

COMPARISON OF TWO TECHNIQUES OF PATTERN RECOGNITION IN THE IMAGE ANALYSIS-BASED WHEAT STALK LENGTH CHARACTERIZATION

Jan Thore Benthien^{a*}, Sabrina Heldner^b, Benjamin Seppke^c, Jan Hörbelt^d

^a Thünen Institute of Wood Research, Leuschnerstraße 91, 21031 Hamburg, Germany, e-mail: jan.benthien@thuenen.de

^b Thünen Institute of Wood Research, Leuschnerstraße 91, 21031 Hamburg, Germany, e-mail: sabrina.heldner@thuenen.de

^c Cognitive Systems Laboratory, Department of Informatics, MIN Faculty, University of Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany, e-mail: seppke@informatik.uni-hamburg.de

^d Faculty of Life Sciences, Rhine-Waal University of Applied Sciences, Marie-Curie-Straße 1, 47533 Kleve, Germany, e-mail: janhoerbelt@web.de

* Corresponding author: e-mail: jan.benthien@thuenen.de

ARTICLE INFO

Article history:
Received: May 2020
Received in the revised form:
May 2020
Accepted: June 2020

Key words:
wheat straw,
ploughless tillage,
stalk length determination,
image analysis,
flow lines-tracing

ABSTRACT

The chaff quality or, more specifically, the distribution of stalk length after straw shredding during wheat harvest with a combined harvester is of significant interest for ploughless tillage practices. The currently applied characterization methods (manual length measurement or the cascade sieve analysis) are time-consuming and labour-intensive. Image analysis-based size characterization has the potential to solve these problems. In this study, two techniques of digital image processing, the well-known method of image moments (rectangular model) and a sub-pixel skeletonization approach (flow lines-tracing), were applied comparatively for stalk length measurement. Upon applying the rectangular model, the analyzed stalks were found to be longer than when the flow lines-tracing algorithm was applied. This was attributed to the unbiased decision criteria of the measuring method. A greater length of the stalk is to be expected in the main stalk than when the length of a twisted or branched stalk is measured from the edge length of a rectangular box that encloses it.

Introduction

Wheat straw is a by-product of wheat harvests, used for different applications (e.g. as bedding for animal husbandry) or ploughed back into the ground as fertilizer. After threshing, the dry straw remaining on the field is finely shredded and spread across the working width of the combine harvester. In particular, in the case of ploughless tillage (e.g., no-till or strip-till farming), high demands are placed on the chaff's chop quality or, more specifically, size distribution. Fine-cut chaff can be more easily and extensively ploughed under and it decomposes more quickly. Thus the possibility of accidental sowing or infection with fungi and

pests is reduced (Voßhenrich, 2000). However, the increase of shredding intensity increases energy consumption and thus costs.

In the context of shredding optimization, effects of changes in the straw chopper's construction or in the chopping settings are currently being investigated with manual measurement or by applying cascade sieve analysis. Such changes include the rotation speed of the cutter block, different activation of chopper elements as counter knives or a friction base on the chaff's chop quality. But sieve fractionation is obviously inadequate, as the shape of straw is not included in this technique, which results in poor reproducibility (Albrecht, 2003). Furthermore, this procedure is quite time-consuming and labour-intensive, and the sample's moisture content impacts the result (Albrecht, 2003). Also, sieve analysis only provides un-specific size information because length, width or thickness are not directly measured. Instead, particle dimensions are compared with the total sieve mesh size, which always results in one-dimensional size information (except for spherical particles). Manual size measurement is time-consuming and labour-intensive, and, additionally, is neither subjective, nor reproducible.

The image analysis-based size measuring technique has the potential to solve these problems and resolve the information limitations as the characterization of the stalks' size distribution is (a) disengaged from physical sample division for mass related evaluation and (b) directly based on each individual stalk length in the sample. Despite these advantages, image analysis-based size measuring has not yet been established in wheat straw shredding optimization and the further development of the combined harvester.

The initial approaches to applying image analysis for the size characterization of ribbon- or fibre-shaped agricultural products determined the stalk length distribution of grass silage (Rechberger et al., 2012). Due to the different characteristics of blades of grass and wheat stalks, these techniques cannot be applied for straw characterization without adaptation.

Within the joint project *Prüfstand zur Vereinzelung von Häckselgütern mit Bilderkennung und Auswertesoftware auf C++ Basis* (Test bench for the separation of shredded particles with image recognition and evaluation software based on C++), of the University of Applied Sciences in Düsseldorf, Institute for Product Development and Innovation – FMDauto (Düsseldorf, Germany) and DLG Test Center Technology and Farm Inputs (Groß-Umstadt, Germany), a vacuum gripper and a short compressed air blast onto a conveyer belt were used for the image acquisition of separate agricultural clippings from 2014 to 2016. Prior to measuring, the applied analysis algorithm can virtually post-separate stalks that lay cross-wise.

Commercially available measuring devices like the Haver CPA 4 Conveyor (Haver & Boecker OHG, Oelde, Germany) seem suitable in general. However, it lacks an appropriate stalk separation and the analysis is limited to the included image evaluation techniques (not specialised in fibre-shaped particles).

What all known approaches have in common is that their sample throughput is too low.

The wood industry was confronted with the same challenges of size characterization in the production of fibreboard that mechanical engineers are facing now in combined harvester development. Manual and sieve-based size measuring techniques were inappropriate for dimensional fibre material characterization. Commercial image analysis-based techniques could not be applied until image acquisition was available for the woolly matted material. Even later, cross-laying fibres could not be properly measured during image acquisition. For the wood industry, this restriction can be solved by applying a compressed air-based separation unit and virtual software-based post-separation of cross-laying fibres (Benthien et al.,

2014). The imaged fibres are skeletonized at sub-pixel precision based on a flow lines-tracing algorithm, which uses the ridges of the image function (Seppke et al., 2015). Presumably this promising approach may be adaptable for the wheat stalk length characterization, as here fibre- or more likely ribbon-shaped material has to be prepared for image acquisition and measured along the length of the main stalk.

The aim of the present study is to support the development of a suitable measuring system for the image analysis-based wheat straw chaff size distribution by comparatively applying the well-known method of image moments (rectangular model) and the novel principle of flow lines-tracing.

Materials and Methods

A sample of wheat straw was characterized by its size, based on the image analysis. A transmitted light scanner (Epson Perfection V750 Pro, Epson America Inc., Long Beach, California, USA) was used for image acquisition. The FibreAnalyser, as a part of the FibreBench working environment (Hamburg University, Department of Informatics, Scene analysis and Visualization, Hamburg, Germany, Version 1.7, Revision 1901), was applied for image analysis. Images (16-bit grey scale) were captured with a resolution of 300 dpi and image size of 254 mm x 203.2 mm, which yielded a pixel edge length of 84.7 μm . Sample preparation, including the separation of the stalks entangled with each other, was done manually. For length measurement the two system's options of pattern recognition, the flow line tracing algorithm and the image moment method (rectangular model), were run respectively. Applying the flow lines-tracing algorithm (described in detail by Seppke, et al., in 2015), only the longest fibre found within each region of interest (ROI) was transferred to the raw data table. Since fibres identified as overlapping were considered branches of the main stalk, they were thus excluded from data evaluation. When using image moment method for length measurement, the projection area of the stalks was also measured, by summing up the area of grey scale pixels above the defined threshold value. The length and projection area measurement was done on a 4-gram sample material (number of stalks = 729), split into 30 scans.

Graphical representation of the data was performed in accordance with DIN ISO 9276-1:2004-09, demonstrating the histogram of the length-related distribution density ($q_1(x)$) and the length-related cumulative distribution ($Q_1(x)$). Length-related means that the measured stalks were sorted in accordance to their length into length classes (baskets). The lengths of the stalks in each length class were summed up and normalized to the sum total of the lengths of all stalks. For a histogram plot, the shares of the cumulated lengths within each length class were divided by each length class width. The area of the consecutive bars, thus, amounts to 1 (or 100% if such amount was normalized initially).

In order to keep the results of length measurement numerically comparable, characteristic values were created from the raw data set. The mean stalk length was calculated by dividing the sum of all stalk length divided by its number. Next, the 10% quantile (D10 or x10), the 50% quantile (D50 or x50), and the 90% quantile (D90 or x90) of the length-related distribution were calculated with the corresponding function of Microsoft Excel (Microsoft Corporation, Redmond (Washington), USA). The D50 equates the median of the length-sorted raw data set of stalk length and gives the length of the stalk so that 50% of the cumulated stalk length is reached. Analogically, the same is true for D10 and D90.

As a measure for the width of the distribution, the span (relative width) is calculated with Equation 1.

$$\text{Span} = (D90 - D10) / D50 \quad (1)$$

where:

- D10 – 10% quantile
- D50 – 50% quantile
- D90 – 90% quantile

A more compact graphical representation of the length-related length distribution than offered by histogram and cumulative distribution is obtained with the box-whisker plot. For this, the quantiles (D10, D50, and D90) were applied in addition to the minimum and maximum stalk length. There the minimum and maximum stalk lengths were displayed as whisker, the D10 and D90 value as box width, and the median indicated by streak in the box. No replicate measurements were taken so that no statement on the reproducibility can be made.

As examples, a narrow stalk, in which only one tracing line is detected (Stalk number 20); a broad stalk, in which several tracing lines were detected (Stalk number 7), and a branched stalk, in which the main stem should be distinguished from the branches (Stalk number 42), were represented.

Results and discussion

Measurements taken using the rectangular model supplied a mean stalk length of 31.1 mm (\pm 20.2 mm), and the flow lines-tracing algorithm resulted in a length of 27.7 mm (\pm 16.3 mm). This means that the stalk length determined based on flow lines-tracing was found to be 3.3% shorter on average than when the rectangular model was applied. The maximal undervaluation of the stalk length determined based on flow lines-tracing, in comparison with the rectangular model, amounts to 94% (Stalk number 15, 29.5 mm (rectangular model), 1.7 mm (flow lines-tracing)) the maximal overvaluation (the greater stalk length in the flow lines-tracing model than in the rectangular model) amounts to 262% (Stalk number 37, 24.4 mm (rectangular model), 88.3 mm (flow lines-tracing)). Minimal and maximal stalk lengths as well as the distribution span are given in Table 1.

Theoretically, flow lines-tracing has to yield larger length values than the rectangular model as it includes the true length-inclusive end-to-end stalk curvatures. In contrast, the rectangular model is only able to cover the direct distance between one end to the other. While no differences were found between the stalk lengths during statistical analysis, significant differences between the individual measurement values, both under- and overvaluations, indicate that at least one of the models is not suitable for the size determination of wheat stalks.

The numeric results (D10, D50, D90) of the rectangular model stalk length measurements and the flow lines-tracing stalk length measurements (Table 1) are illustrated in the graphical representation of the stalk length distribution in Figure 1. The figure shows higher amounts of short stalks in the case of fibre tracing and a higher amount of long stalks in the case of the rectangular model.

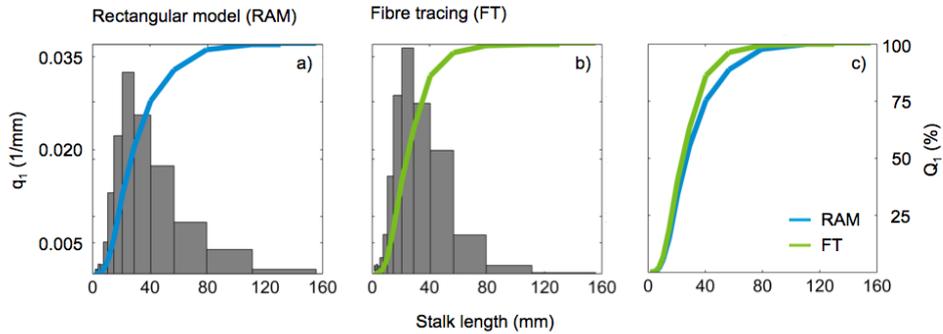


Figure 1. Stalk length distribution

Table 1.
Stalk length at the characteristic values of the length-based cumulative distribution in respect to option of pattern recognition

Characteristic value	Rectangular model		Flow lines-tracing
	Length (mm)	Area (mm ²)	Length (mm)
Minimum	1.4	0.5	1.4
D10	11.5	10.0	10.7
D50	26.3	37.8	24.5
D90	55.8	125.4	48.3
Maximum	154.0	461.6	115.0
Mean value	31.1	58.8	27.7
Span	1.68	-	1.53

Comparing the mean stalk length with the 50% quantile (median), rectangular model and flow lines-tracing each, the presence of skewed distributions (mean \neq median) has to be assumed. This can be seen by looking at the result visualization in Figure 2 (Box-Whisker-Plots). A skewed distribution must also to be assumed for the projection area (sum of the pixel area with a grey scale above defined threshold value), as the stalk's mean projection area ($58.8 \text{ mm}^2 \pm 64.7 \text{ mm}^2$) is different from the 50% quantile (37.8 mm^2). In addition, and for that reason alone, the average-exceeding standard deviation of the stalk's mean projection area point to a skewed distribution.

It is surprising and hardly explicable that the mean values from flow lines-tracing and rectangular model do not differ from each other, while considerable differences were present between the individual stalk lengths. However, these differences can be well worked out by means of the three exemplary stalks in Figure 3. In this figure, the grey scale images of the stalks were merged in the same order of magnitude and further overlaid with the stalk-enclosing rectangle. By providing the start and end-points at the same time, the flow line of the longest tracings path was indicated. The results from length measurement of the three exemplary stalks are presented in Table 2.

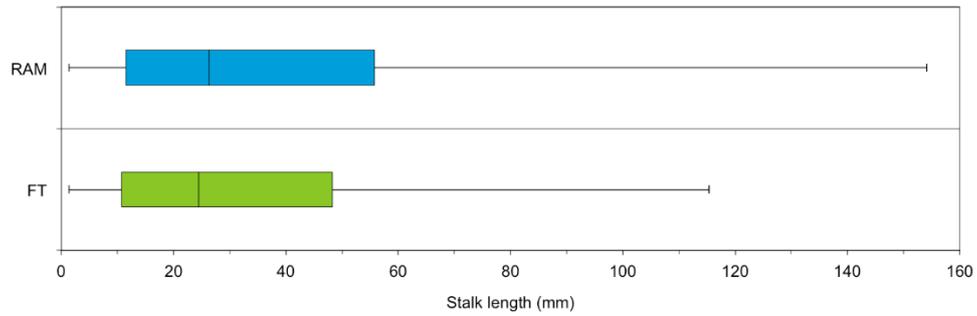


Figure 2. Box-whisker plot of the length-related stalk length distribution with the minimal and maximal value as a whisker edge, D10 and D90 as box edge, and D50 as mark in the box' centre

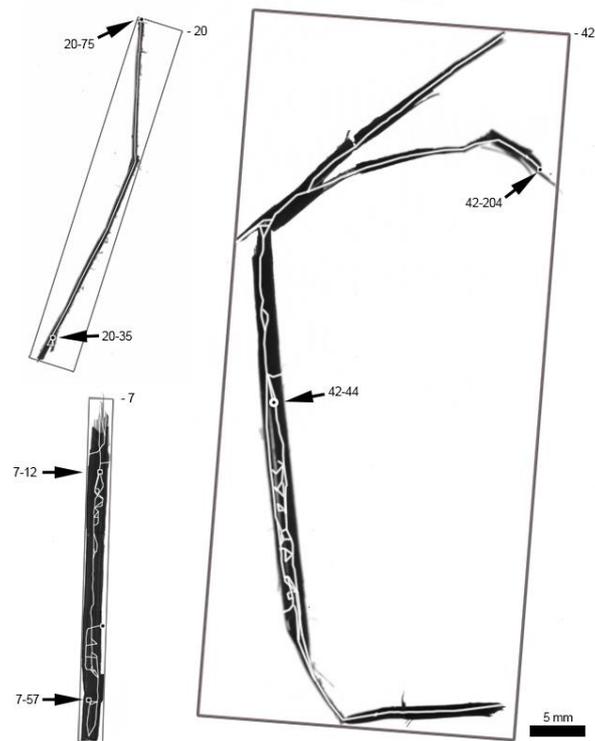


Figure 3. Exemplary stalks (7, 20, 42), overlaid with the stalk-enclosing rectangle and the start and end points of the flow lines indicated

Table 2.

Results from length measurement of three exemplary selected stalks

Stalk number	Rectangular model	Flow lines-tracing		
	Length (mm)	Start point	End point	Length (mm)
7	57.7	7-12	7-57	48.0
20	30.8	20-35	20-75	31.1
42	63.5	42-44	42-204	37.8

For most of the imaged stalks more than just one flow line was found. This can be explained by the much larger width of wheat stalks in comparison to that of wood fibres (factor of 10^3) for which the algorithm was initially designed. Variations in stalk's grey scale were interpreted as overlapping fibres so that they could be post-separated to single fibres with one flow line on each using computer software. Thus, in few cases only the start and end-point of the flow line will match the fibre's tail and reveal the true stalk length. However, the rectangular model will rarely reveal the true stalk length (as seen in the case of Stalk number 7). More often the curvature of the stalk will let the rectangular model undervalue the stalk's length. An extreme overvaluation of the flow line tracing takes place if a curvaceous path is erroneously recognised in a wide stalk.

Conclusions

In the comparative image analysis-based determination of the stalk length, the rectangular model and the flow lines-tracing algorithm were both found to be inadequate for the measurement of wheat straw size distribution. While the flow lines-tracing algorithm is optimized for the analysis of wood fibres and thus software can separate the wide wheat stalks into several single fibres, the rectangular model is obviously unable to determine the true length of the curved stalks. Although the rectangular model is appealing, due to its simplicity and transparency, it will always undervalue the curved stalks. The approach of flow line tracing is promising in general, but has to be adapted for wide stalks with a huge amount of grey scale variations in the stalk's width. A potential advantage of the flow line tracing may be the possibility to recognise and count the branches of the main stalk. The acquisition of the stalk's projection area may in perspective help to calculate the shape factors, e.g. the curling index.

A promising alternative to flow line tracing may be the method of skeletonization. In this approach, the imaged stalk is reduced to its inner skeleton-line. However, this method is unable to post-separate the overlapping stalks so that an intensive mechanical separation prior to image acquisition is required.

The results of this study demonstrate on the one hand that image analysis-based size characterization is a promising technique to measure the size distribution based on the properties of single elements (here: the stalks). On the other hand, it shows, however, that not only the hardware for image acquisition has to be adapted for the individual material characterized, but rather the analysing software as well. For certain materials specialised algorithms cannot be used without any adaptations, despite the promising principals.

Acknowledgements

The authors like to thank Mr. Nick Engehausen (Thünen Institute of Wood Research) for the conscientious execution of sample preparation and image acquisition.

References

- Albrecht, C. (2003). *Konstruktion einer Vorrichtung zur Beschickung von Siebmaschinen (Construction of a device for feeding screening machines)*. Institute of Agricultural Engineering, University of Hohenheim. Stuttgart. Germany.
- Benthien, J.T., Bähnisch, C., Heldner, S., Ohlmeyer, M. (2014). Effect of fiber size distribution on medium-density fiberboard properties caused by varied steaming time and temperature of defibration process. *Wood and Fiber Science* 46(2), 175-185.
- DIN ISO 9276-1:2004-09, Darstellung der Ergebnisse von Partikelgrößenanalysen - Teil 1: Grafische Darstellung (ISO 9276-1:1998) (Representation of results of particle size analysis - Part 1: Graphical representation (ISO 9276-1:1998))
- Rechberger, C., Prankl, J., Oberberger, G., Handler, F., Prankl, H. (2012). "Automatisierte Bestimmung der Partikellängenverteilung von Grassilagen anhand bildanalytischer Methoden" (Automated particle length distribution analysis of processed forage by means of image analysis), in Proceedings of the 70. Internationale Tagung Landtechnik (70th International Conference Agricultural Engineering), 6.-7. November 2012, Karlsruhe, Germany.
- Seppke, B., Bähnisch, C., Benthien, J.T., Heldner, S., Ohlmeyer, M. (2015). "A Concurrent Skeleton-based Approach for the Characterization of Wood Fibers with Sub-pixel Precision for Fiber Board Production", in Proceedings of the 10th International Conference on Mass Data Analysis of Images and Signals (MDA), 11.-14. July 2015, Hamburg, Germany.
- Voßhenrich, H.H. (2000). Beim Häckseln liegt noch vieles im Argen (Still much trouble when coping). *Top Agrar* 7/2000, 68-70.

PORÓWNANIE DWÓCH TECHNIK ROZPOZNAWANIA W CHARAKTERYSTYCE OBRAZOWEJ DŁUGOŚCI ŁODYG PSZENICY

Streszczenie. Jakość plew, a dokładniej rozkład długości łodyg po rozdrobnieniu słomy podczas zbioru pszenicy za pomocą kombajnu zbożowego, ma istotne znaczenie dla uprawy bezorkowej. Stosowane obecnie metody charakterystyki (ręczny pomiar długości lub kaskadowa analiza sitowa) są czasochłonne i pracochłonne. Charakterystyka wielkości w oparciu o analizę obrazu może potencjalnie rozwiązać ten problem. W pracy porównawczej do pomiaru długości łodyg zastosowano dwie techniki cyfrowego przetwarzania obrazu, znaną metodę momentów (model prostokątny) oraz szkieletyzację na poziomie subpikseli (flow lines-tracing). Po zastosowaniu modelu prostokątnego zaobserwowano, że odwzorowane łodygi są dłuższe niż przy zastosowaniu algorytmu flow lines-tracing. Przypisano to niedostosowaniu kryteriów metody do tego zadania pomiarowego. Biorąc pod uwagę łodygę główną, należy spodziewać się większej długości niż w przypadku, gdy pomiar długości skrzyżowanej lub rozgałęzionej łodygi wynika z długości krawędzi prostokątnych ram, które ją otaczają.

Słowa kluczowe: słoma pszenna, uprawa bezorkowa, określanie długości łodygi, analiza obrazowa, flow lines-tracing