A First Public Research Collection of High-Resolution Latent Fingerprint Time Series for Short- and Long-Term Print Age Estimation

Ronny Merkel, *Student Member, IEEE*, Jana Dittmann, *Member, IEEE*, and Claus Vielhauer

Abstract—The creation of publicly available image databases for the signal processing community is a very time-consuming, yet immensely valuable task, enabling scientific progress by providing the opportunity of an objective comparison and reproduction of results. This paper presents for the first time a public research collection of high-resolution latent fingerprint time series for age estimation, captured from a pool of 116 different test subjects. It comprises ten different sets with a total of 2,618 time series (117,384 scans), varying between capturing devices (CWL and CLSM), data types (intensity versus topography), aging periods (short-term aging: 24 h, long-term aging: 0.5 - 3 years) and resolutions (1,270 - 180,142 ppi). Most series are annotated with donor information (age and gender) and capturing conditions (scan parameters, ambient temperature, and humidity). The data are anonymized (using partial prints only) and an organizational revocation mechanism is included to assure non-identifiability of donors in the future. Baseline results for age estimation on all ten sets are provided in the form of correlation coefficients and machine-learning based age estimation (kappa), using 19 features from prior feature spaces as well as new ones (Tamura contrast, Benford's law, and improved dust feature). Classification results exhibit kappa values between 0.51 and 0.85, highlighting the progress made in this very challenging area in recent years and also emphasizing the need of future studies on the issue.

Index Terms—Latent fingerprints, public research collection, age estimation, digitized forensics, computer forensics, fingerprint processing pipeline, baseline performance.

I. INTRODUCTION

AGE estimation is a very important issue in crime scene forensics. In contrast to biometrics, where age is often referred to as the biological age of an enrolled human, the age of crime scene traces refers to the time which has passed between placement of a trace and its capture by the authorities. It is vital for forensic investigators to determine such age, linking a trace to the time of a crime. Especially for latent fingerprints, which are widely used as evidence in criminal investigations, suspects often claim to have been at a crime scene prior to or after the crime [1]. Only if a latent print is able to link an individual to the time of a crime in addition to linking it to the place of the crime, the evidentiary value of the print can be considered strong in court. This challenge of latent print age estimation has been known to forensic investigators for more than 80 years but could not be adequately addressed so far. Recently, non-invasive, high-resolution capturing devices have been introduced to the domain of capturing crime scene traces. In contrast to the classical process of physical or chemical print enhancement and lifting, these techniques do not alter a latent print, therefore allowing for the first time to consecutively capture a single print in regular time intervals and to observe the morphological changes as well as changes in chemical composition and other properties occurring over time. These changes include (among others) the drying of the print (evaporation of water and short-chained fatty acid) as well as a loss in height and structure during degradation (summaries of studies on these degradation processes can be found in [2] and [3]).

Non-invasive, high-resolution capturing devices include Fourier-transform Infrared Spectroscopes capturing chemical print properties, Electric Potential Sensors capturing electrostatic properties as well as optical devices capturing morphological properties such as Coherence Tomography, Chromatic White Light (CWL) sensors and Confocal Laser Scanning Microscopes (CLSM) [2]. However, such non-invasiveness and increased resolution evoke high purchase costs (often higher than €100,000 per device), requiring an intense exchange of datasets between scientific experts to ensure reproducibility and scientific progress. Furthermore, data privacy is an important factor to be considered because the devices can provide very high-resolution print images, which are considered person-related data in most European countries and therefore are subject to specific data protection regulations [4].

The aim of this paper is to provide for the first time a publicly available research collection of high-resolution latent print time series for scientific research (including baseline performances) while at the same time ensuring certain

R. Merkel and J. Dittmann are with the Working Group Multimedia and Security, Department of Computer Science, Otto-von-Guericke-University of Magdeburg, 39016 Magdeburg, Germany (e-mail: merkel@ovgu.de; jana.dittmann@ovgu.de).

C. Vielhauer is with the Department of Informatics and Media, Brandenburg University of Applied Sciences, 14737 Brandenburg an der Havel, Germany (e-mail: claus.vielhauer@th-brandenburg.de).

standards of data privacy protection. It therefore contributes to the IEEE SPS initiative to support open data collections for reproducible research in forensics. It paves the way for organizing challenges and motivates further research in this emerging field. Specific issues to be investigated using the data collected include the following challenges (C1 - C4), of which only the first one is limited to print age estimation:

C1: A very important challenge to be addressed by the provided data collection can be seen in the study and comparison of adequate age estimation features and their limitations such as different statistical, textural or particle-based features as well as dust as a feature. Here, different goals can be investigated, such as age estimation accuracy for different time periods, period lengths and periods with gaps in between.

C2: Another important challenge can be seen in studying the general feasibility of non-invasive, optical CWL and CLSM devices for the analysis of latent prints, which provide very high resolutions and different magnification levels, barely exploited so far in fingerprint forensics. Such feasibility analysis includes the study of sensor artifacts such as light source fluctuations over time, image distortions, aging of the sensor hardware and quality of the captured images. The study of certain particle structures, visible only at microlevel, is furthermore enabled such as the investigation of droplets, dust, external contaminations, crystallizations as well as the microstructure of pores, minutiae and ridge progressions.

C3: The investigation of preprocessing and quality enhancement methods for very high-resolution print images is another important challenge. Within this challenge, differences in intensity and topography print images have to be studied, exploring if topographic images are in general suitable for latent print analysis and can provide additional information, also combining both image types for an increase of the overall available print information. Furthermore, methods for artifact reduction, image quality enhancement, segmentation of certain structures as well as dust reduction can be studied. In that scope, images can also be artificially distorted, e.g. using the StirTrace framework [5], performing noise addition, cropping, rescaling, shifting and pixel-depth reduction to evaluate the impact of these distortions on the image quality and preprocessing. Moreover, the investigation of the impact of lossy compression on the image quality and segmentation performance is enabled. Consecutive temporal images can furthermore be combined in the time domain to improve the overall image quality for a future print detection on challenging surfaces, as suggested in [6] using the degree of persistence.

C4: Studying the impact of external factors on latent prints is another very important challenge, which can be addressed using the specified meta data of the provided collection. Such external factors include environmental influences (such as temperature, humidity, different indoor capturing locations) as well as personal influences (such as donor age and sex). They can be studied in respect to age estimation performance but also in respect to general print visibility and drying behavior, which are also relevant for classical development methods.

In this manuscript, the main focus is on challenge C1, studying the age estimation of latent prints. The overall contribution

of this paper can be summarized as follows:

- A collection of ten different sets of high-resolution latent fingerprint time series (containing 2,618 series comprised of 117,384 partial scans) is compiled to provide a research base for the scientific community. The collection is based on and extended from several smaller studies of [2], [7]–[10], and [11], conducted in the scope of the Digi-Dak [12] project between 2010 and 2015.
- A specific contribution of this paper lies in the aggregation, extension and anonymization of the captured data to ensure privacy protection of the high-quality prints while at the same time allowing the exchange with other researchers. While time series comprising non-anonymous print images have been excluded from the set to adhere to European data privacy regulations, revocation mechanisms have been included on an organizational level for the prints to be published, in case future findings (e.g. identification based on pores) require revocation.
- The collection provides systematic variations of capturing devices (Chromatic White Light CWL sensor [13] vs. Confocal Laser Scanning Microscope CLSM [14]), data types (intensity vs. topography), aging periods (short-term aging of 24 hours vs. long-term aging between 0.5 years and 3 years) and resolutions (between 1,270 ppi and 180,142 ppi; respective dot distances: 20 μm - 0.141 μm).
- Most time series have been annotated with donor information (age and gender) as well as capturing conditions (capturing parameters, ambient temperature / humidity).
- Baseline results are presented for all ten sets in the form of correlation coefficients to mathematical functions (feature level) as well as age estimation accuracies (classification level). Specifically, the prior feature spaces from [2], [7]–[10], and [11] are extended by the Tamura contrast feature as well as ten statistical features based on Benford's Law [15]. Furthermore, the dust feature is improved, achieving significantly better results. In total, 19 different features are applied to the time series. Morphological features (such as ridge thickness, size / amount of pores), which have been shown in [2] and [10] to not be characteristic, have been removed. Classification-level results are extended from short-term aging to long-term periods.

The remainder of this paper is structured as follows: section II gives a short overview over the current state of the art in the field. Section III introduces the ten collected sets of latent print time series and their properties. An evaluation concept based on time series preprocessing, feature computation, age estimation and quality metrics is proposed in section IV, based on which baseline performance results are presented for all collected sets in section V. Section VI concludes the paper and highlights challenges of future research.

II. STATE OF THE ART

Age estimation of latent fingerprints is an important issue to forensic investigations, which has not been adequately addressed for 80 years. From the numerous approaches towards this goal, only a few examples can be provided here.

For a more comprehensive state of the art review, the reader might refer to [2]. A good summary of studies based on chemical print composition is given in [3].

First age estimation approaches include manual print inspection [16] as well as systematic aging experiments under reproduced crime scene conditions such as [1] and [17]. However, their evidential validity is questionable, since it is highly unlikely that all influence factors can be exactly reproduced [18]. Others have tried to study the print visibility over time using physical print development techniques, e.g. [19] and [20]. Furthermore, an Electrostatic Detection Apparatus (ESDA) [21] as well as healing stages of wounds on fingers [22] were investigated. Print luminescence behavior was examined, e.g. in [23] and [24]. Numerous studies have evaluated print degradation using chemical methods such as chromatography and mass spectrometry, e.g. [25]–[27]. Almost all early aging studies have been invasive, altering a print during inspection (e.g. during physical development or dissolving of prints) and therefore did not allow for a systematic study of degradation processes of single prints using time series. None of them was able to provide objective results concerning age estimation performance and error rates.

In recent years, a few preliminary studies have evaluated print aging behavior based on contactless and non-invasive acquisition devices in the chemical [28], [29], luminescence [24], autofluorescence [30] and electrostatic [31] domain, including one police study [32] using cameras and microscopes. Amongst these first contactless approaches are the studies of the Digi-Dak project [2], [7]–[10], and [11], which are based on time series captured with non-invasive, optical Chromatic White Light (CWL) sensors and Confocal Laser Scanning Microscopes (CLSM). They have for the first time provided objective performance measures and error rates for age estimation performance, using a completely digital processing chain. However, no database of fingerprint time series is available at this point in time for the signal processing community, which would allow for extensive and reproducible studies without the significant effort of data collection over several years.

This paper presents a combined collection of ten different sets of latent fingerprint time series captured during the Digi-Dak project research. It is novel in its type (time series instead of single images, considering print age), types of capturing devices (intensity and topography images of CWL and CLSM sensors) as well as its property of very high resolutions (between 1,270 ppi and 180,142 ppi) in comparison to publicly available databases of single latent prints such as [33]. However, the very high resolution of such prints requires specific measures to ensure data privacy protection as well as revocability of the database in the light of potential new identification techniques in the future.

III. DESIGN OF THE DATA COLLECTION

In total, the data collection consists of ten sets of latent print time series, which are varying mainly in respect to the capturing device, aging period as well as resolution and measured area size. An overview over the complete collection can be found in TABLE I. Example print images are given in Fig. 1.

A. General Collection Properties

All prints included in this collection are captured from hard disk platters, which are round disks from hard drives, exhibiting very good reflection properties. They represent the surface type of smooth, planar, non-porous and well-reflecting substrates such as mirrors, glass, displays, credit cards, etc. Other surface types should be studied in future work. In general, the here presented age estimation framework can be similarly applied to them if sufficient contrast between fingerprint and surface can be achieved and if additional distortions (e.g. resulting from structured or curved surfaces) can be removed by additional preprocessing methods or alternative sensors.

Prints were captured from a pool of 116 donors, where each donor contributed only once to each set but often participated in different sets. Although this procedure leads to a substantial variety of different donors for the created sets, the preliminary study of [2] also indicates that the inter-donor variation of latent print degradation speed is only slightly higher than its corresponding intra-donor variation (using feature curve slopes of selected features as print degradation speed). Therefore, intra-donor time series can be considered as being almost equally feasible for age estimation studies as inter-donor series. No pre-treatment of the fingers of donors is enforced (e.g. washing, rubbing the forehead, etc.) and the print application is not controlled to achieve realistic results. A specific selection of fingers is not conducted. In some cases (e.g. for set S1 and set S2), all ten fingers are used from each donor while in other cases, index fingers are often preferred (see section III-B).

The capturing environment is fixed to office locations of normal use representing an indoor scenario and avoiding extreme situations such as very controlled conditions in climatic chambers or the comparatively arbitrary conditions of outdoor environments. For long-term aging, prints are stored in cupboards or boxes. During capture they are (sometimes for several days) exposed to dust and dirt occurring in the office environments. Temperature and humidity vary according to seasonal changes as well as day and night differences. However, during the capture of prints, changes are usually comparatively small and limited to few degrees Celsius in temperature as well as few percentage points of the relative humidity.

For the CWL capturing technique, three similar devices of type FRT MicroProf200 CWL 600 [13] have been used in three non-controlled office environments LocA - LocC, located at the University of Magdeburg, the University of Brandenburg and the Saxony-Anhalt State Office of Criminal Investigation (Germany). Reflected light intensity (pixel depth: 10 bit) as well as topographic height information in meters (pixel depth: 16 bit) have been captured. For the CLSM technique, a single device of type Keyence VK-X110 Confocal Laser Scanning Microscope [14] has been used in a single location capturing the reflected light intensity (pixel depth: 16 bit) as well as topographic information in meters (32 bit). Capture Periods are set to 24 hours for the short-term aging. For long-term

TABLE I
ACQUISITION PARAMETERS OF THE INTRODUCED COLLECTION OF LATENT PRINT TIME SERIES IN RESPECT TO PRINT
APPLICATION, CAPTURING DEVICE AND AMBIENT ENVIRONMENTAL CONDITIONS

Set	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Device	CWL	CLSM	CWL	CWL	CWL	CWL	CLSM	CLSM	CLSM	CLSM
Type of Aging	STA	STA	LTA	LTA	LTA	LTA	LTA	LTA	LTA	LTA
Capture Period	24h	24h	2.5y - 3y	2.5y - 3y	1.7y	1.7y	0.5y	0.5y	0.5y	0.5y
Capture Interval	1h	1h	1w	1w	1w	1w	1w	1w	1w	1w
Print Area (mm)	4x4	1.3x1	5x5	2x2	3.4x3.4	3.4x3.4	1.2x0.9	0.64x0.46	0.225x0.155	0.116x0.080
Pixels	200x200	1024x768	500x500	1000x1000	170x170	170x170	924x668	924x668	824x568	824x568
Dot Distance (μm)	20	1.3	10	2	20	20	1.3	0.69	0.273	0.141
Amount of Series	550	639	20	20	16	16*79	12	12	12	12
Donors	62	64	19	19	16	16	12	12	12	12
Gender (m,f)	361,189	499,140	17,3	17,3	7,9	7,9	10,2	10,2	10,2	10,2
Age (min-max,n/a)	21-63,15	22-83,0	22-45,0	22-45,0	27-55,2	27-55,2	24-46,0	24-46,0	24-46,0	24-46,0
Location (A,B,C)	161,183,206	639,0,0	11,9,0	11,9,0	0,0,16	0,0,16	12,0,0	12,0,0	12,0,0	12,0,0
Temperature (°C)	20.6-30.0	n/a	18.2-32.9	18.4-33.0	22.5-30.9	22.5-31.9	23.9-32.0	23.8-31.9	23.7-32.6	24.0-33.4
Rel. Humidity (%)	28.9-60.0	n/a	25.0-60.0	24.0-59.0	8.1-52.3	6.3-52.3	9.0-36.0	9.0-37.0	9.0-37.0	10.0-38.0
Pub. Images	49	60	7	7	1	1	3	3	3	3

CWL = Chromatic White Light Sensor, CLSM = Confocal Laser Scanning Microscope, STA = Short-term aging, LTA = Long-term aging, h = hours, w = weeks, y = years, A/B/C = Location pseudonym for three non-controlled office environments (University of Magdeburg, University of Brandenburg, Saxony-Anhalt State Office of Criminal Investigation), m = male, f = female, min = minimum, max = maximum, n/a = not available (age information is not provided by the donors in a few cases), Pub. Images = Images available for visual publication (permission granted by donors).
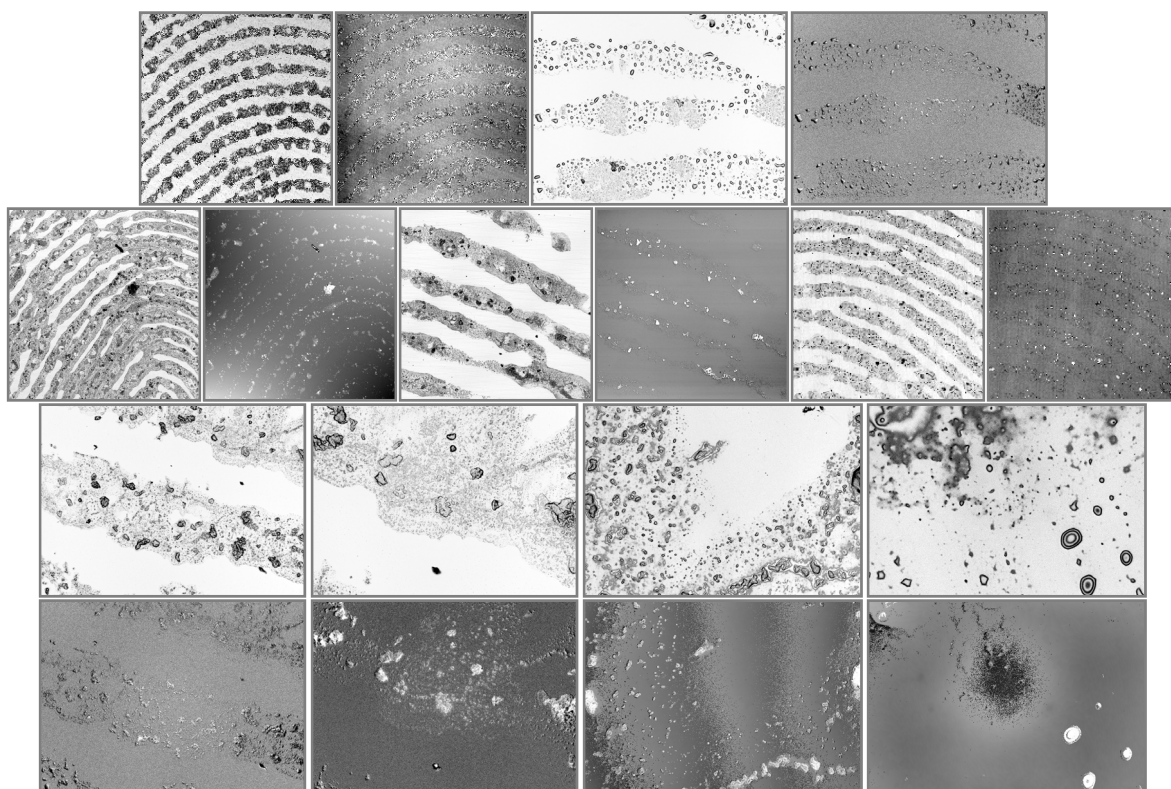

Fig. 1. Examples of (fresh) images of the introduced collection of latent print time series (from left to right). First row: intensity and corresponding topography image of sets S1 and S2. Second row: intensity and corresponding topography image of sets S3 - S5 (S6 is based on the same images as S5, considering several consecutive scans for each point in time). Third row: intensity images of sets S7 - S10 (using the following magnification lenses: 10x, 20x, 50x, 100x). Fourth row: corresponding topography images of sets S7 - S10.

studies, the maximum available period is used, ranging from half a year up to three years. For some sets (e.g. S3, S4), different capturing locations lead to different lengths of the time series, which can vary up to six months.

Capturing intervals are set to 1 hour for short-term aging studies and 1 week for long-term studies. In the latter case,

the first image is captured after several hours or days to exclude short-term aging effects (e.g. drying of the print). The capture periods and intervals of TABLE I are approximations only, the exact absolute age of each print image is specified as ground-truth in the meta data file attached to each individual time series, including donor properties (e.g. age,

gender) and ambient environmental conditions (e.g. temperature, humidity).

B. Set-Specific Properties

Besides the general collection properties and settings specified in TABLE I, the ten sets of time series S1 - S10 have the following set-specific properties:

S1: The first set contains a comparatively large number of time series, captured with the CWL sensor in the scope of short-term aging. It encompassed a certain variety of donors and ages as well as three different capturing locations and is therefore adequate for a general assessment of short-term age estimation performance. Because S1 is a short-term aging set, dust is not an issue, it does not settle on the prints in significant amounts during this short capture period.

S2: The second set can be seen as the CLSM counterpart to S1, providing data of similar amount and variety for the CLSM short-term aging. In opposition to S1, only one device (and therefore only one capturing location) is available. Furthermore, a significantly higher resolution and a smaller print area are used, which is a result of the technical properties of the CLSM device (the provided resolution is the smallest available, using a 10x magnification lens). Similar to S1, dust is not an issue because of the short aging period.

S3: Beginning with S3, all remaining sets are investigating the long-term aging of latent prints. This also means that a print has been drying for several hours or days before its first capture and that dust needs to be considered (except for S6). S3 uses the CWL device and comprises capture periods of up to 3 years, which is a very long time period suitable for adequate investigations of slow changes. Like all long-term aging sets, a print is captured once a week. Because of the significant effort of capturing a print over such a long time-period, only 20 prints are contained in this set (still rendering it the largest available long-term aging set). Moreover, two different locations are present, which is important for the investigation of dust accumulation over time, potentially being location-dependent.

S4: The fourth set is captured from the same latent prints as S3, therefore exhibiting a similar distribution of gender, age, location and ambient climatic conditions. However, a much smaller area is captured in much higher detail (dot distance: 2 μm). Because of this smaller area, a shift in the sensor offset after the first two measurements (measurement error potentially caused by device maintenance) could not be removed. Therefore, the first and second image of a series exhibit a significant shift, whereas all following images are properly aligned. These first two images might still be usable for statistical features (which are often independent of a specific capturing area within the same print) and are therefore also included into the database.

S5: The fifth set covers a shorter long-term aging period of 1.7 years using the CWL sensor. Its difference to S3 and S4 can be seen in the lower resolution (a dot distance of 20 μm is used) and its capture at location C, which is different from the other two sets. It is created by combining the first image of each short-term group of S6.

S6: This set is created from the same samples as S5, hence having a similar gender, age, location and climatic condition as well as similar capturing parameters. Its specialty lies in the capturing protocol for each point in time capturing not only a single print image per week but a consecutive series of 16 images in intervals of one hour. Therefore, S6 is comprised of time series of time series where for each capturing point 16 consecutive images are taken. These consecutive images for each point in time are created to approximate the degradation speed of the print in a very narrow time interval of a few hours to be used as feature value.

S7: The last four sets are created to study the long-term aging using the CLSM device. When using such device, even the smallest magnification of S7 (applying a 10x magnification lens) leads to a higher resolution than the maximum CWL resolution (a CLSM dot distance of 1.3 μm in comparison to the highest CWL dot distance of 2 μ m). The CLSM device is only available for location A and therefore studies are limited to this location. Twelve time series are available over a total capture period of six months.

S8 – S10: The last three sets are similar to S7, only differing in the used magnification, applying 20x, 50x and 100x magnification lenses, which lead to a significantly increased resolution and a very tiny captured print area. Dust is a very important challenge here because a single dust particle can cover large areas of a captured image (see also Fig. 8).

C. Data Privacy Protection and Anonymity

Data privacy protection is an important issue in the European Union. Especially when processing various latent prints with very high capturing resolutions, identification of a print donor can easily be achieved with adequate reference databases. Therefore, making a collection of such prints openly available to researchers would be a clear violation of the data privacy regulations of the European Union [4].

To address this issue, anonymity criteria have been developed in cooperation with the project group "Constitutionally Compatible Technology Design (Provet)" from the University of Kassel (Germany) [34]. According to the report, a partial print image can be considered anonymous if the following requirements are met:

- Not more than two partial images are taken from a single print.
- Each partial image has a maximum size of 5×5 mm.
- In case the pore structure is not clearly visible: the partial image may not contain more than five minutiae.
- In case the pore structure is clearly visible: the partial image may not contain more than three minutiae.
- No special features (e.g. scars) should be visible.

The here presented collection has been designed to only include anonymous fingerprints according to the above summarized specifications of [34]. Minutiae of all images have been manually inspected to adhere to the respective requirements. Therefore, all ten sets are considered as anonymous (no identification can be performed), according to the current state of the art in criminal police services. The requirement of all images covering a maximum print
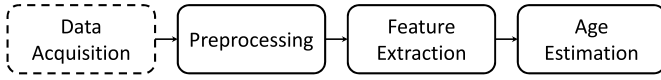
Fig. 2.   General age estimation pipeline from [2].

area of 5 x 5 mm is not an issue for age estimation because preliminary investigations have indicated that a comparatively small area of the print is sufficient for degradation studies [2].

However, when considering future developments in fingerprint identification techniques, there is a realistic chance that at some point in the future it will be possible to perform print identification using the pore structure. In such case, the here presented time series collection might lose its anonymity. Therefore, a mechanism is in place to assure the revocability of the complete database, as described in Supplementary I.

D. Data Format and Spatial Alignment

All print images are stored in a simple binary format, consisting of 4-Byte blocks. The first two blocks of each file denote the height and width of the image followed by the raw pixel data. Intensity and topography series are separately stored. For each time series, a text file provides the properties of the series separated by comma (Comma Separated Value - CSV) such as capture location and date, capture start and end time as well as duration, absolute print age, capturing device and scan parameters for each image. Moreover, temperature and humidity during capture as well as gender and age of the donor are provided.

Prints have been subject to a spatial alignment using pixel correlation maximization, which is mainly important for long-term aging series. While short-term aging prints were fixed only once onto the measurement table and have then been captured consecutively for 24 hours without removal, long-term aging samples were refitted for each measurement, therefore requiring an additional spatial alignment. No further preprocessing is performed to provide print data as close as possible to its original form.

IV. BASELINE PERFORMANCE EVALUATION CONCEPT

To evaluate the performance of the here presented print time series collection, the general age estimation pipeline from [2] is suggested Fig. 2).

All time series are processed using this general pipeline, which can provide a base for adaptability and comparability of the here conducted baseline results in future scientific studies. The data acquisition step has been described in section III for all ten sets of time series. Example histograms showing the gray value distribution of the acquired print images are depicted in Fig. 3. For topography images, print pixels can manifest themselves at both ends of the background pixel distribution (especially for CWL images), because the electromagnetic waves of the optical capturing devices are in some cases directly reflected (e.g. at solid particles, leading to topographic pixels higher than the surface) and in other cases scattered (e.g. at water or fat droplets, leading to topographic pixels lower than the surface).
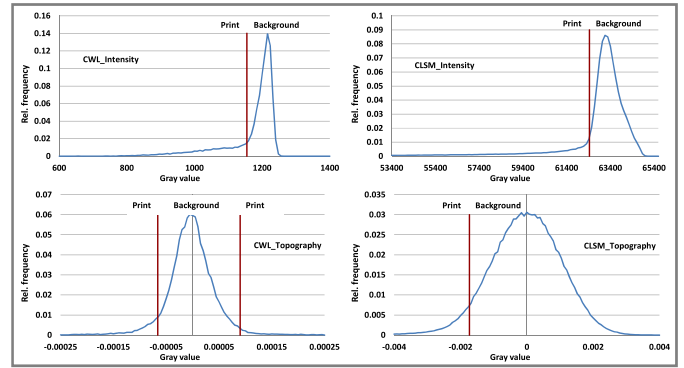

Fig. 3.   Selected histograms showing the gray value distribution of an acquired CWL-intensity (upper left), CWL-topography (lower left), CLSM-intensity (upper right) and CLSM-topography (lower right) image (the border between the partially overlapping fingerprint and background pixels is roughly estimated using vertical lines). For topography images, histograms are depicted after best-fit plane subtraction (for raw topography images before best-fit plane subtraction please refer to Supplementary II).
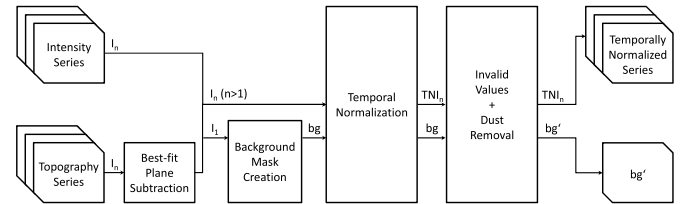

Fig. 4.   Proposed preprocessing steps adapted from [2].

In the remainder of this section, the proposed concept for the other three steps is introduced, including the preprocessing of time series, feature extraction and age estimation. They are based on the earlier work of [2] and considered as a first suggestion. The main aim of this work is to allow other researchers from the signal processing community to improve these processing steps in future studies, addressing the earlier indentified challenges C1 - C4. For such practical comparison, formal quality metrics are proposed and corresponding baseline performances provided.

A. Time Series Preprocessing and Segmentation

Before feature computation, the time series of each set require preprocessing and segmentation steps, which are visualized in Fig. 4. All sets are processed using these steps. For set S6, which consists of separate time sub-series for each point in time n (see also section III-B), the procedure is applied to each sub-series.

In contrast to intensity images, topography data is in the first step planarized using best-fit plane subtraction with least squares approximation [35], [36]. This is necessary to reduce tiny angles in the substrate during measurement table fixation (e.g. by an underlying dirt particle). Because print compounds are very small in their height, small angles during fixation can significantly distort the results (see Fig. 5), which is not an issue for the intensity images, capturing only the reflected intensity and no height information. The best-fit plane subtraction leads to a shift of the topography image pixels, setting the mean height value to zero.
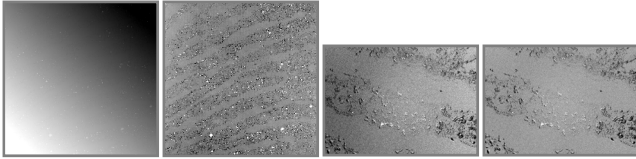
Fig. 5.   Example topography images of set S1 (left) and set S7 (right) before (first and third) and after (second and fourth) best-fit plane subtraction.
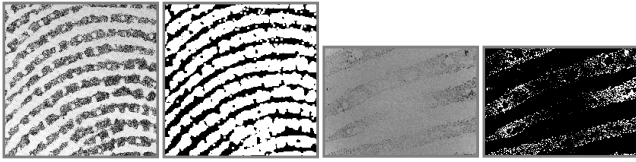

Fig. 6.   Examples of original images (first and third) and corresponding background masks (second and fourth, background depicted in black) of set S1 (left, CWL intensity image) and set S2 (right, CLSM topography image).

In the next step, a background mask *bg* is computed from the first image I_1 of a time series with $1 \leq n \leq n\_max$ consecutive samples. The mask is used to differentiate fingerprint pixels from background pixels throughout the time series. For that purpose, an image I_1 is binarized using data type dependent procedures and thresholds *thresh*:

$$bin = Binarize(I_1, thresh) \tag{1}$$

For intensity images, the Huang method [37] (CWL) and the Triangle method [38] (CLSM) are used for an automated determination of *thresh* (selected histograms of such binarization are given in Supplementary III). Because of their much higher level of noise, topography images are subject to a special print enhancement prior to binarization. CWL topography images are subject to a convolution with a 3 x 3 filter kernel, computing the variance of their neighborhood, which exhibits highest values for small particles and edges as present in prints. It allows for a subsequent binarization with the fixed threshold value of $thresh = 1e^{-9}$. CLSM topography images are convoluted with a 5 x 5 Gaussian blur kernel (sigma = 10) to remove distortions and are subsequently binarized with a fixed threshold value of $thresh = -6.5e^{-4}$ (selected histograms of such binarization are given in Supplementary IV). Such binarization procedures have been found in [2] to perform best out of a variety of techniques. For set S4, preprocessing is started with image I_3 in the here conducted baseline performance evaluation (I_3 is also used for background mask creation) since I_1 and I_2 were subject to significant shifts in the measured area (see also section III-B). The binarized images are furthermore dilated one time to compensate for small binarization errors, increasing the foreground area by one neighboring pixel:

$$bg = dilate(bin, 1) \tag{2}$$

Example background masks are depicted in Fig. 6. It can be observed that the bound around print pixels is comparatively large for CWL images due to their significantly larger pixel size. Additional studies not using dilation operations should be performed in the future (e.g. in the scope of challenge C3) to exactly determine the influence of the dilation operation.
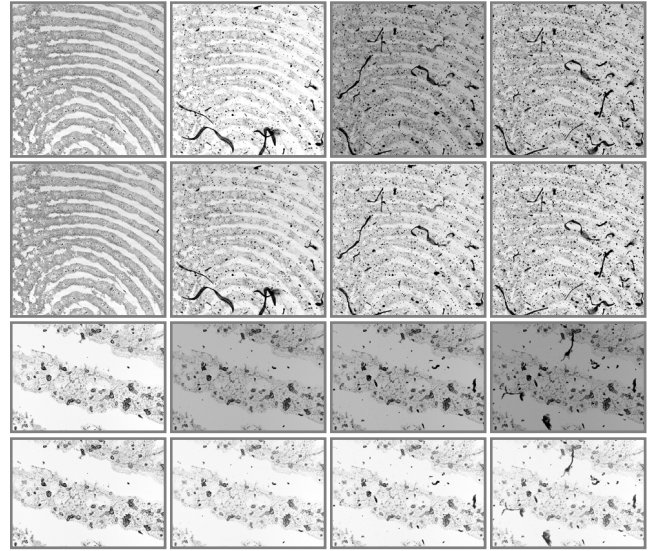

Fig. 7.   Example intensity print time series before and after temporal normalization. First row: example series of set S3 (CWL) with ages 0, 1, 2 and 3 years (left to right); second row: same series after temporal normalization; third row: example series of set S7 (CLSM) with ages 0, 0.2, 0.4, 0.6 years (left to right); fourth row: same series after temporal normalization.

After background mask creation, a print time series is subject to temporal normalization, which reduces differences in the overall brightness of its images. Such differences are caused by fluctuations in the intensity of the light source during continuous service or other distortions. They are significantly higher for the CLSM device in comparison to the CWL sensor. The main idea is to compute the average change in brightness between consecutive images using the non-aging background pixels and subtracting this difference from the complete image to remove light source related changes within the print particles from aging-related changes. This is realized by first determining all background pixels of an image I_n using the background mask *bg*. Sometimes, the background pixels are not completely normal distributed and have a longer tail on one side (caused by few fingerprint pixels being selected together with the background). They are removed by an exclusion of pixels outside the range $[\mu - 2.56 \cdot \sigma; \mu + 2.56 \cdot \sigma]$, repeated 20 times. From the resulting normal distribution of background pixels, the mean value μbg is used for temporal normalization, where a temporally normalized image TNI_n is computed by adding to each image pixel $I_n(x,y)$ the difference of μbg_1 and μbg_n, normalizing the overall brightness of the consecutive images of a time series:

$$TNI_n(x, y) = I_n(x, y) + \mu bg_1 - \mu bg_n \tag{3}$$

Example print time series before and after temporal normalization are depicted in Fig. 7.

While several artifacts from ambient environmental conditions as well as capturing devices are reduced or excluded by background masking using *bg* and temporal normalization, invalid values and dust remain an influence to long-term aging series. Invalid values only occur for the CWL device and mark few pixels where no valid intensity could be retrieved by the capturing device. Dust accumulates over long-time

aging periods and potentially distorts aging tendencies (e.g. if print particles are covered by dust particles). Both distortions are therefore removed by creating a cumulative distortion map *cdm*. Such map is computed from intensity images only (because they are of much higher quality and less distorted than their topography counterparts) and then equally applied to both data types of a time series.

In general, dust reflects less light than fingerprint particles or substrate background in intensity images, therefore manifesting itself at the lower end of a captured image histogram. To retrieve the dust particles of a temporally normalized image TNI_n, a threshold *thresh_m* is manually determined for each long-term aging set S3 - S5 (*thresh_m(S3) = 700, thresh_m(S4) = 640, thresh_m(S5) = 774*) and S7 - S10 (*thresh_m(S7) = 33,000, thresh_m(S8) = 48,000, thresh_m(S9) = 36,500, thresh_m(S10) = 15,000*) by manual inspection of its time series (selected histograms of such binarization are given in Supplementary V). These thresholds are used for binarizing all series of a set, segmenting dust particles (in the future, automated dust detection methods should be studied, e.g. in the scope of challenge C2). With such binarization, the resulting binary image referred to as distortion mask dm_n contains most parts of the dust accumulated in the image. The threshold *tresh_m* is chosen in a way to include as much dust particles as possible, at the cost of also including a lot of fingerprint particles in some cases. However, because only few fingerprint pixels are required for age estimation, it does not seem to negatively influence the age estimation performance if a certain amount of print pixels is removed together with the dust. For the CWL long-term aging series S3 - S5, invalid values reported by the CWL device are also included into the distortion mask dm_n (for S6, dust and invalid value removal is not necessary because only the sub-series are preprocessed which exhibit negligible amounts of dust settlement and invalid values). The cumulative distortion map *cdm(x,y)* is then computed for each pixel location (x,y) of a time series by checking each distortion map dm_n ($1 \leq n \leq n\_max$) of the series if a distortion has occurred at that particular location:

$$cdm(x, y) = \begin{cases} 0, & \sum_{n=1}^{n\_max} dm_n(x, y) = 0 \\ 1, & otherwise \end{cases} \quad (4)$$

Therefore, a *cdm(x,y)* value of zero represents a pixel location which has not been subject to dust or invalid values during the complete time series, whereas all other pixel locations are marked as distorted and are therefore not considered for feature computation. To exclude these distortions and their border regions, the cumulative distortion map is dilated one time (for CWL images) or two times (for CLSM images, exhibiting higher resolutions) and added to the background mask *bg* (marking distorted pixels as belonging to the background):

$$bg'(x, y) = bg(x, y) + Dilate(cdm(x, y), a)$$
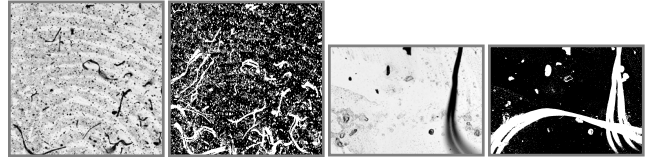$$a = 1(CWL), a = 2(CLSM) \quad (5)$$

Fig. 8. Examples of original images (first and third) and corresponding cdm masks (second and fourth) of set S3 (left, single CWL intensity image of age 3 years and cumulative distortion mask) and set S8 (right, single CLSM intensity image of age 0.4 years and cumulative distortion mask).
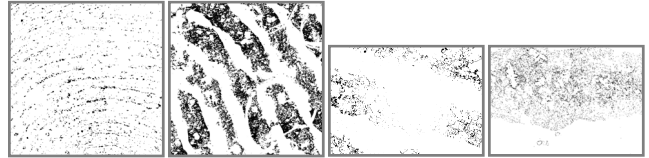

Fig. 9. Example print images after masking with bg' (left to right): example intensity images of series S3 and S4 (CWL), example topography image of series S7 (CLSM), example intensity image of series S8 (CLSM).

Example images of *cdm* are depicted in Fig. 8, print images before and after masking with *bg'* are depicted in Fig. 9.

B. Feature Computation

Within the scope of this paper, 19 different features are computed to provide a baseline performance for each set (see TABLE II for a formalization). They include the previously investigated features of Binary Pixels (F1, capturing the relative frequency of background pixels in a binarized image to measure the disappearance of material during print degradation), the mean print gray value (F2) and standard deviation (F3), the mean variance (F4) and gradients (F5) of local image regions (targeting local changes, e.g. at droplets or print particles) as well as image roughness (F6) and coherence (F7). The here newly investigated features include the Tamura contrast F8 ([39], to investigate if systematic changes can be observed in the texture during aging) as well as ten different features based on Benford's Law [15] (F9 - F18, representing the relative digit frequency of the first digit (0 - 9) of image pixel gray values). The newly investigated dust feature (F19) improves the earlier computed one of [2] by not measuring the amount of image pixels covered by dust but rather counting the amount of dust particles (reducing distortions introduced by the arbitrary size of a settling dust particle).

Giving justification for selecting certain features is particularly hard for the age estimation challenge. Because the complex influences on the aging process are not understood in full detail at this point in time, finding appropriate features is more like a trial and error approach. The aim is therefore to apply a reverse strategy and to potentially conclude from the properties of the characteristic features to the properties of the changes occurring during print aging. The most promising point of reference can be seen in the loss of contrast and an increase of print brightness over time due to the aging process. Therefore, features are mainly of statistical or textural nature, targeting these changes. Morphological features, such as ridge thickness or size of pores, have been shown in [2] to not be very characteristic and are therefore not included. Others, such

TABLE II

INVESTIGATED FEATURES F1 - F19 FOR THE BASELINE PERFORMANCE EVALUATION OF LATENT PRINT AGING TIME SERIES

Feature	Formalization
F1	$F1(TNI_n, bg') = Mean(Binarize(TNI_n, thresh), bg')$ Relative frequency of non-fingerprint pixels (after binarization) within the print area defined by bg', *thresh* is the threshold determined in the background mask creation step (formula (1), topography images are preprocessed similar to the background mask creation by a Variance filter of *kernel* 3x3 (CWL) or a Gaussian Blur filter of kernel 5x5 and sigma = 10 (CLSM).
F2	$F2(TNI_n, bg') = Mean(TNI_n, bg')$ Global mean of pixel gray values not masked by bg', topo results upscaled
F3	$F3(TNI_n, bg') = StdDev(TNI_n, bg')$ Global standard deviation of pixel gray values not masked by bg', topo results upscaled
F4	$F4(TNI_n, bg', kernel) = Mean(LocVar(TNI_n, kernel), bg')$ Mean local variance of pixel gray values not masked by bg', *kernel*: 3x3 (CWL), 5x5 (CLSM), topo results upscaled
F5	$F5(TNI_n, bg', kernel) = Mean\big(Abs\big(Sobel_{XY}(TNI_n, kernel)\big), bg'\big)$ Mean gradients (x- and y-direction) of pixels not masked by bg', *kernel*: 3x3 (CWL), 5x5 (CLSM), topo results upscaled
F6	$F6(TNI_n, bg') = Mean(Abs(TNI_n(x,y) - Mean(TNI_n)), bg')$ Mean surface roughness of pixel gray values not masked by bg', where the roughness of a pixel value is determined by its absolute difference to the image mean, topo results upscaled
F7	$G_X = Sobel_X(TNI_n, kernel), \ G_Y = Sobel_Y(TNI_n, kernel)$ $F7(TNI_n, bg', kernel) = \dfrac{\sqrt[2]{\big(Mean(G_X \cdot G_X, bg') - Mean(G_Y \cdot G_Y, bg')\big)^2 + 4 \cdot \big(Mean(G_X \cdot G_Y, bg')\big)^2}}{Mean(G_X \cdot G_X, bg') + Mean(G_Y \cdot G_Y, bg')}$ Coherence of pixel gray values not masked by bg', *kernel*: 3x3 (CWL), 5x5 (CLSM)
F8	$F8(TNI_n) = \dfrac{\sigma}{(\alpha_4)^m} \quad \alpha_4 = \dfrac{\mu_4}{\sigma^4} \quad \mu_4 = \dfrac{1}{x_{max} \cdot y_{max}} \sum_{x=1,y=1}^{x_{max}, y_{max}} \left(\dfrac{x - \bar{x}}{\sigma}\right)^4 \quad \sigma^4 = \dfrac{1}{x_{max} \cdot y_{max}} \sum_{x=1,y=1}^{x_{max}, y_{max}} (x - \bar{x})^4 \quad m = \dfrac{1}{4}$ Tamura contrast of complete image (from standard deviation σ and curtosis α_4), m = 0.25 represents human perception
F9 - F18	$F9 + d(TNI_n, bg') = Mean\left(\begin{cases}1, FirstDigit(TNI_n(x,y)) = d \\ 0, \qquad\qquad else\end{cases}, bg'\right) \ d \in \{0,1,...,9\}$ Relative frequency of pixel gray values (not masked by bg') with a first (highest) digit d ∈ {0,1,2,3,4,5,6,7,8,9} (Benford's Law)
F19	$F19(dp_{cum\_n}) = BlobCount\big(BlobFilter(dp_{cum\_n}, k)\big) \quad dp_{cum\_n} = \begin{cases} dp_n, & n = 1 \\ dp_n + dp_{n-1}, & n > 1 \end{cases}$ $dp_n = Dilate(BlobFilter(Binarize(TNI_n, thresh_{dust}), k), p)$ Amount of dust particles (*BlobCount*) with size larger than k pixels, extracted with the *BlobFilter* function based on connected components (particles are dilated p times: p(CWL) = 1, p(CLSM) = 2). This feature is only computed for intensity images of long-term aging series (S3 - S5, S7 - S10). Dust particles are segmented in images TNI_n by a manually determined threshold: *thresh$_{dust}$(S3) = 200, thresh$_{dust}$(S4) = 430, thresh$_{dust}$(S5) = 580, thresh$_{dust}$(S7) = 26,350, thresh$_{dust}$(S8) = 48,000, thresh$_{dust}$(S9) = 33,400, thresh$_{dust}$(S10) = 9,300*). The lower bound of the particle size k is also manually determined, according to a set's resolution: *k(S3) = 6, k(S4) = 32, k(S5) = 3, k(S7) = 50, k(S8) = 100, k(S9) = 250, k(S10) = 500*. Extracted dust particles of one image are stored in dp_n. The cumulated image dp_{cum_n} of all prior segmented dust particles is obtained by adding the current image dp_n to its predecessor dp_{n-1} (adding refers to a pixel value output $dp_{cum\_n}(x,y)$ greater than zero if at least one of the two corresponding input pixels $dp_n(x,y)$ or $dp_{n-1}(x,y)$ represents a dust particle.

CWL: Chromatic White Light Sensor, CLSM: Confocal Laser Scanning Microscope, TNI$_n$: Temporally normalized image n of a time series, TNI$_n$(x,y): Specific image pixel of a temporally normalized image n of a time series, Binarize: Binarizes a given image with the provided threshold *thresh* or *thresh$_{dust}$*. Abs: Computation of the absolute value for each image pixel, Mean: Computation of mean pixel gray value for all pixels not masked by bg', StdDev: Computation of pixel variance for all not masked by bg', LocVar: Computation of block-based local variance for overlapping blocks of size *kernel* x *kernel*, Sobel: Computation of block-based Sobel filter for overlapping blocks of size *kernel* x *kernel* in different directions (X, Y or XY). FirstDigit: Computes the first (highest) digit of a number. BlobCount: Computation of amount of segmented blobs. BlobFilter: Segmentation of blobs greater than k pixels based on connected components. Dilate: Application of p dilatation operations. Upscaling refers to the multiplication of a feature value with a factor between 10,000 and 100,000 for sufficient numeric precision. Feature F19 is only applied to intensity long-term aging series (S3 - S5 and S7 - S10), because dust is only relevant over long aging periods and can only be reliably detected in intensity images at this point in time.

as the size of sweat particles, exhibit characteristic changes, similar in trend to their statistical counterparts and seem therefore reducible to them. They are also not studied here.

All features are computed from the temporally normalized images TNI_n of a time series but only for pixel locations marked in bg' as non-background, therefore excluding not only the background but also distortions caused by dust or scan artifacts (see section IV-A). In some cases, only few pixels remain for the feature calculation. However, they are considered as sufficient because most of the distortions have been removed using bg'. For set S6 (which consists of a separate sub-series for each point in time of the main series), the features are computed from the first five values of the sub-series (chosen in respect to preliminary tests), which are used as supporting points for a standard linear regression. The slope of the approximated linear function is used as the feature value of this specific point in time. The set is then processed in a similar way to all other sets.
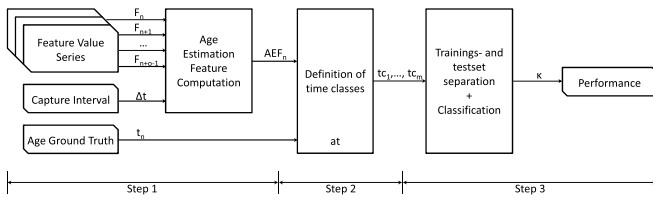
Fig. 10. Proposed age classification steps adapted from [2].

The introduced features F1 - F19 can be applied to intensity as well as topography images of a capturing device, therefore leading to a total of 38 feature combinations. However, the dust feature F19 cannot be computed for all sets and data types. It cannot be computed for the aging series from S1, S2 and S6 (neither intensity nor topography series) because an insufficient amount of dust settles during the captured period of time (for S1 and S2, time series are comprised of a total period of 24 hours; for S6, feature values are computed from five consecutive images, captured with a temporal offset of one hour). Therefore, only 36 feature combinations are available for these sets. Furthermore, the dust feature can generally not be computed for topography series because it is particularly hard to separate dust particles from noise and print particles within these images. This leads to a total of 37 feature combinations for the remaining sets.

C. Experimental Protocols for Age Estimation

Different age estimation strategies are possible, such as age estimation using aging functions or machine-learning based classification. Their results can be compared using common quality measures, e.g. as proposed in section IV-D. According to the findings of [2], machine-learning based classification seems to be more reliable at this point and is used here for baseline performance computation. A general experimental protocol for machine-learning base age estimation can be laid out in three consecutive steps which are visualized in Fig. 10.

Step 1 - Age Information Selection: In this step, the specific *feature derivatives* used for age estimation have to be determined. Features are inputted in the form of feature value series where each feature F has an index n representing its position within the time series (corresponding to the position of the image it is computed from). Apart from using the feature values directly, the first order derivative (representing the speed of change of the feature values) as well as the second order derivative (representing the change of that speed) at a certain point in time can be applied, leading to feature derivatives, here referred to as *age estimation features (AEF)*. The selection of adequate feature derivatives has to be conducted in respect to the investigated challenges C1 - C4 as well as the features to be used. Within the scope of the here presented baseline performance computation, the first derivative is used for all but the dust feature (F19) because it represents the speed of print degradation for a certain print age t_n. For the dust evaluation, the original feature values are used because the amount of dust particles settling on the probe over time is not expected to change during aging, rendering its first derivative constant throughout time. Therefore, the feature

values of F19 are directly used for classification ($AEF_n = F_n$). For approximating the first derivative of a time series of features, a *slope approximation method* (e.g. regression), *slope approximation type* (e.g. linear or logarithmic approximation) as well as the *amount of supporting points* need to be specified. For approximating the first derivative of features F1 - F18 in the scope of the here conducted baseline performance computation, linear regression is applied as slope approximation method and type to compute the slope of the feature curve for the interval $[F_n, F_{n+o-1}]$ where o denotes the amount of supporting points, i.e. the amount of consecutive feature values F_n, F_{n+1}, ..., F_{n+o-1} ($1 < o \leq n$) used for approximation. This interval needs to be large enough to smooth out distortions of the feature curve at single points but small enough to closely resemble the slope of the curve at the particular point in time represented by n. Optimal values for o have been investigated in [2] according to their maximum age estimation performance and are applied here to all sets ($o = 5$ for CWL and $o = 3$ for CLSM images). As a result, a total of n-o+1 age estimation features AEF_n (feature derivatives) can be computed from each feature time series, which are later used for classifying the age of a print at sample points n, comparing it to its respective ground-truth age t_n.

Step 2 - Definition of Age Classes: For classification, age classes ac need to be defined, e.g. in respect to the specific challenge C1 - C4 or application scenario investigated. Different amounts and widths of age classes with or without gaps in between are possible. For the here computed baseline performance, two class problems are selected for study, allowing for a systematic investigation of a basic classification scenario. All available samples of a set are classified in this scenario into those younger or older than a systematically increased *time threshold tt*, iterating through all available capturing points, beginning from a value of one and ending with the maximum capture period of a set ($tt_{n+1} = tt_n + \Delta t$, $\Delta t = 1$ *hour* for all short-term aging series, $\Delta t = 1$ *week* for all long-term aging series). For example, a time threshold of $tt = 2h$ would be considered as a classification of the samples into those younger or older than 2 hours. As a consequence, the point n within a time series and its corresponding ground-truth age t_n are replaced by an age class value $ac = $ *'younger'* or $ac = $ *'older'* in respect to the chosen time threshold tt. Other time thresholds or age classes are possible in the future to provide additional information on the age estimation performance.

Step 3 - Definition of Classification Parameters: For the age estimation performance computation, several classification parameters need to be specified. They include the specific *toolbox* to be used, *classifier* applied and method used for *training and test set division*. In the here suggested experimental protocol, training and test sets are divided in the scope of a standard five-fold cross validation (whereas the first 80% of AEF values from each fold are used for training, leaving the remaining 20% for performance evaluation). The splits are performed manually to include all features of a specific time series either into the training or into the test set to realize independent samples for model training and performance evaluation. For each fold, the training samples are divided into the two age classes $ac = $ *{younger, older}* and the amount of samples of

both classes is equalized by randomly deleting samples from the bigger class until the sizes match. A model is trained for each fold using the WEKA machine-learning toolbox [40] and the LMT classifier (found in [2] as well-performing). The Cohen's kappa classification performance κ [41] is returned and averaged over the five folds. The standard deviation of the kappa values from the different folds is usually below 0.05 (in about 99% of cases for set S1 and 98% of cases for set S2).

D. Used Quality Metrics

As quality metrics, the well known (Pearson) correlation coefficient r and the classification kappa κ are used. To gain a first insight into the characteristics of each feature (e.g. its general trend), the average correlation coefficient r_avrg is computed for all curves of a set in respect to a certain feature. If r_avrg is greater than or equal to zero, a positive trend is assumed, otherwise, the trend is considered negative.

Furthermore, to see how reproducible a trend is in respect to all time series of a certain set (giving a first indication about its later classification performance), the relative frequency of curves $h(|r| \geq 0.8)$ is computed, including all curves with a strong correlation (either positive for $r\_avrg \geq 0.8$ or negative for $r\_avrg \leq -0.8$). A logarithmic function is used for approximation because many aging curves exhibit a logarithmic-like aging trend. More feature-specific approximation functions could also be selected in future work. The dust feature F19 is considered as an exception to this procedure. Because dust is trivially known as not to change in a logarithmic manner, but rather increases in a linear fashion (regular settlement of dust on the probe), a linear approximation is used here.

In a second part of baseline performance evaluation, the classification accuracy of age estimation based on two-class problems is studied (to decide if a print is younger or older than a time threshold tt). The well-known kappa value κ is applied as quality measure, based on a five-fold cross validation and the LMT classifier using the WEKA machine-learning toolbox. The classification performance is presented in respect to different time thresholds tt ($tt = \{1,2,...,23h\}$ for short-term aging and up to $tt = \{1,2,...,29m\}$ for long-term aging) as well as in respect to different feature values and a combination of all available features for each set.

V. EXPERIMENTAL RESULTS

In this section, the experimental results for the baseline performance of the proposed sets of latent print time series are presented. In the first part, the correlation coefficients r_avrg and $h(|r| \geq 0.8)$ are used to analyze the general feasibility of the features by evaluation of the characteristic progression of their aging curves. In the second part, machine-learning based age estimation is applied using two-class problems to evaluate the practical age estimation performance of the features. For classification, the 19 features F1 - F19 are applied, leading to 36 feature combinations for S1, S2 and S6 as well as 37 feature combinations for S3 - S5 and S7 - S10 (see also section IV-B).
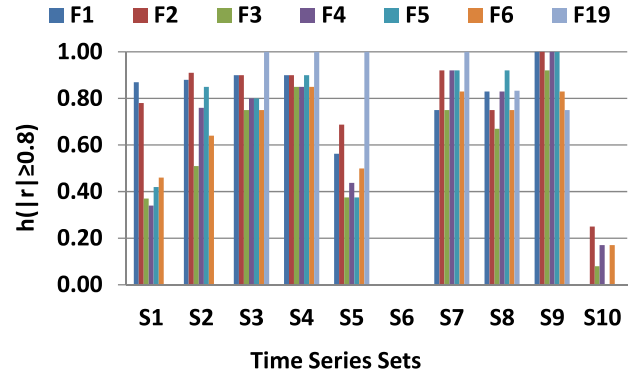

Fig. 11. Relative frequency of aging curves with a strong correlation $h(|r| \geq 0.8)$ for features F1, F2, F3, F4, F5, F6 and F19 (bars from left to right).

A. Feature Feasibility Based on Correlation Coefficients

In this section, the correlation coefficients computed from the feature curves are evaluated in respect to the different features and sets of the collection. The main hypothesis is that features leading to a large amount of curves with a characteristic progression (i.e. a high correlation coefficient to a mathematical aging function) will also lead to a comparatively good performance in the later conducted age classification. The results of the average correlation coefficient r_avrg as well as the relative frequency of curves exhibiting a strong correlation to a logarithmic trend $h(|r| \geq 0.8)$ are provided in Supplementary VI for all features and sets. They show that the general trend of feature curves can be positive or negative, depending on the characteristics of the feature. During print degradation, water as well as short-chained fatty acid evaporate, other components degrade. This leads to a general increase in image brightness due to more light being reflected from the surface (i.e. less light being absorbed, scattered or refracted at water or fat droplets) as well as a decrease in particle height. Therefore, brightness-based features (e.g. F1 or F2) exhibit increased feature values during aging, whereas roughness- or gradient-based features (e.g. F5 or F6) exhibit a decreasing trend.

In respect to the exhibition of characteristic trends, features F1 - F6 as well as F19 perform comparatively well for intensity images throughout most sets, whereas features F7 and F8 as well as the Benford's Law features F9 - F18 perform significantly worse. Set S6 shows particularly poor results suggesting the unfeasibility of the features F1 - F19 for the study of these series of series. The set is therefore not further evaluated in this paper. Topography images perform in general much worse than their intensity counterparts. Especially for sets S8 - S10, the topographic feature curves and their corresponding correlation coefficients cannot even be computed because the applied preprocessing methods are inadequate for such high-resolution topography information, requiring additional methods in the future.

Features with an average correlation coefficient r_avrg close to one (or minus one) usually also exhibit a comparatively high frequency of curves with a strong correlation $h(|r| \geq 0.8)$. The best intensity-image based results are exhibited by F1 - F6 as well as F19 in respect to $h(|r| \geq 0.8)$ and are
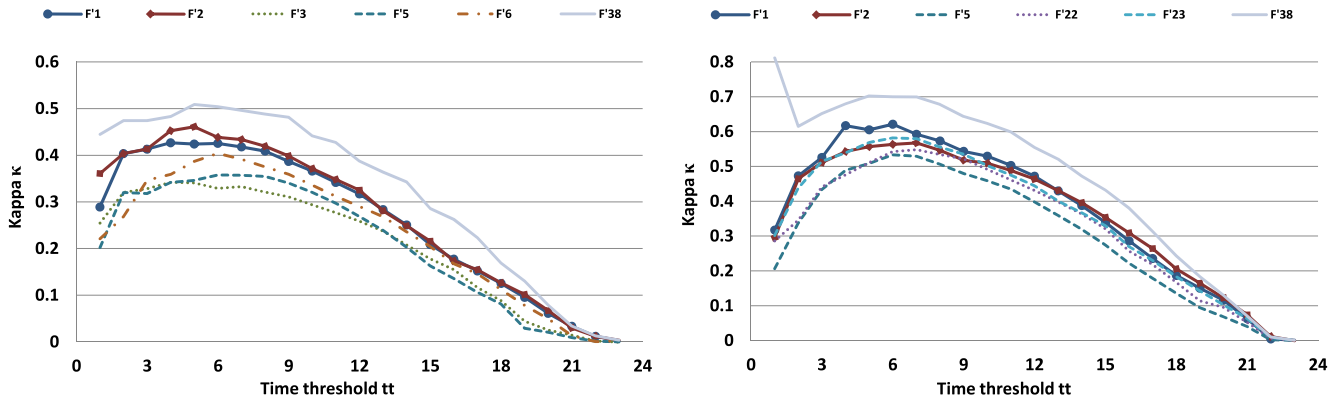
Fig. 12.   Short-term aging classification performance κ of set S1 (CWL, left) and set S2 (CLSM, right) in respect to different time thresholds tt = {1,2,..,23h}. Results are depicted for the combination of all features F'38 as well as the five best performing single features (F'1 - F'8: F1 - F8 based on intensity images, F'9 - F'16: F1 - F8 based on topography images, F'17 - F'26: F9 - F18 based on intensity images, F'27 - F'36: F9 - F18 based on topography images, F'37: F19 based on intensity images is not computed for short-term aging series).

compared in more detail in Fig. 11. The figure shows that apart from the very bad performance of all features for S6 and S10, all selected features perform well on sets S3, S4, S7, S8 and S9, whereas differences occur for sets S1, S2 and S5. For the short-term aging sets S1 and S2, features F1 and F2 perform best, whereas for the long-term aging sets, the dust feature F19 performs best in all cases except for high magnifications (S8 and S9, where it still performs almost as good as many other features). Concluding from such findings, the best classification results should be expected for features F1 and F2 (short-term aging) as well as feature F19 (long-term aging). However, the combination of different features might also lead to promising results. Features performing badly concerning their trend (i.e. having average correlation coefficients close to zero or a low frequency of strong correlations) are expected to perform poorly in the machine-learning based classification.

B. Short-Term Age Estimation Performance Using Machine-Learning

For short-term aging periods of up to 24 hours, the main hypothesis suggests that features F1 and F2 perform best. Features are interdependent to a certain degree but some performance increase is expected for the combination of all features (F'38). The results for the short-term age estimation performance for sets S1 (CWL) and S2 (CLSM) are depicted in Fig. 12 in respect to different time thresholds $tt = \{1, 2, .., 23h\}$. Together with the results of all combined features (F'38), the five best-performing single features are depicted (the results of all features are given in Supplementary VII). It can be seen from the figure that the combined performance of all features performs slightly better than the performance of the best single feature for both devices (with a kappa improvement of about +0.1). Furthermore, a general trend can be seen in that intensity-based features mostly perform better than topography-based ones and that Benford's Law features mostly perform worse than other features (please also refer to Supplementary VII for additional feature curves). Furthermore, the CWL sensor (left

figure) achieves a lower total performance than the CLSM device (right figure). However, in future studies it also has to be investigated to what extent certain influences of the capturing devices are systematically impacting the observed print intensity (e.g. a very strong decrease of light source laser intensity during the measurement period of half a year was observed for the CLSM device, leading to a significant decrease in overall image brightness, which is corrected by temporal normalization). When studying the performances in respect to different time thresholds tt, both devices exhibit the best performance for thresholds between $tt = 4\ h$ and $tt = 8\ h$ (highest performance for both devices is achieved using F'38 and $tt = 5h$: $\kappa\_max(CWL) = 0.51$, $\kappa\_max(CLSM) = 0.70$) with decreasing performance for high thresholds and a certain performance fluctuation for low thresholds which can peak very high in a few cases of the CLSM device (however, these peaks are not considered for the highest performance determination since their cause is unclear at this point). The continuous decrease in age estimation performance for higher time thresholds appears to be mainly caused by the drying process of the prints, where aging speeds become more and more undistinguishable the dryer a print becomes. Concluding from these findings, the age estimation performance is highest for time thresholds between four and eight hours and is decreased with continuous print age. In general, the single features F1 and F2 perform best for both devices, achieving a performance close to the combination of all features F'38 (with an average kappa decrease of about -0.1). Therefore, the investigated features appear to be highly interdependent and very few (if not a single) intensity-based statistical feature(s) might be sufficient for achieving the best overall observed age estimation performance.

C. Long-Term Age Estimation Performance Using Machine-Learning

According to the main hypothesis, long-term aging periods should exhibit worse age estimation results for the features F1 - F18 than the short-term aging periods since the prints have largely dried and less degradation behavior can be observed. However, the capture intervals have also
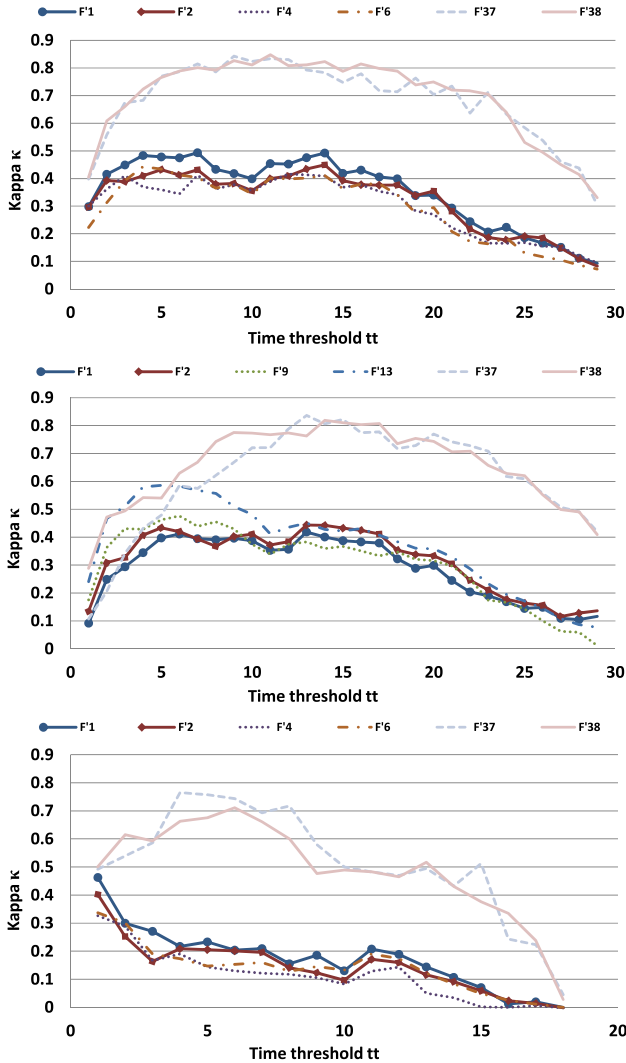
Fig. 13. CWL long-term aging classification performance κ of sets S3 (upper image), S4 (middle image) and S5 (lower image) for the investigated single features F1 - F19 (F'1 - F'8: F1 - F8 based on intensity images, F'9 - F'16: F1 - F8 based on topography images, F'17 - F'26: F9 - F18 based on intensity images, F'27 - F'36: F9 - F18 based on topography images, F'37: F19 based on intensity images) and the combined performance of all features F'38 in respect to different time thresholds tt = {1,2,..,29m}.

become larger, potentially countering such trend. The dust feature F19 should perform particularly well since dust is expected to continuously accumulate on the print over time. The results for CWL long-term aging periods of up to three years (with time thresholds tt = {1,2,..,29m}) are depicted in Fig. 13 for all combined features (F'38) as well as the five best-performing single features (the results for all features are given in Supplementary VIII). They show a similar trend with regard to the short-term aging of Fig. 12 in respect to intensity-based features performing better than topography-based ones, Benford's Law features performing comparatively bad and a combined classification using all features performing better than most single features. However, a big difference can be seen in the dust feature F'37 computed from intensity images, which performs extraordinary well and can alone achieve a similarly good performance than combining all features in F'38. Therefore, the improved dust feature is very promis-

ing, performing much better than any other feature. For the sets S3 and S4 (with a higher resolution than S5), the dust feature achieves very good classification results for a comparatively broad range of time thresholds (between 5 and 20 months; best performance of S3 is achieved for tt = 11m with $\kappa = 0.85$, best performance of S4 is achieved for tt = 14m with $\kappa = 0.82$). For set S5 (exhibiting a slightly lower resolution), a similar trend is present, however less strong, best results are achieved for time thresholds between 2 and 8 months (best performance for tt = 6m with $\kappa = 0.71$). However, only between 17 and 20 differing time series are available for sets S3 - S7, leading only to first qualitative performance indications. The general trend of decreasing age estimation performance for increasing print age can also be observed here, yet not as fast and clear as for the earlier discussed short-term aging.

The CLSM-based long-term aging periods of S7 - S10 (captured over half a year) are investigated in Fig. 14 in respect to time thresholds tt = {1,2,..,6m} exhibiting the age estimation performance for the combination of all features (F'38) as well as the five best-performing single features (the results of all features are given in Supplementary IX). With only 12 time series each, these sets are the least reliable and can only provide a very broad qualitative indication of aging properties. They are varying in their capture magnification. Higher magnifications lead to a smaller size of the measured area which becomes extremely small for S10 (0.116 x 0.080 mm), barely including a few droplets. The results of Fig. 14 (see also Supplementary IX) show that while the combined approach F'38 of using all features still performs best in many cases, other features are often close in performance and the earlier observed differences between intensity and topography-based features as well as between Benford's Law and other features become less clear (whereas topography features can only be computed for S7). Most important, the extraordinarily good performance of the dust feature F'37 has disappeared and the feature performs often worse than other features (especially with increasing magnification). This might be a result of the changes in size and shape of the particles for higher magnifications but can most likely be attributed to the decreasing measured area size, where less and less particles are actually captured for investigation and feature computation (especially for the dust feature, almost no dust particle settled on the investigated area of 0.116 x 0.080 mm during the investigation period). Therefore, several tiles should be captured and stitched together in the future if very high magnifications are chosen or (alternatively) additional features should be designed, focusing on selected particles only. Best performances are achieved for time thresholds of two or three months (S7: $\kappa = 0.55$ for tt = 3m, S8: $\kappa = 0.61$ for tt = 3m, S9: $\kappa = 0.75$ for tt = 2m, S10: $\kappa = 0.44$ for tt = 2m). The general trend of decreasing age estimation performance with increasing time threshold can also be observed, however much less regular than for the prior examined sets.

Concluding from these results for the long-term aging of CWL and CLSM time series, the single dust feature F19 performs especially well, achieving an age estimation performance similar to all features combined (F'38). However, the feature requires a certain minimum measurement area
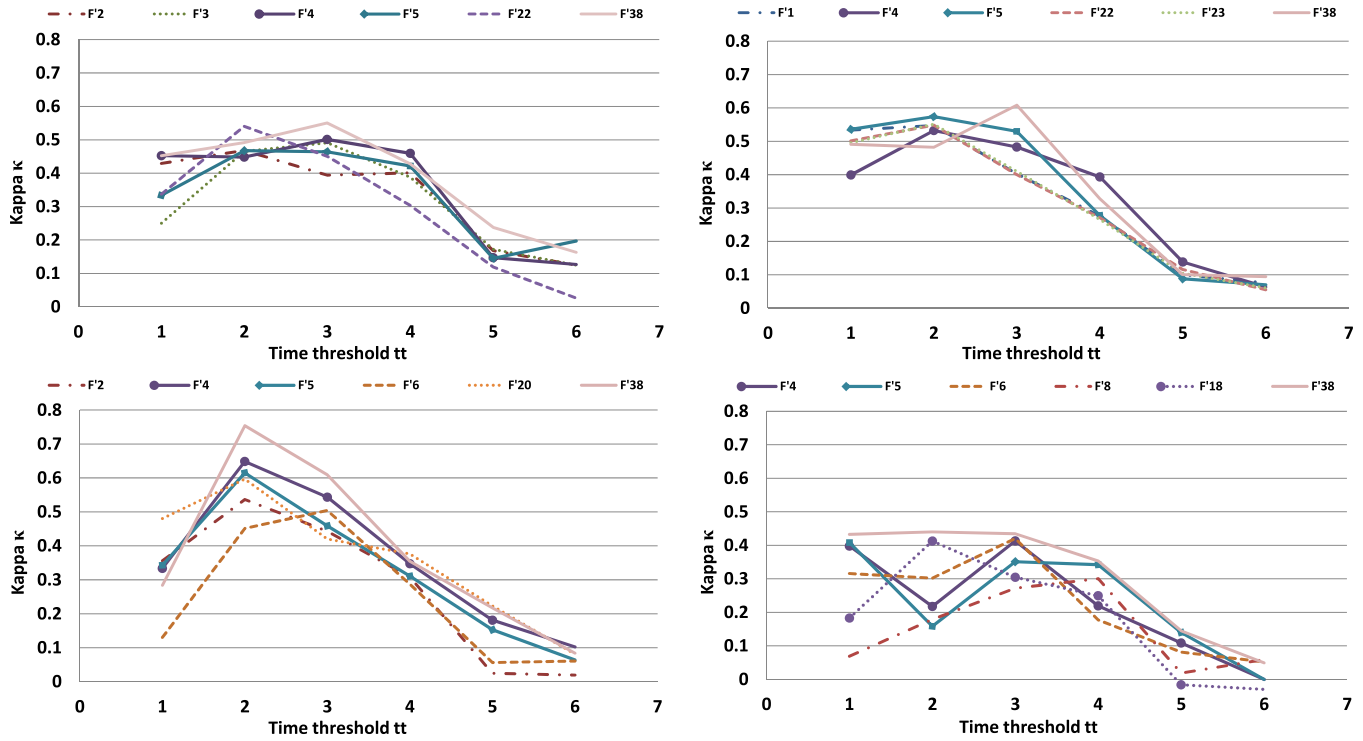
Fig. 14. CLSM long-term aging classification performance κ of sets S7 (upper left image), S8 (upper right image) , S9 (lower left image) and S10 (lower right image) for the investigated single features F1 - F19 (F'1 - F'8: F1 - F8 based on intensity images, F'9 - F'16: F1 - F8 based on topography images, F'17 - F'26: F9 - F18 based on intensity images, F'27 - F'36: F9 - F18 based on topography images, F'37: F19 based on intensity images) and the combined performance of all features F'38 in respect to different time thresholds tt = {1,2,..,6m}.

(on which dust can settle). The very high resolution of the CLSM images does not exhibit such measured area size, not allowing for a sufficient amount of dust particles to settle. Furthermore, dust models have to be studied in the future to create either location-type dependent or crime scene specific dust models for age estimation. The classification performance decreases over time, potentially due to the increased drying of the prints, continuously decelerating the speed of degradation.

D. Discussion

From the overall results it can be concluded that intensity images perform generally better than topography images for the investigated CWL and CLSM devices due to a significantly lower amount of noise and sensor artifacts. Moreover, a certain interdependence among the features is present, with single features often not performing significantly worse than the combination of all features. Therefore, it can be concluded that the different features target a similar characteristic in image changes, which is the loss in overall image contrast. Benford's Law features performed poorly and are rather unfeasible. Features F1 and F2 performed comparatively well throughout most short-term aging series, further highlighting that the characteristic changes during print aging are represented by a general loss of image contrast. Their decreased performance for long-term periods indicates that the drying process of latent prints is strongest in the beginning and eventually becomes very slow, being detectable only over very long time periods. The variance of aging speeds using these features is very high for fresh prints, leading to the conclusion that fresh prints contain either a lot of volatile components (e.g. water) or only very few (e.g. being very dry prints).

After drying for some hours, this variance strongly decreases, leading to slow aging speeds for all prints. As a result, fresh prints with a high aging speed are very well detectible, whereas fresh prints with a low aging speed are difficult to be separated from dried ones. This also explains the fact that age estimation accuracy increases for small time thresholds, peaks at a certain optimal separation performance and then steadily decreases for large time thresholds.

Dust exhibits a rather linear aging characteristic, making it very feasible to complement F1 and F2 for long-term aging periods. The promising results of feature F19, investigated here for the first time, confirm the very high potential of using dust for long-term age estimation. However, location-dependent models about environmental dust have to be established in future studies. Moreover, the magnification dependence of image processing filters has been shown, where very high print magnifications using the CLSM device could not be used for age estimation with the here applied filtering techniques. However, such magnifications could be very valuable when being studied with approaches targeting certain print substance particles (e.g. changes in droplet size and shape) based on additional signal processing techniques in the future.

VI. CONCLUSION AND FUTURE WORK

Current trends in latent print crime scene investigations show the increased application of contactless, non-invasive capturing devices, enabling the acquisition of print time series. In this paper, a first publicly available database of high-resolution latent print time series has been presented. It will allow the signal processing community to study 2,618 time series with a total of 117,384 scans. Consisting of ten sets,

prints from a pool of 116 different donors were included. Varying capturing devices, data types, aging periods and resolutions as well as the annotation with donor characteristics and ambient conditions make this collection a valuable base for comparable research in the field of print age estimation and beyond.

A baseline feature space was proposed based on 19 features from earlier work as well as novel features such as Tamura Contrast, Benford's Law and an improved dust feature. The features were applied to intensity as well as topography data of the time series, leading in best cases to a large amount of correlation coefficients greater than 0.8 and to kappa classification performances between 0.51 and 0.85.

Future work should foremost be seen in the conduct of numerous studies of a wide variety of researchers from the signal processing community for addressing the introduced challenges C1 - C4. Within the scope of these challenges, various aspects of the proposed age estimation pipeline should be improved, including preprocessing techniques for noise and scan artifact reduction, dust removal, quality enhancement, temporal normalization and print segmentation, improved features and feature extraction techniques as well as enhanced age estimation procedures.

Especially, influences from other surface types such as rough, non-planar, porous and poorly-reflecting surfaces as well as influences from the sweat composition and environmental climatic conditions are of high importance. They should either be removed or considered along the steps of the processing pipeline. Temperature and humidity influences can be studied using the annotated meta data. In the case of dust, environmental influences can even be used as a feature if reliable models of dust settlement at different locations can be established. However, addressing the various influences is particularly hard for sweat composition influences since the mere scan images do not provide any information about the specific chemical composition of the prints. Here, additional capturing devices, such as Fourier Transform Infrared Spectrometers, should be studied in the future.

Furthermore, the data collection is not limited to research on print age estimation. General studies on latent print enhancement and segmentation (especially considering contactless acquired high-resolution images), quality comparisons of intensity and topography print data (including data type fusion) as well as studies of print particles and pores should also be conducted. The here presented data collection can therefore be considered a very important and sustainable way of enabling research for the signal processing community. It allows for amply studies on a variety of issues, based on a common data collection, modular processing pipeline and basic quality metrics. At the same time, the applied anonymization and revocation mechanisms allow for a maximum compliance with data protection regulations.

REFERENCES

[1] J. De Alcaraz-Fossoul, C. M. Patris, A. B. Muntaner, C. B. Feixat, and M. G. Badia, "Determination of latent fingerprint degradation patterns—A real fieldwork study," *Int. J. Legal Med.*, vol. 127, no. 4, pp. 857–870, 2013.

[2] R. Merkel, *New Solutions for an Old Challenge: Chances and Limitations of Optical, Non-Invasive Acquisition and Digital Processing Techniques for the Age Estimation of Latent Fingerprints (Dissertation)*, Magdeburg, Germany: O.-v.-G.-Univ. Magdeburg, 2014.

[3] A. Girod, R. Ramotowski, and C. Weyermann, "Composition of fingermark residue: A qualitative and quantitative review," *Forensic Sci. Int.*, vol. 223, no. 1, pp. 10–24, 2012.

[4] *Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals With Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)*, EU Commission, COM Final, Brussels, Belgium, 2012.

[5] M. Hildebrandt and J. Dittmann, "StirTraceV2.0: Enhanced benchmarking and tuning of printed fingerprint detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 833–848, Apr. 2015.

[6] R. Merkel, J. Dittmann, and M. Hildebrandt, "Latent fingerprint persistence: A new temporal feature space for forensic trace evidence analysis," in *Proc. IEEE Int. Conf. Image (ICIP)*, Paris, France, Sep. 2014, pp. 4952–4956.

[7] R. Merkel, S. Gruhn, J. Dittmann, C. Vielhauer, and A. Braeutigam, "On non-invasive 2D and 3D Chromatic white light image sensors for age determination of latent fingerprints," *Forensic Sci. Int.*, vol. 222, no. 1, pp. 52–70, 2012.

[8] R. Merkel, S. Gruhn, J. Dittmann, C. Vielhauer, and A. Braeutigam, "General fusion approaches for the age determination of latent fingerprint traces: Results for 2D and 3D binary pixel feature fusion," *Proc. SPIE*, vol. 8290, p. 82900Y, Apr. 2012.

[9] R. Merkel, J. Dittmann, and C. Vielhauer, "Novel fingerprint aging features using binary pixel sub-tendencies: A comparison of contactless CLSM and CWL sensors," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Tenerife, Spain, Apr. 2012, pp. 7–12.

[10] R. Merkel, K. Otte, R. Clausing, J. Dittmann, C. Vielhauer, and A. Braeutigam, "First investigation of latent fingerprints long-term aging using chromatic white light sensors," in *Proc. 1st ACM Workshop Inf. Hiding Multimedia Secur. (IH&MMSec)*, Montpellier, France, Sep. 2013, pp. 95–104.

[11] A. Krapyvskyy, *Long-Term Aging Analysis of Latent and Printed Fingerprints Using Contactless Confocal Laser Scanning Microscopy (CLSM)*. Otto-Von-Guericke-Univ., Magdeburg, Germany, 2014.

[12] J. Dittmann, C. Vielhauer, M. Ulrich, C. Kraetzer, and T. Hoppe, *Digi–Dak Digitale Fingerspuren*. Magdeburg, Germany: Project Proposal, 2008.

[13] FRT GmbH. *Fries Research & Technology (FRT)*. accessed on 21, 2014. [Online]. Available: www.frtofamerica.com/us

[14] Keyence Corporation. *Keyence Color 3D Laser Scanning Microscope*. Accessed on 21, 2014. [Online]. Available: www.keyence.com/topics/vision/vk/guide.php?lf=KD

[15] F. Benford, "The law of anomalous numbers," *Proc. Amer. Phil. Soc.*, vol. 78, no. 4, pp. 551–572, 1938.

[16] K. Baniuk, "Determination of age of fingerprints," *Forensic Sci. Int.*, vol. 46, no. 1, pp. 133–137, 1990.

[17] J. F. Schwabenland, "Determining the evaporation rate of latent impressions on the exterior surfaces of aluminum beverage cans," *J. Forensic Identificat.*, vol. 42, no. 2, pp. 84–90, 1992.

[18] A. L. McRoberts and K. E. Kuhn, "A review of the case report: Determining the evaporation rate of latent impressions on the exterior surfaces of aluminum beverage cans," *J. Forensic Identificat.*, vol. 42, no. 3, pp. 213–218, 1992.

[19] M. Y. Omar and L. Ellsworth, "Possibility of using fingerprint powders for development of old fingerprints," *Sains Malaysiana*, vol. 41, no. 4, pp. 499–504, 2012.

[20] M. Azoury, E. Rozen, Y. Uziel, and Y. Peleg-Shironi, "Old latent prints developed with powder: A rare phenomenon?" *J. Forensic Identificat.*, vol. 54, no. 5, pp. 534–541, 2004.

[21] C. Kohlepp, "Möglichkeiten und Grenzen der Altersbestimmung daktyloskopischer Spuren und ihre kriminalistische Bedeutung," Wiesbaden, Germany, Tech. Rep., 1994.

[22] V. Stamm, "Methoden zur Altersbestimmung Daktyloskopischer Spuren," Wiesbaden, Germany, Tech. Rep., 1997.

[23] J. M. Duff and E. R. Menzel, "Laser-assisted thin-layer chromatography and luminescence of fingerprints: An approach to fingerprint age determination," *J. Forensic Sci.*, vol. 23, no. 1, pp. 129–134, 1978.

[24] S. Barcikowski, J. Bunte, A. Ostendorf, J. Aehnlich, and R. Herrmann, "Contribution to the age determination of fingerprint constituents using laser fluorescence spectroscopy and confocal laser scanning microscopy," *Proc. SPIE*, vol. 5621, p. 8, Dec. 2004.

[25] N. E. Archer, Y. Charles, J. A. Elliott, and S. Jickells, "Changes in the lipid composition of latent fingerprint residue with time after deposition on a surface," *Forensic Sci. Int.*, vol. 154, nos. 2–3, pp. 224–239, 2005.

[26] G. M. Mong, C. E. Petersen, and T. R. W. Clauss, *Advanced Fingerprint Analysis Project Final Report—Fingerprint Constituents*. Richland, WA, USA: Pacific Northwest National Laboratory, 1999.

[27] C. Weyermann, C. Roux, and C. Champod, "Initial results on the composition of fingerprints and its evolution as a function of time by GC/MS analysis," *J. Forensic Sci.*, vol. 56, no. 1, pp. 102–108, 2011.

[28] K. M. Antoine, S. Mortazavi, A. D. Miller, and L. M. Miller, "Chemical differences are observed in children's versus adultschemical differences are observed in children's versus adults' latent fingerprints as a function of time," *J. Forensic Sci.*, vol. 55, no. 2, pp. 513–518, 2010.

[29] D. K. Williams, C. J. Brown, and J. Bruker, "Characterization of children's latent fingerprint residues by infrared microspectroscopy: Forensic implications," *Forensic Sci. Int.*, vol. 206, nos. 1–3, pp. 161–165, 2011.

[30] A. Van Dam *et al.*, "Oxidation monitoring by fluorescence spectroscopy reveals the age of fingermarks," *Angew. Chem. Int.*, vol. 53, no. 24, pp. 6272–6275, 2014.

[31] P. Watson, R. J. Prance, S. T. Beardsmore-Rust, and H. Prance, "Imaging electrostatic fingerprints with implications for a forensic timeline," *Forensic Sci. Int.*, vol. 209, nos. 1–3, pp. 41–45, 2011.

[32] G. Popa, R. Potorac, and N. Preda, "Method for fingerprints age determination," *Romanian J. Legal Med.*, vol. 18, no. 2, pp. 149–154, 2010.

[33] M. D. Garris and R. M. McCabe. (2000). *Nist Special Database 27: Fingerprint Minutiae From Latent and Matching Tenprint Images*. accessed on Dec. 18, 2015. [Online]. Available: http://www.nist.gov/itl/iad/ig/sd27a.cfm

[34] A. Roßnagel, S. Jandt, M. Desoi, and B. Stach, "Kurzgutachten: Fingerabdrücke in wissenschaftlichen Datenbanken," *Projektgruppe Verfassungsverträgliche Technikgestaltung (Provet)*, Kassel, Germany, 2013.

[35] A. Bjoerck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA, USA: SIAM, 1996.

[36] M. Leich, S. Kiltz, J. Dittmann, and C. Vielhauer, "Non-destructive forensic latent fingerprint acquisition with chromatic white light sensors," *Proc. SPIE*, vol. 7880, p. 78800S, Apr. 2011.

[37] L. K. Huang and M. J. J. Wang, "Image thresholding by minimizing the measures of fuzziness," *Pattern Recognit.*, vol. 28, no. 1, pp. 41–51, 1995.

[38] G. W. Zack, W. E. Rogers, and S. A. Latt, "Automatic measurement of sister chromatid exchange frequency," *J. Histochem. Cytochem.*, vol. 25, no. 7, pp. 741–753, 1977.

[39] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.*, vol. 8, no. 6, pp. 460–473, Jun. 1978.

[40] M. L. G. A. T. U. O. Waikato, *Waikato Environment for Knowledge Analysis (WEKA)*. accessed Nov. 6, 2013. [Online]. Available: http://www.cs.waikato.ac.nz/ml/index.html

[41] J. Carletta, "Assessing agreement on classification tasks: The kappa statistic," *Comput. Linguistics*, vol. 22, no. 2, pp. 249–254, 1996.