Objective paper structure comparison through processing of transmitted light images

Charles E. H. Berger

Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, The Netherlands.

A method for the comparison of paper structure using light transmission images and frequency analysis was developed. The resolution of the light transmission images and the algorithm for the feature extraction was greatly improved to enhance the visibility of peaks in the 2D power spectrum that results from frequency analysis. A comparison method based on correlation measures how well the spectra match as a function of the orientation of the paper, yielding an objective and quantitative measure of similarity between 0 and 1. A technical validation was carried out with 25 different papers showing the potential of this method with common copy papers. Finally, the method was applied in a case.

1 Introduction

Documents offer the forensic examiner a wide variety of properties that can be studied when trying to answer forensic questions. The physical substrate of the document, the paper, is one of them. Several advanced analytical techniques have been applied [1-4], but this paper explores the use of more readily available flatbed scanners to obtain light transmission images. Such images allow one to non-destructively and quantitatively analyze structural traces with certain periodicities left during the paper production process by one or several machines. Spatial frequency analysis of light transmission images has been applied with varying success in the past [5-11] on different types of

paper. In this study we used a high quality scanner which gave good light transmission images for common copy papers. With an improved resolution and feature extraction algorithm, we obtained excellent discrimination of paper structures, without having to make assumptions about the orientation of the paper.

2 Methods

2.1 Feature extraction

A high quality flatbed scanner (CreoScitex Eversmart Jazz) was used to scan 5×5 cm^2 areas of sheets of paper in transmission at high resolution (2540 dpi or 10 μm pixels, giving 5000×5000 pixel images). The sheets of paper were scanned with random orientation, and with either side up. Table 1 shows the 25 different papers that were used in this study. All are white multifunctional 80 g/m² papers; sample 163 is marketed as an inkjet paper, and samples 097 and 121 are recycled papers.

Since the patterns in the paper structure are repetitive, Fourier analysis is ideally suited for analyzing and extracting such repetitive features. Figure 1a shows a transmitted light image and the subsequent image processing steps. The 2D Fourier transform of the full 5000×5000 pixel image was calculated using a freely available implementation [12] of the Fast Fourier Transform (FFT) and its power spectrum is shown in Figure 1b (center 800×800 pixels only). The intensity of any point in this Fourier-transformed image gives the amplitude of the corresponding sinusoidal component in the original image, while its distance to the center gives the frequency, and the vector from the center to the point gives the direction of the component. These properties make the Fourier-transformed image point symmetric.

Though a pattern is already apparent, the overall graininess of the image makes it far from ideal for automatic comparison. Therefore, subsequent image processing is performed in MATLAB® Image Processing Toolbox (The MathWorks Inc., Natick, MA). First, gray scale dilation is applied with a disk (5 pixel radius) as a structuring element. Dilation means setting every pixel within the structuring element (local neighborhood) in the output image to the maximum of that area in the input image. This makes the peaks

more apparent and circular shaped (see Figure 1c). Then the center peak in the image is removed, since it is a common feature for all our FFT images.

Before the next processing step, we define a number of morphological operations. Erosion is the equivalent of dilation, but with the values set to the *minimum* within the structuring element. The morphological opening operation is defined as an erosion followed by a dilation. Top-hat filtering is performed by subtracting the result of the opening operation on an input image, from the input image itself.

Top-hat filtering is applied to images like the one in Figure 1c, to separate peaks from their uneven background, again with a disk as a structuring element (15 pixel radius). Finally, the gray levels of the image are adjusted to remove what is left of the background (see Figure 1d).

2.2 Comparison

Images can be correlated by pair wise multiplying all corresponding pixels in both images, and adding up the results. To make the method robust to the orientation of the pattern in the paper, the correlation for two images A and B is defined as a function of θ, the angle over which A is rotated before correlating it with B:

$$correlation(A, B, \theta) = \frac{\sum_{m,n} R_{mn}\left(A, \theta\right) \cdot B_{mn}}{\sqrt{\sum_{m,n} A_{mn}^2} \sqrt{\sum_{m,n} B_{mn}^2}}, \quad (1)$$

where $R(A,\theta)$ is the rotation of image A over angle θ, and the denominator is a normalization factor. The correlation is maximized for $\theta = \theta_{\max}$ (see Figure 2), which gives the correlation value that will be used in our comparison score. It is the angle for which the compared peak patterns in both images overlap most.

Turning the paper over does not simply result in a mirrored transmission image, so the light transmission image depends on which side of the paper is up. As it is generally not obvious which side of the paper is the “wire” side (the side that was in contact with the sieve during paper production), the method was made robust to whichever side of the paper is up. Every sheet of paper has two images associated with it (front and back), and the final comparison score is defined by the maximum of:

$$correlation(A_{front}, B_{front}, \theta_{\max}) \cdot correlation(A_{back}, B_{back}, \theta_{\max})$$

and (2)

$$correlation(A_{front}, B_{back}, \theta_{\max}) \cdot correlation(A_{back}, B_{front}, \theta_{\max}),$$

which will give us a final comparison score between 0 and 1. A complete lack of overlap of the peak patterns for any angle will give a score of zero, while a score of one will result when there is an angle for which the peak patterns overlap perfectly.

3 Results

3.1 Feature extraction

Figure 3a shows the results of the feature extraction algorithm applied to light transmission images (5×5 cm^2) of 25 different multipurpose papers. The numbers in the image are the labels of the papers in our collection. Figure 3b shows the results for the reverse side of the papers. It is evident from the images that there are clear differences between the different papers, as well as between the different sides of the papers. Variation in the paper structure is caused by the paper sieve as well as other machines involved in the paper production process [10]. This variation obviously increases the individualizing value of the extracted patterns.

The influence of the sample size of the paper was studied by reducing it from 5×5 cm^2 to 4×4, 3×3, 2×2, and finally to 1×1 cm^2. Figure 4a shows the results of the feature extraction for the reduced sample sizes (paper 003). The correlation with the corresponding reference sample was consistently higher for sample sizes down to 1×1 cm^2 (see Figure 4b), and did not drop significantly for sample sizes down to 2×2 cm^2. The feature extraction therefore seems rather robust to sample size, in case 5×5 cm^2 of paper area is not available.

3.2 Comparisons at the sheet level

To study the intra-source and inter-source variation, 5 light transmission images (5×5 cm^2) were obtained from either side of the 25 papers (one sheet was used per paper). The orientation and position of the papers was varied randomly.

Figure 5 gives the resulting scores of all possible 7750 comparisons, where comparison scores vary between zero (white) and one (black). The intra-source comparisons give the darker “shark tooth” pattern along the diagonal. Dotted lines indicate comparisons of differently named papers that came from the same factory, forming the clusters {083,008,163}, {097,095,153}, and {111,108}.

Histograms of all intra-source (‘true’) comparisons, and inter-source (‘false’) comparisons are given in Figure 6.

3.3 Comparisons at the pack level

In an experiment analogous to the one described in the previous section, 5 separate sheets were taken from each paper, but from the same pack of paper. Intra-source measurements were now performed on those 5 individual sheets rather than on just one sheet as before. Clearly this is more relevant for the forensic scenario in which a sheet of paper is to be compared to *e.g.* a pack of paper at a suspect’s home.

The comparison results in Figure 7 are very similar to those in Figure 5. Figure 8 shows that the histograms of the intra-source and inter-source comparisons now have some overlapping values.

3.4 Case example

The method presented in this paper was applied for investigational purposes in a case where someone was suspected of producing 13 documents which he claimed came from 13 different sources. Figure 9 shows the comparison results for the 13 documents, prepared over a period of 4 years. Two clusters of documents with a very similar paper structure can clearly be discerned.

4 Discussion

While Figure 6 showed a perfect discrimination between intra-source and inter-source comparisons, Figure 8 shows some overlap of comparison score distributions. This is caused by the increased intra-source variation when the source is more realistically defined as a pack of paper, and not as a single sheet of paper. It is tempting to derive a value for the likelihood ratio: the value of the evidence for the proposition that 2

sheets of paper come from the same pack (rather than from a random pack). But the collection of 25 papers is not designed to be representative of packs of paper in any population. Therefore, the present work is only a technical forensic validation because the present collection of papers is not suitable for evidence evaluation purposes.

Unless one or more of the paper producers have a very dominant market share, there is reason for optimism about the strength of the evidence obtained with this method. Other advantages are that it is relatively fast, robust, and low-tech.

5 Conclusion

The presented method for the comparison of paper structure using light transmission images and frequency analysis performs well, even for common copy papers which formed a problem up to now. The resolution of the light transmission images and the algorithm for the feature extraction was greatly improved to enhance the visibility of peaks in the 2D power spectrum. The method is comparatively cheap and easy to use, and entirely non-destructive.

References

[1.] J.A.W. Barnard, D.E. Polk, B.C. Giesses, Forensic identification of paper by elemental analysis using scanning electron microscopy. Scanning Electron Microsc 8 (1975) 519–27.

[2.] H.A. Foner, N. Adan, The characterization of papers by X-ray diffraction (XRD): measurement of cellulose crystallinity and determination of mineral composition. J Forensic Sci Soc 23 (1983) 313–21.

[3.] J. Andrasko, Microreflectance FTIR techniques applied to materials encountered in forensic examination of documents. J Forensic Sci 41 (1996) 812–23.

[4.] R. Sugita, H. Ohta, S. Suzuki, Identification of Photocopier Paper by Pyrolysis Gas Chromatography. The 4th Annual Meeting of Jpn Assoc Tech Iden Japan, 1999.

[5.] H. Praast, L. Goettsching, Analysis der siebmarkierung im durchlight. Das papier 41 (1987) 105–120.

[6.] M. Shinozaki, Y. Tajima, S. Miyamoto, Paper "Formation" Image Analysis. Jpn J Paper Tech 39 (1996) 24–8.

[7.] T. Enomae, S. Kuga, Paper formation analysis of light transmission images acquired by desk-top flat-bed image scanner. The 47th Annual Meeting of the Japan Wood Res Soc; 1997.

[8.] M. Shinozaki, Frequency Analysis of Paper Formation. Jpn TAPPI J 53 (1999) 914–25.

[9.] M. Shinozaki, Y. Tajima, S. Miyamoto, An evaluation method for paper formation based on light transmission distribution and its spatial frequency analysis. J Soc Fiber Sci Tech Jpn 55 (1999) 383–92.

[10.] Miyata H, Shinozaki M, Nakayama T, Enomae T. A Discrimination Method for Paper by Fourier Transform and Cross Correlation. J Forensic Sci 47 (2002) 1-8.

[11.] O. Comte, D. Dessimoz, L. Lanzi, S. Marquet, W. Mazzella, Paper discrimination by fast Fourier transform. Poster abstract, 4th European Academy of Forensic Science, 2006.

[12.] www.4N6site.com

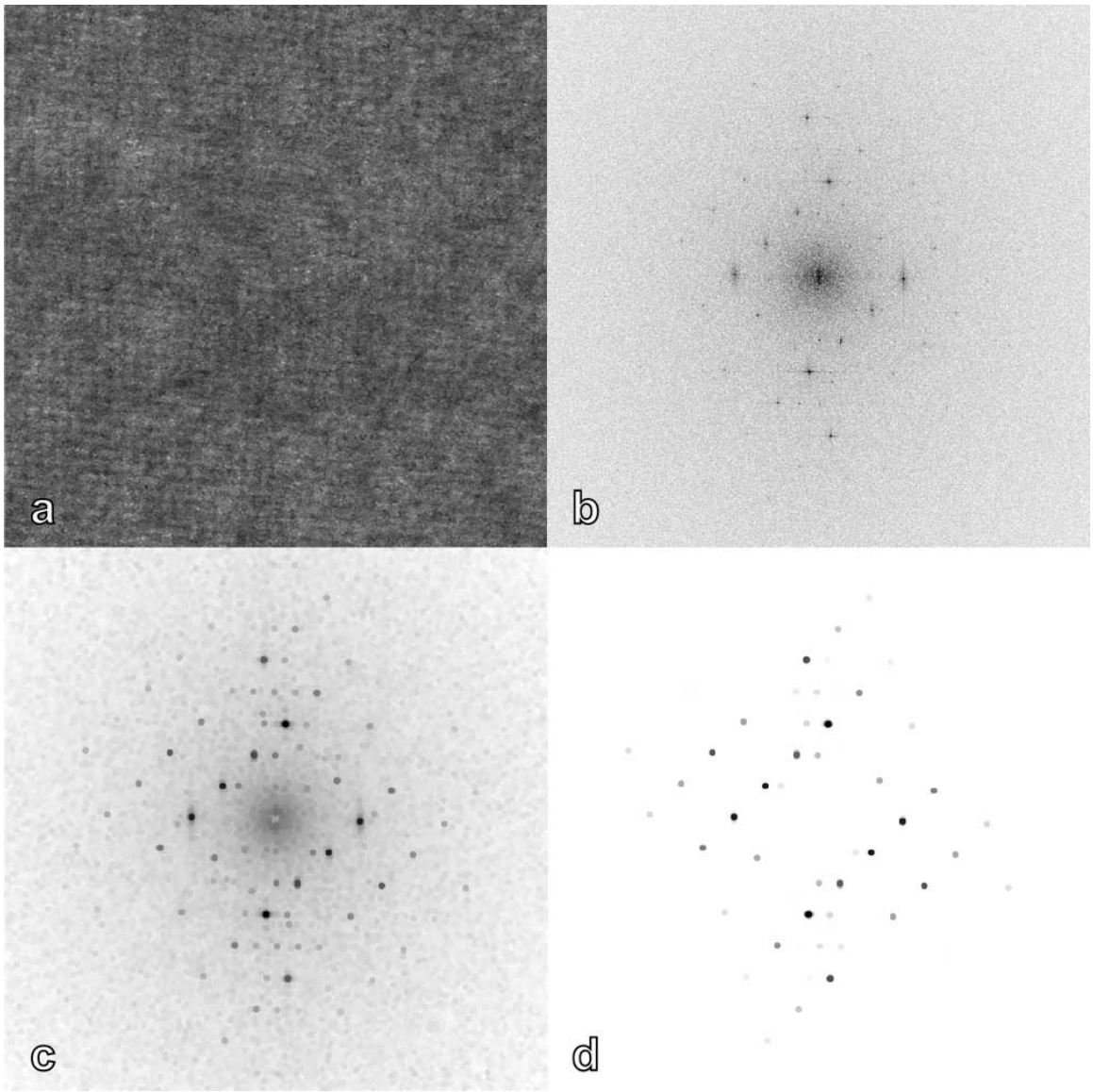

Figure 1

(a) Transmitted light image showing the structure of a paper.

(b) Power spectrum of the two-dimensional Fourier Transform of the transmitted light image (density gives amplitude).

(c) Grayscale dilation of the power spectrum image and removal of its center.

(d) Top-hat filtered version of the previously dilated image.

The images have been rescaled in size and contrast for clarity.

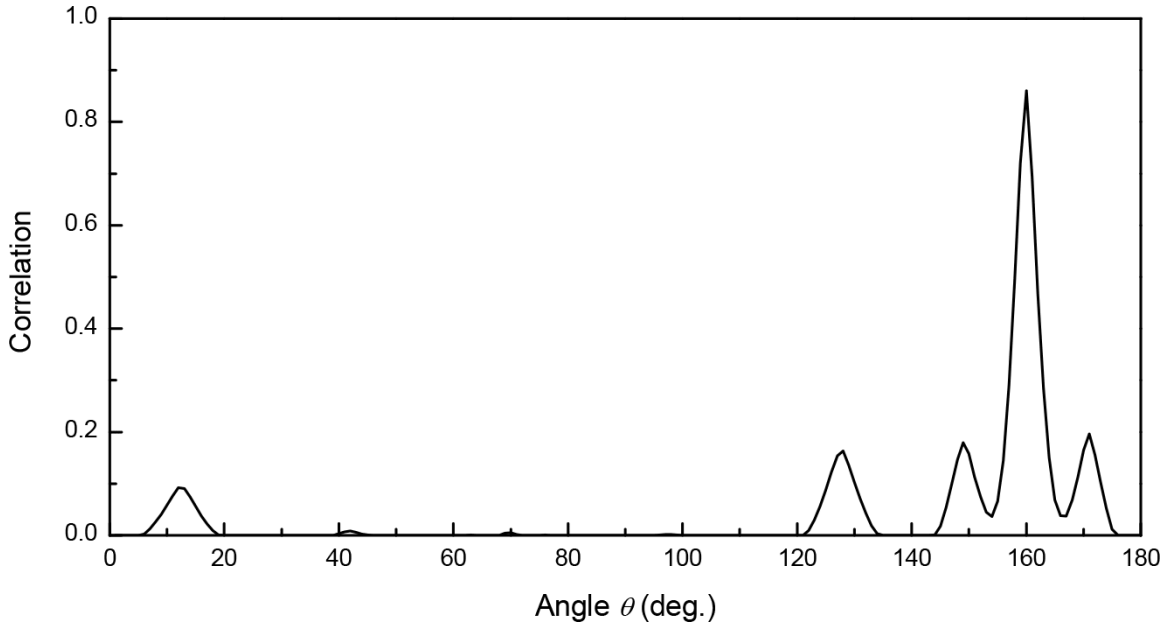

Figure 2

The correlation for two images as a function of the angle θ over which the one image is rotated relative to the other (see Equation 1).

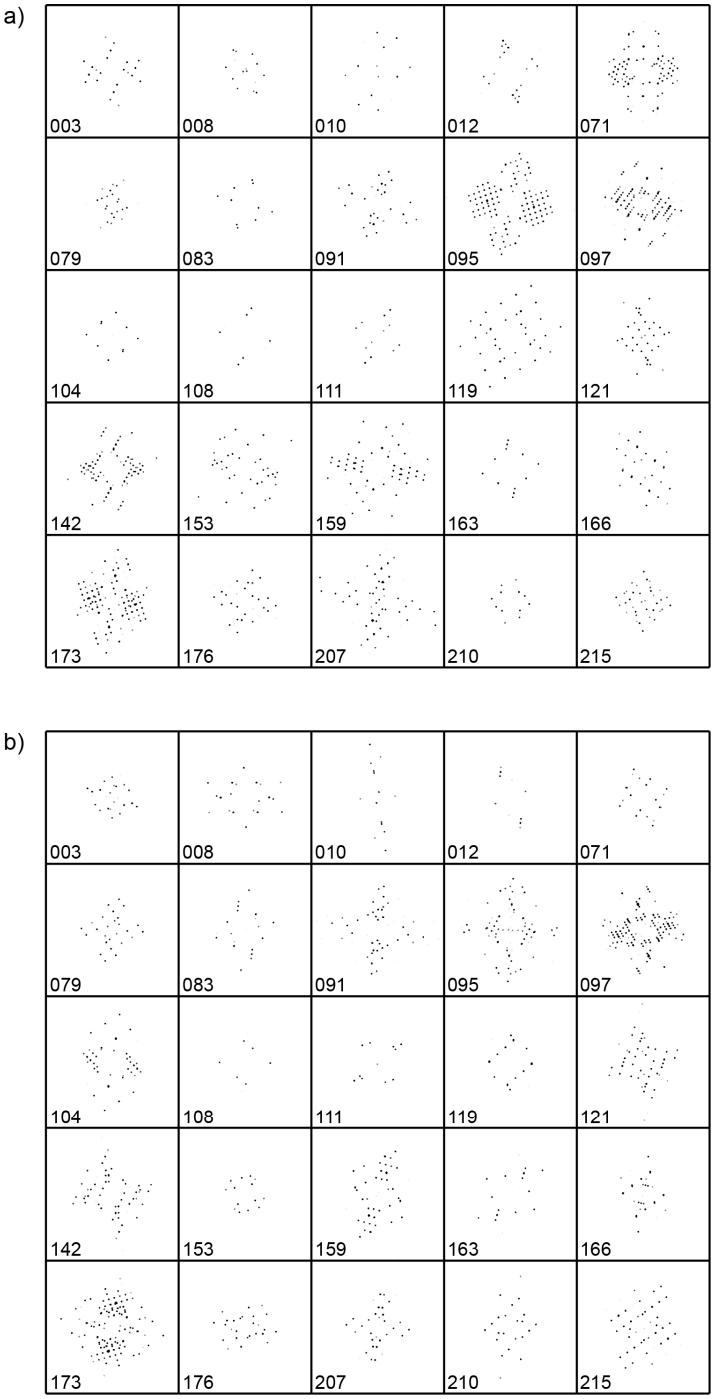

Figure 3

(a) Features extracted by the feature extraction algorithm for 25 different copy papers.

(b) Idem, for the reverse side of the sheets.

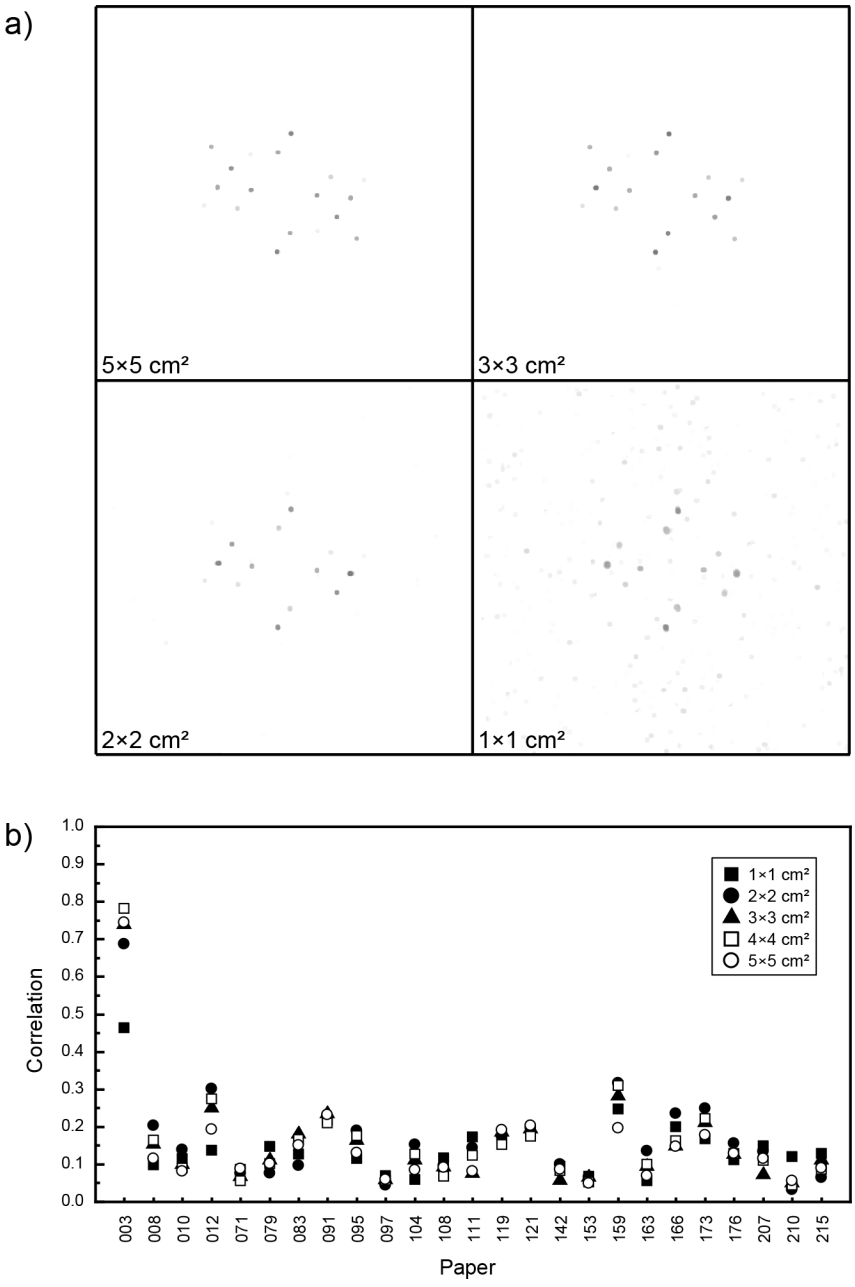

Figure 4

(a) Results of the feature extraction applied to a sample of paper 003 at sample sizes of 5×5 cm^2, 3×3, 2×2, and finally 1×1 cm^2.

(b) Correlations of the sample of paper 003 at various sizes with the 25 papers in the collection.

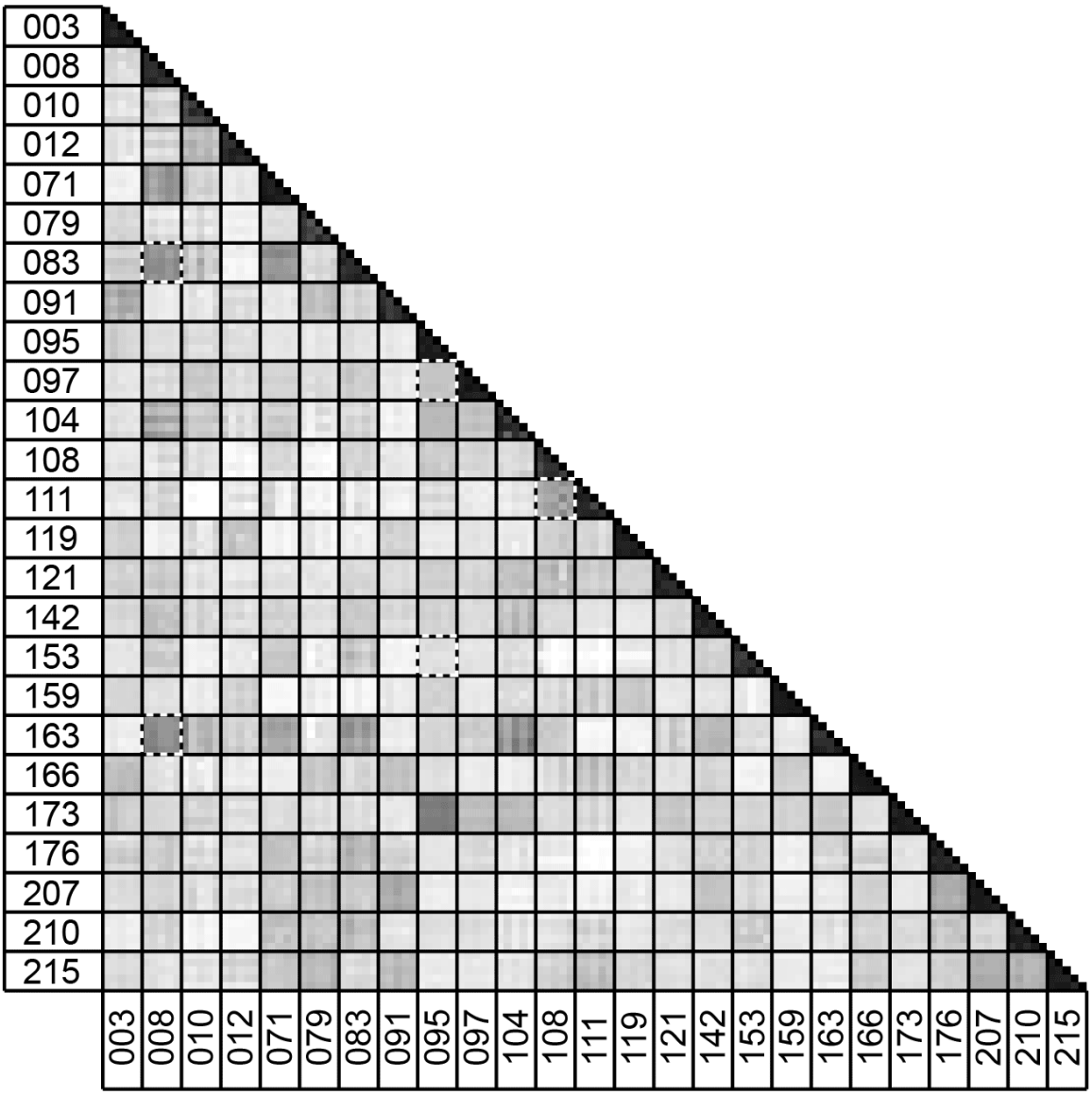

Figure 5

All 7750 comparison scores at the sheet level, where higher comparison scores give darker pixels. Five transmitted light images were obtained for both sides of the 25 sheets of paper. These 250 images gave 250 within-sheet comparison scores, and 7500 between-sheet comparison scores. The comparison scores of images compared with themselves is 1 by definition, giving black pixels along the diagonal. Dotted lines indicate comparisons of differently named papers that came from the same factory.

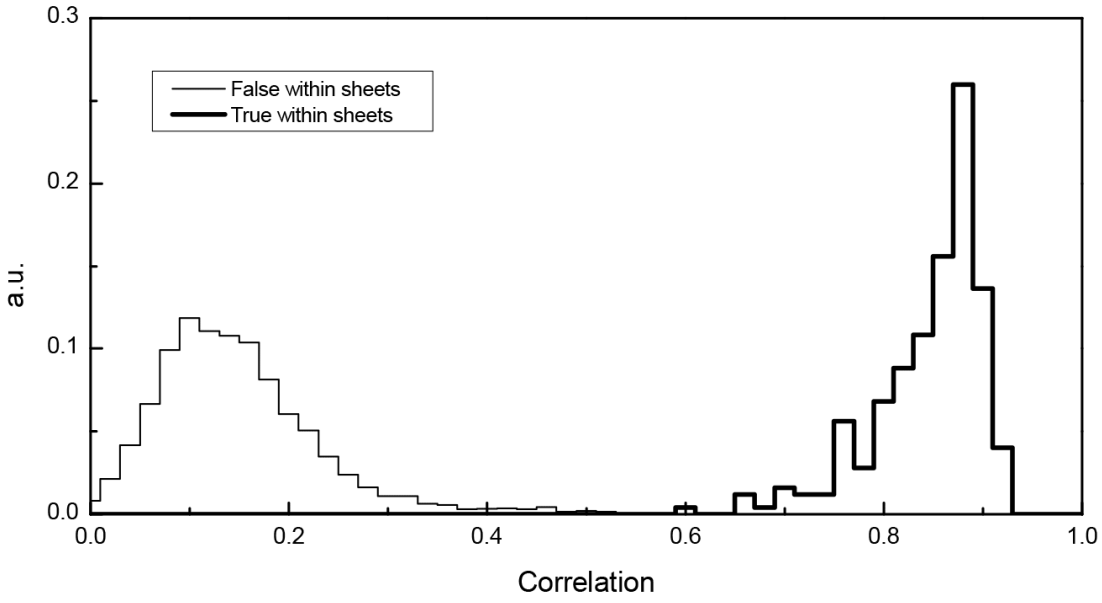

Figure 6

Histograms showing the distribution of all intra-source (within-sheet) and inter-source (between-sheet) comparison scores at the sheet level.

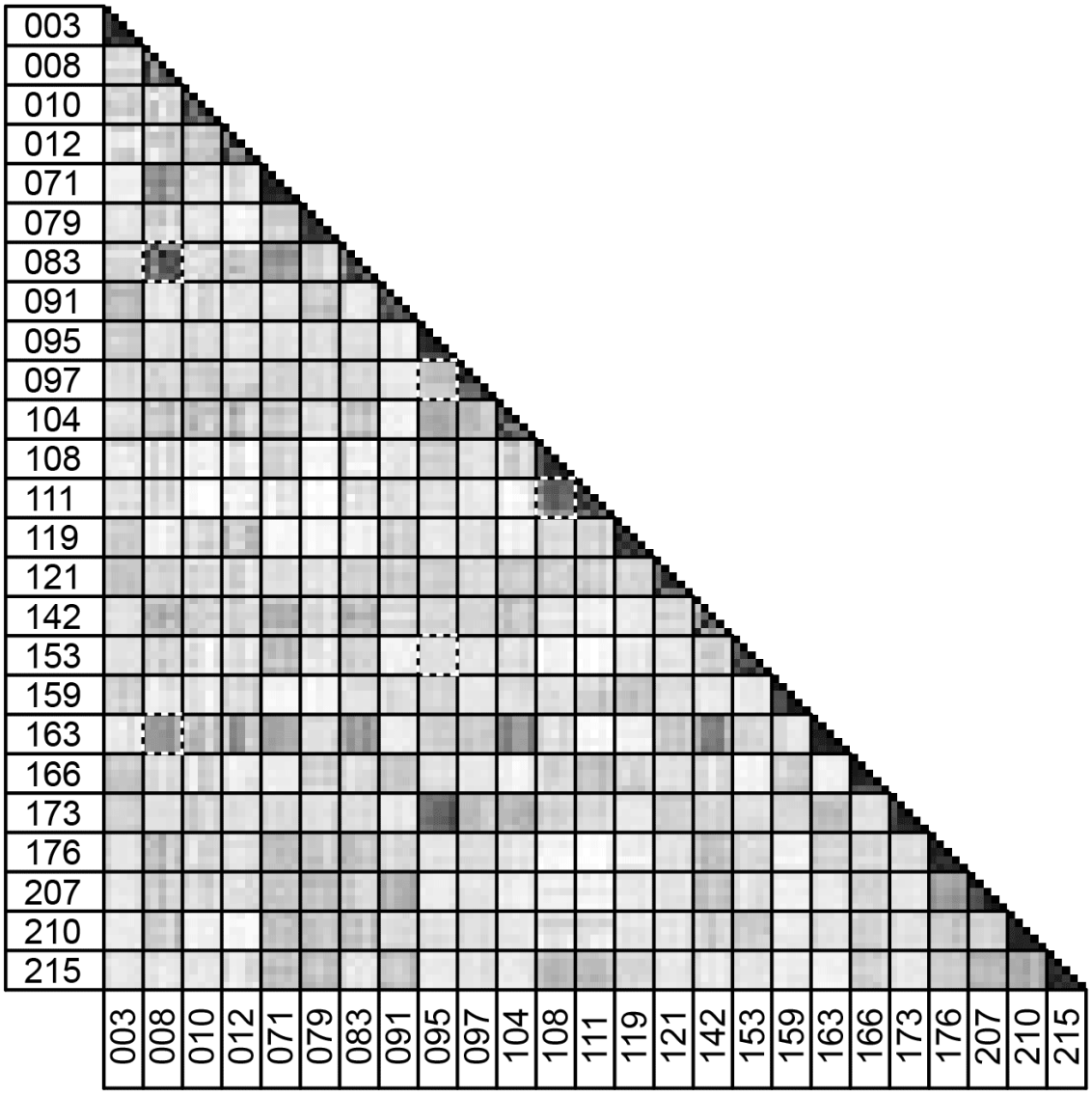

Figure 7

All 7750 comparison scores at the pack level, where higher comparison scores give darker pixels. Five transmitted light images were obtained for both sides of 5 sheets taken from each of the 25 papers. Dotted lines again indicate comparisons of differently named papers that came from the same factory.

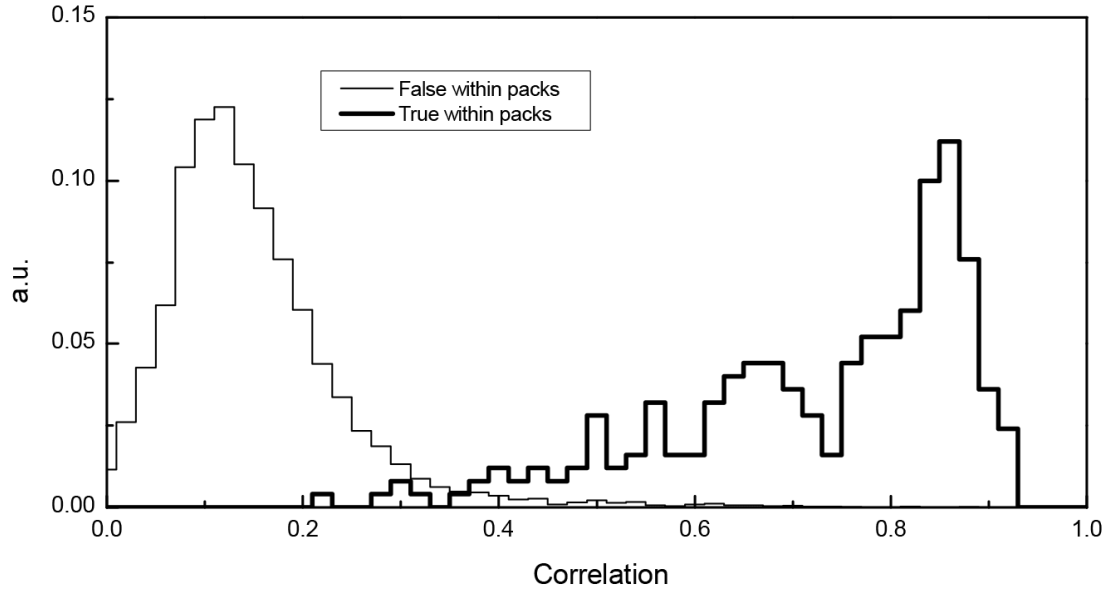

Figure 8

Histograms showing the distribution of all intra-source (within-pack) and inter-source (between-pack) comparison scores at the pack level.

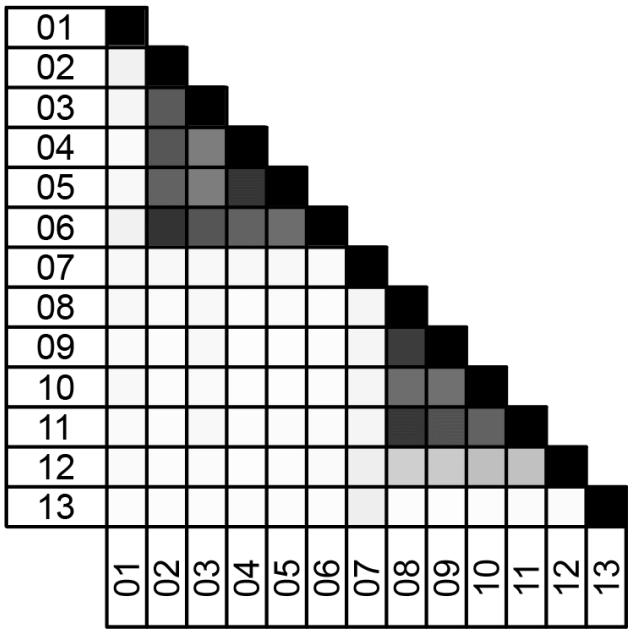

Figure 9

Comparison scores for 13 documents in a case where someone was suspected of being the source of 13 documents which he claimed came from 13 different sources. The documents were prepared over a period of 4 years.