POLYNOMIAL PULP FIBER MODELING

Heikki Huttunen, Hannu Oinonen, Jyrki Selinummi, Pekka Ruusuvuori, and Ville Voipio

Department of Signal Processing
Tampere University of Technology, Tampere, Finland

ABSTRACT

This paper proposes an automatic measurement algorithm for calculating statistics of pulp fiber lengths and curvature. The fibers are extracted from digital images, and a parametric model is fit to the extracted objects. The modeling approach improves the robustness of the detection and enables connecting broken fiber parts into full length curves. The developed algorithm is compared to an alternative method, and concluded to yield similar results while providing additional features, such as curvature estimates, of the detected fibers.

Index Terms— Pulp, Fiber, Image Analysis, Length Distribution, Image Modeling

1. INTRODUCTION

Measurement of the properties of pulp fibers is widely used in paper making industry for improving process efficiency. By far the most common measurement method is based on *off-line* assessment, where samples of the pulp material are first manually acquired at constant time points, and processed for analysis by scanning. Subsequently, the sample is measured for different properties either by hand, or with two or three dimensional automated image analysis, the methods of which have been thoroughly studied (see, e.g., [1, 2, 3, 4, 5, 6]). The off-line nature of the measurement allows for high accuracy data, but results in low speed processing.

This paper considers automatic image analysis for *in-line* measurements, where the target images are automatically acquired from pulp-fiber solution [7] using rapid laser flash illumination. The benefits of the in-line approach are evident: the processing is faster, the feedback is more responsive, the possibility of a sampling error is eliminated, and it requires a minimal amount of manual labor. On the other hand, the produced images have low signal to noise ratio, and the extent of fiber overlap becomes uncontrollable, since no manual preprocessing of the sample is performed.

The approach described in this paper relies on modeling the fibers using polynomial curves. There exists vast literature on curve fitting for images [1, 2], and it has previously been used for cotton fiber modeling [8]. However, the curved appearance of cotton fibers required piecewise polynomials and designated control points, unlike our more straightforward approach. Methods for fiber measurements also exist [9], but a polynomial model for the fiber appearance enables analytical calculation of many quantities, such as the curvature. The introduced method allows for efficient detection of the fibers in noisy images, and can also be used as a tool for joining fiber parts that have been falsely broken due to low quality input images.

2. FIBER SEGMENTATION

A typical input image after histogram equalization in shown in Figure 2 (a). The example illustrates the challenges of in-line image analysis: the image is noisy, fibers overlap, and it is difficult even for a human observer to reliably separate individual fibers. Figure 1 presents the framework of the proposed system, consisting of five functional blocks. The procedure starts with an optional preprocessing step. The second step, segmentation, extracts the fibers from the background. Thirdly, resulting objects are filtered by their shape. Then, skeletonization and subsequent pruning is applied to thin the fibers and remove short erroneous branches. Finally, a parametric model is fit to the skeleton data, after which required features can be measured from the resulting fibers.

The extraction of the fiber foreground areas from the darker background consists of two steps: denoising and segmentation step. Preprocessing enhances the image data allowing for more reliable segmentation. However, the required procedure depends on various conditions, such as the imaging hardware, exposure time, etc., which essentially determine the required filtering operations. In our case, the acquired images appear as shown in the Figure 2 (a). The noisy appearance of the image is obvious, and our choice of pre-processing filters consists of a combination of a median filter (for filtering the impulsive noise) and a Gaussian lowpass filter (for smoothing the additive noise.

The performance of the segmentation step is critical to the performance of the subsequent stages of the method. Out method is based on P-tile thresholding due to its simplicity and adjustability. We did experiment with more elaborate methods including the Otsu threshold [10] and energy minimization with the graph cuts [11], but they did not seem to improve the overall perfomance. This is probably due to the subsequent robust fiber modeling stage that makes a smooth

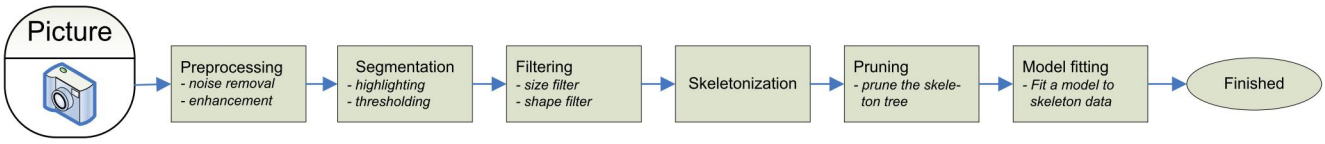

Fig. 1: Flowchart of proposed system.

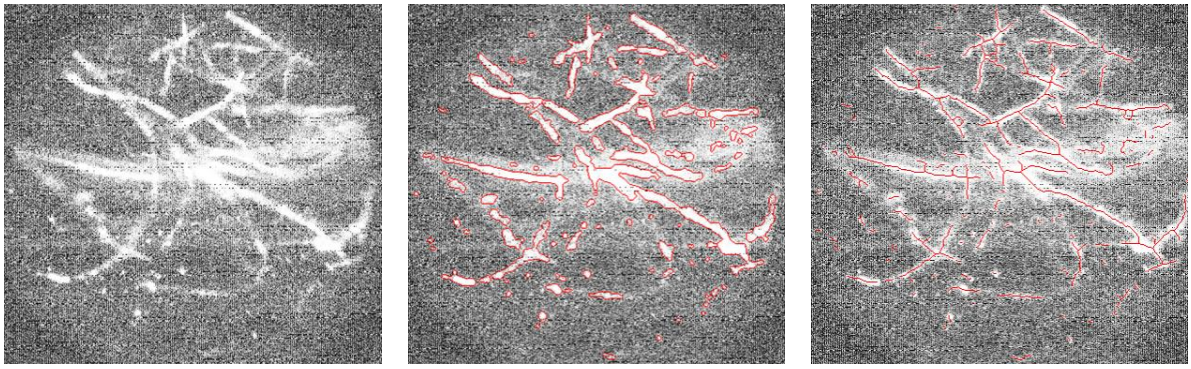

Fig. 2: Fiber detection example. (a): Typical fiber image after histogram equalization, (b): segmentation result superimposed over the original, (c): pruned skeleton of segmentation result.

segmentation result obsolete. The segmentation procedure for separating the fibers from the background was performed in the following way. First, we applied the standard top-hat filter highlighting local maxima to emphasize narrow objects, such as the fiber pixels. Second, we used adaptive P-tile-thresholding to set P percent of pixels as background, separating fiber pixels from the background. This adaptive algorithm proceeds iteratively as follows:

1. Threshold the image using P-tile-thresholding with $P = 80\%$.
2. Measure the ratio between foreground and background pixels outside the round area of interest.
3. If the ratio exceeds a constant C, increase P by a predefined value, and return to step 1. Otherwise, accept the thresholding result.

This iterative method successfully locates the fiber pixels, since there should be only a few foreground pixels outside the area of interest. Therefore, when the value of parameter P is too small, the thresholding tends to overestimate the foreground area, and the value of P must be set higher. This adaptation was required since the number of the fibers (foreground pixels) varied greatly from image to image.

After global thresholding, excess objects are often present in the binarized images. Unlike true large objects, these are caused by noise, and can be removed based on their size. In our test images, 15 pixels was found suitable limit for minimum object size, but an optimal limit has to be determined experimentally. Figure 2 (b) shows the segmentation result superimposed on the original histogram equalized image.

The ideal representation of a fiber is a thin rigid curve. The segmentation, however, produces only a binary map of fiber area, but this map can be transformed into a set of curves by skeletonization. Again due to noise, the skeletonized image contains a number of extra branches correctable by pruning. The pruning simply removes branches according to certain criteria, for example the branch length (5 pixels in our data), the limit of which can only be obtained by experiment. This pruned skeleton is illustrated in Figure 2 (c).

3. FIBER MODELING

After skeletonization, the individual fibers have to be extracted. In our approach, we model the fiber structure with a polynomial parametric model, which is straightforward to implement, but simultaneously offers flexibility to represent various fiber shapes. The modeling starts by separating the fiber mesh at each intersection point. This ensures that the shape of each fiber segment is simple enough to be described with a low order polynomial.

The polynomial models both horizontal and vertical coordinates as a polynomial of a real valued parameter $t \in [0, 1]$ as $\mathbf{x}(t) = (x(t), y(t)) \in \mathbf{R}^2$. More specifically, suppose that we have labeled a connected segment of the fiber skeleton that consists of N pixels with coordinates $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$. Then the parametric linear

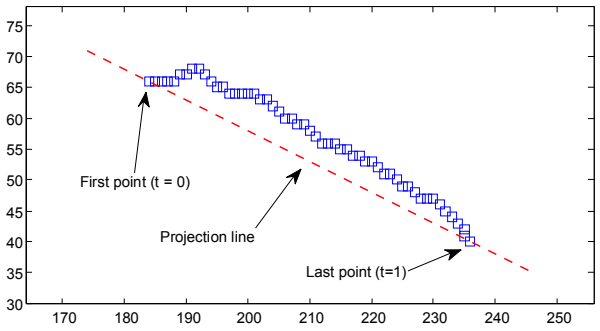

Fig. 3: An example of sorting the skeleton pixels by projection.

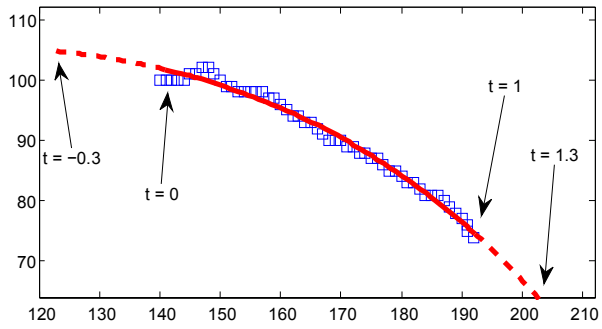

Fig. 4: The result of fitting a second order least squares model into the fiber segment of Figure 3. The dashed lines illustrate how the fiber would be extrapolated in the catenation step.

model of the M^{th} order is given by $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{e}$ with

$$\mathbf{x} = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_N & y_N \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^M \\ 1 & t_2 & t_2^2 & \cdots & t_2^M \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_N & t_N^2 & \cdots & t_N^M \end{pmatrix}$$

and

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_{01} & \theta_{02} \\ \theta_{11} & \theta_{12} \\ \vdots & \vdots \\ \theta_{M1} & \theta_{M2} \end{pmatrix}.$$

The vector $\mathbf{e} \in \mathbf{R}^N$ is the residual noise term that the model does not explain.

In order to fit the model into the fiber segment data, the discrete parameter values $t_0, t_1, \ldots, t_{N-1}$ have to be defined for each data point. A natural choice is to associate the endpoints into the values $t_1 = 0$ and $t_N = 1$, but also the rest of the points require an association. Literature proposes to treat t as another unknown parameter and estimate the values from the data [2], or to traverse the chain of pixels and increase the value according to the distance to the previous point [1]. The latter approach is more straightforward and more suitable for our problem. However, as we later consider joining non-connected fibers, the order of traversing through all points may not be obvious in all cases. Alternatively, we can exploit the simplicity of the structure of the fibers by projecting the pixels onto the line passing through its endpoints as Figure 3 illustrates. After mapping the points into the one-dimensional subspace, a suitable value $t_k \in [0, 1]$ for each $k = 1, 2, \ldots, N$ can be determined according to the linear distance from the two endpoints.

The solution of the model parameter matrix $\boldsymbol{\theta}$ can be found by the standard least squares (LS) approach (see e.g., [12]): $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$, from which we get the actual model for the fiber segment under consideration: $\tilde{\mathbf{x}} = \mathbf{H}\hat{\boldsymbol{\theta}}$. An example of a second order LS model of a fiber is shown in Figure 4. Note that it is straightforward to extend the LS framework into a weighted version, which can take, for example, the curvature at ends of the fiber more efficiently into account.

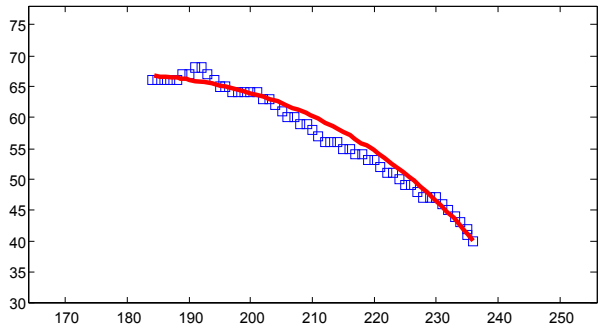

Fig. 5: The result of fitting a second order least squares model with weights $w_k = (1 - \min(t_k, 1 - t_k))^{10}$ into the fiber segment of Figure 3.

In some cases, the fiber endpoints tend to have a higher curvature than the middle area. This can be compensated by assigning a higher weight to the endpoints using weighted least squares: $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{W}\mathbf{x}$. As an example, the weight can be assigned by the rule: $w_k = 1 - \min(t_k, 1 - t_k)$, for $k = 1, 2, \ldots, N$ which is one minus the distance to the nearest endpoint. The balance between endpoints and the center can be adjusted by raising the weights into higher power. An example of the second order fit with no weights is shown in Figure 4, and the corresponding fit with weights $w_k = (1 - \min(t_k, 1 - t_k))^{10}$ is shown in Figure 5. The latter figure clearly gives more emphasis to the curvature near the endpoints.

Prior to fitting the model to the fiber segments, the original skeleton mesh was cut at each crossing point. In addition, the original segmentation produces components that have clearly been cut incorrectly into parts. To correct for these errors, fibers are connected by rejoining parts based on the polynomial model. This procedure is the following.

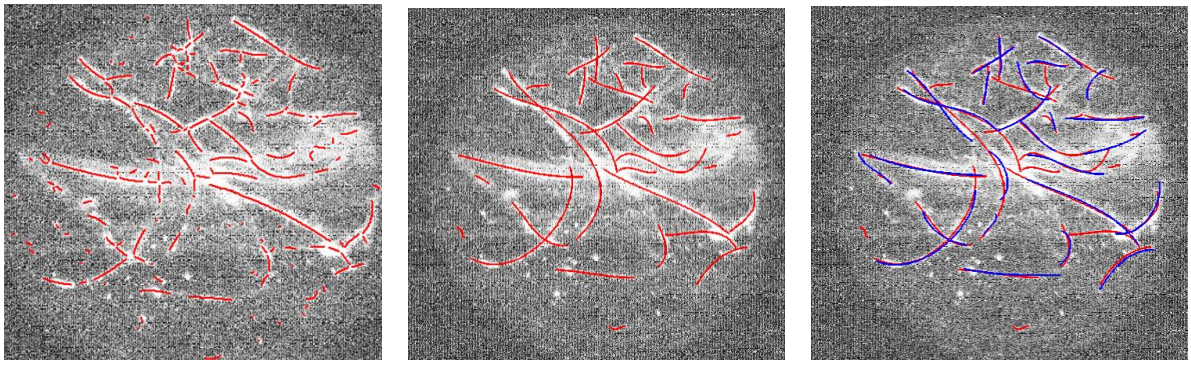

Fig. 6: Model fitting. (a): Fitting the model to pruned skeleton, (b): final result after combining separate parts of the skeleton mesh. (c): Result of manual (blue) segmentation overlaid on top of figure (b).

1. Select an end of a non-catenated fiber, which is closer to an end of another fiber than a threshold T_f. If no matching fiber ends are found, stop the algorithm. If the selected fiber end is close to many other ends, proceed with all the fibers and select the best match for catenation in step 6.
2. Extrapolate both fibers towards each other. Note that the fiber model is defined with $t \in [0, 1]$, and the extrapolation is done with values $t < 0$ or $t > 1$, see Figure 4.
3. Find a location where the extended fibers are closest to the original endpoints of the original non-extended fibers. Note that there is an analytical solution for the minimal distance.
4. At the location found in the previous step: Calculate tangent angles of both of the fibers from the polynomial model.
5. Fit the model to the complete fiber (pixels of both fibers). Calculate the fitting residual $\|\mathbf{e}\|$.
6. Catenate the fibers if the error is less than a threshold T_e, and the difference of the tangents is less than a threshold T_t.
7. Return to step one.

This procedure essentially joins fiber segments, whose endpoints are close to each other and whose orientations near the endpoints match. One run on the algorithm only catenates fibers that have not been joined before, so if the input images are very noisy with several gaps in many fibers, the procedure can be repeated. The number of repetitions depends on the application; we repeated the catenation twice for all images. An example of the catenation process is shown in Figure 6.

4. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed method, we manually annotated all fibers from ten test images similar to the one in Fig. 2 (a). In all, the images contain 211 annotated fibers. The proposed method finds altogether 319 fibers from the same images, and the excess ones are mostly shorter than any of those annotated manually. The result for one of the images is shown in Fig. 6 (c). The figure shows that especially longer segments are correctly recognized, and that the differences are mainly with the short ones. However, due to the nature of the measurement, manual annotation is also very error-prone and requires subjective interpretation.

Despite the difficulties with the ground truth, the important thing for the application is not the correct segmentation of individual fibers, but instead the length and curvature distribution over a larger set of images. These are shown for the ten test images in the histograms of Figure 7. The length of a fiber can now be calculated analytically from the polynomial model $\mathbf{x}(t) = (x(t), y(t))$ as $L(\mathbf{x}) = \int_0^1 \sqrt{(x'(t))^2 + (y'(t))^2}\, dt$. The analytical representation allows also a precise calculation of the curvature, but for the sake of comparison, we define it in the conventional manner in paper industry, $\kappa = L(\mathbf{x})/\|\mathbf{x}(0) - \mathbf{x}(1)\|$, i.e., as the ratio of the curve length and the distance of its endpoints.

Instead of a thorough one-to-one comparison of fibers we concentrate only on the statistical properties of the results of the manual and automatic method. This is because an accurate measure of sensitivity and specificity would require finding a match between found fibers in the two results. Since the matching is non-trivial and error-prone, we skip this analysis. Moreover, the manufacturer is not interested in individual fibers, but rather on the statistics on a longer period.

The histograms of Fig. 7 are very similar to each other both in terms of both distributions. The most significant difference is in the length histogram, where the manual annotation has significantly less fibers than the automatic one. However, this is probably partly due to human perception, which tends to discard too small objects as noise.

In the length histogram, also a dashed vertical line is shown. This is the estimate of mean fiber length obtained by

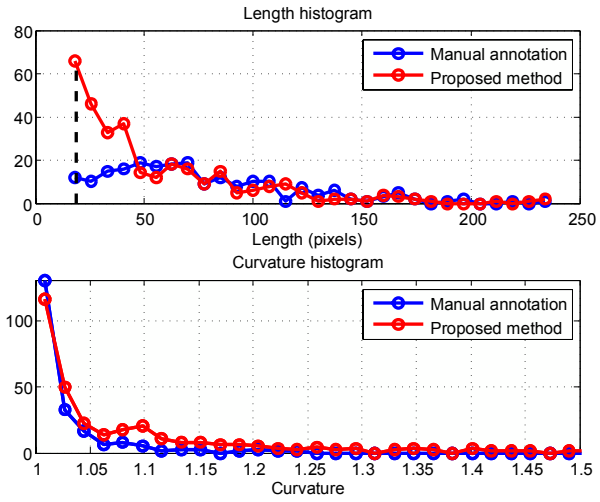

Fig. 7: Histograms of length (top) and curvature (bottom) distributions. The dashed line in the upper plot is the result of the traditional method [9] for the average fiber length.

a more traditional method [9], which estimates fiber lengths as twice the ratio of the number of all foreground pixels to the number of fiber endpoints (from the skeleton; 2 (c)). It is clearly seen, that the standard method underestimates the mean length of the fibers. This is mostly due to a large number of noisy short fiber segments in the skeleton, which our method joins together. Moreover, the proposed method also enables the estimation of fiber curvatures, which is of interest and not possible to assess using the traditional method.

The average processing time on a modern PC is approximately 7 seconds per image. Although this is not strictly real time and many frames are skipped, the method is still fast enough to capture the essential changes in the statistics of the process.

5. CONCLUSIONS

In this paper, we presented a method for pulp fiber detection and measurements based on images taken *in-line* during pulp manufacturing. The fibers overlap, and input images suffer from heavy imaging noise, making simple thresholding based analysis infeasible. By applying a model based approach after preprocessing, we enabled reliable fiber detection and curvature measurements. The mode represents each fiber as a parametric polynomial, which enables their analytical treatment. The analytical properties are used for deciding which fiber segments should be joined based on extrapolating the fitted curves, and studying their tangent vectors.

Despite the difficulty of comparing the result with the ground truth, we have shown that the statistics of the result are close to those of manual annotation. The key differences are in the length distribution, where manual annotation has fewer short fibers. However, this is largely due to inaccuracy of the human segmentation with small fibers, which may be indistinguishable from noise to human observer.

6. REFERENCES

[1] F.S. Cohen and J.-Y. Wang, "Part I: Modeling image curves using invariant 3-D object curve models-a path to 3-D recognition and shape estimation from image contours," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 1, pp. 1–12, Jan 1994.

[2] F.S. Cohen, Z. Huang, and Z. Yang, "Invariant matching and identification of curves using B-splines curve representation," *IEEE Trans. Image Process.*, vol. 4, no. 1, pp. 1–10, Jan 1995.

[3] K. Robb, O. Wirjadi, and K. Schladitz, "Fiber orientation estimation from 3D image data: Practical algorithms, visualization, and interpretation," *Int. Conf. Hybrid Intell. Syst.*, 2007.

[4] J. Bian and S. Qiu, "Pulp fibre recognition based on curvelet transform and skeleton tracing algorithm," in *IEEE Int. Conf. Industr. Electr. Appl.*, May 2007, pp. 2534–2538.

[5] S. Svensson and M. Aronsson, "Using distance transform based algorithms for extracting measures of the fiber network in volume images of paper," *IEEE Trans. Syst., Man, Cybern. B*, vol. 33, no. 4, pp. 562–571, Aug. 2003.

[6] E.L.G. Wernersson, A. Brun, and C.L. Luengo Hendriks, "Segmentation of wood fibres in 3D CT images using graph cuts," *Proc. ICIAP 2009*, vol. 5716, 2009.

[7] J. Jääskeläinen, "Method for illuminating particles contained in a medium for optical analysis, and optical particle analyser, U.S. patent, application # 10/365448," Feb 2003.

[8] H. Wang, "Fiber property characterization by image processing," M.S. thesis, Texas Tech University, 2007.

[9] J.C. Russ, *The image processing handbook*, CRC Press, Inc., Boca Raton, FL, 3 edition, 1999.

[10] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62 –66, Jan. 1979.

[11] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222 –1239, Nov 2001.

[12] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, Englewood Cliffs, NJ, 1993.