

Towards Automated Forensic Pen Ink Verification by Spectral Analysis

Michael Kalbitz^{1,2(✉)}, Tobias Scheidat^{1,2}, Benjamin Yüksel¹,
and Claus Vielhauer^{1,2}

¹ Department of Informatics and Media,
University of Applied Sciences Brandenburg,
Magdeburger Str. 50, 14770 Brandenburg an der Havel, Germany
michael.kalbitz@th-brandenburg.de

² Faculty of Computer Science, Otto-von-Guericke-University Magdeburg,
Universitätsplatz 2, 39106 Magdeburg, Germany

Abstract. Handwriting analysis plays an important role in crime scene forensics. Tasks of handwriting examination experts are for example identification of the originator of a given document, decision whether a signature is genuine or not, or to find strokes added to an existing writing in hindsight. In this paper we introduce the application of an UV–VIS–NIR spectroscope to digitize the reflection behavior of handwriting traces in the wavelength range from 163 nm to 844 nm (from ultraviolet over visible to near infrared). Further we suggest a method to distinguish ink of different pens from each other. The test set is build by 36 pens (nine different types in four colors each). From the individual strokes feature vectors are extracted which allows for a balanced classification accuracy of 96.12% using L1-norm and 95.26% for L2-norm.

Keywords: Handwriting forensics · Spectral analysis · Ink determination

1 Introduction

Handwriting still plays an important role in criminal forensics. Besides the question of who wrote a given document, there is also the question of whether some text or strokes were subsequently added with a different pen or ink. In forensic cases, such alterations are typically found in fabricated testaments or bank transfer forms.

Non-destructive examination is important for forensic work, since multiple investigations, of the same relevant substrate material, are often needed. In some cases, for example, a document must be examined for fingerprints after the handwriting analysis has been completed. Some conventional investigations, however, modify the evidence by physical or chemical treatment. That means that the investigation is partly destructive in nature. To address this drawback, there has been increasing interest in using surface scanning for forensic analysis in the digital domain.

While much research in forensics deals with the integrity and authenticity of digital media, there is also the question of whether digital representation might actually improve forensic investigation. Hildebrandt et al. [4], for example, show that a spectroscopic scan can make latent fingerprints visible. Another benefit of this approach is that it is possible to do both fingerprint and handwriting investigations with one scan.

In this paper, we propose a new method for ink verification in forensic documents by means of automated spectral analysis and classification. We introduce a new method based on the gradients of the spectral power histogram, present our first results of intra-class as well as inter-class feature distribution, and threshold-based classification.

This paper is structured as follows: Sect. 2 provides an overview of related work. The proposed process pipeline is described in Sect. 3. Detailed descriptions of the test set and the evaluation are presented in Sect. 4. Section 5 includes a short conclusion and the prospects for future work.

2 Related Work

To decide whether a handwritten document has been modified, it is important to be able to classify the writing instrument used, and identify handwriting strokes that have been added at a later date to an existing writing trace. In [7] Silva et al. present a method based on linear discriminant analysis to classify different types and brands of pens (five brands of ballpoint, two brands of roller ball, and three brands of gel). Here, data is acquired by an infrared spectroscope. The authors report correct recognition of type and brand in 99.5% and 100% of the test cases, respectively. These are good results, but at the described approach, the forensic expert has to define the measure points manually. In [3], Denman et al. use Time-of-Flight Secondary Ion Mass Spectroscopy for non-destructive analysis of organic and inorganic ink components. Their recognition rates for organic components and inorganic components were 84.4% and 91.1%, respectively.

An overview of typical components found in pen inks is given in [6]. Here, the author explains that pastes for ballpoint pens are still the most commonly used inks, but that there are some interesting alternatives, such as those used for roller handle or gel pens. In [5], the authors classify black gel pens and propose methods for estimating the age of a written document. In [2], Adam et al. investigate 25 black ballpoint pens from the UK market. They propose a principal component analysis to identify specific inks. Most of these approaches, however, are based on chemical investigations of the ink in question, and some are invasive in nature. Our question remains: is chemical knowledge of an ink really needed to verify or identify it?

The handwriting probes for our experimental evaluation were acquired using a UV-VIS-NIR spectroscope [1] (hereafter *UV-VIS-NIR*). Here, UV stands for ultra-violet radiation, VIS stands for visible light, and NIR stands for near infrared radiation. This industrial surface examination sensor can capture luminance values at 2048 discrete levels for wavelengths between 163 nm and 844 nm,

that is, for wavelengths in the ultraviolet, visible, and near infrared ranges. The measurement of each single point results in 2048 values, one for each discrete level within the wavelength range provided by UV–VIS–NIR. The results between 163 nm and 200 nm, however, are quite noisy, due to oxygen absorption of the ultraviolet radiation.

3 Methodology

The Methodology is based on a four-step processing pipeline introduced by Vielhauer in [8]. The processing pipeline is shown in Fig. 1, and described in the following subsections in context of the suggested methodology.



Fig. 1. Processing pipeline

3.1 Data Acquisition

With best knowledge of the authors, there is no public test database available, containing spectral information of different inks. By this reason, an appropriate testset was created. For preparation of handwriting traces we use 9 types of pens with four color instances each (blue, black, red, green). A list of all 36 pens is given in Table 1. Since we do not need complete handwriting traces to determine the used ink, the probes are simplified to a single stroke of approximately 5 mm per pen on white copy paper (80 g/m²).

Table 1. List of inks in our test set

Pen id	Brand id	Type	Color
1	1	Ballpoint pen	Blue, black, red, green
2	2	Four color ballpoint pen	
3	3	Ballpoint pen	
4	4	Gel rollerball	
5	4	Liquid ink rollerball	
6	4	Fineliner	
7	4	Fibre tip-pen	
8	5	Liquid gel pen	
9	6	Four color ballpoint pen	

During acquisition, each stroke is measured by the UV-VIS-NIR. We use a scan area of $2.8 \times 1.6 \text{ mm}^2$ with a dot distance of $100 \mu\text{m}$ and an integration time of 300 ms. This means, in total we obtain $28 \times 16 = 448$ measure points per stroke. For each measure point the sensor provides a vector of 2048 entries, each of them represents the reflection power in one of 2048 wavelength bins w_0, \dots, w_{2047} . The structure of the UV-VIS-NIR data is shown in Fig. 2.

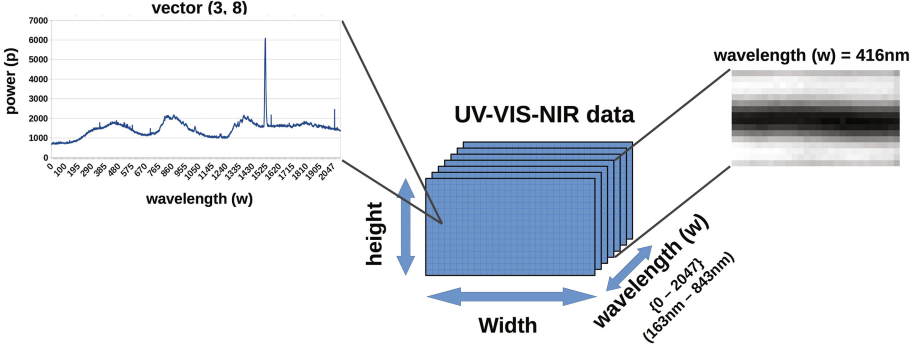


Fig. 2. Structure of the sensor data: for each of the stroke samples (right), 28/16 measurement points in width/height respectively (middle) are taken. For each point, the wavelength response is provided as a power histogram of 2048 wavelength bands (left).

3.2 Pre-processing

The aim of the pre-processing is to extract measure points which contain parts of the stroke. Figure 3 show the suggested pre-processing chain. We identified five steps for our pre-processing:

1. remove measurement drop-outs
2. built a reference vector
3. normalize the wavelength
4. extract writing trace
5. export data

The measurement drop-outs addressed in the first step occur, if the reflected energy is considered insufficiently low by the sensor. In Fig. 3 such errors are the black pixels in the image on top. If drop-out occur, this values are replaced by the median for each wavelength (of the 2048) from the neighbors of this measure point. The pseudo code for this step is given in Algorithm 1.

The second step is to built a reference vector. Therefore it is needed to determine which measure points are background (white paper) and which are foreground (ink of stroke). The reference vector represents for each wavelength the reflection behavior of the used lamp and the carrier material.

Algorithm 1. Remove measure errors

```

for all mp in measure-points do
  if norm(mp) == 0 then
    mp ← median-from-neighbors(mp)
  end if
end for

```

Based on this acquisition technique we have 2048 images of the scanned area where each image represents one wavelength. Classic segmentation approaches will to the best knowledge of the authors not work with this kind of data. Based on color theory we know that a white object needs to have a high reflection behavior in the visible light area. In knowledge of this, and the fact that the background of the handwriting samples is white copy paper we search in the scanned data for those 20% measured points having the highest power (in sum over all 2048 wavelengths) as shown with the pseudo-code in Algorithm 2. This threshold is used, because we want to ensure that at least this area will be not covered by the stroke. The background points from a scan are red marked in Fig. 3. Based on the 20% measure points having highest power a reference vector is created by using the median value calculated for each wavelength.

Algorithm 2. Background detection

```

smp ← sort-by-max-norm(measure-points)
smp20 ← get-first-20-percent-entries(smp)
for i = 0; i < count-wavelength; i++ do
  reference-vector(i) ← get-median-over-all-entries(smp20, i)
end for return reference-vector

```

The third step is to normalize the wavelengths for each measure point, based on the reference wavelengths. This will be done by division of each measure point by the reference background value at the same wavelength. After normalisation, we observe normalized wavelength response values in the range of 0.18 to 1.16, former they are in a range of 642 to 7806. An example for a normalised measure point is shown in the middle of Fig. 3. Tests with different normalization methods such as min max mapped to $[0 \dots 1]$ does not show significant changes of the contribution of the latter computed distances.

The fourth step extract the relevant measure points from the handwriting strokes. With the knowledge that our acquired stroke has 28 columns (as shown in Fig. 2) an iteration from left to right combined with a search for the measure point having the highest power is carried out. The selected measure points from the stroke are also red marked in the lower portion of Fig. 3.

In the fifth step the relevant measure points are exported to a csv-file. This was done to decrease the processing time of the feature extraction. Thus, there are 28 vectors for each of the 36 strokes. An exemplary file is shown in the bottom of Fig. 3. In the first column is a index, the second column contains the

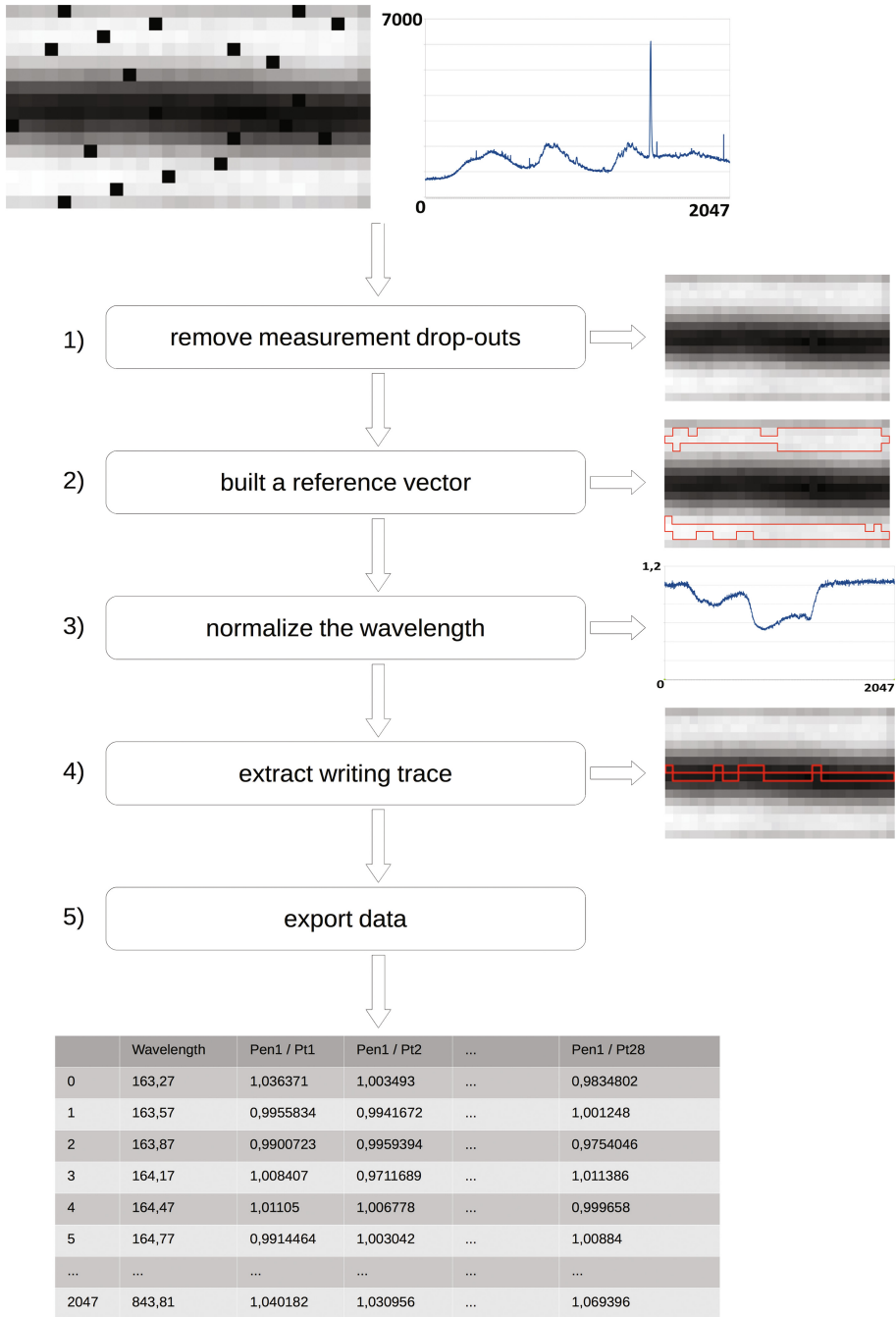


Fig. 3. Pre-Processing approach on the example of pen 1. From top to bottom: raw data, pre-processing steps with interim results, pre-processing result. Measure points, which are detected in the single steps, are marked with a red border. (Color figure online)

specific wavelength, and in the last (28) columns are the values of the selected measure points for the specific wavelength.

3.3 Feature Extraction

Due to the observed variance in the energy response of the wavelength bands, we decide to use the gradients as features. To simplify the calculation, we assume a linear gradient between two neighbored wavelengths by using Eq. 1. In which w_{i-1}/p_{i-1} is the previous wavelength/power and w_{i+1}/p_{i+1} is the next wavelength/power in the vector. In this approach, the wavelength gradient is estimated by a numerical differentiation. The exact gradient can not be calculated because we have only values of a function and not the function itself.

$$m = \frac{p_{i-1} - p_{i+1}}{w_{i-1} - w_{i+1}} \quad (1)$$

To determine the difference between two measure points, the distance between this vectors is calculated by means of the absolute value norm (*L1-norm*) or euclidean-norm (*L2-norm*) for each of the 2048 vector components. This comparison is carried out for all 28 measure points of any individual stroke (intra class comparison) as well as for all 28 measure points of one single stroke against each measure point of all other strokes (inter class comparison).

3.4 Classification

The classification of the suggested methodology is a verification based on a binary decision that decides whether two given feature vectors can be assigned to the same class (intra class, verified) or not (inter class, declined). The decision is based on a threshold estimated by the Equal Error Rate (EER).

4 Experimental Investigation

The experimental investigation is focused on the classification accuracy. The data acquisition, pre-processing, and feature extraction are carried out as described in Sect. 3.

4.1 Test Methodology

The test set consists of 36 strokes written using 9 different pen types (each in the color black, blue, red, and green). From every stroke 28 measure points are extracted. This procedure results in 1008 measure points. For the evaluation we distinguish between measure points from the same stroke (intra class) and measure points from two different strokes (inter class). Since distances are determined, we treat the comparisons symmetrically, i.e. when measure point 1 has been compared to point 2, we do not need to compare 2 to 1 again. The comparisons are based on L1-norm, and L2-norm. The procedure of intra class

measure points	1	2	3	4	5	...	28
comparison 1-28		1-2	1-3	1-4	1-5	...	1-28
comparison 29-55			2-3	2-4	2-5	...	2-28
comparison 56-81				3-4	3-5	...	3-28
⋮						⋮	
comparison 376-377							26-27 26-28
comparison 378							27-28

Fig. 4. Intra test method for one stroke

comparisons is shown in Fig. 4. We compare each measure point from a stroke with each other measure point from the same stroke. The number of resulting comparisons (c) for the complete intra test can be calculated by Eq. 2 where p is the number of used pens (36), and n is the number of extracted measure points (28).

$$c = p * \frac{n * (n - 1)}{2} \quad (2)$$

For the intra test 378 comparisons are made per stroke or a total of 13 608 comparisons. The method for the inter class comparison is illustrated in Fig. 5. Here, we compare each measure point of each stroke with each measure point of all other strokes. That means 980 comparisons for each measure point (28 measure points * 35 pens). Overall the amount on comparisons for the inter class test can be calculated with the formula 3 where p is the number of used pens (36), n is the number of extracted measure points (28), and $intra$ is the number of intra test comparisons. This means over all we made 493 920 inter class comparisons.

$$c = \frac{p * n * (p * n - 1)}{2} - intra \quad (3)$$

The distribution (over all comparisons) for L1-norm and L2-norm are shown in Figs. 6 and 7.

4.2 Evaluation

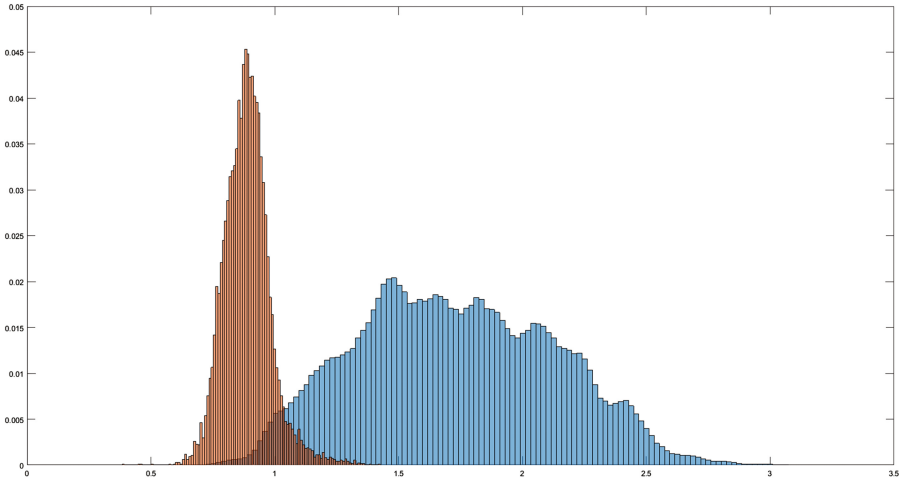
Based on the distribution of the distances for intra and inter classes we calculate FMR and FNMR for different threshold which map all occurring distances. The plots are shown in Fig. 8 for L1-norm and Fig. 9 for L2-norm. FMR means False Match Rate that happen if the distance is less than the selected threshold but it is not in the same class. FNMR means False None Match Rate that happens if the distance is greater than the threshold but is in the same class. The intersection

		stroke 1				
		S1P1	S1P2	S1P3	...	S1P28
stroke 2	S2P1	S1P1-S2P1	S1P2-S2P1	S1P3-S2P1	...	S1P28-S2P1
	S2P2	S1P1-S2P2	S1P2-S2P2	S1P3-S2P2	...	S1P28-S2P2
	S2P3	S1P1-S2P3	S1P2-S2P3	S1P3-S2P3	...	S1P28-S2P3

	S2P28	S1P1-S2P28	S1P2-S2P28	S1P3-S2P28	...	S1P28-S2P28
stroke 3	S3P1	S1P1-S3P1	S1P2-S3P1	S1P3-S3P1	...	S1P28-S3P1
	S3P2	S1P1-S3P2	S1P2-S3P2	S1P3-S3P2	...	S1P28-S3P2
	S3P3	S1P1-S3P3	S1P2-S3P3	S1P3-S3P3	...	S1P28-S3P3

	S3P28	S1P1-S3P28	S1P2-S3P28	S1P3-S3P28	...	S1P28-S3P28
⋮		⋮	⋮	⋮	⋮	⋮
stroke 36	S36P1	S1P1-S36P1	S1P2-S36P1	S1P3-S36P1	...	S1P28-S36P1
	S36P2	S1P1-S36P2	S1P2-S36P2	S1P3-S36P2	...	S1P28-S36P2
	S36P3	S1P1-S36P3	S1P2-S36P3	S1P3-S36P3	...	S1P28-S36P3

	S36P28	S1P1-S36P28	S1P2-S36P28	S1P3-S36P28	...	S1P28-S36P28

Fig. 5. Inter test method on the example of stroke 1.**Fig. 6.** Distribution of L1-norm (orange = intra class, blue = inter class). The distributions are relative represent, that means it show the percentage frequency for each distance (Color figure online)

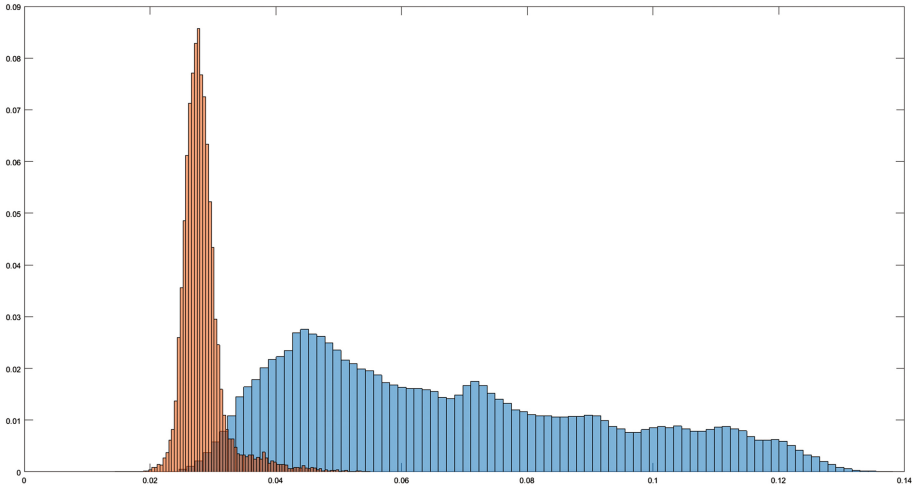


Fig. 7. Distribution of L2-norm (orange = intra class, blue = inter class). The distributions are relative represent, that means it show the percentage frequency for each distance (Color figure online)

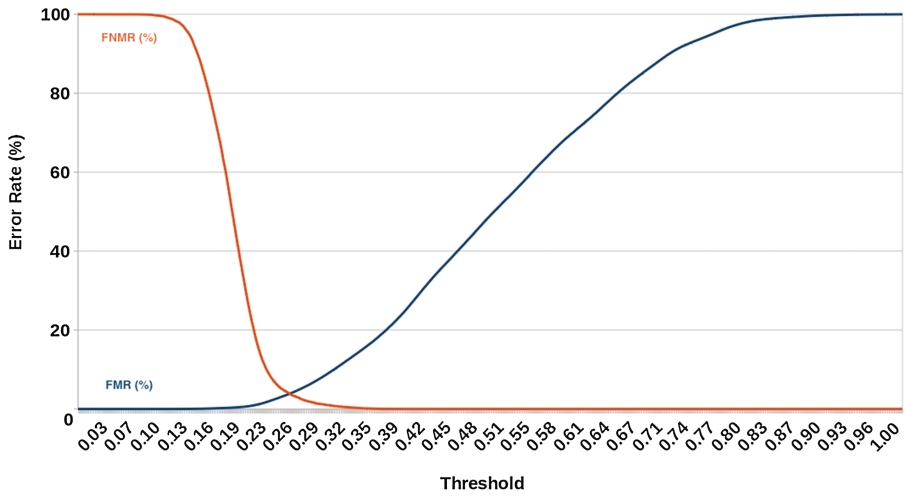


Fig. 8. FMR and FNMR for L1-norm, EER = 3.88%, Threshold = 0.256

of the both curve is the EER. Based on the EER we got a balanced classification accuracy of 96.12% for L1-norm, and 95.26% for the L2-norm.

Relative confusion matrices for L1/L2-norm are shown in Tables 2 and 3 determined on the base of the threshold of the EER for the intra and inter classification accuracy.

Table 2. Confusion matrix for L1-norm (threshold = 1.0656)

	Intra class	Inter class
Test positive	96.17%	3.83%
Test negative	3.94%	96.06%

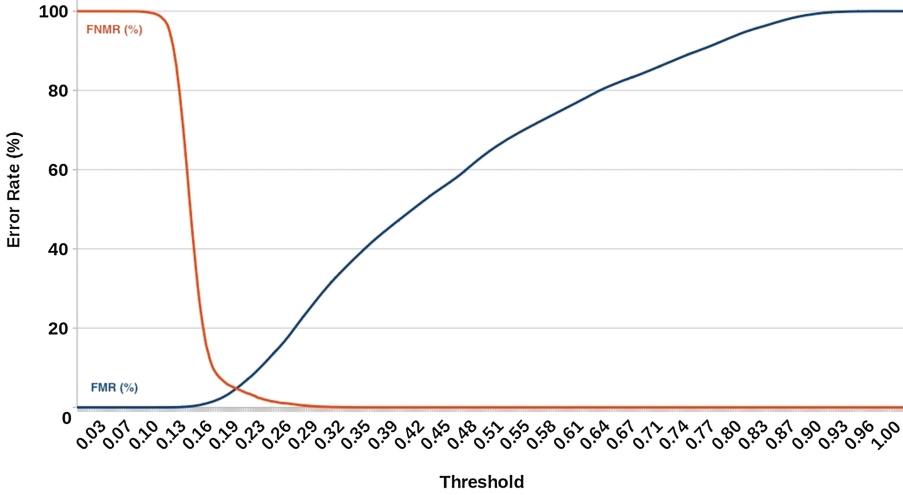


Fig. 9. FMR and FNMR for L2-norm, EER = 4.74%, Threshold = 0.192

Table 3. Confusion matrix for L2-norm (threshold = 0.03496)

	Intra class	Inter class
Test positive	95.27%	4.73%
Test negative	4.76%	95.24%

4.3 Results in Context of the State of the Art

The test set described in [7] consists of 100 blue ink pens, containing ten different instances per brand of ballpoint pens (five brands), roller ball pens (two brands) and gel pens (three brands). Best result determined for the identification of pen type described in [7] amounts 99.5%. Compared to the highest classification accuracy of 96.12% calculated in this paper there is a marginal difference. In contrast to our automatic approach, the measure points are preselected by hand. Keeping in mind, since the values are based on varying test sets containing different compilations of pen types, colors, and brands, the results are not comparable to each other directly. However, the evaluation shows feasible methodology and test setup for ink investigation within handwriting forensic context. Further, the results of our initial research are based on a simple

approach by using the linear gradients between neighbored frequencies as features. On the other side, data acquired by the UV–VIS–NIR provides comprehensive information to develop additional features in future work to increase classification accuracy further more.

5 Conclusion and Future Work

Ink verification by spectral analysis seems to be promising. Even based on one simple feature we observe first quite good results for both intra and inter class determination. Based on our results, it is possible to verify different inks automatically in a scan with an EER of 3.88% (L1-norm). To do an identification of the ink there is the need to establish a database with reference scans for as many as possible inks.

The challenge for the intra class determination is that the amount of tests is just a fraction of the inter tests (nearly 3%). A false positive ratio of 3.83% represent absolute more measure point recognize as false positive as true positive. To decrease the number of false positive, it should be investigated, if it would useful to classify on a vote from several measure points instead of a single measure point. Another possibility could be to add more features to get a higher sensitivity.

Future work could be include investigations on overlapped strokes as well as on the influence of different writing conditions, such as black permanent marker, text marker or colored paper. Further, additional features will be designed and implemented in future. Based on those, possibilities of machine learning methods can be studied for ink identification. Another future topic is the research on effects of inks aging (duration between writing time and acquisition time) on recognition behavior. Furthermore, methods should be investigated to identify the ink type and/or brand based on a scan of the whole document.

Acknowledgments. This work has been funded by the German Federal Ministry of Education and Research (BMBF, contract no. FKZ 03FH028IX5). Authors would like to thank the staff of the FRT GmbH for their support and fruitful discussions on relevant features for ink identification. Further, we thank all colleagues of the two research groups at both universities for their advices.

References

1. FRT GmbH - The Art of Metrology (2016). <https://frtmetrology.com/en>. Accessed 16 May 2017
2. Adam, C.D., Sherratt, S.L., Zholobenko, V.L.: Classification and individualisation of black ballpoint pen inks using principal component analysis of uv-vis absorption spectra. *Forensic Sci. Int.* **174**(1), 16–25 (2008). <http://www.sciencedirect.com/science/article/pii/S0379073807001326>
3. Denman, J.A., Skinner, W.M., Kirkbride, K.P., Kempson, I.M.: Organic and inorganic discrimination of ballpoint pen inks by ToF-SIMS and multivariate statistics. *Appl. Surf. Sci.* **256**, 2155–2163 (2010)

4. Hildebrandt, M., Makrushin, A., Qian, K., Dittmann, J.: Visibility assessment of latent fingerprints on challenging substrates in spectroscopic scans. In: Decker, B., Dittmann, J., Kraetzer, C., Vielhauer, C. (eds.) CMS 2013. LNCS, vol. 8099, pp. 200–203. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-40779-6_18](https://doi.org/10.1007/978-3-642-40779-6_18)
5. Liu, Y.Z., Yu, J., Xie, M.X., Liu, Y., Han, J., Jing, T.T.: Classification and dating of black gel pen ink by ion-pairing high-performance liquid chromatography. *J. Chromatogr. A* **1135**(1), 57–64 (2006). <http://www.sciencedirect.com/science/article/pii/S0021967306017808>
6. Petrov, P.: Classification Pen by Type of Ink (2017). <http://blogadney.eu/classification-pen-by-type-of-ink/>. Accessed 17 May 2017
7. Silva, C.S., de Borba, F.S.L., Pimentel, M.F., Pontes, M.J.C., Honorato, R.S., Pasquini, C.: Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis. *Microchem. J.* **109**, 122–127 (2013)
8. Vielhauer, C.: Biometric User Authentication for IT Security: From Fundamentals to Handwriting. Springer, New York (2006)

Digital Forensics and Watermarking

16th International Workshop , IWDW 2017, Magdeburg,

Germany, August 23-25, 2017, Proceedings

Kraetzer, C.; Shi, Y.-Q.; Dittmann, J.; Kim, H.J. (Eds.)

2017, XI, 422 p. 174 illus., Softcover

ISBN: 978-3-319-64184-3