2002 BELTWIDE COTTON CONFERENCES, ATLANTA, GA – JANUARY 8-12

IMAGE ANALYSIS FOR COTTON FIBERS
PART I: LONGITUDINAL MEASUREMENT

Y. Huang and B. Xu
The University of Texas at Austin
Austin, TX

Abstract

This paper presents the image analysis algorithms developed for a microscopic system that specifically evaluates cotton properties from longitudinal views. The system utilizes a motorized stage to transport a sample slide to grab sequential images of fiber segments, transversely scans the detected fibers in the images, and outputs both fineness and maturity measurements of cotton based on the statistic data of the massive scans across fibers. The experiments showed that the longitudinal data were repeatable and consistent with the data obtained from other testing methods.

Introduction

Much research has demonstrated that important cotton properties are measurable from the microscopic images of cotton fibers captured in both longitudinal and cross-sectional views [(Bargeron, J.D. 1986), (Barker, R.L. and Lyons, D.W., 1979), (Matic-Leigh, R. and Cauthen, D., 1994), (Thibodeaux, D.P. and Evans, J. P., 1986), (Thibodeaux, D. P. and Price, J. B., 1988), (Xu, B., Pourdeyhimi, B. and Sobus, J., 1993), (Xu, B. and Ting, Y., Part I 1996), (Xu, B. and Ting, Y., Part II 1996), (Xu, B., Wang, S. and Su, J., 1999)]. To take longitudinal measurements, fibers are usually cut into short segments/snippets, spread on a glass slide, and imaged by a video camera through a light microscope at the magnification of approximately 450x. Due to inherent variability of cotton, a reliable prediction of any cotton properties must be based on the measurements of a large number of fibers. Thus, the sample slide needs to be automatically transported on the microscope to permit as many fibers as possible to be imaged at different positions. The movements of the slide makes it difficult to keep all the images well focused. Fibers with fuzzy edges may cause errors in fiber width measurements because locating appropriate edges of a fuzzy fiber is problematic. Fibers often cross or touch each other, and measurements taken at a joint of two fibers can seriously distort the width data of these two fibers. Since cotton fibers are convoluted, the edges of fibers are often heavily curved, which adds errors to the estimation of the scanning directions across the fibers. The accuracy of the measurement is also dictated by other problems, such as merging scanned segments belonging to the same fiber. In this paper, we introduce the algorithms that deal with these problems in an effort to develop a reliable and efficient imaging system for measuring cotton fineness and maturity from longitudinal views.

Methodology

Adaptive Thresholding

A fiber image taken by a CCD camera under a transmitting light source can be formatted to an 8-bit bitmap that has grayscales from 0 (black) to 255 (white). In the image, the area covered by fibers has significantly lower grayscales than the unblocked area, and pixels near fiber edges have intermediate grayscales, forming a transition band between the fiber and the background. The width of the transition band varies with the lighting and focusing conditions. Fibers with fuzzy edges have much wider transition bands, which create problems in identifying fiber edges correctly. Figure 1 shows two images of the same fibers with different focusing and lighting conditions. Fibers in Figure 1*b* appear much coarser than those in Figure 1*a* although physically they are identical. To ensure reliable geometrical measurements of fibers, fibers in a grayscale image must be reasonably and consistently identified regardless of the image capturing conditions.

Thresholding an image is a process to separate objects from the background by using a criterion to sort every pixel in the image. A threshold can be simply calculated based on the average brightness of the image and is applied to the entire image. It was found that the mean threshold was appropriate when an image was sharp and the background is bright (Figure 1*a*), but tended to overestimate unfocused fibers in a dark image (Figure 1b). The average width of the fibers in Figure 1*a* was 1.44 times larger than that in Figure 1*b* when the images were thresholded using the mean values. The reason for overestimating fibers in an unfocused image with the mean value is that pixels in transition bands have much lower grayscales than the mean value, and are all classified as fibers pixels. The mean grayscale of an image does not correspond to the realistic edges of unfocused fibers.

The threshold should be dynamically adjusted to compensate the changes in the focusing and lighting condition, which can be reflected in the histogram of the image. A histogram is a distribution of the pixels against the grayscale, revealing allocations of pixels belonging to the fibers and background. The histograms of the two images in Figure 1 are presented in

Figure 2. As shown in Figure 2*a*, the histogram of the sharp and bright image in Figure 1*a* has two distinct peaks that appear at the two ends of the grayscale and the pixel counts decrease towards the center of the grayscale gradually. The two peaks correspond to the major portions of the fiber and background pixels (G_f and G_b), and the distributions between the two peaks are dictated by transition bands of fiber edges and variations in background. In this case, the mean grayscale (G_{mean}) seems to be an optimal threshold that separates the fibers and the background.

Figure 2*b* shows the histogram of the image in Figure 1*b*. It can be seen that when the image is dark and out of focus, the valid range of the histogram is compressed and the change in distributions occurs unevenly in the fiber and background regions. The histogram compression moves much more pixels into the fiber peak so that fibers are more likely to be overestimated when the mean grayscale G_{mean} is used as a threshold. Therefore, the new threshold should be formed by adjusting G_{mean} with the valid range of a histogram. For the unfocused image (Figure 1*b*), the valid range of the histogram was decreased by ($G_{max} - G_b$), and the threshold may be calculated as follows:

$$\text{Threshold} = G_{mean} - c^*(G_{max} - G_b),$$

where c is a coefficient that can be experimentally determined. In this study, a value around 2.0 was found to be effective for the coefficient. Figure 3 shows the binary images of the original images in Figure 1 using this adaptive thresholding method. The two binary images become much more consistent even thought the corresponding grayscale images captured under very different focusing and light conditions. The average width of the unfocused and dark fibers is just 1.08 larger than that of the focused and bright fibers.

Edge Tracing

To measure the widths of fibers, fibers need to be scanned from edge to edge transversely. Edge tracing is a step to identify the edge pixels of fibers for the transverse scanning. To present a binary image, we use “1” to stand for a black pixel (fiber) and “0” for a white pixel (background). When a black pixel is located, the algorithm needs to determine whether it is an edge pixel, and if so, what are the increments in the x and y directions for locating the next edge pixel. The judgment can be made from the presence of the eight neighbors of this active pixel (see a gray block in Figure 4). There are 256 possible combinations regarding the presence of the eight neighbors. As shown in Figure 4, a ranking number from 0-7 is assigned to every neighbor.

A code representing the neighbors of the current pixel can be calculated with the statuses and ranking numbers of the eight neighbors:

$$\text{Code} = \sum_{i=0}^{7} P_i \times 2^i$$

where P_i is a binary variable indicating the presence status of the i^{th} neighbor. P_i=1 when the i^{th} neighbor is a black or fiber pixel. Otherwise, P_i=0. For each code, one can analyze the available neighbors around the active pixel, and determine if this pixel is an edge pixel and which neighbor should be traced next. The 256 codes and the corresponding neighboring situations constitute a look-up table that provides a quick solution to the computer when it traces fiber edges. Since some pixels are simply noise on the edge (a small branch or a short bridge), they should not be traced, and the corresponding codes should not be included in the look-up table. Figure 5 gives some examples of these invalid codes when the current pixel is bounded with 1 to 4 isolated black pixels, or surrounded by all eight neighbors. The first four codes in the figure represent cases where the current pixel and its neighbors form branches (codes 1 and 5) and bridges (codes 49 and 165), and the last code (255) indicates a non-edge pixel because the active pixel is encompassed by all black pixels.

If a code represents a valid edge pixel, the algorithm finds the next pixel and the x and y increments from the active pixel. Since the increments are direction dependent, four separate look-up tables can be created based on the four tracing directions: up-right, up-left, down-right or down left. Figure 6 shows the valid codes in the look-up table in the up-right direction. One code in this table indicates that a new edge pixel can be traced from the active position in this direction, and the x and y increments are given by the relative position of the new pixel. In some cases, a pixel may have multiple neighbors that can be traced in the tracing direction (see code 31 in the up-right direction). The next neighbor to be traced is chosen in the counterclockwise order. When the neighbors around the active pixel do not match any of these templates, a look-up table in another direction is called for further comparisons. The current tracing is terminated if the validity checking from the four look-up tables is false.

Transverse Scanning

Transverse scanning is a step to determine the fiber width at an edge point in a direction perpendicular to the edge of the fiber. To conduct transverse scanning along a fiber, an interval that controls the spacing of two consecutive scans must be given and the direction of the fiber edge at a scanning point needs to be estimated. Suppose the current edge pixel is $0(x_0, y_0)$

and the next edge pixel $1(x_1, y_1)$ (see Figure 7*a*). The distance between $0(x_0, y_0)$ and $1(x_1, y_1)$ is equal to the scanning interval. The edge direction at $1(x_1, y_1)$ can be approximated by the slope of the line connecting these two pixels. The transverse scanning starts at $1(x_1, y_1)$ and follows a direction perpendicular to the calculated edge direction until it meets an edge pixel $2(x_2, y_2)$ on the other side of the fiber. The distance between $1(x_1,y_1)$ and $2(x_2, y_2)$ is a local measure of the fiber width. Figure 7*a* shows the transverse scanning at a 10-pixel interval. Since cotton fibers are convoluted, the edges of fibers are often curved. The calculated directions at some scanning pixels may not realistically indicate the edge directions, causing errors in the width measurements. This error can be reduced by rescanning the fiber using the directions calculated from its axis. After the initial scanning, the middle points of all the scans are connected, and the direction perpendicular to the middle axis at each middle point is used to search for edge pixels on both sides of the axis. It can be seen from Figure 7*b* that the line connecting two middle points gives a direction that more precisely represents the local directions of both sides of the fiber. This double scanning algorithm also generates much more uniform scans on both sides of the fiber.

When immature cotton fibers exist in the image, the transverse scanning must deal with gaps of various sizes within fiber regions. Immature fibers have thinner walls and therefore are more transparent than mature fibers (Figure 8*a*). After thresholding, the bright portions in the immature fibers may become holes (white pixels) in the binary image (Figure 8*b*). In order to discern holes within a fiber from background gaps between fibers, the grayscale information of those areas in the original image must be utilized. As shown in Figure 8*a*, the grayscale of immature fibers are differentiably lower than the background. If a discontinuity occurs during the transverse scan on the binary image, the corresponding pixel in the original image is checked. If the grayscale is lower than the background (G_b), the pixel is considered as fiber holes and the scanning continues. The transverse scans over the immature fiber were also shown in Figure 8*a*. The detection of these highly transmissive regions is particularly useful for identifying dead cotton fibers.

Scan Validation

Validation is a step to avoid or remove improper scans during or after the transverse scanning. Improper scans are those on various debris and intersections of two or more fibers (Figure 9*a*). Since fibers may intercept each other at any angle, crossing detections need to be conducted at multiple stages. During the scanning, each scan is examined with the following rules:

1. The length of the scan should not exceed the preset limits for the minimum and maximum widths of cotton fibers. In this study, the lower and upper limits were set to 3*um* and 65*um*, respectively. The two limits give a sufficient range for valid cotton widths, but exclude scanning on extremely small or large objects in the image.
2. The length difference between two consecutive scans should not exceed 50%. A sudden change in scan length suggests a joint of another object to the current fiber.
3. The direction change between two consecutive scans should not exceed 20^{o}. A sudden change in scan direction also suggests a crossing point with another fiber.

If a scanning procedure violates one of these rules, it will be terminated. A new scanning procedure starts at the next valid edge pixel. Because of the interruptions, a fiber may have multiple scanned segments of various lengths. The starting and ending points of each segment are registered (Figure 9*a*). After the entire image is processed, all the scans need to be further checked to find fiber crossing sections and short debris undetected during the scanning with the following rules:

1. The coefficient of variance (*CV*) of the scan lengths of one scanned segment should not exceed 0.65. When the *CV* of a segment is higher than 0.65, the segment is considered having too large variations in scan width that may be caused by touching or crossing sections of two objects.
2. The length of one scanned segment should be three times larger than the average width of the segment. Thus, scans on short debris can be deleted.

If a scanned segment violates either rule, all the scans in this segment will be deleted. Figure 9*b* presents transverse scans of fibers after applying these validation rules to Figure 9*a*, demonstrating that the validation is crucial for improving the accuracy of the data.

Merging

One of the important tasks to be done in the longitudinal analysis is to assess the maturity of cotton fibers based on the fiber convolutions [2, 4]. To analyze the longitudinal convolution, a fiber should be scanned to have multiple twists so that the adequate information about fiber width variations can be obtained. Since the transverse scanning along a fiber may be interrupted by joints with other objects, the scanned fiber may contain several scanned segments, each of which may not be long enough to have a complete twist. Hence, separate segments belonging to the same fiber need to be connected.

In order to avoid false connections between two segments, two parameters are used to judge whether the connection can proceed. The first parameter is the distance between the ending point of the first segment and the starting point of the second segment. If the distance is within three times of the average width of the first segment, the two segments are then considered for merging. Figure 10 shows an example of the segment merging. For the ending point of segment 1, only the starting points of segments 2 and 3 are within the given range, and therefore they are the candidates for a possible connection. The second parameter is the angle between the two segments at their connecting ends. Assume that the directions of segments 1 and 2 are α_1 and α_2. If the difference between α_1 and α_2 is within a given tolerance (20^o), i.e., $|\alpha_2 - \alpha_1| < 20^o$, the connection between the segments should be forwarded. This requirement eliminates segment 3 in Figure 10 for the connection with segment 1. The third parameter is the direction (α_3) of the line connecting the two ends. α_3 should be within angles α_1 and α_2. This requirement prevents connecting two parallel segments which are close to each other but not from the same fiber. There are four correct connections made in Figure 10.

Experiment

The experiment of longitudinal measurements of cotton fibers were performed on a customized imaging system that includes a video zoom microscope, a B/W CCD camera, and a frame grabber. The imaging system can also analyze fiber cross sections, which will be reported in the next paper of this series. The image resolution was 1.86*um*/pixel. The microscope was equipped with a motorized stage that automatically transports the sample slide to allow the camera to grab fiber images at many different positions. During the travel to the next position, fibers in the grabbed image were scanned and the image was discarded. The sample preparation was done with a special fiber cutter and spreader. Fibers were cut into 0.5 mm long segments, and then randomly spread on a microscope slide.

A software package was developed to implement the algorithms discussed above. For each scanned fiber, the software calculates the number of scans (N_s), the length of the scanned segments (L_s), the maximum (W_{max}), minimum (W_{min}), mean (W_{mean}) and standard deviation (W_{sd}) of fiber widths, and the number of twists (N_t). N_t is counted by the alternations of the maximum and minimum widths along a fiber axis. After the scanning of an entire slide, the software can output the statistics and distributions of all the data. The number of the scanned fibers (T_f), varying from 1000-4000, depends on the density of fiber segments spread on the slide. These outputs provide direct measurements for both fiber fineness and maturity of the analyzed samples.

It is important to know the correlations of the data from this longitudinal measurement method with the data obtained from other testing methods. Seven varieties of cottons were tested by the imaging system for both longitudinal and cross-sectional measurements, and by the Advanced Fiber Information System (AFIS). Figure 11 shows the correlation of the average width (W_{mean}) and the average perimeter of cross sections (P) of the cottons, and the correlation of the average the W_{mean} with the AFIS fineness data (F). Both analyses prove reasonably good correlations among these methods.

Since the imaging system can produce a large quantity of transverse scanning data from one slide, reliable distributions of fiber widths can be calculated. Figure 12 shows the distributions of fiber widths of the four international-calibration-standard cottons labeled as A21, C37, D6 and G21. D6 and G21 have very similar mean widths, but their width distributions are significantly different. G21 has a much broader distribution and therefore a larger variance than D6. It is expected that G21 contains more immature fiber than D6 because immature fibers have higher convolutions. A21 and C37 have very similar distribution shapes but different peak locations. Statistically, A21 is coarser than C37.

The repeatability of the data generated from the software was tested by repeatedly scanning a sample slide under the same condition. The coefficient of variance (*CV*) of the data from multiple tests was used to evaluate the repeatability. Table I gives the data of a sample slide that was scanned six times. The *CV*s of the data from different tests are under 4.5%, which demonstrates a good consistency of the data. Since the scanning of each test could not be started exactly at the same position on the sample slide, the images grabbed in different tests contained different portions of fibers, and the *CV*s might be originated primarily from the inherent variations of fibers.

A test on a bale of cotton was conducted to investigate the level of variability of the longitudinal data across the cotton bale through multiple samplings, and to find out how many fibers need to be scanned so that the data are sufficient and reliable for estimating the attributes of the whole bale. 25 samples were taken from the bale at different places, and three slides were prepared for each sample. Table II shows the results of the 25 samples. For one sample, each measurement in the table is the average of all fibers on three slides. Due to the difficulty in controlling the densities of fibers on the slides, the total number of scanned fibers (T_f) for a sample varied greatly, which in turn influences the number of scans (N_s), the scanned length (L_s), and the number of twists (N_t) on single fibers. Thus, the *CV*s of these four parameters do not solely reflect the variability of cotton. However, the *CV*s of the four width measurements, W_{max}, W_{min}, W_{mean} and W_{sd}, more closely link to the variability of cottons, and they are significantly higher than those of fibers on one slide (Table I).

When the data from all the 25 samples (177,016 fibers on 75 slides) were pooled together, the *CV*s of the fiber width measurements could be calculated with different numbers of fibers. Figure 13 shows the curves of W_{mean} and W_{sd} against the T_f of the fibers taken into account. Initially, the *CV* increases with the T_f, meaning that sampling more fibers adds information about the variability of cotton in the bale. After the T_f passes 3×10^4, the *CV* tends to approach to a stable level with a much smaller scale of fluctuations. This means that increasing T_f in the calculations does not increase the variability information. To overcome the variability of samples, a bale of cotton should be sampled to have at least 3×10^4 fibers in the calculation.

Summary

This paper presented the image-processing algorithms developed specifically for analyzing longitudinal images of cotton fibers in an automatic system. The effort in the research was made to improve the accuracy and efficiency of the system. The adaptive thresholding method reduces the errors arising from unfocused fibers. The look-up table increases the efficiency of tracing fiber edges. The double-scanning algorithm enhances the accuracy of transverse scans of fibers. The validation rules prevent false scans from being included in the output. The merging algorithm links short segments that belong to the same fiber so that fiber twists can be evaluated. The longitudinal data generated from the system showed good correlations with the data obtained from the other methods.

Acknowledgement

We are grateful for the financial support provided by Texas Food and Fiber Commission to this research. We would like to thank Mr. Eric Hequet, the International Textile Center at Lubbock, and Dr. D. Thibodeaux, the Southern Regional Research Center at New Orleans, for the cotton samples and the fiber cross-section and AFIS data used in the research.

References

Bargeron, J.D. (1986), Preliminary Investigation of The Measurement of Cotton Fibers with the Peyer Texlab System: Comparability and Reliability, *Textile Res. J.*, **56**, 121-124.

Barker, R.L. and Lyons, D.W., (1979), Determination of Fiber Cross-Sectional Circularity from Measurements Made in A Longitudinal View. *Journal of Engineering for Industry*, **101**, 59-64.

Matic-Leigh, R. and Cauthen, D., (1994), Determining Cotton Fiber Maturity by Image Analysis. *Textile Res. J.*, **64**, 534-544.

Thibodeaux, D.P. and Evans, J. P., (1986), Cotton Fiber Maturity by Image Analysis, *Textile Res. J.*, **56**, 130-139.

Thibodeaux, D. P. and Price, J. B., (1988), Reference Method for Determination of the Maturity of Cotton Fibers. *Melliand Textilberichte*, **70**(4), 243-246.

Xu, B., Pourdeyhimi, B. and Sobus, J., (1993), Fiber Cross-Sectional Shape Analysis Using Imaging Techniques, *Textile Research Journal*, **63**, 717-730.

Xu, B. and Ting, Y., (1996), Fiber Image Analysis, Part I: Fiber Image Enhancement, *J. of Textile Institute*, **87**, 274-283.

Xu, B. and Ting, Y., (1996), Fiber Image Analysis, Part II: Measurement of General Geometric Properties of Fibers. *J. of Textile Institute*, **87**, 274-283.

Xu, B., Wang, S. and Su, J., (1999), Fiber Image Analysis, Part III: Autonomous Separation of Fiber Cross Sections, *Journal of Textile Institute*, **90**, 288-297.

Table I Repeatability Test

Test	T_f	N_s	L_s (*um*)	W_{max} (um)	W_{min} (um)	W_{mean} (um)	W_{sd} (um)	N_t
1	1669	50.23	171.89	19.46	10.92	15.25	4.44	2.25
2	1694	50.30	171.94	18.81	10.42	14.67	4.41	2.36
3	1743	51.83	178.74	19.04	10.44	14.76	4.43	2.34
4	1715	53.40	184.67	19.02	10.27	14.70	4.52	2.38
5	1576	52.44	180.13	18.81	9.98	14.82	4.56	2.54
6	1593	52.56	179.34	19.06	10.17	14.67	4.57	2.44
CV (%)	4.03	2.49	3.12	1.40	3.27	1.59	1.46	4.41

Table II Longitudinal Data of Samples from the Same Cotton Bale

Sample	**T_f**	**N_s**	**L_s (*um*)**	**W_{max} (um)**	**W_{min} (um)**	**W_{mean} (um)**	**W_{sd} (um)**	**N_t**
1	9438	42.19	146.43	20.57	12.22	16.24	4.32	2.01
2	9586	41.55	144.51	20.35	12.08	16.06	4.24	2.01
3	5840	46.78	162.75	19.62	11.04	15.21	4.43	2.25
4	6202	49.48	171.92	19.60	10.82	15.06	4.51	2.37
5	4918	50.17	172.36	19.68	11.33	15.34	4.18	2.42
6	6825	46.43	160.62	19.38	10.96	15.05	4.36	2.24
7	6354	49.19	170.24	19.86	11.32	15.51	4.39	2.26
8	9080	44.97	156.45	20.21	11.69	15.86	4.42	3.04
9	7157	48.58	169.30	19.36	10.80	14.93	4.38	2.40
10	7213	47.27	163.38	19.39	11.07	15.12	4.27	2.31
11	7770	47.15	163.62	19.63	10.95	15.22	4.49	2.26
12	5280	48.89	169.33	19.42	10.85	15.03	4.40	2.34
13	6993	46.80	161.92	19.83	11.26	15.37	4.38	2.26
14	6311	49.83	172.06	19.46	10.74	15.05	4.48	2.46
15	7302	42.84	149.58	19.11	10.84	14.84	4.29	2.18
16	6883	44.83	156.31	18.95	10.83	14.80	4.23	2.19
17	7262	47.83	165.99	19.30	10.69	14.91	4.45	2.33
18	7600	44.87	156.23	18.99	10.82	14.76	4.24	2.23
19	6908	48.35	166.72	19.96	11.19	15.45	4.51	2.25
20	7984	46.57	162.10	19.68	11.23	15.35	4.31	2.25
21	6973	46.89	162.48	19.58	11.11	15.21	4.34	2.32
22	6832	47.12	164.38	19.33	11.06	15.07	4.21	2.26
23	8758	44.01	152.94	18.97	10.79	14.78	4.25	2.15
24	6388	45.99	160.05	20.18	11.77	15.82	4.29	2.13
25	5159	43.65	152.46	20.68	12.11	16.28	4.37	2.21
CV (%)	17.31	5.13	4.90	**2.41**	**4.04**	**2.91**	**2.23**	8.37

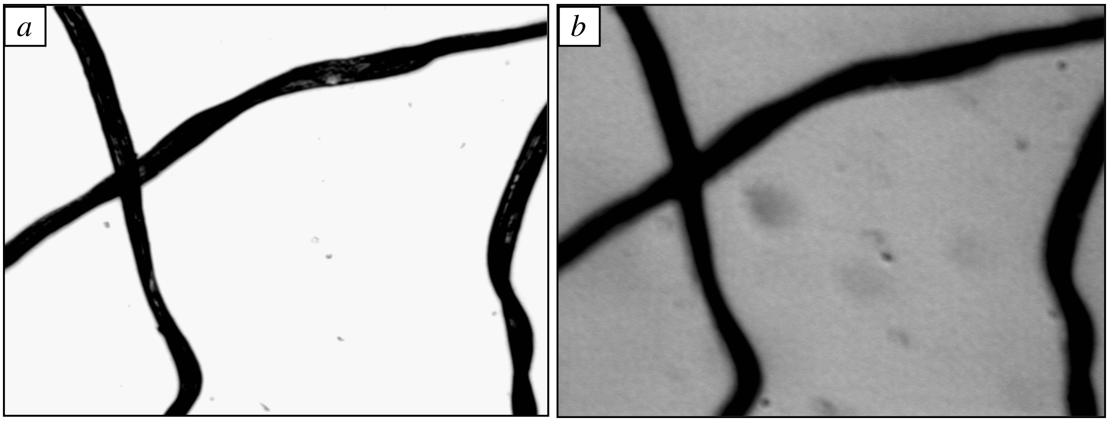


Figure 1. Images with Different Focusing and Lighting Conditions, *a*: a focused, bright image; *b*: an unfocused, dark image.

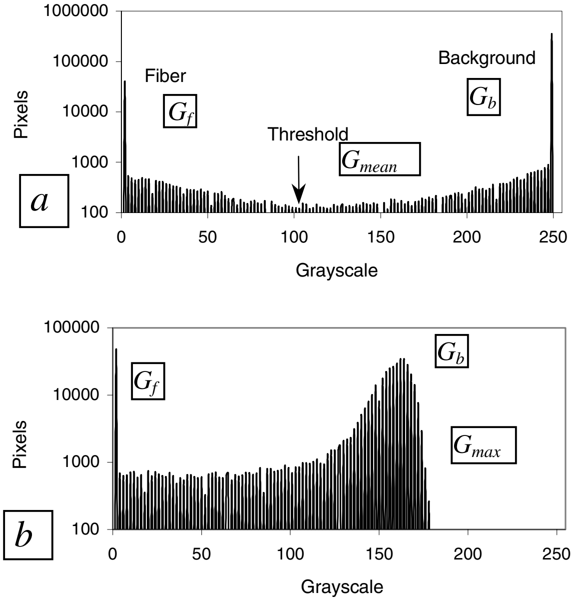


Figure 2. Histograms of the Images in Figure 1.

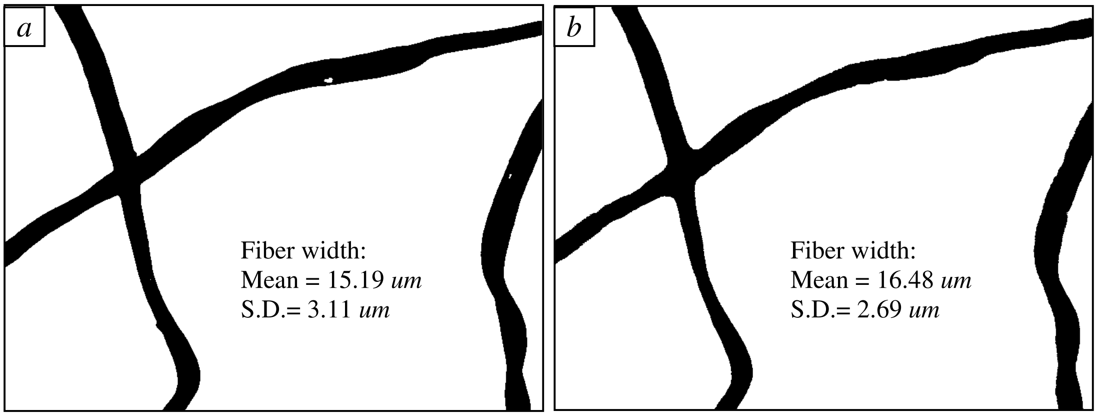


Figure 3. Adaptive Thresholding of the Images in Figure 1.

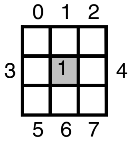


Figure 4. Numbering Eight Neighbors of the Active Pixels (grayed).

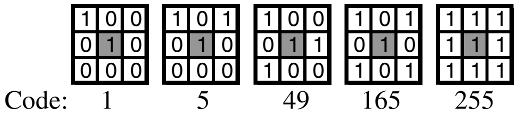
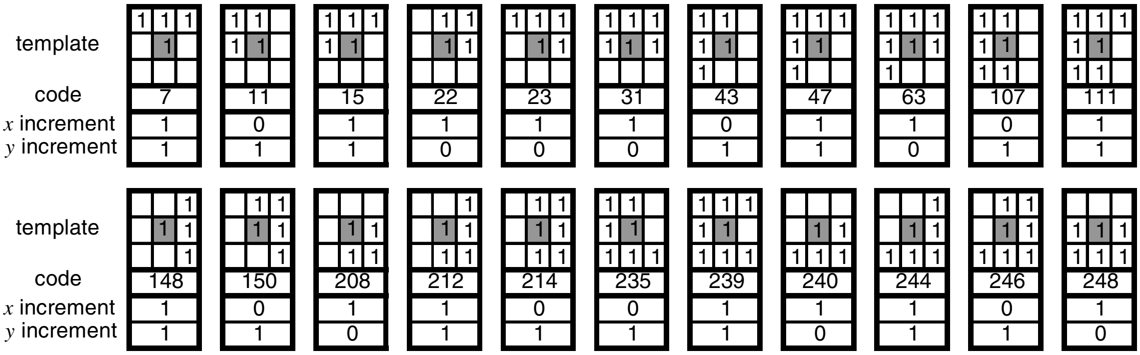


Figure 5. Examples of Invalid Codes.



code	7	11	15	22	23	31	43	47	63	107	111
x increment	1	0	1	1	1	1	0	1	1	0	1
y increment	1	1	1	0	0	0	1	1	0	1	1

code	148	150	208	212	214	235	239	240	244	246	248
x increment	1	0	1	1	0	0	1	1	1	0	1
y increment	1	1	0	1	1	1	1	0	1	1	0

Figure 6. Look-Up Table for Edge Tracing in the Up-Right Direction.

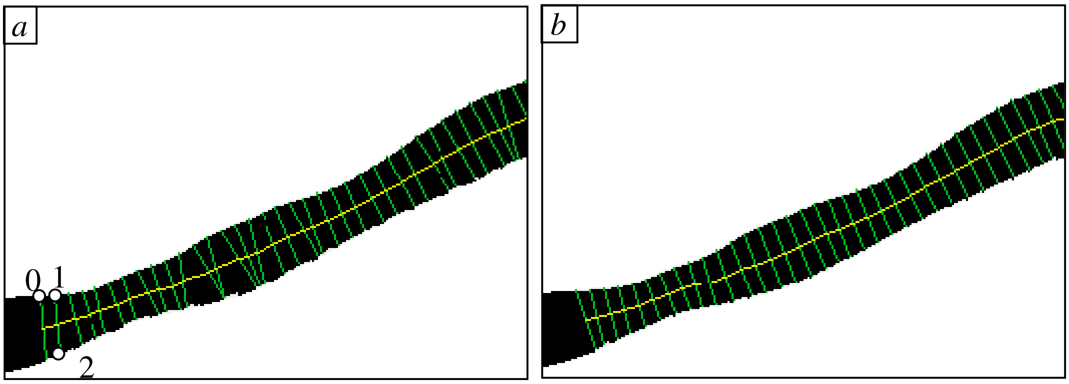


Figure 7. Transverse Scanning.

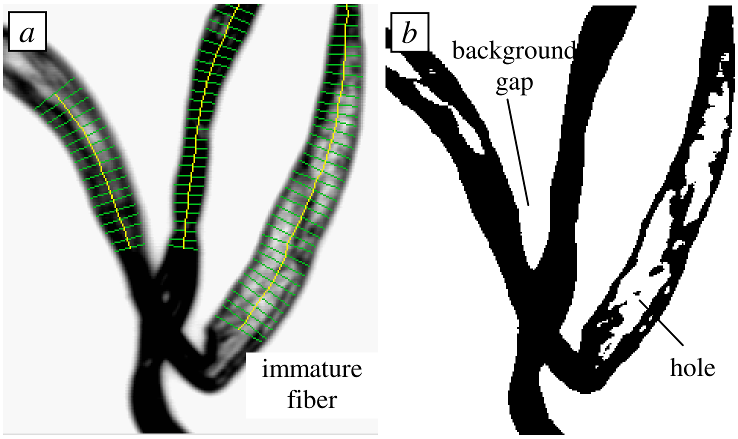


Figure 8. Transverse Scanning Over Immature Fibers.

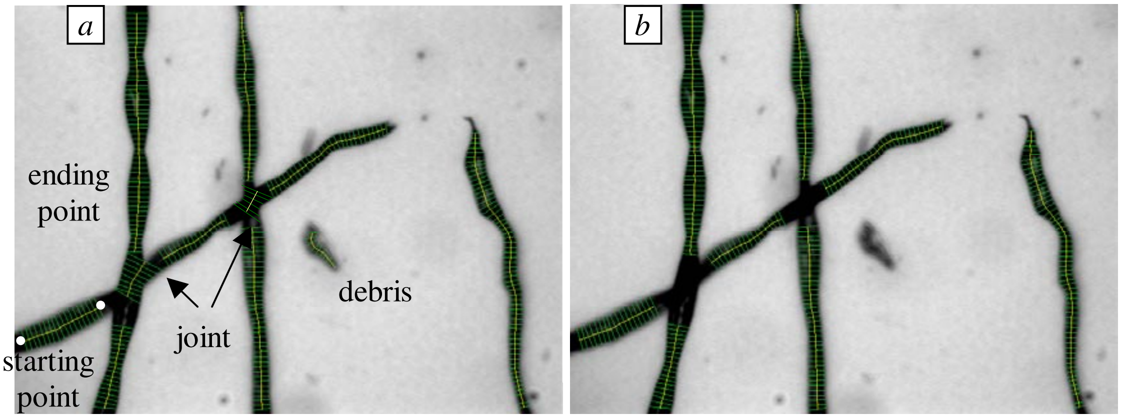


Figure 9. Scan Validation, *a*: before validation; *b*: after validation.

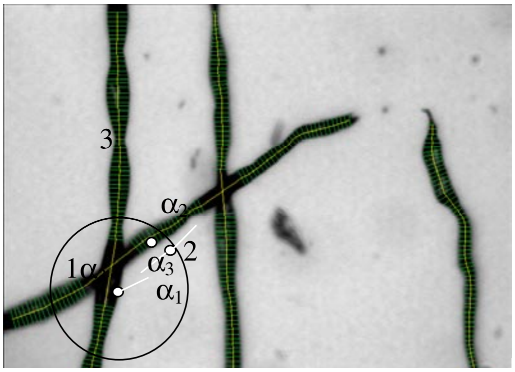


Figure 10. Merging of Scanned Segments.

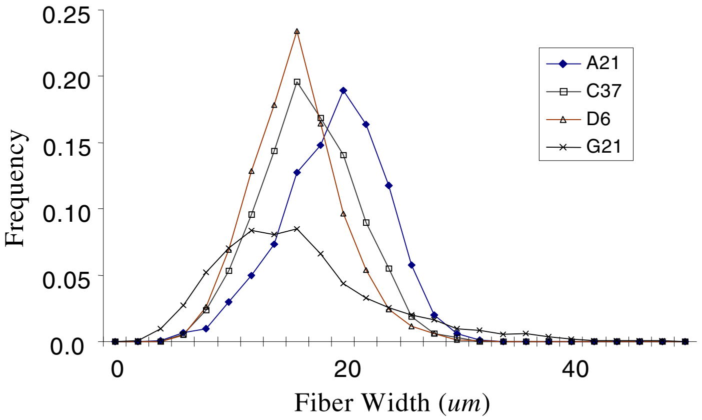


Figure 11. Correlations of Longitudinal Data with Cross-sectional Data (*a*) and AFIS Data (*b*).

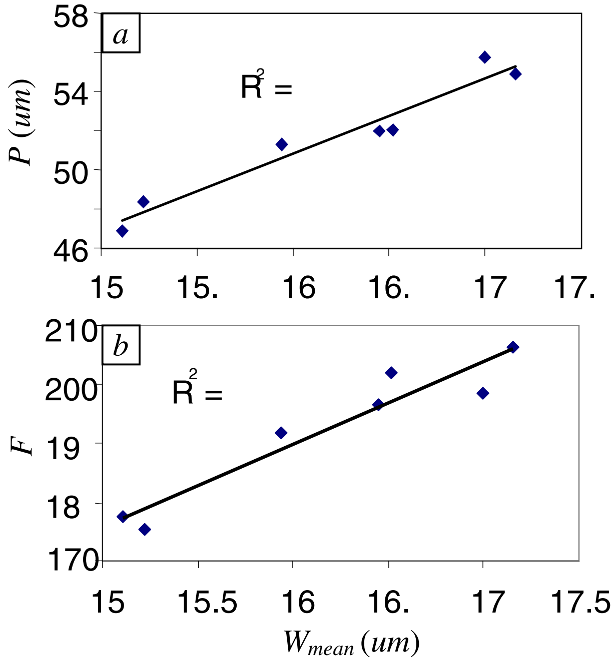


Figure 12. Distributions of Fiber Widths.

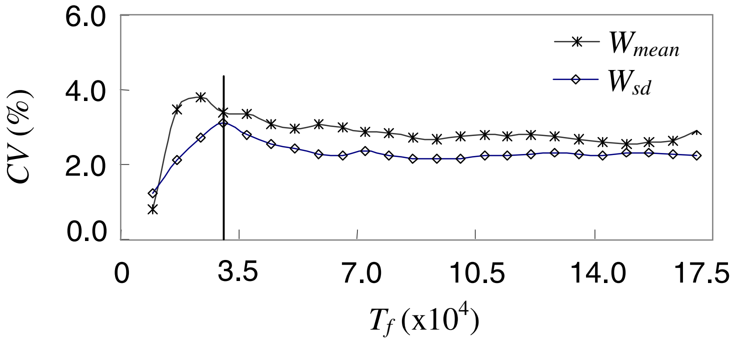


Figure 13. Variations of Fiber Widths in Different Sets of Samples.