The WANDA Measurement Tool for Forensic Document Examination

Merijn VAN ERP, Louis G. VUURPIJL, Katrin FRANKE, Lambert R.B. SCHOMAKER

1 NICI, P.O. Box 9104, 6500 HE Nijmegen, the Netherlands

{M.vanErp, vuurpijl}@nici.kun.nl

2 Fraunhofer IPK, Department of Pattern Recognition,

Pascalstrasse 8-9, 10587 Berlin, Germany

katrin.franke@ipk.fhg.de

3 Department of Artificial Intelligence, University of Groningen

Grote Kruisstraat 2/1, 9712 TS Groningen, the Netherlands

schomaker@ai.rug.nl

Abstract

This paper introduces the WANDA Measurement tool (WAM) for forensic document examination. The WANDA system is a desktop workbench that supports the user in the complete process of measuring characteristic features in handwritten documents. By using technologies like plug-ins, XML, and client/server modularity, a system was created that is easy to maintain, portable, and highly adaptable. Within WANDA, the WAM is the tool for interactively measuring the handwriting features. The WAM is based on the results of a comparison study between two writer-identification systems, Script and FISH. It incorporates nine measurements identical to those of FISH and a new *allograph* measurement. Furthermore, its intuitive new user-interface reduces the steep learning curve and streamlines the working process. The precision of the WAM was assessed by comparing the obtained features to measurements made by forensic experts using FISH. This was further validated by comparing handwriting on paper with corresponding scanned offline images. Both validations show that the WAM performs correctly. First usability tests with expert and novice users show that the WAM is easy to use and yields equivalent measurement values to those produced with FISH.

1 Introduction

Since 1986, the Forensic Informationsystem on Handwriting (FISH) [Phillip, 1996] is being used by the German law enforcement agency Bundeskriminalamt (BKA). Various other forensic institutes such as the US Secret Service and the Dutch NFI also make use of this tool. The FISH system is a handwriting analysis and writer identification system, which enables the user (a forensic handwriting expert) to measure the features of a piece of handwriting. The handwriting, annotated with the measured features, will then be compared with earlier collected and measured pieces of handwriting in order to find the best matches. Schomaker & Vuurpijl (2000) concluded in a comparison between FISH and the Dutch system Script [de Jong et al., 1994] that, though FISH is an excellent system when it comes to writer identification results, the user-interface of FISH should be improved. Furthermore, a technological update was necessary to improve its portability and ease of maintenance.

In a joint initiative led by the Frauenhofer IPK, an international group of institutes has developed a prototype for a new handwriting analysis and writer identification system, christened WANDA [Franke et al., 2003][Franke et al., 2004]. Similar initiatives for a modern forensic handwriting examination tools that are in progress can be found in [Shrihari & Leedham, 2003][LumenIQ, 2004][Pikaso, 2004]. WANDA provides a large update to FISH. This update consists of technologies like network access, client/server modularity, exchangeable plug-ins and human readable XML messaging and storage. The new system was built in Java, using highly portable public domain software and no proprietary data formats. These technical updates ensure that the system is up to date with current developments, platform independent, and easy to expand and maintain. Furthermore, WANDA has a new user interface that conforms to the standard interfaces that current computer users will be familiar with.

In this article we will focus on one specific tool in the WANDA workbench, the WANDA Measurement tool (WAM). The WAM enables the user to perform interactive feature measurements on handwriting. The first priorities of the WAM were to improve user friendliness, to reduce the amount of labor involved during measurements, to reduce the amount of subjective influence by the user on the measurements, and to allow novice users to be able to work with the system, while retaining the possibility of fast expert usage.

In this paper, first the WAM user-interface will be illustrated. Next, the *allograph* measurement, a new measurement introduced in the WAM, is described. Subsequently, we will present our validation of the acquired measurement values by comparing them to FISH measurements from images from the BKA databases. A further validation of measurement features was performed by comparing data from different sources. Finally, the usability study that was performed on the WAM to validate the user-interface is discussed.

2 The WAM user-interface

One of the main goals of WANDA was a modern, easy-to-use user-interface. In this Section we will describe how this was accomplished for the WAM interface. To better understand the WAM and its user interface, it is necessary to see how the WAM fits into the WANDA workbench. When a user has a piece of handwriting that needs to be examined and identified, some steps will be taken before the WAM needs to be used.

First, the user will scan the handwriting. This scan will be stored in the WANDA database and will be the basis for all subsequent steps. The user can now retrieve the scan from the database. From the scanned image, the user can now select a so-called *region of interest* (ROI). An ROI is the part of the handwriting that the user is currently working on. The concept of the ROI was introduced for multiple reasons. There may suspected that more than one writer produced the handwriting, only certain pieces may be interesting, and it is often easier to work on only part of the entire image to get a more detailed view. Note that the user can easily select the entire image as a ROI if that is preferred.

When a ROI is selected, the user can employ the rest of the WANDA workbench on it, including the WAM. However, it is prudent to use the image pre-processing tool of WANDA before the WAM is used. That way, the image can be cleared of any parts that do not belong to the handwriting and that can interfere with precise measurements. After this, the scanned handwriting is ready to be measured by the WAM.

The WAM interface consists of a main window and a measure window. The main window in the WAM (see Figure 1) provides an overview of all measurements that have been performed on the current ROI. The measurements are color-coded by type, to enable the user to quickly identify them and to avoid cluttering the image too much.

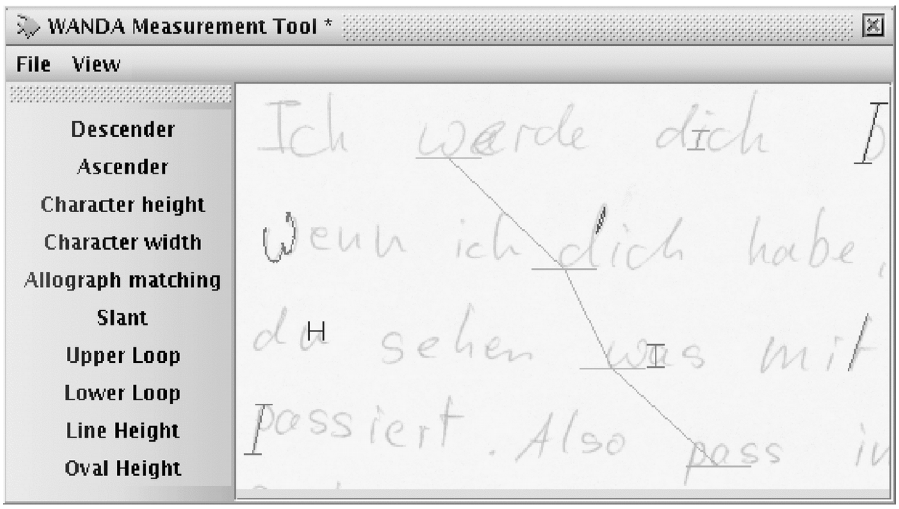


Figure 1: The main window of the WANDA measurement tool (WAM).

The buttons provide a clear overview of the available functions of the tool. They also provide mouse support for all commands, which is easier for novice users (especially of the Windows generation) than accelerator keys. Next to mouse support, accelerator keys are still provided for expert users to allow fast and efficient interactions. The measurement data is presented in a table that can be accessed via the menu. This data, calculated per measurement type, consists of the number of measurements done in this ROI, their average value, and its standard deviation.

Measurements are done one type at a time. No specific order is necessary and a user does not have to finish all measurements of a certain type of measurement before starting another type. An unfinished type can always be returned to later.

When the user wants to initiate a measurement type, the button corresponding to this type needs to be pressed. A new window, the measure window, is opened. This window is similar for all measurements (see Figure 2). It shows of the ROI containing the handwriting to measure on (number 3 in Figure 2). To measure a feature, the user has to select a character to measure this feature on from the drop down list (1). Now, the user can start measuring a character matching the one chosen. During this process the user is supported by the instruction window (3), a WAM interface novelty that guides the user step-by-step through the measurement process. This allows novice users to start measuring right away without hindering experts in any way. Finally, the three control buttons (4, 5, and 6) can be used respectively to quit measuring this type of measurement, to stop a currently active measurement and start over, and to accept and finish the current measurement.

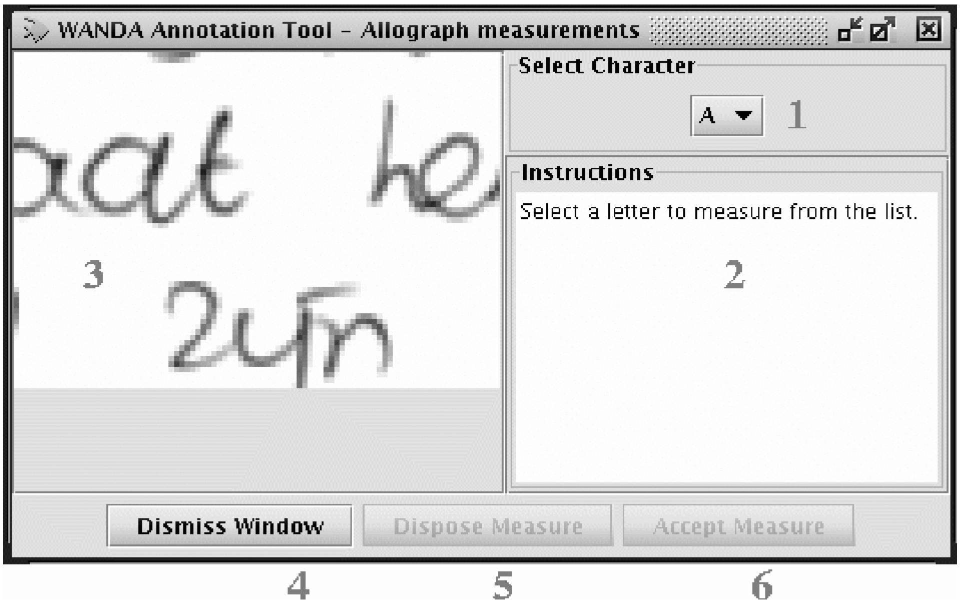
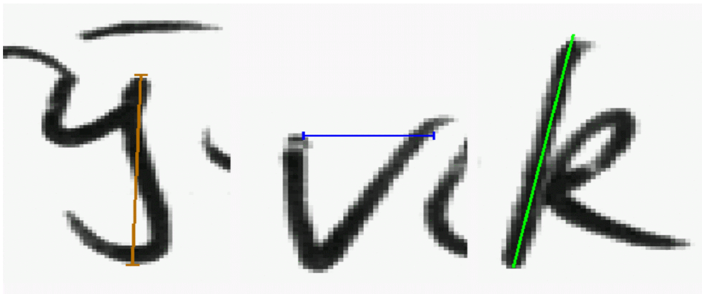


Figure 2: The measure window of the WANDA measurement tool (WAM).

The WAM has ten different measurements. The basic measurements consist of the various character heights (ascenders, descenders, corpus height, and height of oval characters), the slant of characters, and the character width. If present in the handwriting, the WAM allows the users to measure (upper and lower) loops of characters. As the only non-character based feature, the WAM also provides the measurement of the average distance between the baselines in a piece of handwriting. The last measurement, the new allograph measurement, is discussed in detail in the next Section. The other nine measurements are identical to the measurements performed in FISH. We will give a brief summary of the measurements below. A more detailed description of the supported measurements is given in [Phillip, 1996].



(a) Height (b) Width (c) Slant

Figure 3: Typical examples of the height, width and slant measurements.

The ascenders and descenders are measured from the top to the bottom of the vertical strokes (not necessarily the top and bottom of the letter). The character heights are measured on the vertical strokes extending from the baseline to the corpus line in (e.g. in the 'n', 'u' an 'i'), while the oval height measures the distance of the top of the oval part of letters like 'a', 'o' and 'd' to the bottom of that oval. All measurements start at the first ink pixel and end at (including) the last ink pixel (for example, see Figure 3a for a descender).

The line height, the distance between two consecutive baselines, is measured from the bottom pixel of a letter on the first line to the bottom pixel of a letter on the consecutive line.

The width of a character is measured on 'cupped' letters (e.g. 'u' and 'n'). It is the distance from the right edge of both cup-ends (i.e. the inside of the letter on the left side and the outside of the letter on the right, see Figure 3b).

The slant (see Figure 3c) is performed by drawing a line on an ascender (preferably, though descenders are allowed). The angle between the drawn line and the x-axis is considered the slant of the letter.

Finally, loops are measured by calculating three values based on the trace of inside edge of the loop. By searching the inner bound of the loop, the WAM will determine the longest and shortest length across and return as measured values the longest length, the ratio between the longest and shortest length (the form of the loop) and the slant of the longest length line. These three values are used to characterize the loop.

3 Recognition-based measurements

In the WAM, a new measurement was introduced, the allograph match [Vuurpijl & Schomaker, 1997]. Allographs are specific character forms or styles. Allographs are somewhat comparable with fonts like, for example, 'Times New Roman' and 'Helvetica', but on a character by character base. So, an allograph of an 'a' is a very specific way in which an 'a' can be written. Every person has his or her own set of allographs that often contain some subset of characters that is unique for that person. As such, allographs are very helpful in identifying the writer of a piece of handwriting. The allograph measurement of the WAM can be used to collect these allographs. During this measurement, the user is asked to trace a letter in the questioned document by pen-tablet or mouse. This creates an on-line trajectory of the allograph of that letter. The trajectory is then matched to a database of on-line allograph prototypes and the WAM will present the best matches for the user to choose from (see Figure 4). The label of this best matching prototype can then be used as one of the features in writer identification. The WAM uses prototypes instead of actual allographs for two reasons. First, it greatly reduces the amount of computation time that is required to match the allographs of different documents. Now only labels have to be matched. Second, it results in a more objective measurement. The prototypes are fixed, while the traces that users produce will vary from person to person.

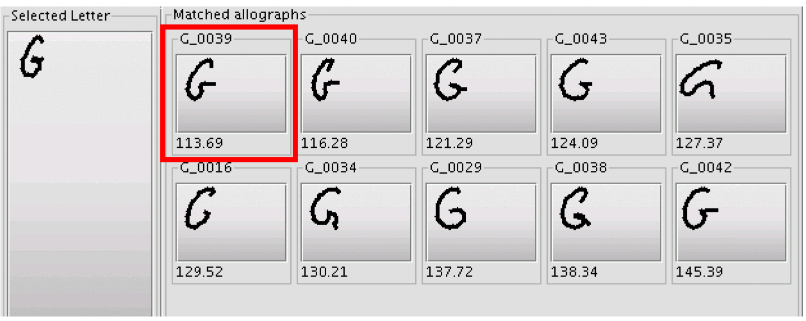


Figure 4: The presentation of the matching allograph prototypes.

The allograph match was added to the WAM as a proof of concept for recognition-based measurements. The idea is that if the program is able to recognize the written text, it can automatically and objectively measure all needed features for writer identification. As an added advantage, it opens up the well-researched area of pattern recognition techniques, techniques that are often directly applicable to writer identification. Another example of this can be found in [Schomaker, et al., 2003]).

4 Validation of measurements

One of the hard requirements of the WANDA project was that the resulting measurements should be backward compatible with the FISH measurements. FISH has been used for over a decade and a large database of cases has evolved. Not meeting the compatibility restraint would make this huge pool of data useless. Therefore, the measurements from WANDA should very closely resemble the measurements from FISH. A validation procedure was conducted to ensure that the results are compatible. Two kinds of validation were performed: one on the images that were measured and one by comparing the measurement results. The first validation assesses whether the scanned image on which WANDA measures the features is equal to the actual handwriting on paper.

The image validation was done by measuring features on paper and the same features from the image file. Measurements on paper were performed with both a measure with a (coarse) millimeter scale and a high-resolution microscope (very precise). Those were compared to the digitized version (300 dpi) of the image of which the number of pixels for the features was determined and to the results of measuring those features in the WAM (in tenths of millimeters). All results proved precise within 0.85 mm (i.e. less than one 300 dpi pixel size) deviation between the ink on paper and the digitized ink.

The second validation, that on the measurement results, was done to ensure the backward compatibility of Wanda to FISH. For this validation, the measurement results of Wanda and FISH for the same measurements were compared. The validation was done on 11 scans provided by the BKA. For each scan, the original

FISH features were available. All FISH measurements were carefully redone in the WAM tool. Next the results of the Wanda tool were compared to the results the FISH tool gave for these scans. Some points on the precision of this validation must be noted. First, the provided FISH scans were in 200 dpi, while the WANDA scans were in 300 dpi. We chose to measure with this difference, as 300 dpi should be the lowest resolution on which WANDA measurements are to be taken. Lower resolutions lead to lower precision and the currently available scanners are capable of handling 300 dpi without any problem. For this validation however, this means that even a perfect compatibility can still result in a deviation of up to two (300 dpi) pixels.

The second point to note is that only had the average of each type of measurement (e.g. ascenders and slant) per document measured in FISH was available. Individual measurement results (compared to the actual measurement image) are not stored in FISH. Therefore, only average scores could be compared.

Finally, it is unavoidable that some measurements done in WANDA differed slightly (one or two pixels) in placement compared to FISH. At times the scans were not clear on the (very) precise location of a measurement, so a subjective choice had to be made.

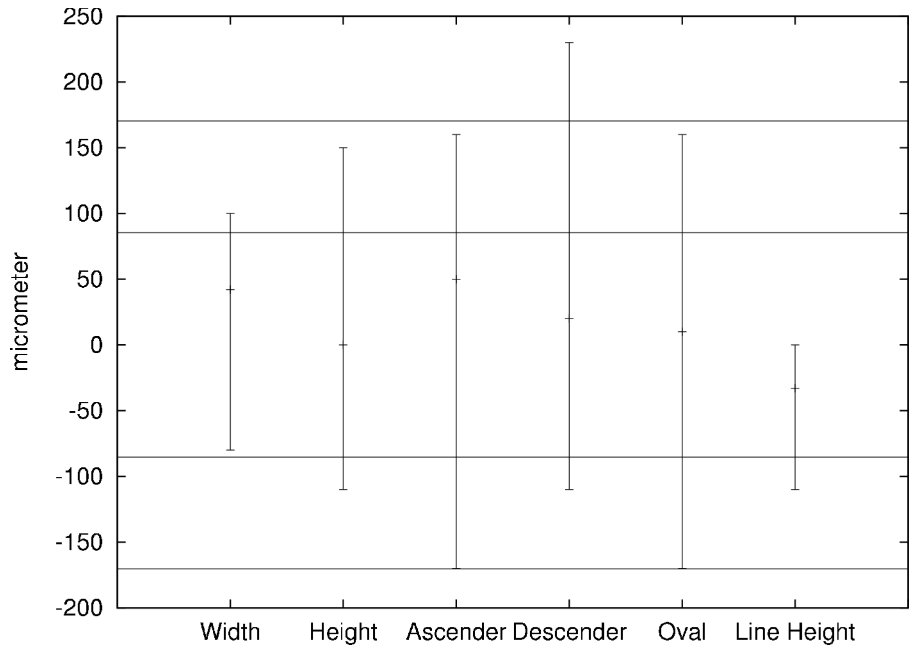


Figure 5: The difference between the Wanda measurement results and the FISH results. The bars represent the range in which all differences over all documents fell. The horizontal lines signify a difference of one and two pixels in either way.

Figure 5 shows the results of the validation. The y-axis shows the difference between the WANDA and the FISH measurements (i.e. the WANDA measurement minus the FISH measurement). For each type of measurement that could be validated, a vertical bar is displayed. This bar indicates the range over which this difference occurs. The small dot in the middle of each bar indicates the average difference between each pair of measurements. Note that the slant and loop measurements are missing as they have a different scale and the resolution change has a different impact on them. The results of those validations were comparable in conclusion with those shown.

From the results we can conclude that, except for the descender, the differences between FISH and WANDA fall within the 2-pixel limit for all measurement types. This is as expected with the resolution change and a good compatibility between WANDA and FISH. Also note that each average difference falls well within the 1-pixel range, indicating that the differences measured are divided more or less equally over “too high” and “too low”. What is not so clear from the figure is that the extreme ranges are mostly due to single occurrences. In case of the descender for example, the high point of the range is outside the 2-pixel boundary. However, there was just 1 document on which the results were off like that. The next instance in this range lies over a millimeter lower.

Based on these validation results, keeping in mind the sub-optimal conditions under which they were acquired, we conclude that WANDA is backward compatible with FISH.

5 Usability study

The WAM is a technological and ergonomic update on the interactive measuring of handwriting features. The technical part of the WAM is checked by validating its results (see Section 4) and by testing it. The ergonomic part is tested by having actual users use the tool. Since the project aimed at a system that was usable by both experts and novices, the usability study should contain both. Four experts and ten novice users tested the system. The experts were the users of FISH from the BKA (2) and the NICI (2). Their input was invaluable as they were capable of judging WANDA compared to its predecessor FISH. The ten novice users on the other hand had no experience with writer identification or handwriting measurements at all. Also, without a system to compare WANDA to, their views on the system would be from an entirely different angle than those of the experts.

The usability study for the WAM was run over 5 documents per user. For every document the user had to measure 10 instances of every kind of measurement. The exceptions were the two loop measurements, as not all handwriting styles include loops. Loop measurements were only to be taken if enough (5 at least) of a kind were present in the handwriting. The novice users had no previous encounter with either the Wanda system or writer identification in general. They were given the Wanda measurement tool manual to read first. After that they were given a brief instruction in how to use the rest of the Wanda system, so they could focus all their attention on the measurement tool and not be distracted by the preliminary activities. Except for the manual they were not given any further information on the measurement tool. The test subjects were asked to give a list of good and bad points of the WAM. The novice users were also asked if they encountered any difficulties after reading the manual and their results were checked on errors to see if they understood the instructions correctly.

The overall results of the usability test were very encouraging. With only the manual and the instruction window of the measurement tool itself to help them, the novice users were able to quickly start measuring. The number of errors made was low and consistent with errors made through misinterpreting the instructions rather than a confusing interface. Furthermore, after the first document was processed, the rest of the documents were measured considerably faster. On average per user, each document after the first was done in half the time (20 to 30 minutes) it took to measure the starting document (40 minutes to an hour). Moreover, the measure time for the later documents was approximately equal to the time an expert takes to measure it. This indicates that the novice users learned most of how to use the system after measuring a single document. In all fairness, it should be noted that the expert users produced the measurements in a more precise and careful manner.

Two measurements did show errors that are harder to correct. Both the line height measurement and the character width were often measured incorrectly. While part of this can be explained by misreading the instructions, some interesting remarks can be made.

- Line measurement

  The line measurement in its present form appears to be rather confusing for most novice users. It differs from the other measurements in two ways. First, it is the only measurement not done on characters. While this does not make an important difference in the technique that is used to measure it, most users failed to see that a line measurement and a character measurement are basically the same. The resulting error was that users would use the same line over and over to make new line height measurements, something they never did with a similar measurement on a character. It is obvious that the instructions need to contain one or more very clear rules on this subject. However, seeing that it was done wrong by so many users this may not suffice.

The second difference between the line height and other measurements lies in the user being able to measure more than one line height in a single measurement. In practice, this open-ended nature meant that some change had to be made to the normal order of the measurement. In the case of the WAM we chose to keep the moment at which the system start to search for the ink based on the mouse-clicks to the end of the process of a measurement. This works fine with the other measurements, but with the line height this makes it difficult for the user to see if an error was made until it is too late. This is the most obvious user interface design error we were able to find during the usability test.

- Width measurement

The width measurement is a tough measurement, even for experts. The ideal width measurement is taken on letters that contain a cupped form and is performed at the edge of the open end of the cup (see Figure 3b). First of all, this is hard to explain to a novice user. Second, the measurement is rather strict in what is to be measured (as to be compatible with FISH). Many forms of handwriting (especially when it is sloppy and/or cursive) produce forms of the ’n’ and ’u’ (the preferred subjects for a character width measurement) that are not suitable for this measurement. This makes the explanation of the character width to novice users even harder to accomplish. We have not found a satisfactory solution for this problem yet.

We also asked the users if they thought that the WAM worked well overall and whether they found some points that needed improvement. All users agreed that the WAM itself was easy to use and they saw no major problems. As can be expected, number of smaller issues was addressed. The larger part of these concerned typical programming errors that can readily be corrected. The rest of the remarks were small suggestions of improvement and annoyances. These were of a more personal taste (mostly a suggestion was only put forth by one person), but we are still evaluating their impact.

The results of the usability study were overall positive. Both novice users and experts found the system easy to use, clear, and intuitive. The novice users were able to use the system correctly after studying the manual and with the help of the instruction window.

6 Conclusion

WANDA is a technical and ergonomic update of the FISH writer identification system. WANDA incorporates state-of-the-art technologies resulting in an up-to-date, flexible, and open system. This makes it easy to maintain, portable, and easy to extend.

The WAM tool is the part of WANDA that upgrades the interactive feature measurements. The WAM has got a new user-interface that uses color and textual cues to give the user a clear and functional overview of the current state and the possibilities of the tool. It also provides a proof-of-concept for recognition-based measurements in the form of the new allograph measurement. The validation of the

WAM shows that the system yields the proper results and that the interface design is sound and usable by both expert and novice users.

References

[Franke et al., 2003] Franke, K., Schomaker, L.R.B., Veenhuis, C., Taubenheim, C., Guyon, I., Vuurpijl, L.G., van Erp, M. & Zwarts, G. (2003). WANDA: A generic framework applied in forensic handwriting analysis and writer identification. In A. Abraham, M. Köppen, and K. Franke (eds.) *Design and Application of Hybrid Intelligent Systems, Proceedings of the 3rd International Conference on Hybrid Intelligent Systems(HIS '03)* (pp. 927-938). Amsterdam, the Netherlands: IOS Press.

[Franke et al., 2004] Franke, K., Schomaker, L.R.B., Veenhuis, C., Vuurpijl, L.G., van Erp, M. & Guyon, I. (2003). WANDA: A common ground for forensic handwriting examination and writer identification. *ENFHEX news – Bulletin of the European Network of Forensic Handwriting Experts*, 1/04, 23 - 47.

[de Jong et al., 1994] Jong, W. C. de, Kooij, L. N. Kroon van der & Schmidt, D. P. (1994). Computer aided analysis of handwriting, the NIFO-TNO approach. In *4th European Conference for Police and Government Handwriting Experts*.

[Srihari & Leedham, 2003] Sargur Srihari and Graham Leedham (2003), A survey of computer methods in forensic document examination, In *Proceedings of the 11th International Graphonomics Society Conference* (pp. 278-281). Scottsdale, USA.

[LumenIQ, 2004] LumenIQ Homepage (n.d.). Retrieved Febuary 27, 2004, from http://www.lumeniq.com

[Phillip, 1996] Phillip, M. (1996). *Fakten zu FISH, Das Forensische Informations-System Handschriften des Bundeskriminalamtes - Eine Analyse nach über 5 Jahren Wirkbetrieb*. Wiesbaden, Germany: Kriminaltechnisches Institut 53, Bundeskriminalamt.

[Pikaso, 2004] Pikaso Software Inc.'s Write-On website (n.d.) Retrieved Febuary 27, 2004, from http://www.pikaso.com/write-on/main.htm

[Schomaker, et al., 2003] Schomaker, L., Bulacu, M. & Erp, M. Van (2003). Sparse-parametric writer identification using heterogeneous feature groups. In *Proceedings of the International Conference on Image Processing* (pp. 545-548). Barcelona, Spain: IEEE Press.

[Schomaker & Vuurpijl, 2000] Schomaker, L. & Vuurpijl, L. (2000). *Forensic writer identification: A benchmark data set and a comparison of two systems*. Nijmegen, the Netherlands: Nijmegen Institute for Cognition and Information.

[Vuurpijl & Schomaker, 1997] Vuurpijl, L. & Schomaker, L. (1997). Finding structure in diversity: A hierarchical clustering method for the categorization of allographs in handwriting. In *4th International Conference on Document Analysis and Recognition (ICDAR'97)* (pp. 387–393). Ulm, Germany.