Database and Data Management Requirements for Equalization of Contactless Acquired Traces for Forensic Purposes

Stefan Kirst, Martin Schäler

Otto-von-Guericke University of Magdeburg
Universitätsplatz 2
39106 Magdeburg
stefan.kirst@iti.cs.uni-magdeburg.de
martin.schaeler@iti.cs.uni-magdeburg.de

Abstract: The importance of fingerprints and microtraces within the field of criminalistics and forensics is well known. To increases probative values of such evidences several techniques of lifting, enhancement and feature extraction exist. An upcoming field is the contactless acquisition of traces, because the integrity of traces is preserved. Hence, dealing with digital representations of traces is challenging because of the amount of data and their complexity. A further issue from such an acquisition method is the potential presence of perspective distortions, which we already started to deal with in [KCDV12]. Within the scope of a productive use of contactless acquisition methods, pre-processing steps like the equalization come along. In this short motivational paper we want to give a perspective on requirements for an underlying database and database management system to support the methods of [KCDV12] as a potential real-case scenario. Thereby, we point out possible starting points for parallelization potential and frequent queries, especially when using filter masks.

1 Introduction

In criminalistics fingerprints and microtraces are important evidence but the traditional way of capturing them, like the usage of brush and powder, alters the trace. A solution are contactless methods, that assure a non-invasive and repeatable acquisition. However, these processes struggle with distortions due to the perspective of the sensor or non-planar surfaces. Variations of features like the relative position of minutiae, ridgeline distances or length and width of microtraces are possible changes due to perspective distortion. As an outcome, algorithms based on affected features may not work properly. There are non-distorting techniques like the non diffractive element based approach for smooth curved surfaces in [KM09], but those methods are only applicable on specially shaped surfaces. Our recent achievements [KCDV12] in equalizations of such distortions using confocal microscopy[1] demonstrate the resolvability of this challenge. After evaluating our approach

[1]for all scans the confocal laser microscope ”‘Keyence VK-X110”’ [Key] was used; intensity and topography data were used for equalization

on different surfaces and shapes, we accomplished to use our methods on fingerprints and microtraces[2]. Significant reduction of a perspective distortion induced error in equalized images could be presented.
Providing information of detected distortions as well as equalized images we are able to support a forensic investigator's work. Since an equalization may take up to ten minutes[3] a productive application of these techniques in productive scenarios might be exhausting as crime scenes with a single trace seem to be unusual. The delays in all measurements scale when using faster systems, but will not dissolve entirely.
So far our recent work on equalization of non-planar surfaces is concentrated on qualitative results. In this paper aspects of database- and database management systems are motivated using the outlined scenario. Based on experiences of our first exemplary evaluations of our new methods, we highlight the potential and challenges that arise from a database point of view, when it comes to performance issues in the productive use of forensic methods.
After a short summary of our equalization methods we concentrate on the following points of interest:

- potential of parallelization for inter- and intra trace computations
- frequent requests, under the perspective of filter masks

2 Equalization of forensic traces on non-planar surfaces - an application scenario

The following presentation of the basic background is necessary for later explanations on amount of data and suggestions to improve our approach. Besides, the conclusions of this paper rely on our experience on methods for equalization. Hence, a general knowledge of our approach and its capabilities is mandatory.

In [KCDV12], we dealt with perspective distortions as they arise in contactless scans of non-planar or non-perpendicular placed surfaces to preserve topological features. These distortions alter the topology[4] of traces, so that topology based processing methods, like NBIS[5] [WGT^{+}] or ridge density based sex determination [Gun07], might produce unreliable results. The presented approach divides an inhomogenously shaped surface into small quasi planar surfaces. After equalizing them separately, we try to reassemble them according to their previous topology. The slopes are determined by analyzing the topography data in a blockwise manner. We were able to adapt these techniques on fingerprints and microtraces. For evaluation purposes we introduced a landmark based relative error to show the impact of our introduced pre-processing methods. This error describes the distance of fixed points, called landmarks, in comparison to the very same points in an undistorted image. First evaluations of our approach indicate significant reductions of this

[2] a partial human hair, see figure 2 and table 1
[3] equalization of a fingerprint(20 degrees, area: 14,26mm x 9,49mm): 10.30 minutes - using: Intel(R) Core(TM) i7-2670QM @ 3.1GHz, 8GB RAM; javaVM @ 4GB, WDC WD6400BPVT-60HXZ
[4] e.g. ridgeline distance, relative position of minutiae
[5] "'NIST Biometric Image Software"' - see [online]: http://www.nist.gov/itl/iad/ig/nbis.cfm

relative error for all scans after the equalization. Due to reasons of space, we are not able to present more information on the equalization process, as well as on the evaluation procedure, and kindly refer to our previous work [KCDV12].

At first, we applied our approach on different types of surfaces[6] before we used it on combination of traces (see table 1).

type of trace	**test object**	**number of scans**	**area**
latent fingerprint right pointing finger	platter	4 scans (angles: 0, 20, 40, 60)	14266,36µm x 10972,92µm
partial human hair	quasi-planar application on platter	4 scans (angles: 0, 20, 40, 60)	1346,73µm x 4333,62 µm

Table 1: test objects: surface with traces [KCDV12]

The results for the equalization of fingerprints based on a landmark based relative error were significant[7] (see figure 1).

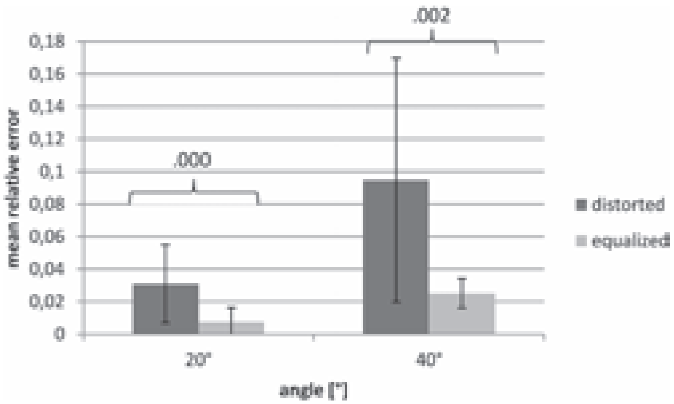

Figure 1: results for the relative error in scans of a fingerprint in different angles - minutiae-based set of landmarks [KCDV12]

The application of our approach on microtraces is promising as well. It should be noted, that all equalizations on this microtrace were calculated using only the surface of the human hair itself, which is a quiet challenging approach regarding to its rough and flaky surface. We were also able to reduce the relative error distinctly for the human hair[8] (see figure 2).

[6]a planar surface in different angles, two curved surfaces, a spherical surface - see [KCDV12]
[7]significances are shown in curly braces over each bar in the diagram
[8]no standard deviation or significance could be calculated

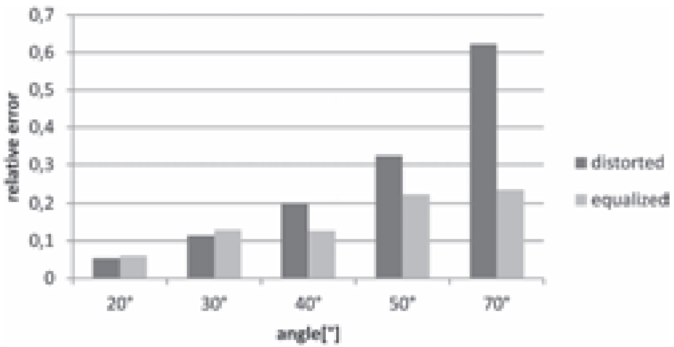

Figure 2: equalization of a human hair on a platter using its length as a metric [KCDV12]

By recovering the topology of surfaces and applied traces we can support an investigators work with additional useful information. Despite long acquisition times[9], the equalization of a surface takes an inconvenient amount of times as well. Hence, to enable productive use of our results we have to address both challenges. In this paper, however, we focus on potential improvements for the equalization process. For more results and an examplary presentation of distorted and equalized images, see [KCDV12].

2.1 Summary of the used concept

First of all, we introduced a definition of types of perspective distortion. These types describe the presence and dimension of perspective distortion starting with no distortion up to irregular distortions in two dimensions[10].
Using this graduation we then formalized the sensors view on a surface with traces by forming tuples of feature classes following the scheme used in [KLD06]. These classes are:

- the surface, the trace is applied on
- the trace pattern, that describes the trace itself
- the gradients of the surface
- and environmental influences

Based on this formalization the three main steps in figure 3 of an equalization were developed.

[9] using the "'Keyence VK-X110, a scan of a fingerprint of the test set of table 1 easily exceeds 20 hours

[10] two dimensional perspective distortions can not be projected into a distortion free representation, since none of the surfaces principal curvatures is zero

Figure 3: equalization pipeline

First the **topography data is filtered** by subtracting a Gaussian filtered version of the topography data to manage outliers and noise, because topography data is the base of our equalization approach. The resulting filter mask covers all useable areas for the following **blockwise determination of gradients**. In each block all pixel to pixel slopes are measured and combined into a block global gradient. The results of all blocks on a coherent homogenously distorted area form a global gradient, that is used for the **rescaling** process **of the original data**.
It should be mentioned that all the investigators decisions still base upon the original unprocessed data. The new equalized versions are only a helpful support to provide a well-founded different view on a trace.

To evaluate the improvement of an equalization we introduced a landmark based relative error to compare distorted and equalized representations of a scan. Therefore, landmarks were placed precisely within a test set. Subsequently, all distances between all landmarks were computed and compared to a planar scan to describe their deviation.

This proceeding requires topography and intensity or color data, which in our case were captured with the confocal laser microscope "‘Keyence VK-X110"’ [Key]. All methods are implemented in Java.

3 Statistics on Equalization and a Potential Productive Use

The following section gives an overview on the amount of data, that is used and created during an equalization. We also look into the calculation times and deduct requirements on database systems and database management systems.
As shown in our previous work, this first evaluation on equalization of non-planar surfaces gives a promising outlook on its potential of topology correction [KCDV12].
Future test sets need to be significantly greater in order to form a general conclusion and to approach to a potential productive use in the field of forensics. For a potential daily use the acceptance of theses methods and contactless fast acquisition of traces are preconditioned.

Such an application of equalization methods for a daily use with more than one crime scene or a crime scene with a high rate of traces might result in a great number of traces. Assuming an emergence of 1000 traces a huge amount of data and process time will be needed. Supporting a forensic investigator and his work, as the main goal of the equalization process, fast and structured access to all traces is needed. The actual needed amount of time for a contactless and perspectively distorted trace to be read and equalized is about minutes. However, based on our contacts with local, federal and international law enforcement agencies, seconds instead of minutes would be practicable instead. Below an outlook is given to motivate techniques in database systems and database management, that support fast file transfers and parallel calculations.

All statistical information used in the following section were documented using VisualVM [VVM][11]. If not mentioned otherwise these statistics were collected on the equalization of 8 consecutive processed combinations of traces (see table 1).

In Table 2, the amounts of data for an equalization of non-planar surfaces with applied traces are presented for the acquired area, the raw data and the resulting maximum heap data for a consecutive processing of the given test set (see table 1).

	minimum	**maximum**
scan area	1342.80µm x 1824.89µm	9390.39µm x 14250.64µm
intensity data	5.4 MB	74 MB
topography data	5.4 MB	74 MB
used heap data	0 MB (starting point)	1.05 GB

Table 2: summary of the minima and maxima of the acquired area, the amounts of used raw and heap data

The size of used raw data[12] for given test set (see table 1) vary from 10,8MB[13] to 148MB[14]. The whole process splits into the import of raw data und the actual equalization calculations. After the import and during determination of distortion and rescaling, the amount of used data easily increases by the factor of 7. In our tests the resulting amount of used heap space during a single equalization rose up to 1.05GB[15]. So the overhead of data are not a few kilobytes, but rather hundreds of megabytes up to gigabytes.

The data we deal with in this case are the raw intensity and topography data in their original form, as well as filtered, subtracted and binarized topography data, resulting filter masks, data about block calculation/ positioning and the rescaled versions of intensity/ topography data.

Despite large acquisition times due to measurement speed of the used equipment, the processing of this data in a daily use seems also impractical as we analyze the CPU times in the following. Table 3 summarizes the maximum, minimum and mean CPU times for the equalization of the given test set(see table 1).

[11] on a system, using: Intel(R) Core(TM) i7-2670QM @ 3.1GHz, 8GB RAM; JavaVM @ 4GB, WDC WD6400BPVT-60HXZ

[12] data size is given in total: sum of size of intensity- and topography data

[13] scan of a partial human hair at 70 degrees (area: 1.35mm x 1,83mm)

[14] complete scan of a fingerprint at 20 degrees (area: 14,26mm x 9,40mm)

[15] value refers to the equalization of a complete scan of a fingerprint at 20 degrees (area: 14,26mm x 9,40mm)

amout of time	equalization	import	computation
maximum	10.30min	9.27min	2.50min
minimum	0.39min	0.28min	0.08min
mean	2.65min	2.07min	0,59min

Table 3: maximum, minimum and mean CPU times: entire equalizations, import of raw data and calculation steps; based on scans of given test set, see table 1

The filter masks for equalizations of microtraces were created in a previous manual step, so that additional data and CPU time might be addable. Since the resulting filter masks for the human hair only cover the small area of the hair itself, the calculation part of the equalization does not take a long time.

The information in the following figure 4 were measured using VisualVM[16] [VVM]. The average CPU time is visualized for hot spot classes.

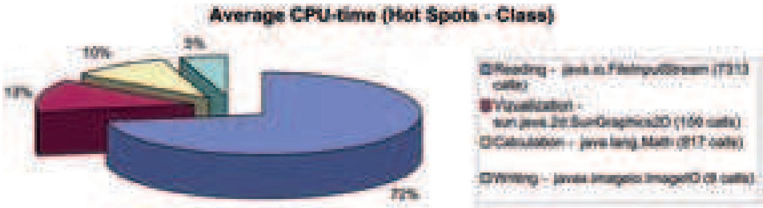

Figure 4: Average CPU-time: Hot Spots - Class

The biggest amount of time is needed to just read original and cached data. Of course the used system is to blame for that, however imagining not only a single processing of a trace surface at a time but several queries, the throughput will increase distinctly, since the majority of calls is also focussed on read operations.

Subsequently, we highlight requirements and starting points for potential improvements for our approach.

3.1 Frequent Data and Filter Masks

As shown above, reading data takes most of the time, so the access on all this data and cached data needs to provided with a high performance. Especially a fast mapping of the filter mask onto the filtered topography data would increase the access to all calculation blocks without a need for a try and error approach for every possible block.
Among other things, Table 3 contains the computation times within an equalization. The equalization of fingerprints takes the most time within our test set, because ridgelines and their potential of a good diffuse reflection creates a big amount of usable topography data.

[16]The average CPU-time was calculated for 8 consecutive equalizations (based on scans of given test set, see table 1)

On the other hand we are forced to use small block sizes (20 to 20 pixels or below) to be able to place them on the thin ridgelines, which significantly increases the number of blocks and thereby the calculation time. This dependency between the number of blocks and the present trace or surface endorses an internal mapping of filter masks on the topography data.

3.2 Potential Improvement of Parallelization

As pointed out in Table 3 already the import of the original intensity and topography data takes a decent amount of time. Hence to that amount, a parallel import of raw data (intensity and topography data) is obvious.

As mentioned above the necessary CPU time depends on the number of used blocks, which itself depends on the amount of usable areas in the topography data[17], the size and resolution of the scanned surface and on the used block size. The bigger blocks are on an evenly shaped part of the acquired surface, the more reliable the result of the determination of gradients will be. Since big consecutive areas in the topography data are rare the block size is limited, which results in a greater number of blocks.
Since the blockwise calculation of gradients is done sequentially, more time can be spared by parallelizing their computation. An additional benefit from this approach is that with all valid blocks the largest coverable area of topography can be determined. The number of blocks affects the potential of parallelization, though.

Another aspect of potential parallelization is the parallel equalization of multiple traces. Starting with just a few blocks for every surface to check on their slopes, traces with better quality and less perspective distortions can be processed first to provide an investigator with more reliable results first.

3.3 Forensic Data Treatment

For forensic purposes, it is necessary that no original data may be altered or deleted to preserve its integrity and to assure the confirmability of the whole process. Possible solutions might be the use of forensical file formats like AFF4 [CS10] or the invertible fragile watermarking approach from [SSM^{+}11], which could be used on non-trace areas of the image to obtain the integrity without altering any trace information. This also implies that a forensical secure way of logging and access of data is mandatory. When it comes to logging - all parameters, used functions and the equalization result need to be stored.

[17] in case of quasi-planar applied traces, the slope of the substrate might be used as well; using a HDD platter in the present test set eliminates this option due its high reflection characteristics

4 Summary and outlook

Perspective distortions are one of the main problems of contactless trace acquisition. On the other hand the benefits of integrity-preservation and repeatability make such non-invasive acquisition methods valuable for forensic purposes. When aiming for a productive use of the presented pre-processing methods, adjustments to the underlying architecture are important to provide an investigator with fast and secure additional information within a decent amount of time.

What are the requirements on a database and a database management system to fulfill the requirements of the presented scenario?
In the previous section we gave a perspective on the amounts of data and used processing time, that are necessary for the equalization of non-planar surfaces in order to deal with perspective distortions. Especially matching of data when using filter masks or addressing calculation blocks is a important aspect. Furthermore the potential of parallelization for blockwise gradient determination and parallel equalization of multiple traces were motivated to be supported by the database and database management system. Not least the storage of all used data and logging of all processing steps and parameters is mandatory to cope with the requirements of a forensic investigation. That includes integrity and authenticity issues as well as further security aspects.

Acknowledgments

The work in this paper has been funded in part by the German Federal Ministry of Education and Science (BMBF) through the Research Program under Contract No. FKZ:13N10818, FKZ:13N10817 and FKZ:13N10816.

References

[CS10] Michael Cohen and Bradley Schatz. Hash based disk imaging using AFF4. *Digit. Investig.*, 7:S121–S128, August 2010.

[Gun07] Dr. S. Gundagin. Sex Determination from Fingerprint Ridge Density. *Internet Journal of Medical Update*, 2007.

[KCDV12] Stefan Kirst, Eric Clausing, Jana Dittmann, and Claus Vielhauer. A first approach to the detection and equalization of distorted latent fingerprints and microtraces on non-planar surfaces with confocal laser microscopy. pages 85460A–85460A–12, 2012.

[Key] Keyence. Keyence VK-X110. [online] http://www.keyence.de/products/microscope/laser/vkx100_200/vkx100_200_specifications_1.php, last seen: 02.07.2012.

[KLD06] Stefan Kiltz, Andreas Lang, and Jana Dittmann. Klassifizierung der Eigenschaften von Trojanischen Pferden. D-A-CH Security 2006, Düsseldorf, Deutschland, Syssec, ISBN: 3-00-018166-0, Patrick Horster (Eds.), 2006.

[KM09] Peiponen K.-E. Kuivalainen, K. and K. Myller. Application of a diffractive element-based sensor for detection of latent fingerprints from a curved smooth surface. In *Measurement Science and Technology 20(7),077002*, 2009.

[SSM+11] Martin Schäler, Sandro Schulze, Ronny Merkel, Gunter Saake, and Jana Dittmann. Reliable provenance information for multimedia data using invertible fragile watermarks. In *Proceedings of the 28th British national conference on Advances in databases*, BNCOD'11, pages 3–17, Berlin, Heidelberg, 2011. Springer-Verlag.

[VVM] VisualVM Version 1.3.5. [online] http://visualvm.java.net last seen: 12.11.2012.

[WGT+] Craig I. Watson, Michael D. Garris, Elham Tabassi, Charles L. Wilson, R. Michael Mccabe, Stanley Janet, and Kenneth Ko. User's Guide to NIST Biometric Image Software (NBIS).