

On Image Processing and Pattern Recognition for Thermograms of Watermarks in Manuscripts – A First Proof-of-Concept

Dominik Hauser¹, Matthias Beckmann^{2,3}(✉), Günther Koliander⁴, and
H. Siegfried Stiehl¹

¹ Department of Informatics, Universität Hamburg, Germany
{dominik.hauser,hans-siegfried.stiehl}@uni-hamburg.de

² Center for Industrial Mathematics, University of Bremen, Germany
matthias.beckmann@uni-bremen.de

³ Dept. of Electrical and Electronic Engineering, Imperial College London, U.K.

⁴ Acoustics Research Institute, Austrian Academy of Sciences, Austria
guenther.koliander@oeaw.ac.at

Abstract. Watermarks in historical manuscripts are figural shapes serving as tokens for provenance research (e.g. scribe identification, dating, papermill attribution, scribe-papermaker relation, trading, etc.) in Humanities such as Musicology. As of today, they come in a variety of formats: digitized handtracings and rubbings, X-ray based imagery and, more recently, thermograms acquired with infrared (IR) cameras – all of which have been made accessible via image data bases in libraries or archives like the watermark information system (WZIS). A key use case from a scholar’s perspective is the search for similar or even equal watermarks in whatever digitized data collections. Non-surprisingly, the prerequisite is the availability of a versatile, reliable, and user-friendly tool comprising methods from digital image processing (IP) and pattern recognition (PR). In our paper, we focus on bridging the gap between digitized thermograms of music manuscripts and watermark classification for similarity-based search through (i) a state-of-the-art (SOTA) analysis, (ii) a resulting conceptual design based on well-understood SOTA as well as novel methods, (iii) an easy-to-use implementation, and (iv) an experimental validation as Proof-of-Concept (PoC). The current system performance is characterized using thermograms recently made openly available within the DRACMarkS project as well as WZIS. The experimental results clearly demonstrate success in bridging the existing gap hence also setting a baseline for an as yet lacking benchmark.

Keywords: Historical Document Analysis · Image Processing · Document Analysis Systems

1 Introduction

Watermarks in historical manuscripts are not just a by-product of manual paper-making but a figurative mark of papermills – thus a key to provenance research.

They are physically generated by weaving thin wires in a framed sieve composed of a grid of chain lines and orthogonal laid lines. During the manual process of paper production both the metallic watermark and wires leave their trace in the paper pulp via indentation. Hence watermarks are discernible in manuscripts through e.g. backlighting or made visible through digital imaging techniques.

Needless to mention that visible (analogue) watermarks attracted researchers early on since 19th century with C.M. Briquet, G. Piccard and K.T. Weiss as pioneers on documenting, characterizing, annotating, and archiving (see e.g. [24], for a highly recommendable introduction). However, although digital watermark archives and information systems are available now, e.g. via the Bernstein portal⁵, no unique or even standardized watermark taxonomy exists – not to mention the lack of e.g. an ontology from a vision science perspective.

In our case of thermography for music manuscripts, the primary visual content of a digital thermogram is both a full or (due to folding or cutting of a paper sheet) partial watermark and the typically regular grid of bolder chain and thinner laid lines. Both watermarks and chain/laid lines are objects of interest for musicologists, paper historians, and material scientists (see [12] for paper and material aspects).

One advantage of thermography for manuscripts as introduced in [17] with off-the-shelf IR cameras (wavelength range e.g. 1500 – 5500 nm, lower wavelengths were shown to be suboptimal [11]) clearly is that neither hand-written notes nor hand-drawn/pre-fabricated staves are captured due to wavelength properties of used writing/printing ink (which holds for recto and verso hence imaging of bleed-through is omitted). In other words, thermography measures true watermark shape – as compared to e.g. subjective handtracings – based upon the interaction between IR camera, paper, and IR metrology lab set-up. On the downside thermogram quality is severely degraded: Low resolution (640 × 512 pixels in our case), low signal-to-noise ratio (SNR), spatially varying background texture and intensity gradients (due to e.g. material and lab handling of page). Even more, digitized watermarks (or, synonymously, strokes) may lack 4-/8-connectivity and are subject to translation, scaling, and rotation in image domain. Moreover, a rather large number of watermark classes as well as subclasses per class (viz shape variants) exist, whereas the sample size per (sub-)class may vary significantly. As already mentioned above, neither an international standardized visual taxonomy nor an ontology could have been achieved – taken all together: The nature of the image data itself and sparse/imbalanced training data at hand hinders a deep learning approach as of today.

Alternatively, based on our SOTA analysis (see Section 2), we demonstrate in our PoC the value of both a processing chain of well-understood IP methods up to binarization (see Section 3.1) and a Radon-based feature representation scheme for binarized watermarks (see Section 3.2) which amounts to classification fidelity beyond one based upon Euclidean vector space embedding (see Section 4 for validation via the DRACMarkS dataset⁶ with 461 thermograms (as

⁵ <https://www.memoryofpaper.eu/>

⁶ <https://www.oeaw.ac.at/acdh/projects/schubert-watermarks>

of August 2023) of which two visually distinct watermark classes were selected for classification experiments). Finally, we summarize our findings and discuss future research directions (see Section 5).

2 State of the Art

Although watermark detection, extraction and classification from digitized hand-tracings as well as manuscripts has been an active area of research since decades, as yet *no* work on an end-to-end (E2E) processing chain for thermograms has been published to the best of our knowledge. However, our SOTA deep dive [6] resulted in a list of well-understood IP and PR methods already applied to different watermark imagery serving as candidates for our PoC.

Apart from the fathers of study of (analogue) watermarks – coined filigranology – Zamperoni’s early work [29] laid the foundation for computer-assisted chain/laid line suppression, contrast enhancement, and binarization. Significant advancements were later made by Whelan and Soille [27, 28], who improved watermark extraction using Fourier transform and morphological operators, emphasizing the suppression of chain and laid lines. Rauber [22] focused on content-based image retrieval using watermarks encompassing both manual and semi-automatic processes, including contrast enhancement and contour refinement. Additionally, Hiary and Ng [8–10] contributed by using morphological operators and the Radon transform for highlighting the chain lines as a feature. Deep learning techniques for watermark classification have also been explored on digitized handtracings [21, 26], but face challenges due to sparse and imbalanced training data (see above) in the domain of thermography.

Building on our SOTA, it is important to stress that (i) among the various research efforts and publications in the field of watermark extraction from manuscripts, none have specifically addressed the use of thermography data and (ii) the nature of thermography data clearly suggests a preference for well-understood IP/PR methods over deep learning. Hence we aim at developing, implementing, and validating a PoC in the novel context of thermography data that (i) yields convincing results for watermark extraction and classification and also (ii) shows the potential of current thermography as a digitization technique.

3 Proof of Concept

As briefly mentioned above, given our systematic SOTA study we focus on (i) an E2E processing chain of well-understood IP/PR methods but also on (ii) a novel Radon transform based representation scheme with proven translation and scaling invariance for binarized watermarks enabling effective and accurate classification with only linear discriminant functions. Prime design criteria for our PoC were ease-of-use, transparency, understandability, and modularity. In the following we will elucidate key theoretical, methodical, and practical issues of our PoC (followed by an experimental validation in Section 4).



Fig. 1: Low-contrast page (left) from DRACMarkS raw dataset "Mus.Hs. 27666 Bl. 50_1" [15] and clipping result (right)

3.1 Image Processing (IP)

Our PoC design and implementation is specifically tailored to the DRACMarkS dataset aiming at (i) accommodating for the given image quality (see Section 1) and (ii) achieving effective watermark binarization. Certain assumptions about the images are incorporated as optional parameters in the PoC: all images 1) feature a high-intensity background (due to the heating plate), 2) are captured using the same camera and lens, 3) contain locally varying intensity degradations, and 4) show vertically oriented chain lines. Given both these valid assumptions and data quality at hand, a carefully designed combination of preprocessing methods is indispensable and described as follows.

In the DRACMarkS dataset [15], all images exhibit a high-intensity background (assumption 1)), which may skew results from both global and local image processing methods, hence making effective preprocessing essential.

Conventional cropping methods proved inadequate due to a non-rectangular, often trapezoidal shape of page images, and the risk of truncating watermarks near page edges. A simple yet effective method for the DRACMarkS dataset involves intensity clipping using a threshold on the bimodal image histogram to eliminate the high background intensities. The threshold is heuristically set as

$$\text{clip}_{\text{threshold}} = \text{mean}_{\text{img}} + \max(\text{mean}_{\text{img}} - \text{median}_{\text{img}}, c) \quad (1)$$

with $c = 300$ based on experiments (see Fig. 1 (right) for result).

The next step involves extracting chain lines using the progressive probabilistic Hough transform [16], which is essential for camera calibration (assumption 2)) including distortion correction for 25 mm focal length. After chain line extraction, a simple cropping based on projection profiles effectively removes most of the residual background without harming watermarks positioned near the page edge.

Regarding intensity degradations (assumption 3)), a non-linear median filter with an experimentally determined window size ($7 \times$ stroke width) is applied in order to locally normalize intensity variations while preserving fine detail like (e.g. neighboring) watermark strokes or chain lines.

Since tightly spaced laid lines are not relevant in our case, suppressing them is beneficial. Previous research [28,29] has shown that a notch filter in the frequency domain effectively suppresses these high frequency lines. The filter parameters (e.g. of a Gaussian) are determined by the spacing and orientation of the laid lines.

As the last preprocessing step, Contrast Limited Adaptive Histogram Equalization (CLAHE) [20] is applied for further contrast enhancement easing watermark detection. We use the root mean square (RMS) contrast as a measure of contrast, which corresponds to the standard deviation of image intensities. The clipping limit for CLAHE is set inverse to the contrast of the image. Thus an image with lower contrast receives more pronounced enhancement (see Fig. 2 (left)).

In order to extract watermarks from thermograms, treating them not as (dis-)connected contours but as strokes of a certain width proved beneficial. For detection we employ spatial anisotropic Gabor filters

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{(x-x_0)_{\theta_0}^2}{\sigma_x^2} + \frac{(y-y_0)_{\theta_0}^2}{\sigma_y^2}\right)} e^{i(2\pi F_0(x \cos(\omega_0) + y \sin(\omega_0)) + P)} \quad (2)$$

with the filter response acting as a multiplicative weight for the image's intensity values (see Fig. 2 (right)). Equation (2) shows the complex Gabor filter, as in [19], and the parameters⁷ are set such that the filter is tailored to watermark strokes. Hence an even Gabor filter (the real part with zero offset), similar to filters used for detecting text lines of a certain height and orientation [7], is used. The choice of an even Gabor filter ensures that the positive part of the cosine wave matches the width of the watermark strokes (see Section 4), resulting in a peak output when positioned directly on the strokes. The user-set watermark stroke width defines the spatial frequency of the filter.

To detect strokes of different orientation, we use eight filters, oriented from 0° to 157.5° in 22.5° increments, each with an anisotropic Gaussian envelope for preferred orientation detection. After convolution with these eight Gabor filters, aggregating the responses effectively becomes a challenge. A simple average of

⁷ (x_0, y_0) – center coordinates of the Gaussian; θ_0 – rotation angle giving $(x - x_0)_{\theta_0} = (x - x_0) \cos(\theta) + (y - y_0) \sin(\theta)$ and $(y - y_0)_{\theta_0} = -(x - x_0) \sin(\theta) + (y - y_0) \cos(\theta)$; σ_x^2, σ_y^2 – variance in x-/y-direction; (u_0, v_0) – spatial frequency with magnitude $F_0 = \sqrt{u_0^2 + v_0^2}$ and direction $\omega_0 = \arctan\left(\frac{v_0}{u_0}\right)$; P – offset of the sinusoid

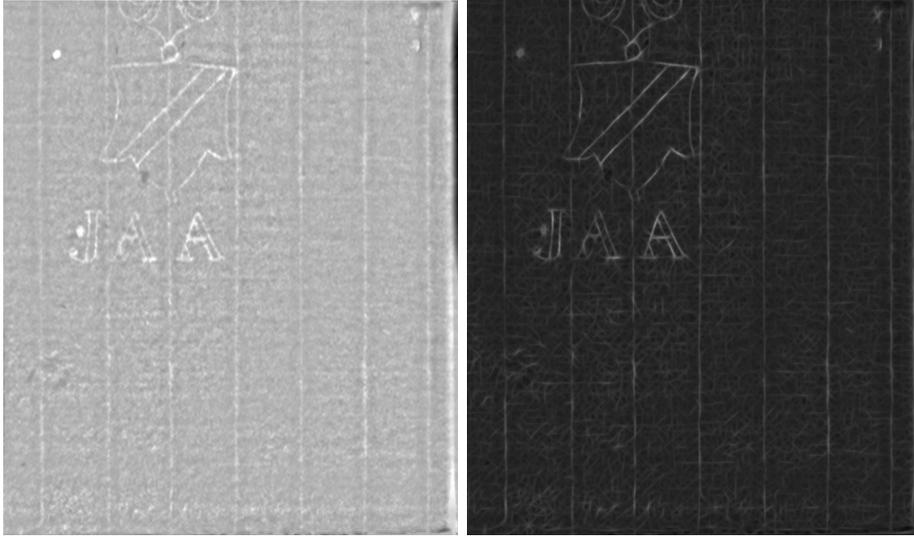


Fig. 2: Results of (left) preprocessing (camera calibration, laid line suppression, contrast enhancement and normalization) data in Fig. 1 (right) and (right) after Gabor filtering of preprocessed data from (left)

all outputs might diminish the significance of the most relevant orientation for each watermark pixel. Thus, we assign each pixel the maximum value from all filter responses, prioritizing real watermark strokes. The result may contain spurious short line segments, representing errors which have a smaller filter response compared to those from the watermark strokes.

In order to mitigate these errors, we then apply mathematical morphology, specifically the black top-hat operator (cf. [25]), which is effective in isolating small objects against the background. This operator is defined as the difference between the image's closing and the original image (refer to [5] for details on morphological operators). The size of the cross-shaped structuring element matches the given width of watermark strokes and its shape is chosen for its superior performance. The result further enhances watermarks while reducing spurious errors.

As a final IP step, our binarization converts the result from the black top-hat operator, in short black hat image, into a binary watermark point set. Global thresholding techniques prove ineffective due to the dominance of zero-valued pixels, and adaptive thresholding struggles with high noise levels in the lower value range of the black hat image, leading to errors. Therefore we apply the hysteresis thresholding proposed by Canny [4] after simple histogram equalization stretching the histogram of the black hat image and thus simplifying threshold selection. For the DRACMarkS dataset [15], we identified effective thresholds after equalization through comparative analysis of a subset consisting of two images from each of 39 pages (78 images in total). While still resulting in some

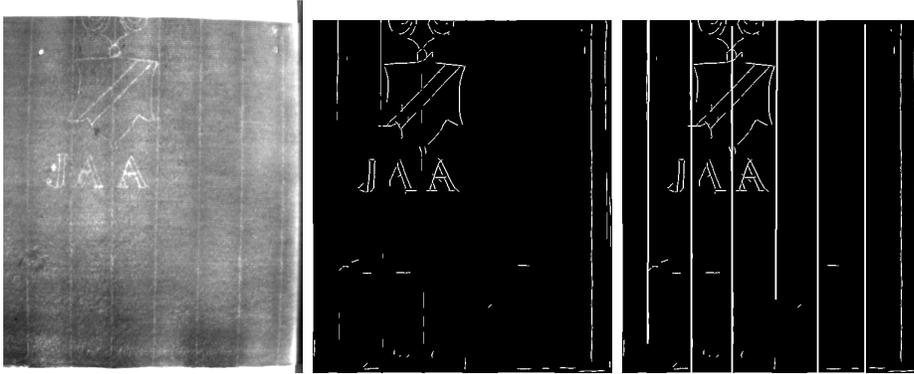


Fig. 3: Result of binarization with masked (middle) and overlaid (right) chain-lines of preprocessed raw data (left)

false positives, these thresholds accommodate well for the variability in image quality within the DRACMarkS dataset [15].

In the end, we present two outputs for the binarized watermark (see Fig. 3): The first one masks nearly all detected chain lines, while the second one overlays them (e.g. for providing valuable insights into paper making).

3.2 Pattern Recognition (PR)

Given the limited access to labelled thermograms of watermarks that can serve as training data, the determination of similarities of watermarks, which is essential for any PR approach, requires a suitable feature representation of a given binarized watermark. Moreover, we require that the PR results are invariant with respect to translation and scaling of the watermark and are tolerant against still remaining spurious short line segments, cf. Fig. 3 (middle).

For this purpose, we apply the so-called Radon Cumulative Distribution Transform (R-CDT) introduced in [13]. As a first step, for an image, modelled as a non-negative function $f: \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}$ normalized by $\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) dx dy = 1$, the Radon transform

$$\mathcal{R}f(t, \theta) = \int_{\ell_{t, \theta}} f(x, y) d(x, y)$$

is computed, where $\ell_{t, \theta} = \{(x, y) \in \mathbb{R}^2 \mid x \cos(\theta) + y \sin(\theta) = t\}$ denotes the unique straight line with signed distance $t \in \mathbb{R}$ to the origin and normal direction $\mathbf{n}_\theta = (\cos(\theta), \sin(\theta))$ for $\theta \in [0, \pi)$. In a second step, for each angle $\theta \in [0, \pi)$ the cumulative distribution transform (CDT) \hat{p}_θ of the univariate Radon projection $p_\theta = \mathcal{R}f(\cdot, \theta)$ is computed such that

$$\int_{-\infty}^{\hat{p}_\theta(s)} p_\theta(t) dt = \int_{-\infty}^s r(t) dt$$

for a fixed reference signal $r: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ satisfying $\int_{\mathbb{R}} r(t) dt = 1$, typically chosen as the characteristic function $r = \chi_{[0,1]}$ of the interval $[0, 1]$. Note that \hat{p}_θ corresponds to the optimal transport map that transports r to p_θ while minimizing the transport cost and can be computed based on the cumulative distribution functions of p_θ and r . In this way, any normalized image $f: \mathbb{R}^2 \rightarrow \mathbb{R}_{\geq 0}$ is mapped to its R-CDT $\hat{\mathcal{R}}f: \mathbb{R} \times [0, \pi) \rightarrow \mathbb{R}$.

This feature representation is tailored to our requirement of invariance with respect to translation and scaling of the watermark in the sense that the R-CDT enables linear separability of classes that are generated from template images by certain transformations. To be more precise, assume that we are given two image classes \mathbb{F}, \mathbb{G} that are generated by normalized template images f_0, g_0 via

$$\begin{aligned} \mathbb{F} &= \{f \mid \forall \theta \in [0, \pi) \exists h_\theta \in \mathbb{H} : \mathcal{R}f(t, \theta) = h'_\theta(t) \mathcal{R}f_0(h_\theta(t), \theta) \forall t \in \mathbb{R}\} \\ \mathbb{G} &= \{g \mid \forall \theta \in [0, \pi) \exists h_\theta \in \mathbb{H} : \mathcal{R}g(t, \theta) = h'_\theta(t) \mathcal{R}g_0(h_\theta(t), \theta) \forall t \in \mathbb{R}\}, \end{aligned}$$

where \mathbb{H} is a convex group of increasing diffeomorphisms $h: \mathbb{R} \rightarrow \mathbb{R}$ with respect to function composition. Then, we can show that the transformed image classes in R-CDT space

$$\hat{\mathbb{F}} = \{\hat{\mathcal{R}}f(\cdot, \theta_0) \mid f \in \mathbb{F}\}, \quad \hat{\mathbb{G}} = \{\hat{\mathcal{R}}g(\cdot, \theta_0) \mid g \in \mathbb{G}\}$$

are linearly separable if $\theta_0 \in [0, \pi)$ is chosen such that $\mathcal{R}\mathbb{F}(\cdot, \theta_0) \cap \mathcal{R}\mathbb{G}(\cdot, \theta_0) = \emptyset$. This is in contrast to [13], where the R-CDT is considered for *all* $\theta \in [0, \pi)$.

Of particular importance are the cases $\mathbb{H} = \{h \mid h(t) = t + c \text{ for some } c \in \mathbb{R}\}$, which leads to translation of images, i.e.,

$$\mathbb{F} = \{f \mid f(x, y) = f_0(x + c_x, y + c_y) \text{ for some } c_x, c_y \in \mathbb{R}\},$$

and $\mathbb{H} = \{h \mid h(t) = ct \text{ for some } c > 0\}$, which yields mass preserving scaling of images, i.e.,

$$\mathbb{F} = \{f \mid f(x, y) = c^2 f_0(cx, cy) \text{ for some } c > 0\}.$$

Both choices for \mathbb{H} are convex groups of diffeomorphisms with respect to function composition and, hence, we can linearly separate two image classes in R-CDT space using only *one* angle $\theta_0 \in [0, \pi)$ if the two classes were generated by translation and/or mass preserving scaling of two distinct template images. Note that, so far, image rotation cannot be covered by our theory. However, our numerical experiments in Section 4 show that linear separability in R-CDT space can also be achieved for rotation and even shear when increasing the number of angles.

To deal with discrete images $I \in \mathbb{R}^{N \times M}$ consisting of $N \times M$ pixels, we use a standard discretization [1] of the Radon operator in Python mapping I onto the Radon sinogram $\mathcal{R}I \in \mathbb{R}^{J \times K}$, where J corresponds to the number of parallel lines per angle and K is the number of angles, where we typically have $J = \lceil \sqrt{2}N \rceil$ for $M = N$. Thereon, the CDT is discretized using spline interpolation, leading to the R-CDT sinogram $\hat{\mathcal{R}}I \in \mathbb{R}^{J \times K}$, which serves as feature representation of I for the subsequent application of classical pattern recognition algorithms like,

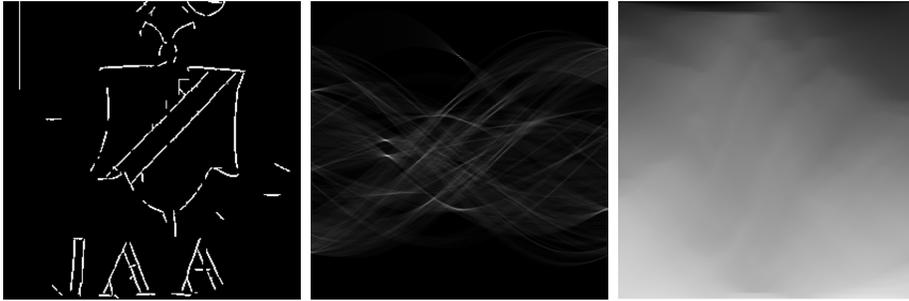


Fig. 4: Radon (middle) and R-CDT (right) sinograms as feature representation of a cropped binarized watermark (left)

e.g., support vector machines (SVMs) for classification, k-means clustering or principle component analysis (PCA) for the sake of dimensionality reduction.

For illustration, Fig. 4 shows the binarized watermark from Fig. 1 manually cropped to 280×280 pixels along with its Radon and R-CDT sinogram both of 396×180 pixels.

4 Experimental Validation

Image Processing. Next up we provide details on (i) our experimental data and strategy, (ii) so far achieved IP/PR results as well as on (iii) constraints on the parameter space. The ideal approach to our validation would require a benchmark along with a comprehensive dataset and performance metrics, yet the sparse coverage of thermography data in prior studies hinders benchmarking. Moreover, the lack of ground truth from experts having traced watermarks from thermography data force us to rely on subjective rating-based image quality assessment (IQA) with an ordinal scale (see [2]). Our validation task is further complicated by the varying quality of DRACMarkS data [15] ranging from high to low quality (see below). Our IQA aligns with the digitization guidelines provided by the Deutsche Forschungsgemeinschaft [3] and the Federal Agencies Digital Guidelines Initiative [23].

In terms of our IQA, the focus is on the following criteria: contrast between watermark and background, manuscript noise and degradations, as well as the discernibility of the watermark, particularly its completeness. To systematically assess quality, we have devised a rating scheme with four categories according to these criteria.

From the overall 461 DRACMarkS manuscript pages, a publicly accessible subset of 187 pages of six manuscripts [15] was rated by us. The categorization, based solely on our IQA, comprises 80 images rated as *low*, 58 as *medium*, and 14 as *high* quality. The remaining 35 images were unrated due to the absence of discernible watermarks. Note that ‘high quality’ within this raw dataset might

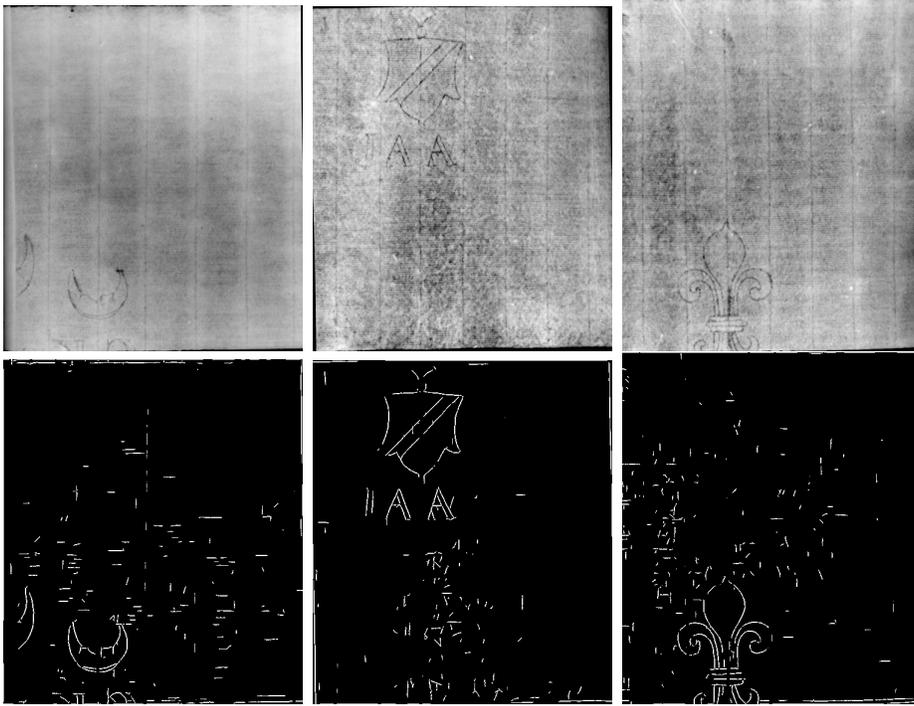


Fig. 5: Samples of preprocessed high-quality DRACMarkS data (top) and PoC results (bottom) for moon (left), blazon (middle) and lily (right) watermarks

not meet the same standard when compared to postprocessed thermography datasets (e.g. WZIS [14]).

Low quality: Raw images often lack discernible watermarks or are obscured by noise and poor contrast. The final PoC outputs typically exhibit substantial errors and are as yet unsuitable for further processing.

Medium quality: Images show degradations like strong brightness variations that do not affect the watermark. The PoC output yields watermarks with an acceptable watermark-to-error (in pixels, respectively) ratio, although spurious line segments remain. Note that error here refers to the spurious line segments scattered throughout the image.

High quality: Images contain minimal degradations and noise, but show a high watermark-to-background contrast. The PoC output captures watermark strokes effectively, with few errors like gaps (see Fig. 5).

In terms of even higher quality, the Staatsbibliothek zu Berlin⁸ (SBB) has provided thermography data to WZIS, which we also used for our validation. The SBB data display a noticeable improvement in quality compared to those of DRACMarkS, especially in terms of noise, contrast, and brightness variations

⁸ <https://staatsbibliothek-berlin.de/en/about-the-library/departments/music>

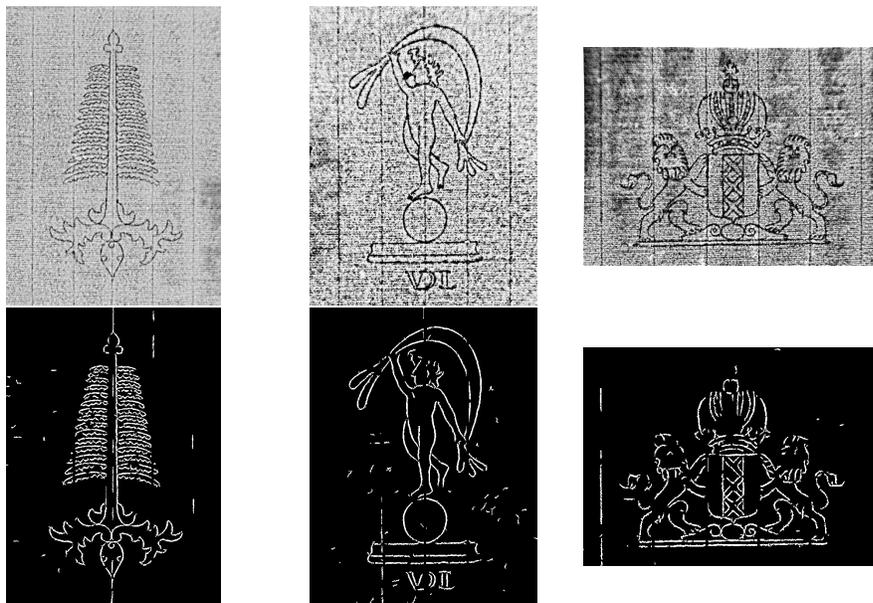


Fig. 6: Samples of original high-quality SBB/WZIS data (top) and PoC results (bottom) for watermarks of different figural shape on different backgrounds

and are cropped to show only the watermark. However, the lack of metadata detail about the digitization process and postprocessing of SBB’s raw data impedes direct comparison to raw data from DRACMarkS.

Our PoC, developed for the DRACMarkS raw data, needs only slight adjustments when applied to the SBB data, e.g. intensity clipping and camera calibration are not necessary. The resulting successful binarization of watermarks demonstrates the versatility and robustness of our PoC. Our results are highly satisfactory, see Fig. 6 for our PoC output based on three SBB thermograms^{9,10,11} with masked chain lines, exceptionally low error rates, and nearly complete binarized watermarks.

In closing the multi-stage IP part, regarding ease-of-use it is worth stressing that *all* parameters except one are (i) optimized for the DRACMarkS dataset via experimentation and (ii) fixed for all experiments with both datasets (also the one of SBB/WZIS). The *only* user-set parameter is the stroke width of watermarks (e.g. 3 pixels for DRACMarkS and 5 pixels for SBB).

Classification. To illustrate the effectiveness of our proposed feature representation for binarized watermarks, we now focus on a binary classification task

⁹ https://www.wasserzeichen-online.de/wzis/?ref=DE0960-BachSt191a_186

¹⁰ https://www.wasserzeichen-online.de/wzis/?ref=DE0960-Martines1M_III

¹¹ https://www.wasserzeichen-online.de/wzis/?ref=DE0960-Telemann21737_200_23v

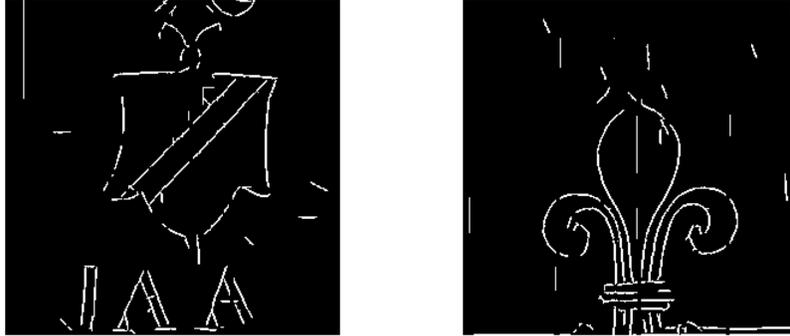


Fig. 7: Binarized blazon/lily watermarks serving as templates for dataset via data augmentation with random translation, scaling, rotation and shear

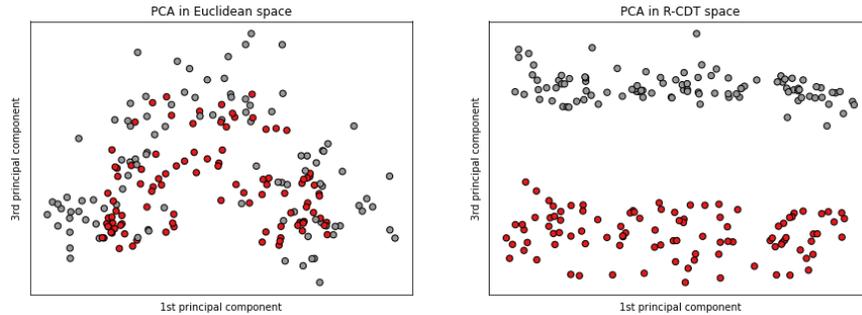


Fig. 8: Result of PCA for fully augmented watermark dataset (Fig. 7) in both Euclidean (left) and R-CDT (right) space

and consider two different watermarks from manuscripts of the DRACMarkS dataset [15], which we binarized and manually cropped to 280×280 pixels, see Fig. 7. Each of those serves as the template image of a class of watermark images, which we aim to classify. To this end, each template image is zero-padded to a size of 560×560 pixels and, afterwards, randomly scaled by a factor between 0.75 and 1.25, rotated by an angle between -10° and 10° , horizontally sheared with shear angle between -0.25 rad and 0.25 rad and, finally, translated horizontally and vertically by up to ± 125 pixels in both directions. With this procedure, we create an augmented dataset of two classes consisting of 100 watermarks each.

Fig. 8 visualizes the dataset using principal component analysis (PCA) with dimension 3, for ease of visualization projected onto the plane spanned by the first and third principal direction. The PCA depicted in Fig. 8 (left) uses the canonical representation in the Euclidean space $\mathbb{R}^{560 \times 560} \cong \mathbb{R}^{313600}$, whereas the PCA in Fig. 8 (right) uses our proposed feature representation based on the R-CDT with only one angle $\theta_0 = 0^\circ$, leading to a much lower-dimensional representation in \mathbb{R}^{794} , referred to as R-CDT space, which amounts to a reduc-

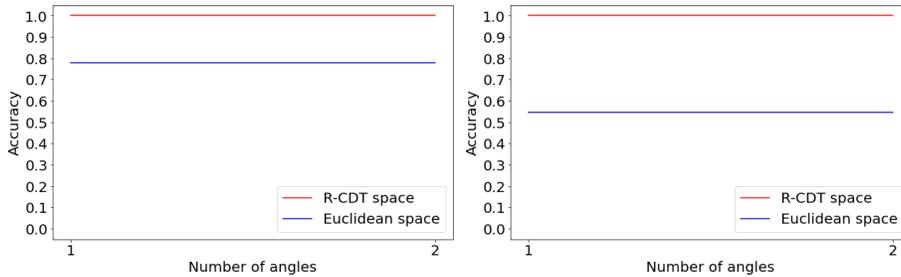


Fig. 9: Test accuracy given two-class-problem with hard-margin SVM for 50 (left) and 5 (right) training samples in Euclidean vs. R-CDT space

tion of approximately 367 times. We clearly observe that the classes are linearly separable in R-CDT space, but *not* in Euclidean space.

To confirm this observation, we trained hard-margin support vector machines (SVMs) with linear kernel (cf. [18]) to solve this binary classification task. The test accuracy is shown in Fig. 9, where we use 50% and 5%, respectively, of the watermarks for training and the complement for testing. As expected, we observe that in Euclidean space the SVM is not able to separate the classes, whereas in R-CDT space the SVM yields a perfect result, both using only one or two angles $\theta \in \{0^\circ, 90^\circ\}$. Moreover, while the classification result degrades in Euclidean space when decreasing the number of training images from 50 to 5, the classification is still perfect in R-CDT space.

In an additional experiment, simulating the standard use case of similarity-based search of watermarks in an available database like WZIS [14], we consider the binary classification task for two distinct classes of visually similar watermarks from different manuscripts of the DRACMarkS dataset [15]. The reason for simulations is the trivial fact that existing watermark databases do currently not contain our novel R-CDT representation. As of now, we identified 20 instances of both the blazon and lily watermark depicted in Fig. 7, which we binarized and manually cropped to 280×280 pixels, cf. Fig. 10, leading to a test dataset of in total 40 binarized DRACMarkS watermarks. The classification is based on five thermograms^{12,13,14,15,16} of each the blazon and lily watermark from WZIS, which we also binarized using our PoC and cropped to 280×280 pixels. Thereon, each image is randomly scaled by a factor between 0.75 and 1, rotated by an angle between -5° and 5° , horizontally sheared with shear angle between -0.25 rad and 0.25 rad and, finally, translated horizontally and vertically by up to ± 25 pixels in both directions to create a training dataset of in total 40 watermarks.

¹² https://www.wasserzeichen-online.de/wzis/?ref=DE0960-Schubert13_5

¹³ https://www.wasserzeichen-online.de/wzis/?ref=DE0960-SchubertF32_4

¹⁴ https://www.wasserzeichen-online.de/wzis/?ref=DE0960-Schubert41_9

¹⁵ https://www.wasserzeichen-online.de/wzis/?ref=DE0960-BeethovenArtaria138_48

¹⁶ https://www.wasserzeichen-online.de/wzis/?ref=DE4795-GoS20_f.4



Fig. 10: Results of classification via R-CDT space for blazon/lily watermarks in binarized DRACMarkS data (green indicates correct whereas red false labeling)

For solving the binary classification task, we use our proposed R-CDT feature representation with two angles $\theta \in \{0^\circ, 90^\circ\}$ followed by a hard-margin SVM with linear kernel. The predicted labels of the DRACMarkS test watermarks are shown in Fig. 10, where we achieve an accuracy of 92.5%.

For comparison, we also trained soft-margin SVMs with Gaussian kernel (cf. [18]) based on the conventional Euclidean representation of watermarks, thereby relaxing the strict condition of linear separability in Euclidean space. In this case, the predicted label is constantly ‘blazon’ so that the accuracy is only 50% and, hence, SVM classification in Euclidean space fails.

5 Conclusion

Starting out from our systematic SOTA study and following a theory-driven system engineering approach, we proposed a novel E2E processing pipeline as a PoC and validated it through initial well-designed experiments with thermography data of varying quality. Key design criteria of our concept were the use of raw thermography data, transparency and understandability of selected IP/PR methods, ease-of-use, flexibility in chaining of methods, mastering of complexity via a controllable parameter regime, and definition of baseline performance facing the lack of benchmarks. On the one hand, our PoC clearly demonstrates its validity while on the other hand it uncovers current limits (e.g. spurious short

line segments) to be alleviated by improved methods. Furthermore, our results documented their dependence on image quality – a trivial fact that will spur further research on IR sensor characteristics including noise modeling, IR-paper interaction, IR signal detection etc. Regarding prevalent digitized watermark hand tracings as well as current Deep Learning (DL) research, we also propose to scrutinize the option of training appropriate DL models for single tasks in the processing chain, e.g. binarization. Moreover, an in-depth analysis of generalizability and scalability of our concept is to be conducted w.r.t. a concerted thermography lab set-up and a larger test dataset from various archives. In closing, the watermark research community is well-advised to develop an interdisciplinary roadmap towards computational filigranology.

Acknowledgments. Partial funding of our research by the Austrian Academy of Sciences within the Heritage Science Austria project DRACMarkS is gratefully acknowledged. We are indebted to the team members of the DRACMarkS project, as well as Dr. E. Frauenknecht (WZIS, Stuttgart) and Dr. M. Rebmann (SBB, Berlin) for their support. Moreover we are grateful to DFG for funding the invitational roundtable on “Thermography Digitization of Watermarks in Music Manuscripts” end of October 2023 in Vienna. Also we feel obliged to appreciate constructive reviews. Finally we devote our paper to pioneer Piero Zamperoni (1939-1998).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Adler, J., Kohr, H., Ringh, A., Moosmann, J., Ehrhardt, M., Lee, G., Verdier, O., Karlsson, J., Palenstijn, W., Öktem, O., Chen, C., Andrade Loarca, H., Lohmann, M.: odgroup/odl: ODL 0.7.0 (2018). <https://doi.org/10.5281/zenodo.1442734>
2. Alaei, A., Bui, V., Doermann, D., Pal, U.: Document image quality assessment: A survey. *ACM Computing Surveys* **56**(2), 1–36 (2023)
3. Althenhöner, R., Berger, A., Bracht, C., Klimpel, P., Meyer, S., Neuburger, A., Stäcker, T., Stein, R.: DFG-Praxisregeln "Digitalisierung" - Aktualisierte Fassung 2022. Published in (2023). <https://doi.org/10.5281/zenodo.7435724>
4. Canny, J.: A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8**(6), 679–698 (1986). <https://doi.org/10.1109/TPAMI.1986.4767851>
5. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*. Pearson (2018)
6. Hauser, D.: From Digital Thermography Images of Music Manuscripts towards Binarized Watermarks: A Proof of Concept. Master’s thesis, Universität Hamburg, Department of Informatics (2023)
7. Hauser, D., Kassens, C., Stiehl, H.S.: On learning-free detection and representation of textline texture in digitized documents. In: *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM)*. pp. 203–212 (2022). <https://doi.org/10.5220/0010801300003122>
8. Hiary, H.: Paper-based watermark extraction with image processing. Ph.D. thesis, University of Leeds (2008), available at <https://etheses.whiterose.ac.uk/1355/1/hazem.pdf> [last accessed: 04.09.2023]

9. Hiary, H., Ng, K.: Watermark: From paper texture to digital media. In: First International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS'05) (2005). <https://doi.org/10.1109/AXMEDIS.2005.50>
10. Hiary, H., Ng, K.: A system for segmenting and extracting paper-based watermark designs. *International Journal on Digital Libraries* **6**, 351–361 (2007)
11. Hufnagel, S., Rastonis, V., Fuchs, R., Herzog, A.: Paper trails: Hyperspectral imaging of watermarks in Icelandic paper. In: *Care and conservation of manuscripts*. vol. 18, pp. 213–226 (2021)
12. Hufnagel, S., Sigurðardóttir, Þ., Ólafsson, D.: Paper Stories – Paper and Book History in Early Modern Europe. De Gruyter (2023). <https://doi.org/doi:10.1515/9783111162768>
13. Kolouri, S., Park, S.R., Rohde, G.K.: The Radon cumulative distribution transform and its application to image classification. *IEEE Transactions on Image Processing* **25**(2), 920–934 (2016). <https://doi.org/10.1109/TIP.2015.2509419>
14. Landesarchiv Baden-Württemberg: WZIS - Wasserzeichen Informationssystem (2010), available at <https://www.wasserzeichen-online.de/wzis/index.php> [last accessed: 04.09.2023]
15. Loose-Einfalt, K., Koliander, G., Lindmayr-Brandl, A., Gubsch, C., Gulewycz, P., Peterlechner, M.: Digitized watermarks in the music manuscripts of Franz Schubert (2023). <https://doi.org/10.5281/zenodo.8305575>
16. Matas, J., Galambos, C., Kittler, J.: Robust detection of lines using the progressive probabilistic Hough transform. *Computer Vision and Image Understanding* **78**(1), 119–137 (2000)
17. Meinschmidt, P., Märgner, V.: Multispectral image archiving of watermarks in historical papers. In: *IS&T Archiving Conference*. pp. 92–96 (2011). <https://doi.org/10.2352/issn.2168-3204.2011.8.1.art00023>
18. Mohri, M., Rostamizadeh, A., Talwalkar, A.: *Foundations of Machine Learning*. MIT Press (2018)
19. Movellan, J.R.: Tutorial on Gabor filters. Tech. rep., Institute for Neural Computation, University of California San Diego (2002), available at <https://inc.ucsd.edu/mplab/75/media//gabor.pdf> [last accessed: 04.09.2023]
20. Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K.: Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing* **39**(3), 355–368 (1987)
21. Pondenkandath, V., Alberti, M., Eichenberger, N., Ingold, R., Liwicki, M.: Identifying cross-depicted historical motifs. In: *16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. pp. 333–338 (2018)
22. Rauber, C.: *Acquisition, archivage et recherche de documents accessibles par le contenu: application à la gestion d'une base de données d'images de filigranes*. Ph.D. thesis, University of Geneva (1998). <https://doi.org/10.13097/archive-ouverte/unige:142942>
23. Rieger, T., Phelps, K.A., Beckerle, H., Brown, T., Frederick, R., Mitrani, S., Breen, P., Breitbart, M., Williams, D., Triplett, R., Horsley, M.: *FAGDI Technical Guidelines for Digitizing Cultural Heritage Materials: Third Edition* (2023), available at <https://www.digitizationguidelines.gov/guidelines/digitize-technical.html> [last accessed: 04.09.2023]
24. Rückert, P., Hodeček, S., Wenger, E.: Bull's head and mermaid: The history of paper and watermarks from the middle ages to the modern period.

- Booklet and catalogue of the exhibition presented by the Landesarchiv Baden-Württemberg, Hauptstaatsarchiv Stuttgart and the Austrian Academy of Sciences, Kommission für Schrift- und Buchwesen des Mittelalters, Vienna, Stuttgart and Vienna (2009), available at https://bernstein.oeaw.ac.at/twiki/pub/Main/ProjectExhibitions/bernstein_2009_book_en.pdf [last accessed: 18.01.2024]
25. Serra, J.: Image Analysis and Mathematical Morphology. Academic Press (1982)
 26. Shen, X., Pastrolin, I., Bounou, O., Gidaris, S., Smith, M.H., Poncet, O., Aubry, M.: Large-scale historical watermark recognition: dataset and a new consistency-based approach. In: 25th International Conference on Pattern Recognition ICPR. pp. 6810–6817 (2020). <https://doi.org/10.1109/ICPR48806.2021.9412762>
 27. Whelan, P.F., Soille, P.: Watermark extraction in paper samples. In: Optical Engineering Society of Ireland & Irish Machine Vision and Image Processing Joint Conference. pp. 287–299 (1998)
 28. Whelan, P.F., Soille, P., Drimbarean, A.: Real-time registration of paper watermarks. *Real-Time Imaging* **7**(4), 367–380 (2001)
 29. Zamperoni, P.: Wasserzeichenextraktion aus digitalisierten Bildern mit Methoden der digitalen Bildsignalverarbeitung. *Das Papier* (Darmstadt) **43**(4), 133–143 (1989)