Binarization of Historical Watermark Images

Anna N. Lantink A.N.Lantink@student.tudelft.nl

Introduction

- **Watermarks (Fig. 1): images that appear in historical paper.** They identify the producer of the paper [1].
- Watermarks provide data on a document's origins.
  - No tool is publicly available to automatically visually analyze watermarks.
- **Binarization: process of categorizing pixels into the watermark foreground, and non-watermark background** [2].
  - Isolates the watermark's shape.
  - Difficult to binarize degraded images, such as historical watermarks (Fig. 1)
- **This raises the question:**

To what extent can thresholding techniques be effective in binarizing watermark images with degraded quality, and how do different algorithms compare to each other?

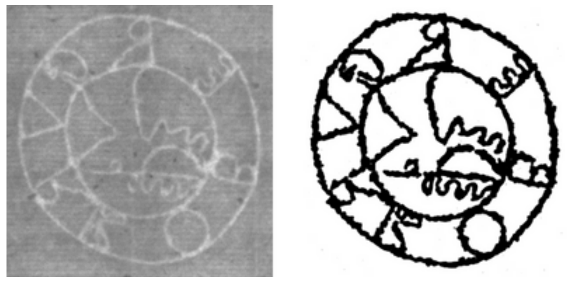

Figure 1: A raw watermark, (left) and a perfectly binarized watermark (right).

Results

- Quantitative (Table 1): Overall, metrics are poor. Different metrics have different results.
- Qualitative (Tables 3-4): Poor agreement on which algorithm performs best. Substantial agreement that watermark is present but contains background [13].

	F1 Score ($\times 10^{-2}$)	PSNR	NRM ($\times 10^{-2}$)	MPM ($\times 10^{-2}$)
Algorithm	Mean	Mean	Mean	Mean
Contrast [4]	13.95	8.38	37.23	5.46
Logical Adaptive [6]	20.46	**12.16**	37.88	1.83
Background Estimation [7]	20.20	6.68	21.51	10.03
Proposed Algorithm	24.21	11.47	34.38	**1.82**
Niblack [9]	12.17	3.16	27.14	24.83
Watermark Prototype [10]	11.49	2.84	29.20	27.38
Color Histograms [5]	20.50	6.40	19.24	9.18
Entropy [3]	**30.72**	9.62	**19.13**	2.47
Otsu [8]	15.50	4.15	21.55	16.24

Table 1: Quantitative results. The row in blue is the baseline. Bold represents a top performing result.

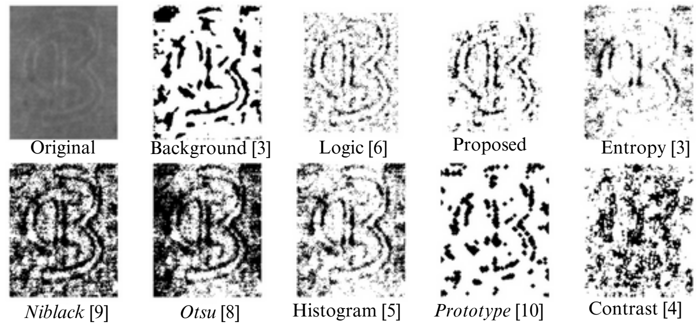

Figure 3: The binarized results for a test set image. Italics represent baseline algorithms.

Algorithm	Percentage (%) of Overall Selection
Contrast [4]	0.90
Logical Adaptive [6]	**34.65**
Background Estimation [7]	10.90
Proposed Algorithm	30.97
Niblack [9]	7.43
Watermark Prototype [10]	5.28
Color Histograms [5]	6.94
Entropy [3]	2.08
Otsu [8]	0.83

Table 2: The frequency that participants choose an algorithm as best. The row in blue is the baseline. Bold represents the most selected algorithm.

Rating	Percentage (%) for Statement 1	Percentage (%) for Statement 2
Strongly Agree	12.01	1.32
Agree	**45.56**	12.08
Neutral	20.90	16.74
Disagree	18.61	**48.75**
Strongly Disagree	2.92	21.11

Table 3: The frequency of likert ratings across users and images. Statement one is: the complete watermark is shown, and statement two: the non-watermark background is not present.

Conclusion

- The selected algorithms usually find the watermark, but fail to fully separate watermark from background.
- The selected algorithms are not effective to a significant extent in binarizing degraded watermark images.

Future Work

- Testing out non-thresholding binarization algorithms.
- Combining the thresholding algorithms with denoising. This could significantly improve the amount of misclassified foreground pixels.

Methodology

Dataset

- Data provided by the German Museum of Books and Writing.
- 235 raw watermark images.
- Split: 66% training, 17% validation, and 17% test data.

Algorithm Selection

- **Specialized algorithms**: entropy [3], contrast [4], color histograms [5] logic [6], and background estimation [7].
- **Baselines**: commonly used algorithms: Otsu [8] and Niblack [9], and an algorithm from a watermark matching prototype [10].
- **The proposed algorithm.**

Proposed Algorithm

1. Generate low and high detail binarized images.
2. Iterate through all pixels in low detail.
3. Take a window around each pixel from the high detailed image.
4. Add all pixels in window to the final image.

Figure 2: A synthetic watermark image

Experiments

- **Qualitatively**: Survey where participants chose which algorithm performed best and rated its performance.
- **Quantitatively**: The F1, PSNR, NRM, and MPM [11] were calculated. Evaluated with synthetic data, using noised images of drawings [12].

Resources

[1] L. Muller, "Understanding paper: Structures, watermarks, and a conservator's passion," https: //harvardartmuseums.org/article/understanding-paper-structures-watermarks-and-a-conservator-s-passion, 2021
[2] N. Chaki, S. H. Shaikh, and K. Saeed, Exploring Image Binarization Techniques, vol. 560. in Studies in Computational Intelligence, vol. 560. New Delhi: Springer India, 2014. doi: 10.1007/978-81-322-1907-1.
[3] C. A. B. Mello and A. H. M. Costa, "Image Thresholding of Historical Documents Using Entropy and ROC Curves," in Progress in Pattern Recognition, Image Analysis and Applications, vol. 3773, A. Sanfeliu and M. L. Cortés, Eds., in Lecture Notes in Computer Science, vol. 3773. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 905–916. doi: 10.1007/11578079_93.
[4] Bolan Su, Shijian Lu, and Chew Lim Tan, "Robust Document Image Binarization Technique for Degraded Document Images," IEEE Trans. on Image Process., vol. 22, no. 4, pp. 1408–1417, Apr. 2013, doi: 10.1109/TIP.2012.2231089.
[5] A. V. S. Rao, G. Sunil, N. V. Rao, T. S. K. Prabhu, L. P. Reddy, and A. S. C. S. Sastry, "Adaptive Binarization of Ancient Documents," in 2009 Second International Conference on Machine Vision, Dubai, UAE: IEEE, 2009, pp. 22–26. doi: 10.1109/ICMV.2009.8.
[6] M. Kamel and A. Zhao, "Extraction of Binary Character/Graphics Images from Grayscale Document Images," CVGIP: Graphical Models and Image Processing, vol. 55, no. 3, pp. 203–217, May 1993, doi: 10.1006/cgip.1993.1015.
[7] B. Gatos, I. Pratikakis, and S. Perantonis, "Adaptive degraded document image binarization," Pattern Recog- nition, vol. 39, no. 3, pp. 317–327, Mar. 2006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/ S0031320305003821
[8] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Trans. Syst., Man, Cybern., vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.
[9] W. Niblack, An introduction to digital image processing. Englewood Cliffs, N.J: Prentice-Hall International, 1986, pp. 115-116.
[10] D.-M. Banta, S. Kho, A. N. Lantink, A.-R. Marin, and V. Petkov, "A watermark recognition system: An approach to matching similar watermarks," http://resolver.tudelft.nl/uuid:e8dfbd63-ae54-4159-b786-d1d8c64dc827, 2023.
[11] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "DIBCO 2009: document image binarization contest," IJDAR, vol. 14, no. 1, pp. 35–44, Mar. 2011, doi: 10.1007/s10032-010-0115-7.
[12] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch objects?," ACM Trans. Graph., vol. 31, no. 4, pp. 1–10, Aug. 2012, doi: 10.1145/2185520.2185540.
[13] N. Wongpakaran, T. Wongpakaran, D. Wedding, and K. L. Gwet, "A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples," BMC Medical Research Methodology, vol. 13, no. 1, p. 61, Dec. 2013. [Online]. Available: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-6111