities versus differences, subjectivity, and bias), natural variation, line quality, quality versus quantity; what forensic experts need from the document analysis community; what the document analysis community needs to understand about FHEs work; existing systems and system problems; application of the Bayesian approach to forensic evidence evaluation (i.e. using the Likelihood Ratios a measure of the strength of evidence), and reporting by means of a verbal conclusion scale. The state-of-the-art of automatic handwriting/signature analysis systems is also focused where the emphasis is on the internal working of these systems along with the future directions in this regard. The purpose is to familiarize the forensic experts about working of automatic systems.

On the second day, the workshop is organized where researchers from handwriting examination and pattern recognition communities present their novel researches. This volume contains the proceedings of the AFHA 2013 workshop. Thirteen submissions were received and after a single-blind-peer review process, ten papers were accepted for this volume.

The first paper, '*Some Observations on Handwriting from a Motor Learning Perspective'* discusses the dynamics of signatures in the light of recent findings in motor learning, according to which a signature is a highly automated motor task and, as such, it is stored in the brain as both a trajectory plan and a motor plan. It conjectures that such a stored representation does not necessarily include the entire signature, but can be limited to only parts of it, those that have been learned better and therefore are executed more automatically than others.

The second paper, '*Offline Handwriting Acquisition under Controlled and Uncontrolled Conditions*' discusses the offline handwriting acquisition under controlled and uncontrolled conditions for research purposes. The paper emphasizes that for forensic purposes, it is preferred to start building databases with forensically relevant data. This is because handwriting samples that make up the current publicly available databases have all been collected under controlled conditions.

The third paper '*Oriented Local Binary Patterns for Writer Identification*' presents an oriented texture feature set, based on local binary patterns (LBP), and apply it to the problem of offline writer identification using the ICDAR 2011 and ICHFR 2012 writer identification contest datasets.

The fourth paper '*Chinese Handwritten Writer Identification based on Structure Features and Extreme Learning Machine*' proposes an approach for writer identification of Chinese handwriting using Chinese character structure features (CSF) and extreme learning machine (ELM). To extract the features embedded in Chinese handwriting characters, special structures have been explored according to the trait of Chinese language.

The fifth paper '*Dissimilarity Representation for Handwritten Signature Verification*' discusses the dissimilarity representation (DR) approach where proximity among patterns constitute the classification space. The paper provide various scenarios where similar concept has been applied by forensic Questioned Document Examination (QDE) experts, when proximity between questioned signatures and a set of templates lead to the authentication decision.

The sixth paper '*Multi-script Off-line Signature Verification: A Two Stage Approach*' presents a technique for off-line English, Hindi (Devnagari), and Bangla (Bengali) signature verification by initially identifying the script type and then applying verification. This paper highlights that better results could be achieved when the script is identified in advance.

The seventh paper '*Off-Line Signature Verification based on Ordered Grid Features: An Evaluation*' presents and evaluates an offline signature modeling which attempts to advance a grid based feature extraction method uniting it with the use of an ordered power set. More specifically, this work represents the pixel distribution of the signature trace by modeling specific predetermined paths having Chebyshev distance of the two, as being members of alphabet subsets-events.

The eighth paper '*Towards Automated Hyper-spectral Document Image Analysis*' provides an overview of the applications of hyper-spectral imaging with focus on solving pattern recognition problems, especially handwriting analysis and signature verification.

The ninth paper '*Fusing Modalities in Forensic Identification with Score Discretization*' proposes a method of score fusion based on discretization. It is evaluated considering the signatures and fingerprints.

The tenth paper '*Joint Glossary of Forensic Document Examination and Pattern Recognition*' introduces an open scientific glossary, based on the MediaWiki engine, to the forensic examination and pattern recognition communities. The purpose is to enable the development of a shared conceptualization among the two communities.

We would like to thank the authors for their paper submission, our program committee members for their reviews and active participation in various activities concerning tutorial and workshop, and the AFHA 2013 workshop chairs for their advice and guidance throughout the endeavor.

The AFHA 2013 PC-chairs,

August 2013.

Committees

Program and Organization Chairs

Muhammad Imran Malik,
German Research Center for Artificial Intelligence, Kaiserslautern, Germany

Marcus Liwicki,
German Research Center for Artificial Intelligence, Kaiserslautern, Germany
University of Fribourg, Switzerland

Linda Alewijnse,
Netherlands Forensic Institute, The Hague, the Netherlands

Michael Blumenstein,
Professor, Griffith University, Southport QLD 4215, Australia

Charles E.H. Berger,
Netherlands Forensic Institute, The Hague, the Netherlands

Reinoud D. Stoel,
Netherlands Forensic Institute, The Hague, the Netherlands

Bryan Found,
Chief Forensic Officer, Victoria Police Forensic Services Department, Australia

Program Committee

Angelo Marcelli, U Salerno

Giuseppe Pirlo, U Bari

Javier Ortega-Garcia, U A Madrid

Julian Fierrez, U A Madrid

Katrin Franke, NIS Labs

Loris Nanni, U Bologna

Miguel Ferrer, ULPGC

Réjean Plamondon, E P Montreal

Sargur N. Srihari, U Buffalo

Takashi Matsumoto, Waseda U

Wataru Ohyama, Mie U Japan

Xiaohong Chen, China

Zeno Geradts, NFI

TABLE OF CONTENTS

Some Observations on Handwriting from a Motor Learning Perspective

Angelo Marcelli, Antonio Parziale, Rosa Senatore
Natural Computation Laboratory, DIEM
University of Salerno
Fisciano (Sa), Italy
{amarcelli, anparziale, rsenatore}@unisa.it

***Abstract*—We discuss the dynamics of signatures in the light of recent findings in motor learning, according to which a signature is a highly automated motor task and, as such, it is stored in the brain as both a trajectory plan and a motor plan. We then conjecture that such a stored representation does not necessarily include the entire signature, but can be limited to only parts of it, those that have been learned better and therefore are executed more automatically than others. Because these regions are executed more automatically than others, they are less prone to significant variations depending on the actual writing conditions, and therefore should represent better than other regions the distinctive features of signatures. To support our conjecture, we report and discuss the results of experiments conducted by using an algorithm for finding those regions in the signature ink and eventually using them for automatic signature verification.**

***Index Terms*—motor learning and execution; stability region; signature verification;**

I. Introduction

According to the daily experience, a coordinated sequence of "elementary" movements is acquired and executed faster and more accurately the more it is practiced. Early in learning, actions are attention demanding, slow and less accurate, whereas after long-term practice performance becomes quick, movements are smooth, automatic, and can be performed effortlessly, using minimal cognitive resources.

Studies on motor control have shown that selection, execution and learning of the movements needed to perform a motor task involve several brain areas and motor subsystems, but their activation and cooperation depend on the kind of movements that are being made and on the effector that is being used [1].

Indeed, when a child starts learning handwriting by copying letters or words, he attempts several trajectory patterns in order to replicate the same shape of the letters, selecting the points to reach through the visual system, and performing the appropriate sequence of movements through the motor system. During the initial phase of learning, the movements are quite straight and aimed to reach a sequence of points (as in Figure 1a). The executed motor plan is corrected according to the information provided by the visual and proprioceptive feedback, so that the actual trajectory corresponds to the desired one, and the lowest energy is spent by the muscular subsystem involved. As learning proceed, simple point-to-point movements become continuous, curved and smoother,

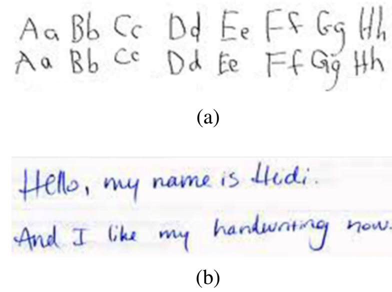


Fig. 1. Handwriting samples, written by a child (a) and a skilled writer (b).

the motor sequence comes to be executed as a single behavior and is performed automatically, using minimal cognitive resources (as in Figure 1b).

There is also strong evidence, supported by the results of several experimental studies on motor learning, that a given sequence of actions is learned from different perspectives. It has been observed, first by Lashley [2] and then by Hebb [3], that a generic movement, learned with one extremity, can be executed by different effectors. Furthermore, other studies have shown that writing movements learned through the dominant hand could be repeated using different body parts, such as non-dominant hand, the mouth (with the pen gripped by teeth) and foot (with the pen attached to it), even if the subject had essentially no previous experience writing with any of this body parts [4], [5]. Despite the different muscular and skeletal systems used and, even though the movements are not smooth, it can be observed that the writing production follows the same trajectory in all conditions [4] (see Figure 2). The ability to perform the same movement pattern by different muscular systems is called "motor equivalence". It suggests that movements directed to perform a task are stored in the brain in two ways: in an abstract form (effector-independent) related to the spatial sequence of points representing the trajectory plan, and as a sequence of motor commands (effector-dependent) directed to obtain particular muscular contractions and articulatory movements.

Other studies on motor learning have shown that when the untrained hand is used to perform a given sequence, learned with long-term practice with the other hand, performances are poor, but this is not true for a newly learned sequence [6], supporting the hypothesis that early in learning the execution of the motor task is more based upon the trajectory plan (effector independent), whereas late in learning upon the sequence of motor commands (effector-dependent).

Execution of voluntary movements requires the interaction between nervous and musculoskeletal systems, involving several areas, from the higher cortical centers to the motor circuits in the spinal cord [7].

In seeking to understand all the breadth and facets of motor learning, many researchers have used different approaches and methods, such as genetic analysis, neuroimaging techniques (such as fMRI, PET and EEG), animal models and clinical treatments (e.g. drugs administration and brain stimulation). These studies have provided a large body of knowledge that has led to several theories related to the role of the central nervous system in controlling and learning simple and complex movements. According to the results reported by neuroimaging and experimental studies on motor learning, several cortical and subcortical structures, including the basal ganglia, cerebellum, and motor cortical regions, are thought to be critical in different stages and aspects in the acquisition and/or retention of skilled motor behaviors.

In order to locate which brain area, or areas, underlie effector-independent representation of handwriting, Rijntjes and colleagues [8] carried out an fMRI study to examine patterns of brain activation associated with signing, using either the hand or the big toe. Their results showed the involvement of the parietal cortex in general, and posterior parietal cortex and occipitotemporal junction in particular, in the representation of written letter forms.

More recently, other neuroimaging studies have investigated the dynamics and functional connectivity of brain networks associated with learning a novel sequence of hand stroke movements to write ideomotor character [9]. Their results also suggest that a novel sequence of movements is initially mapped to form an internal representation of the sequence that is progressively encoded and refined subcortically (in the basal ganglia and in the cerebellum) as performance improves.

The imaging data reported by other studies on motor learning support the notion that distinct regions of the basal ganglia participate in different stages of learning. These studies report increased activity within the striatum (the input nucleus of the basal ganglia), in particular within the associative striatum and sensorimotor striatum early and late in learning, respectively. However, although there is solid evidence that the initial learning of many skills depends on the striatum, there are contrasting results in the literature regarding to the role of the sensorimotor striatum in automatic responding. For example, whereas some fMRI studies reported increased activity in the sensorimotor striatum with extended training, others reported decreased activity. Moreover, Turner and colleagues [10] reported that temporary inactivations of sensorimotor regions of the internal segment of the globus pallidus (a basal ganglia nucleus whose activity depends on the sensorimotor striatum) did not impair the ability of monkeys to produce previously learned motor sequences. Therefore, these results

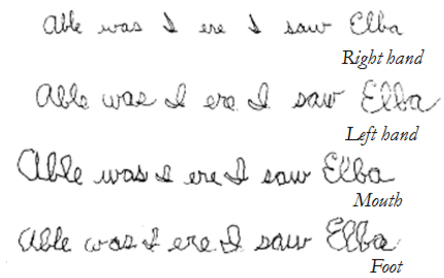


Fig. 2. A sentence written by the same writer using different body parts. Reproduced from [4].

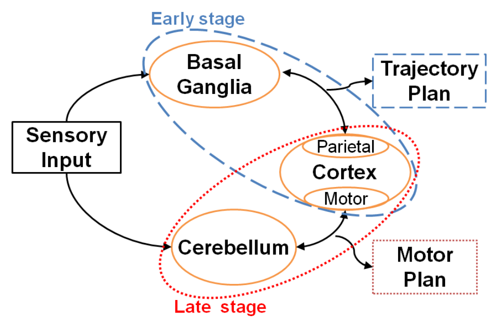


Fig. 3. Neural scheme of the model for procedural motor learning.

sustain the hypothesis that the basal ganglia play an important role in the initial stage of learning, whereas it is not well-established their importance in the final stage of learning.

With regard to the cerebellum, many studies report increased activity within the cerebellar cortex during learning, and increased activity within the dentate nucleus (an output nucleus of the cerebellar circuitry) until automaticity is achieved. A detailed review of the imaging studies whose results are here cited can be found in [11].

According to these results, we have proposed a neural scheme, based on the hypothesis that acquiring new motor skills requires two phases, in which two different processes occur:

- during the early stage, humans learn the spatial sequence associated to the motor task in visual coordinates, i.e. the sequence of points to reach in order to generate the ink trace.
- during the late, automatic phase, the sequence of motor commands in motor coordinates is acquired and comes to be executed as a single behavior.

The neural scheme for motor learning is shown in Figure 3 and incorporates the parietal and motor cortex, basal ganglia and cerebellum [12].

Sensory information is provided by an input module (sensory input in the figure) to the cerebral cortex, basal ganglia and cerebellum. The parietal association cortex releases signals that specify the position of targets in extrapersonal space (according to the studies conducted by Andersen and Zipser [13] and Rijntjes [8]). Therefore, the basal ganglia, interacting with the parietal cortex, select the next target point in the sequence. In turn, parietal cortex sends this information to the cerebellum that, interacting with the motor cortex, selects the appropriate motor command.

This model fits with the our hypothesis that motor learning follows two distinct phases. During the early phase of learning, the model learns the spatial sequence in visual coordinates (i.e. the sequence of points to reach in order to realize the motor task) through the interactions between the basal ganglia and the parietal cortex. The spatial sequence is then converted into motor commands through the interactions of the cerebellum and the motor cortex. Therefore the cerebral cortex, basal ganglia and cerebellum initially would work in parallel. The basal ganglia, through the associative striatum, are involved in the acquisition of the spatial sequence and the cerebellar cortex starts working to acquire the motor sequence. As learning proceeds, the sequence of motor commands in motor coordinates is acquired and stored in the dentate nucleus.

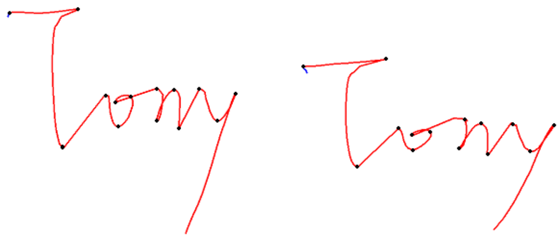
II. Signatures and motor learning

The neural scheme illustrated in the previous section suggests that after the learning, i.e. when the movement is executed fluently, the sequence of motor command is executed as a single movement. It suggests also that the more a movement is repeated the better is learned, i.e. the more it is automated. When applied to handwriting, the model suggests that the ultimate goal of the learning is that of producing a repertoire of completely automated movements in correspondence of the most frequently used sequences of characters. Such a repertoire depends on the sequences of characters the writer is most familiar with, which triggers the learning, and the sequences of the corresponding motor commands. Thus, the handwriting style emerges from both those aspects, the former being mainly language and cultural dependent, the latter being dependent on the physical and cognitive motor skills of the subject. Accordingly, different subjects may develop different repertoires of completely automated movements, either because the sequences of characters for which a completed automated movement is learned are different or because a different sequences of motor commands are learned for a given sequence of characters. When a completely automated movement has been learned for an entire message, multiple executions of such a movement should produce similar results, the difference between them being mainly influenced by the effector-dependent encoding of the learned sequence rather than from the effector-independent one. On the other hand, when more than one completely automated movement needs to be used for encoding the entire message, further variability may be observed in multiple execution of the same movements because the movements introduced for smoothing the transition between two successive completely automated movements are planned on the fly during the execution, and therefore may vary in both the effector-independent and the effector-dependent component.

What do these observations suggest in case of signatures? A signature is a movement the subject is very familiar with, that has been learned through repeated practice, and therefore it will have triggered a learning process whose final result is the repertoire of completely automated movements used by the subject while signing. If the entire signature is encoded in a single completely automated movement, it is expected that signatures produced by using the effector under the same condition result in very similar traces. In such a condition, in fact, the effector-independent part of the movement does not change because it has been completely learned and the effector-dependent component is supposed to be the same during all the execution. On the contrary, if the signature is produced by executing more than one completely automated movement, repeated execution may produce different traces, even under the assumption that the effector is used under the same condition, because there will be differences in the movements, and therefore in the traces, for connecting two successive completely automated movements. It follows from the observations reported above that whatever (dis)similarity measure is adopted for deciding whether a signature is genuine or not, it should be handled with care. In particular, it can be used successfully only after it has been decided which one are the parts of the signature that correspond to the execution of completely automated movements, and only the (dis)similarity between those parts of the signatures at hand should be evaluated by the adopted measure, because only those parts are expected to be ”stable” across multiple executions of the signature. In other words, the signature verification should be conducted by weighting differently the (dis)similarity between ”stable” regions and the (dis)similarity between other regions of the signature. In the following sections, we will briefly illustrate a procedure we have designed for finding the stability regions and then results obtained in a signature verification experiment.

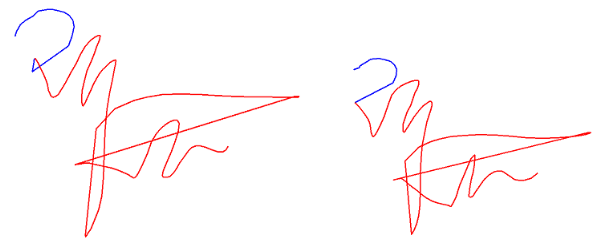
III. Finding the Stability Regions

It follows from our definition of stability regions that they are sequences of strokes produced as a single behavior and therefore should be embedded into any execution of the signature. Let us recall that a completely learned movement is stored in two forms, a sequence of target points, and a sequence of motor commands, and that the former is effector-independent, while the latter is effector-dependent. When the same effector is used in multiple executions, therefore, the only source of variability is the actual state of the effector, which may give raise to local variations in the shape of the ink traces. These traces, however, are composed of the same number of strokes and aimed at reaching the same sequence of target points. Assuming such a perspective, the stability regions are the longest common sequences of similar strokes found in two signatures, where similar means that they are aimed at reaching the same sequence of target points by following the same path. The method we have developed for finding the stability regions assumes that the signature signal has been segmented into a sequence of strokes, and the detection of the stability regions is achieved by an ink matcher that finds the longest common sequences of strokes with similar shapes between the inks of



(a) Genuine n. 16 (b) Genuine n. 19

Fig. 4. Genuine signatures produced by the user n. 22. The stability region is in red.



(a) Genuine n. 8 (b) Genuine n. 10

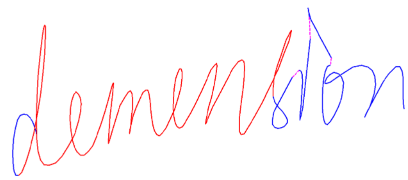
Fig. 5. Genuine signatures produced by the user n. 40. The stability region is in red.

a pair of signatures [14]. For deciding when two sequences are similar enough, i.e. when they match, the method exploits the concept of saliency that has been proposed to account for attentional gaze shift in primate visual system [15]. The rationale behind this choice is that, by evaluating the similarity at different scales and then combining this information across the scales, sequence of strokes that are globally more similar than other will stand out in the saliency map. The global nature of the saliency guarantees that its map provides more reliable estimation of trace similarity with respect to that provided by local criteria, as it is usually proposed in the literature [16]. According to the definition of stability regions, one would expect that the sequences of similar strokes provided by the ink matching appear in all the signatures. In practice, however, both the stroke segmentation and the ink matching may introduce errors, in locating the segmentation points (i.e. estimating the trajectory) and/or deciding when a sequence of strokes is similar to another (i.e. estimating the motor plan), that may produce different stability regions for the set of signatures. To decide which sequences correspond to the stability regions, we consider that longer stability region correspond to longer sequence of elementary movements executed in a highly automated fashion. Because the level of automation is the result of the learning process described above, and because the learning is an individual feature, long stability regions are more subject specific than short ones. Accordingly, we remove the stability regions that are subsequences of longer ones.

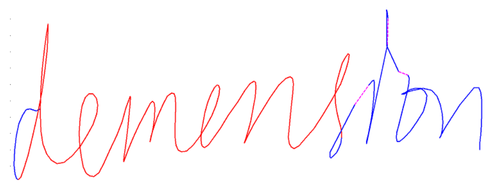
IV. Experimental results

We have two experimental results to support our conjecture about the role of stability regions in signature learning and execution and their effectiveness in signature verification. In both cases, the experiments were conducted on the SVC2004 dataset, adopted in the literature for writer verification/identification [17].

The first one was carried on by 3 subjects independently. They were provided with a written definition of stability regions in terms of sequence of strokes and asked to find them between 100 pairs of genuine signatures previously segmented by our algorithm. We then compared their outputs and removed 13 pairs for which there was some disagreement among them. This 87 pairs were then processed as above and the provided output compared with the one provided by the experts. In all the cases we have found a perfect correspondence between the machine and the expert. As an illustration of the results, the figures 4-6 show the stability regions found by the algorithm in case of signatures of different complexity. Figure 4 shows two signatures produced without any pen-up and pen-down occurring between the beginning and the end of the signature. The two traces are divided into the same number of strokes, and the stroke segmentation points, represented in figure as black dot, are located on the shape so as to roughly preserve their relative positions. According to our model, thus, the subjects concluded that the two shapes have been generated by the same motor plan, because it aims at reaching the same sequence of target points (estimated by the relative position of the segmentation points, as described in [18]) by means of the same sequence of elementary movements with the same time superimposition between successive ones (as estimated by the similarity between sequence of strokes). In this case, one would expect the algorithm to find just one stability region covering the whole signature, as it happens. Figure 5 shows two signatures produced by another writer without lifting the pen, as in the previous case, but with the end-effector in a different initial condition. Again, by looking at the segmentation points and at the similarity between sequence of strokes, the experts (and the machine as well) concluded that there was a difference in the initial parts of the signature (depicted in blue in the figure) and therefore they were not include in the stability region. Eventually, figure 6 depicts two long and complex signatures produced by a third writer.



(a) Genuine n. 14



(b) Genuine n. 18

Fig. 6. Genuine signatures produced by the user n. 6. The stability region is in red, the pen-up in magenta.

Because of the pen-up within the trace, depicted in magenta, and according to our conjecture, we expect that this signature is less automated and that stability regions may be found only during pen-down, as it happens. When requested to explain why they did not include the beginning of the ink trace (in blue) in the stability regions, the experts told us that the movement at the beginning of the sequence were very different, since in the first case the first stroke was directed top left, while in the second it was directed to left, showing also a sign of hesitation at the very beginning, as the subject started a movement directed down-left and suddenly corrected it. Similarly, in the first case it appears to be a stop-and-go or an hesitation while drawing the letters. In both cases, they were interpreted as sign of difference between the sequence of strokes constituting the motor plan.

The second result comes from a signature verification experiment we have designed and performed on the same dataset [14]. In such an experiment, we have used the stability regions provided by our algorithm for both selecting the genuine signatures to be used as reference and classifying the questioned signatures as both genuine or forged. Each questioned signatures was compared with the stability regions of the references. If a match was found, the similarity between the sequence(s) of strokes of the stability region(s) in the reference and the matching sequences of strokes in the questioned was compared with two thresholds, to decide whether the questioned was genuine or not. Despite this very simple decision criteria, and the exploitation of shape information only for measuring the similarity between sequence of strokes, the experimental results showed that our method was the 5th among the 15 methods considered in the final ranking, but also that it exhibited the lowest standard deviation of the performance. This latter finding suggests that the method captures the common aspects of signatures as they derive from the model, and therefore is quite robust in providing similar performance independently of the distinctive signing habit of each subject. Even more interesting, most of the errors are found in case of signatures with many pen-up and pen-down, and whose stability regions are made of a few strokes, further supporting our claim that the more the signature is automated the longer are the stability regions.

All together, those results show that stability regions, as we have defined and implemented them, do seems to exist and that they can represent a promising way to root signature verification within the framework of motor learning and execution.

V. Conclusions and future directions

We have discussed some recent findings in neurocomputational modeling of motor learning and execution and suggested that they may provide a new perspective for handwriting analysis. Under such a perspective, we have conjectured that signatures are represented as a motor plan, stored in a distributed fashion between the basal ganglia and the cerebellum, which encodes both the target points to be reached and the motor program to execute for producing the desired handwriting. From this conjecture we have derived a definition of stability regions by globally evaluating the traces shape similarity by means of a saliency map.

Our conjecture is supported by two experiments showing that: human subjects may actually find stability regions that fits with our definition and that such regions provide a plausible estimate of the motor plans used to produce the observed traces; the proposed algorithm finds the same stability regions as the human subjects; the stability regions may be used for both selecting the reference signature and performing signature verification, providing very promising results even when a very simple criterion is used to decide whether a questioned signature is genuine or not.

In the future we will investigate to which extent our model can deal with disguising writers. We would also like to understand whether there is any relation between legibility and learning of signatures.

References

[1] M. Kawato, "Internal models for motor control and trajectory planning," *Current Opinion in Neurobiology*, vol. 9, pp. 718–727, 1999.

[2] K. Lashley, "Basic neural mechanisms in behavior," *Psychological Review*, vol. 37, pp. 1–24, 1930.

[3] D. O. Hebb, *The organization of behavior: a neuropsychological theory*. New York: Wiley, 1949.

[4] M. H. Raibert, *Motor control and learning by the state space model*. Cambridge: Artificial Intelligence Laboratory, MIT, 1977.

[5] A. M. Wing, "Motor control: mechanisms of motor equivalence in handwriting," *Current Biology*, vol. 10, pp. 245–248, 2000.

[6] M. K. Rand, O. Hikosaka, S. Miyachi, X. Lu, and K. Miyashita, "Characteristic of a long-term procedural skill in the monkey," *Experimental Brain Research*, vol. 118, pp. 293–297, 1998.

[7] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, *Principles of Neural Science*. McGraw-Hill, 2000.

[8] M. Rijntjes, C. Dettmers, C. Buchel, S. Kiebel, R. Frackowiak, and W. C., "A blueprint for movement: functional and anatomical representations in the human motor system," *Journal of Neuroscience*, vol. 19, no. 18, pp. 8043–8048, 1999.

[9] B. A. Sweet, J. L. Contreras-Vidal, B. Rasmus, and A. Braun, "Neural substrates of graphomotor sequence learning: A combined fMRI and kinematic study," *Journal of Neurophysiology*, vol. 103, no. 6, pp. 3366–3377, 2010.

[10] R. S. Turner, K. McCairn, D. Simmons, and I. Bar-Gad, *The basal ganglia VIII (Advances in behavioral biology, vol. 56)*. Springer, 2005, ch. Sequential motor behavior and the basal ganglia, pp. 563–574.

[11] R. Senatore, *The role of Basal Ganglia and Cerebellum in Motor Learning: A computational model*. University of Salerno: PhD Thesis, 2012.

[12] R. Senatore and A. Marcelli, "A neural scheme for procedural motor learning of handwriting," in *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*, 2012, pp. 659–664.

[13] R. A. Andersen and D. Zipser, "The role of the posterior parietal cortex in coordinate transformations for visual-motor integration," *Canadian Journal of Physiology and Pharmacology*, vol. 66, pp. 488–501, 1988.

[14] A. Marcelli, S. Fuschetto, and A. Parziale, "Modeling stability in on-line signatures," in *International Graphonomics Society (IGS), 2013*, 2013, pp. 135–138.

[15] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 11, pp. 1254–1259, 1998.

[16] D. Impedovo and G. Pirlo, "Automatic signature verification: The state of the art," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 38, no. 5, pp. 609–635, 2008.

[17] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll, "Svc2004: First international signature verification competition," in *Biometric Authentication*. Springer, 2004, pp. 16–22.

[18] A. Marcelli, A. Parziale, and A. Santoro, "Modeling handwriting style: a preliminary investigation," in *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*. IEEE, 2012, pp. 411–416.

Offline Handwriting Acquisition under Controlled and Uncontrolled Conditions

Linda Alewijnse
Netherlands Forensic Institute
Department of Digital Technology and Biometrics
The Hague, The Netherlands
l.alewijnse@nfi.minvenj.nl

***Abstract*—This paper gives a description of offline handwriting acquisition under controlled and uncontrolled conditions for research purposes. The data collection task is an underestimated part in the process of developing signature verification or handwriting identification systems. There is a continuous need for new, unpublished data to train and evaluate new algorithms. Handwriting samples that make up the current publicly available databases have all been collected under controlled conditions. However, good quality data is still limited.**

On the contrary, research databases constituted of case related biometric data in general are scarce. To suit forensic purposes, it is preferred to start building databases with forensically relevant data. When verification and identification systems are trained on this type of material, the output will be more suited for forensic examination purposes. The challenges in this area are considered.

Keywords—offline data, data collection, signature verification, forensic handwriting examiner

I. Introduction

Signature verification is a biometric technique with promising results for the near future for implementation within the forensic handwriting examination. In the past 10 years rapid developments are made within the pattern recognition discipline [1]. Implementing analysis tools in the forensic practice is the next challenge. Before an automated signature verification or handwriting identification system can be implemented, the forensic community must be ascertained that the systems are trained, evaluated and validated by correct environmental conditions.

Collecting and selecting handwriting samples for research purposes is often an underestimated task. The number of publicly available databases with handwriting is limited, so new data must be collected regularly. Data are primarily collected to provide information regarding a specific topic. Therefore, data must be in accordance with the objective of the study. The overall performance of a biometric technology is eventually influenced by the quality of the input data.

A. *Learning from the past*

The following example illustrates the importance of sample design and sample selection to suit the purpose of the study. In 2002, Srihari and colleagues [2] conducted a study to test the principle of individuality of handwriting. Handwriting samples were collected from 1500 individuals. The dataset was representative for the US population with respect to gender, age, ethnicity, handedness, etc. The automated system CEDAR-FOX was used to evaluate the handwriting, and could identify the writer of a particular sample with 98 percent confidence. Inferring these statistics over the entire U.S. population, writer identification can be established with 96 percent confidence.

Saks [3] commented on this study by arguing that to test individuality, a better sampling design would have been to gather a representative sample of clusters of writers, with each cluster composed of highly similar writers. Only then, the data would have been discriminative of highly similar handwriting. And it would have been repeatable if the same effect was observed between the clusters. The choice of data by Srihari and colleagues was not adequate for testing the hypothesis that handwriting is individual.

In a response to this, Durina and colleagues [4] conducted a study in which samples of writing were obtained from 52 writers and their teachers who were taught the same copybook style at the same Catholic elementary school approximately 4 decades ago. The research addressed the criticisms that earlier studies on the individuality of handwriting did not include populations from homogeneous writing communities. It demonstrated that there is a high degree of inter-writer variation among writers, even in populations where the driving forces for variation are low. In spite of the size of the dataset, it was better fit for purpose to investigate the uniqueness of handwriting.

B. *Learning from each other*

In the past years, from 2009 until 2013, different datasets with signatures as well as handwriting are collected by the Netherlands Forensic Institute for the Signature Competition (SigComp) [5]. This competition allows researchers and practitioners from academia and industries to compare performance on signature verification on new and unpublished datasets. Because all participating parties in the competition are provided with the same data, results are comparable. While the competition provides an overview of involved parties and shows the performance of the available systems to the forensic community, the pattern recognition researchers are more concerned about which features are most discriminative. The SigComp provides a platform to bridge the gap between the two communities.

Two years ago, in 2011, a group of researchers from different fields of expertise started the discussion about how to bridge the gap between the two communities and to signal the challenges. Computer programmers learned how a forensic handwriting examination is carried out and examples of real casework are described. Forensic scientists got an overview of state-of-the-art automatic verification systems. Recent advances are comparing the performance with Minimum Cost of Log Likelihood Ratios [6], the task of reporting a probabilistic output score, and the addition of disguised signatures in new datasets. Nevertheless, much work needs still to be done in order of bringing together researchers in the field of automated handwriting analysis and signature verification and experts from the forensic handwriting examination community.

The scope of the competition changes each year. In the end, when automated systems are meant to aid the FHE in the examination or as an objective tool. The first competition was focused on skilled forgeries. After that, disguised signatures were added to the questioned signatures. Last year we've provided different scripts, i.e. Dutch and Chinese signatures. The consequence of the changing focus of the competition allows the developers to improve their algorithms and benefit from new and unpublished handwriting data.

II. OBJECTIVE

Three scenarios for handwriting data collection can be distinguished: 1) The samples are collected under controlled conditions, e.g. let the participants write on the same make of paper, with the same writing instrument, in similar writing position, etc., 2) spontaneous writings are collected from participants by gathering their writings from the past, and 3) forensic handwriting samples from casework are shared, either anonymously or by an online evaluation platform.

Topics that are covered in this paper are:

- offline and online data
- requirements of the dataset
- controlled versus uncontrolled conditions
- research data versus forensic data

The first part of the paper describes the most favorable and pragmatic approach for offline handwriting sample collection. The second part stresses the importance of data collection under uncontrolled conditions. Furthermore, this paper calls for exploring the possibilities of using forensic datasets to further develop automated systems.

III. METHOD

Two categories capturing a person's handwriting can be distinguished, namely, offline and online. The online modality is discussed here very shortly, because this data is not available to the forensic handwriting examiner. It is useful for biometric identification and finding the new features or feature combinations that are most discriminative. Handwriting examiners will in particular be interest in offline systems and therefore offline data acquisition is described more in detail.

A. *Online data*

Online data collection requires an electronic writing tablet and recording software. Most often WACOM tablets are used to collect handwriting samples, but since pen-input devices getting more widespread this might change on short term. The online handwriting is captured with an electronic writing tablet and stored digitally in x, y, and z-positions as a function of time.

B. *Offline data*

Offline handwriting data is a representation of the handwriting in as a scanned image. It has been demonstrated [7] the FHE's can infer dynamic information, such as writing velocity and pen pressure, from the static trace. Writing velocity is reflected in line quality, pen pressure differences and blunt beginnings and endings of stroke. The pen pressure is not useful for the examiner as an absolute measure, since it is not only writer specific but strongly depends on extrinsic factors. It is only writer specific if other conditions such as writing surface and writer instrument are constant. The indentation of the paper shows the handwriting examiner if the ink was deposited by a natural course of writing or by forced writing.

For offline data collection all that is needed is a pen, a piece of paper and a scanner. To aid the writer, a guiding line or box can be used. The easiest and practical solution is to use an underlying sheet of paper with the lineation or boxes printed with a black, bold line. No lineation or bounding boxes must strike trough the writings. In this way, the data is kept 'clean' and less effort for data preparation is needed.

Fig. 1 Offline specimen signatures collected under controlled conditions.

C. Data requirements

The requirements for a high-quality offline dataset of handwritten samples are summed up below. A formal data collection process is necessary as it ensures that gathered data are both defined and accurate and that decisions based on arguments embodied in the findings are valid [7].

The first list proposed shows which requirements of the dataset are advised for training and evaluating automated systems. Additionally there is a list of extra requirements which are important for forensic handwriting researchers. The summed information is necessary for forensic handwriting examiners to get a better understanding of the data used in experiments. In general, the data must reflect the variation of handwriting in the relevant population, and intra-writer variation must represent reality.

Pattern recognition data requirements:

- Substantial number of specimen writers
- Substantial number of simulators
- High resolution scans of the written samples, preferably 400 dpi.
- Suitable format (PNG format would be preferable. This lossless format will retain information from images when re-opened and re-saved. The PNG format also creates smaller file size but without the quality loss of a GIF-file)
- Cropping of the image
- Assign an identification code as filename
- Compatibility with earlier collections

Additional forensic requirements:

- Writer sex, age, handedness, level of education, and profession
- Cultural origin (for signatures) or copybook system (for handwritten text)
- Substantial amount of questioned writing (e.g. half a page of text)
- Substantial amount of reference writing (number of reference signatures or number of lines of text)
- Specification of conditions of forgery and/or disguised
- Time span over which the data was collected

IV. FORENSIC HANDWRITING DATA

A. Collecting existing specimens

One way of acquiring relevant data is to collect existing writings. Such handwriting can consist signatures on agreements, receipts, cheques, passports, etcetera. In short, it can comprise handwriting, which is comparable to the reference material in casework. All factors that are considered by forensic handwriting examiners are in the dataset: natural variation in the writings, different surfaces, different writing instruments, different time period and the samples are written under different mental circumstances. Both intrinsic and extrinsic factors are represented. Participants are not approached to write something, but provide the researcher with their previously written material.

a)
b)
c)
d)

Fig. 2 Examples of collected specimen signatures written under uncontrolled conditions: a) A signature that was written under a declaration form, b) two overlapping signatures with restricted space for signing, c) signature on a receipt that was written in a standing writing position, and d) signature on an ID-document, dating from 5 years ago.

B. Case related data

The best would be using forensic casework data to evaluate and validate automated systems, but legal aspects regarding privacy form an obstacle. One possible solution for sharing forensic samples is to facilitate access at an online evaluation platform. BEAT [8] is a project that is funded by the European Commission, under the Seventh Framework Programme and is offering such an approach. The goal of the project is to propose a framework of standard operational evaluations for biometric technologies. Unfortunately, it is not available for forensic biometrics yet.

Simulated data can be used in the training phase of system development, because the ground truth of the origin is known. The evaluation phase should at least contain case related data. However, the validation of the system should completely be performed with real casework samples.

V. CONCLUSION AND DISCUSSION

Where biometric systems usually have access to high quality and uniform data, in forensic practice the trace under investigation is often characterized by poor quality. This is not represented by the currently existing handwriting databases.

Since input data determines the overall performance of the automated system, a next step in bridging the gap between the pattern recognition community and forensic handwriting examiners should logically involve the use of samples that were written under uncontrolled circumstances. The condition of the dataset has its effect on the systems' performance on that trace and accordingly influences the strength of the evidence.

REFERENCES

[1] M. Caligiuri and L. Mohammed, "The Neuroscience of Handwriting: Applications for Forensic Document Examination," CRC Press, 2012.

[2] S.N. Srihari, S-H Cha, H. Arora, and S. Lee, "Individuality of handwriting", J Forensic Sci, vol. 47(4), pp. 856—872, 2002.

[3] M. Saks, Authors' Response in the J Forensic Sci, vol. 48(4), July 2003.

[4] M.E. Durina and M.P. Caligiuri, "The Determination of Authorship from a Homogenous Group of Writers," Journal of ASUDE, vol. 12, nr. 2, 2010.

[5] M.I. Malik, M. Liwicki, L. Alewijnse, and W. Ohyama, "ICDAR2013 Competitions on Signature Verification and Writer Identification for On- and Offline Skilled Forgeries (SigWiComp2013)," in press.

[6] M. Liwicki et al., "Signature Verification Competition for Online and Offline Skilled Forgeries (SigComp2011)," in Document Analysis and Recognition (ICDAR), 2011 International Conference on, pp. 1480-1484, 2011.

[7] D. Meuwly and R.N.J. Veldhuis, "Forensic biometrics: From two communities to one discipline," IEEE Conference publications BIOSIG 2012, Darmstadt Germany, pp. 1-12, Sep 2012.

[8] Information available at www.beat-eu.org

Oriented Local Binary Patterns for Writer Identification

Anguelos Nicolaou
Institute of Computer Science and
Applied Mathematics University of Bern
Neubrückstrasse 10
3012 Bern, Switzerland
Email: anguelos.nicolaou@gmail.com

Marcus Liwicki and Rolf Ingolf
Document, Image and Voice Analysis (DIVA) Group
University of Fribourg
Bde des Perolles 90
Fribourg Switzerland
Email: firstname.lastname@unifr.ch

***Abstract*—In this paper we present an oriented texture feature set and apply it to the problem of offline writer identification. Our feature set is based on local binary patterns (LBP) which were broadly used for face recognition in the past. These features are inherently texture features. Thus, we approach the writer identification problem as an oriented texture recognition task and obtain remarkable results comparable to the state of the art. Our experiments were conducted on the ICDAR 2011 and ICHFR 2012 writer identification contest datasets. On these datasets we investigate the strengths of our approach as well its limitations.**

I. Introduction

A. Local Binary Patterns

Local binary patterns (LBP) were broadly popularized in 2002 with the work of Ojala et al [1] as a texture feature set extracted directly on grayscale images. As well demonstrated by Ojala, the histogram of some specific binary patterns is a very important feature-set. LBP are inherently texture features, but they have been used in a very broad range of applications in Computer Vision (CV), many of which exceed the typical texture recognition tasks. In 2004, Ahonen et al [2] used successfully LBP for face recognition. In 2007, Zhao et al [3] extended the operator as a 2D plus time voxel version of LBP, called VLBP, and used them successfully for facial gesture recognition. In 2009, Whang et al [4] combined LBP features with HOG features to address the problem of partial occlusions in the problem of human detection.

B. Writer Identification

While graphology, i.e. the detection of personality traits based on handwriting, has been associated with bad science [5] and has failed to provide experimentally sound significant results [6], handwriting style can be considered an invariant attribute of the individual. Writer identification has traditionally been performed by Forensic Document Examiners using visual examination. In recent decades there is an attempt to automate the process and codify this knowledge in to automated methods. In 2005, Bensefia et al [7] successfully used features derived from statistical analysis of graphemes, bigrams, and trigrams. In 2008, He et al [8] used Gabor filter derived features and in 2010 Du et al [9] introduced LBP on the wavelet domain. Even-though the method of Du uses LBP for feature extraction in writer identification, the similarities end there. Our method makes no assumptions specific to handwriting and treats the problem as a generic oriented binary texture classification problem. The extent to which handwriting contains invariant characteristics of the writer is an open question. While forensic document examiners have been tested in detecting disguised handwriting by Bird et al [10], Malik et al [11] have started to address the issue of different writing styles for automated offline writer identification systems. It remains an open question whether handwriting style can provide us with real biometric markers, invariant to the sample acquisition conditions. By preserving the generic attributes of our method, we can safely avoid addressing many complications that are specific to handwriting analysis and writer detection.

II. LBP Feature Set

Although writer identification seems to require scale invariant features, scale sensitive features might be suited as well. Writers tend to write with a specific size, therefore the scale of the texture tends to be directly dependent on the sampling rate. The task of writer identification is almost always done with respect to a dataset, where the sampling rate is defined or at least known when performing feature extraction. It is feasible and probably worth the effort of resampling all text images to a standard sampling resolution, rather than improvising a scale invariant feature-set. Our feature-set as is the norm, is derived from the histogram of occurring binary patterns.

A. The LBP operator

LBP were defined in [1] as a local structural operator, operating on the periphery of a circular neighborhood. LBP are encoded as integers, which in binary notation would map each sample on the periphery to a binary digit. As can be seen in Fig. 1 and (2), LBP are defined by the radius of the circular neighborhood and the number of pixels sampled on the periphery. The sampling neighborhood $N_{r,b}$ is formally defined in (1).

$$\forall n, \phi : n \in [0..b-1] \wedge \phi = (n * 2 * \pi)/b$$
$$\forall f(x_1, x_2) : R^2 \implies \{0, 1\}$$

$$N_{r,b}(I(x,y), n) = I(x + sin(\phi) * r, y + cos(\phi) * r) \quad (1)$$

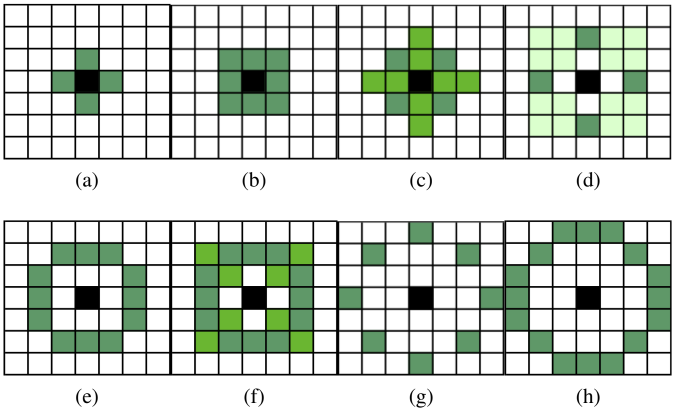


Fig. 1: Indicative LBP operators: $LBP_{1,4}$ (a), $LBP_{1,8}$ (b), $LBP_{1.5,8}$ (c), $LBP_{2,8}$(d),$LBP_{2,12}$ (e), $LBP_{2,16}$ (f), $LBP_{3,8}$ (g), $LBP_{3,16}$ (e). Dark green represents pixels with 100% contribution, green represents pixels with 50%, light green pixels with 25%, and black is the reference pixel.

$$\begin{aligned} LBP_{r,b,f}(x,y) = f(N_{r,b}(I(x,y),n) * 2^n, I(x,y))+ \\ f(N_{r,b}(I(x,y),n-1) * 2^{n-1}), I(x,y)) + ... \\ ... + f(N_{r,b}(I(x,y),0) * 2^0, I(x,y)) \end{aligned} \quad (2)$$

When defined on grayscale images, LBP are obtained by thresholding each pixel on the periphery by the central pixel. Because we worked on binary images as input, a lot more operations than greater or equal (thresholding) were possible as a binary operation. We generalized our definition of the LBP in (2), to consider the boolean operator marked as f a third defining characteristic of the LBP operator $LBP_{r,b,f}$ along with the radius r and the number of samples b.

We took into account several factors for selecting the appropriate LBP binary operator. In what concerns the bit count, a bit-count of 8 presents us with many benefits. Implementation wise, the LBP transform is an image that uses one byte per pixel. Its histogram has 256 bins providing a high feature-vector dimentionality and good discriminative properties. Additionally, containing the distinct LBP count to 256, guaranties highly representative sampling in relatively small surfaces of text.

B. The LBP function

While LBP are traditionally derived from grayscale images, when dealing with text, its better to use binarized text images as input, thus avoiding all information coming from the text background. We considered many different binary operations and chose the binary operator "equals" (3) as $f()$ in (2).

$$f(x_{ceter}, x_{periphery}) = \begin{cases} 1 & : x_{ceter} = x_{periphery} \\ 0 & : x_{ceter} \neq x_{periphery} \end{cases} \quad (3)$$

"Equals" as a boolean function on an image means true for any background pixel in the peripheral neighborhood of a background pixel, true for any foreground pixel in the peripheral neighborhood of a foreground pixel, and false for everything else. When using the "equals" function as the binary function in a 8 bit-count LBP, all pixels with only foreground and only background have an LBP value of 255. By suppressing (ignoring) the 255 bin, we make the LBP histogram surface invariant. All occurrences left in the histogram represent the pixels in the border between foreground and background. The core of the feature set comprises of the 255 histogram bins normalized to a sum of 1. This normalization renders the features derived from the histogram invariant to the number of signal pixels in the image.

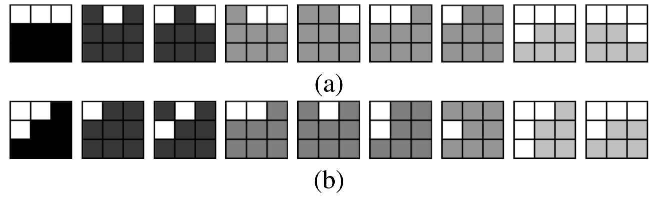


Fig. 2: LBP edge patterns. In (a) the top-edge contributing patterns and in (b) the top-left edge contributing patterns can be seen. Contribution: black 100%, dark gray 50%, gray 25%, and light gray 12.5%

C. Redundant Features

Having the normalized 255 bins from the histogram as the core of the feature set, we calculate some redundant features that will amplify some aspects of the LBP we consider significant in the writer identification task. Our goal is to have a feature-set discriminative enough to work well with naive classifiers such as nearest neighbor or, even more, classify writers by clustering the samples without any training.

The first redundant feature group we use, is edge participation. We consider each pattern to have a specific probability of belonging to an edge of a specific orientation; from now on we call that contribution. The sum of the number of occurrences of each pattern, multiplied by its contribution factor makes up the oriented edge occurrences. In Fig. 2a all top-edge patterns can be seen along with their probability, in 2b we can see the patterns of the top-left-edge patterns and their probabilities which are derived from the top-edge patterns by rotating them counter-clock-wise. By rotating the contributing patterns of the top-edge, we can obtain the contributing patterns of all eight edge-orientations. We also add the more general edge-orientations: horizontal, vertical, ascending, and descending as separate features which are calculated as the sum of the respective pair of edge-orientations. In the end we calculate the two final aggregations of perpendicular and diagonal, which are the sum of horizontal and vertical and respectively ascending and descending. In total we obtain 14 edge-features, which we then normalize to a sum of 1. One of our aims when introducing these redundant features is to enhance characteristics that have been associated with writer identification such as text slant.

The second redundant feature-group we implemented are the rotation invariant hashes. We grouped all patterns, so that each pattern in a group can be transformed in to any other pattern in that group by rotating. When having an 8 sample LBP, the distinct rotation invariant patterns are 36 in total [1]. Some pattern groups contain only one pattern eg. pattern 0, while other groups contain up to 8 patterns, such as all one bit true patterns 1,2,4,8,16,32,64,128. We took the number of occurrences for each group in the input image and normalized them to a sum of 1, thus providing 36 rotation invariant features. A complementary feature-group to the rotation invariant patterns is what we named rotation phase. For each group of rotation invariant features, we took the minimum, with respect to the numeric value, pattern in the group and designated it as group-hash. The number of clockwise rotations each pattern needs to become its groups-hash, is what we call the rotation phase. By definition, the distinct phases in an LBP image, are as many as the number of samples of the LBP. The frequency of all phases normalized to the sum of 1, provides us with 8 more redundant features that are complementary to the rotation invariant hashes.

A third group of redundant features we introduced to our feature-set is what we called beta-function as defined in (4) along with the bit-count of every pattern.

$$\begin{aligned}
&\forall n \in [1..bitcount] \\
&\forall lbp \in [0..2^{bitcount-1}] \\
&d(lbp, n) = \begin{cases} 1 & : \text{bit } n \text{ is set in } lbp \wedge \\ & \quad \text{bit } n-1 \text{ is not set in lbp} \\ 0 & : otherwise \end{cases} \\
&\beta(lbp) = \sum_{n} d(lbp, n)
\end{aligned} \tag{4}$$

When the sample count is 8, the β function, has up to 5 distinct values. The histogram of the β function (5 bins) normalized to a sum of 1 and the histogram of the bit-count of every pattern normalized to 1 as well, are the last redundant feature-group we defined. The β function becomes an important feature when the LBP radius is greater than pen stroke thickness. In those situations, e.g. a β count of one, would indicate the ending of a line, and a β count of three or four would indicate lines crossing.

If we put it all together, we have 255 bins of the histogram, plus 36 rotation invariant features, plus 8 rotation phase features, plus 14 edge-features, plus 5 β function features, plus 9 sample-count features, makes a total of 327 features; these are the proposed feature-set. The redundant features make the features well suited for naive classifiers. By setting the 255 histogram bin to 0, the feature set ignores all non signal areas in the image. The normalization of all bins to a sum of 1, as well as the nullification of the last bin, renders our feature set invariant with respect to non signal areas (white).

D. The Classifier

Once we transform a given set of images into feature vectors, we can either use them as a nearest neighbor classifier or perform clustering on them. While clustering seems to be a more generic approach, it is constrained by the need to process all samples at the same time. Such a constraint makes the clustering approach very well suited for research purposes but hard to match any real world scenarios. The construction of the classifier consists of four steps. In the first step, we extract the image features. In the second step, we rebase the features along the principal components of a given dataset by performing principal components analysis. This step might, in a very broad sense of the term, be considered training because our method acquires information from a given dataset. In the third step we scale the rebased features by a scaling vector which was defined by evolutionary optimization on the train-set. The optimization process is also performed on a given dataset and should also be considered as a training stage. While it is not required, it makes more sense that both training steps are performed on the same dataset. The fourth and last step is to calculate the L1 norm on the scaled and rebased feature vectors. Steps two and three can be combined in to a linear operation on the feature space and in many aspects should be viewed as a statistically derived heuristic matrix. Our classifier, as was implemented, has two inputs, a subject dataset and a reference dataset. The output consists of a table where each row refers to a sample in the subject dataset and contains all samples in the reference dataset ranked by similarity to the specific sample. When benchmarking classification rates of our method, we can simply run our classifier with an annotated dataset as both object dataset and reference dataset. In this case, the first column contains the object sample and the second column contains the most similar sample in the dataset other than its self. The rate at which the classes in the first column agree to the classes in second column, is the nearest neighbor classification rate.

E. Scale Vector Optimisation

Describing in detail the optimization process of the scaling vector would go beyond the scope of this paper. In brief we optimized using an evolutionary algorithm. We used as input the 125 most prominent components of the features and the id of the writer of each sample. We optimized using the ICHFR 2012 writer identification competition dataset [13] which contains 100 writers contributing 4 writing samples each. Individuals of the algorithm were modeled as vector of continuous scaling factors for each feature in the feature space. The fitness function was based on the classification rate a nearest neighbor classifier obtains when the feature space is scaled by each individual. The stoping criteria was set to 2000 generations, and each generation had 20 individuals. Suitable parents were determined by the rank they obtained in the generation.

III. Experimental Procedure

In order to have a proper understanding of our methods performance, its robustness, and its limitations, we conducted a series of experiments. We used two datasets for our experiments: the dataset from the ICDAR 2011 writer identification contest [12], hereafter 2011 dataset and the dataset from the ICHFR 2012 writer identification challenge [13], hereafter 2012 dataset. The 2011 dataset has 26 writers contributing samples in Greek, English, German, and French with 8 samples per writer. The 2012 dataset has 100 writers, contributing samples in Greek and English with 4 samples per writer. While the 2011 dataset was given as the train set for the 2012 contest, we used them in the opposite manner. In order to avoid overfitting during the optimization step, we deemed the "harder" dataset, containing more classes and less samples per class, was better suited for training.

A. Performance

As previously described, our method consist of four stages: feature extraction, principal components analysis, scaling vector optimization, and L1 distance estimation. Steps two and

TABLE I: Performance Results. Various modalities of our method's results on the 2011 dataset [12] and state of the art methods performance for reference

NAME	Nearest Neighbor	Hard Top-2	Hard Top-3	Hard Top-5	Hard Top-7	Soft Top-5	Soft Top-10
Tsinghua	99.5%	97,1%	NA	84.1%	44.1%	100%	100%
MCS-NUST	99.0%	93.3%	NA	78.9%	38.9%	99.5%	99.5%
Tebessa	98.6%	97.1%	NA	81.3%	50.0%	100%	100%
No PC, No train	**96.63%**	87.02%	79.33%	63.94%	28.84%	98.56%	99.04%
PC, No train	98.56%	91.35%	84.62%	68.27%	34.62%	98.56%	98.56%
PC, Train	**98.56%**	95.19%	91.83%	84.13%	50.48%	99.04%	99.04%

three require a training dataset, while steps one and four are totally independent of any data. In TABLE I analytical scores of our method in various modalities can be seen. Apart from the nearest neighbor accuracy we also add the hard TOP-N and soft TOP-N criteria [12], [13]. The soft TOP-N criterium is calculated by estimating the percentage of samples in the test set that have at least one sample of the same class in their N nearest neighbors. The hard TOP-N criterium is calculated by estimating the percentage of samples in the test set that have only samples of the same class in their N nearest neighbors. More in detail, in TABLE I we can see various versions of our method and their performance as well as some state of the art methods for reference. Methods Tsinghua, MCS-NUST and Tebessa [14] are the top performing methods from the ICDAR 2011 writer identification contest. We must point out that our method had a vastly superior train set, consisting of 400 samples, and we had access to the test set while working. Our method has two parts that were optimized on our train set, the 2012 dataset. The first is the principal components of the train set and the second is the scaling of the feature space. No PC, No train is the raw feature space without any training, just the features in an L1 nearest neighbor setup. PC, No train is the feature space rebased along the principal components of the the train set in a L1 nearest neighbor setup. PC, Train is the feature space rebased along the principal components of the the train-set and scaled along the optimized vector in a L1 nearest neighbor setup. As we can see our method almost reaches the overall performance of the state of the art when it incorporates the full trained heuristics but it also provides very good results in its untrained form.

B. Qualitative Experiments

Apart from providing a comprehensive accuracy score that is comparable to other methods, in order to describe the strength and limitations of our method, we performed a series of experiments that simulate frequently appearing distortions to the data.

1) Rotation: Text orientation, is a text image characteristic that is definitely affected by the writer. Under controlled homogeneous experimental conditions of data acquisition, text orientation should depend only on the writer. Quite often in real life scenarios we have no way of knowing whether an image has been rotated or not and to which extent. One of the important characteristics of a writer identification system is the robustness against moderate rotation. We address this issue by an experiment where we try to recognize samples of a dataset with rotated versions of the database. More specifically we took the 2012 dataset and we rotated its samples by 1° from -20° to 20°. We obtained our measurements by classifying the original 2012 dataset with the the rotated versions. In Fig. 3 the rotation sensitivity of our method can be demonstrated. Two different measurements can be seen. The first one, noted as Sample Self Recognition, is the the nearest neighbor including the test sample. Sample Self Recognition rate will be by definition 100% when no rotation occurs. The second measurement, marked as Nearest Neighbor is the accuracy of nearest neighbor excluding the first occurrence. Nearest Neighbor is by definition the accuracy when no rotation occurs. As can be seen in Fig. 3 our method demonstrates some rotation tolerance from -5° to $+5^\circ$ with sustainable accuracy rates, but performance drops significantly beyond this limit[1]. It is also worth noticing the fact that -1° and $+1^\circ$ rotations perform slightly worst than -2° to $+2^\circ$; a possible explanation for this could be aliasing phenomena.

2) Downsampling: As we stated previously, in most real world scenarios, the sampling resolution will be known to a writer identification system, but not always controlled as sometimes the data are acquired by external sources or at different times. We devised an experiment that demonstrate the behavior and limitations of our method in what concerns the resolution. We took the ICDAR 2011 Writer Identification dataset and rescaled it to various scales from 100% down to 10%. As can be seen in Fig. 4 we obtained three measurements. The first, marked as Self Recognition Unscaled Sample, is the nearest neighbor when classifying the initial dataset with the subsampled dataset as a database. The second, marked as Nearest Neighbor Unscaled Sample, is the second nearest neighbor when classifying the initial dataset with the subsampled dataset as a database. We presume that the first nearest neighbor will always be the same sample in different scales and therefore disregard it for this measurement. The third measurement, named Nearest Neighbor Scaled Sample, is the accuracy of the second nearest neighbor when classifying the scaled dataset with the scaled dataset a database. The first two measurements describe the sensitivity our method has in comparing samples of different sampling resolution and therefore scale as well, while the third measurement demonstrates how well our method would work on datasets of lower resolution. We should also point out that the optimization process was performed on the original resolution. As we expected and can be seen in Fig. 4, we find that our method has no tolerance in comparing samples from different sampling rates. We can also

[1] samples rotated by more than 5° could be manually corrected during sample aquisition

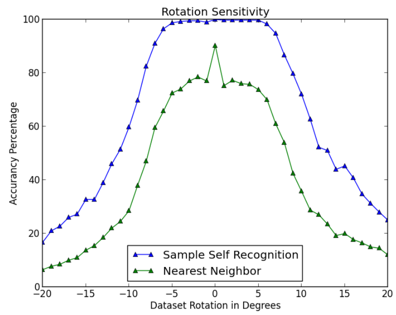


Fig. 3: Rotation Sensitivity

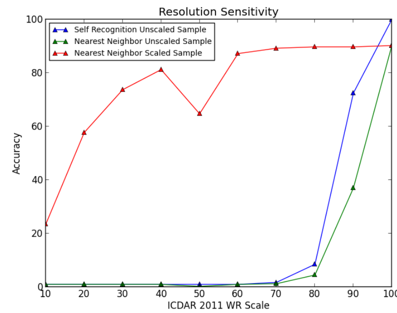


Fig. 4: Resolution/scale sensitivity

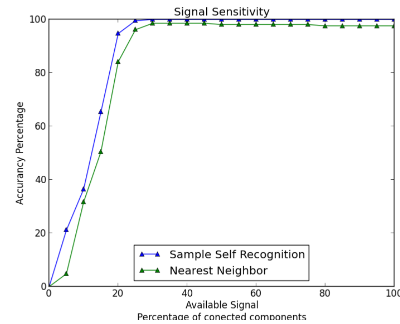


Fig. 5: Grapheme quantity sensitivity

conclude that our method has tolerance to lower than standard resolutions, but benefits mostly from higher resolutions. The out of the norm measurement in Nearest Neighbor Scaled Sample posed us with a puzzle. The most probable explanation is that it is related to aliasing but is worth investigating more.

3) Removing Graphemes: A very important characteristic of writer identification methods is how much text is required to in order to reach the claimed accuracy. We conducted an experiment to answer specifically this question. Our strategy was to create group datasets that vary only on the amount of signal (text) and then compare results on these datasets. As the primary dataset we took the ICDAR 2011 writer identification dataset, because it provides us with relatively large text samples. In order to quantify the available signal, we took the 2011 dataset and for each image in the dataset, we produced 20 images with different amounts of connected components from the original image. Due to the very high locality of our feature set, the fact that we removed connected components instead of text lines should be negligible and at the same time it gave us quite higher control over the signal quantity. As can be seen in Fig. 5 the results are quite surprising. Instead of having a gradual drop in performance, the performance is unaffected down to 30% of the graphemes, bellow that point, performance drops linearly.

4) Writer vs Writing Style: We submitted an earlier version of our method to the SigWiComp2013 competition. The goal of the writer identification part of the competition, is to measure the performance of writer identification systems, when the handwriting style has been altered. A sample dataset was made available by the organizers of the competition. The dataset contained 55 writers contributing 3 text samples each and each sample written a different writing style. Having access to the sample dataset, we performed a simple experiment to determine whether our features encapsulate writer biometrical information or simply the writing style. We separated the dataset of 165 samples in to left and right halves. We then performed a pair matching of the left halves to the right halves based on the nearest neighbor classification. We obtained two measurements, first the percentage of left-samples having an assigned right-sample written by the same writer (55 classes), and second the percentage of left-samples having the specific sample's complementary right-half as the nearest neighbor (165 classes). The writer identification rate was **87.27%**, while the specific sample recognition rate was **86.06%**. By definition the writer identification rate is greater or equal to the sample recognition rate. We performed a one tail t-test on the results on 165 sample-classifications and obtained a p-value of 0.3734, which by all standards make the recognition-rates indistinguishable. This experiment indicates that for our method any two samples written in different writing styles are as different regardless of whether they were written by the same individual or not. From a forensic perspective, these measurements imply that our method does not distinguish between disguised writing style and natural writing style.

IV. Discussion

In this paper we presented a powerful feature set that summarizes any texture on a binary image as a vector of 327. We use our feature extraction method to produce a database from any given dataset with handwriting samples and use it as a nearest neighbor classifier. In order to improve our classifier we performed PCA on a specific dataset and linearly transformed the feature space. We also scaled the features by a scaling vector in order to increase the impact of the features that contribute to correct classifications on our test set. Both these improvements can be combined in to single matrix with which we multiply all feature vectors. This single matrix should be viewed as a heuristic matrix statistically derived from the 2012 dataset. It is also valid to think of this matrix as the result of a supervised learning process. The idea is to calculate this matrix once per type of texture we want to classify. In the context of western script handwriting, we obtained the matrix from the 2012 dataset and used it in benchmarking our method on the 2011 dataset, our qualitative experiments, and our submission to SigWiComp2013 [11]. When comparing the experimental results to the state of the art, we can not obtain a perfectly fair comparison. The state of the art methods participated in a blind competition with a very small train-set, although we could maybe assume that participants had access to larger third-party datasets as well. Since datasets of competitions are published after the competitions, the only way to have a perfectly fair comparison to the state of the art is to participate in those competitions. A comparison of the untrained classifier (96.63%) to the state of the art (99.5%) is quite unfair towards our method. On the other hand, a comparison of our trained classifier (98.56%) to the state of the art (99.5%) is a bit unfair towards the state of the art. In the authors opinion, a fair comparison would be a lot closer to the trained classifier than to the untrained. The performance of the untrained classifier demonstrates clearly the potency of our feature set. The qualitative experiments were not performed with forensics in mind, except for the last one, writer vs writing style. In writer vs writing style we tried to determine the extent to which our feature set can deal with disguising writers; the quick answer is, no our method can not deal with disguising writers. There are many subtleties in the conclusions that can be drawn from the writer vs writing style experiment about what phenomena is that our features model. One could even say that our method is more about texture similarity than about writer similarity; assuming there are biometric features in handwriting, the proposed feature set does not seem to encapsulate them. From a software engineering perspective the approach of treating writer identification as a distance metric instead of a classifier [12] seems more efficient and modular, it allows for simplification and standardization of benchmarking. The fact that the proposed features encapsulate no structural information what so ever, makes them a very good candidate for fusion with other feature sets.

ACKNOWLEDGMENT

The first author of this paper would like to thank Georgios Louloudis for his precious insights on the subject of writer identification and performance evaluation. The first author would also like to thank Muhammad Imran Malik for his effort and assistance in the participation of this method on the SigWiComp2013 competition.

REFERENCES

[1] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[2] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Computer Vision-ECCV 2004*. Springer, 2004, pp. 469–481.

[3] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915–928, 2007.

[4] X. Wang, T. X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 32–39.

[5] G. A. Dean, I. W. Kelly, D. H. Saklofske, and A. Furnham, "Graphology and human judgment." 1992.

[6] A. Furnham, "Write and wrong: The validity of graphological analysis," *The Hundreth Monkey and Other Paradigms of the Paranormal*, pp. 200–205, 1991.

[7] A. Bensefia, T. Paquet, and L. Heutte, "Handwritten document analysis for automatic writer recognition," *Electronic letters on computer vision and image analysis*, vol. 5, no. 2, pp. 72–86, 2005.

[8] Z. He, X. You, and Y. Y. Tang, "Writer identification of chinese handwriting documents using hidden markov tree model," *Pattern Recognition*, vol. 41, no. 4, pp. 1295 – 1307, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320307004037

[9] L. Du, X. You, H. Xu, Z. Gao, and Y. Tang, "Wavelet domain local binary pattern features for writer identification," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 3691–3694.

[10] C. Bird, B. Found, and D. Rogers, "Forensic document examiners skill in distinguishing between natural and disguised handwriting behaviors," *Journal of forensic sciences*, vol. 55, no. 5, pp. 1291–1295, 2010.

[11] M. I. Malik, M. Liwicki, L. Alewijnse, W. Ohyama, M. Blumenstein, and B. Found, "Signature verification and writer identification competitions for on- and offline skilled forgeries (sigwicomp2013)," in *12th Int. Conf. on Document Analysis and Recognition*, Washigton, DC, USA, 2013, p. n.A.

[12] G. Louloudis, N. Stamatopoulos, and B. Gatos, "Icdar 2011 writer identification contest," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*. IEEE, 2011, pp. 1475–1479.

[13] G. Louloudis, B. Gatos, and N. Stamatopoulos, "Icfhr 2012 competition on writer identification challenge 1: Latin/greek documents," in *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*. IEEE, 2012, pp. 829–834.

[14] D. Chawki and S.-M. Labiba, "A texture based approach for arabic writer identification and verification," in *Machine and Web Intelligence (ICMWI), 2010 International Conference on*. IEEE, 2010, pp. 115–120.

Chinese Handwritten writer identification based on Structure Features and extreme learning machine

Jun Tan
Guangdong Province Key Laboratory of Computational Science
Sun Yat-Sen University
GuangZhou,PR China 510275
Email: mcstj@mail.sysu.edu.cn

JianHuang Lai
School of Information and Science and Technology
Sun Yat-Sen University
GuangZhou,PR China 510275
Email: stsljh@mail.sysu.edu.cn

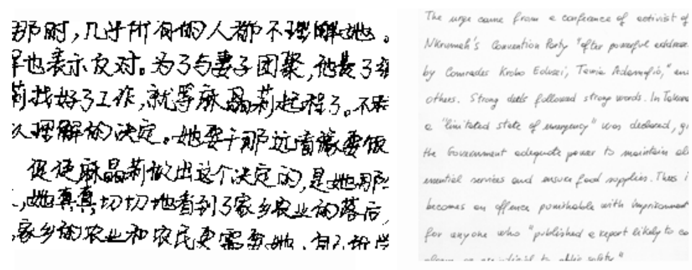
Wei-Shi Zheng
School of Information and Science and Technology
Sun Yat-Sen University
GuangZhou,PR China 510275
Email: wszheng@ieee.org

***Abstract*—In this paper,we propose a new approach for writer identification of Chinese handwritten.In our method, we deal with writer identification of Chinese handwritten using Chinese character structure features(CSF) and extreme learning machine(ELM).To extract the features embedded in Chinese handwriting characters, special structures have been explored according to the trait of Chinese handwriting characters,where 20 features are extracted from the structures, these features constitute patterns of writer handwriting. We also combine structure features with extreme learning machine (ELM) as a new scheme for writer recognition, ELM is single hidden layer feed forward networks (SLFN), which randomly chooses the input weights and analytically determines the output weights of SLFN. This algorithm learns much faster than traditional popular learning algorithms. Experimental results demonstrate CSF/ELM method can achieve better performance than other traditional schemes for writer identification.**

I. Introduction

As one of the most important methods in the biometric individual identification, writer identification has been widely used in the fields of bank check, forensic, historic document analysis, archaeology, identifying personality [1], many approaches have been developed [2], [3]. According to the different input methods, writer identification is commonly classified into on-line and off-line.

Compared with its on-line counterpart, off-line writer identification is a rather challenging problem. Chinese characters are ideo graphic in nature[4], Chinese characters can be expressed in at least two common styles, such as in block or in cursive.In block style, there is an average of 810 strokes. Meanwhile there are more strokes in cursive style. According to [5], in Chinese characters, the complication structures are mostly affected by multi stokes of each character. Additionally, as shown in Fig.1, the stroke shapes and structures of Chinese characters are quite different from those of other languages such as English, which makes it more difficult to identify Chinese handwriting [2]. The approaches proposed for English handwriting writer identification is no longer suitable for the case of Chinese handwritings [6], [7]. In this paper, we propose Chinese structure feature(CSF) as algorithm of feature extraction and combine CSF with extreme learning machine (ELM) as a new scheme for writer identification.



(a) Sample of Chinese handwriting (b) Sample of English handwriting

Fig. 1: Samples of the Chinese and English handwritings

A. Related Work

The process of writer identification consists of three main parts: preprocessing, feature extraction and identification (or matching). The feature extraction and matching are the two major topics in the literature of writer identification.

Given a free style handwritten document, a preprocessing is often required. Segmentation is an indispensable step in preprocessing. Some methods have been proposed to segment characters recently [8], we proposed a method for Chinese character segmentation based on nonlinear clustering[9].

Handwriting features except CSF feature[10] others such as texture, edge, contour and character shape have been widely studied recently. Several researchers [11], [12] proposed to take the handwriting as an image containing special texture, and therefore regarded writer identification as the texture identification. Among them, Zhu [11] adopted 2-D Gabor filtering to extract the texture features, while Chen et al. [12] used the Fourier transform. Xu and Ding[13] proposed a histogram-based feature to identify writer, called grid microstructure feature which is extracted from the edge image of the scanned images.

In [10], we propose a method for extracting Chinese structure features(CSF). Despite good performance, one serious drawback is that, it only compare one by one sample using algorithm of Similarity Matching, and it cannot classify multi-samples to different writer class. Several classifier methods have been developed to overcome the problem.

Once discriminant features have been extracted, they are submitted to a classifier whose task is to identify writer that they represent. The widely used classifiers at least include Weighted Euclidean Distance(WED) classifier [2], [11], [12], Bayesian model,BP neural networks [3], likelihood ranking [14], SVM[15]. For matching singleton non-sequential features such as texture, edge and contour, Weighted Euclidean Distance (WED) has been shown to be effective by the experiments. In [3], both Bayesian classifiers and neural networks were used as the classifiers. Imdad [15] use Steered Hermite Features to identify writer from a written document, and the algorithm takes Support Vector Machine for training and testing.

The traditional algorithms for this issue such as back-propagation (BP) need many iterative steps to calculate the optimal values of the input weights and the output weights, so their speeds are very slow in general. ELM [16] is an efficient and practical learning mechanism for the single-hidden-layer feed-forward neural networks. ELM[17] can learn the input weights and the output weights by directly calculating the Moore−Penrose generalized inverse matrix of the hidden layer output matrix of the neural network instead of using the iterative steps. So, it is necessary to perform efficient features extraction on the one hand, and to take steps to reduce the training/testing time on the other hand[18]. ELM is an efficient algorithm which tends to reach the smallest training error, obtain smallest norm of weights, produce best generalization performance, and runs faster than the iterative algorithms[19].

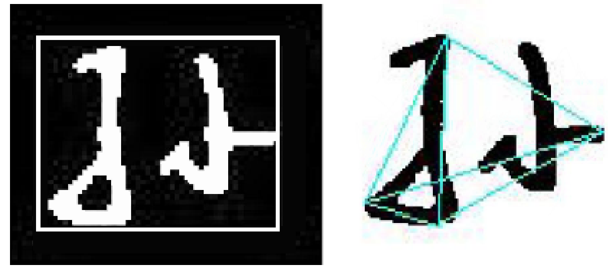
The rest of the paper is organized as follows. In Sect. II, we first briefly review algorithms of CSF feature extraction, ELM is briefly explained in Sect.III. Our proposed scheme in Sect.IV. We analyze the experimental results in Sect.V. Finally, the conclusion is given in Sect. VI.

II. Chinese Structure Features(CSF)

Features are directly extracted from each single character. Since the stroke shapes and structures of Chinese characters are quite different from those of other languages such as English, where the handwriting characteristics are embedded, we propose to utilize the stroke shapes and structures for handwriting identification.

Through a number of experiments, we discover that the discriminatory handwriting characteristics lie in the two structures[10]. They are the bounding rectangle and a special quadrilateral which we call TBLR quadrilateral, as shown in Fig.2(a) and Fig.2(b) respectively.

The following nine Chinese Structure features(CSF) are obtained from the bounding rectangle as shown in Table I. $F1$: The ratio of the width to the height of the bounding rectangle A; $F2, F3$: The relative horizontal and vertical positions of the gravity center; $F4, F5$: The relative horizontal and vertical gravity centers; $F6, F7$: The distance between the gravity center $G_1(x_1, y_1)$and the geometric center$G_2(x_2, y_2)$, and the slope of the line connecting them; $F8$: The ratio of the foreground pixel number to the area of the bounding rectangle; $F9$: The stroke width property,where P_t is the binary pixel



(a) Bounding rectangle (b) TBLR quadrilateral

Fig. 2: Two special structures of Chinese handwriting character. (a) Bounding rectangle. (b) TBLR quadrilateral.

TABLE I: CSF feature from the bounding rectangle

ith	Eqs.	Comments
1	A_w/A_h	A_w andA_h are the width and height of A.
2	$\frac{\sum_{i=1}^{A_w} i\times P_x(i)}{\sum_{i=1}^{A_w} P_x(i)}$	Foreground pixel number $i-$th vertical$P_x(i)$
3	$\frac{\sum_{j=1}^{A_h} j\times P_y(j)}{\sum_{j=1}^{A_h} P_y(j)}$	Foreground pixel number $j-$th horizontal$P_y(i)$
4	$F2/A_w$	$F2$ is 2th CSF feature
5	$F3/A_h$	$F3$ is 3th CSF feature
6	$\|G1 - G2\|$	Gravity center $G_1(x_1, y_1)$
7	$(y_2 - y_1)/(x_2 - x1)$	Geometric center$G_2(x_2, y_2)$
8	$\frac{\sum_{i=1}^{A_w}\sum_{j=1}^{A_h} \times P(i,j)}{A_w \times A_h}$	Foreground pixel number $P(i, j)$
9	$\frac{\sum_{i=1}^{A_w}\sum_{j=1}^{A_h} \times P(i,j)}{\sum_{i=1}^{A_w}\sum_{j=1}^{A_h} \times P_t(i,j)}$	Binary pixel after refining$P_t(i, j)$

after refining the preprocessed image A. Given a structuring element $B = \{C, D\}$ consisting of two elements C and D, the refining operation keeps repeating the hit-or-miss operation,

$$A \circledast B = (A\Theta C) - (A \oplus \hat{D}) \quad (1)$$

until convergence, i.e., the change stops.

Similarly, from the TBLR quadrilateral, we can obtain the following seven CSF features as shown in Table II. $F10$: The ratio of the area of the top half part $S_{up}$to the area of the whole quadrilateral S ; $F11$: The ratio of the area of the left half part S_{left} to S ; $F12$: The cosine of the angle of the two diagonal lines,where$a$ and b are the direction vectors of the two diagonal lines respectively. The$F10, F11, F12$ measure the global spatial structure of the character. $F13$: The ratio of foreground pixel number P_{inner} within the $TBLR$ quadrilateral to the total foreground pixel number P_{total} . It

TABLE II: CSF features from the TBLR quadrilateral

ith	Eqs.	Comments
10	S_{up}/S	S_{up} is the area of the top half part.
11	S_{left}/S	S_{left} is the area of left half part.
12	$\cos(a, b)$	a and b are the direction vectors of diagonal
13	P_{inner}/P_{total}	Foreground pixel number P_{inner}
14	P_{inner}/S_{TBLR}	Area of the $TBLR$ quadrilateral S_{TBLR}
15	P_{left}/P_{total}	Foreground pixel number of left half part P_{left}
16	P_{top}/P_{total}	Foreground pixel number of top half part P_{top}

measures the global degree of stroke aggregation. $F14$: The ratio of the P_{inner} to the area of the $TBLR$ quadrilateral S_{TBLR}; $F15$: The ratio of foreground pixel number of the left half part P_{left} within the $TBLR$ quadrilateral to P_{total}; $F16$: The ratio of foreground pixel number of the top half part P_{top} within the $TBLR$ quadrilateral to P_{total}.

Apart from the above sixteen features, we obtain another four CSF features as follows:

$F17$: The number of connected components. This feature measures the joined-up writing habit. $F18$: The number of hole within the character. $F19$: The number of stroke segments. It can be obtained by deleting all crossing point of a character, and the number is the total segment number. $F20$: The ratio of the longest stroke segment to the second longest stroke segment, where the stroke segments are obtained the same as that of $F19$.

III. Extreme learning machine(ELM)

For N arbitrary distinct samples (X_i, T_i), where $X_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$and $T_i = [t_{i1}, ti2, \ldots, t_{im}]^T \in R^m$, standard SLFN with$\hat{N}$ hidden neurons and activation function g(x) are mathematically modeled as follow:

$$\sum_{i=1}^{\hat{N}} \beta_i g(W_i \cdot X_j + b_i) = O_j, j = 1, 2, ..., N \tag{2}$$

where $W_i = [w_{i1}, w_{i2}, ..., w_{in}]^T$ is the weight vector connecting the ith hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{im}]^T$ is the weight vector connecting the ith hidden neuron and the output neurons, and b_i is the threshold of the ith hidden neuron. The numbers of input and output neurons are represented using n and m respectively. $W_i \cdot X_j$ denotes the inner product of W_i and X_j. The output neurons are chosen linear in this experiment.

The architecture of ELM classifier is shown in Fig.3. In the training procedure, the SLFN with $\hat{N}$ hidden neurons with activation function $g(x)$ can approximate these N samples with zero error means that $\sum_{i=1}^{\hat{N}} \|o_j - t_j\| = 0$ i.e., there exist$\beta_i, W_i$and b_i such that

$$\sum_{i=1}^{\hat{N}} \beta_i g(W_i \cdot X_j + b_i) = t_j, j = 1, 2, ..., N \tag{3}$$

The above N equations can be written compactly as:

$$\mathbf{H}\beta = \mathbf{T} \tag{4}$$

where

$$\mathbf{H}(W_1, ..., W_{\hat{N}}, b1, ..., b_{\hat{N}}, X_1, ..., X_N) = \begin{bmatrix} g(W_1 \cdot X_1 + b_1) & \ldots & g(W_{\hat{N}} \cdot X_1 + b_{\hat{N}}) \\ \vdots & \vdots & \vdots \\ g(W_1 \cdot X_N + b_1) & \ldots & g(W_{\hat{N}} \cdot X_N + b_{\hat{N}}) \end{bmatrix}_{N \times \hat{N}} \tag{5}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\hat{N}}^T \end{bmatrix}_{\hat{N} \times m} \quad and\ \mathbf{T} = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{6}$$

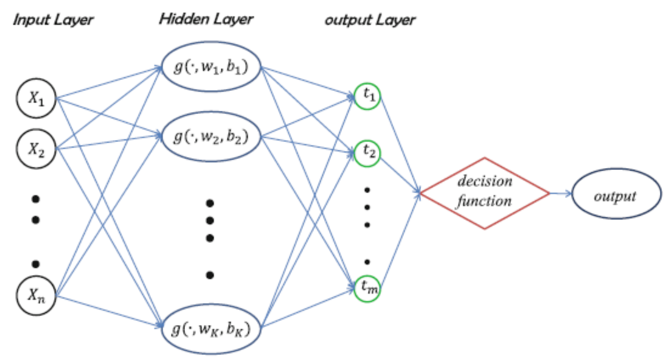


Fig. 3: The structure of Extreme Learning Machine classifier.

$\mathbf{H}$ is called the hidden layer output matrix of the neural network; the ith column of $\mathbf{H}$ is the ith hidden neuron's output vector with respect to inputs $X_1, X_2, \ldots, X_N$[16].

If the number of hidden neurons is equal to the number of distinct training samples,i.e.$\hat{N} = N$, matrix $\mathbf{H}$ is square and invertible, and SLFN can approximate these training samples with zero error. However, in most cases the number of hidden neurons is much less than the number of distinct training samples, $\hat{N} \ll N$, so $\mathbf{H}$ is a non square matrix and there may not exist $W_i, b_i, \beta_i (i = 1, \ldots, \hat{N})$ such that $\mathbf{H}\beta = T$. Thus, one may need to find specific $\hat{W}_i, \hat{b_i}, \hat{\beta}_i (i = 1, \ldots, N)$ such that

$$\|\mathbf{H}(\hat{W}_1, \ldots, \hat{W}_{\hat{N}}, \hat{b_1}, \ldots, \hat{b}_{\hat{N}})\hat{\beta} - \mathbf{T}\| = \min_{W_i, b_i, \beta_i} \|\mathbf{H}(W_1, \ldots, W_{\hat{N}}, b_1, \ldots, b_{\hat{N}})\beta - \mathbf{T}\| \tag{7}$$

the smallest norm least squares solution of the above linear system is:

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T} \tag{8}$$

Algorithm ELM: Given a training set $\aleph = (\mathbf{x}_i, \mathbf{t}_i) | x_i \in R^n, t_i \in R^m, i = 1, \ldots, N,$ activation function $g(x)$ and hidden neuron number $\hat{N}$,

Step 1: Assign arbitrary input weight $\mathbf{w}_i$ and bias b_i, $i = 1, \ldots, \hat{N}$

Step 2: Calculate the hidden layer output matrix $\mathbf{H}$.

Step 3: Calculate the output weight $\hat{\beta}$ by Eqs.(8),

where $\mathbf{H}$ and $\mathbf{T}$ are defined as Eqs. (5) and (6).

IV. Our Scheme

Some of the results in this paper were first presented in [10]. In this paper, we present more technique details effectiveness of CSF/ELM approach. Fig.4 demonstrate the flowchart of the proposed approach. There are three main steps for Chinese handwritten writer identification. The first step is handwritten image preprocessing, which removes noises and normalizes the images into the same size. The second step is feature extraction, which finds effective representation of the difference of writers in handwritten. Instead of using complex feature extraction methods, we propose Chinese structure features(CSF) for feature extraction. The last step is to apply ELM learning method to classify different writers.

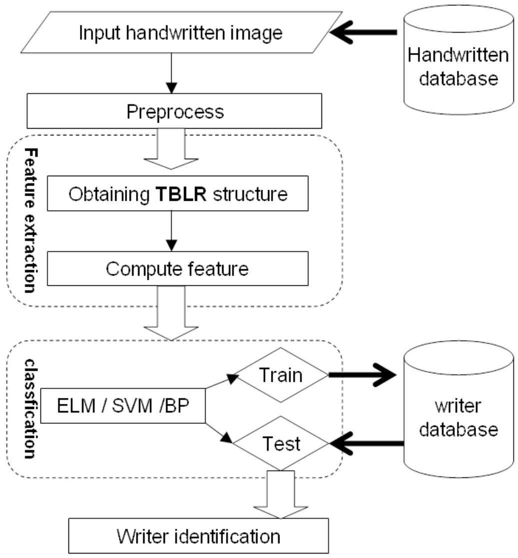


Fig. 4: The flowchart of the proposed approach.

For example, the entire process of CSF/ELM-based handwritten writer recognition is as follows:

- Step 1: The appropriate training strategy based on the selected training set, we randomly selected from a handwritten database as part of the training set $TrainSet = Si, i = 1, ..., N$, where N is the total of training samples, and the remaining samples as the test set;
- Step 2: Image pre-processing for training set, through the noise removal and standardization;
- Step 3: CSF feature extraction method to extract the optimal recognition feature vector, 20 features are extracted from structures of Bounding Box and TBLR quadrilateral;
- Step 4: ELM train phase, using **Algorithm ELM** and Eqs. (5)(6)(9), set the input weight parameters arbitrary $\mathbf{w}_i$ and bias b_i,$i = 1, \ldots, \hat{N}$ randomly, get the hidden layer output matrix **H** and the output weight β, the model of training has been trained, the training process is complete;
- Step 5: ELM test phase. Testing the model parameters obtained from the training model, and then we can obtain the actual output through the test image by the Eqs. (3), to identify the writer.

V. Experiments and Analysis

A. Handwritten Database

To test the performance of the proposed method in the writer identification, we do some experiments over 2 Chinese handwritten databases: SYSU [10] and KAIST Hanja1 [20]. Among them, SYSU database which was generated and collected by ourselves as follows, 245 volunteers were asked to sign his (or her) name and one of the others names twice, and a correction of 950 Chinese characters are obtained. The KAIST Hanja1 database contains 783 frequently used Chinese characters, where each character consisting of 200 samples written by 200 writers respectively. Fig.5 and Fig.6 shows some samples of the databases.

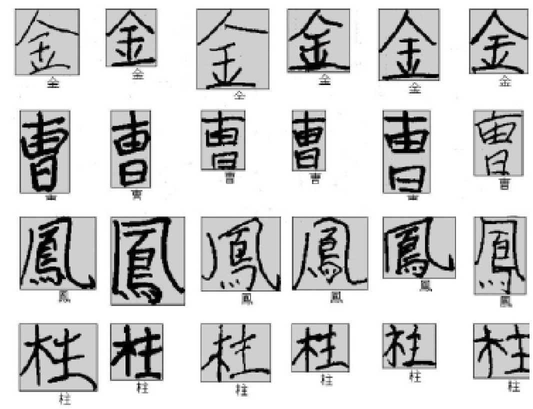


Fig. 5: Examples of the HanjaDB1 database.

Fig. 6: Examples of the SYSU database.

B. Comparative results

In experiment, because the features may have large differences in value, in order to avoid large values of features to submerge the contributions of the small value of features, all samples were normalized between 0 and 1 before sending to the learning algorithms as input.

We compare the proposed method with three well-known methods including the methods in[11], [12], [13]. Each of the compared methods is well-adjusted/trained to generate the best results. Both the recognition accuracy of writer identification and average time cost are reported and compared.

Furthermore, learning method is also an important problem. We used learning methods(ELM,BP[3],SVM[15]) for testing, and the average training time and the average test time is calculated. The cost time of the experiments is shown in table III. From the table, we can see that SVM[15] and BP[3] training times are relative much more than the ELM training time,the average total time of ELM, SVM, BP are 1.3001, 31.395, 34.801 seconds respectively. Therefore, the method of ELM has the highest speed.

TABLE III: Time cost (seconds) of different learning methods

Method	Average training time	Average test time	Average total time
ELM	1.412	0.0312	1.3001
SVM[15]	32.05	0.063	31.395
BP[3]	36.37	0.382	34.801

TABLE IV: Recognition accuracies of different methods

	Top1		Top5		Top10		Top20	
Method	Hanja	Sysu	Hanja	Sysu	Hanja	Sysu	Hanja	Sysu
Gabor[11]	54.3	53.2	61.3	64.2	67.2	68.8	72.2	73.9
Fourier[12]	42.7	52.3	54.5	61.7	64.8	70.2	72.3	74.5
GMSF[13]	74.5	75.4	82.7	83.2	85.6	88.4	87.2	91.4
Our	81.3	82.6	86.3	87.5	91.2	92.2	95.4	98.5

Finally, we compare our scheme with approaches [11], [12], [13]in the Chinese handwritten database Hanja and SYSU. These approaches use different features and classifiers. method[11] using Gabor feature and WED classifier, GMSF method[13] using GMSF feature and Weighted Chi-square, and Fourier feature and Mathematical expectations classifier are for method[12]. In the comparison, Table IV gives the Top-1, Top-5, Top-10 and Top-20 recognition accuracies of the four methods. We show the recognition accuracies of the algorithms on the handwritten database in Table IV. From the table, we can see the accuracies of our method are higher than the others. The lowest accuracy of method[12] is 42.7%, and our method has the similar accuracy of method[13]. It is obvious that our method is more effective for identifying writer in Chinese handwriting.

VI. CONCLUSION

In this paper, we list some results of the literatures. Gabor and wavelet features[11] used in the traditional methods of Chinese writer identification are affected greatly by the normalization and the arrangement of characters in texture blocks. Differently, CSF feature uses original handwriting images and tries to find out the writing structure of the writer in local regions. The ELM is a classifier used to train a single hidden layer neural network. From the experimental results. We can see that, Table IV includes the performance of our method and some other methods for Chinese writer identification. The recognition accuracy of our method using the CSF/ELM seems better than the existing methods for Chinese writer identification. compared with traditional learning algorithm, ELM has faster speed, better generalization performances. The effectiveness of CSF/ELM for Chinese writer identification is proved by the experiments.

It is expectable that our approach can be used for multilingual handwriting including western handwritings and arabic number.

ACKNOWLEDGMENT

This work was supported by Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program and Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University, the Technology Program of GuangDong (2011B061300081).

REFERENCES

[1] S. Impedovo and G. Pirlo. Verification of handwritten signatures: an overview. In *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pages 191–196. IEEE, 2007.

[2] Z. He, Y.Y. Tang, and X. You. A contourlet-based method for writer identification. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 1, pages 364–368. IEEE, 2005.

[3] M. Bulacu, L. Schomaker, and L. Vuurpijl. Writer identification using edge-based directional features. *writer*, 1:1, 2003.

[4] L. Shan. Passport to chinese: 100 most commonly used chinese characters, book 1, 1995.

[5] F.H. Cheng. Multi-stroke relaxation matching method for handwritten chinese character recognition. *Pattern recognition*, 31(4):401–410, 1998.

[6] A. Schlapbach and H. Bunke. Off-line handwriting identification using hmm based recognizers. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 654–658. IEEE, 2004.

[7] L. Schomaker and L. Vuurpijl. Forensic writer identification: A benchmark data set and a comparison of two systems [internal report for the {Netherlands Forensic Institute}]. 2000.

[8] L.Y. Tseng and R.C. Chen. Segmenting handwritten chinese characters based on heuristic merging of stroke bounding boxes and dynamic programming1. *Pattern Recognition Letters*, 19(10):963–973, 1998.

[9] J. Tan, J.H. Lai, and C.D.; Wang. A new handwritten character segmentation method based on nonlinear clustering. *NEUROCOMPUTING*, 89(8):213–219, 2012.

[10] J. Tan, J.H. Lai, C.D. Wang, and M.S. Feng. A stroke shape and structure based approach for off-line chinese handwriting identification. *International Journal of Intelligent Systems and Applications (IJISA)*, 3(2):1, 2011.

[11] Y. Zhu, T. Tan, and Y. Wang. Biometric personal identification based on handwriting. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 797–800. IEEE, 2000.

[12] Q. Chen, Y. Yan, W. Deng, and F. Yuan. Handwriting identification based on constructing texture. In *Intelligent Networks and Intelligent Systems, 2008. ICINIS'08. First International Conference on*, pages 523–526. IEEE, 2008.

[13] Lu Xu, Xiaoqing Ding, Liangrui Peng, and Xin Li. An improved method based on weighted grid micro-structure feature for text-independent writer recognition. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 638–642. IEEE, 2011.

[14] G. Zhu, Y. Zheng, D. Doermann, and S. Jaeger. Signature detection and matching for document image retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):2015–2031, 2009.

[15] A. Imdad, S. Bres, V. Eglin, H. Emptoz, and C. Rivero-Moreno. Writer identification using steered hermite features and svm. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 2, pages 839–843. IEEE, 2007.

[16] G.B. Huang, Q.Y. Zhu, and C.K. Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1):489–501, 2006.

[17] N.Y. Liang, G.B. Huang, P. Saratchandran, and N. Sundararajan. A fast and accurate online sequential learning algorithm for feedforward networks. *Neural Networks, IEEE Transactions on*, 17(6):1411–1423, 2006.

[18] X.Z. Wang and C.R. Dong. Improving generalization of fuzzy if–then rules by maximizing fuzzy entropy. *Fuzzy Systems, IEEE Transactions on*, 17(3):556–567, 2009.

[19] G.B. Huang, D.H. Wang, and Y. Lan. Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, 2(2):107–122, 2011.

[20] I.J. Kim and J.H. Kim. Statistical character structure modeling and its application to handwritten chinese character recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(11):1422–1436, 2003.

Dissimilarity Representation for Handwritten Signature Verification

George S. Eskander, Robert Sabourin, and Eric Granger
Laboratoire d'imagerie, de vision et d'intelligence artificielle
Ecole de technologie supérieure, Université du Québec
Montréal, QC, Canada
Email: geskander@livia.etsmtl.ca, robert.sabourin@etsmtl.ca, eric.granger@etsmtl.ca

***Abstract*—Signature verification (SV) systems authenticate individuals, based on their handwritten signatures. The standard approach for such systems employ feature representations (FR), where features are extracted from the signature signals and classifiers are designed in the feature space. Performance of FR-based systems is limited by the quality of employed feature representations and the quantity of training data. The dissimilarity representation (DR) approach is recently introduced to pattern recognition community, where proximity among patterns constitute the classification space. Similar concept has been applied by forensic Questioned Document Examination (QDE) experts, where proximity between questioned signatures and a set of templates lead to the authentication decision. Recently, few automatic SV systems are proposed to simulate the QDE approach, by employing DR-based pattern recognition methods. In this paper, we explore different scenarios for employing the DR approach for replacing and/or enhancing the standard SV systems. A general framework for designing FR/DR based systems is proposed, that might guide the signature processing research direction to new areas.**

I. Introduction

Signature Verification (SV) systems verify that a signature sample belongs to a specific writer. Signature signals can be acquired either online or offline. For online systems, signature dynamics such as velocity, pressure, stroke order, etc., are acquired during the signing process. Special pens and tablets are employed for the online acquisition task. On the other hand, for offline systems, signature images are scanned, after the signing process. Only static information are extracted from the signature images, producing a harder pattern recognition problem [1].

Standard SV systems employ feature-based pattern recognition approaches. Discriminative features are extracted from the signature signals, so that each signature is represented as a vector in the Feature Representation (FR) space. The classifiers are then designed in the feature space. Simply, accuracy of such systems relies on to which extend the employed feature representation is discriminative and stable. Signature representations of different users may have high similarities, when features are not discriminative enough. Also, representations of the same writer may differ significantly, when features are not stable. Besides quality of features, enough training data is required to design reliable classifiers in the feature space. The training samples should represent a wide range of genuine signatures and possible forgeries, for all system users. For real world applications, e.g., banking systems, the number of users could be very high and there is a high risk of forgery. The enrolling signature samples, available for designing such systems, are mostly few and no samples of forgeries are available. With these limitations, it is a challenge to extract informative feature representations and to design feature-based classifiers, that absorb the intra-personal variabilities while detecting both the forgeries and the inter-personal similarities.

The Dissimilarity Representation (DR) approach for pattern recognition is recently introduced, by Elzbieta Pekalska and Robert P.W. Duin., [2]. The rational behind this concept is that modeling the proximity between objects may be more tractable than modeling the objects themselves. To this end, dissimilarity measures are computed and considered as features for classification. The dissimilarity measures can be derived in many ways, e.g. from raw (sensor) measurements, histograms, strings or graphs. However, it can also be build on top of a feature representation [3].

In the field of forensic science, similar concept has been applied by the Questioned Document Examination (QDE) experts. A questioned handwritten sample is associated to a specific writer, if it is similar to a set of reference templates of his handwritings. Degree of similarity is determined by comparing a set of graphonomic features, extracted from both the questioned and template samples.

Recently, some automatic SV systems are proposed to simulate the QDE approach [4]-[7]. Distances between intra-personal training samples are computed and used as intra-class samples. Similarly, distances between inter-personal training samples are computed and used as inter-class samples. The produced distance samples are used to train a single two-class classifier, that distinguishes between intra-class and inter-class distances.

The DR approach, besides it enabled automating the forensic expert manual tasks, it alleviates some of the limitations of the FR-based design approach. First, a distance sample is generated for every pair of the original training samples, so it results in a much higher number of samples. This property alleviates the shortage of training data required to model the signatures. Second, dissimilarities between signature signals maybe more discriminative and stable than the feature representations. This is why the QDE experts build their decisions on the dissimilarity between questioned and template samples, and not on the absolute measurements of the questioned sample. Finally, the DR-approach could be applied to develop global classifiers, that are valid for all current and future users. This concept is known as Writer-Independent (WI) systems, developed by Siteargur N. Srihari et al., [8], and Santos and Sabourin et al., [4]. Instead of building a single writer-dependent (WD) classifier for each user using his enrolling signatures, a single global classifier is designed by learning the dissimilarities between signatures of all users. The rational behind the WI approach is: while it is impossible to model a feature-based class distribution that is valid for current and future users, the statistical models for inter-sample distances are generic and can be generalized for users whose signature samples are not used for training.

In this paper, we argue that the DR approach can be applied in different scenarios, in order to design more robust classifiers. It can enable the design of new family of classification systems, such as global and hybrid global/user-specific classifiers. Also, the DR approach can be employed, as an intermediate design tool, for enhanced performance of standard feature-based systems.

In the next section, the DR approach is illustrated, and a general framework for designing FR/DR based systems is proposed. Section III surveys the existing implementations of the DR approach to the offline signature classification area, and relates them to the proposed framework. Section IV discusses possible directions and areas where the DR approach can be applied.

II. General Framework For Dissimilarity-Based Classification

Although the DR is a general approach, where dissimilarity measures can be derived directly from patterns, e.g., raw (sensor) measurements, graphs, etc., we discuss here a special case where the DR is build on top of a feature representation (FR). This approach is suitable for the offline signature classification task, as many techniques of feature extraction are already proposed [1].

Figure 1 illustrates a DR constituted on top of a FR. Assume a system is designed for M different users, where for any user m there are R prototypes (templates) $\{p_{mr}\}_{r=1}^{R}$. Also, a user n provides a set of J questioned signature images $\{Q_{nj}\}_{j=1}^{J}$. The dissimilarity between a questioned sample Q_{nj} and a prototype p_{mr} is $D^{Q_{nj}p_{mr}}$. In case that questioned and prototype samples belong to the same person, i.e., $n = m$, the dissimilarity sample is an intra-personal sample (black cells in Figure 1). On the other hand, if questioned and prototype samples belong to different persons, i.e., $n \neq m$, then the dissimilarity sample is an inter-personal sample (white cells in Figure 1).

Perfect dissimilarity representation implies that all of the intra-class distances have zero values, while all of the inter-class distances have large values. This occurs when the employed dissimilarity measure absorbs all of the intra-class variabilities, and detects all of the inter-class similarities.

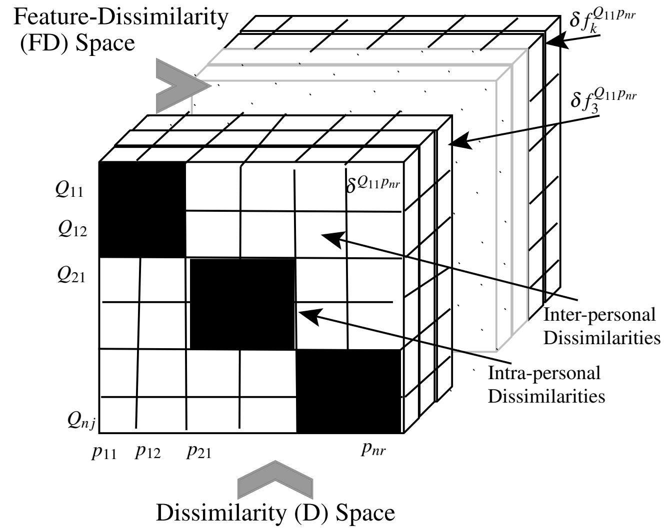


Fig. 1. Illustration of a Dissimilarity Representation (DR) built on top of a feature representation: black and white cells represent intra-personal and inter-personal dissimilarities, respectively. The third dimension represents the Feature Dissimilarity (FD) space, where dissimilarities between prototype and query signatures are measured by the distance between their feature representations. The dissimilarity cells may produce a simple dissimilarity matrix or a Dissimilarity (D) space, where distances to prototypes constitute the space dimensions. Values of FD-space vector elements control the value of corresponding dissimilarity cell.

To design a reliable classifier that works in a DR space, it is not mandatory to achieve a perfect representation, but only a discriminative one. The degree of ease to design a reliable classifier depends on the discriminative power of the representation. Accordingly, it is more important to carefully design the DR, then the classifier design comes in a next step.

In case of the DR is build on top of a FR, quality of the resulting DR relies on the quality of features that constitute the FR, and on the applied dissimilarity measure. For instance, assume the feature representations $F^{Q_{nj}} = \{f_k^{Q_{nj}}\}_{k=1}^{K}$ and $F^{p_{mr}} = \{f_k^{p_{mr}}\}_{k=1}^{K}$, are extracted from the query sample Q_{nj} (from user n) and a prototype p_{mr} (of user m), respectively. Also, consider the Euclidean distance $\delta^{Q_{nj}p_{mr}}$ as a measure of dissimilarity:

$$\delta^{Q_{nj}p_{mr}} = \sqrt{\sum_{k=1}^{K} (\delta f_k)^2}, where\ \delta f_k = \| f_k^{Q_{nj}} - f_k^{p_{mr}} \| \quad (1)$$

It is obvious that, the overall distance between feature representations of the two samples is controlled by the individual feature components, and on the reference prototypes. Accordingly, features and prototypes should be properly selected, in order to minimize the intra-personal dissimilarities and to maximize the inter-personal dissimilarities. Moreover, dissimilarity measures other than the Euclidean distance can be investigated for better dissimilarity representations.

After designing a discriminative representation, classifiers can be designed in the resulting space. Different forms of dissimilarity representation spaces can be employed. More

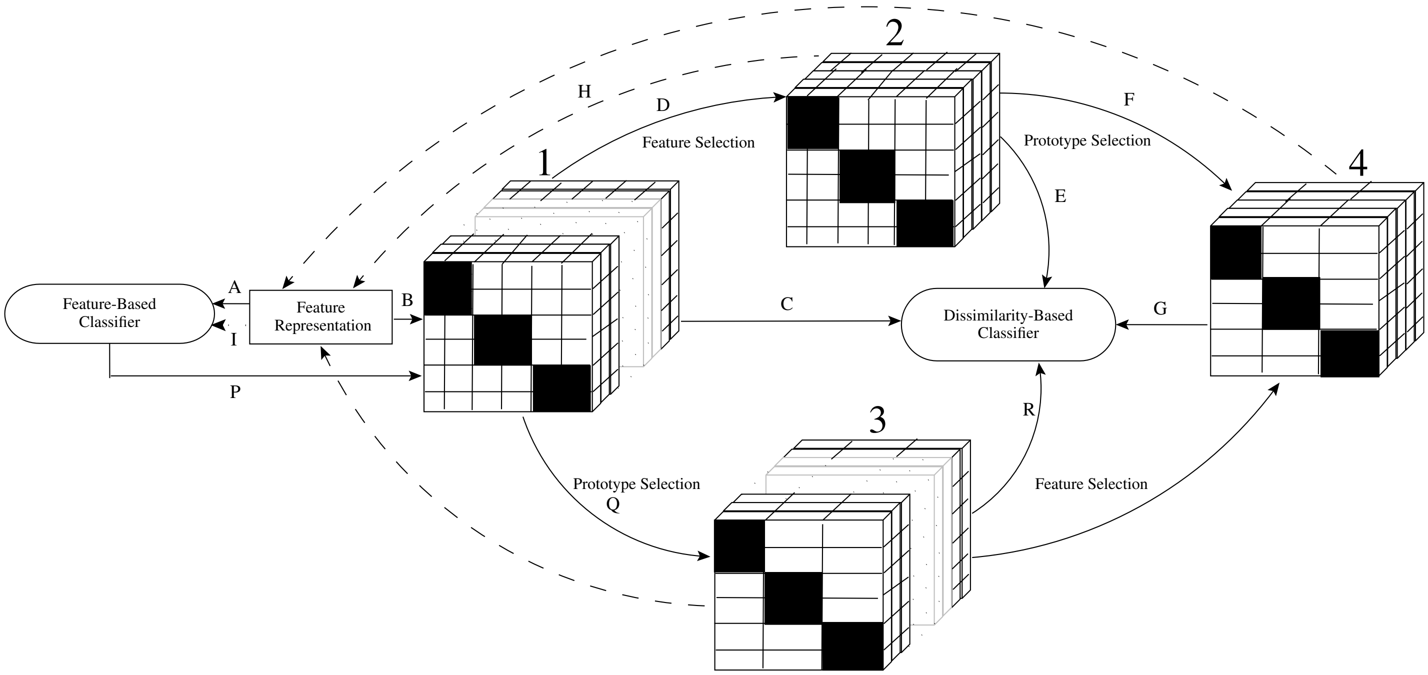


Fig. 2. General framework for designing classification systems based on the Dissimilarity Representation (DR) approach: Block 1–full FD-space-full D-space, Block 2–reduced FD-space-full D-space, Block 3–full FD-space-reduced D-space, Block 4–reduced FD-space-reduced D-space.

specifically, three different forms of dissimilarity representations (DR) can be constituted:

- Dissimilarity matrix: the matrix of all distances, where a row D^{nj} represents distances between a query j that belongs to a specific user n, with respect to the prototypes of all users:

$$D^{nj} = \{\delta^{Q_{nj}p_{11}}, .., \delta^{Q_{nj}p_{mr}}, .., \delta^{Q_{nj}p_{MR}}\}. \quad (2)$$

where $m \in [1, M]$ and $r \in [1, R]$.

- Dissimilarity space (D-Space): the dissimilarity matrix is projected on a space, where each row of the matrix is represented as a vector D^{nj} in this space. By other words, each dimension of the D-space is the distance to a specific prototype.
- Feature-Dissimilarity space (FD-Space): the dissimilarity matrix is embedded in an Euclidean space, where dimensions of this space are the dissimilarities of feature values. In the FD-space, a vector $d^{Q_{nj}p_{mr}}$, has same dimensionality as that of the original feature space, where $d^{Q_{nj}p_{mr}} = \{\delta f_k^{Q_{nj}p_{mr}}\}_{k=1}^{K}$. The length of a vector $d^{Q_{nj}p_{mr}}$ is equivalent to $\delta^{Q_{nj}p_{mr}}$, given by Eq. 1.

We argue that, classifiers can be designed in any of the aforementioned dissimilarity representation spaces. Moreover, the different tasks for feature selection, prototype selection, and classifiers design, can be done in different spaces, whenever translation between spaces is possible. This strategy permits applying a massive number of pattern recognition techniques, with multiple combinations of space transitions. We propose that new techniques for pattern recognition might be developed based on this strategy. In this context, the DR approach is employed either as a tool for enhancing the standard FR-based systems (for feature/prototype selection), or to design reliable dissimilarity-based classification systems (when classifiers are designed in a DR space).

Figure 2 illustrates a general framework for designing classification systems based on the DR approach. The standard approach is to extract feature representations from the training samples, and design classifiers in the feature space (path A in the Figure). However, the DR approach can be employed in different scenarios for either build new family of classifiers in DR-based spaces, or to enhance the performance of standard feature-based classifiers. More specifically, dissimilarities can be computed on top of a feature representation, and are used to constitute different types dissimilarity representations (DR), e.g., dissimilarity matrix, D-space, or FD-space (path B). The resulting representation could be constituted on top of a huge number of feature extractions, and based on large number of prototypes. The intra-personal (black cells) and inter-personal (white cells) dissimilarities, should be discriminative enough in order to design a DR-based classifier (path C). In case that the DR is not enough informative, feature selection and/or prototype selection can be applied for enhanced representation. For instance, feature selection can be employed in a FD-space (path D). In literature, there are many methodologies of feature selection that can be applied to select the most discriminative and stable features. The resulting DR is constituted on top of a sparser feature representation, however, redundancy in prototypes may exist (block 2). A classifier can be then designed in the resulting space (path E), or a prototype selection step is done (path F) producing a more compact representation (block 4). Surely, classifiers designed in the sparse and compact representation, are lighter and more accurate (path G). Also, order of the feature/prototype selection processes can be reversed (see the bottom part of the Figure). It is obvious that, it is more logical to run the feature selection process in the FD-space, however, the D-space is more suitable for prototype selection task. The classifier design process can be implemented based on different DR (dissimilarity matrix, D-space, or FD-space).

Besides that the DR approach can be employed to design dissimilarity-based classifiers, it can be considered as an intermediate tool for building reliable feature-based classifiers. Good features and/or prototypes can be selected in a dissimilarity-based space, then the representation is translated back to a sparser and more informative feature space (dotted paths, like path H-I). On contrary, FR-based classifiers can be designed and they are considered as an intermediate tool, to design reliable DR-based classifiers. In such case, multi-classifier systems can be designed, where FR-based classifiers are used to produce the dissimilarity measures, that are needed to build the DR (path P).

III. Current Implementations to Offline Signature Systems

The first application of the dissimilarity learning to biometrics, and more specifically, to the behavioral handwritten biometrics is proposed by Jain, A.K. et al., [10]. The dissimilarity between handwritten digits is measured by the amount of deformation required to restore a query sample to its stored prototype. This approach is extended to the author identification problem by Cha and Srihari [11], where distance statistics are used for classification. Later, similar concept is applied to the handwritten signature images. Here we list and categorize some of these implementations, and relate them to the proposed framework for DR-based classification shown in Figure 2.

A. Writer-Dependent Systems

The Writer-Dependent (WD) approach seeks to build a single classifier for each user based on his enrolling signatures. The DR concept is first introduced to design WD-SV systems, by Siteargur N. Srihari et al., [8]. Correlation between high dimensional (1024-bits) binary feature vectors, is employed as a dissimilarity measure. For a specific user, distances among every pair of his training samples, are determined to represent the intra-class samples. Also, distances between samples of the specific user and some forgeries are computed to represent the inter-class samples. The authors tried different classification strategies: one-class, two-class, discriminative, and generative classifiers. This implementation is a realization of the path B-C in Figure 2, where classifiers are designed based on the statistics of the dissimilarity matrix.

Later, Batista et al., [13] applied the dissimilarity learning concept to produce reliable WD-SV systems. A feature-based one-class classifier is built by producing user-specific generative models using Hidden Markov models (HMMs). To increase the system accuracy, a two-class discriminative model is build in DR space. The HMMs models are considered as prototypes, and samples are projected to a D-space by considering the likehood to each HMM generative model as a similarity measure. SVM classifies are then designed in the produced D-space. This implementation is a realization of the path APC in Figure 2. Also, the authors employed the AdaBoost method for classifier design in the D-space. This later implementation achieves prototype selection, while building the classifier, which is a realization of the path APQR in the Figure.

B. Writer-Independent Systems

Instead of building a single writer-dependent (WD) classifier for each user using his enrolling signatures, a single writer-independent (WI) classifier is designed by learning the dissimilarities between signatures of all users. This concept is impossible to realize by means of the standard FR approach. However, it is possible to model the class distributions of intra-class and inter-class dissimilarities, by employing the DR approach. A single ”global” classifier can be designed to model, or to discriminate between, these classes. If a huge number of samples are used to build the global DR-based classifier, it is statistically valid that the resulting model generalizes for users whose samples are not included in the training set.

The WI concept is proposed by Siteargur N. Srihari et al., [8], and Santos and Sabourin et al., [4]. While the first group used the correlation between binary features as a distance measure, the second group employed the Euclidean distance between graphometric feature vectors. This implementation is a realization of the path BC in Figure 2, where the classifiers are designed in the FD-space. Improved implementation of this concept is proposed where different dissimilarity spaces are generated based on different feature representations, and classification decisions taken in each space are fused to produce the final decision [5], [6]. This scenario can be considered as generation of different instances for path BC, and fusion is done in the score or decision levels.

More recently, Rivard et al., [7] extended the idea to perform multiple feature extraction and selection. In this work, information fusion is also performed at the feature level. Multiple graphometric features are extracted based on multiple size grids. Then, the features are fused and pairwise distances between corresponding features are computed to constitute a high dimensional feature-dissimilarity space, where each dimension represents dissimilarity of a single feature. This complex representation is then simplified by applying the boosting feature selection approach (BFS) [12]. A sparser and more discriminative FD-space is produced by applying BFS with multi-feature extraction. This scenario can be considered as realization of path BDE in Figure 2. As the resulting WI classifier recognizes all users, even the users who are enrolled after the design phase, so the feature representation embedded in the WI classifier is considered as a global ”population-based” representation.

C. Adaptation of Writer-Independent Systems

Recently, some work is done to combine advantages of both WI and WD approaches. Eskander et al., [14] extends on the system in [7] by adapting the population-based representation to each specific user, with the aim of reducing the classification complexity. While the first WI stage is designed in a FD-space, the following WD stage is designed in a standard feature space. Accordingly, the final WD classifier is FR-based classifier, that avoids storing reference signatures for enhanced security. Simulation results on two real-world offline signature databases (the Brazilian DB and GPDS public DB) confirm the feasibility and robustness of the proposed approach. Only a single compact classifier produced similar level of accuracy (Average Error Rate of about 5.38% and 13.96% for the Brazilian and the GPDS databases, respectively) as complex WI and WD systems in literature. This scenario is a realization of path BDHI in Figure 2.

IV. Research Directions

The aforementioned implementations represent a subset of large number of possible FR/DR combinations. Future research may investigate the unvisited scenarios of the proposed framework. For instance, combinations of global/user-specific, generative/discriminative, one-class/two-class systems can be designed. Also, all of the tasks for feature selection, prototype selection, classifier design, etc., can be employed in either feature space, dissimilarity matrix, FD-space, and D-space. Selection of the working space for each step, should depend on the specific requirements and constraints of the design problem and on the application itself. For example, in [14], features are selected in a FD-space as that provides a way to select reliable feature representations. Then, the classifiers are designed in a standard feature space, to avoid the need for storing signature templates for verification. Besides the large number of possible combinations and translations between the different spaces, there is also a wide range of pattern recognition techniques and tools that can be tested with the proposed framework. This includes different methods for feature extraction and selection, prototype selection, classifiers, etc.

From the application perspective, the proposed framework can be utilized for other applications, rather than the standard SV systems. For example, the Signature Identification (SI) systems that identify a producer of a signature sample, can be designed based on the DR-approach. Prototypes of all system users can be considered to build a classification D-space. Another example of systems, that imply a challenging design problem, is the signature-based bio-cryptographic systems. In these systems, cryptographic keys of encryption and digital signatures, are secured by means of handwritten signatures. It is a challenging to select informative features, signature prototype, and system parameters, for encoding reliable signature-based bio-cryptographic systems, based on the standard FR approach. Instead, recently, we proposed a methodology to design such systems, by means of the DR approach [15]. Features are selected in the FD-space and prototypes are selected in the D-space. Some of the system parameters such as length of the cryptographic key, are optimized in the different spaces.

V. Conclusions

In this paper, the dissimilarity approach for pattern recognition is considered to design signature verification (SV) systems. A general framework is proposed, for designing classification system based on a mixture of feature and dissimilarity representations. This framework imparts additional flexibility to the pattern recognition (PR) area. Combinations of transitions between different feature and dissimilarity spaces are suggested. Some of the existing implementations to the SV problem, are surveyed and related to the proposed framework. There are, however, a wide range of methodologies and applications that might benefit from the proposed approach, that opens a door for new research directions.

Acknowledgment

This work was supported by the Natural Sciences and Engineering Research Council of Canada and BancTec Inc.

References

[1] D. Impedovo and G. Pirlo., Automatic signature verification: the state of the art. *IEEE Transactions on SMC, Part C: Applications and Reviews*, vol.38, no.5, pp.609-635,2008.

[2] Elzbieta Pekalska , Robert P.W. Duin. Dissimilarity representations allow for building good classifiers. *PR Letters*, vol.23, no.8, pp.161-166, 2002.

[3] Robert P.W. Duin , Marco Loog, Elzbieta Pekalska , and David M.J. Tax. Feature-Based Dissimilarity Space Classification. *Proceedings of the 20th International conference on Recognizing patterns in signals, speech, images, and videos (ICPR'10)*, pp.46-55, 2010.

[4] C. Santos, E. Justino, F. Bortolozzi, R. Sabourin. An off-line signature verification method based on document questioned experts approach and a neural network *Proceedings of 9Th IWFHR International Workshop on Frontiers in Handwriting Recognition (IWFHR'04)*, pp.498502, 2004.

[5] L. Oliveira, E. Justino, R. Sabourin. Off-line signature using writer-independent approach. *IJCNN*, pp.25392544, 2007.

[6] D. Bertolini, L. Oliveira, E. Justino, R.Sabourin. Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers. *PR*, vol.43, no.1, pp.387396, 2010.

[7] Rivard, D, Granger, E and Sabourin, R., Multi-Feature extraction and selection in writer-independent offline signature verification. *IJDAR*, vol.16, no.1, pp.83-103, 2013.

[8] Sargur N. Srihari, Aihua Xu and Meenakshi K. Kalera. Learning Strategies and Classification Methods for Off-Line Signature Verification. *Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR'04)*,

[9] Elzbieta pekalska, Robert P.W. Duin, Pavel Paclk Prototype selection for dissimilarity-based classifiers. *PR*, vol.39, pp.189208, 2006.

[10] Jain, A.K. and Zongker, D. Representation and recognition of handwritten digits using deformable templates. *IEEE Transactions on PAMI*, vol.19, no.12, pp.1386-1390, 1997.

[11] S. Cha., Use of distance measures in handwriting Analysis. *PhD Thesis, State University of New York at Buffalo*, USA, 2001.

[12] K. Tieu and P. Viola., Boosting image retrieval. *International Journal of Computer Vision*, vol.56, no.1, pp.17-36, 2004.

[13] L. Batista, E. Granger and R. Sabourin. Applying Dissimilarity Representation to Off-Line Signature Verification . *International Conference on PR (ICPR)*, pp.1293-1297, 2010.

[14] Eskander, G.S., Sabourin, R. and Granger, E., Hybrid Writer-Independent–Writer-Dependent Offline Signature Verification System. *IET-Biometrics Journal, Special issue on Handwriting Biometrics*, doi: 10.1049/iet-bmt.2013.0024, in press, 2013.

[15] Eskander, G.S., Sabourin, R. and Granger, E., On the Dissimilarity Representation and Prototype Selection for Signature-Based Bio-Cryptographic Systems. *2nd Intel. Workshop on Similarity-Based Pattern Analysis and Recognition (SIMBAD2013)*, York, UK, 3-5 July 2013, LNCS, vol.7953, pp.265-280.

Multi-script Off-line Signature Verification: A Two Stage Approach

Srikanta Pal
School of Information and Communication Technology, Griffith University, Gold Coast Australia, Email: srikanta.pal@griffithuni.edu.au

Umapada Pal
Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India, Email: umapada@isical.ac.in

Michael Blumenstein
School of Information and Communication Technology, Griffith University, Gold Coast, Australia, Email: m.blumenstein@griffith.edu.au

***Abstract*—Signature identification and verification are of great importance in authentication systems. The purpose of this paper is to introduce an experimental contribution in the direction of multi-script off-line signature identification and verification using a novel technique involving off-line English, Hindi (Devnagari) and Bangla (Bengali) signatures. In the first evaluation stage of the proposed signature verification technique, the performance of a multi-script off-line signature verification system, considering a joint dataset of English, Hindi and Bangla signatures, was investigated. In the second stage of experimentation, multi-script signatures were identified based on the script type, and subsequently the verification task was explored separately for English, Hindi and Bangla signatures based on the identified script result. The gradient and chain code features were employed, and Support Vector Machines (SVMs) along with the Modified Quadratic Discriminate Function (MQDF) were considered in this scheme. From the experimental results achieved, it is noted that the verification accuracy obtained in the second stage of experiments (where a signature script identification method was introduced) is better than the verification accuracy produced following the first stage of experiments. Experimental results indicated that an average error rate of 20.80% and 16.40% were obtained for two different types of verification experiments.**

Keywords—Biometrics; off-line signature verification; multi-script signature identification.

I. INTRODUCTION

Biometrics are the most widely used approaches for personal identification and verification. Among all of the biometric authentication systems, handwritten signatures, a pure behavioral biometric, have been accepted as an official means to verify personal identity for legal purposes on such documents as cheques, credit cards and wills [1].

In general, automated signature verification is divided into two broad categories: static (off-line) methods and dynamic (on-line) methods [2], depending on the mode of handwritten signature acquisition. If both the spatial as well as temporal information regarding signatures are available to the systems, verification is performed using on-line [3] data. In the case where temporal information is not available and the system can only utilize spatial information gleaned through scanned or even camera-captured documents, verification is performed on off-line data [4].

Considerable research has previously been undertaken in the area of signature verification, particularly involving single-script signatures. On the other hand, less attention has been devoted to the task of multi-script signature verification. Very few published papers involving multi-script signatures, including non-English signatures, have been communicated in the field of signature verification.

Pal et al. [5] introduced a signature verification system employing Hindi Signatures. The direction of the paper was to present an investigation of the performance of a signature verification system involving Hindi off-line signatures. In that study, two important features such as: gradient feature, Zernike moment feature and SVM classifiers were employed. Encouraging results were obtained in this investigation. In a different contribution by Pal et al. [6], a multi-script off-line signature identification technique was proposed. In that report, the signatures involving Bangla (Bengali), Hindi (Devnagari) and English were considered for the signature script identification process. A multi-script off-line signature identification and verification approach, involving English and Hindi signatures, was presented by Pal et al. [7]. In that paper, the multi-script signatures were identified first on the basis of signature script type, and afterward, verification experiments were conducted based on the identified script result.

Development of a general multi-script signature verification system, which can verify signatures of all scripts, is very complicated. The verification accuracy in such multi-script signature environments will not be as successful when compared to single script signature verification [10]. To achieve the necessary accuracy for multi-script signature verification, it is important to identify signatures based on the type of script and then use an individual single script signature verification system for the identified script [10]. Based on this observation, in the proposed system, the signatures of three different scripts are separated to feed into the individual signature verification system. On the other hand to get a comparative idea, multi-script signature verification results on a joint English, Hindi and Bangla dataset, without using any script identification, is also investigated.

The remainder of this paper is organized as follows. The multi-script signature verification concept is described in Section II. Section III introduces the notable properties of Hindi and Bangla script. The Hindi, Bangla and English signature database used for the current research is described in Section IV. Section V briefly presents the feature extraction techniques employed in this work. The classifier details are described in Section VI. The experimental settings are presented in Section VII. Results and a discussion are provided in Section VIII. Finally, conclusions and future work are discussed in Section IX.

II. MULTI-SCRIPT SIGNATURE VERIFICATION CONCEPT

When a country deals with two or more scripts and languages for reading and writing purposes, it is known as a multi-script and multi-lingual country. In India, there are officially 23 (Indian constitution accepted) languages and 11 different scripts.

In such a multi-script and multi-lingual country like India, languages are not only used for writing/reading purposes but also applied for reasons pertaining to signing and signatures. In such an environment in India, the signatures of an individual with more than one language (regional language and international language) are essentially needed in official transactions (e.g. in passport application forms, examination question papers, money order forms, bank account application forms etc.). To deal with these situations, signature verification techniques employing single-script signatures are not sufficient for consideration. Therefore in a multi-script and multi-lingual scenario, signature verification methods considering more than one script are necessarily required.

Towards this direction of verification, the contribution of this paper is twofold: First, multi-script signature verification considering joint datasets as shown in Figure 1, the second is identification of signatures based on script, and subsequent verification for English, Hindi and Bangla signatures based on the identified script result. A diagram of this second verification mode is shown in Figure 2.

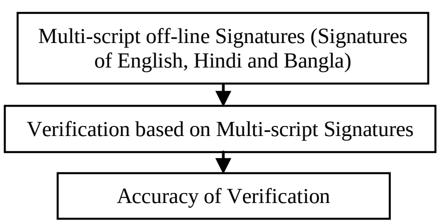


Figure 1. Diagram of signature verification considering a joint dataset.

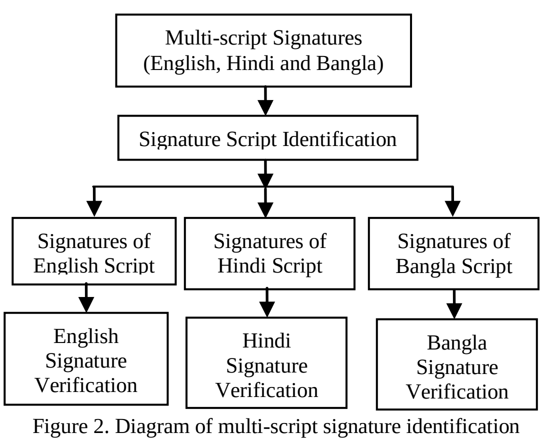


Figure 2. Diagram of multi-script signature identification and verification based on English, Hindi and Bangla signatures.

III. PROPERTIES OF HINDI AND BANGLA SCRIPT

Most of the Indian scripts including Bangla and Devanagari have originated from ancient Brahmi script through various transformations and evolution [8]. Bangla and Devanagari are the two most accepted scripts in India. In both scripts, the writing style is from left to right and there is no concept of upper/lower case. These scripts have a complex composition of their constituent symbols. The scripts are recognizable by a distinctive horizontal line called the 'head line' that runs along the top of full letters, and it links all the letters together in a word. Both scripts have about fifty basic characters including vowels and consonants.

IV. DATABASE USED FOR EXPERIMENTATION

A. Hindi and Bangla Signature Database

As there has been no public signature corpus available for Hindi and Bangla script, it was necessary to create a database of Hindi and Bangla signatures. The Hindi and Bangla signature databases used for experimentation consisted of 50 sets per script type. Each set consists of 24 genuine signatures and 30 skilled forgeries. Some genuine signature samples of Hindi and Bangla, with their corresponding forgeries, are displayed in Table 1 and Table 2.

B. GPDS English Database

Another database, consisting of 50 sets from GPDS-160 [9], was also utilised for these experiments. Each signature set of this corpus consists of 24 genuine signatures and 30 simple forgeries. The reason 50 sets were used from the GPDS on this occasion, is due to the fact that the Bangla and Hindi datasets described previously were comprised of 50 sets each, and it was considered important to have equivalent signature numbers for experimentation.

TABLE 1. SAMPLES OF HINDI GENUINE AND FORGED SIGNATURES

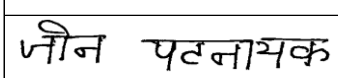
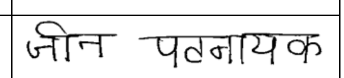
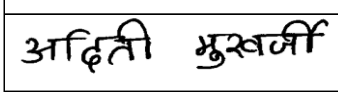
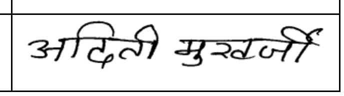
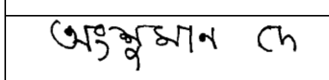
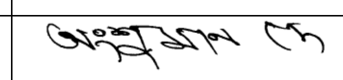
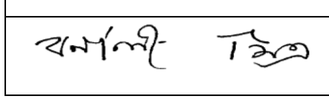
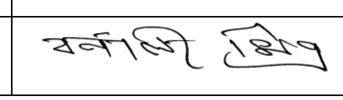
Genuine Signatures	Forged signatures
	
	

TABLE 2. SAMPLES OF BANGLA GENUINE AND FORGED SIGNATURES

Genuine Signatures	Forged signatures
	
	

V. FEATURE EXTRACTION

Feature extraction is a crucial step in any pattern recognition system. Two different types of feature extraction techniques such as: gradient feature extraction and the chain code feature are considered here.

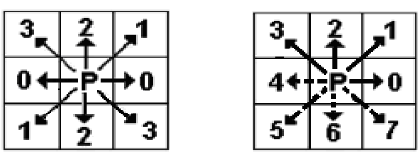
A. Computation of 576-dimensional gradient Features

576-dimensional gradient features were extracted for this research and experimentation, which are described in paper [7].

B. 64-Dimensional Chain Code Feature Extraction

The 64-dimensional Chain Code feature is determined as follows. In order to compute the contour points of a two-tone image, a 3 x 3 window is considered surrounding the object point. If any one of the four neighbouring points (as shown in Fig. 3 (a)) is a background point, then this object point (P) is considered as a contour point. Otherwise it is a non-contour point.

The bounding box (minimum rectangle containing the character) of an input character is then divided into 7 x 7 blocks. In each of these blocks, the direction chain code for each contour point is noted and the frequency of the direction codes is computed. Here, the chain code of four directions only [directions 1 (horizontal), 2 (45 degree slanted), 3 (vertical) and 4 (135 degree slanted)] is used. Four chain code directions are shown in Fig. 3 (b). It is assumed that the chain code of directions 1 and 5, 2 and 6, 3 and 7, 4 and 8, are the same. Thus, in each block, an array is obtained of four integer values representing the frequencies, and those frequency values are used as features. Thus, for 7 x 7 blocks, 7 x 7 x 4= 196 features are obtained. To reduce the feature dimensions, after the histogram calculation into 7 x 7 blocks, the blocks are down-sampled with a Gaussian filter into 4 x 4 blocks. As a result, 4 x 4 x 4 = 64 features are obtained for recognition. To normalize the features, a maximum value of the histograms from all the blocks, is computed. Each of the above features is divided by this maximum value to obtain the feature values between 0 and 1.



(a) (b)

Figure 3. Eight neighbours (a) For a point P and its neighbours (b) For a point P the direction codes for its eight neighbouring points.

VI. Classifier Details

Based on these features, Support Vector Machines (SVMs) and the Modified Quadratic Discriminant Function (MQDF) are applied as the classifiers for the experiments.

A. SVM Classifier

For this experiment, a Support Vector Machine (SVM) classifier is used. The SVM is originally defined for two-class problems and it looks for the optimal hyper plane, which maximizes the distance and the margin, between the nearest examples of both classes, named support vectors (SVs). Given a training database of M data: $\{x_m| \; m=1,..., M\}$, the linear SVM classifier is then defined as:

$$f(x) = \sum_j \alpha_j x_j \cdot x + b$$

where $\{x_j\}$ are the set of support vectors and the parameters α_j and b have been determined by solving a quadratic problem [11]. The linear SVM can be extended to various non-linear variants; details can be found in [11, 12]. In these proposed experiments, the Gaussian kernel SVM outperformed other non-linear SVM kernels, hence identification/verification results based on the Gaussian kernel are reported only.

B. MQDF Classifier

The Modified Quadratic Discriminant Function is defined as follows [13].

$$D(X) = (N + N_0 + n - 1) In\left[1 + \frac{1}{N_0\sigma^2}\left[\|X - M\|^2 - \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_i \frac{N_0}{N}\sigma^2}\right]\right] + \sum_{i=1}^{k} \mathrm{In}\left(\lambda_i + \frac{N_0}{N}\sigma^2\right)$$

where X is the feature vector of an input character; M is a mean vector of samples; Φ_i^T is the i[th] eigen vector of the sample covariance matrix; λ_i is the i[th] eigen value of the sample covariance matrix; k is the number of eigen values considered here; n is the feature size; σ^2 is the initial estimation of a variance; N is the number of learning samples; and N_0 is a confidence constant for s and N_0 is considered as 3N/7 for experimentation. All the eigen values and their respective eigen vectors are not used for classification. Here, the eigen values are stored in descending order and the first 60 (k=60) eigen values and their respective eigen vectors are used for classification. Compromising on trade-off between accuracy and computation time, k was determined as 60.

VII. Experimental Settings

A. Settings for Verification used in 1[st] Stage of Experiments

The skilled forgeries were not considered for training purposes. For experimentation, random signatures were considered for training purposes. For each signature set, an SVM was trained with 12 randomly chosen genuine signatures. The negative samples for training (random signatures) were the genuine signatures of 149 other signature sets. Two signatures were taken from each set. In total, there were 149x2=298 random signatures employed for training. For testing, the remaining 12 genuine signatures and 30 skilled forgeries of the signature set being considered were employed. The number of samples for training and testing for these experiments are shown in Table 3.

Table 3. No. of Signatures used per set in 1st Phase of Verification

	Genuine Signature	Random Signatures	Skilled Forgeries
Training	12	298	n/a
Testing	12	n/a	30

B. Settings for Verification used in 2^{nd} Stage of Experiments

1) Settings for Signature Script Identification

150 sets of signatures (50 sets of English, 50 sets of Hindi and 50 sets of Bangla) were used for signature script identification. 30 sets of signatures from each script were considered for training, and the remaining 20 sets were considered for testing purposes. The number of samples for training and testing used in experimentation of the identification approach are shown in Table 4.

TABLE 4. SIGNATURE SAMPLES USED FOR SCRIPT IDENTIFICATION PHASE.

	English Signatures		Hindi Signatures		Bangla Signatures	
	Genuine	Forged	Genuine	Forged	Genuine	Forged
Training	720	900	720	900	720	900
Testing	480	600	480	600	480	600

2) Settings for Signature Verification after Signature Script Identification

The verification task in the second stage was explored separately for English signatures, Hindi signatures and Bangla signatures based on the identified script result. Signature samples (30 sets from each script) that were considered for training purposes in signature script identification were not used for the individual verification task. Only the correctly identified samples from 20 sets (used for the testing part in identification) were considered for verification. For each signature set, an SVM was trained with 12 genuine signatures. The negative samples for training were 95 (19x5) genuine signatures of 19 other signature sets.

VIII. RESULTS AND DISCUSSION

A. Experimental Results

1) First Verification Experiments

In this stage of experimentation, 8100 (150x54) signatures involving English, Hindi and Bangla scripts were employed for training and testing purposes. At this operational point, the SVMs produced an AER of 20.80%, and an encouraging accuracy of 79.20% was achieved in this first mode of verification.

2) Second Verification Experiments

In this stage of verification the signatures are identified based on their script and subsequently, the identified signatures are applied separately for verification. In the signature script identification stage, only 64-dimensional chain code features were used because a slightly better accuracy was obtained when compared to the gradient feature. The MQDF classifier was also taken into account in the script identification step applying chain code features for a better accuracy, but MQDF did not achieve the better result as compared to SVMs in this study. To get a comparative idea, script identification results using two different classifiers with chain code features are shown in Table 5. An accuracy of 93.08% is achieved at the script identification stage by using the SVM classifier. The accuracy of Bangla, English and Hindi are 85.19, 95.74 and 98.33% respectively. Confusion matrices obtained using SVM classifiers, and the 64-dimensional chain code features investigated, are shown in Table 6.

TABLE 5. ACCURACY OBTAINED USING SVM AND MQDF CLASSIFIERS

Classifiers	**Identification Accuracy (%)**
SVMs	93.08
MQDF	82.45

TABLE. 6. CONFUSION MATRIX OBTAINED USING THE CHAIN CODE FEATURE AND SVM CLASSIFIER

	Bangla	English	Hindi
Bangla	920	19	141
English	27	1034	19
Hindi	10	8	1062

Based on the outcomes of the identification phase, verification experiments subsequently followed. Verification results obtained for individual scripts were calculated on 93.08% (identification rate) accuracy levels. In this phase of experimentation, the SVMs produced an overall AER of 21.10%, 13.05% and 15.05% using English, Hindi and Bangla signatures respectively. The overall verification accuracy obtained for the second major experiments (identification plus verification) was 83.60% (average of 78.90% of English, 86.95% of Hindi and 84.94% of Bangla).

B. Comparision of Performance

From the experimental results obtained, it was observed that the performance of signature verification in the second set of experiments (identification and verification) was encouraging compared to the signature verification accuracy from the first experiment set (verification only). Table 7 demonstrates the accuracies attained in the first experiment set as well as separate verification results for English, Hindi and Bangla from the second experiment set.

TABLE 7. VERIFICATION ACCURACIES RESULTING FROM DIFFERENT EXPERIMENTS

Verification Techniques		**Accuracy (%)**
Experiment Sets	**Dataset Used**	
1^{st} experiment	**English, Hindi and Bangla**	79.20
2^{nd} experiment	**English**	78.90
	Hindi	86.95
	Bangla	84.94

In the second stage of verification, the overall accuracy is 83.60% (Avg. of 78.90%, 86.95% and 84.94%) which is 4.40 (83.60-79.20) higher than the accuracy in the first stage. The comparison of these two accuracies is shown in Table. 8.

TABLE 8. ACCURACY IN DIFFERENT PHASES OF VERIFICATION

Verification Experiment	Verification Accuracy (%)
Without Script Identification	79.20
With Script Identification	83.60

From the above table it is evident that verification accuracy with script identification is much higher than without script identification. This increased accuracy is achieved because of the proper application of the identification stage. This research clearly demonstrates the importance of using identification in multi-script signature verification techniques.

C. *Error Analysis*

Most of the methods used for signature verification generate some erroneous results. In these experiments, a few signature samples were mis-identified in both the identification and verification stages. Few of the confusing signature samples obtained in the signature script identification stage using the SVM classifier are shown in Figures 4, 5 and 6. Three categories of confusing samples are generated by the classifier. The first category illustrates a Bangla signature sample treated as a Hindi signature sample. The second one illustrates an English signature sample treated as a Bangla signature sample and the third one illustrates a Hindi signature sample treated as a Bangla signature sample.

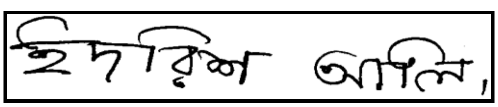


Figure 4. Bangla sample treated as Hindi

Figure 5. English treated as Bangla

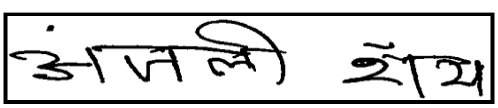


Figure 6. Hindi Signature treated as Bangla

IX. CONCLUSIONS AND FUTURE WORK

This paper provides an investigation of the excellent performance of a multi-script signature verification technique involving English, Hindi and Bangla off-line signatures. The novel approach used in a multi-script signature verification environment with the combination of a custom Hindi and Bangla off-line signature dataset provides a substantial contribution to the field of signature verification. In such a verification environment, the proper utilization of a script identification technique, which substantially affects the verification accuracy, indicates an important step in the process. The comparatively higher verification accuracy obtained in the second experimental approach is likewise a substantial contribution. The gradient feature, chain code feature as well as SVM and MQDF classifiers were employed for experimentation. The idea of a multi-script signature verification approach, which deals with an identification phase, is a very important contribution to the area of signature verification. The proposed off-line multi-script signature verification scheme is a new investigation in the field of off-line signature verification. In the near future, we plan to extend our work considering further sets of signature samples, which may include different languages/scripts.

X. ACKNOWLEDGMENTS

Thanks to my colleague Mr. Nabin Sharma for his help towards the preparation of this paper.

REFERENCES

[1] R. Plamondon and G. Lorette, “Automatic signature verification and writer identification - the state of the art”, Pattern Recognition, pp.107–131, 1989.

[2] S. Madabusi, V. Srinivas, S. Bhaskaran and M. Balasubramanian, “On-line and off-line signature verification using relative slope algorithm”, in proc. International Workshop on Measurement Systems for Homeland Security, pp. 11-15, 2005.

[3] D. Impedovo, G. Pirlo, “ Automatic signature verification: The state of the art”, IEEE transactions on Systems, Man, and Cybernetics part-C, vol. 38, no. 5, pp. 609–635, 2008.

[4] M. Kalera, S. Srihari and A. Xu. “Offline signature verification and identification using distance statistics”, International Journal on Pattern Recognition and Artificial Intelligence, pp.1339-1360, 2004.

[5] S. Pal, U. Pal, M. Blumenstein, “Hindi Off-line Signature Verification”, in proc. International Conference on Frontiers in Handwritten Recognition, ICFHR 2012, Bari, Italy, pp. 371-376.

[6] S. Pal, A. Alaei, U. Pal, M. Blumenstein, “Multi-Script off-line signature identification” , in proc. International Conference on Hybrid Intelligent Systems, pp. 236-240, 2012.

[7] S. Pal, U. Pal, M. Blumenstein, “Hindi and English off-line signature identification and verification”, in proc. International Conference on Advances in Computing. pp. 905–910, 2012.

[8] B. B. Chaudhuri and U. Pal, “An OCR system to read two Indian language scripts: Bangla and Devnagari (Hindi)”, in proc. International Conference on Document Analysis and Recognition, pp. 1011–1015, 1997.

[9] M. A. Ferrer, J. B. Alonso, and C. M. Travieso, “Offline geometric parameters for automatic signature verification using fixed-point arithmetic”, IEEE transactions on Pattern Analysis and Machine Intelligence, 27:993–997, 2005.

[10] S. Pal, U. Pal and M. Blumenstein, “A Two-Stage Approach for English and Hindi Off-line Signature Verification”, International workshop on Emerging Aspects in Handwritten signature processing, 2013(Acceoted).

[11] V.Vapnik, “The Nature of Statistical Learning Theory”, Springer Verlang, 1995.

[12] C. Burges, “A Tutorial on support Vector machines for pattern recognition”, Data Mining and Knowledge Discovery, pp.1-43, 1998.

[13] F. Kimura, K. Takashina, S. Tsuruoka and Y. Miyake, “Modified quadratic discriminant function and the application to Chinese character recognition”, IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 9, pp 149-153, 1987.

Off-Line Signature Verification based on Ordered Grid Features: An Evaluation

Konstantina Barkoula, George Economou
Physics Department
University of Patras
Patras, Greece
email: kbarkoula@gmail.com, economou@upatras.gr

Elias N. Zois, Evangelos Zervas
Electronics Engineering Department
Technological and Educational Institution of Athens
Egaleo, Greece
e-mail: {ezois, ezervas}@teiath.gr

***Abstract*— A novel offline signature modeling is introduced and evaluated which attempts to advance a grid based feature extraction method uniting it with the use of an ordered powerset. Specifically, this work represents the pixel distribution of the signature trace by modeling specific predetermined paths having Chebyshev distance of two, as being members of alphabet subsets-events. In addition, it is proposed here that these events, partitioned in groups, are further explored and processed within an ordered set context. As a proof of concept, this study progresses by counting the events' first order appearance (in respect to inclusion) at a specific powerset, along with their corresponding distribution. These are considered to be the features which will be employed in a signature verification problem. The verification strategy relies on a support vector machine based classifier and the equal error rate figure. Using the new scheme verification results were derived for both the GPDS300 and a proprietary data set, while the proposed technique proved quite efficient in the handling of skilled forgeries as well.**

Grid Features, Power Set, Ordering, Signature Verification

I. Introduction

Automated handwritten signature verification systems (ASVS) remain up to now an accepted way for humans to declare their identity in many application areas including civilian ones [1], [2], [3], [4]. ASVS are separated into two major categories based on the method that the signature is obtained. Both online and offline ASVS must cope with the evidence that the process of creating handwritten signatures, even when they originate from a well trained genuine writer, will carry natural variations, defined as intra-writer variability [5]. It is adopted that the online ASVS are generally more efficient when compared to offline. A commonly used figure of merit which is employed in order to characterize the efficiency of ASVS is the equal error rate (EER) which is calculated from the ROC or DET plots of both types of error rates.

The goal of an offline ASVS is to efficiently transform an image into a mathematical measurable space where it will be represented by means of its corresponding features [6]. Next, the features are feeding computational intelligence techniques and pattern recognition classifiers which will decide, after appropriate training and testing procedures, if a signature under query belongs to a claimed writer [7], [8]. According to the experimental protocol followed, there are two major approaches which have been applied to off-line ASVS; writer dependent (WD) and writer-independent (WI). The WD approach uses an atomic classifier for each writer. The WI approach uses a classifier to match each input questioned signature to one or more reference signatures, and a single classifier is trained for all writers [9], [10].

Feature extraction is considered to be one of the most challenging tasks when ASVS are designed. An important feature extraction philosophy which attracts increasing interest, exploits the signature using a coarse or fine detail grid which is imposed upon the image. Among others, examples of grid based feature extraction can be found in the work provided by references [10], [11], [12], [13], [14], [15], [16], [17], [18] and [19].

In another work provided by Tselios, Zois, Nassiopoulos and Economou [20], a grid based feature extraction method was developed which represents the signature trace by taking into account the histogram of specific pixel path transitions along predefined paths within pre-confined Chebyshev distances of two (F_{CB2} feature). The feature extraction concepts have been advanced by describing these paths in a way in which they can be viewed as symbols transmitted by a discrete space random source. The combination of the produced F_{CB2} symbols defines the message or event that the random source sends out when a certain sequence of signature pixels is accounted. They are treated according to the event concept, reported in standard set and information theory and they are complemented along with their corresponding probabilistic moments [21]. In this work and in order to further increase our signature discriminating capability the potential messages-events of the F_{CB2} paths are organized in sub-groups of independent tetrads. Each tetrad is organized according to its ordered powerset with respect to inclusion [22]. The outcome of this procedure provides an attempt to model the handwriting process in concordance with basic elements of information and coding theory.

The distributions of the now ordered transition paths in the new feature space are used to code the signature image. In the case study presented here a WD verification scheme is followed which comprises of the training and testing phase. Verification results have been drawn with the use of two databases, the GPDS300 and a proprietary one by means of the false acceptance, false rejection and the equal error rate (EER) figure of merit. The rest of this work is organized as follows: Section 2 provides the database details and the description of the feature extraction algorithm. Section 3 presents the experimental verification protocol which has been applied. Section 4 presents the comparative evaluation results while section 5 draws the conclusions.

II. Database And Feature Extraction Procedure

A. Database Description

The proposed feature extraction modeling has been studied with the use of two databases of 8-bit grey scale signatures: a Greek signers' database (CORPUS1) [20] and GPDS-300 (CORPUS2) [12]. CORPUS1 comprises of a domestic Greek collection of 105 genuine and 21 simulated forgery signature samples for each of the 69 signers of the database. Genuine samples were acquired in a one month time frame. CORPUS2 contains 24 genuine signatures and 30 simulated forgeries for each of the 300 signers of the database and is publicly available. During the experimental process, two schemes of randomly selected training and testing samples were used for comparison with the outcomes of contemporary research in the field. In the first scheme, 12 genuine and 12 simulated-forgery reference samples per writer are used, while in the second scheme 5 genuine and 5 simulated forgery reference samples are used. The remaining samples are used for testing.

B. Preprocessing

In order to produce the binary form of the acquired signatures the following preprocessing steps have been carried out: thresholding using Otsu's method [6], skeletonization, cropping and segmentation. This procedure is expected to reduce a number of side effects of the writing instruments variations. The result is the generation of the most informative window (MIW) of the image. The features are extracted either from the whole MIW of the signature or from segments of signature's MIW with the use of the equimass sampling grid method [14]. Equimass sampling grid segmentation provides strips of the signature with uniform size of signature pixels instead of the trivial distance grid segmentation which provides segments of equal area. The result is depicted in Fig. 1. In this work the feature vector is generated from the 'S2' scheme used in [20].

C. Alphabet Description

Fig. 2 depicts the alphabet which is defined as a set of symbols, emerging from the F_{CB2} description according to [12]. To be more specific, F_{CB2} alphabet is the set of transition paths of three consecutive pixels under the constraint of having the first and third pixels restrained to a Chebyshev distance equal to two.

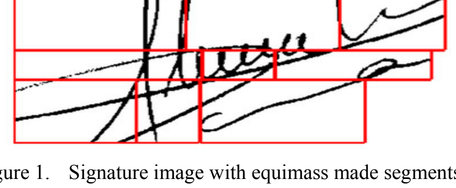


Figure 1. Signature image with equimass made segments

Since, in offline signatures, signature-pixel ordering is unknown, the ordered sequence of the pixels cannot be estimated. This note diminish the number of queried F_{CB2} transition paths, in a 5x5 pixel grid window, with center pixel each black pixel of signature's image, to the sixteen independent transition paths presented in Fig. 2. In this case study only the F_{CB2} paths have been taken into account. It is advantageous in our case to explicitly treat the notion of the signature pixels indexes (i,j) as a transformation of sequences produced by the source. As a consequence, the feature extraction grid can be identified as a discrete space – discrete alphabet source.

D. Ordered Event Modeling

Let the triad (Ω, B, P) indicate the probability space on which all the potential outcomes are identified. By definition Ω is the sample space upon which a discrete digital source transmits alphabet symbols. The source may transmit either single symbols or sets of them (events) from a 16 symbol alphabet as figure 2 illustrates. Let B a sigma field (the event space) that encloses all potential occurrences of symbols combinations from the F_{CB2} alphabet. That is, B is the largest possible σ-field [23] which is the collection of all subsets of Ω and is called the power set. Finally, let P be the corresponding distributions of the σ-field.

In order to evade the problem of 2^{16} space management Ω is grouped into T subsets $\{\Omega_t\}_{t=1,\ldots,T}$ and we define the sub-s-fields B_t as the power sets for each Ω_t. In this work we choose to group the 16-F_{CB2}(i) components into ensembles of four tetrads (call it hereafter F_4-collection) thus resulting to an early set of 4×2^4=64 possible event combinations. From the complete set of all the possible ensembles of the F_4 collection only 87 orthogonal cases shall be enabled along with their corresponding probabilities. From a mathematical point of view the signature image is analyzed into four major subspaces where each of them is composed of 16 orthogonal dimensions. The term orthogonal denotes that each symbol in a sub-alphabet space of a F_4 tetrad cannot be derived as any combination of the same subspace F_4 symbols. This constraint provides each signature with 87 different F_4 orthogonal tetrad event sets, found through exhaustive search. Fig. 3 provides the F_{CB2} alphabet along with a F_4 orthogonal collection. As a proof of concept, the orthogonal F_4 collection #44, selected randomly is illustrated in figure 3.

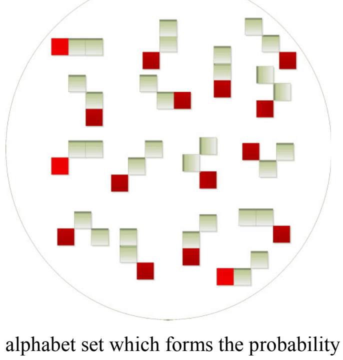


Figure 2. F_{CB2} alphabet set which forms the probability space Ω .

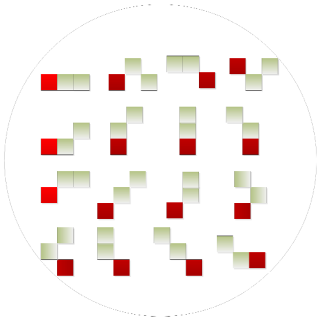


Figure 3. One F_4 collection of tetrads (#44). Each horizontal tetrad is considered to form a subspace in the original 16-dimnensional feature space and consequently generates a powerset of events

Finally, each one of the four F_4 power-sets of figure 3b is evaluated by ordering the elements of the powerset with respect to inclusion. Fig. 4 provides a graphical explanation of one powerset in line with the proposed modeling. In order to illustrate the method with clarity, figure 4 has been created which shows the powerset of the #44 F_4 collection with respect to inclusion. The indexes x, y, z, w are associated with one tetrad's elements of the F_4 collection. For each arrow in figure 4 there is a corresponding probability evaluated for every segmented image. Thus, the overall dimensionality of the feature vector for one F_4 collection is equal to 32 (4+12+12+4) for each image segment.

According to the exposed material, a discrete source, designated as S_n, can be defined by its transmitted set of symbols-events which are now members of an ordered F_4 collection. This novel modeling of the feature generation process is an evolution of the previous method as it was described in [20]. It attempts to model the distribution of the signature pixel paths as an information source and to associate events of ordered paths (arrows as seen in fig. 4) along with their corresponding first order probabilities.

E. *Creation of the ordered feature vector*

To make this work robust a short description is provided for generating the ordered feature components. According to the material exposed in sections II_C, II_D, each one of the preprocessed image segments is scanned top-down and left-right to identify its signature pixels. Let us denote with the labels One (O) and Two (T) a conjugated pair of 5×5 moving grids with the property that their topological centers are distant by a Euclidean distance of one. Then for each signature pixel the {O, T} grids are imposed. Next, detection of discrete events at both {O, T} grids is performed followed by the evaluation of the corresponding ordered probabilities, as described in fig. 4. In addition, fig. 5 presents in a graphical manner the generation of a feature component namely the {X, XY}. In this work the overall feature dimensionality is 128 due to the selection of the segmentation preprocessing steps.

III. CLASSIFICATION PROTOCOL

On the grounds of proofing the proposed concept and according to the discussion exposed in section II the training

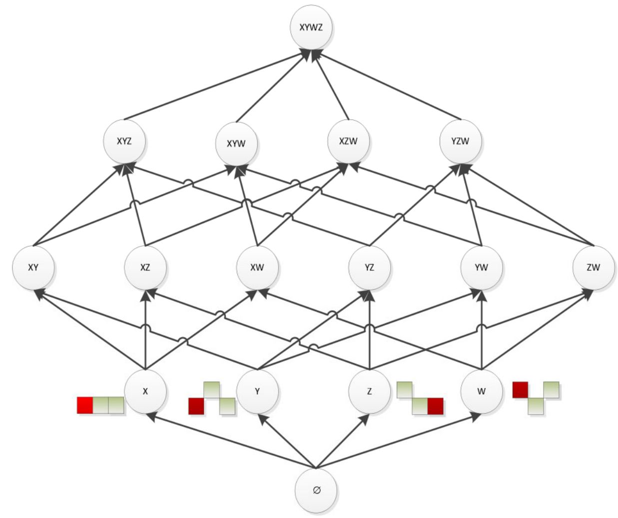


Figure 4. Power set for one subspace (the first horizontal line of fig. 3) of the #44 F_4 collection ordered with respect to inclusion

phase of the WD verification scheme follows: for each writer, *#nref* reference samples of genuine along with an equal number of simulated-forgery signature samples are randomly chosen in order to train the classifier. The "S2" image segmentation scheme combines the features calculated on the whole signature image as well as the relevant 2x2 equimass segmentation grid [20]. These features supply the classifier training section without assuming any additional processing. The classifier used is a hard-margin two class support vector machine (SVM) classifier using radial basis kernel. Selection of the training samples for the genuine class was accomplished using randomly chosen samples according to the hold-out validation method. The remaining genuine and simulated forgery signatures feature vectors, drawn using the same F_4 collection, feed the SVM classifier directly for testing. The SVM output apart from the binary class decision provides a score value which is equal to the distance of the tested sample from the SVM separating hyperplane. The operating parameters of the SVM have been determined through exhaustive search. It is noted that there is a wide area of rbf sigma values that the system has the reported results.

Evaluation of the verification efficiency of the system is accomplished with the use of a global threshold on the overall SVM output score distribution. This is achieved by providing the system's False Acceptance Rate (FAR: samples not belonging to genuine writers, yet assigned to them) and the False Rejection Rate (FRR: samples belonging to genuine writers, yet not classified) functions. With these two rates, the receiver operator characteristics (ROC) are drawn by means of their FAR/FRR plot. Then, classification performance is measured with the utilization of the system Equal Error Rate (EER: the point which FAR equals FRR).

IV. RESULTS

According to the discussion presented above, FAR, FRR and the relevant EER rates, are evaluated for (a) CORPUS 1 and and (b) CORPUS 2 with five and twelve reference

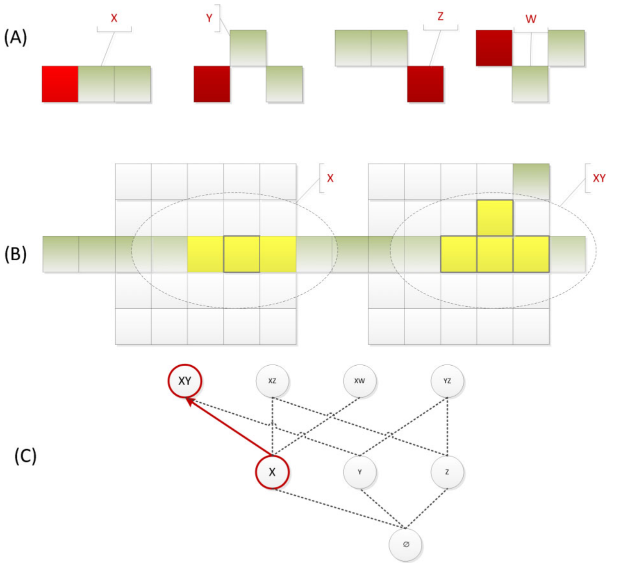


Figure 5. (A) One set of the #44 F_4 collection as depicted in fig. 3b. (B) left and right grids labeled as One (O) and Two (T) respectively imposed on a signature trace (mark with green shadowed pixels) and corresponding events activated. For illustration purposes the topological grids have a distance of 7 instead of 1 that is followed at the actual feature extraction method. (C) Ordered event detection is designated between the red circles and feature component update along red line.

samples for both genuine and forger class. The corresponding results are presented in Table I by means of the mean FAR, FRR and EER values. The letters G and F in Table I designate the genuine and skilled forgery samples respectively. In addition, the ROC curves are presented for both databases in fig. 6 along with their corresponding EER defined as the cross section of the ROC curves and the diagonal.

Our results are compared to recently published relevant figures. The reported results for CORPUS 1 are compared with the results relevant to those reported in [12] for feature level simulated forgery verification tests using 'S2' scheme using (a) nref=5 and (b) the mean value of nref=10 and nref=15 tests for comparison with our test using nref=12. The comparison results are presented in Table II. Concerning CORPUS 2, we present in Table III, the results of recently reported research work using nref=5 and nref=12, along with the results of the current approach.

V. Conclusions

In this work a handwritten model based on the powerset of an ordered event topology with respect to inclusion is considered as a tool for offline signature verification. A number of verification experiments based on an SVM classifier have been carried out in two signature databases namely the GPDS and a proprietary one. Primary verification results indicate that the proposed feature extraction method has an appealing aspect; As a comment on the efficiency of the method one can state that in the case of the Corpus 1 a substantial improvement is observed while in the case of Corpus 2 the results are comparable with those of the

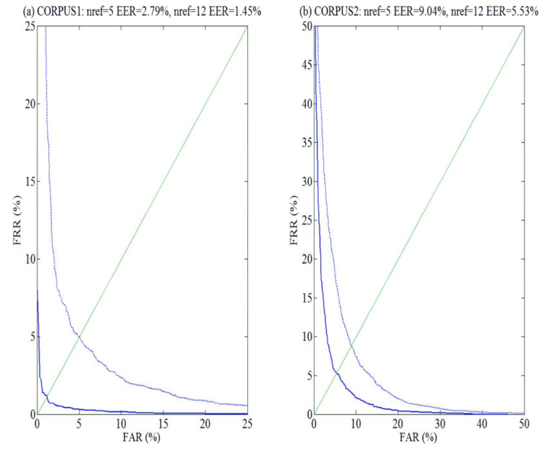


Figure 6. ROC curves with the corresponding EER for corpuses 1, 2.

literature. Since the approach described in this case study is preliminary it is anticipated that further exhaustive research will unveil important conclusions with respect to the modeling of handwriting. However a number of various other models and experimental setups including i.e. the dissimilarity framework [10] need to be examined in order to verify the effectiveness of the proposed approach.

TABLE I. Verification Efficiency (%)

Experimental Set	FAR	FRR	EER
CORPUS 1, #nref=5 (GF)	2.18	3.29	2.79
CORPUS 2, #nref=5 (GF)	13.03	5.23	9.04
CORPUS 1, #nref=12 (GF)	1.13	1.60	1.45
CORPUS 2, #nref=12 (GF)	7.73	3.45	5.53

TABLE II. Comparing EER with Approach [20]

Experimental Set	EER (%)
[20] #nref=5 (GF)	9.16
Proposed #nref=5 (GF)	2.79
[20] #nref=12 (GF)	4.65
Proposed #nref=12 (GF)	1.45

TABLE III. Comparing EER with Various Approaches (%)

Method	EER		EER
[20] #nref=5 (GF)	12.32	***Proposed #nref=5***	9.04
[12] GPDS-100 nref=5 (GF)	12.02		
[19] #nref=13 (only G)	4.21		
[20] for nref=12 (GF)	6.2	***Proposed #nref=12***	5.53
[12] # ref = {10G, 15F}	8.26		
[13] #refn=12 (GF)	13.76		
[24] #nref=12 (GF)	15.11		
[25] # nref=12 (only G)	15.4		

References

[1] R. Plamondon and S. N. Srihari, "On-line and off-line handwriting recognition: A comprehensive survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 63-84, 2000.

[2] F. Leclerc and R. Plamondon, "Automatic Signature verification: the state of the art-1989-1993", International Journal of Pattern Recognition and Artificial Intelligence, vol. 8, pp. 643-660, 1994.

[3] D. Impedovo, and G. Pirlo, "Automatic signature verification: The state of the art, "IEEE Transactions on Systems Man and Cybernetics" vol. 38, pp. 609-635, 2008.

[4] L. Batista, D. Rivard, R. Sabourin, E. Granger, and P. Maupin, "State of the art in off-line signature verification". In Verma, B., Blumenstein, M. (eds.) Pattern Recognition Technologies and Applications: Recent Advances, pp. 39–62, 2008.

[5] M. C. Fairhurst, "Signature verification revisited: promoting practical exploitation of biometric technology", Electron. Commun. Eng. J., vol. 9, pp. 273–280, 1997.

[6] R. C. Gonzalez and R. E. Woods, "Digital Image processing", Addison Wesley, Reading, 1992.

[7] S. Theodoridis and K. Koutroumbas, "Pattern Recognition", Academic Press, 2009.

[8] R. O. Duda and P. E. Hart, Pattern classification. New York: John Wiley and Sons, 2001.

[9] D. Bertolini, L. S. Oliveira, E. Justino, and R. Sabourin, "Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers". Pattern Recognition, vol. 43, pp. 387–396, 2010.

[10] D. Rivard, E. Granger and R. Sabourin, "Multi feature extraction and selection in writer independent off-line signature verification", International Journal on Document Analysis and Recognition, vol. 16, pp. 83-103, 2013.

[11] V. K. Madasu, and B. Lovell. "An Automatic Off-line Signature Verification and Forgery Detection System", in Verma, B., Blumenstein, M. (eds.) Pattern Recognition Technologies and Applications: Recent Advances, pp. 63–89, 2008.

[12] J. F. Vargas, M. A. Ferrer, C. M. Travieso and J. B. Alonso, "Off-line signature verification based on grey level information using texture features", Pattern Recognition, vol. 44, pp. 375-385, 2011.

[13] R. Kumar, J. D. Sharma, and B. Chanda, "Writer independent off-line signature verification using surroundedness feature", Pattern Recognition Letters, vol. 33, pp. 301-308, 2012.

[14] D. Impedovo, G. Pirlo, L. Sarcinella, E. Stasolla, and C. A. Trullo, "Analysis of Stability in Static Signatures using Cosine Similarity", in: Proc of International Conference on Frontiers in Handwriting Recognition, pp. 231-235, 2012.

[15] B. H. Shekar, and R. K. Bharathi, "LOG-Grid based off-line signature verification", in Fourth International Conference on signal and image processing, 2012. S. Mohan, S., Kumar, S.S. (eds), LNEE 222, pp. 321-330, 2013.

[16] J. P. Swanepoel, and J. Coester. "Off-line signature verification using flexible grid features and classifiers fusion", in International Conference on Frontiers in Handwriting Recognition, pp. 297-302. , 2012.

[17] M. K. Kalera, S. Shrihari and A. Xu, "Offine line signature verification using distance statistics", International Journal of Pattern Recognition and Artificial Intelligence, vol. 18, pp. 1339-1360, 2005.

[18] A. Gilperez, F. Alonso-Fernandez, S. Pecharroman, J. Fierrez-Aguilar, and J. Ortega-Garcia, "Off-line signature verification using contour features", in International Conference on Frontiers in Handwriting Recognition ICFHR 2008.

[19] M. Parodi, J. C. Gomez, and A. Belaid, "A circular grid-based rotation invariant feature extraction approach for off-line signature verification", in 11th International Conference on Document Analysis and Recognition, pp. 1289-1293. 2011.

[20] K. Tselios, E. N. Zois, E. Siores, A. Nassiopoulos, and G. Economou, "Grid-based feature distributions for off-line signature verification". IET Biometrics, vol. 1, pp. 72-81, 2012.

[21] T. M. Cover, and A. Y. Thomas, "Elements of information theory", 2nd ed. John Wiley and Sons (2006).

[22] P. R. Halmos, "Naive set theory". The University Series in Undergraduate Mathematics. van Nostrand Company, Princeton, 1960.

[23] Hazewinkel, M.: Encyclopedia of Mathematics, Springer, 2001.

[24] V. Nguyen, Y. Kawazoe, T. Wakabayashi, U. Pal, and M. Blumenstein, "Performance Analysis of the Gradient Feature and the Modified Direction Feature for Off-line Signature Verification", in Proc. 2010 Int Conf on Frontiers in Handwriting Recognition, pp. 303-307, 2010.

[25] M. B. Yilmaz, B. Yanikoglu, C. Tirkaz, and A. Kholmatov, "Offline signature verification using classifier combination of HOG and LBP features", in Proc. Int Joint Conf on Biometrics, pp. 1-7, 2011.

Towards Automated Hyperspectral Document Image Analysis

Zohaib Khan, Faisal Shafait and Ajmal Mian
School of Computer Science and Software Engineering
The University of Western Australia, 35 Stirling Highway, CRAWLEY, 6009
Email: zohaib@csse.uwa.edu.au, faisal.shafait@uwa.edu.au, ajmal.mian@uwa.edu.au

***Abstract*—Hyperspectral imaging and analysis refers to the capture and understanding of image content in multiple spectral channels. Satellite and airborne hyperspectral imaging has been the focus of research in remote sensing applications since nearly the past three decades. Recent use of ground-based hyperspectral imaging has found immense interest in areas such as medical imaging, art and archaeology, and computer vision. In this paper, we make an attempt to draw closer the forensic community and image analysis community towards automated forensic document examination. We believe that it has a huge potential to solve various challenging document image analysis problems, especially in the forensic document examination domain. We present the use of hyperspectral imaging for ink mismatch detection in handwritten notes as a sample application. Overall, this paper provides an overview of the applications of hyperspectral imaging with focus on solving pattern recognition problems. We hope that this work will pave the way for exploring its true potential in the document analysis research field.**

Keywords*—Multispectral imaging, Hyperspectral imaging, Hyperspectral document analysis, forensic document examination, ink mismatch detection

I. Introduction

Human eye exhibits a trichromatic vision. This is due to the presence of three types of photo-receptors called *Cones* that are sensitive to different wavelength ranges in the visible range of the electromagnetic spectrum [1]. Conventional imaging sensors and displays (like cameras, scanners and monitors) are developed to match the response of the trichromatic human vision so that they deliver the same perception of the image as in a real scene. This is why an RGB image constitutes three spectral measurements per pixel. Most of the computer vision applications do not make use of the spectral information and directly employ grayscale images for image understanding. There is evidence that machine vision tasks can take the advantage of image acquisition in a wider range of electromagnetic spectrum capturing more information in a scene compared to only RGB data. Hyperspectral imaging captures spectral reflectance information for each pixel in a wide spectral range. It also provides selectivity in the choice of frequency bands. Satellite based hyperspectral imaging sensors have long been used for astronomical and remote sensing applications. Due to the high cost and complexity of these hyperspectral imaging sensors, various techniques have been proposed in the literature to utilize conventional imaging systems combined with a few off-the-shelf optical devices for hyperspectral imaging.

Strictly speaking, an RGB image is a three channel multispectral image. An image acquired at more than three specific wavelengths in a band is referred to as a *Multispectral Image*. Generally, multispectral imaging sensors acquire more than three spectral bands. An image with a higher spectral resolution or more number of bands is regarded as a *Hyperspectral Image*. There is no clear demarcation with regards to the number of spectral bands/resolution between multispectral and hyperspectral images. However, hyperspectral sensors may acquire a few dozen to several hundred spectral measurements per scene point. For example, the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) of NASA has 224 bands in 400-2500nm range [2].

During the past several years hyperspectral imaging has found its utility in various ground-based applications. The use of hyperspectral imaging in archeological artifacts restoration has shown promising results. It is now possible to read the old illegible historical manuscripts by restoration using hyperspectral imaging [3]. This was a fairly difficult task for a naked eye due to its limited capability, restricted to the visible spectral range. Similarly, hyperspectral imaging has also been applied to the task of material discrimination. This is because of the physical property of a material to reflect a specific range of wavelengths giving it a spectral signature which can be used for material identification [4]. The greatest advantage of hyperspectral imaging in such applications is that it is non-invasive and thus does not affect the material under analysis compared to other invasive techniques which inherently affect the material under observation.

Despite the success of hyperspectral imaging in solving various challenging computer vision problems in recent years, its use in the document image analysis research has remained largely unexplored. In this paper, we intend to draw the attention of the document analysis and forensics community towards this promising technology. We believe that there is a huge potential in hyperspectral imaging to solve various challenging document image analysis problems, especially in the forensic document examination domain. First, we present in Section II a brief survey on the applications of hyperspectral imaging in the field of pattern recognition. Then, some of our recent work on forensic document examination using hyperspectral imaging is discussed in Section III. The paper is concluded with some hints about directions for future research in Section IV.

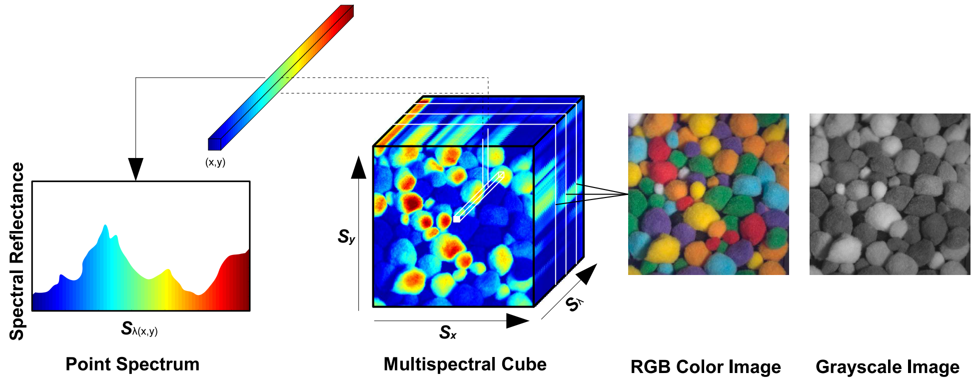


Fig. 1. A hyperspectral image is represented as a 3D cube (shown in pseudo-colors in center). Each slice of the cube along the spectral dimension S_λ is regarded as a channel or a band. Point spectrum on the spectral cube at the (x, y) spatial location (left). An RGB image and a grayscale image rendered from the hyperspectral cube (right).

II. Hyperspectral Imaging and Applications

A hyperspectral image has three dimensions: two spatial (S_x and S_y) and one spectral (S_λ) (see Figure 1). The hyperspectral data can be represented in the form of a *Spectral Cube*. Similarly, a hyperspectral video has four dimensions – two spatial dimensions (S_x and S_y), a spectral dimension (S_λ) and a temporal dimension (t). The hyperspectral video can be thought of as a series of *Spectral Cubes* along temporal dimension. Hyperspectral imaging has been applied in various areas, some of which are listed in Table I. In the following, we provide a brief survey of the applications of hyperspectral imaging in pattern recognition. The scope of our survey is limited to the multispectral and hyperspectral imaging systems used in ground-based computer vision applications. Therefore, high cost and complex sensors for remote sensing, astronomy, and other geo-spatial applications are excluded from the discussion.

TABLE I. Applications of hyperspectral imaging in different areas.

Areas	Applications
Art and Archeology	Analysis of works of art, historical artifact restoration
Medical Imaging	MRI imaging, microscopy, biotechnology
Military	Surveillance, access control
Pattern Recognition	Material identification, biometrics
Remote Sensing	Crop monitoring, mineralogy, water observation

A. Biometrics Applications

The bulk of biometric recognition research revolves around monochromatic imaging. Recently, different biometric modalities have taken advantage of hyperspectral imaging for reliable and improved recognition. The images can cover visible, infrared, or a combination of both ranges of the electromagnetic spectrum (see Figure 2). We briefly discuss the recent work in palmprint, face, fingerprint, and iris recognition using hyperspectral imaging.

Palmprints have emerged as a popular choice for human access control and identification. Interestingly, palmprints have even more to offer when imaged under different spectral ranges. The line pattern is captured in the visible range while the vein pattern becomes apparent in the near infrared range. Both line and vein information can be captured using a multispectral imaging system such as those developed by Han et al. [5] or Hao et al. [6]. The underlying principle of a multispectral palmprint imaging device is to use a monochromatic camera with illumination sources of different colors. Images of a palm are sequentially captured under each illumination within a fraction of a second.

Multispectral palmprint recognition system of Han et al. [5] captured images under four different illuminations (red, green, blue and infrared). The first two bands (blue and green) generally showed only the line structure, the red band showed both line and vein structures, whereas the infrared band showed only the vein structure. These images can be fused and features extracted for subsequent matching and recognition. The contact-free imaging system of Hao et al. [6] acquires multispectral images of a palm under six different illuminations. The contact-free nature of the system offers more user acceptability while maintaining a reasonable accuracy. Experiments show that pixel level fusion of multispectral palmprints has better recognition performance compared to monochromatic images. The accuracy achievable by multispectral palmprints is much higher compared to traditional monochromatic systems.

Fingerprints are established as one of the most reliable biometrics and are in common use around the world. Fingerprints can yield even more robust features when captured under a multispectral sensor. Rowe et al. [7] developed a multispectral imaging sensor for fingerprint imaging. The system comprised of illumination source of multiple wavelengths (400, 445, 500, 574, 610 and 660nm) and a monochrome CCD of 640x480 resolution. They showed that MSI sensors are less affected by moisture content of skin which is of critical significance compared to the traditional sensors. Recognition based on multispectral fingerprints outperformed standard fingerprint imaging.

Face recognition has an immense value in human identification and surveillance. The spectral response of human skin is a distinct feature which is largely invariant to the pose and expression [8] variation. Moreover, multispectral images of faces are less susceptible to variations in illumination sources and their directions [9]. Multispectral face recognition systems generally use a monochromatic camera coupled with a *Liquid Crystal Tunable Filter* (LCTF) in the visible and/or near-infrared range. A multispectral image is captured by electronically tuning the filter to the desired wavelengths and acquiring images in a sequence.

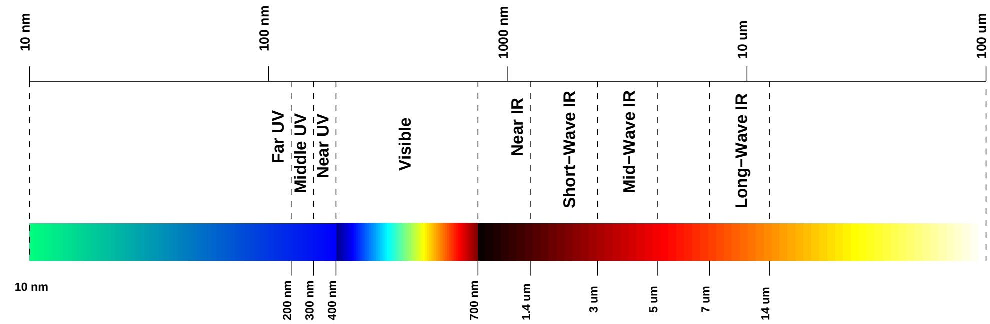


Fig. 2. The electromagnetic spectrum.

Iris is another unique biometric used for person authentication. Boyce et al. [10] explored multispectral iris imaging in the visible electromagnetic spectrum and compared it to the near-infrared in a conventional iris imaging systems. The use of multispectral information for iris enhancement and segmentation resulted in improved recognition performance.

B. *Material Identification*

Naturally existing materials show a characteristic spectral response to incident light. This property of a material can distinguish it from other materials. The use of multispectral techniques for imaging the works of arts like paintings allows segmentation and classification of painted parts. This is based on the pigment physical properties and their chemical composition [3]. Pelagotti et al. [11] used multispectral imaging for analysis of paintings. They collected multispectral images of a painting in UV, Visible and Near IR band. It was possible to differentiate among different color pigments which appear similar to the naked eye based on spectral reflectance information.

Gregoris et al. [12] exploited the characteristic reflectance of ice in the infrared band to detect ice on various surfaces which is difficult to inspect manually. The developed prototype called *MD Robotics' Spectral Camera system* could determine the type, level and location of the ice contamination on a surface. The prototype system was able to estimate thickness of ice (<0.5mm) in relation to the measured spectral contrast. Such system may be of good utility for aircraft/space shuttle ice contamination inspection and road condition monitoring in snow conditions.

Multispectral imaging has critical importance in magnetic resonance imaging. Multispectral magnetic resonance imagery of brain is in wide use in medical science. Various tissue types of the brain are distinguishable by virtue of multispectral imaging which aids in medical diagnosis [13].

Clemmensen et al. [14] used multispectral imaging to estimate the moisture content of sand used in concrete. It is a very useful technique for non-destructive in-vivo examination of freshly laid concrete. A total of nine spectral bands was acquired in both visual and near infrared range. Zawada et al. [15] proposed a novel underwater multispectral imaging system named *LUMIS* (Low light level Underwater Multispectral Imaging System) and demonstrated its use in study of phytoplankton and bleaching experiments.

Spectrometry techniques are also widely used to identify the fat content in pork meat, because it has proved significantly cheaper and more efficient than traditional analytical chemistry methods [16]. For that purpose, near-infrared spectrometers are used that measure the spectrum of light transmitted through a sample of minced pork meat.

Last but not least, multispectral imaging has also important applications in defense and security. For instance, Alouini [17] showed that multispectral polarimetric imaging significantly enhances the performance of target detection and discrimination.

III. Forensic Document Examination using Hyperspectral Imaging

Hyperspectral imaging (HSI) has recently emerged as an efficient non-destructive tool for detection, enhancement [18], comparison and identification of forensic traces [19]. Such systems have a huge potential for aiding forensic document examiners in various tasks. Brauns et al. [20] developed a hyperspectral imaging system to detect forgery in potentially fraudulent documents in a non-destructive manner. A more sophisticated hyperspectral imaging system was developed at the National Archives of Netherlands for the analysis of historical documents in archives and libraries [21]. The system provided high spatial and spectral resolution from near-UV through visible to near IR range. The only limitation of the system was its extremely slow acquisition time (about 15 minutes) [22]. Other commercial hyperspectral imaging systems from Foster & Freeman [23] and ChemImage [24] also allow manual comparison of writing ink samples. Hammond [25] used visual comparison in Lab color mode for differentiating different black inks. Such manual analysis of inks cannot establish the presence of different inks with certainty, because of inherent human error. Here we will demonstrate a promising application of hyper-spectral imaging for automated writing inks mismatch detection that we have recently proposed [26]. The work is based on the assumption that same inks exhibit similar spectral responses whereas different inks show dissimilarity in their spectra. The phenomenon is illustrated in Figure 3. We assume that the spectral responses of the inks are independent of the writing styles of different subjects.

Using our hyperspectral imaging setup (see [26] for details), a database comprising of 70 hyperspectral images of a hand-written note in 10 different inks by 7 subjects was collected[1]. All subjects were instructed to write the same sentence, once in each ink on a white paper. The pens included

[1]UWA Writing Ink Hyperspectral Image Database
http://www.csse.uwa.edu.au/%7Eajmal/databases.html

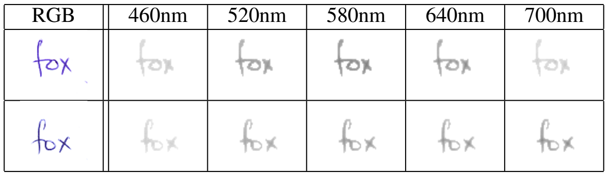


Fig. 3. The above images highlight the discrimination of inks offered by hyperspectral images. We show a selected number of bands at specific wavelengths for two different blue inks (for the word 'fox'). Notice that only the pixels belonging to the writing pixels are shown and the pixels of the background are masked out. A closer look allows one to appreciate that hyperspectral imaging captures subtle differences in the inks, which are enhanced, especially at higher wavelengths.

5 varieties of blue ink and 5 varieties of blank ink pens. It was ensured that the pens came from different manufacturers while the inks still appeared visually similar. Then, we produced mixed writing ink images from single ink notes by joining equally sized image portions from two inks written by the same subject. This made roughly the same proportion of the two inks under question.

The mixed-ink images were pre-processed (binarization [27] followed by spectral response normalization) and then fed to the k-means clustering algorithm with a fixed value of $k = 2$. Finally, based on the output of clustering, segmentation accuracy was computed as

$$\text{Accuracy} = \frac{\text{True Positives}}{\text{True Positives + False Positives + False Negatives}}$$

The segmentation accuracy is averaged over seven samples for each ink combination C_{ij}. It is important to note that according to this evaluation metric, the accuracy of a random guess (in a two class problem) will be 1/3. This is different from common classification accuracy metrics where the accuracy of a random guess is 1/2. This is because our chosen metric additionally penalizes false negatives which are useful to quantify in a segmentation problem.

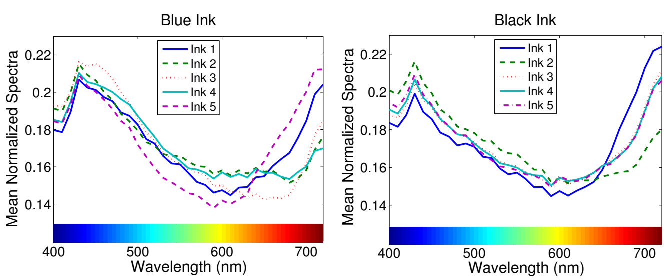


Fig. 4. Spectra of the blue and black inks under analysis. Note that at some ranges the ink spectra are more distinguished than others.

Figure 4 shows the average normalized spectra of all blue and black inks, respectively. It was achieved by computing the average of the spectral responses of each ink over all samples in the database. It can be observed that the spectra of the inks are distinguished at different ranges in the visible spectrum. A close analysis of variability of the ink spectra in these ranges reveals that most of the differences are present in the high-visible range, followed by mid-visible and low-visible ranges.

We now inspect how hyperspectral information can be beneficial in discrimination of inks. We compare the segmentation accuracy of HSI with RGB in Figure 5. As expected,

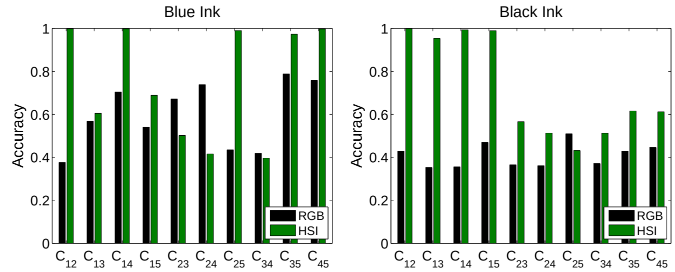


Fig. 5. Comparison of RGB and HSI image based segmentation accuracy.

HSI significantly improves over RGB in most of the ink combinations. This results in most accurate clustering of ink combinations C_{12}, C_{14}, C_{12}, C_{25}, C_{35} and C_{45}. In case of black inks, ink 1 is highly distinguished from all other inks resulting in the most accurate clustering for all combinations C_{1j}. However, it can be seen that for a few combinations, HSI does not show a remarkable improvement. Instead, in some cases, it is less accurate compared to RGB. These results encouraged us to further look at HSI in detail in order to take advantage of the most informative bands. The results of different feature (band) selection methods for this problem are detailed in [26]. Overall, the results showed that use of a few selected bands further improved discrimination between most of the ink combinations.

We now present some qualitative results on segmentation of blue and black ink combinations. The original images of a combination of two blue inks (C_{34}) and black inks (C_{45}) are shown are in Figure 6. RGB images are shown here for better visual appearance. The ground truth images are labeled in pseudo-colors, where green pixels represent the first ink and red pixels represent the second ink.

The clustering based on RGB images fails to group similar ink pixels into the same clusters. A closer look reveals that all of the ink pixels are falsely grouped into one cluster, whereas most of the boundary pixels are grouped into the other cluster. This implies that typical RGB imaging is not sufficient to discriminate inks that appear visually similar to each other. On the other hand, segmentation based on HSI is much more effective compared to RGB. It can be seen that the majority of the ink pixels are correctly grouped in HSI in accordance with the ground truth segmentation. Note that the k-means clustering algorithm used in this work is rather basic. The use of more advance clustering algorithms has the potential of further improving the accuracy of ink segmentation.

IV. Conclusion and Outlook

This paper presented an overview about different applications of hyperspectral imaging in pattern recognition. We also demonstrated a sample application of HSI in document image analysis, where it was possible to discriminate between two visually similar inks using hyperspectral images of the documents. This is the first reported work on using automatic document image analysis methods in combination with hyperspectral imaging to address forensically relevant issues in questioned document examination. In future, it will be interesting to see whether spectral imaging can aid in writer identification. Since it is possible to identify hand writings

Fig. 6. Example test images. For a visual comparison of RGB and HSI mismatch detection accuracy, we purposefully selected two hard cases.

by the texture [28] or ink-deposition traces [29], a promising research direction would be to investigate whether these feeble variations in ink strokes are reflected in the spectral response of the inks. In addition, ink or document aging is a phenomenon that can be observed in a more effective manner using spectral imaging. During the aging process, the chemical properties of ink and paper change due to various environmental factors. Spectral imaging can potentially capture subtle differences in inks or paper due to aging. These are just a few application examples where HSI can potentially provide solutions to some major practical problems in document analysis. We hope that this work will open up many exciting possibilities for tackling forensic document examination problems with a new perspective.

Acknowledgment

This research was supported by ARC Grant DP110102399.

References

[1] P. R. Martin, "Retinal color vision in primates," in *Encyclopedia of Neuroscience*. Springer, 2009, pp. 3497–3501.

[2] P. Shippert, "Introduction to hyperspectral image analysis," *Online Journal of Space Communication*, vol. 3, 2003.

[3] S. Baronti, A. Casini, F. Lotti, and S. Porcinai, "Principal component analysis of visible and near-infrared multispectral images of works of art," *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 103–114, 1997.

[4] B. Thai and G. Healey, "Invariant subpixel material detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 3, pp. 599–608, 2002.

[5] D. Han, Z. Guo, and D. Zhang, "Multispectral palmprint recognition using wavelet-based image fusion," in *Proc. International Conference on Signal Processing*. IEEE, 2008, pp. 2074–2077.

[6] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multispectral palm image fusion for accurate contact-free palmprint recognition," in *Proc. International Conference on Image Processing*. IEEE, 2008, pp. 281–284.

[7] R. K. Rowe, K. Nixon, and S. Corcoran, "Multispectral fingerprint biometrics," in *Proc. IEEE SMC Information Assurance Workshop*. IEEE, 2005, pp. 14–20.

[8] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1552–1560, 2003.

[9] H. Chang, A. Koschan, M. Abidi, S. G. Kong, and C.-H. Won, "Multispectral visible and infrared imaging for face recognition," in *Proc. Computer Vision and Pattern Recognition Workshops*. IEEE, 2008, pp. 1–6.

[10] C. Boyce, A. Ross, M. Monaco, L. Hornak, and X. Li, "Multispectral iris analysis: A preliminary study," in *Proc. Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2006.

[11] A. Pelagotti, A. Del Mastio, A. De Rosa, and A. Piva, "Multispectral imaging of paintings," *IEEE Signal Processing Magazine*, vol. 25, no. 4, pp. 27–36, 2008.

[12] D. Gregoris, S. Yu, and F. Teti, "Multispectral imaging of ice," in *Proc. Canadian Conference on Electrical and Computer Engineering*, vol. 4. IEEE, 2004, pp. 2051–2056.

[13] T. Taxt and A. Lundervold, "Multispectral analysis of the brain using magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 13, no. 3, pp. 470–481, 1994.

[14] L. H. Clemmensen, M. E. Hansen, and B. K. Ersbøll, "A comparison of dimension reduction methods with application to multi-spectral images of sand used in concrete," *Machine Vision and Applications*, vol. 21, no. 6, pp. 959–968, 2010.

[15] D. G. Zawada, "Image processing of underwater multispectral imagery," *IEEE Journal of Oceanic Engineering*, vol. 28, no. 4, pp. 583–594, 2003.

[16] H. H. Thodberg, "A review of bayesian neural networks with an application to near infrared spectroscopy," *IEEE Transactions on Neural Networks*, vol. 7, no. 1, pp. 56–72, 1996.

[17] M. Alouini, "Target detection and discrimination through active multi-spectral polarimetric imaging," in *Computational Optical Sensing and Imaging*. Optical Society of America, 2005, pp. 1–8.

[18] S. Joo Kim, F. Deng, and M. S. Brown, "Visual enhancement of old documents with hyperspectral imaging," *Pattern Recognition*, vol. 44, no. 7, pp. 1461–1469, 2011.

[19] G. Edelman, E. Gaston, T. van Leeuwen, P. Cullen, and M. Aalders, "Hyperspectral imaging for non-contact analysis of forensic traces," *Forensic Science International*, vol. 223, pp. 28–39, 2012.

[20] E. B. Brauns and R. B. Dyer, "Fourier transform hyperspectral visible imaging and the nondestructive analysis of potentially fraudulent documents," *Applied spectroscopy*, vol. 60, no. 8, pp. 833–840, 2006.

[21] R. Padoan, T. A. Steemers, M. Klein, B. Aalderink, and G. de Bruin, "Quantitative hyperspectral imaging of historical documents: technique and applications," *ART Proceedings*, 2008.

[22] M. E. Klein, B. J. Aalderink, R. Padoan, G. De Bruin, and T. A. Steemers, "Quantitative hyperspectral reflectance imaging," *Sensors*, vol. 8, no. 9, pp. 5576–5618, 2008.

[23] foster + freeman, http://www.fosterfreeman.com/index.php.

[24] ChemImage, http://www.chemimage.com/.

[25] D. L. Hammond, "Validation of lab color mode as a nondestructive method to differentiate black ballpoint pen inks*," *Journal of forensic sciences*, vol. 52, no. 4, pp. 967–973, 2007.

[26] Z. Khan, F. Shafait, and A. Mian, "Hyperspectral imaging for ink mismatch detection," in *Proc. International Conference on Document Analysis and Recognition (ICDAR)*, 2013.

[27] F. Shafait, D. Keysers, and T. M. Breuel, "Efficient implementation of local adaptive thresholding techniques using integral images," *Document Recognition and Retrieval XV*, pp. 681 510–681 510–6, 2008.

[28] K. Franke, O. Bunnemeyer, and T. Sy, "Ink texture analysis for writer identification," in *Proc. IEEE Workshop on Frontiers in Handwriting Recognition*, 2002, pp. 268–273.

[29] K. Franke and S. Rose, "Ink-deposition model: The relation of writing and ink deposition processes," in *Proc. IEEE Workshop on Frontiers in Handwriting Recognition*, 2004, pp. 173–178.

Fusing Modalities in Forensic Identification with Score Discretization

Y.L. Wong, S. M. Shamsuddin, S. S. Yuhaniz
Soft Computing Research Group
Universiti Teknologi Malaysia
81310 Johor, Malaysia
yeeleng28@gmail.com,mariyam@utm.my,sophia@utm.my

Sargur N. Srihari
Department of Computer Science and Engineering
University at Buffalo,The State University of New York
Buffalo, NY 14260 USA
srihari@cedar.buffalo.edu

Abstract*—The fusion of different forensic modalities for arriving at a decision of whether the evidence can be attributed to a known individual is considered. Since close similarity and high dimensionality can adversely affect the process, a method of score fusion based on discretization is proposed. It is evaluated considering the signatures and fingerprints. Discretization is performed as a filter to find the unique and discriminatory features of each modality in an individual class before their use in matching. Since fingerprints and signatures are not compatible for direct integration, the idea is to convert the features into the same domain. The features are assigned an appropriate matched score, MS_{bp} which are based to their lowest distance. The final scores are then fed to the fusion, FS_{bp}. The top matches with FS_{bp} less than a predefined threshold value, η are expected to have the true identity. Two standard fusion approaches, namely *Mean* and *Min* fusion, are used to benchmark the efficiency of proposed method. The results of these experiments show that the proposed approach produces a significant improvement in the forensic identification rate of fingerprint and signature fusion and this findings support its usefulness.

***Keywords*—forensic; multimodal; discretization; matching scores; fusion; identification**

I. INTRODUCTION

The goal of forensic analysis is that of determining whether observed evidence can be attributed to an individual. The final decision of forensic analysis can take one of three values: identification/no-conclusion/exclusion. Biometric systems have a similar goal of going from input to conclusion but with different goals and terminology: biometric identification means determining the best match in a closed set of individuals and verification means whether the input and known have the same source. While biometric systems attempt to do the entire process automatically, forensic systems narrow-down the possibilities among a set of individuals with the final decision being made by a human examiner. Automatic tools for forensic analysis have been developed for several forensic modalities including signatures [1], fingerprints [2], handwriting [3], and footwear prints or marks [4]. In both forensic analysis and biometric analysis more than one modality of data can be used to improve accuracy [5], [6]. Examples of the need to combine forensic evidence in forensic analysis are: signature and fingerprints on the same questioned document, pollen found on the clothing of an assailant together with human DNA [7], multiple shoe-prints in a crime scene [8], etc. In this paper we explore how evidence of different modalities can be combined for the forensic decision. Biometric identification systems such as token based and password based identification systems, unimodal identification recognizes a user, by "who the person is", using a one-to many matching process (1:M) rather than by "what the person carries along". Conventional systems suffer from numerous drawbacks such as forgotten password, misplaced ID card, and forgery issues. To address these problems, unimodal based identification was developed and has seen extensive enhancements in reliability and accuracy of identification. However, several studies have shown that the poor quality of image samples or the methodology itself can lead to a significant decreasing in the performance of a unimodal based identification system [9], [10], [11]. The common issues include intra-class variability, spoof attack, non-universality, and noisy data. In order to overcome these difficulties in unimodal identification, *multimodal based identification systems* (MIS) have been developed. As the name suggests, in an MIS the identification process is based on evidence presented by multiple modality sources from an individual. Such systems are more robust to variations in the sample quality than unimodal systems due to the presence of multiple (and usually independent) pieces of evidence [12]. A key to successful multimodal based system development for forensic identification, is an effective methodology organization and fusion process, capable to integrate and handle important information such as distinctiveness characteristic of an individual. Individual's distinctive characteristics is unique to forensic. Therefore, in this paper, the *multi-matched scores based discretization* method is proposed for forensic identification of an individual from different modalities. Compared to previous methods, the proposed method is unique in the sense that the extracted features correspond to the individuality of a particular person which are discretized and represented into standard sizes. The method is robust and capable to overcome dimensionality issues without requiring image normalization. The low dimension and standardized features make the design of post-processing phase (classifier or decision) straightforward. Moreover, the clear physical meanings of the discretized features are meaningful and distinctive, and be used in more complex systems (e.g., expert systems for interpretation and inference).

II. Related Work

In identification systems, fusion takes into account a set of features that can reflect the individuality and characteristics of the person under consideration. However, it is difficult to extract and select features that are discriminatory, meaningful and important for identification. Different sets of features may have better performance when considering different groups of individuals and therefore, a technique is needed to represent for each sample set of features. In this paper, multi-matched scores fusion based discretization is proposed for forensic identification to represent the distinctiveness in multi-modalities of an individual.

A. Representation of individuality features

Extracting and representing relevant features which contains the natural characteristics of an individual is essential for a good performance of the identification algorithms. Existing multimodal based identification systems make the assumptions that each modality feature set from an individual is local, wide-ranging, and static. Thus, these extracted feature sets are commonly fed to individual matching or and classification algorithms directly.

As a result, the identification system becomes more complex, time consuming, and costly because a classifier is needed for each modality. Furthermore, concatenating features from different modalities after the feature extraction method leads to the need of comparing high dimensional, heterogeneous data which is a nontrivial issue. However, much work has been proposed to overcome the dimensional issues in extracted features such as implementation of normalization techniques after extraction. Careful observation and experimental analysis need to be performed in order to improve the performance of identification. Too much of normalization will diminish the originality characteristic of an individual from different modality images. Thus, another process is needed to produce a more discriminative, reliable, unique and informative feature representation to represent these inherently multiple continuous features into standardized discrete features (per individual). This leads to the *multi-matched score fusion discretization* approach introduced in this paper which is explored in the context of forensic identification of different modalities for distinguishing a true identity of a person.

B. The discretization algorithm

Discretization is a process whereby a continuous valued variable is represented by a collection of discrete values. It attracted a lot of interest from and work in several different domains [13], [14], [15]. The discretization method introduced here is based on discretization defined in [16].

Given a set of features, the discretization algorithm first computes the size of interval, i.e., it determines its upper and lower bounds. The range is then divided by the number of features which then gives each interval upper and lower approximation. The number of intervals generated is equal to the dimensionality of the feature vectors, maintaining the original number of extracted features from different extraction methods in this study. Subsequently, a single representation value for each interval, or cut, is computed by taking the midpoint of the lower approximation, $Approx_{lower}$ and upper approximation, $Approx_{upper}$ interval. Algorithm 1 shows the discretization steps discussed above.

Algorithm 1: Discretization Algorithm

Require: Dataset with f continuous features, D samples and C classes;
Require: Discretized features, D';
for each individual **do**
  Find the Max and the Min values of D samples
  $numb\_bin = numb\_extracted\_feature$
  Divide the range of Min to Max with $numb\_bin$

  Compute representation values, $RepValue$:

  for each bin **do**
    Find the $Approx_{lower}$ and $Approx_{upper}$
    Compute the midpoints of all $Approx_{lower}$ and $Approx_{upper}$
  end for

  Form a set of all discrete values, $Dis\_Features$:

  for 1 to $numb\_extracted\_feature$ **do**
    for each bin **do**
      if (feature in range of interval) **then**
        $Dis\_Feature = RepValue$
      end if
    end for
  end for
end for

C. Processing and extraction of Signature and Fingerprint

For signature, the input image is first binarized by adaptive thresholding, followed by morphology operations (i.e., remove and skel) to get the gray level of clean and universe of discourse signature image (UOD) as illustrated in Fig. 1. The UOD of signature is extracted using geometry based extraction approach [17], which is based on 3x3 window concept. The process is done on individual window instead of the whole image to give more information of the signature image icludes the positions of different line structures.

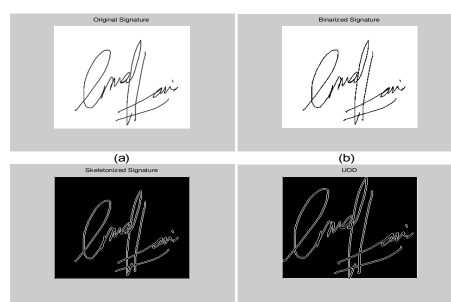


Fig. 1. Examples of preprocessed signature image (a)Original image (b)Binarized image (c)Skeletonized image (d)UOD.

For fingerprint, two types of manutia points namely termination and bifurcation points are extracted using Minutia based extraction approach. Fig. 2 shows the block diagram of minutia based extraction process. Fingerprint image are binarized, thinned and false minutia are removed to extract the region of interest (ROIs). Finally, the extracted ROI for fingerprint and UOD for the signature are fed to the discretization.

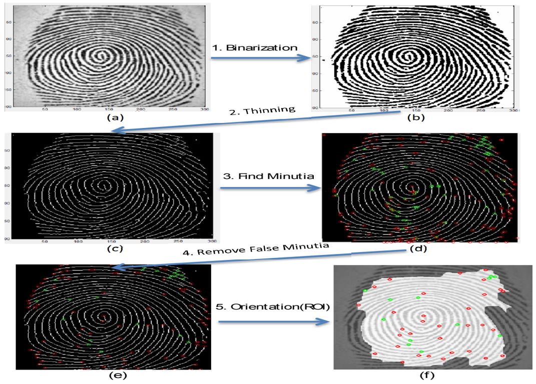


Fig. 2. Examples of preprocessed fingerprint image (a)Original image (b)Binarized image (c)Thinned image (d)Minutia Points (e)False Minutia removed (f)ROI.

Unimodal extraction and the discretization step are illustrated in Table I for signature data for individual 1, and Table II for the fingerprint data for the same individual. In each of these tables, the feature values are divided into predefined number of bins, which is based on the number of features for each modality image.

In the top portion of these tables, for each bin, the lower and upper values are recorded in columns two and three respectively, and bin, $RepValue$, the average of lower and upper values, is recorded in column four. Max and Min values are highlighted in bold face.In the bottom portion of the table, the discretized features for signature and fingerprint are displayed. These tables shows an example of how the actual feature sets from individual are discretized. As it can be seen from the Table I, the feature values, 35.259 occurs for every column of the nine features for the signature data of the same individual.This means that the first individual is uniquely recognized by this discriminatory value. A similar discussion holds for Table II, where the set of discriminatory values for fingerprint data for first individual, obtained from four different images is 104.

The selected features are the representation values (Discriminatory features, DF of an individual) that describe the unique characteristics of an individual which will be used for matching process. In matching module, the distance between the discretized values with the stored feature values are computed by Euclidean Distance equation as defined in (1).

$$ED_{bp} = \sum_{i=1}^{N} \left(Df_{bp,i} - Df_{bp,i}^{(r)} \right) \tag{1}$$

Where $Df_{bp,i}$ represents ith discretized feature of new modality image meanwhile $Df_{bp,i}^{(r)}$ defines the ith discretized feature of reference modality image in stored template and bp represents either behavioral or phisiological trait of the individual. The ith total number of features extracted from a single modality image is denoted by N. Let $X_{sign} = ED_{sign}(x)$, where $X_{sign} = (x_1, ...x_d)$ denotes a distance for discretized signature features and $Y_{finger} = ED_{finger}(y)$, where $Y_{finger} = (y_1, ...y_d)$ is a distance for the discretized fingerprint features. The lowest distance for signature can be denoted as $min[ED_{sign}(x)]$ and lowest distance for fingerprint can be defined as $min[ED_{finger}(y)]$. Then, we define the modality features with the lowest distance as match score-1,($MS_{bp} = 1$), the second modality features with the second lowest distance as $MS_{bp} = 2$ and so on. bp here defines either behavioral(i.e., signature) or phisiological(i.e., fingerprint) trait of the individual. Then, the match score, $MSbp$ is fed to the fusion approach.

D. Multi-modality fusion

After matching, the matched scores of signature and fingerprint are fed to the fusion method. Let X_{sign}=$MS_{sign}(1)$,$MS_{sign}(2)$,...$MS_{sign}(n)$ denotes the computed signature match scores and $Yfinger$=$MS_{finger}(1)$,$MS_{finger}(2)$,...$MS_{finger}(n)$ defines the computed match scores for fingerprint.In this work, the final fused score, FS_{bp} of the individual are computed using Equation (2), where k represents the number of different modalities of an individual. The MS for fingerprint and signature are combined and divided by k to generate a single score which is then compared to a predefined threshold to make the final decision.

$$FS_{bp} = \frac{MS_{sign} + MS_{finger}}{k} \tag{2}$$

Fusion approaches, namely Mean, $MeanFS_{bp}$ and Min, $MinFS_{bp}$ fusion as defined in (3) and (4) are chosen for comparisson to show the efficiency of the proposed method on multi-modalities identification.

$$MeanFS_{bp} = (xMS_{sign} + yMS_{finger})/2 \tag{3}$$

$$MinFS_{bp} = min(MS_{sign}, MS_{finger}) \tag{4}$$

Finally, the FS_{bp} is forward to next phase for identification. In identification process of one-to-many matching (1:M), FS_{bp} is compared with the predefined identification threshold, η in order to identify the individual from M individuals. In this work, the identity of a person is identified if,

$$FS_{bp} \leq \eta \tag{5}$$

III. Experimental Results

The performance of this work is performed using ROC curve which consists of Genuine Acceptance Rate (GAR) of a system mapped against the False Acceptance Rate (FAR). In this work, GAR is equal to 1-FRR. Fig. 1 shows the performance of Unimodal identification for signature and fingerprint. Discretization is applied in this experiment. No normalization and fusion methods are implemented. The performance of the identification for both discretized signature and fingerprint and non-discretized dataset is compared.

TABLE I

EXAMPLE OF DISCRETIZATION PROCESS FOR SIGNATURE FEATURES OF FIRST INDIVIDUAL

LOW and UPPER BIN for Individual: 1
MIN Value 10.8096 MAX Value 98.8273

Bin	Lower	Upper	RepValue
0	**10.8096**	20.5893	15.69945
1	20.5893	30.3691	25.4792
2	30.3691	40.1488	35.259
3	40.1488	49.9286	45.0387
4	49.9286	59.7083	54.8184
5	59.7083	69.4881	64.5982
6	69.4881	79.2678	74.3779
7	79.2678	89.0476	84.1577
8	89.0476	**98.8273**	93.9374

DISCRETIZED DATA

f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	Class	
15.69945	15.69945	**35.259**	**35.259**	15.69945	15.69945	15.69945	25.4792	25.4792	1s	**Discriminatory**
54.8184	64.5982	93.9374	**35.259**	15.69945	**35.259**	54.8184	45.0387	25.4792	1s	**Value is**
25.4792	**35.259**	**35.259**	25.4792	25.4792	54.8184	**35.259**	45.0387	45.0387	1s	**35.259**
64.5982	**35.259**	25.4792	25.4792	**35.259**	74.3779	45.0387	15.69945	15.69945	1s	**for 1st ind.**

TABLE II

EXAMPLE OF DISCRETIZATION PROCESS FOR FINGERPRINT FEATURES OF FIRST INDIVIDUAL

LOW and UPPER BIN for Individual : 1
MIN Value 55 MAX Value 195

Bin	Lower	Upper	RepValue
0	**55**	69	62
1	69	83	76
2	83	97	90
3	97	111	104
4	111	125	118
5	125	139	132
6	139	153	146
7	153	167	160
8	167	181	174
9	181	**195**	188

DISCRETIZED DATA

f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_10	Class	
104	90	**104**	132	**104**	118	**104**	**104**	62	146	1f	**Discriminatory**
90	132	**104**	132	160	90	146	146	160	188	1f	**Value is**
76	90	**104**	132	160	62	**104**	**104**	160	181	1f	**104**
90	**104**	118	132	146	132	76	62	160	160	1f	**for 1st ind.**

From ROC graph, clearly defines that the use of discretization on the unimodal dataset enhances the overall performance of identification significantly over the performance of identification without discretization. Due to efficiency of the discretization method on unimodal identification, thus, the same technique is applied to multimodal identification in order to improve the accuracy of identification on multiple modalities.

Fig. 2 and Fig. 3 below shows the performance of ROC graph for two different fusion methods namely Mean fusion rule and Min method with the implementation of Z-Score normalization and matched scores fusion based discretization approach on multiple modalities. From the ROC graph depicted in Fig. 2, it can be seen that the implementation of the proposed method based discretization on the multi-modalities fusion of signature and fingerprint shows a better performance than the standard signature and fingerprint identification system. At FAR of 0.1%, 1.0%, and 10.0%, the implementation of the proposed method which is based on discretization has a GAR of 96.9%, 98.9%, and 99.9% respectively, where the performance is better than the Z-score normalization and Mean fusion on signature and fingerprint modalities, 93.5%, 93.7%, and 96.4%. Fig. 3 shows the GAR performance on Min fusion based Z-score normalization and the proposed multi-matched score

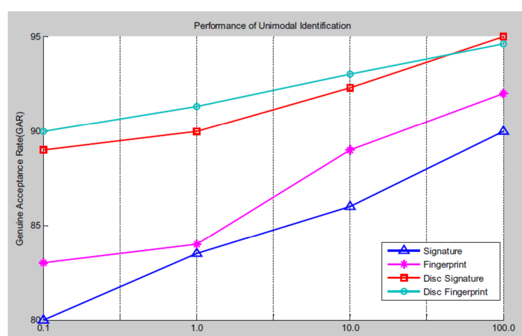


Fig. 3. Performance of uni-modality identification.

based discretization. Again, in Fig. 3, interestingly, the proposed method based on discretization on signature and fingerprint modalities yields the best performance over the range of FAR. At 0.1%, 1.0%, and 10.0% of FAR, the Min fusion method works the best with proposed method, 95.0%, 97.99%, and 99.40% respectively. Therefore, it can be summarized that the used of discretization and proposed fusion of fingerprint and signature modalities generally performs well over the use of normalization and conventional fusion approaches for personal identification.

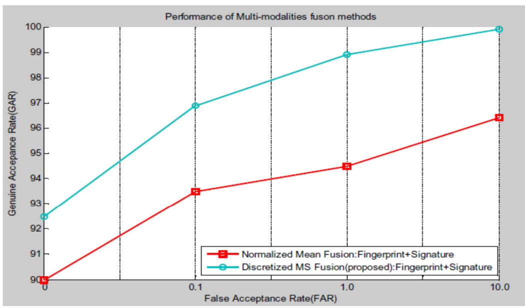


Fig. 4. Performance of Multi-modality fusion methods for signature and fingerprint.

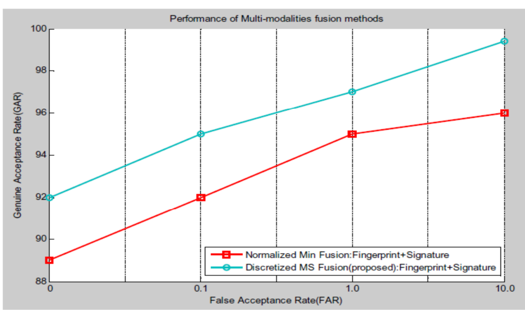


Fig. 5. Performance of Multi-modality fusion methods for signature and fingerprint.

IV. Conclusion

A key to successful multimodal based system development for forensic identification, is an effective methodology organization and fusion process, capable to integrate and handle important information such as distinctiveness characteristic of an individual. In this paper, the match scores discretization is proposed and implemented on different modality datasets of an individual. The experiments are done on signature and fingerprint datasets, which consist of 156 students (both female and male) where each student contributes 4 samples of signatures and fingerprint. Ten features describing the bifurcation and termination points of fingerprint, were extracted using Minutia based extraction approach whereas signature is extracted using Geometry based extraction approach. In matching process, each template-query pair feature sets is compared using Euclidean distance. Two fusion approaches namely Mean and Min fusion are performed to seek for the efficiency of the proposed method in Multimodal identification. The experimental results show that the proposed multi-matched scores discretization perform well on multiple set of individual traits, consequently improving the identification performance.

Acknowledgment

This work is supported by The Ministry of Higher Education (MOHE) under Research University Grant (GUP) and Mybrain15. Authors would especially like to thank Universiti Teknologi Malaysia, Skudai Johor Bahru MALAYSIA for the support and Soft Computing Research Group (SCRG) for their excellent cooperation and contributions to improve this paper.

References

[1] S. N. Srihari, "Computational methods for handwritten questioned document examination," *National Criminal Justice Research Report*, 2010.

[2] C. Su and S. Srihari, "Evaluation of rarity of fingerprints in forensics," *Advances in Neural Information Processing Systems*, vol. 23, pp. 1207–1215, 2010.

[3] S. N. Srihari and K. Singer, "Role of automation in the examination of handwritten items," in *Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on*. IEEE, 2012, pp. 619–624.

[4] Y. Tang, H. Kasiviswanathan, and S. N. Srihari, "An efficient clustering–based retrieval framework for real crime scene footwear marks," *International Journal of Granular Computing, Rough Sets and Intelligent Systems*, vol. 2, no. 4, pp. 327–360, 2012.

[5] J. Gonzalez-Rodriguez, J. Ortega-Garcia, and J.-L. Sanchez-Bote, "Forensic identification reporting using automatic biometric systems," in *Biometric Solutions*. Springer, 2002, pp. 169–185.

[6] J. Gonzalez-Rodriguez, J. Fierrez-Aguilar, D. Ramos-Castro, and J. Ortega-Garcia, "Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems," *Forensic science international*, vol. 155, no. 2-3, pp. 126–140, 2005.

[7] K. J. Craft, J. D. Owens, and M. V. Ashley, "Application of plant dna markers in forensic botany: Genetic comparison of¡ i¿ quercus¡/i¿ evidence leaves to crime scene trees using microsatellites," *Forensic science international*, vol. 165, no. 1, pp. 64–70, 2007.

[8] Y. Tang, S. N. Srihari, H. Kasiviswanathan, and J. J. Corso, "Footwear print retrieval system for real crime scene marks," in *Computational Forensics*. Springer, 2011, pp. 88–100.

[9] A. Jain and A. Ross, "Introduction to biometrics," *Handbook of Biometrics*, pp. 1–22, 2008.

[10] A. Ross, K. Nandakumar, and A. Jain, "Introduction to multibiometrics," *Handbook of Biometrics*, pp. 271–292, 2008.

[11] N. Solayappan and S. Latifi, "A survey of unimodal biometric methods," in *Proceedings of the 2006 International Conference on Security and Management*, 2006, pp. 57–63.

[12] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern recognition letters*, vol. 24, no. 13, pp. 2115–2125, 2003.

[13] H. Liu, F. Hussain, C. L. Tan, and M. Dash, "Discretization: An enabling technique," *Data mining and knowledge discovery*, vol. 6, no. 4, pp. 393–423, 2002.

[14] R. Ahmad, M. Darus, S. M. H. Shamsuddin, and A. A. Bakar, "Pendiskretan set kasar menggunakan taakulan boolean terhadap pencaman simbol matematik," *Journal Teknologi Maklumat & Multimedia*, pp. 15–26, 2004.

[15] B. O. Mohammed and S. M. Shamsuddin, "Feature discretization for individuality representation in twins handwritten identification," *Journal of Computer Science*, vol. 7, no. 7, pp. 1080–1087, 2011.

[16] A. Muda, S. Shamsuddin, and M. Darus, "Invariants discretization for individuality representation in handwritten authorship," *Computational Forensics*, pp. 218–228, 2008.

[17] K. Huang and H. Yan, "Off-line signature verification based on geometric feature extraction and neural network classification," *Pattern Recognition*, vol. 30, no. 1, pp. 9–17, 1997.

Joint Glossary of Forensic Document Examination and Pattern Recognition

Inés Baldatti
Banco Central de la República Argentina
Buenos Aires, Argentina
Email: inesbaldatti@gmail.com

Erika Griechisch
University of Szeged, Hungary
H-6720 Szeged, Árpád tér 2.
Email: grerika@inf.u-szeged.hu

***Abstract*—In this paper we introduce an open, scientific glossary which uses MediaWiki engine to the forensic examiner and pattern recognition scientific communities. Besides our aim to find editors from these communities who contribute to extend the glossary and make it as complete as possible, we would like to translate the terms from English to other languages, e.g. Portuguese, German, Chinese, Japanese, Arabian. The contribution can be started with translating the existing words at the glossary. The second part of our work when the glossary become be more completed, will consist into create the very understanding and useful glossary.**

I. Introduction

In the last few years there were more and more communication and joint research between the forensic examiners and pattern recognition scientists. These two communities can efficiently work together, if they understand the terms from both sciences. We have seen and heard on meetings, conferences, workshops, and during discussions it is a necessity to have a useful, extendable glossary, and dictionary which helps the common work.

Our aim is to create a glossary and a dictionary with the important terms of the forensic science for the forensic document examiners and pattern recognition experts in different languages. We are considering even the different expressions among countries who share the same language in order to obtain a better understanding into our fields no matter where we are from. Thinking and hoping this work will be a useful tool for both: the forensic and pattern recognition communities.

The long name of the glossary is *Glossary of Forensic Document Examination and Pattern Recognition* and the short name is GoFDER. The site is availabe on the http://projects.dfki.uni-kl.de/gofder/index.php URL. Figure 1 shows the logo of the glossary which depicts an important tool of the forensic scientist.

A. Related work

An offline glossary from 1999 is [2] which integrates first time in publishing terms from forensic science. The Forensic Science Central is a great contribution with links and forum, but its own glossary [3] contains only a small portion of definition, there is no dictionary and it is not extendible. Similar holds for the website of ThinkQuest [4], in addition there are less terms and the terms there were not described by experties, the reference marked on the page is only Google search. The multilingua lexicon of European Network of Forensic Science Institutes (ENFSI) is a great contribution created by several forensic institute, founded in 1999. It contains several words in many languages, but it is only a dictionary, without explanation, descriptions.

B. Authors

The first author Inés Rosa Baldatti is a forensic document examiner, analyst of Payment Systems at the Central Bank of Argentina Republic. The second author Erika Griechisch is a PhD student at the University of Szeged (Institute of Informatics), her topic is online signature verification. Further authors are every colleague who would like to contribute.

1) Beginning: The authors met at First International Workshop on Automated Forensic Handwriting Analysis (AFHA) in 2011, Beijing, China. In that opportunity it was clear the necessity to get a good understanding and a fluid contact among professionals worldwide. So, we decide to make this work, that takes a long time and dedication, and responsibility.

2) Creating: We read papers and books and websites, and extract the terms from them keeping the meaning.

II. Technical background

Nowadays a glossary or dictionary which is available only in printed version is not really useful. An online glossary is more useful than a printed one and it can be easily printed if it is necessary. Several criteria should meet, which are feasible only if the glossary is an online one.

Fig. 1. The logo of the project

Fig. 2. WikiEditor: WikiText and Preview (above), Changes (below)

The most important criterion is the expandability of the glossary. There will be always terms which can be added, not necessary because there incompleteness of the glossary itself rather the expension of the scientist. Regarding the expansion, it is much easier to correct errors in an existing online website than a book.

Other consideration was the possibility of localization of the glossary. We know for the translation part it is necessary to be clean and easy otherwise we can expect only few translators.

In this section we introduce and describe the MediaWiki package and its extensions which were most suitable for our purposes.

A. MediaWiki

MediaWiki (MW) is a versatile package, written in PHP and originally developed for use on Wikipedia since 2002. It is very widespread, well-documented, can handle any kind of media easily (links, images, videos, etc), moreover it is open-source which makes it easier to discover and fix any kind of bugs. The MediaWiki package quite flexible and further functionalities can be easily integrated to a basic MediaWiki website.

MediaWiki uses a markup language called *wikitext* to use basic formatting so the users without knowledge of HTML can edit the pages easily. We added the WikiEditor extension to the GoFDER website, which allows the users to see 3 different view during the editing. The first one is the plain wikitext, the second is a preview, the third one shows the differences between the previous version of the page and the current (edited) version. Thus if someone is new in wikitext markup language, s/he can simply check the Preview before submission, see Figure 2.

In order to achieve our goals, to create a multilingual glossary with dictionary, a basic MediaWiki is not sufficient. Thus we added other extension to improve the efficiency. In the following we describe each of them.

B. Semantical MediaWiki

Semantical MediaWiki (often noted by SMW) is an extension of MediaWiki that helps to search, organise, tag, browse, evaluate and share the wiki's content [8] since 2005.

While a traditional Wikipedia site contains text which is useful and can be processed easily by humans, it is not easy to understand or evaluate for a computers. The Semantical MediaWiki helps to extend a capability of a Wiki site by adding annotations, which makes wiki a collaborative database.

Semantical MediaWiki itself has several extension too, we added the Semantic Glossary to our website, which helps to describe terms with the Terminology page of the wiki. The reader of a page just point to a word with the mouse and if the Terminology page contains that word, it's description will appear in a small box below the word.

C. Translate extension

The Translate extension makes MediaWiki a powerful tool to translate every kind of text. [7]

It runs inside MediaWiki and has many features for translators, however its usage is very simple. After a page is marked with the `<translation>` tag, the extension automatically splits the text between the translation tag to translation units. The arrangment of the translation units can be approved or redefined if it is necessary. According to the default settings each translation unit is one paragraph. After the arrangement of the translation units are saved, the page can be translated via a translation tool. Figure 3 and 4 show two view of a translation page, translators can use which are more convinient for them.

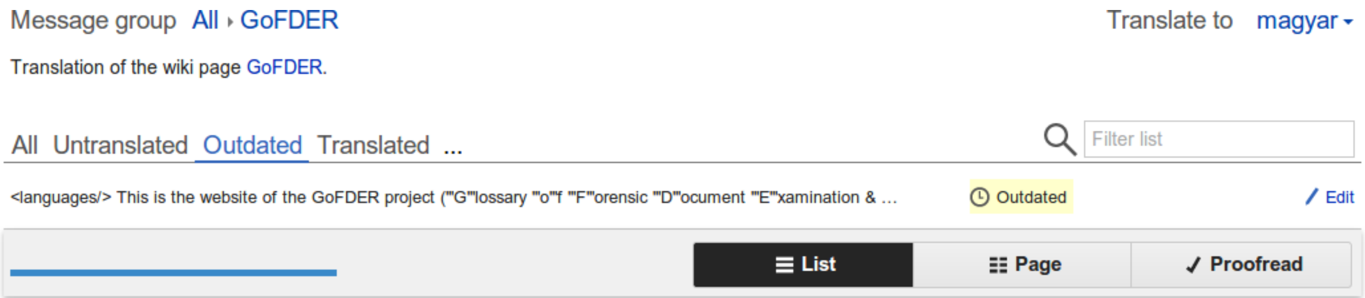


Fig. 3. Translate a page: *List* view

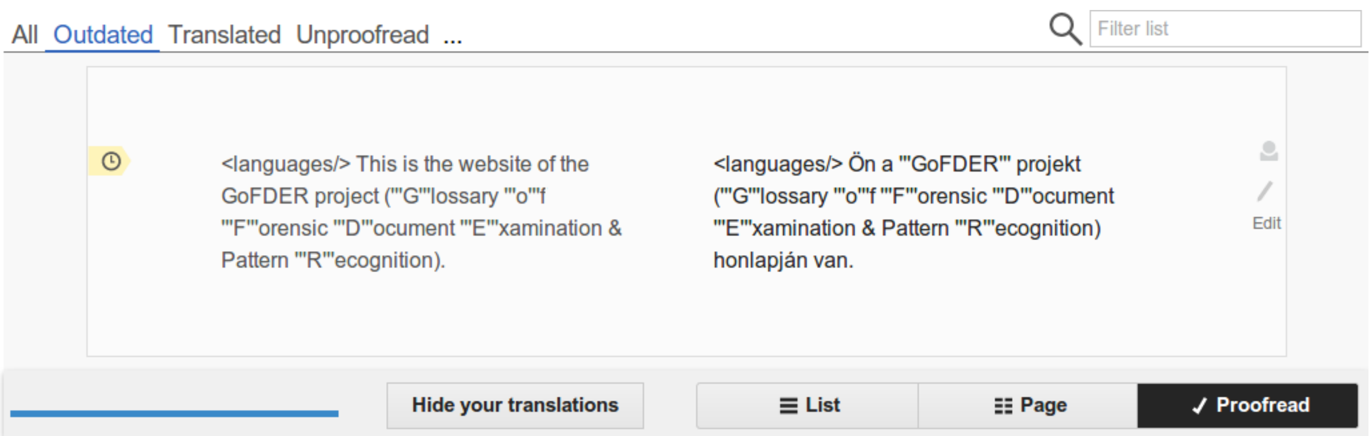


Fig. 4. Translate a page: *Proofread* view

Users can choose the language they intend to translate from a list. On the GoFDER website there is a predefined list on the top of all the translatable webpages which shows the languages we primarily intend to find translators (the first author of this paper is responsible mainly for Spanish translations, the second is about Hungarian). Nevertheless there is no language restrictions, contributors are welcome to translate to any language.

If an original English wikipage (which is marked for translation) has any changes, on the top of the page there will be a note about that and each unit which is effected will be marked as outdated translation.

III. Contribution

Recently (end of May, 2013) the glossary has 20 pages (terms), the dictionary has 561 English words with Spanish translation and some Hungarian translation as well.

A. New terms

We encourage forensic document examiners to extend the glossary and the dictionary as well, comment the recent terms on the discussion pages, ask if something is not clear or complete. We want to keep our glossary professional, so on our wiki site it is not possible to edit or modify pages without registration. After registration and login, users can modify the pages.

B. Translation

We would like to translate the terms from English primarily to Spanish, German and Hungarian. In the same way to increase the dictionary, we plan to add new languages as well. We hope to attain contributors who can create the Portuguese, Chinese, Japanese, Arabian, etc translation.

The translations are available for a page if it is marked for translation. Every registered user can translate sites, but only translation administrators can mark pages for translation and confirm completed translations. If a user creates a page and s/he is member of the Translation group, s/he can mark it for translation. if s/he is not member of this group, she can ask someone from the Translation group.

More details about editing and translating are available on the *Contribution* page of the website.

IV. Conclusion

Here we presented the conception and initation of our work. The technical background of the glossary have been prepared. However we are still open for suggestions and ideas to improve the website. From now on the main part of the project is to add new terms, specify and extend the existing ones, add examples and explanation as many as possible.

We expect as an outcome of our common effort with the contributors that we can provide a useful, up-to-date and beneficial glossary for the scientific community.

Acknowledgement

The second author was supported by the TÁMOP-4.2.4B/2-11/1-2012-0001 project.

References

[1] GoFDER website (May 2013 –) http://projects.dfki.uni-kl.de/gofder/
[2] John C. Brenner: *Forensic Science Glossary*, 1999, CRC Press (184 pages), http://www.crcpress.com/product/isbn/9780849311963
[3] Forensic Science Glossary of Forensic Science Central (May 2013) http://www.forensicsciencecentral.co.uk/glossary.shtml
[4] ThinkQuest Glossary (May 2013) http://library.thinkquest.org/04oct/00206/text_glossary.htm
[5] MULTILINGUA lexicon (May 2013) http://www.ies.krakow.pl/multilingua/welcome
[6] MediaWiki website (May 2013), http://www.mediawiki.org
[7] Translate Extension (May 2013) http://www.mediawiki.org/wiki/Extension:Translate
[8] Semantic MediaWiki (May 2013), http://semantic-mediawiki.org

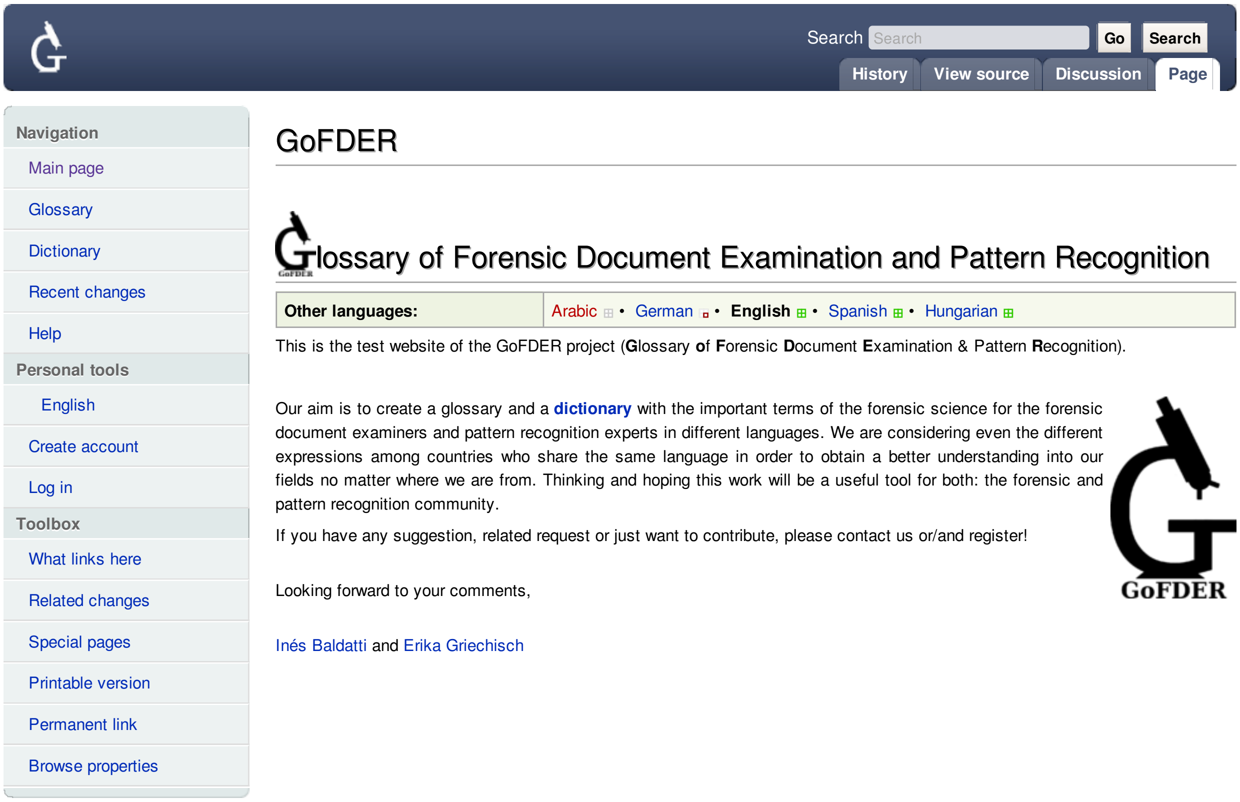


Fig. 5. Main page

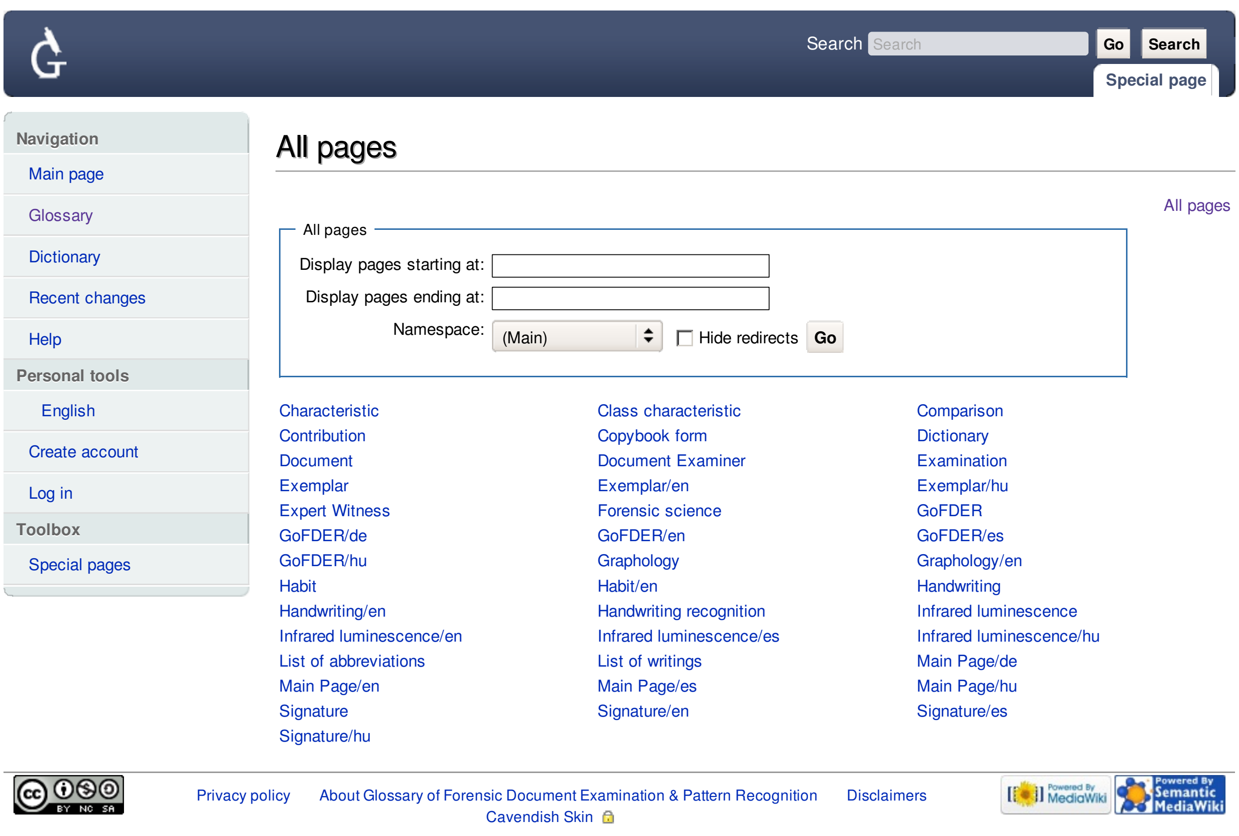


Fig. 6. Glossary – All pages