agronomy

MDPI

Article

Morphometry of the Wheat Spike by Analyzing 2D Images

Mikhail A. Genaev [1,2,*], Evgenii G. Komyshev [1], Nikolai V. Smirnov [2], Yuliya V. Kruchinina [1], Nikolay P. Goncharov [1] and Dmitry A. Afonnikov [1,2]

[1] Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, Novosibirsk 630090, Russia
[2] Novosibirsk State University, Novosibirsk 630090, Russia
* Correspondence: mag@bionet.nsc.ru

Received: 27 May 2019; Accepted: 12 July 2019; Published: 17 July 2019

Abstract: Spike shape and morphometric characteristics are among the key characteristics of cultivated cereals associated with their productivity. Identification of the genes controlling these traits requires morphometric data at harvesting and analysis of numerous plants, which could be automatically done using technologies of digital image analysis. A method for wheat spike morphometry utilizing 2D image analysis is proposed. Digital images are acquired in two variants: a spike on a table (one projection) or fixed with a clip (four projections). The method identifies spike and awns in the image and estimates their quantitative characteristics (area in image, length, width, circularity, etc.). Section model, quadrilaterals, and radial model are proposed for describing spike shape. Parameters of these models are used to predict spike shape type (spelt, normal, or compact) by machine learning. The mean error in spike density prediction for the images in one projection is 4.61 (~18%) versus 3.33 (~13%) for the parameters obtained using four projections.

Keywords: image analysis; machine learning; algorithms; computer vision

1. Introduction

Spike shape and morphometric characteristics are among the key characteristics of cultivated cereals associated with the important agronomic traits, such as yield, non-brittle spike, and easy threshing. Biologists and breeders are interested in the spike characteristics, such as length, number of spikelets per spike, number of kernels per spike, size and shape of kernels, and their weight per spike as well as spike shape and density, presence or lack of awns, and spike and awn color [1]. Spikes differ in their shape, size, density, awnedness, color, and so on. The spike shape is controlled by a set of genes; their study will make it possible to purposefully create new cultivars with improved yield characteristics, easier threshing, and resistance to environmental factors [2]. The wheat inflorescence is a compound spike with a length of 5–17 cm and longer; it consists of a rachis and spikelets. The rachis consists of segments each bearing with spikelets. In turn, each spikelet comprises two–five florets and only one to three of them develop into grains.

An important goal of the breeding and genetic experiments is a rapid and accurate assessment of the target plant parameters, referred to as phenotyping [3]. The spike length and density as well as the number of spikelets and kernels per spike, weight of 1000 grains, and several other characteristics are of a great relevance for breeders and geneticists [4,5]. The kernel shape is also a useful breeding trait since it determines the flour yield along with the grain size and uniformity, thereby contributing to grain commercial value [6]. The characteristics of spike morphology also form the background of genus *Triticum* L. [3,7] and related species [8] taxonomies. Currently, the spike characteristics in most studies are assessed by an expert via a visual analysis and measurements, which is rather time-consuming;

the more so as the modern experiments involve tens of thousands of plants [5,9]. Correspondingly, automation of this laborious and time-consuming process is relevant for the science and breeding. The efficiency of plant phenotyping can be increased by technologies of digital image analysis [10–12]. These technologies were applied for both kernel size and shape morphometry [13–16] and analysis of the spike traits [17–20].

The methods for digital image analysis of spike characteristics are also developed and allow for solving of different problems. Grillo et al. [17] developed a method for the wheat variety identification using glumes size, shape, color, and texture characteristics obtained from image analysis. Makanza et al. [18] designed the software allowing for determination of ear length and width as well as estimation of maize grain size and weight. Pound et al. [19] used deep learning to count wheat spikes and assess the number of spikelets per spike using images of wheat plants taken in glasshouse conditions. However, Pound et al. in their work did not estimate the morphological characteristics of spikes, such as their length, width, and type. As for the deep learning algorithms, their use, in turn, requires a large number of annotated spike images (several tens of thousands) to train the neural network parameters. Hughes et al. [20] determined wheat spike and grain morphometric parameters from X-ray micro computed tomography data. This method is highly accurate when determining the fine characteristics of spike and grain shapes but requires a special device for recording tomographic images. Kun et al. [21] proposed morphometry of wheat spikes via image processing. The authors utilized 2D images to assess various characteristics, such as spike length and awn number and length, and classified the spike shape type according to its length-to-width ratio. The spike parameters were used for their classification according to cultivars with the help of back propagation neural network. An advantage of the last approach is in that a rather simple imaging protocol applicable to common use is employed in spike morphometry. However, the software for image processing is unavailable. In addition, the description of wheat spike shape based on the length-to-width ratio is a significant simplification. Thus, a method for wheat spike morphometry that would utilize a commonly available imaging protocol to assess a wide range of the spike shape and size characteristics with the help of an available software is still of high relevance.

Here, we propose a method for determining the quantitative characteristics of spike shape based on analysis of its 2D images. This method utilizes digital processing of the images of the spikes detached from plants. The images are recorded under laboratory conditions without any special requirements to photographic equipment and light conditions. Analysis of the detached spikes allows for digitizing of the already existing spike specimens. This method is based on outlining the spike contour and further analysis of the spike size and shape. This method makes it possible to extract several traits associated with spike shape and its awns. The proposed approach has shown high performance in identifying the image regions pertaining to the spike and its awns. Several regression models are proposed for describing the shape of wheat spike. These techniques do not require any large training samples for parameter selection (unlike the deep learning methods [19]). Parameters of these models have been used to predict the spike characteristics, such as the index of spike density and its shapes. The proposed method is available at http://wheatdb.org/werecognizer.

2. Materials and Methods

2.1. Imaging

The imaging protocols for further analysis of spikes were proposed earlier [22]. See Supplementary file (Section 1) for a detailed description. The spike is captured on a blue background ('table' protocol) or is vertically fixed with a clip holder ('clip' protocol). The holder allows spikes to be fixed at different angles relative to the spike axis. We used this to obtain four projections of a spike, which further allowed us to improve the accuracy in assessing the spike type and density (see Results). Each image contains a ColorChecker Mini Classic target (https://xritephoto.com/camera) for color correction. This correction allows for avoiding color shifts in the images, which result from differences in the

lighting conditions [23]. Another advantage in using the color scale is its standard size (68 × 108 mm), allowing for assessment of image scale. Figure 1 shows examples of spike images.

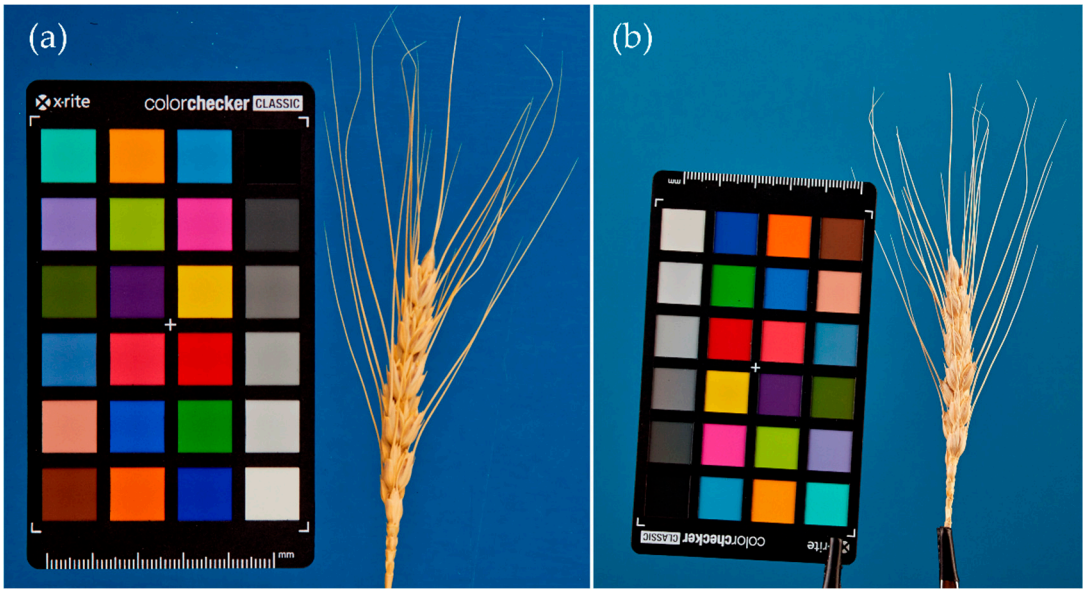

Figure 1. Images of spikes captured by two different protocols: (**a**) On a table and (**b**) fixed with a clip holder.

2.2. *Identifying Spike and Awns in Images*

2.2.1. Image Preprocessing

Images were processed using the OpenCV (Open Source Computer Vision Library; https://opencv.org) software package [24]. At the first stage, we identified the color scale for the image (Figure 1, left). The color scale may be tilted relative to the image vertical axis and its plane is not always perpendicular to the optical axis of the lens in the case of 'clip' protocol. The scale was identified using its reference image to obtain several descriptors by searching for key points. The descriptor in the spike image closest in the Hamming distance was determined for each descriptor of the reference image. The regions corresponding to different colors of the palette were identified by aligning a calibrated image of the scale with the reference one using a RANSAC algorithm [24].

Having identified the color scale, we correct the image color with the help of a method used in epiluminescence microscopy [25]. The shift in colors in spike image relative to the reference was assessed using D_{col} parameter (see Supplementary file, Section 2): The higher its value, the more significant is the color distortion. If D_{col} is close to zero, the image contains almost no color distortion. For details of the algorithm of color scale identification and correction, see Supplementary file, Section 2.

The image scale (pixel size, mm) was calculated from the ratio of the known color scale area to its area in the image (taking into account the correction for its orientation). Further, the ColorChecker area was excluded from analysis.

At the next stage of image preprocessing, we blurred the image using a Gaussian filter with a kernel of 3 × 3 to reduce noise and removed the fragments of clip holder (for 'clip' protocol). An enlarged fragment of the image showing the result of the smoothing algorithm operation is given in Supplementary Materials (Figure S3, Supplementary file).

2.2.2. Image Binarization

The image was binarized after its conversion to a HSV (Hue, Saturation, Value) color space, which separates the color information from luminosity and more stably characterizes color at different illumination levels. As has been earlier demonstrated, the HSV color space best fits the recognition

of objects on a uniform background [26], and in particular, for detecting the plant contours [27]. The intervals of HSV channels used for binarizing the image into spike and background regions were selected using a training sample (see below Section 2.2.4). The resulting image was segmented into the pixels of background and the remaining pixels belonging to the regions of spike and of awns. After eliminating the ColorChecker region and segmentation according to the spike color, the contour that was largest in its area corresponded to spike and awns and was selected in the image. The contours smaller in area, which corresponded to rubbish and glumes (Figure 2d), were discarded from further analysis.

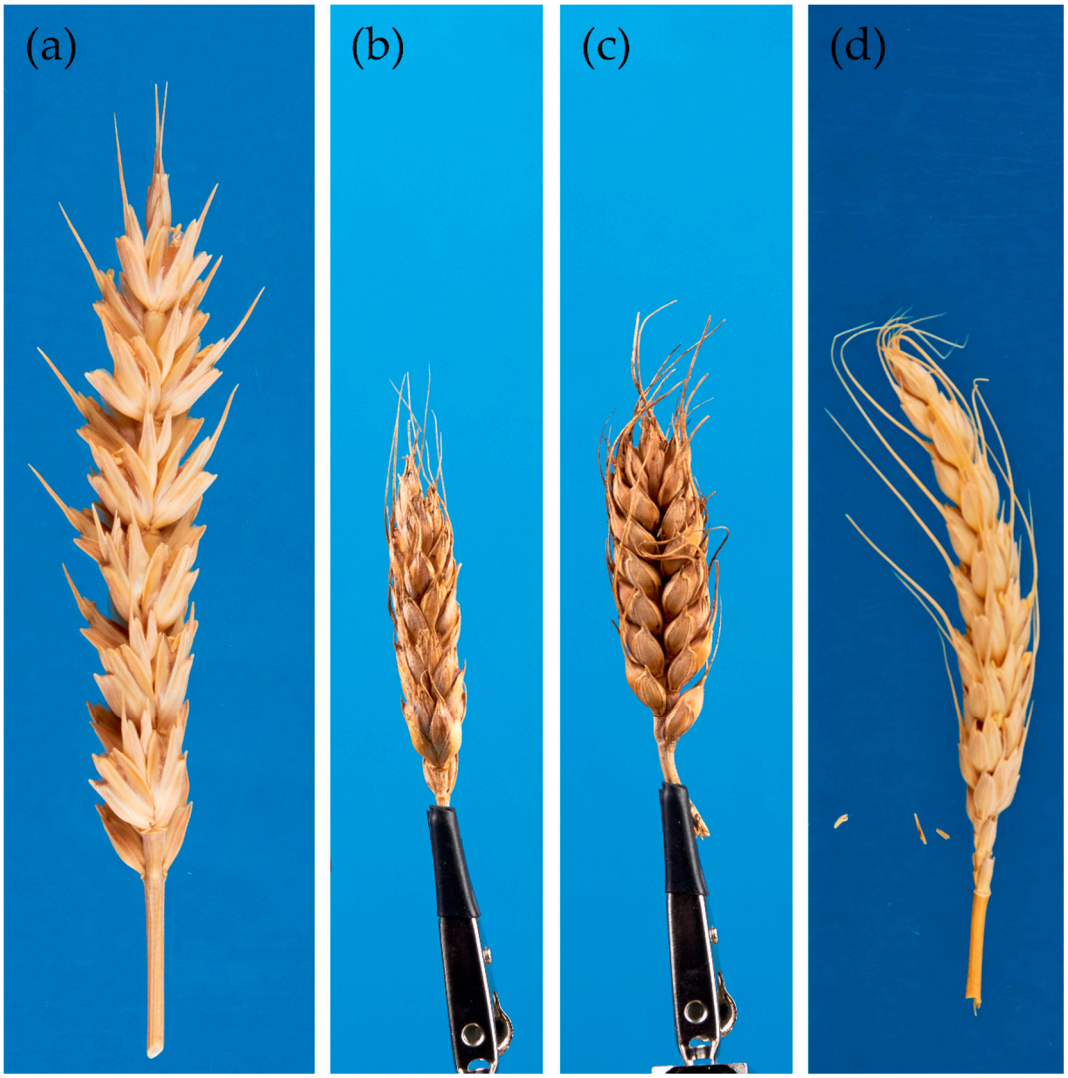

Figure 2. Awnedness types of spikes: (**a**) awnless, (**b**) awnletted, (**c**) half-awned, and (**d**) and short-awned.

2.2.3. Awn Identification

The awns are colored similar to the spike. In some images, awns intensively intersect and even stick together forming a bundle, which significantly hinders identification of individual awns and even makes it impossible in some cases. Correspondingly, a two-stage algorithm was used.

At the first stage, the pixels of the awn skeleton are identified. Since the awns are much thinner as compared with the spike region (body), a partial skeletonization algorithm was used for identifying the awn skeleton. The pixels at the spike–background boundary were iteratively erased. The threshold of erasure iterations, n_{Ei}, was used; this threshold was selected using a training sample of images (see below). The erasure process was stopped in any spike/awn region when its thickness reduced to

a single pixel. After n_{Ei} erasure iterations were completed, the spike/awn regions with a unit thickness were assigned to the awn skeleton. All pixels of the spike/awn region were then classified relatively to the spike body and awns. For this purpose, each pixel of the boundary was provided with a linked list of the pixels, which included the pixels removed at the subsequent iterations and were located in the direction perpendicular to the erased spike–background boundary. In the course of contour refining, the lists for the adjacent boundary pixels were pooled when a removed pixel at a certain iteration appeared to be common for two of them. The lists of the pixels that contained a pixel belonging to the awn skeleton were assigned to the awn region. The remaining pixels were regarded as belonging to the spike body.

2.2.4. Selecting Parameters for Identification of Spike and Awn Regions in Image

Operation of the algorithm for identification of spike and awn regions depends on the following parameters:

(1) The target values of HSV channels and the ranges of their acceptable deviation for image binarization into the background and spike/awn regions and

(2) The number of iterations (n_{Ei}) to the stoppage of skeletonization algorithm.

To select these parameters, we used a sample of 93 spike images of F_2 hybrids between the near-isogenic line of Australian common wheat cultivar Triple Dirk (Triple Dirk B) and the Chinese wheat *Triticum yunnanense,* King ex S.L. Chen (syn. *T. spelta* ssp. *yunnanense* (King ex S.L. Chen) N.P. Gontsch.), accession KU 506 acquired using both the 'table' and 'clip' protocols. Each spike was classified according to the types of awnedness based on an expert estimation into awnless, awnletted, half-awned, and short-awned (Figure 2a–d, respectively) [2]. Table 1 lists the distribution of spike images in this sample according to the types of protocol and awnedness pattern. The total awn and spike body pixels were manually marked for each of these images. The images were randomly divided into the test (30 images) and training (63 images) samples so that the ratios of awn types in spikes in both samples were approximately equal.

Table 1. Distribution of the spikes in sample according to awnedness types.

Awnedness Type	Number of Images	'Clip' Protocol	'Table' Protocol
Awnless	16	10	6
Awnletted	4	4	0
Half-awned	14	14	0
Short-awned	59	36	23

The Jaccard index, J [28], also known as Intersection over Union (IoU) [29], was used for assessing performance of the algorithm for image segmentation into the background and spike:

$$J(A,\ B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (1)$$

Here, A denote the pixels of the image region generated by segmentation using the designed algorithm and specified values of its parameters and B, the manually marked pixels of the image region. We calculated the Jaccard indices: J_e is the binarization accuracy for the whole spike with awns; J_b, recognition accuracy for the pixels of spike body (the spike region minus the pixels of awns); and J_a, the accuracy of awn recognition. While selecting the parameters, we optimized the mean J_e for test sample and after the optimization, independently estimated J_b and J_a using the test sample. Genetic algorithm [30] was used for optimization. The blocks of parameters (individuals) included sets of seven target HSV colors and their ranges (dH, dS, and dV). The blocks could exchange (crossing over) target colors with linked ranges. The population size varied from 20 to 100 individuals.

2.2.5. Estimating the Effect of the Degree of JPEG Compression on Segmentation Accuracy

In this work, we used the JPEG format for analyzing images. This format is a flexible and efficient technique for compressing digital images and makes it possible to achieve a high degree of compression with a insignificant loss in image quality perception for an unaided human eye [31]. Nonetheless, a JPEG image compression leads to a certain loss of information. The loss is the higher, the higher the degree of compression. The loss of information may have a negative effect on the further analysis, in particular, a high degree of compression leads to a considerable loss in accuracy in morphometry of biological objects [32]. The loss of information when forming a JPEG file is specified by quality factor (*QF*), varying in the range of 1–100: The higher the quality, the smaller is the information loss, and correspondingly, the degree of compression [33].

All images used in this work were captured as RAW files with a Canon 600D digital SLR and converted into JPEG using Capture One v.7 with the following parameters: Quality = 90 and ICC Profile = sRGB IEC61966-2.1. We have estimated the effect of the JPEG file quality specified during compression of the initial image on the accuracy of image segmentation. We have analyzed the spike images of two plants of the F_2 hybrids between the Triple Dirk B and the Chinese wheat *T. yunnanense*, one awned and one awnless (Figure S4a,c; Supplement file). For this purpose, we got a TIFF file (without compression) and a series of JPEG files with a quality varying from 1 to 100 (11 variants for each) for each of the analyzed images.

For each of the obtained images, the spike and awns were automatically segmented into ColorChecker regions using the above-described technique. The Jaccard index, *J*, in the comparison of segmentation results for JPEG (with compression) and TIFF (without compression) images was independently estimated. Additionally, *J* was estimated when jointly segmenting the spike and awn regions. This allowed us to assess the effect of image quality on segmentation accuracy.

2.2.6. Awn Quantitative Characteristics

Once the awn and spike regions are selected, the characteristics of spike awnedness are calculated, including the total awn area, S_a (mm^2), determined as the number of the corresponding pixels multiplied by the area of 1 pixel; the number of awns, N_a, as the number of contacts between spike and awn regions (number of awn bases); total awn length, L_a (mm), as the total length of the pixels that from the skeleton of awn region; and the mean awn length, $l_a = L_a/N_a$ (mm). Since part of the images contained considerably intersected and/or overlapped awns, it was impossible to distinguish individual awns; correspondingly, we did not assess the lengths of individual awns (Figure 2b–d).

2.3. Spike Morphometry

2.3.1. Identifying and Straightening Spike Contour

Broken lines were sometimes formed after awn erasing at the sites where they contacted the spike body. In this case, we smoothed the contour using an algorithm for computing elliptic Fourier descriptors [34]. After the awn regions were removed, descriptors for 70 harmonics were computed to recognize the spike contour and further uses them for determining the points of the smoothed contour.

The spike axis was approximated with a broken line iteratively constructed of the segments of spike body. At the first stage, the center of mass and the main axes of the ellipsoid approximating the contour were determined for the pixels of the contour. The major axis corresponded to the direction of the rachis in the center of mass. The perpendicular to the major axis divides the spike region into two parts. At the second stage, an analogous procedure was applied to each part. As a result, each part was also divided by the corresponding axis into two parts; the next iteration stage was again applied to each part. The procedure was successively performed until the number of segments exceeded 20. The centers of mass of each constructed segment determined the broken line that approximated the spike axial line. At the last stages of iteration, the segment size across the axis in some cases could be larger than along the axis. In this case, an axis co-directional to the major axis of the segment

constructed at the previous stage, but passing the center of mass of the current segment was used as the axis of this segment.

After the spike axial line is determined, the spike contour is straightened so that this axis is vertical with its upper end at the spike tip and its lower end, at its base and the distance of the pixels of the transformed contour from the axial line are equal to the distances of the corresponding pixels of the initial contour. This transformation makes it possible to remove the deformation of spike contour caused by its bending. The size and quantitative characteristics of spike shape are determined in the straightened spike contours.

2.3.2. Integral Characteristics of Spike Shape

The characteristics of spike shape fall into several categories. The first group comprises integral shape characteristics, such as spike length L_e (mm), which is approximated by the length of spike axial line; P_e (mm), the perimeter of spike without awns; S_e (mm^2), the area of spike region; for the square index, SQI = $S_e/(L^2)$ is the ratio of the spike area to its squared length.

The other integral parameters are described below.

Circularity index I reflects the degree to which the shape of the contour is close to a circle; its value varies from 0 to 1, with the unity value for an ideal circle:

$$C = \frac{4\pi \times area}{perimeter^2} \tag{2}$$

The perimeter is longer for the contours with numerous convexities on its surface, while the circularity index acquires lower values. In such cases, it is reasonable to use a roundness index, R, since this value is independent of such irregularities of perimeter:

$$R = \frac{4 \times area}{\pi[Major\ axis]^2} \tag{3}$$

Rugosity index, Rg, is determined as the ratio of the perimeter of the contour to the convex perimeter:

$$Rg = \frac{Ps}{Pc} \tag{4}$$

where Ps is the perimeter of the contour and Pc, its convex perimeter, also known as the least convex hull, i.e., the least convex figure that contains all points of an image.

Solidity index, S, is the ratio of the contour's area to the area of its convex hull:

$$S = \frac{Contour\ Area}{Convex\ Hull\ Area} \tag{5}$$

2.3.3. Model of Sections

The first model for describing the spike shape was a set of sections determined by the perpendiculars to spike's axial line with a step of 1/21 (20 sections + 1) of the spike length. Two distances for the pixels of the contour were determined for each perpendicular (from each side of the axial line). Thus, this model was defined by 40 parameters.

2.3.4. Model of Quadrilaterals

The contour of a spike placed horizontally is representable as two quadrilaterals (Figure 3)—upper and lower ones with a common base. The left and right sets of edges are approximated by two quadrilaterals with one adjacent side, their base, being equal to the sum of the spike axis intervals (spike length L_e). The geometry of the upper quadrilateral is determined by the four following independent parameters (Figure 3):

x_{u1} is the distance from the spike tip to projection B' of top B onto base AD;
x_{u2} is the distance of B' to projection C' of top C onto base AD;
y_{u1} is the distance of top B to its projection B' onto base AD; and
y_{u2} is the distance of top C to its projection C' onto base AD;
Distance x_{u3} from projection C' to spike base D is computable from the spike length as $x_{u3} = L_e - x_{u2} - x_{u1}$.

Analogous parameters x_{b1}, x_{b2}, x_{b3}, y_{b1}, and y_{b2} are determined for the lower (bottom) quadrilateral (Figure 3).

The procedure used for selecting parameters of the quadrilateral that most adequately describes the spike shape consisted of the following. The perpendicular to the spike axis (Figure 3, dashed line) was constructed for each pixel of the spike boundary, i. The height of the perpendicular is y_i. The cross of the perpendicular with an edge of the quadrilateral determines the height y_{qi}. This procedure was performed for all pixels of the boundary to calculate the below value:

$$DSq = \sum_i \left(y_i - y_{qi}\right)^2 \quad (6)$$

which, at a fixed L_e, depends on four parameters, $DSq = DSq(x_{u1}, x_{u2}, y_{u1}$, and $y_{u2})$. We selected parameters x_{u1}, x_{u2}, y_{u1}, and y_{u2} so that DSq was minimized. Levenberg–Marquardt algorithm [35] implemented in the Apache commons-math3 library 3.6.1 (class org.apache.commons.math3.fitting.leastsquares.LevenbergMarquardtOptimizer) was used for DSq minimization. The algorithm converges in dozens of iterations (50 on the average); the time to find the parameters numerically using one Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz processor is about 25–50 ms. The parameters were independently selected for the upper and lower quadrilaterals.

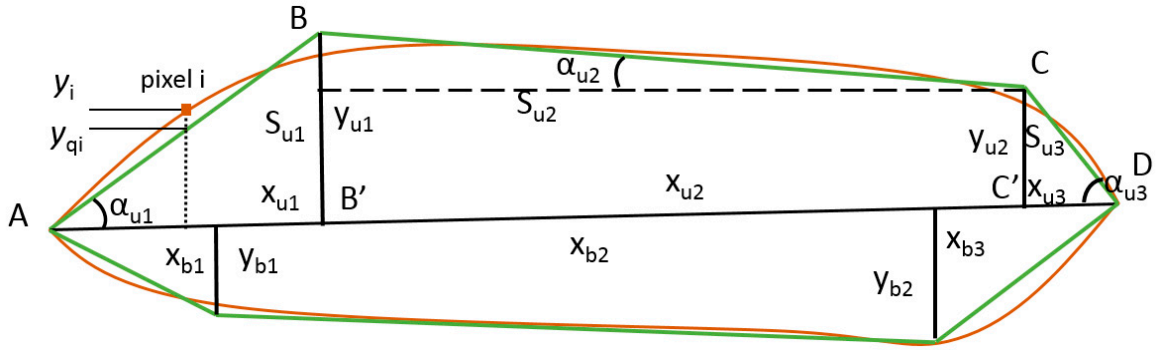

Figure 3. Representation of the spike shape as two quadrilaterals. Black horizontal line shows the spike axial line; brown line, spike contour; and green lines, quadrilaterals approximating spike contour. The main parameters characterizing geometry are shown for the upper quadrilateral. The dashed line is perpendicular to the axis through each pixel i of the contour; the height values for ith pixel, y_i, an edge of quadrilateral, y_{qi}, are shown.

We have also estimated several derived parameters for these two quadrilaterals:

α_{u1} is the inclination of edge AB relative to the base of the upper quadrilateral (degrees);
α_{u2} is the inclination of edge BC relative to the base of the upper quadrilateral (degrees);
α_{u3} is the inclination of edge CD relative to the base of the upper quadrilateral (degrees);
t_{1u} is the tangent of angle α_{u1};
t_{2u} is the tangent of angle α_{u2};
t_{3u} is the tangent of angle α_{u3};
S_{u1} is the area of triangle ABB' (mm^2);
S_{u2} is the area of trapezium $BB'C'C$ (mm^2);

S_{u3} is the area of triangle DCC' (mm^2);
S_u is the area of the upper quadrilateral (mm^2); and
y_{um} is the mean height of the upper quadrilateral (mm).

The parameters for the lower quadrilateral were calculated in an analogous manner.

The following variables were also calculated for both quadrilaterals:

$A_{Ix2} = (x_{u1} - x_{b1})2/x_{u1} + (x_{u2} - x_{b2})2/x_{u2} + (x_{u3} - x_{b2})2/x_{u3}$ is the asymmetry index for the lengths of segments (mm);

$A_{Iy2} = (y_{u1} - y_{b1})2/y_{u1} + (y_{u2} - y_{b2})2/y_{u2}$ is the asymmetry index for the heights of the segments (mm); and

$A_{Ixy2} = A_{Ix2} + A_{Iy2}$ is the total asymmetry index (mm).

2.3.5. Radial Model

Parameters of the radial model are calculated as follows: 360 rays are drawn from the center of mass of the spike contour starting from the direction of its major axis with a step of 1 degree to the point of the contour closest to a given ray. The lengths of the 360 intervals, thus constructed, represent the radial model.

The total list of all spike morphometric characteristics is shown in Supplementary file (Table S1).

2.4. Predicting Spike Density Index and Type of Its Shape

2.4.1. Sample of Spike Images

We assessed the efficiency of our approach by predicting the spike shape type. For this purpose, the spike images of 249 plants, annotated manually, were extracted from the SpikeDroid database [22]. We used digitalized data of 1245 spike images of eight hexaploid wheat species, one artificial amphyploid and one intraspecific F_2 hybrid population (Table 2). These wheat species have the spikes of contrasting shapes, which are controlled by well-studied genes [36]. The main spike of each plant was chosen; five images of each spike were acquired in one projection using a 'table' protocol and four, using a 'clip' protocol). The characteristics of the plants and spikes of the images, which were used for training and testing the spike shape recognition, are listed in Supplementary file (Table S2).

Table 2. Wheat species and hybrids used in prediction of spike density index and shape type.

Species	Number of Plants	Number of Images
Triticum compactum Host	63	315
F_2 Triple Dirk B × KU506 *Triticum yunnanense*	52	260
Triticum aestivum L.	50	250
Triticum antiquorum Heer ex Udacz.	20	100
Triticum sphaerococcum Perc.	19	95
Triticum spelta L.	18	90
Amphyploid speltiforme	9	45
Triticum yunnanense King ex S.L. Chen	9	45
Triticum macha Dekapr. et Menabde	9	45

The sample comprises the plants of over 20 genotypes from nine countries (Table S2). The numbers of spelt, normal, and compact spikes are 67, 72, and 110, respectively. The sample contains plants with different types of awnedness, namely, awnless (114 plants), awnletted (64), awned (33), and half-awned (38). The average spike length for different genotypes varies from 3.46 to 11.17 cm. The number of spikelets varies from 14 to 24 with an average of 13 to 45 by 1 dm. Thus, this sample represents a wide variety of spikes relative to their shape, size, and type of awnedness. For each plant, the spike shape type was determined by an expert, its length was measured, and the number of spikelets per spike was counted.

Examples of the images of each spike shape type are shown in Supplementary file (Figures S5–S7). Density index D [37], where $(A - 1)$ is the number of spikelets per spike without the apical spikelet and B, the length of spike rachis (cm), displays a high correlation with spike shape.

$$D = \frac{10(A-1)}{B} \tag{7}$$

2.4.2. Methods for Predicting Spike Characteristics

To predict the density index, we used the random forest method (RandomForestRegressor, sklearn.ensemble software package, scikit-learn library [38]). The spike shape type was determined with the help of LogisticRegression and RandomForestClassifier of the same library.

For prediction, we used 44 parameters of the quadrilateral model and seven indices that describe the spike geometric characteristics (perimeter, spike area, total awn area, circularity index, roundness, solidity, and rugosity). The number of parameters for the model of sections and radial model was reduced to 10 using principal component analysis (the largest components were selected from the sample in Table 2 so that they finally accounted for over 90% of the total variance). Thus, we analyzed 71 parameters for each image.

In the case of 'clip' protocol, each spike is represented in four projections. We pooled the parameters of all four images by ranking projections according to a decrease in spike width (parameter y_{um}). As a result, we got 284 parameters (parameters 1 to 142 describe the frontal side of the spike and 143 to 284, its lateral side).

To estimate the characteristics of spike shape, we selected the most significant parameters according to their predictive capacity for spike density estimation. The parameters were ranked according to their significance based on the averaged results of eight different approaches: (1) Linear regression coefficients (Linear Regression); (2) regression coefficients with lasso regularization (Lasso); (3) regression coefficients with ridge regularization (Ridge); (4) selection of parameters according to stability criterion [39] using RandomizedLasso, sklearn library (Stability); (5) ranking according to recursive feature elimination (RFE) [40]; (6) ranking based on average entropy reduction computed when constructing the trees of solutions by random forest method (RF) using RandomForestRegressor class, sklearn.ensemble software package; (7) ranking based on the correlation between regressor and target parameters (Corr) using f_regression class in sklearn.feature_selection package; and (8) ranking based on the maximal information coefficient (MIC), which takes into account nonlinear dependences between parameters [41] using the MINE (maximal information-based nonparametric exploration) statistics of minepy package. The estimate for a parameter in all these strategies varies in the range of 0 to 1.

2.4.3. Assessing Prediction Performance

To estimate prediction performance for density index, we used the mean absolute error (MAE) and mean absolute percent error (MAPE), calculated as

$$\mathrm{MAE} = \frac{1}{M}\sum\nolimits_{j=1}^{j=M}\left|n_j - n'_j\right| \tag{8}$$

$$\mathrm{MAPE} = \frac{100\%}{M}\sum\nolimits_{j=1}^{j=M}\left(\left|n_j - n'_j\right|/n_j\right) \tag{9}$$

where M is the number of spikes; n_j is the density index calculated manually; and n'_j is the predicted density index value. In addition, we assessed the Pearson correlation coefficient for predicted and expert-assessed values.

The prediction performance for spike classification according to their shape was assessed using measure $F1$, calculated as the mean harmonic for precision and recall [42]:

$$F1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (12)$$

where *TP, TN, FP,* and *FN* correspond to the number of true positive, true negative, false positive, and false negative solutions.

The data were divided into training and test samples at a ratio of 70% to 30%. The prediction accuracy was assessed using cross-validation. To compare different methods for predicting the spike characteristics, the MAE, MAPE, and $F1$ values were averaged over five iterations of the cross-validation for both the training and test samples. However, the parameters were preserved during cross-validation and their values with the mean errors are given in Section 3.3.

2.5. Analysis of F_2 Hybrid Plants

We analyzed F_2 hybrids between a near-isogenic line of the Australian common wheat cultivar Triple Dirk (Triple Dirk B) and the Chinese wheat *Triticum yunnanense* King ex S.L. Chen (syn. *T. spelta* ssp. *yunnanense* (King ex S.L. Chen) N.P. Gontsch.) accession KU 506. All samples included 120 plants grown in a hydroponic greenhouse under standard air humidity, temperature, and light conditions, namely, the day length of 18 h, moisture content of 65%–70%, and temperature of +20/+25 °C (night/day). The Triple Dirk B plants have normal spike shape versus KU 506 with a dense spelt spikes. The spike images of these plants were mainly captured using a 'table' protocol.

3. Results

3.1. Assessing the Recognition Accuracy for Awn and Spike Regions

After selecting the optimal parameters for the algorithm, the achieved recognition accuracy of spike body and awns for the test sample were $J_b = 0.925$ and $J_a = 0.660$, respectively, and for the training sample, $J_b = 0.932$ and $J_a = 0.634$.

Color correction had no significant effect on the recognition accuracy of spike and awns: the mean Jaccard index, J_b, for the spike body segmentation with color correction was 0.925 and for awns with color correction, mean J_a was 0.679.

Figure 4 shows examples of the results of algorithm application to the spike with identifier 6450 from the SpikeDroid database [22]. The estimated recognition performance for spike body and awns in this image was $J_b = 0.963$ and $J_a = 0.796$, respectively. As is evident from Figure 4C, the main share of unclassified pixels concentrated at the awn ends, while the pixels for the larger part of the awns were correctly identified.

Figure 4. Stages of the algorithm for identifying awns by the example of a spike of F_2 hybrid between the Triple Dirk B and Chinese wheat *T. yunnanense* ID no. 6450: (**a**) Spike image on blue background; (**b**) binarized spike pattern; and (**c**) initial spike image with recognized awn regions (marked red) and spike sections (marked blue).

We estimated the effects of various factors associated with the protocols of image acquisition and processing on the accuracy of awn identification, including the scale of imaging (number of pixels per unit captured area), type of the protocol ('table' or 'clip'), and spike projection (front or side) for the latter protocol.

First and foremost, we studied the effect of image spatial resolution on the accuracy of awn identification. As has emerged, the sample of images that we used for assessing the accuracy of detecting the contours contains four scales of spatial resolutions (Supplementary file, Figure S8). All images acquired using the 'table' protocol reside along one line (scale 4, orange circles, highest resolution). The remaining resolution variants (from the lowest resolution scale 1 to medium resolution scale 3) correspond to the images of the 'clip' protocol. Thus, the differences in the image spatial

resolution are inevitable for these protocols and the use of the color scale for determining the image resolution is justified.

Figures 5 and 6 show the distribution of parameter J for identification of the spike body and awns for all 93 images.

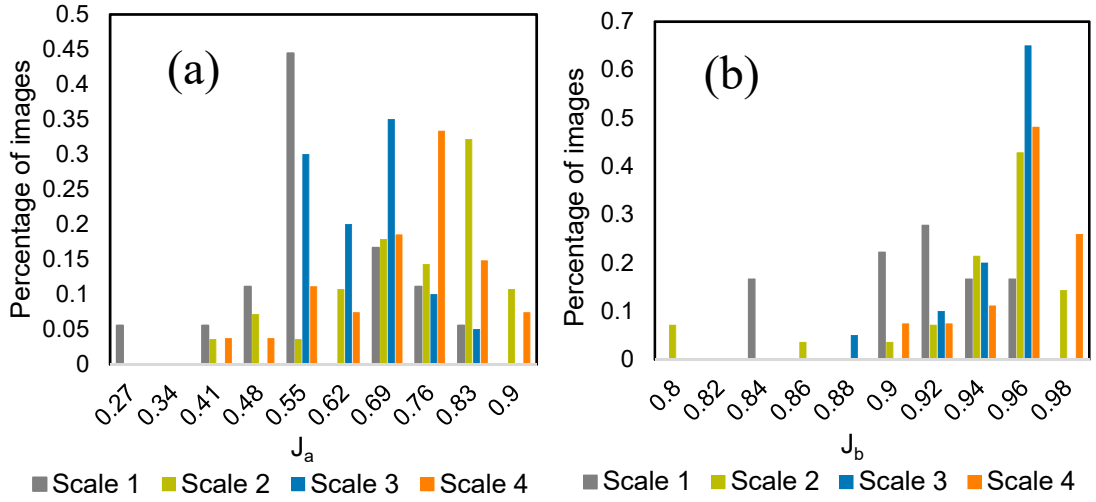

Figure 5. Distribution of Jaccard index J for (**a**) the awns and (**b**) spike bodies grouped according to the scale of spatial resolution.

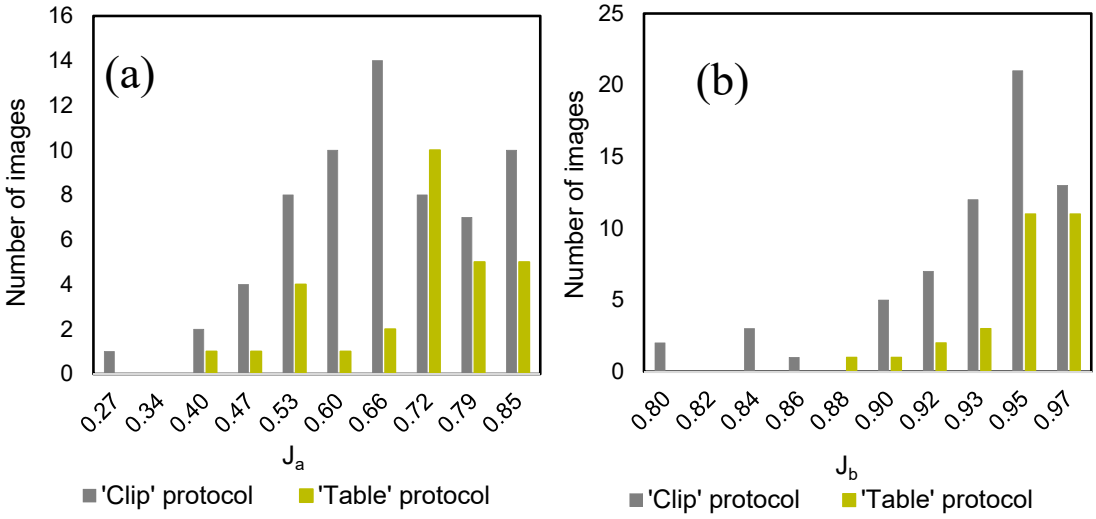

Figure 6. Distribution of the Jaccard indices, J, for (**a**) awns and (**b**) spike body grouped according to the used protocols ('clip' or 'table').

The estimate for the recognition accuracy of awn regions, J_a, falls into the range of 0.27–0.9. For resolution scale 1, the distribution of accuracy is shifted towards the smaller side (Figure 5) and amounts on the average to approximately 0.549; this is logical since it is the lowest resolution. As for the remaining scales (scales 2, 3, and 4), the J_a distributions differ to a lesser degree: Their mean values are 0.695, 0.607, and 0.676, respectively. The recognition accuracy for the spike body, J_b, falls into the range of 0.8–0.98 with means of 0.901, 0.928, 0.938, and 0.945 for scales 1–4, respectively. Note that J_b displays the evident trend of an increase in the recognition accuracy of spike body with an increase in the spatial resolution, unlike J_a.

The second factor putatively influencing the algorithm performance is the type of protocol used. In the 'table' variant, the distance between the camera and object in a set of shots is fixed, while the spike is placed horizontally on the surface of a table in the plane perpendicular to the lens axis. In the case of 'clip' protocol, the spike may deviate from the plane perpendicular to the lens axis because it is

bent or fixed along the axis differing from the vertical. We analyzed the distribution of J for the images grouped according to the used protocols (Figure 6).

The results shown in Figures 5 and 6 demonstrate that the image scale significantly influences the recognition accuracy of both the spike body and awns and the type of protocol used has a significant effect on the identification accuracy for spike body, but not for awns.

The results of analysis of the degree of image quality on segmentation accuracy in the case of JPEG compression demonstrate that the J value when segmenting the ColorChecker and spike body regions is above 0.99 at a QF value of 50 and higher. This means that the image quality in the QF range of 50–100 has almost no effect on the recognition accuracy of ColorChecker and spike body regions. Even for $QF = 1$, the J values for these two regions are >0.98.

The recognition accuracy for the awn pixels depends on QF to a greater degree. In this case, J significantly depends on the awnedness of spike. If the area of awns is large enough, $J > 0.95$ for $QF \geq 90$. As for $QF \geq 50$, the J value is over 0.93. For an awnless spike, $J < 0.9$ as early as $QF = 90$ and drops to 0.83 at $QF = 50$. Nonetheless, the recognition quality for $QF = 90$ (the parameter used in our work) looks quite acceptable.

Note that the performance of our method makes it possible to process one image using one Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz processor on the average over 3 min (with switched off options of debugging information output and image saving at intermediate processing stages).

3.2. Analyzing Awnedness Parameters for Spike Sample

We have analyzed the total awn area (mm^2), S_a. The significant differences between S_a distributions for a sample of 46 images (without repeating different projections) were observed for the images of awnletted and short-awned spikes ($\chi^2 = 22.64$, $p < 0.05$); awnless and short-awned ($\chi^2 = 24.75$, $p < 0.05$); and short-awned and half-awned ($\chi^2 = 18.09$, $p < 0.05$).

Figure 7 shows the distribution of the awnedness types relative to S_a: The short-awned spikes have the largest range; note that they have both large and small values of the total awn area. The largest number of spike images with this awnedness type has a medium area value ($180 < S_a < 222$ mm^2). The awnletted spikes have the minimum total awn area followed by the awnless and half-awned spikes. However, the obtained S_a distributions for different spike shape types considerably overlap, with only one S_a interval housing a single awnedness type, short-awned spikes (264 $mm^2 < S_a$).

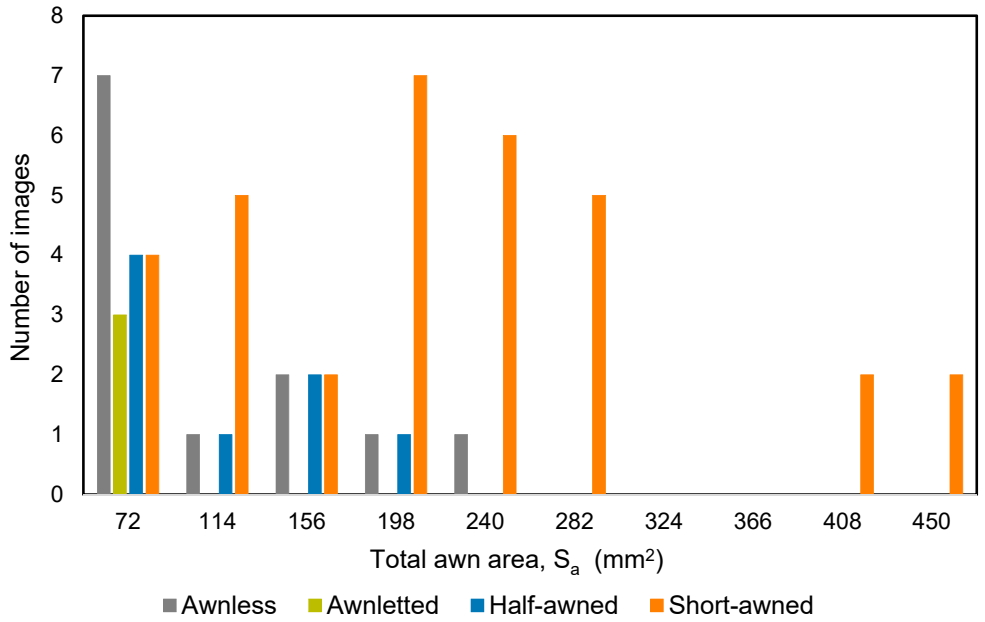

Figure 7. Histogram of the spike distribution according to awnedness type for the sample used to compute the total awn area.

3.3. Predicting Density Index and Type of Spike Shape

The distribution of spikes from the samples shown in Table 2 according to the manually assessed density index is shown in Figure 8. As obvious, the compact spikes display highest density; for normal and spelt spikes the density is within the range of 10–30. The distributions of the last two types considerably overlap. An analogous significant overlapping of the distributions of the number of spikelets and spike length is observable for these spike shape types (Figure 8B). It is also evident that these distributions for the compact spikes are considerably different.

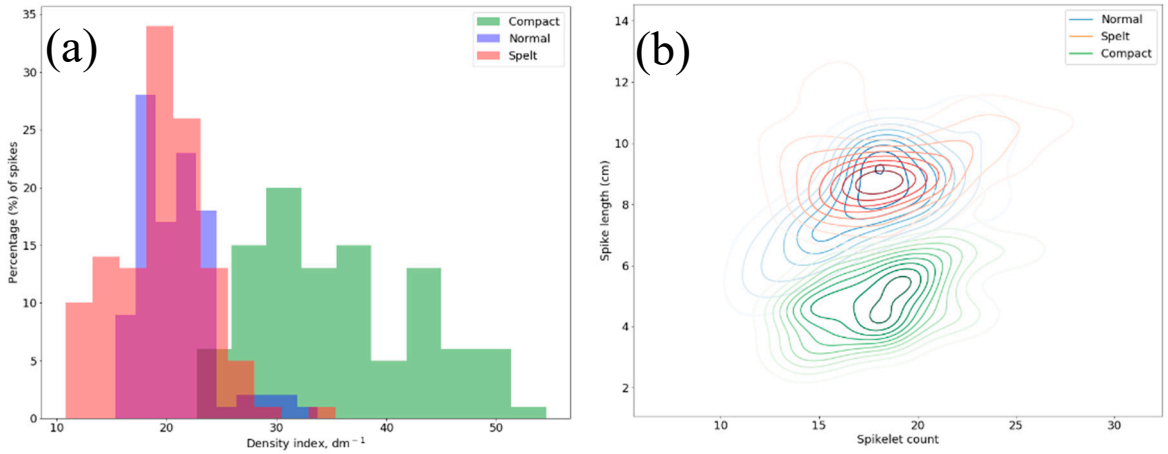

Figure 8. (**a**) Histogram of spike distribution according to density index and (**b**) diagram of spike distribution according to the number of spikelets per spike and spike length. Color is associated with spike shape: Green denotes a compact spike; blue, normal; and red, a spelt one.

To predict the spike density, we selected variables characterizing spike morphology using several criteria (as is described in Section 2.4.2). The variables were ranked based on the averaging of eight significance characteristics. Then, the variables were added one by one in the descending order of mean significance value using both training and test samples. Regression was performed using the random forest method. While adding variables, the MAE values were assessed for the test and training data.

The training involved 12 most significant parameters. The number of parameters for training was selected based on the training curves (Supplementary file, Figure S9), showing the dependence of MAE in predicting spike density index on the number of the best parameters used for training by the random forest method.

We ranked the parameters according to their significance when predicting the spike density index. The results of selection of the 12 best parameters for 'table' (one projection) and 'clip' (four projections) protocols are listed in Tables 3 and 4, respectively.

The parameters with a high significance for both protocols are spike length and circularity index. Interestingly, these traits appear significant for almost all projections in the case of 'clip' protocol.

The errors of prediction for density index and spike shape type (spelt, normal, or compact) are listed in Tables 5 and 6, respectively. The training involved the parameters listed in Table 3 for the images acquired according to 'table' protocol and in Table 4 for the 'clip' images. The mean absolute error in predicting density index in the test sample (MAE test) is 4.61 for one projection and 3.33 for four projections. The Pearson correlation coefficient for the predicted values and expert estimates in the test sample (R) is 0.51 for one projection and 0.74 for four projections. The classification models were trained by logistic regression (lr) and random forest (rf) methods.

Table 6 lists the errors of prediction for spike shape type averaged over five cross-validation iterations. The estimates for these parameters for each cross-validation iteration are listed in Table S4 (Supplementary file). We attempted to use the predicted density index (lr_density_pred and rf_density_pred) as an additional trait, but had no gain in accuracy. The values of $F1$ measure for

performance of spike shape classification in test sample is 0.85 (F1_rf) for four projections and 0.78 (F1_lr) for one projection. These accuracy estimates demonstrate that utilization of the information about four projections increases the prediction performance by 7% and decreases MAE for spike density prediction by 1.28, which corresponds to 4.25%.

Table 3. Estimation of the measures of significance of spike quantitative characteristics for predicting spike density based on analysis of the images acquired by 'table' protocol: characteristics are listed in column 1; significance values, in columns 2–9; and their means, in column 10. The significance measures are described in Methods, Section 2.4.2.

Parameter	Linear Regression	Ridge	Lasso	Stability	RFE	RF	Corr	MIC	Mean
L	0.93	0.07	0.04	0.74	0.81	1	1	1	0.7
R	0	0.75	0	0.68	1	0.06	0.65	0.93	0.51
x_{u2}	1	0.1	0.11	0.61	0.73	0.07	0.5	0.88	0.5
t_{b1}	0	1	1	0.04	0.95	0.01	0	0.34	0.42
P	0	0.07	0.07	1	0.66	0.06	0.49	0.71	0.38
profile_8	0	0.51	0.55	0.86	0.88	0.03	0.07	0.13	0.38
y_{b1}	0	0.7	0.7	0.21	0.92	0.05	0.04	0.37	0.37
y_{u1}	0	0.82	0.78	0.02	0.75	0.03	0.01	0.41	0.35
profile_1	0	0.46	0.45	0.44	0.71	0.18	0.07	0.51	0.35
radial_3	0	0.01	0.01	0.92	0.2	0.21	0.31	0.92	0.32
y_{b2}/L	0	0	0	0.48	1	0.13	0.33	0.6	0.32
C	0	0.2	0	0.34	1	0.05	0.36	0.64	0.32

Table 4. Estimation of the measures of significance of spike quantitative characteristics for predicting spike density based on analysis of the images acquired by 'clip' protocol (the corresponding projections are parenthesized); characteristics are listed in column 1; significance values, in columns 2–9; and their means, in column 10. The significance measures are described in Methods, Section 2.4.2.

Parameter	Linear Regression	Ridge	Lasso	Stability	RFE	RF	Corr	MIC	Mean
x_{b2} (projection 2)	1	0.15	0.13	0.75	1	0	0.46	0.68	0.52
x_{u2} (projection 2)	0.98	0.06	0.04	0.87	0.9	0.04	0.63	0.63	0.52
C (projection 3)	0	0.24	0	0.81	0.53	1	0.62	0.94	0.52
y_{b1}/L (projection 3)	0	0.01	0	0.9	0.59	0.86	0.86	0.8	0.5
R (projection 3)	0	0.38	0	0.94	0.64	0.01	0.84	0.96	0.47
L (projection 4)	0.42	0.06	0.01	0.5	1	0.01	0.79	0.79	0.45
R (projection 4)	0	0.32	0	0.88	0.52	0.01	0.81	1	0.44
y_{u1}/L (projection 3)	0	0.01	0	0.94	0.71	0.02	1	0.71	0.42
L (projection 3)	0.11	0.08	0.02	0.26	1	0.2	0.73	0.76	0.4
y_{u1}/L (projection 4)	0	0.01	0	1	0.48	0.01	0.84	0.8	0.39
R (projection 2)	0	0.19	0	0.8	0.49	0.03	0.63	1	0.39
x_{u2} (projection 1)	0.58	0.05	0	0.52	0.73	0.03	0.43	0.65	0.37

Table 5. Estimates of prediction performance for of spike density index for one ('table') and four ('clip') projections ($MAE_{training}$ and MAE_{test}, mean absolute errors for training and test samples; $MAPE_{training}$ and $MAPE_{test}$, mean absolute percent errors for training and test samples; $R^2_{training}$ and R^2_{test}, Pearson correlation coefficient between the predicted and expert estimates for training and test samples).

Performance Measure	'Clip'	'Table'
$MAE_{training}$	1.48	1.88
MAE_{test}	3.33	4.61
$MAPE_{training}$	6.19	7.77
$MAPE_{test}$	13.28	17.80
$R^2_{training}$	0.95	0.94
R^2_{test}	0.75	0.52

Table 6. Estimates of prediction performance for spike shape type for one ('table') and four ('clip') projections (F1_lr and F1_rf, the *F*1 measure for the accuracy of spike classification according to density by logistic regression and random forest methods; *F*1 lr_density_pred and F1 rf_density_pred, the *F*1 measure for the classification accuracy using an additional parameter, predicted density index, by logistic regression, and random forest methods).

Performance Measure	'Clip'	'Table'
F1_lr	0.82	0.78
*F*1 lr_density_pred	0.83	0.77
F1_rf	0.85	0.72
*F*1 rf_density_pred	0.84	0.76

Table 7 lists the estimates of confusion matrix values for the classification of spike shape in the test sample for four projections obtained by the random forest method. It shows that compact and spelt spikes are distinguished with the least number of errors. As for the normal spikes, they are most frequently predicted as belonging to two other types (lowest prediction quality). Interestingly, the spikes belonging to normal and spelt ones give the highest number of mutual false predictions. These results agree with the histograms in Figure 8A, which are based on expert estimates.

Table 7. Confusion matrix for classification of spike shape in the test sample of images (see Table 2) in four projections ('clip' protocol).

	Compact Observed	Normal Observed	Spelt Observed
Compact predicted	158	13	0
Normal predicted	8	79	20
Spelt predicted	0	20	77

3.4. Analysis of F_2 Hybrid Plants

The results of comparison between spike length estimated manually and *L* parameter obtained using 2D images is represented in Supplementary file Figure S10. The Pearson correlation coefficient between two values is 0.808 ($p < 0.01$), MAE = 0.75 cm, MAPE = 0.09. We predicted spike density index for this sample also. The Pearson correlation between predicted and manually estimated *D* values is 0.69, MAE = 2.59, MAPE = 11.26, which is similar to the values estimated for the spike test samples (Table 5). The distribution of predicted *D* values for F2 hybrid plants shown in Figure 9. This distribution is bimodal. Interestingly, two groups of spikes with $D < 26$ (94 plants) and $D > 26$ (26 plants) correspond well to the values of density index from parent plants (see Figure 9).

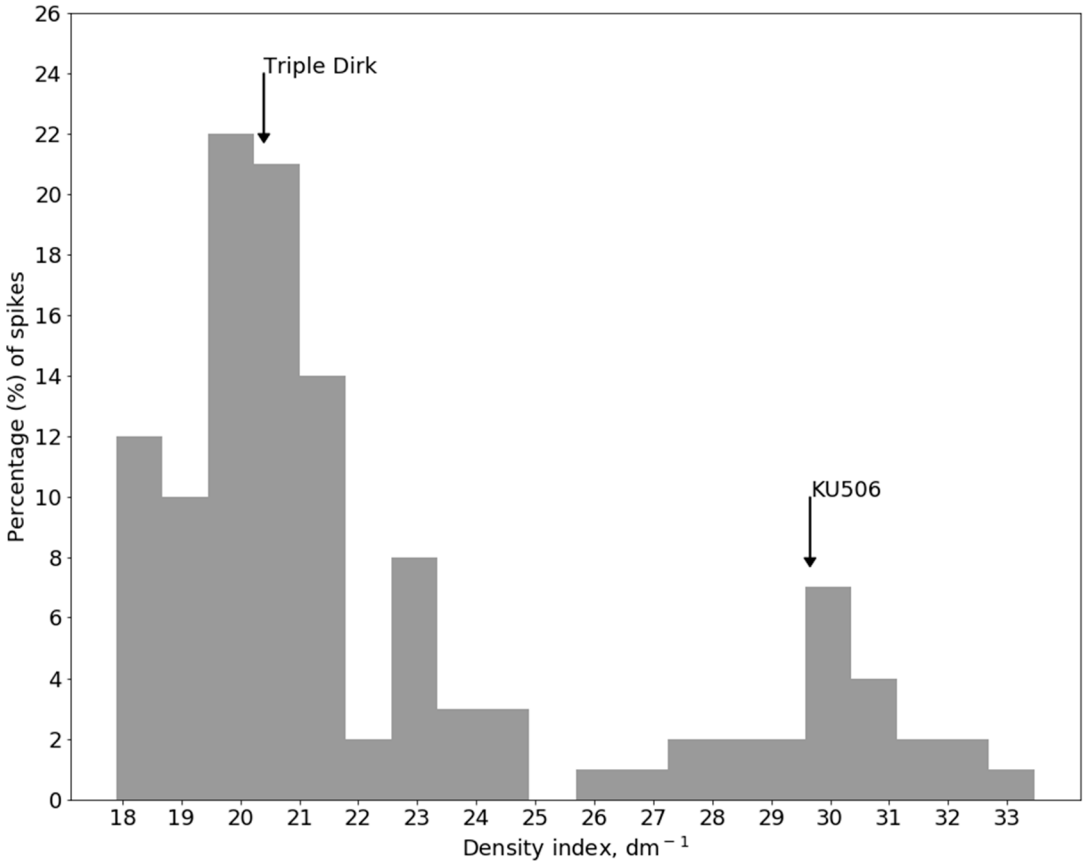

Figure 9. Distribution of the spike density values predicted using 2D images for the F_2 hybrids between a near-isogenic line of the Australian common wheat cultivar Triple Dirk (Triple Dirk B) and Chinese wheat *Triticum yunnanense* King ex S.L. Chen (syn. *T. spelta* ssp. *yunnanense* (King ex S.L. Chen) N.P. Gontsch.) accession KU 506. The X axis shows the density index and the Y axis is the percentage of spikes. Arrows point to the parent *D* values, amounting to 20.40 for Triple Dirk B and 29.66 for KU 506.

4. Discussion

The insight into the genetic control of wheat spike characteristics requires morphometry of a large number of accessions. This process can be automated utilizing various methods for image analysis. However, it is desirable that the corresponding methods are applicable to mass use and utilize simple and inexpensive phenotyping techniques. When elaborating our approach, we followed these criteria; in particular, we used the protocol that did not require any expensive equipment and gives the images of acceptable quality. An advantage of the protocol is that it allows for analysis of the spikes from different collections.

Analysis of the identification performance for spike region has shown that the image scale has the most significant effect on the background/spike and spike body/awns segmentation accuracy: The farther the spike from the camera lens, the higher is the segmentation error. The imaging resolution considerably differs depending on the protocol for image acquisition; in 'table' protocol, the camera is considerably closer to the object so the resolution is high, whereas in 'clip' variant, the distance to the object may vary, making the resolution lower. Since the focal length was fixed, the resolution depended only on the distance to the imaged object.

Color correction of images has no significant effect on the segmentation accuracy. Most likely, this is the result of standard illumination during spike capturing [22]. Nonetheless, the option of color correction, available in our method, allows the colors in images to be more objectively estimated and assists in considerable improvement of data processing under conditions of poorly controllable illumination.

However, the captured images may have some flaws. For example, spikes have the awns going beyond the spike plane, which can interfere with the focusing, making the resulting image of the spike or its parts blurry (Figure 2). This is surmountable by an increase in the depth of focus, manual focusing, and capture of several images to discard the defect images. Nonetheless, it has turned out that the image quality even for such images is sufficient to assess the spike shape, thereby demonstrating that our method is suitable for the screening of spikes.

A smaller J_a value as compared to J_b is in part associated with a small area of awns relative to the overall spike and concurrently with a larger number of boundary pixels for the awn region. Thus, cost of the error at binarization stage is considerably higher for the awn region as compared with the spike body region. In addition, the very manual marking of high-resolution images is laborious and does not allow the fluctuations in identification of spike boundaries to be avoided.

Here, we propose a method for assessing the parameters of spike shape. These parameters are arbitrary dividable into "general" characteristics reflecting the overall shape (for example, spike circularity, area, and length). We also additionally introduced the ad-hoc models for description of spike shape. Two of them (model of sections and radial model) were formulated based on most general considerations and their parameters have no specificity associated with the typical wheat spike shape. The third model, which describes the spike as two quadrilaterals with a common side, may be a more illustrative generalization of spike shape. The fact that a considerable number of parameters used in this model are among those significant for predicting the spike density (Tables 3 and 4) suggests that this model is the most adequate. The general spike characteristics and the parameters of the model of quadrilaterals are prevalent among the significant parameters. Figures S5–S7 (Supplementary file) shows how the model approximates two quadrilaterals for the spikes of different shapes as well as the indices with the largest contributions to density index predictions,

In our work, we have assessed the prediction accuracy for spike density and shape type based on the quantitative characteristics obtained by analyzing spike images. Two regression methods, logistic regression and random forest, were selected since the volume of training sample (in our case, ~200 spikes) is not critical for them as compared with neural networks or deep learning neural networks, which require thousands and tens of thousands of images to gain a good result [19].

The logistic regression–based prediction of spike density gives a MAE value of 4.61 (using one spike image projection). Note that the prediction accuracy increases (MAE = 3.33) when the data for four projections are used, although the number of parameters for prediction remains the same (however, these 12 parameters include the characteristics obtained by analyzing different spike projections).

The accuracy of spike shape classification into three types (spelt, normal, and compact) using the random forest algorithm ($F1$ measure) is 0.78 for one spike projection and 0.85 for four projections. Interestingly, the latter value is close to the estimate for classifying spikes of four wheat varieties by Kun et al. [21], which amounts to 0.88 (neural networks using 12 input parameters). A difficulty in solving this problem is that the spelt and normal spikes may be rather similar, as is shown, in particular, by a considerable overlapping of the corresponding spike density distributions (Figure 8). Further accumulation of annotated images and use of deep learning neural network methods will be helpful in solving this problem.

Results of F_2 hybrid plants demonstrated high correlation between spike length values estimated manually and from 2D image analysis. This allows us to hope for the effective use of the proposed method to assess the character of the hybrid segregations in interspecific wheat hybridization.

5. Conclusions

Summing up, we have proposed a method allowing for automated assessment of quantitative spike shape and awnedness characteristics by analyzing the spike images captured with the help of standard protocols. While the efficiency of this method in its current version is improvable at certain stages by optimization of processing algorithms and time, the already available image segmentation accuracy is quite acceptable for further determination of spike shape characteristics.

It has been shown that the mean error of spike density prediction for the images in the plane of table amounts to 4.61 (~18%) for the test sample and to 3.33 (~13%) for the prediction involving all four projections. In an automated spike classification into three types, the value of measure *F*1 for the images captured according to 'table' protocol is 0.78 versus 0.85 for the 'clip' variants with analysis of four projections.

The method can be useful for assessing the spike characteristics in breeding and genetic studies as well as for analyzing the spikes in collections.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4395/9/7/390/s1, Supplementary file in PDF format: 1. Protocol for Image Capture; 2. Algorithm for Color Scale Identification; 3. Supplementary tables; 4. Supplementary figures; Table S1. Parameters determined by the application for recognizing spike shape; Table S2. Characteristics of the plants and spikes the images of which were used for training and testing the method for recognition of spike shape; Table S3. The effect of JPEG image quality on the accuracy of segmentation; Table S4. The performance measures for spike shape classification and density estimation for each of the five cross-validation iterations; Figure S1. 'Clip' protocol; Figure S2. 'Table' protocol; Figure S3. A zoomed-in fragment of the image illustrating the result of operation of the smoothing algorithm using a Gaussian filter; Figure S4. Images of spikes and the results of segmentation for two plants of the F2 hybrids between common wheat Triple Dirk B and *Triticum yunnanense* with different awn types; Figure S5–7. Stages of algorithm operation for compact, normal and spelt spikes; Figure S8. Ratio of the total pixels of awns to the total awn area (mm^2) grouped according to imaging scale; Figure S9. Dependence of the mean absolute error (MAE) in predicting spike density index on the number of best parameters used for training with random forest method; Figure S10. Scatterplot diagram for the main spike length for Triple Dirk B × KU506 F_2 hybrid plants estimated manually and from the 2D image analysis.

Author Contributions: Conceptualization, M.A.G., D.A.A. and N.P.G.; methodology, M.A.G., E.G.K. and D.A.A.; software, E.G.K., N.V.S. and M.A.G.; validation, E.G.K. and M.A.G.; formal analysis, M.A.G.; investigation, M.A.G., E.G.K. and N.V.S.; resources, Y.V.K., and N.P.G.; data curation, Y.V.K. and N.P.G.; writing—original draft preparation, M.A.G., E.G.K. and D.A.A.; writing—review and editing, M.A.G., N.P.G. and D.A.A.; visualization, M.A.G. and E.G.K.; supervision, D.A.A. and N.P.G.; project administration, M.A.G.; funding acquisition, M.A.G.

Funding: The work was supported by the Russian Science Foundation (project no. 17-74-10148). Bioinformatics data analysis was performed with the help of computational resources of the Bioinformatics Joint Access Center, supported by budget project no. 0324-2019-0040.

Acknowledgments: The authors are grateful to Galina Chirikova for translating this manuscript from Russian to English. We are grateful to anonymous reviewers for their helpful comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Genaev, M.; Komyshev, E.; Smirnov, N.; Kruchinina, Y.; Goncharov, N.P.; Afonnikov, D. *The International Comecon List of Descriptors for the Genus Triticum L.*; VIR: Leningrad, Russia, 1984. (In Russian)
2. Konopatskaia, I.; Vavilova, V.; Blinov, A.; Goncharov, N.P. In Spike morphology genes in wheat species (*Triticum* L.). *Proc. Latv. Acad. Sci. Sect. B. Nat. Exact. Appl. Sci.* **2016**, *70*, 345–355. [CrossRef]
3. Goncharov, N.P. Genus *Triticum* L. taxonomy: The present and the future. *Plant Syst. Evol.* **2011**, *295*, 1–11. [CrossRef]
4. Börner, A.; Schäfer, M.; Schmidt, A.; Grau, M.; Vorwald, J. Associations between geographical origin and morphological characters in bread wheat (*Triticum aestivum* L.). *Plant Genet. Resour.* **2005**, *3*, 360–372.
5. Guo, Z.; Zhao, Y.; Röder, M.S.; Reif, J.C.; Ganal, M.W.; Chen, D.; Schnurbusch, T. Manipulation and prediction of spike morphology traits for the improvement of grain yield in wheat. *Sci. Rep.* **2018**, *8*. [CrossRef] [PubMed]
6. Goriewa-Duba, K.; Duba, A.; Wachowska, U.; Wiwart, M. An Evaluation of the Variation in the Morphometric Parameters of Grain of Six *Triticum* Species with the Use of Digital Image Analysis. *Agronomy* **2018**, *8*, 296. [CrossRef]
7. Hammer, K.; Filatenko, A.A.; Pistrick, K. Taxonomic remarks on *Triticum* L. and × *Triticosecale* Wittm. *Genet. Resour. Crop. Evol.* **2011**, *58*, 3–10. [CrossRef]
8. Matsuoka, Y.; Nishioka, E.; Kawahara, T.; Takumi, S. Genealogical analysis of subspecies divergence and spikelet-shape diversification in central Eurasian wild wheat *Aegilops Tauschii* Coss. *Plant Syst. Evol.* **2009**, *279*, 233–244. [CrossRef]

9. Li, Y.; Cui, Z.; Ni, Y.; Zheng, M.; Yang, D.; Jin, M.; Chen, J.; Wang, Z.; Yin, Y. Plant density effect on grain number and weight of two winter wheat cultivars at different spikelet and grain positions. *PLoS ONE* **2016**, *11*, e0155351. [CrossRef]
10. Afonnikov, D.; Genaev, M.; Doroshkov, A.; Komyshev, E.; Pshenichnikova, T. Methods of high-throughput plant phenotyping for large-scale breeding and genetic experiments. *Russ. J. Genet.* **2016**, *52*, 688–701. [CrossRef]
11. Giuffrida, M.V.; Chen, F.; Scharr, H.; Tsaftaris, S.A. Citizen crowds and experts: Observer variability in image-based plant phenotyping. *Plant Methods* **2018**, *14*, 12. [CrossRef]
12. Fahlgren, N.; Gehan, M.A.; Baxter, I. Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* **2015**, *24*, 93–99. [CrossRef] [PubMed]
13. Tanabata, T.; Shibaya, T.; Hori, K.; Ebana, K.; Yano, M. SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiol.* **2012**, *4*, 1871–1880. [CrossRef] [PubMed]
14. Komyshev, E.; Genaev, M.; Afonnikov, D. Evaluation of the SeedCounter, a mobile application for grain phenotyping. *Front. Plant Sci.* **2017**, *7*, 1990. [CrossRef] [PubMed]
15. Wu, W.; Zhou, L.; Chen, J.; Qiu, Z.; He, Y. GainTKW: A Measurement System of Thousand Kernel Weight Based on the Android Platform. *Agronomy* **2018**, *8*, 178. [CrossRef]
16. Strange, H.; Zwiggelaar, R.; Sturrock, C.; Mooney, S.J.; Doonan, J.H. Automatic estimation of wheat grain morphometry from computed tomography data. *Funct. Plant Biol.* **2015**, *42*, 452–459. [CrossRef]
17. Grillo, O.; Blangiforti, S.; Venora, G. Wheat landraces identification through glumes image analysis. *Comput. Electron. Agric.* **2017**, *141*, 223–231. [CrossRef]
18. Makanza, R.; Zaman-Allah, M.; Cairns, J.; Eyre, J.; Burgueño, J.; Pacheco, Á.; Diepenbrock, C.; Magorokosho, C.; Tarekegne, A.; Olsen, M. High-throughput method for ear phenotyping and kernel weight estimation in maize using ear digital imaging. *Plant Method.* **2018**, *14*, 49. [CrossRef]
19. Pound, M.P.; Atkinson, J.A.; Wells, D.M.; Pridmore, T.P.; French, A.P. Deep learning for multi-task plant phenotyping. In Proceedings of the IEEE International Conference on Computer Vision (ICCVW), Venice, Italy, 22–29 October 2017; pp. 2055–2063.
20. Hughes, N.; Askew, K.; Scotson, C.P.; Williams, K.; Sauze, C.; Corke, F.; Doonan, J.H.; Nibau, C. Non-destructive, high-content analysis of wheat grain traits using X-ray micro computed tomography. *Plant Methods* **2017**, *13*, 76. [CrossRef]
21. Bi, K.; Jiang, P.; Li, L.; Shi, B.; Wang, C. Non-destructive measurement of wheat spike characteristics based on morphological image processing. *TCSAE* **2010**, *26*, 212–216.
22. Genaev, M.A.; Komyshev, E.G.; Fu, H.; Koval, V.S.; Goncharov, N.P.; Afonnikov, D.A. SpikeDroidDB-information system for annotation of morphometric characteristics of wheat spike. *VOGiS* **2018**, *22*, 132–140. (In Russian) [CrossRef]
23. Berry, J.C.; Fahlgren, N.; Pokorny, A.A.; Bart, R.S.; Veley, K.M. An automated, high-throughput method for standardizing image color profiles to improve image-based plant phenotyping. *PeerJ* **2018**, *6*, e5727. [CrossRef] [PubMed]
24. Kaehler, A.; Bradski, G. *Learning OpenCV 3: Computer vision in C++ with the OpenCV library*; O'Reilly Media Inc.: Sebastopol, CA, USA, 2016.
25. Quintana, J.; Garcia, R.; Neumann, L. A novel method for color correction in epiluminescence microscopy. *Comput. Med. Imag. Grap.* **2011**, *35*, 646–652. [CrossRef] [PubMed]
26. Shaik, K.B.; Ganesan, P.; Kalist, V.; Sathish, B.; Jenitha, J.M.M. Comparative study of skin color detection and segmentation in HSV and YCbCr color space. *Proc. Comp. Sci.* **2015**, *57*, 41–48. [CrossRef]
27. Kumar, P.; Miklavcic, S. Analytical study of colour spaces for plant pixel detection. *J. Imaging* **2018**, *4*, 42. [CrossRef]
28. Jaccard, P. The distribution of the flora in the alpine zone. 1. *New Phytol.* **1912**, *11*, 37–50. [CrossRef]
29. Everingham, M.; van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
30. Whitley, D. A genetic algorithm tutorial. *Stat. Comput.* **1994**, *4*, 65–85. [CrossRef]
31. Wallace, G.K. The JPEG still picture compression standard. *IEEE Transact. Consum. Electron.* **1992**, *38*, xviii–xxxiv. [CrossRef]
32. McEntee, M.F.; Nikolovski, I.; Bourne, R.; Pietrzyk, M.W.; Evanoff, M.G.; Brennan, P.C.; Tay, K.L. The effect of JPEG2000 compression on detection of skull fractures. *Acad. Radiol.* **2013**, *20*, 712–720. [CrossRef]

33. Fidler, A.; Skaleric, U.; Likar, B. The impact of image information on compressibility and degradation in medical image compression. *Med. Phys.* **2006**, *33*, 2832–2838. [CrossRef]
34. Kuhl, F.P.; Giardina, C.R. Elliptic Fourier features of a closed contour. *Comput. Graph. Image Process.* **1982**, *18*, 236–258. [CrossRef]
35. Press, W.H. *Numerical Recipes in Fortran 77: The Art of Scientific Computing*; Cambridge University Press: Cambridge, UK, 1992.
36. Swaminathan, M.; Rao, M. Macro-mutations and sub-specific differentiation in Triticum. *Wheat Inf. Serv.* **1961**, *13*, 9–11.
37. Flaksberger, K.A. Pshenitsi-rod Triticum, L. Wheats-genus Triticum L. In *Cultivated Flora of the USSR. Bread Cereals—Wheat*; Wulff, E.V., Ed.; Gosselkhozgiz: Moscow, Russia, 1935; pp. 17–434. (In Russian)
38. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
39. Meinshausen, N.; Bühlmann, P. Stability selection. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2010**, *72*, 417–473. [CrossRef]
40. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]
41. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting novel associations in large data sets. *Science* **2011**, *334*, 1518–1524. [CrossRef] [PubMed]
42. Van Rijsbergen, C.J. *Information Retrieval*, 2nd ed.; Butterworths: London, UK, 1979; ISBN 0-408-70929-4.