Text Removal Using Wavelet Transform and Morphological Operations

An Approach for the Removal of Text and Ink Artifacts from Historical Watermark Images

Diana Banță[1]

Supervisor(s): Dr. Martin Skrodzki[1]**, Dr. Jorge Martinez Castaneda**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: Diana Banță
Final project course: CSE3000 Research Project
Thesis committee: Dr. Martin Skrodzki, Dr. Jorge Martinez Castaneda, Dr. Christoph Lofi

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Abstract

Watermarks have an essential role in identifying the origins and age of specific documents. However, this is often a laborious process. One of the main issues in automatic watermark segmentation is the presence of text that obstructs it, making it difficult to properly reconstruct a watermark. Image processing and machine learning techniques face limitations, requiring time, training of data, or manual parameter selection. This research introduces a new method using wavelets transform to locate and remove text from a watermarked image, while preserving the underlying watermark. This method manages to outperform classic image processing techniques for the case when text is thicker than the watermark outline.

1 Introduction

Watermarks are images or symbols embedded in paper that can be observed by directly shining light through the sheet of paper. They are used by historians and researchers to identify the origins of historic documents or works of art [1]. In order to be able to use a watermark to identify a document, it is important to be able to retrieve as much of its original shape and form as possible, as these are its recognizing features. This is not always possible however, since the paper may be old, or may contain artifacts such as tears and folding marks.

Historic watermarks were embedded in laid paper, using manually crafted molds [1]. These molds, in addition to the watermark shape, presented vertical and horizontal lines which were imprinted on the pa- per together with the watermark. One such example can be seen in Figure 1a.

Additionally, one major problem is the presence of text on the paper, which occludes the original form of the watermark and leads to loss of information from it. Such examples can be seen in Figures 1a and 1b, where the text makes it difficult to see the entire shape of watermark. Therefore, it is important that special care is taken when creating algorithms for segmenting watermark images with text, since losing too much of the underlying watermark could make it impossible to recognize the watermark and retrieve information on given documents.

Since automation is needed to speed up and ease the process, multiple algorithms for watermark retrieval were developed. Previous work includes extraction and segmentation of watermarks from individual papers, but few work exists for an integrated recognition system. Such system could receive an image of a watermark and then return similar watermarks from large databases. This would help historians with understanding the context and meaning of watermarks without having to manually search through vast collections.

This research builds upon a watermark recognition system in its initial stages that was developed with the previously mentioned goal [2]. It seeks to improve over past work by proposing an approach that combines wavelet transform and morphological operations for the process of text removal. This idea was inspired by the promising results this system has obtained when adapting the approach in [3] for line removal.

Consequently, the main question this research seeks to tackle is:

- How effective is the joint use of wavelet transform and morphological operations in the removal of text from watermark images, and how does it compare to algorithms using morphological operations and contrast enhancement?

This question can be split into several sub-questions that will aid in reaching a conclusive answer, specifically:

1. How efficient are existing approaches involving morphological operations and contrast enhancement in identifying and removing text from the watermarked images?
2. Under which conditions could wavelet transforms be used to identify and remove text from the watermarked images?
3. How does an approach involving wavelet transforms and morphological operations perform compared to an approach involving morphological operations and contrast enhancement in identifying and removing text from the watermarked images?

The results are evaluated in relation to different watermark segmentation approach, the only one that explicitly addresses handwritten text removal in watermark images by using image processing techniques [4]. Additionally, the results are computed using watermark images provided by the German Museum of Books and Writing [1].

The performance is assessed based on criteria such as original watermark conservation and quantity of text removed successfully. To allow for appropriate evaluation for each criteria, the dataset to be used consists of synthetically generated images of watermarks with different amounts of overlapping ink text. This gives control over variables while also simulating a realistic instance of watermark images.

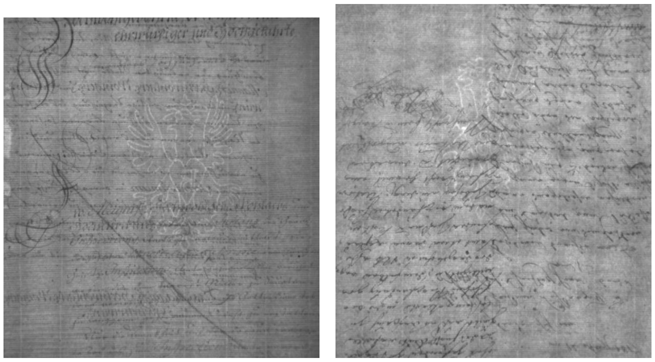
The results obtained in this manner show very good results, especially for processing images with thick text that obstructs the watermark. This is a promising result, since most other systems face limitations for such images.

2 Background

Retrieving watermarks from images can oftentimes be difficult due to factors such as paper irregularities or stains. Additionally, it can be the case that the watermark is overlapped by text or ink. In the context of historic watermark recognition, text represents scripts handwritten in black ink. Text can be deteriorated to the point of illegibility, and it may fully or partly overlay a watermark embedded in paper.

The images considered in this research consist of watermarks partly or fully intersected by text, with different degrees of visibility. The watermark can be placed in different parts of the image and may be of different sizes and intensities. The same applies for the overlapping text. Examples of such watermark images can be observed in Figure 1a and 1b.

[1] https://www.dnb.de/EN/Ueber-uns/DBSM/dbsm_node.html#sprg315370

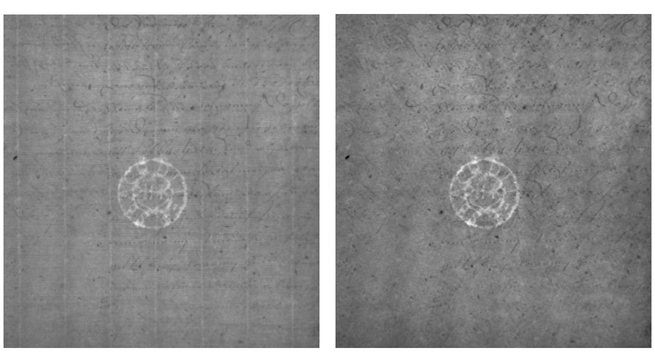


(a) (b)

Figure 1: (a) Watermark image fully overlapped by text, (b) Watermark image partially overlapped by text

While Figure 1a is fully covered by text, Figure 1b is more difficult to distinguish due to the intensity and thickness of the text. Even as a human, it is not immediately clear that the two images are not of the same watermark. It can be observed that the necks of the birds have different thickness and orientation, and the two B's on the birds' chests have different shapes. For this example, it was enough to observe these differences. However, if these details were missing, it would have been far more difficult to tell the difference between the two watermarks. Since every detail matters, it is very important to preserve as much as possible from the original watermark when trying to eliminate unwanted image artifacts. Therefore, it is important to carefully select algorithms that are capable of achieving such things.

Figure 2a shows a different watermark image, where text is present, but its intensity is not enough to hinder the watermark. However, vertical lines are clearly visible at equal distances, and horizontal lines can also be observed at a closer look. These are not part of the watermark, so they should be eliminated, ideally, but they intersect the watermark right through the middle.



(a) (b)

Figure 2: (a) Watermark with visible lines, (b) Watermark after line removal

Figure 2b shows the same image after line removal with the combined wavelet-Fourier approach has been applied [3]. It can be observed how the vertical line intersecting the watermark through its middle part in Figure 2a has disappeared, as well as all the other vertical and horizontal lines in the image, however, the watermark itself has not been altered.

Wavelets are wave-like oscillations or functions used to decompose signals into multiple scales, allowing analysis at multiple levels of detail [5]. They are often used in image compression, as they are able to effectively approximate the image at different resolutions. Since they can capture contrast and intensity differences in the image, they are useful for isolating certain regions of an image. Wavelet transform decomposes the image into three coefficients, namely vertical, horizontal and diagonal, as well as an image approximation at that level. These coefficients store information regarding the intensity changes occurring in the image in different directions on different scales [6].

There are plenty of other methods used for image denoising, including *morphological operations*. These are operations applied to an image in order to adjust its pixels based on its neighbors. This leads to changes in contrast, intensity, or image sharpness [7]. Relevant for this paper are *dilation* and *erosion* operations, which aim to add, respectively remove pixels from object bounds within an image. This depends on the size and shape of the *structuring element* being used, and which is superimposed on the image. The effect of dilation is that of expanding the object. Erosion is the opposite operation, its effect being of shrinking a given object within an image [5].

Contrast stretching is another image processing operation that has as purpose to expand the pixel intensity ranges within an image in such a way that it spans the entire range, while maintaining the relative differences between intermediate values [7]. The following section presents previous work done using these concepts in the context of text removal from historic watermark images, and the existing limitations.

3 Related Work

Some proposed algorithms make use of machine learning techniques [8], [9]. A major drawback in such algorithms, however, is the need for additional data. Specifically, in order to process a watermark, an already processed image of the watermark is required. Since this data is not always available, this type of approach becomes infeasible.

Several other approaches were developed relying on image processing techniques. An overview on different such contributions was given [4], also explaining the drawbacks and limitations of these approaches. Specifically, most algorithms face difficulties with processing images containing ink or text, but also with removing the vertical and horizontal lines that are caused by the process of imprinting a watermark on paper.

The same thesis proposed a bottom-up approach consisting of morphological operations and contrast enhancement techniques that aimed to successfully segment watermarks in these cases [4].

The thesis introducing this approach includes another alternative method. However, this method was avoided as it consists of machine learning techniques requiring training of data.

The bottom-up method starts with an unprocessed image, and ends with a binarized output of the segmented watermark. It consists of foreground interference removal and background estimation.

Foreground interference has the goal of removing the interference caused by the overlapping ink by means of contrast stretching and morphological dilations and erosions. The most important process in this part is that of deriving the ideal size of a structuring element used to remove the foreground interference without blurring the watermark underneath. This structuring element is then used for dilation and erosion of the image, this image being returned and used for background estimation.

Background estimation is computed by means of sequential contrast stretching and sequential morphological opening operations on the image obtained previously. An essential element is that it uses the image intensity to derive an ideal parameter to use for these operations. The concrete details on the implementation for this part are elaborated on in the paper. The final result for the overall algorithm is the difference between the image obtained in foreground interference and the one resulted from background estimation.

However, the approach faced limitations when given watermarks that were not clearly visible, as well as images with thick, dark interference.

One major drawback of image processing techniques, and especially morphological operations, is that they tend to alter entire regions in the image. As a result, details in the watermark design may be lost as well when applying such operations. Since the watermark may be difficult to distinguish even in the original image, it is very important to keep it as unchanged as possible during this process. Subsequently, even though the approach in [4] outperforms other related work, limitations in the extent of applicability and efficiency still exist. Therefore, it would be beneficial to find faster techniques that can address these issues.

A first version of an automatic watermark recognition system that uses exclusively image processing techniques has been developed, with promising results [2]. The system was created using digitized watermarks provided by the German Museum of Books and Writing. The watermarks within this dataset are similar with the ones used in previous work that originate from the Bernstein Project [10].

This system performs watermark segmentation in four steps, specifically pre-processing, denoising, thresholding and post-processing. The last two steps consist of binarization of the segmented watermark, which is not relevant for text removal.

Pre-processing only involves inverting the image in order to make the watermark appear darker than the background and interference. Then denoising consists of sharpening, line removal, contrast stretching, followed by BM3D and Kuwahara filtering. The motivation and implementation details for this algorithm can be found in the source paper [2].

While this watermark recognition system faces similar issues as mentioned previously, it was noticed that the combined wavelet-Fourier transform approach presented in [3] proved highly effective in removing the horizontal and vertical lines from a given image while maintaining the quality of the underlying watermark. This method works by repeatedly applying Fourier transform in the wavelet domain, and then blurring particular corresponding coefficients using Gaussian blur. Then the image is reconstructed by using the refined coefficients from the wavelet space.

This method has inspired the question of how efficient such an approach would be in other scenarios compared to the already existing image processing approaches. Since no experimental work has been done yet with this idea for text removal, adapting and applying this concept could possibly show new and efficient ways in which the current limitations can be addressed.

Given the high flexibility and scalability of wavelet transforms [7], and the general property of text pixels having lower intensity values than the rest of the image, wavelet transform could be used to isolate these pixels. Once isolated, morphological operations could be applied locally on these regions to remove the text from a given image, without altering the underlying watermark. Such an approach could be an important step in improving denoising of images in general cases, without needing to employ machine learning approaches for increased accuracy.

4 Methodology

Two algorithms have been implemented and tested using a synthetically generated dataset. The first algorithm is an adaptation of a method making use of morphological operations and contrast stretching to remove text from watermark images [4]. This is the only such approach that explicitly tackles text in the context of watermark images, and it claims good results.

Alternatively, the approach introduced by this paper has been inspired from the concept of the algorithm in [3]. The algorithm was used with the purpose of line removal, with good results and high watermark preservation (Figure 2b). This is possible due to the application of Fourier transform within the wavelet domain in order to blur the coefficients corresponding to vertical image lines. *Fourier transforms* are image representations as sine and cosine waves of different amplitudes and frequencies. They are capable of identifying periodic signals and patterns within the image [5].

However, the scope of this paper is text removal, which differs from lines in intensity values, watermark obstructing levels, and absence of periodicity. Since applying Fourier transform seems to be less effective in the process of text removal, its use in text localization could represent a matter for future work, and stay out of the scope of this paper. However, the idea of applying operations in the wavelet domain could still be of interest. This is because wavelets achieve high detail preservation and access, and are effective in cases of non-periodic artifacts. [5]

Subsequently, the new method using wavelet and morphological operations is implemented and tested against the algorithm proposed in [4]. In order to do this, a synthetic dataset of watermark images has been created by overlapping images of laid paper, watermark outlines, and handwritten text images. The performance of the two algorithms is assessed by means of peak signal-to-noise ration, structural similarity index, and mean square-error [11], [12].

This section covers the approach used to generate a synthetic dataset for training and testing the algorithms, followed by a breakdown of each algorithm to be evaluated. Lastly, the evaluation metrics and procedure are detailed.

4.1 Dataset Creation

In order to have access and control over text and watermark variables, a synthetic dataset of images was created. This seeks to imitate the general appearance of an image with a watermark overlapped by text, while also having access to information needed to assess the performance of the text removal. In order to achieve this, three datasets of different types of images were used: one with background image, one with watermark images, and one with handwritten text. To make sure that an image generated from these elements is a relevant imitation of a non-synthetic image, the data used for each of these datasets was selected to resemble individual parts from the non-synthetic images.

The dataset of background images consists of background images from the watermark dataset provided by the German Museum of Books and Writing, cropped to exclude the existing watermark. These images present different levels of general noise, artifacts, and relative pixel intensity (Fig. 3).

The dataset of watermarks was created from random binarized watermak images from the same database. The binarized watermarks have different levels of thickness and detail, which ensures that they can resemble both images where the watermark is very easy to observe, but also images where the watermark quality is highly altered (Fig. 4).

The text images were obtained from three public datasets of handwritten documents images, two of which consist of historic handwritten documents [13], [14], [15]. These images were binarized as well and had their background set as transparent. However, since text is darker, the foreground consisted of black pixels, while the background had the white ones. Since binarization removed different amounts of data from the original images, the remaining text is of different thickness and detail preservation, which helps to account for various types of text found in watermark images (Fig. 5).

After creating the datasets of backgrounds, watermarks and text, the synthetic dataset was generated by randomly picking one image from each dataset, and overlapping them. This was done by resizing the selected watermark image instance by a random scale, then reducing its transparency by a random factor. This ensures that the watermark has a random visibility level within the image. Lastly, a random set of coordinates was selected within the background image, and the watermark was overlapped on that image area. The same process was used for the randomly selected text image, but its placement was selected to be within a range that overlaps the watermark by at least 25%.

A text file was created containing information about the images used to generate each of the synthetic images. For each watermark and text image, this information included their respective overlaying coordinates, as well as their scaling and transparency factors.

Each image in the dataset consists of a watermark overlapped partly or fully by text with thickness higher or lower than its own, on a paper region with arbitrary intensity characteristics. Thus, the synthetic dataset generated following this approach is extremely diverse in contents, accounting for a very wide range of watermark images overlapped by text. Figure 6 presents an example of a synthetic watermark generated by applying this process for the components in Fig. 3-5. Additionally, Figure 9 shows an example of a non-synthetic watermark image for comparison.

4.2 Implementation of Baseline Algorithm

The first algorithm that was implemented consisted of an adaptation of an approach using morphological operations and contrast enhancement techniques [4]. This algorithm was chosen as it is the only approach explicitly addressing text removal in watermark retrieval, and it claims high performance when dealing with such cases.

In the original paper, the algorithm consists of two parts: foreground interference removal and background estimation. Since the focus is on text removal, only the first half of the algorithm has been reproduced, as it is the only one addressing this issue.

Foreground removal is done by performing contrast stretching on the initial image and approximating the gray level value with the same amount of zero pixels as the percentage of black pixels in the stretched image. Then dilation is applied with increasing structuring element sizes until there are no pixels with intensity value lower than the the gray level previously derived.

However, some adaptations were made in order to be able to reproduce the results from this approach as best as possible. This is because the original document does not explicitly state the size of the final erosion operation from foreground interference. Not knowing all the values of parameters being used leads to erratic behavior for some inputs. Thus, through trial and error it was found that the most similar outputs to the examples in the original paper were found by choosing an erosion size lower by 2 than the computed dilation structuring element size.

Nevertheless, this raises a new issue, namely in the case when the approximated dilation size is lower or equal than 3. In those cases, the erosion with structuring element of lower size either does not have any effect on the image, or will throw an error.

A solution to this issue was found by applying another dilation with a structuring element of size increased by 1, and then an erosion with the same structuring element.

4.3 Implementation of Wavelet-Morphology Algorithm

The algorithm consists of two parts: text localization and text removal.

Text localization makes use of the general intensity properties of text and watermark. Namely, the text is usually one of the darkest parts of the paper, while the watermark generally has high pixel intensities. Since text tends to be darker even than the background, there are sudden changes of in-

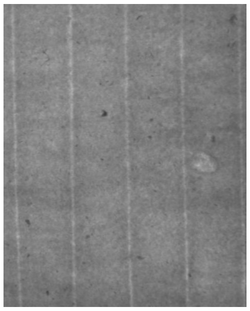


Figure 3: Example of background image

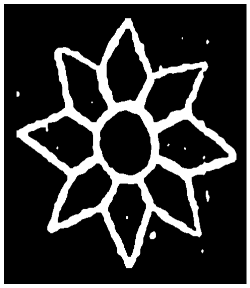


Figure 4: Example of watermark image

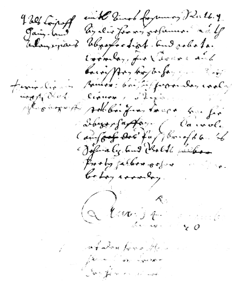


Figure 5: Example of handwritten text image

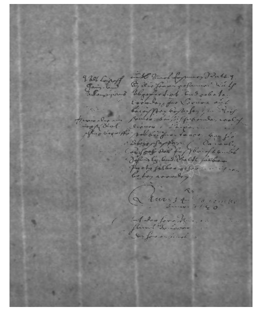


Figure 6: Example of synthetic watermark image created from Fig. 3-5

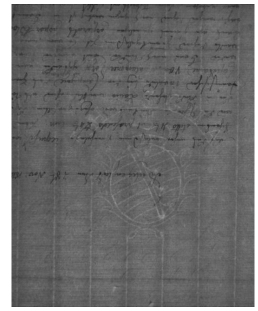


Figure 7: Example of non-synthetic watermark image

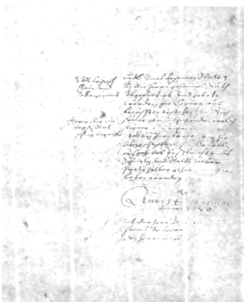


Figure 8: Image approximation in wavelet domain.

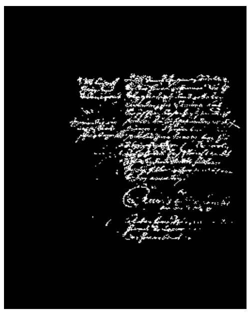


Figure 9: Candidate text pixel intensities in Figure 6.

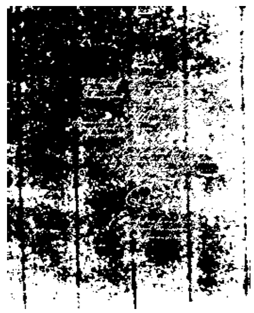


Figure 10: Pixel intensities above mean intensity value.

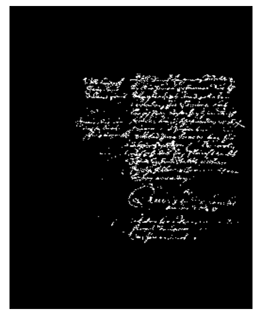


Figure 11: Intersection of the images from Figure 9 and Figure 10.

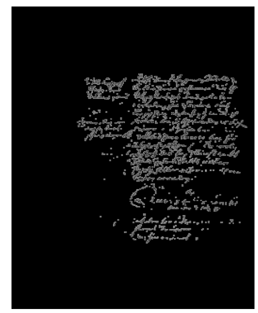


Figure 12: Image containing just the text pixels from Figure 6.

tensity where text intersects the background. Thus, wavelet decomposition can be used to identify these locations.

This is done by using the approach presented in [16], of identifying candidate text pixels within the image. Explicitly, a new image is created as the average of squared intensities from all the wavelet coefficients at that decomposition level. The mean intensity and standard deviation of this image are obtained and summed. Then, a binarized image is created consisting only of the pixels with intensities larger than this sum (Figure 9).

However, since the watermark is generally lighter than the background of the image, it is possible that the binarized image contains pixels from the intersection between watermark and background. Thus, a new binarized image is created by eliminating the pixels with higher intensity values than the average intensity value of the image approximation (Figure. 8) within the wavelet domain. An example of such an image is given in Figure 11.

Lastly, since the goal is to isolate the text and not interfere with the watermark, a mask is created by intersecting the two binarized images computed before. This way, the locations with high contrast are kept, but only where the pixels have low intensity values. The results computed in this manner can be seen in Figure 12.

Extracting the text is then performed by creating a binary mask for the original image. In this mask, all pixels are set to zero except for those identified in the high-contrast, low-intensity regions from the previous step, which are set to 255. Applying the mask keeps the zero pixels unchanged while setting all the other pixels to the intensity values of the corresponding pixels from the original image.

This mask image consists of text pixels with low intensity values and the neighboring background pixels which have higher values. The intensity distribution can be observed more clearly by computing the histogram of the mask segment. Two peaks are displayed in the histogram, corresponding to the most frequent intensity values within the image. In order to separate text from background, a threshold intensity value is chosen between the two peaks. This value is calculated as the average of the intensity values at the two local maxima of the peaks.

The next step is reducing the contrast between text and background. This is done by replacing the text pixels with the average intensity value of the surrounding pixels from the background within the mask. This value is computed by first applying the threshold value obtained earlier to set all text pixels to zero. However, it may be the case that the mask does not cover enough pixels from the background to be able to approximate its intensity value accurately. For this reason, the surface of the mask is increased by means of dilation, and the contents of the mask are set to the corresponding values in the original image.

Nevertheless, it may still be the case that the mask contains more text pixels than background pixels. This is verified by computing the ratio between the number of background pixels and the total number of non-zero pixels in the mask. If the mask has less than 50% background pixels, computing the average from the neighboring pixels may lead to an inaccurate approximation of the background intensity. Thus, in such cases, the average is computed from the entire initial image.

Having the locations of text pixels and the value they should be reassigned to, there is only one thing that needs to be considered: There are still pixels at the intersection between text and background that are too dark to be background and too light to be text. Additionally, the image may contain areas of different intensities, so assigning one value for all text pixels may create new areas of high contrast in the image.

Therefore, special care needs to be taken in order to ease these changes in contrast. For this, each pixel within the mask is assigned a value considering its current intensity value. Specifically, pixels with high intensity values could be part of the watermark, so should be preserved more than pixels of high intensities, which should be fully changed. However, high pixel intensities may also correspond to artifacts, and so by keeping their values the same it may happen to accidentally create more artifacts, which may be detrimental for segmenting the watermark. Therefore, the pixel intensity values are computed as shown in Algorithm 1.

Algorithm 1 Adjust Pixel Intensity for Each Pixel in the Image Mask

Input: *image_mask* the mask of the image
mean_val the mean intensity value of the background
max_mask the maximum intensity value within the mask
threshold the threshold value for segmenting the text

for each pixel in *image_mask* **do**
  if $pixel\_val > \frac{mean\_val + max\_val}{2}$ **then**
    $pixel\_val \leftarrow \frac{2 \cdot pixel\_val + mean\_val}{3}$
  else if $pixel\_val > threshold$ **then**
    $pixel\_val \leftarrow \frac{pixel\_val + mean\_val}{2}$
  else
    $pixel\_val \leftarrow mean\_val$

This algorithm makes the transition from text to background smoother by modifying the pixel values in a more gradual manner. However, there could still be the case of having pixels of very different intensity values near each other. Thus, in order to make the transitions between each neighboring pixel even smoother, dilation is applied for the computed mask. The last step consists of replacing the text pixels with the values just obtained.

Since the intersection between text and background may happen more gradually, and may imply a small change of contrast in multiple neighboring pixels, a thresholding value is computed to also include pixels with values close to the old threshold. Since the overall contrast of an image can vary from case to case, the new threshold is chosen depending on the old value. Specifically, the new threshold is equal to the minimum between the old value increased by 12.5% and the old value increased by 15. This is done in order to keep the threshold low enough to properly segment text regions, while also accounting for the differences in image intensity values.

The resulting image after applying all the steps of the proposed algorithm can be observed in Figure 13.

4.4 Evaluation Criteria

The algorithm is developed using three initial datasets, each containing 50 images of laid paper backgrounds, watermarks, and handwritten text images respectively. From these, 350 synthetic images were generated in batches of size 50, each having different scale and transparency ranges for the watermark and text. These images were used to develop the two algorithms. Subsequently, an additional set of batches were generated for evaluation. The performance of the approaches will be assessed and compared based on two criteria: watermark conservation and quantity of text removed successfully.

The watermark conservation is evaluated by comparing the percentage of pixel intensities that were lost from the initial image, as well as the overall relative contrast the watermark has compared to the rest of the image. This is performed by making use of the coordinates at which the original watermark image has been overlaid on the original background image. This is done to create a mask for extracting the intensity values of the processed watermark.

Assessing the quantity of text successfully removed is done by using the coordinates of the text images as a mask to compare the change in contrast in that area. The corresponding image obtained by overlapping the watermark with the background but without the text is therefore computed. It is used in order to estimate expected contrast and pixel intensities at the intersection between text and watermark from the original image.

In order to measure these aspects appropriately, several metrics are employed. These include peak signal-to-noise ratio (PSNR), which evaluates the overall difference between maximum image signal and background noise, and structural similarity index (SSIM), which measures differences in the general structure of an image [11]. These are particularly useful for evaluating watermark conservation, with higher values representing better results. Moreover, mean square-error (MSE) [17] may be effective when measuring text removal, since it represents the pixel intensity differences between the two images. This value should be lower for better results.

A final score for all of these metrics for each approach is then computed by taking the average values corresponding to each image in the evaluation set.

5 Results

The evaluation was computed using four different types of datasets, each characterized by the relative thickness of text and watermark. Each dataset consisted of 50 synthetic images, except for the 'Very Thick Text' one, which only had 35.

A 'thin text' watermark image is any image where the text is thinner or almost similar to the watermark (Figure 14). These are generally images where the watermark is easily visible to the human eye. 'Very thin text' images consist of text

Table 1: Comparison of Evaluation Metrics of Proposed Approaches. _wav denotes the values of metrics computed for the proposed wavelet algorithm, while _ip denotes the metrics computed for the algorithm from literature. The values in bold correspond to the better score for each metric in each category.

Type of Dataset	**Evaluation Criteria**	**SSIM_wav**	**SSIM_ip**	**MSE_wav**	**MSE_ip**	**PSNR_wav**	**PSNR_ip**
Thin Text	Watermark Conservation	**0.9209**	0.8704	**9.0006**	9.6138	**39.2003**	39.1286
	Text Removed	**0.9354**	0.8973	7.4648	**6.3803**	39.8297	**40.6988**
	Original Image Preservation	**0.9862**	0.9728	**1.6130**	1.7505	47.0520	**47.2130**
Very Thin Text	Watermark Conservation	0.8727	**0.8746**	**17.7264**	17.8141	**36.7704**	36.1226
	Text Removed	0.8846	**0.9197**	13.5218	**10.5727**	37.5582	**38.4590**
	Original Image Preservation	**0.9673**	0.9664	**3.7932**	4.4782	**43.4034**	42.7634
Thick Text	Watermark Conservation	**0.9261**	0.7964	**9.1501**	14.4470	**39.4024**	36.9138
	Text Removed	**0.9394**	0.8370	**7.6228**	10.0531	**39.8929**	38.5258
	Image Preservation	**0.9832**	0.9499	**2.2164**	3.3884	**45.8250**	43.6180
Very Thick Text	Watermark Conservation	**0.8986**	0.7609	**10.1473**	13.8461	**38.4659**	37.3121
	Text Removed	**0.9153**	0.8177	**8.4766**	9.0970	**39.2919**	39.0829
	Total Image Preservation	**0.9792**	0.9498	**2.1388**	2.6673	**45.4478**	44.6648

that is significantly thinner than the watermark. These images contain visible watermarks, but the text may be difficult to distinguish.

The proposed wavelet algorithm performs the worst for these two datasets, and especially for the 'Very thin text' watermarks. Regardless, the SSIM value is still higher for the overall image, indicating that the algorithm did not alter the image too significantly. The Text Removed criterion is the one with the lowest scores. This happens most likely due to the dilation operations performed to the mask with the purpose of including more background in the mask (Figure 15).

Alternatively, the approach in [4] obtains its best results for these types of images. However, the computed results show a clear trend for the other two types of datasets. A 'Thick Text' image is an image where the text is thicker than the watermark, and it overlaps it partly or fully. However, the outline of the watermark is still visible by the human eye. On the other hand, the 'Very Thick Text' dataset contains images where text is either significantly thicker than the watermark, or it fully overlaps it making it difficult to distinguish it, or both (Figure 16). The computed results suggest that the algorithm proposed in this paper is able of processing such instances (Figure 17).

Ultimately, the algorithm introduced in this paper yields good results in processing different types of inputs. It manages to outperform the baseline algorithm for a majority of input data. Since the two algorithms seem to complement each other, since one yields lowest scores when the other achieves its highest, the possibility of combining the two could be considered in the future.

6 Responsible Research

Ethics and reproducibility of work are of utmost importance when conducting any sort of research. To ensure these elements were met throughout this research, several actions have been taken.

First of all, the GitLab repository containing all the code and implementation details has been made available and is linked in this paper [2]. Then, the methods used have been presented with supporting evidence, examples and visual representations. The datasets used for generating synthetic data have been mentioned, and, with the exception of the watermark dataset, are publicly available [13],[14],[15] and licensed under a Creative Commons Attribution 4.0 International License [3]. The former is not publicly available as it has been provided for the purpose of this research by the German Museum of Books and Writing. However, since the scope of this research is on text removal, the watermark symbols could be substituted by binarized images of any simple shape without influencing the results obtained by the algorithm. The same is true for the background images used, any image of an old document could suffice in reproducing the work presented in this paper due to the extensive focus on text and not on the watermark segmentation.

Subsequently, necessary steps have been taken in order to ensure that this research was conducted conforming to ethical aspects, and in a way that can be recreated if needed.

7 Conclusions and Future Work

This paper introduced a promising approach to text removal from historic watermark images. Watermarks are symbols or images embedded in paper that allow historians and researchers to identify the origins of documents. One significant issue hindering the process of watermark segmentation and identification is the presence of paper artifacts and overlapping elements such as ink and text. So far, proposed approaches have faced difficulties and limitations in performing text removal from watermark images, especially when the text obstructing the image is thick.

This paper has presented an approach adopting a concept from a different algorithm and adjusting it to be applied in

[2] https://gitlab.ewi.tudelft.nl/cse3000/2023-2024-q4/Skrodzki_Castaneda/dbanta-Automated-processing-of-scanned-historic-watermar

[3] https://creativecommons.org/licenses/by/4.0/

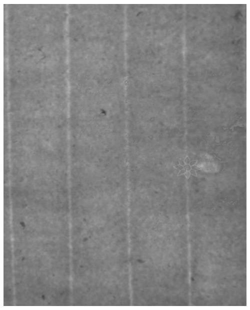


Figure 13: Output after applying text removal for Figure 6.

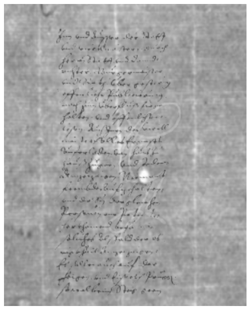


Figure 14: Example of image in the 'Thin Text' dataset.

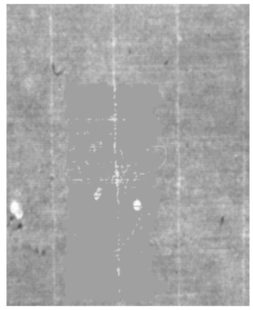


Figure 15: Result of wavelet-based algorithm on Figure 15.

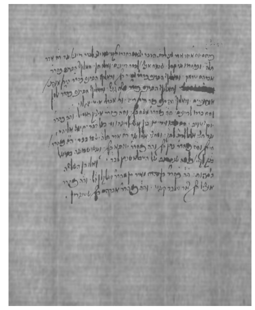


Figure 16: Example of image in the 'Very Thick Text' dataset.

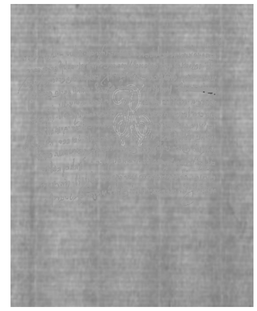


Figure 17: Result of wavelet-based algorithm on Figure 16.

another setting. Specifically, this research has been inspired by the application of a combined wavelet-Fourier transform approach [3] in the context of line removal. This approach proved useful in removing the chain lines of watermarked paper, while preserving the features of the watermark. This was done by applying Fourier transform within the wavelet domain, and blurring out the periodic vertical coefficients. Since text if different from lines by its lack of periodicity, its intensity values, and levels of watermark obstruction, adaptations from this method had to be found.

The most important element, which has been fundamental to this algorithm, was the idea of applying operations within the wavelet domain. This is because wavelets can decompose images and signals with very high detail conservation and access. Thus, instead of identifying the patterns of periodic lines, the algorithm computes the coefficients of the wavelet decomposition, which was inspired by [16] and makes use of intensity characteristics of text for localizing it. Then, morphology is used locally within the identified region, leading to good watermark level of conservation.

This has been compared with another approach that estimates and then removes the text obstructing the watermark by means of morphological operations, dynamic thresholding, and contrast enhancement [4]. This approach is the only that was found to explicitly address and perform text removal reliably only by means of image processing.

The results obtained by comparing the proposed approach with the existing algorithm indicate very good result for processing images where text is thick and heavily obstructing the underlying watermark. However, some limitations exist for processing images with very thin text width, which could be looked more into in future work. Another direction for future work could be the use of Fourier transforms within the wavelet domain in order to improve text localization.

Additionally, the synthetic dataset creation pipeline may be of use in training some future AI-based text removal algorithm. This is because of the wide range of random images that follow the general aspect of watermark images overlapped by text that can be obtained. Nevertheless, the dataset creation could also be extended to include multiple types of text per image and more diverse contrast-noise images. This has been out of scope for this research, but may bring useful contributions if extended in the future.

Nevertheless, the presented system is successful in tackling increased image levels and high text obstruction, outperforming the best image processing approach presented for the context of historic watermark recognition. Therefore, this may bring significant improvements for existing watermark recognition systems, which in turn could be of great aid for historians and researchers.

References

[1] L. Müller, "Understanding paper: Structures, watermarks, and a conservator's passion," *Harvard Art Museums*, May 2021.

[2] D. Banţă, S. Kho, A. Lantink, A. Marin, and V. Petkov, "A watermark recognition system: An approach to matching similar watermarks," 2023, last accessed 28 May 2024. [Online]. Available: http://resolver.tudelft.nl/uuid:e8dfbd63-ae54-4159-b786-d1d8c64dc827

[3] B. Münch, P. Trtik, F. Marone, and M. Stampanoni, "Stripe and ring artifact removal with combined wavelet - fourier filtering," *Opt. Express*, vol. 17, no. 10, pp. 8567–8591, May 2009. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?URI=oe-17-10-8567

[4] H. A. Hiary, "Paper-based watermark extraction with image processing," Ph.D. dissertation, The University of Leeds, School of Computing, 2008.

[5] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 3rd ed. Cengage Learning. [Online]. Available: https://www.researchgate.net/publication/220695728_Image_processing_analysis_and_and_machine_vision_3_ed

[6] M.-S. Song, "Wavelet image compression," in *Contemporary Mathematics*, D. Han, P. Jorgensen, and D. Larson, Eds. American Mathematical Society, vol. 414, pp. 41–73. [Online]. Available: http://www.ams.org/conm/414/

[7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. New Jersey: Prentice-Hall, 2002.

[8] X. Shen, I. Pastrolin, O. Bounou, S. Gidaris, M. H. Smith, O. Poncet, and M. Aubry, "Large-scale historical

watermark recognition: dataset and a new consistency-based approach," *CoRR*, vol. abs/1908.10254, 2019. [Online]. Available: http://arxiv.org/abs/1908.10254

[9] Y. Qi, W. R. Huang, Q. Li, and J. L. DeGange, "Deeperase: Weakly supervised ink artifact removal in document text images," 2020, pp. 3522–3530.

[10] "Bernstein—the memory of papers project." [Online]. Available: https://www.memoryofpaper.eu/BernsteinPortal/appl_start.disp

[11] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th International Conference on Pattern Recognition*, pp. 2366–2369, ISSN: 1051-4651. [Online]. Available: https://ieeexplore-ieee-org.tudelft.idm.oclc.org/document/5596999#full-text-header

[12] H. Sheikh and A. Bovik, "Image information and visual quality," vol. 15, no. 2, pp. 430–444, conference Name: IEEE Transactions on Image Processing. [Online]. Available: https://ieeexplore-ieee-org.tudelft.idm.oclc.org/document/1576816

[13] J. A. Sánchez, "Bentham dataset r0." [Online]. Available: https://zenodo.org/record/44519

[14] J. A. Sanchez, V. Romero, A. H. Toselli, and E. Vidal, "ICFHR2016 competition on handwritten text recognition on the READ dataset," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, pp. 630–635. [Online]. Available: https://ieeexplore.ieee.org/document/7814136/

[15] I. Rabaev, B. K. Bakarat, and J. El-Sana, "The pinkas dataset." [Online]. Available: https://zenodo.org/records/3569694

[16] S. A. Ali and A. T. Hashim, "Wavelet transform based technique for text image localization," vol. 2, no. 2, pp. 138–144. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405609X16300124

[17] C. C. Beckner, Jr. and C. L. Matson, "Using mean-squared error to assess visual image quality," vol. 6313, p. 63130E. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2006SPIE.6313E..0EB