ATHENA: Automatic Text Height ExtractioN for the Analysis of old handwritten manuscripts

Ruggero Pintus
Yale University
ruggero.pintus@yale.edu

Ying Yang
Yale University
ying.yang.yy368@yale.edu

Holly Rushmeier
Yale University
holly.rushmeier@yale.edu

***Abstract*—A massive digital acquisition of huge sets of deteriorating historical documents is mandatory due to their value and delicacy. The study and the browsing of such digital libraries is becoming crucial for scholars in the Cultural Heritage field, but it requires automatic tools for analyzing and indexing those dataset items. We present here a layout analysis method to perform automatic text height estimation, without the need of any kind of manual intervention and user defined parameters. It proves to be a robust technique in the case of very noisy and damaged handwritten manuscripts. The effectiveness of the method is demonstrated on a huge heterogeneous corpus of medieval manuscripts, with different writing styles, and affected by other uncontrollable factors, such as ink bleed-through, background noise, and overtyping text lines.**

I. Introduction

Nowadays, manuscripts are being digitized at an increasing rate. In the Cultural Heritage field this activity becomes much more important, since a huge number of historical documents are deteriorating day by day, and their digital preservation is compulsory due to their value and delicacy. Moreover, a digital collection of such documents represents an invaluable database that woudn't otherwise be available to the public, whether they are experts, tourists or people keen on art. The amount and importance of the information contained in this variety of different language manuscripts results in an increased interest in developing tools to explore, browse and enjoy them in a more comprehensive manner. Digital libraries all over the world have yet to be exploited electronically for consulting, exchange and distant access purposes. However, to produce a fast, electronic, searchable form for a document, it must first be indexed. Some recent works [1], [2] show the utility of hyperlinks to browse digital collections. It stands to reason that both the massive digitization and indexing of thousands of pages require completely automatic tools. Thus the document layout analysis field plays a significant role, being a fundamental step of any document image understanding system.

Although efficient algorithms exist that cope with printed documents, analyzing old handwritten manuscripts poses some more difficult challenges. Those documents are affected by problems of ageing and have looser layout formatting requirements. Further, their physical structure, containing text, capital letters, portraits, ornamental bands and graphical contents, is even harder to extract due to other uncountable uncontrollable factors, such as holes, spots, writing from the verso appearing on the recto (ink bleed-through), ornamentations, background noise, touching text lines, different writing styles, and so forth. It follows that the segmentation of historical handwritten documents is still an open research field, and, to our knowledge, a completely automatic solution has not been presented yet. A common problem is the initial estimation of the text height. While there are some good automatic techniques that find this value for printed documents with clear inter-line spacing, this task grows more difficult as inter-lines become narrower (e.g., in medieval manuscripts). Due to the ease of manually performing this task, some techniques [3], [4], [5] ask the user to input a rough estimation of the text height. However, user intervention becomes infeasible in the case of massive datasets with a high number of different documents and high variability of text size.

We present here a parameter-free, automatic method to perform text height estimation. Given the image of a manuscript page, a multi-scale representation is first produced. Then, for each sub-image at each level, a robust, frequency-based descriptor is computed. Finally, a voting procedure finds the predominant spatial frequency in the document page, whose period is the value of the text height. It proves to be an efficient and reliable technique in the case of very noisy and damaged old handwritten manuscripts. Here we list the major contributions of the proposed approach.

Frequency-based descriptor. A new local image descriptor based on a frequency analysis of the y-axis projected profile of the normalized image autocorrelation function.

Multi-scale framework. A multi-level approach with a voting procedure to exploit spatial consistency between frequency-based image descriptors at different scale levels.

Evaluation. To assess our method, we present an extensive evaluation of the proposed algorithm, applied to a huge heterogeneous corpus content.

II. Related work

Document analysis is one of the most studied fields in image processing. A huge amount of work has been presented to deal with segmentation [6], line extraction [7], char and word spotting [8], [9], and classification of handwritten manuscripts [10], [11]. An exhaustive review is far outside the scope of the paper, and the reader is referred to various recent surveys [12], [13]. Here, we discuss only the techniques closely related to ours.

Integration profiles. Commonly used approaches to determine text height estimation are based on Projection Profiles [14], [15], [16], XY-CUT algorithm [17], and Run Length Smearing Algorithm (RLSA) [18]. They are all based on different ways to directly integrate the original image along

rows, columns or, rarely, diagonal directions. They are based on a priori strong assumptions, and short lines or very narrow lines with overlapping descenders and ascenders will produce a weak signal. While these approaches are mainly used for printed documents, some papers adapted them to handwritten ones with little overlap between lines and moderately skewed texts [19], [20]. Although these solutions are typically faster, they are very sensitive to noise, and not robust enough to be directly applicable to a generic handwritten, possibly damaged manuscript, with generic layout rules, and other irregularities. Further, they are not completely automatic approaches, because, to avoid local minima in the analysis of projection profiles, some manually defined parameters are needed [21], [22].

Local descriptors. Recent works perform handwritten text characterization by extracting orientation-based features, such as a histograms of oriented gradients [23], Gabor descriptors [24], scale invariant features [25], and an autocorrelation function [3], [5], allowing analysis of documents with unspecified layout structure. The main issue is that all mentioned techniques require user intervention, either to train some classifiers [25], or to manually set some parameters that are strictly dependent on the document text height, such as the neighborhood radius in Garz et al. [4] or the size of kernel windows in Mehri et al. [3]. Although these solutions are very robust to noise, manually adjusted parameters limit the range of their applicability, and make them unsuitable for massive, non-homogeneous corpus.

Multi-scale representations. Exploiting a multi-resolution representation and a frequency-based framework is a well-known approach in image analysis (e.g., [26]), which has been applied to a plethora of applications. In the specific field of document layout analysis these methods are typically used to segment document images scanned from newspapers and journals [27], [28]. Recently, Almeida et al. [29] use wavelets to reduce ink show-through noise in scanned images. Joutel et al. [30] presents a multi-level curvelets decomposition of ancient document images for indexing linear singularities of handwritten shapes; it allows for applications such as manuscripts dating, expertise and authentication of its author, style and period.

Our contribution. Instead of relying on projection profiles directly obtained from the original image, we compute the y-axis profile of the normalized autocorrelation function, which is more robust. It is independent of document brightness and contrast, and skewed text. Instead of defining some parameter values to properly deal with local maxima and minima in the profile, we analyzed it by extracting its discrete fourier coefficients, and by estimating the most predominant spatial frequency in a parameter-free manner; this is robust to noise in the projection profile as well. A complete and reliable automatic solution is achieved by integrating this local image representation into a multi-scale framework, where a descriptor is computed at different scale levels. An extensive evaluation is presented, proving the robustness and reliability of the proposed method, and showing that it is well-suited for huge digital libraries with a high variability of layouts, syles and levels of conservation.

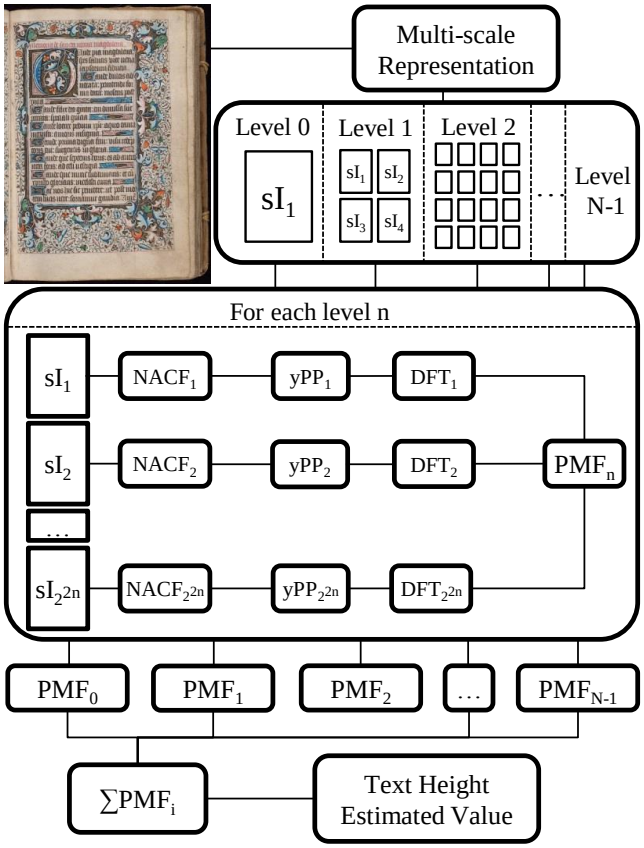


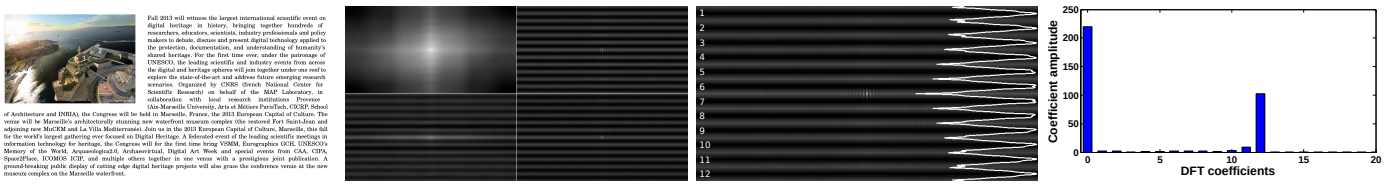
Fig. 1: **Algorithm pipeline.** Given an input image we compute its multi-level representation. After estimating the text height Probability Mass Function (PMF) for each level, we obtained the final estimation by a voting framework across all levels.

III. Technique overview

In Fig. 1 we outline the pipeline of the proposed technique. The algorithm is given an input manuscript image. First we produce a N-level multi-scale representation; at level n, we split each original image in 2^{2n} small sub-images. Then, we analyze these levels separetely. For each of their sub-images we compute the normalized autocorrelation function (NACF), and we integrate this signal to obtain its y-axis projection profile (yPP). We find the main periodicity of the yPP by applying the Discrete Fourier Transform (DFT). We use the information corresponding to the highest DFT coefficient from all sub-images to compute, for level n, an estimation of the text height in terms of probability mass function (PMF). Finally, we exploit the coherence between levels to find the final estimation of the page text height, by accumulating all the PMFs from all levels.

IV. Text height estimation

In this section we explain in detail the proposed technique; for display purpose only, we use the sample image in Fig. 2a to show all the steps of our approach. In general, the input data is unconstrained; the only requirement is that it is an image of a manuscript containing text. It can have figures, ornamentations, capital letters, portraits, touching and overlapping texts, and



(a) (b) (c) (d)

Fig. 2: **Frequency-based descriptor.** Given a sample image (a), for each sub-image at each sub level we compute the normalized autocorrelation function (NACF)(b). We show the NACF integration along x axis to obtain the y-axis projection profile signal of the top-right sub-image at level 1 (c), and its discrete fourier coefficients (d).

can be affected by background noise, ink bleed-through and other kinds of damage due to ageing. The only mild assumption is that either the input text is quasi-horizontal, or that a pre-processing step is applied in order to correct the overall page orientation. This could be easily done by employing the well known Rose of Directions method [31]; here, in section V, we also present an alternative solution to correct orientation. In our case, however, since we use operators that are very robust to skewed texts, we will see how this is a very relaxed constraint, and how typical acquisition setups do not require any alignment correction at all. Hence, to produce the results in section VI, we do not use any orientation correction.

A. Multi-scale representation

First, we compute a multi-scale representation of the input image. Considering a particular level n we split the original image in 2^{2n} small sub-images. The number of levels must be fixed; it must be uncorrelated with the acquisition resolution, and independent of the text height, the layout and the structure of the manuscript page. Since multi-scale analysis is based on consistencies across different levels, the more levels, the more robust is the algorithm. However, given an arbitrary high level value, the probability that a sub-image contains one or more text lines tends exponentially to zero. We experientially found that a 5-level multi-scale representation is a reliable parameter value.

B. Single level analysis

After building the multi-scale representation, we perform a separate analysis of each obtained level. We start by computing the normalized autocorrelation function of each sub-image at that level. The autocorrelation function for a two dimensional signal is defined by:

$$ACF\left(x,y\right)=\sum_{\alpha\in\Omega}\sum_{\beta\in\Omega}I\left(\alpha,\beta\right)I\left(\alpha+x,\beta+y\right) \quad (1)$$

The autocorrelation value at position (x,y) is the sum of the products of the grayscale image values $I\left(\alpha,\beta\right)$ and the pixel values after a translation of (x,y). The normalized autocorrelation function $(NACF)$ is:

$$NACF\left(x,y\right)=\frac{ACF\left(x,y\right)-min_{ACF(x,y)}}{max_{ACF(x,y)}-min_{ACF(x,y)}} \quad (2)$$

where min and max are the minimum and maximum values of the autocorrelation function. Fig. 2b shows the normalized autocorrelation functions of sub-images at level 1. We can clearly see the difference between sub-images that contain text lines or figures.

To extract the spatial periodicity of the patterns that correspond to text regions, we compute the y-axis projection profile of the $NACF$. In Fig. 2c we superimpose the $NACF$ and the profile (white curve) of the top-right sub-image at level 1. We analyze its frequency footprint by computing the Discrete Fourier Transform (DFT) coefficients. After discarding the constant component (i.e., 0-index coefficient), the coefficient with the highest amplitude corresponds to the predominant spatial frequency. In other words, if the maximum amplitude coefficient has index n, it means that the signal has n periods inside the studied domain. After computing the DFT of the profile in Fig. 2c, in Fig. 2d we plot the amplitude of the first 100 coefficients. The $12th$ coefficient has the highest amplitude, i.e., the profile in Fig. 2c has 12 periods. For each sub-image, the size in pixels of that period, obtained by dividing the sub-image height by the number of periods, is a possible candidate value for the text height estimation at that particular level.

Now we have to merge the information of all the sub-images. In a histogram we accumulate the amplitude of the 2^{2n} most relevant coefficients, and the index with the highest amplitude integral is the winner for the current level. For instance, in Fig. 2b, whose amplitudes fall into histogram bin 12, while there is only a single amplitude in bin 1. However, due to the discrete nature of the performed analysis, we do not want to produce a single level value for the text height estimation. On the other hand, for each level n, we prefer to build the following Gaussian probability mass function (PMF) of the text height random variable t:

$$PMF_n\left(t\right)=w_n\times e^{-\frac{1}{2}\frac{\left(t-\mu_n\right)^2}{\sigma_n^2}} \quad (3)$$

$$\mu_n=\frac{th_n^{min}+th_n^{max}}{2},\sigma_n^2=\left|th_n^{max}-\mu_n\right|^2 \quad (4)$$

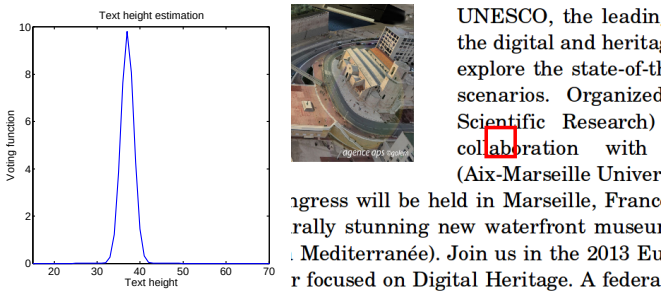
where $th_n^{min}=height_n/\left(i_n+0.5\right)$, $th_n^{max}=height_n/\left(i_n-0.5\right)$, i_n is the winner coefficient index, and $height_n$ is the height of the level sub-image. The level-based normalized weight w_n:

$$w_n=\frac{1}{C_n\times width_n}\sum_{\chi=1}^{C_n}A_n^{\chi} \quad (5)$$

serves to make the PMFs from different levels comparable. C_n is the number of coefficients that contribute to the winner index i_n, A_n^x is one of their corresponding amplitudes, and $width_n$ is the width of the sub-image at level n.

C. Multi-level analysis

The result of the previous step is a bunch of N Gaussian probability mass functions. Each one gives a per-level estimation of the possible text height value for the analyzed page. We want to combine all these PMFs, considering the property that sub-images containing text at different levels produce similar expected values, even if the number of periods or the corresponding amplidutes are different. We compute a voting function by accumulating all the PMFs, and the value t_E corresponding to its maximum is the final text height estimation for the manuscript page. Fig. 3a shows the multi-level voting function obtained by accumulating PMFs from the sample image in Fig. 2a. In Fig. 3b we draw a square with the edge size equal to the corresponding estimated t_E.



(a) Voting function (b) Estimated text height

Fig. 3: **Multi-level analysis.** A voting function is obtained accumulating all the N Gaussian probability mass functions from all levels (a). The height corresponding to the maximum of this function is the estimated text height (t_E). To check the quality of the algorithm outcome, a square with edge size equal to t_E is drawn over the original text (b).

D. Implementation

The direct computation of the autocorrelation function as expressed in equation 1 is computationally inefficient. However, we can use the Plancherel theorem, which allows us to more efficiently express the equation in terms of the image Fourier transform [3]. We have found that the contribution of the level 0 to the computation of the final text height is generally very poor. Since its analysis is the most computationally expensive, by discarding that level we obtain a significant speed up without changing the output result. Based on the properties of the normalized autocorrelation function and its y-axis profile, we can apply an outlier pruning strategy in the single level PMF computation step. Image parts that contain figures don't have a main direction, so their NACF is typically a homogeneous signal with a high value at the center pixel; its profile is a curve with one high central peak, and a decreasing behaviour as a function of $\left|\frac{1}{x}\right|$. In these cases, the index of the most relevant DFT coeffient is 1 (index 0 is the constant coefficient), i.e., one period in the studied domain. Since we are looking for spatial periodicities, we avoid accumulating all these coefficients with index $i_n \leq 1$.

V. Orientation Correction

The algorithm described above works well for images without strong skew. But in a more general case where the lines of the texts within images of scanned documents could have a certain amount of skew, we need to deskew the input images before applying our algorithm. This sub-section describes a simple but efficient pre-processing step that determines the text skew and orientation.

The main idea is based on the fact that the text within the test images has obvious vertical patterns with respect to one viewing direction (see Fig. 4 (b) and (c)). Thus we can calculate the skew angle for a given input image by detecting the straight lines within it and looking into the statistics of the angles between these line segments and the x- or y-axis. More specifically, given a test image, we convert it into a binary image. Note that a number of image binarization algorithms have been proposed [32], [33] and that we use the method by Otsu et al. [32] in this paper. After that, we utilize the recent line detection technique by Gioi et al. [34] to detect all the line segments in the binary image. Assuming that (x_1, y_1) and (x_2, y_2) are the coordinates of the two endpoints of the i-th line segment and, without loss of generality, $y_2 \geq y_1$, we define the angle θ_i formed by this line segment and y-axis as

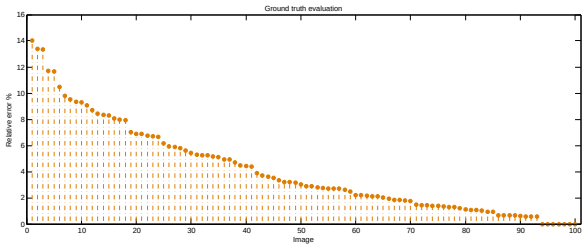
$$\theta_i = \begin{cases} \arccos\left(\dfrac{y_2 - y_1}{\|(x_2 - x_1, y_2 - y_1)\|}\right) & \text{if } x_1 \leq x_2 \\ -\arccos\left(\dfrac{y_2 - y_1}{\|(x_2 - x_1, y_2 - y_1)\|}\right) & \text{otherwise} \end{cases} \quad (6)$$

where $\|\cdot\|$ stand for the l^2-norm, and $\theta_i \in [-90°, 90]°$ is actually the angle between the vectors $(x_2 - x_1, y_2 - y_1)$ and (0, 1). Finally, we build a histogram of all θ_i with N bins and consider as the image skew angle θ the angle that corresponds to the peak of the histogram. Let $f(\alpha_i)$ be the frequency for the i-th histogram bin centered at α_i degrees. Then the skew angle θ is given by $\theta = \arg\max_{\alpha_i} f(\alpha_i)$. The sign of θ represents the direction of skew. That is, if $\theta < 0$, we need to rotate the image clockwise by $-\theta$ degrees to make the textlines parallel to the x-axis; otherwise, we make a counter-clockwise rotation by θ degrees.

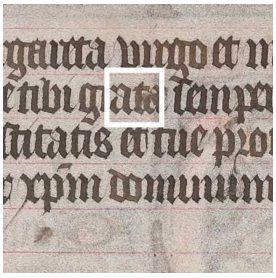
VI. Results

We tested our algorithm on 21 Medieval manuscripts (6922 pages), written by hand before 1500 AD; they are from Yale University's Beinecke Rare Book and Manuscript Digital Library [35] (a set of scripts is available [36] to download the book database). Those books are very different from each other, in terms of acquisition resolution, level of conservation, amounts of figures or ornamentations, and writing styles. Our technique was implemented on Linux using C++ and the OpenCV library [37]. Our benchmarks were executed on a PC with 8 Intel Core i7-3630QM CPU @ 2.40GHz processors, 12GB RAM, and a NVidia GeForce GTX 650M.

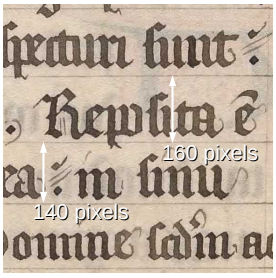
Given an input page, it is very difficult to manually define a correct and unique text height, because it changes across the single page or even across a single line. Both when it



(a) Relative error



(b) Image n.1 in (a)



(c) Text height variability

Fig. 4: **Ground truth.** Given a 100 image dataset, the maximum error is 14% as seen in (a), and corresponds to the square edge size t_E in (b). (c) shows text variability across a single page.

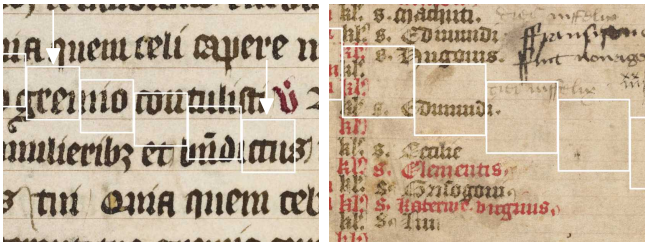


Fig. 5: **Visual evaluation.** We present some check images, corresponding with acceptable (a) or unacceptable results (b).

is manually set [3], [4] and in our automatic estimation, the reliability of the text height value is inversely proportional to character size variability. We performed two different kinds of evaluations, to understand if this computed value is below an acceptable error or not. In the first case, we produced a ground truth dataset; we randomly took 100 images from all the datasets, with different resolutions, text heights and types of manuscript, and manually measured the text height for each of them. We then compared those values with the ones automatically computed by the algorithm. In Fig. 4a we plot the relative error of image i as $\epsilon_i = 100\frac{|t_E^i - \tilde{t}_E^i|}{\tilde{t}_E^i}$, where t_E^i and $\tilde{t}_E^i$ are respectively the automatic and manual estimated values. In the plot we sort the errors in a descendent order. All the relative errors are under 15%, and in Fig. 4b we show the image corresponding to the highest relative error, in which we have drawn a square of edge size equal to $\tilde{t}_E^i$; the automatic estimated value well depicts the spatial periodicity of the analyzed text, and it has a reasonable size for a general layout analysis approach [3], [4]. Further, in Fig. 4c we show how difficult it is for the user to choose a good height value; the spatial text period in two adjacent lines varies from 140 to 160 pixels, with a difference of about 15% between them. Thus, a 14% maximum relative error is an acceptable outcome. However, this evaluation is only practical for a small subset of images. We would like to check all the thousands images in the studied books. This could be done only in a visual manner and in a computer assisted framework. Hence, for each image, after computing the text height value t_E, the algorithm draws a pattern of nine squares with edge equal to t_E. The original image with these overlapping squares helps the user to quickly estimate if the analysis result is visually reasonable. To understand the reliability of such evaluation, in Fig. 5 we show details of some checked images, corresponding with both acceptable and unacceptable results. We highlight with arrows the squares that were particularly helpful to us in marking the outcome as a good one. Table II shows a high rate of good estimations; all the books are over 94%, with some even achieving 100% accuracy.

Fig. 6: **Illuminated manuscript pages.** We present the original image of the page and two highlighted parts; the squares have edge lengths equal to the automatically estimated text height t_E.

Typical pages of illuminated manuscripts are shown in Fig. 6. They contain text in two different colors, capital letters of different types and sizes, the parchment background, other figures inside the text and ornaments. The images could also contain the dark acquisition background, and other visible parts of the book. We present both the original images of the page and one or more highlighted parts, with a square of edge size equal to the estimated text height t_E. This helps demonstrate the conditions of the whole analysis domain, and to visually appreciate the quality of the output. Although the result is good, the pages in Fig. 6 are not so challenging, since they are very well preserved and do not contain any kind of noise. In Fig. 8 and Fig. 9, we present most of the problems that arise when dealing with very old handwritten manuscripts. The two pages in Fig. 8a are affected by significant ageing and very bad preservation conditions; the result shows how the proposed frequency-based descriptor is able to extract the main image directions even at the presence of a very noisy signal. It is also robust to low constrast signals, as shown in Fig. 8b, where the ageing makes the ink almost disappear. Due to the value of these rare books, the acquisition setup is very controlled, both for preservation of their integrity, and to produce the best possible digital images. However, some texts could come out not perfectly horizontal, mainly because the text could be skewed compared to the page edges. In our tests

we never used an automatic skew correction, and Fig. 8c proves how the proposed technique is able to properly deal with such skew texts. The multi-scale framework is convenient when we need to cope with other extreme but unrare situations, such as a low number of text lines (Fig. 9a) or a small percentage of text in a page with a lot of figures and other non-text elements (Fig. 9b). Two extreme cases are presented in Fig. 9c; in one case, only a small part of the text is visible, and, in the other, the page is very damaged and contains a lot of comments written in different styles. We also demonstrate how our approach is robust to the well known problem of ink bleed-through, which makes the writing from the verso appear on the recto (Fig. 9d). The page on the right in Fig. 9d is affected by bleed-through, it contains very few text lines, a lot of noise and other handwritten signs. Since the presented method aims at finding the most predominant spatial periodicity in the page, we have seen that it fails when there are some concurrent high amplitude frequencies. This occurs when the text is not organized in a regular manner, or, in other words, when the inter-line spacing has high variability. The failures (bad images) in table II are always similar to those in Fig. 8d; on the left the bad estimated text height value clearly depends on the groups of three text lines, while the case of the right contains both a non-regular text line pattern and an additional comment part in the bottom, written (perhaps by a different author) with a completely different style.

To evaluate the performance of the method for text orientation correction, we arbitrarily rotated the 100 randomly selected images and applied the method to the rotated images to obtain the skew angles θ. In this experiment, we fix the number of histogram bins at $N = 360$. Fig. 7 (a) shows the absolute angle error/difference between the computed skew angles θ and their corresponding groundtruth rotation angles. From this figure, we see that our orientation correction method is able to find the skew angle at a satisfactory rate, that is, with an angle error of less than 1.5 degrees, up to 94% of the time. After comparing the white horizontal reference line in Fig. 7 (b) and the text orientation, it is clear that the error of 1.5 degrees is trivial and thus absolutely acceptable for practical applications. Fig. 7 (c) shows a close-up of the image that corresponds to the highest angle error of 5.54 degrees. Although the error is 5.54 degrees, our method actually outputs the expected skew angle because we can, upon close inspection, observe that the strokes of the texts are approximately perpendicular to the white horizontal reference line. In addition, it is worth mentioning here that the proposed text height estimation algorithm can tolerate a skew of up to 6 degrees (see Fig. 8 (c)).

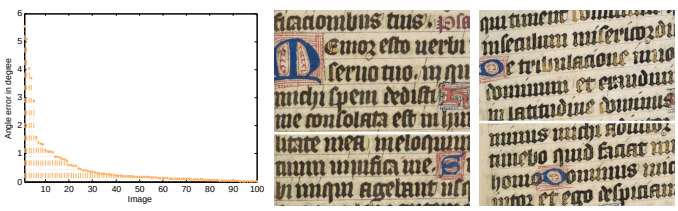


Fig. 7: **Text orientation correction.** (a) Angle error. Close-up of an image with angle error of 1.58 degrees (b) and 5.54 degrees (c).

Book Name	# Figure Pages	Precision	Recall
BeineckeMS310	23	0.92	1.0
BodleianMSGoughLiturg.3	5	0.45	1.0
BodleianMSLaudMisc.204	16	0.88	1.0
Walters34	26	0.89	1.0

TABLE I: Precision and Recall values in the retrieval of those pages that contain only figures.

Although the automatic text height estimation is just the first important step to building a completely automatic layout analysis framework, this simple output can lead to some interesting results. It turns out that the text size measured across all the pages of a single book is somehow consistent, while the text height estimation for pages without any text is random. By exploiting the text height and the color statistical distribution (average and variance) across the same book, we can distinguish between pages that contain text and pages that contain only figures. In table I we show the precision/recall results after applying this segmentation to books having pages with only figures. The true positive are the pages well segmented that contain only figures, while the false positive/negative are bad segmented pages that respectively do not/do contain only figures. In our experiments the recall value equal to 1 because we do not have any false negatives. Another possible application could be a tool that gives scholars an ordered list of pages based on the computed statistical distribution; i.e., users can analyze a book by first sorting its pages according to the level of image or text content.

VII. Conclusions

We have presented a method to perform automatic text height estimation, without the need of any kind of manual intervention and user defined parameters. We have tested our algorithm on a large heterogeneous corpus of 21 medieval books, containing almost seven thousands pages. With an average per-page computation time of 5 seconds and more than 99% good text height estimations, it has proved to be very robust and reliable in the case of very noisy and damaged manuscripts, with different writing styles, text sizes, image resolutions, levels of conservation, and affected by other countless uncontrollable factors, such as holes, spots, ink bleed-through, ornamentation, background noise, and touching text lines. Future work will use this technique as a first step of a more general document layout analysis framework, which will exploit the capabilities of the presented frequency-based local descriptor. Due to the intrinsic parallel nature of the presented analysis, a GPU-based implementation is straightforward, and would make it more suitable for processing a large database.

Acknowledgment

This work was supported by the Digitally Enabled Scholarship with Medieval Manuscripts (DESMM) project funded by the Mellon Foundation (http://ydc2.yale.edu/).

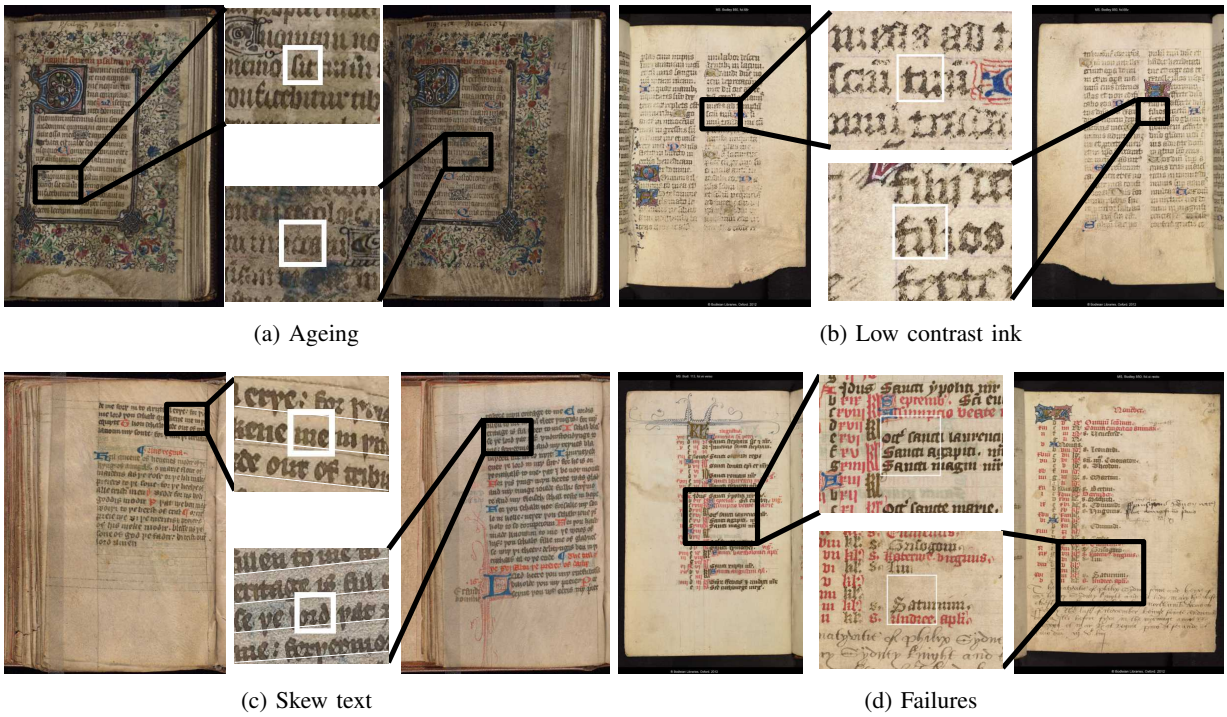
References

[1] X. Wang and E. Keogh, "Augmenting historical manuscripts with automatic hyperlinks," in *ISM'09.*, 2009, pp. 571–576.

[2] F. Le Bourgeois and H. Kaileh, "Automatic metadata retrieval from ancient manuscripts," *Lecture notes in computer science*, pp. 75–89, 2004.

Book Name	Avg. Resolution WxH	Pages	Non-text	Text	Good	Bad	Time
BeineckeMS310	2886 x 3794	309	32	277	266 (96%)	11 (4%)	12m (2sec/pg)
BeineckeMS10	2324 x 3127	187	13	174	174 (100%)	0 (0%)	4m (1sec/pg)
BeineckeMS109	1822 x 2416	270	19	251	251 (100%)	0 (0%)	4m (1sec/pg)
BeineckeMS360	1654 x 2083	382	11	371	371 (100%)	0 (0%)	3m (0.5sec/pg)
BeineckeMS748	2313 x 3232	8	0	8	8 (100%)	0 (0%)	11s (1sec/pg)
BeineckeMS525	2053 x 2855	46	4	42	42 (100%)	0 (0%)	1m (1sec/pg)
BodleianMSBodley113	5329 x 7487	315	12	303	292 (96%)	11 (4%)	75m (14sec/pg)
BodleianMSBodley850	5370 x 6959	246	16	230	217 (94%)	13 (6%)	41m (10sec/pg)
BodleianMSDouce18	5167 x 7155	534	17	517	505 (98%)	12 (2%)	2h27m (14sec/pg)
BodleianMSGoughLiturg.3	5278 x 6786	257	36	221	219 (99%)	2 (1%)	1h (14sec/pg)
BodleianMSLaudMisc.204	5170 x 7013	286	36	250	246 (98%)	4 (2%)	1h38m (21sec/pg)
BodleianMSliturg.e.17	5237 x 7201	224	13	211	209 (99%)	2 (1%)	36m (10sec/pg)
MarstonMS22	3814 x 2574	121	5	116	116 (100%)	0 (0%)	4m (2sec/pg)
Osborna44	2336 x 3025	483	13	470	463 (98%)	7 (2%)	12m (2sec/pg)
Osbornfa1	3487 x 4240	400	3	397	396 (99%)	1 (1%)	28m (4sec/pg)
Walters34	2050 x 3139	658	78	580	574 (99%)	6 (1%)	12m (1sec/pg)
Walters102	2291 x 3359	222	14	208	208 (100%)	0 (0%)	6m (2sec/pg)
Admont43	2882 x 4347	358	0	358	358 (100%)	0 (0%)	17m (3sec/pg)
Admont23	2815 x 4286	591	0	591	591 (100%)	0 (0%)	27m (3sec/pg)
CologneErzbisch127Ka	3480 x 4491	625	10	615	614 (99%)	1 (1%)	16m (2sec/pg)
CologneErzbisch128Kb	3072 x 3840	400	1	399	399 (100%)	0 (1%)	5m (1sec/pg)
# books - 21		6922	333	6589	6519 (99%)	70 (1%)	10h (5sec/pg)

TABLE II: Text height estimation statistics.



(a) Ageing (b) Low contrast ink

(c) Skew text (d) Failures

Fig. 8: **Challenging samples and failures.** These pages are affected by the following imperfections: (a) strong ageing; (b) low contrast ink; (c) skewed text. In (d) the algorithm fails due to a lack of a predominant spatial frequency.

[3] M. Mehri, P. Gomez-Krämer, P. Héroux, and R. Mullot, "Old document image segmentation using the autocorrelation function and multiresolution analysis," in *IS&T/SPIE Electronic Imaging*, 2013.

[4] A. Garz, A. Fischer, R. Sablatnig, and H. Bunke, "Binarization-free text line segmentation for historical documents based on interest point clustering," in *DAS 2012*, 2012, pp. 95–99.

[5] N. Journet, J.-Y. Ramel, R. Mullot, and V. Eglin, "Document image characterization using a multiresolution analysis of the texture: application to old documents," *IJDAR*, vol. 11, no. 1, pp. 9–18, 2008.

[6] C. Grana, D. Borghesani, and R. Cucchiara, "Picture extraction from digitized historical manuscripts," in *Proceedings of the ACM International Conference on Image and Video Retrieval*. ACM, 2009, p. 22.

[7] S. Jindal and G. S. Lehal, "Line segmentation of handwritten gurmukhi manuscripts," in *DAR*. ACM, 2012, pp. 74–78.

[8] M. Diem and R. Sablatnig, "Recognizing characters of ancient manuscripts," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2010, pp. 753 106–753 106.

[9] I. Z. Yalniz and R. Manmatha, "An efficient framework for searching text in noisy document images," in *DAS*, 2012, pp. 48–52.

[10] G. Louloudis, B. Gatos, I. Pratikakis, and C. Halatsis, "Text line detection in handwritten documents," *Pattern Recognition*, vol. 41, no. 12, pp. 3758–3772, 2008.

[11] Y. Leydier, F. Lebourgeois, and H. Emptoz, "Text search for medieval manuscript images," *Pattern Recognition*, vol. 40, no. 12, pp. 3552–

(a) Few text lines

(b) Small percentage of text

(c) Significant damage

(d) Ink bleed-through

Fig. 9: **Challenging samples.** These pages are affected by the following imperfections: (a) few text lines; (b) a small quantity of text; (c) significant damage; (d) ink bleed-through.

3567, 2007.

[12] N. Sharma, U. Pal, and M. Blumenstein, "Recent advances in video based document processing: a review," in *DAS*, 2012, pp. 63–68.

[13] L. Likforman-Sulem, A. Zahour, and B. Taconet, "Text line segmentation of historical documents: a survey," *IJDAR*, vol. 9, no. 2-4, pp. 123–138, 2007.

[14] M. Bulacu, R. van Koert, L. Schomaker, T. van der Zant *et al.*, "Layout analysis of handwritten historical documents for searching the archive of the cabinet of the dutch queen," in *ICDAR*, 2007, pp. 357–361.

[15] V. Shapiro, G. Gluhchev, and V. Sgurev, "Handwritten document image segmentation and analysis," *Pattern Recognition Letters*, vol. 14, no. 1, pp. 71–78, 1993.

[16] A. Antonacopoulos and D. Karatzas, "Document image analysis for world war ii personal records," in *Proc. Document Image Analysis for Libraries, 2004.*, 2004, pp. 336–341.

[17] S. Khedekar, V. Ramanaprasad, S. Setlur, and V. Govindaraju, "Text-image separation in devanagari documents," in *ICDAR*, vol. 2, 2003.

[18] Y. Wang, I. T. Phillips, and R. M. Haralick, "A study on the document zone content classification problem," in *DAS*, 2002, pp. 212–223.

[19] I. Bar-Yosef, N. Hagbi, K. Kedem, and I. Dinstein, "Line segmentation for degraded handwritten historical documents," in *ICDAR*. IEEE, 2009, pp. 1161–1165.

[20] A. Zahour, B. Taconet, P. Mercy, and S. Ramdane, "Arabic hand-written text-line extraction," in *ICDAR*. IEEE, 2001, pp. 281–285.

[21] A. K. Jain and A. M. Namboodiri, "Indexing and retrieval of on-line handwritten documents," in *ICDAR*, 2003, p. 655.

[22] E. H. Ratzlaff, "Inter-line distance estimation and text line extraction for unconstrained online handwriting," in *Workshop on Frontiers in Handwriting Recognition*, 2000, pp. 33–42.

[23] R. Minetto, N. Thome, M. Cord, N. J. Leite, and J. Stolfi, "T-hog: An effective gradient-based descriptor for single line text regions," *Pattern Recognition*, 2012.

[24] V. Eglin, S. Bres, and C. Rivero, "Hermite and gabor transforms for noise reduction and handwriting classification in ancient manuscripts," *IJDAR*, vol. 9, no. 2-4, pp. 101–122, 2007.

[25] A. Garz, M. Diem, and R. Sablatnig, "Local descriptors for document layout analysis," in *Advances in Visual Computing*. Springer, 2010, pp. 29–38.

[26] C. L. Sabharwal and S. Subramanya, "Indexing image databases using wavelet and discrete fourier transform," in *Proceedings of the 2001 ACM symposium on Applied computing*. ACM, 2001, pp. 434–439.

[27] Y.-L. Qiao, Z.-M. Lu, C.-Y. Song, and S.-H. Sun, "Document image segmentation using gabor wavelet and kernel-based methods," in *ISSCAA*, 2006, pp. 5–pp.

[28] A. Lemaitre, J. Camillerapp, and B. Coüasnon, "Multiresolution cooperation makes easier document structure recognition," *IJDAR*, vol. 11, no. 2, pp. 97–109, 2008.

[29] M. S. Almeida and L. B. Almeida, "Nonlinear separation of show-through image mixtures using a physical model trained with ica," *Signal Processing*, vol. 92, no. 4, pp. 872–884, 2012.

[30] G. Joutel, V. Eglin, and H. Emptoz, "A complete pyramidal geometrical scheme for text based image description and retrieval," in *Image and Signal Processing*. Springer, 2008, pp. 471–480.

[31] N. Journet, R. Mullot, J.-Y. Ramel, and V. Eglin, "Ancient printed documents indexation: a new approach," in *Pattern Recognition and Data Mining*. Springer, 2005, pp. 580–589.

[32] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[33] J. Sauvola and M. Pietikinen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225 – 236, 2000.

[34] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: A fast line segment detector with a false detection control," *PAMI, IEEE Transactions*, vol. 32, no. 4, pp. 722–732, 2010.

[35] Beinecke, "Beinecke rare book and manuscript library," Yale University, 2013. [Online]. Available: http://beinecke.library.yale.edu/

[36] ——, "21 book database - download scripts," Yale University, 2013. [Online]. Available: http://hdl.handle.net/10079/cz8w9v8

[37] OpenCV, "Opencv - open source computer vision library," 2013. [Online]. Available: http://opencv.org/