

Article

# Image-Based On-Panicle Rice [*Oryza sativa* L.] Grain Counting with a Prior Edge Wavelet Correction Model

Liang Gong <sup>1,\*</sup>, Ke Lin <sup>1</sup> , Tao Wang <sup>1</sup>, Chengliang Liu <sup>1,\*</sup>, Zheng Yuan <sup>2</sup>, Dabing Zhang <sup>2</sup> and Jun Hong <sup>2</sup>

<sup>1</sup> School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; chris\_lin@sjtu.edu.cn (K.L.); wangtao12@sjtu.edu.cn (T.W.)

<sup>2</sup> School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; zyuanyuan@sjtu.edu.cn (Z.Y.); zhangdb@sjtu.edu.cn (D.Z.); hongjun0699@sjtu.edu.cn (J.H.)

\* Correspondence: gongliang\_mi@sjtu.edu.cn (L.G.); chlliu@sjtu.edu.cn (C.L.); Tel.: +86-21-3420-7061 (L.G.); +86-21-3420-4350 (C.L.)

Received: 28 February 2018; Accepted: 16 May 2018; Published: 7 June 2018



**Abstract:** The number of rice grains on a panicle is an important index for variety screening during high-quality rice [*Oryza Sativa* L.] breeding. For an in-vivo image-based measurement, the occlusion and overlapping among grains are the major challenges in non-destructive precise phenotyping of the on-panicle grains. In order to tackle these challenges, this paper describes a correction-model-referred on-panicle grain counting method based on the area of the rice panicle and its edge contour wavelet analysis. First, we assume that a deterministic correlation exists between the number of grains of the panicle and the traits of its edge contour morphology, which reflects the extent to which the grains are occluded. Second, a method for coarsely estimating grain number per panicle is proposed based on the projective area of the panicle in the image and the average area of a rice grain. Finally, a correction model which is built with the average wavelet frequency of the edge contour of the panicle is employed to correct the estimated value of the grain number. Two randomly selected cases are investigated in detail, showing that computation accuracy with a correction model is increased by 26% and 23% respectively when compared to that of the naive area-based computation. In conclusion, we reveal and validate the relationship between the number of grains of the panicle and the fluctuation frequency of its edge contours. Further, experiments show that errors caused by overlapping and occlusion scenarios can be alleviated with the estimation and correction hybrid models, achieving an average accuracy of 94% compared to the results of manual counting.

**Keywords:** rice panicle; phenotyping; grain counting; image processing; wavelet analysis

## 1. Introduction

Rice [*Oryza Sativa* L.] is one of the world's major food crops, and is the staple food for more than half of the world's population [1]. Major goals of current agricultural research are to maximize the yield potential, grain quality, and stress resistance of rice grown in today's changing climate [2]. Therefore, new technologies need to be developed to accelerate the breeding process through more advanced phenotyping and genotyping methods [3]. Currently, there are many high-throughput phenotyping platforms to measure plant growth and morphology [4]. These platforms can be divided into three categories depending upon the measurement environment: platforms for laboratory-, greenhouse-, or field-based measurements [5]. The Australian Plant Phenotypic Group Facility (APPF) combines digital image processing technology, large-scale computing technology and robotics, which has been successfully applied to high-throughput precision modeling and prediction of grain biomass of greenhouse plants [6]. PhenoFab is a high-tech phenotyping platform developed

by the KeyGene company, which is a greenhouse service operation that combines a high-throughput, non-invasive technology with trait interpretation to exploit phenotypic variation [7]. Other representative high-throughput phenotyping platforms include Phenoscope [8], Glyph [9], Phenodyn [10], GROWSCREEN-Rhizo [11], FLUORO [12], GROWSCREEN [13] and PHENOPSIS [14].

The grain yield of a rice plant is closely related to panicle traits such as the total number of grains per panicle, the number of filled grains per panicle, and the seed setting rate (the ratio between the filled grains and the total grains). Hence, obtaining phenotypic traits of rice panicles plays an extremely important role in rice high quality breeding research [15]. Among these traits, the number of grains per panicle directly reflects the yield of rice. Therefore, achieving accurate and efficient measurements of the number of rice grains is of great significance for rice breeding.

The Rice Phenotyping Digital Observation Platform developed by HUSTC (Huazhong University of Science and Technology, Wuhan, China) used roller-compacted threshing mechanism and image processing method to measure rice yield-related traits [16]. However, this method requires destructive sampling, in which rice grains are detached from the branches, undermining the integrity of the plant. Since rice panicle architecture is also a key target of selection when breeding for yield and grain quality, it is essential to retain the integrity of the panicle to simultaneously obtain the panicle traits such as the primary branch length, internode numbers, and exertion thickness [17].

In general, non-destructive counting methods are image-based and require computational methods to extract the phenotypes of interest quickly and efficiently from the digital images. As a result, software development for image analysis is an important, emerging area in plant phenotyping; for example, Lobet et al. (2013) created an online database for plant image analysis software tools, which could help researchers select the most suitable software for their research [18]. These tools tend to be specific to particular plant species and/or specific phenotypes, and new tools must be developed and refined in order to expand the number of plant species and traits that can be computationally phenotyped.

Using computational image analysis to count grains on intact panicles of cereal crops is a particularly challenging problem because the grains often overlap one another and block some grains from being visible to the camera. Similar problems are encountered in blood cell counting studies and have previously been addressed in medical research using either a feature-based method [19] or a data-driven method using a neural network [20]. However, the feature-based method, commonly used when the image features can be explicitly described, cannot be applied to counting rice grains or many other seeds because they overlap more extensively than blood cells, causing more severe problems with invisibility. In addition, although a data-driven method using a neural network might achieve satisfactory results, this approach involves too high levels of computational complexity and memory consumption to be employed by portable and affordable phenotyping instruments.

An alternative to feature-based methods or neural networking is to estimate the number of grains on a panicle based on the overall structure and outline of the panicle. To develop an image-based counting method for rice grains, Zhao and coworkers [15] analyzed the correlation between the panicle area, the branch length and the number of grains per panicle. This work demonstrated that there is a strong correlation among these three traits, and that, compared to branch length, the rice panicle area is more strongly correlated with grain numbers. P-trap, a plant image analysis software for grain counting was jointly developed by the Algarve University in Portugal and the French Institute of Research and Development. This software uses the area ratio between a rice grain and a panicle to calculate the number of rice grains in a panicle, and corrects for potential overlap between grains using a fixed coefficient  $k$  [21]. Using a fixed value of coefficient  $k$ , which is not altered according to the differing extent of overlap among individual varieties, might result in inaccuracies when applied to rice varieties with very dense panicles.

To overcome the disadvantages of using the constant coefficient  $k$ , this paper proposes a method of edge contour wavelet analysis to correct for overlap and occlusion among the grains. Because the contours of each rice grain are curved, the morphology of the panicle outline is affected by the grain

edges and the contour geometric features can be used to estimate the grain number. Fortunately, the wavelet provides a multi-scale representation of the contour geometrics with a computable coefficient, which can be used to estimate the number of grains even in situations in which grains partially overlap. Specifically, the frequency of fluctuations in the edge of the panicle was used to characterize the degree of overlap between rice grains, and instead of using a correction factor with a fixed value  $k$ , the wavelet frequency ratio  $f'/f_0$  was used because it varies with the grain density of the panicle. When this method was used to count the number of rice grains in 50 genetically diverse panicles, the results indicated that the average counting error was 6% when compared to manual counting results. Using the edge contour frequency ratio can effectively characterize the degree of overlap between rice grains, which is better than using a fixed coefficient to correct for overlap when counting rice grains. Our method is implemented in MATLAB 2016b and is available at <https://github.com/ChrisLinSJTU/rice-counting>.

## 2. Materials and Methods

### 2.1. System Design

Intuitively the overlapping and occlusion among on-panicle grains are able to be described with the complexity metrics of the grain edges, and the grain edge fluctuations correspond to a pattern that determines the visibility of the grains. Therefore, the occlusion and overlap can be represented by the one-dimensional wavelet expression of the grain edges, which can be modeled as a correction coefficient to quantitatively evaluate the grain visibility. This inspired a more precise method for counting the on-panicle grains, i.e., a coarse area model estimates the grain numbers, and an incremental edge wavelet model gives a correction coefficient to improve its accuracy.

The process that was used for counting rice grains with edge contour wavelet analysis is shown in Figure 1. First, a panicle image was taken in a Seed & Panicle Phenotyping Instrument. The Seed & Panicle Phenotyping Instrument is an image-based portable rice phenotype platform. Its shape is a  $60 \times 40 \times 50$  cm rectangle, with a tablet computer above it for photographing plants and analyzing the morphological characteristics of the plant. Panicles were held on the platform using black round magnets that were fixed in transparent plastic cubes. Second, after the panicles were photographed, preprocessing operations were applied to the image, including binarization, circle detection, corrosion and expansion, and hole filling. Third, the area  $m(k)$  and the main axis of each connected region was calculated. Then, the sequence  $s^a$  of distance from region edge points to the corresponding main axis in each region was calculated. Next, a 1-D continuous wavelet transformation was applied to each  $s^a$ , and the average frequency  $f'^a$  was obtained. Then, based on measurements of the average area of rice grains  $m_0$  and the average fluctuation frequency of a rice grain's edge without overlap  $f_0$ , the formula  $m(k)/m_0 \times f'^a/f_0$ , was used to calculate the number of rice grains in each connected region. Finally, the number of rice grains in each region was summed to estimate the total number of rice grains in the image. A flow chart of the rice counting algorithm is shown in Figure 1.

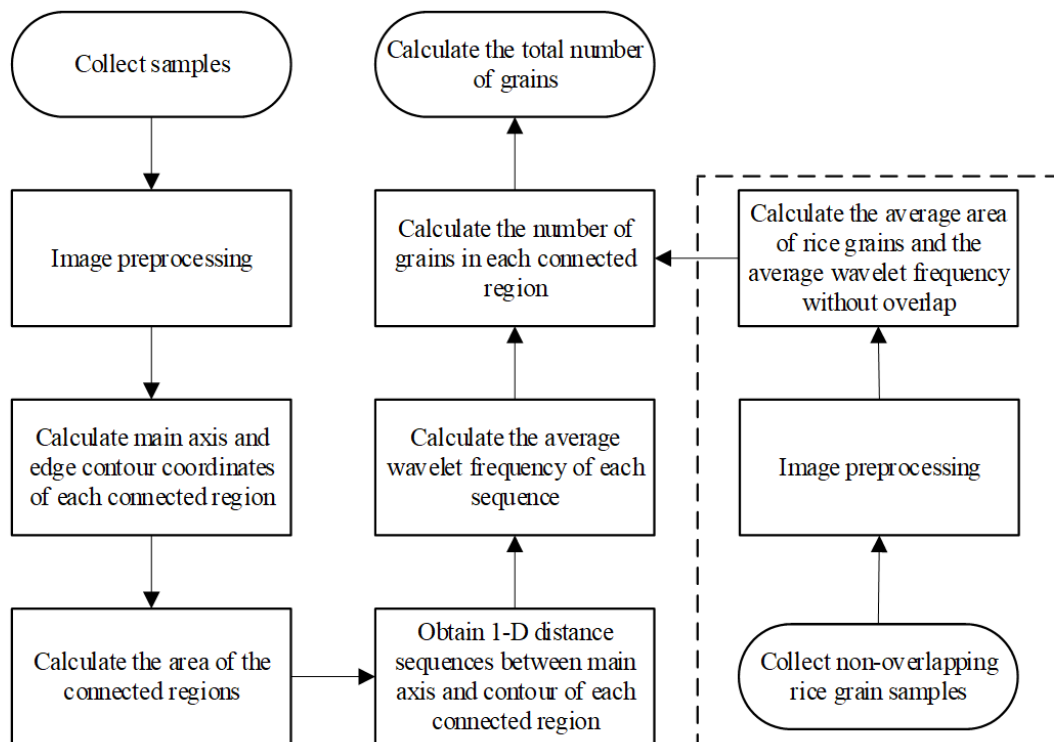


Figure 1. Flow chart of rice counting algorithm.

## 2.2. Preparation

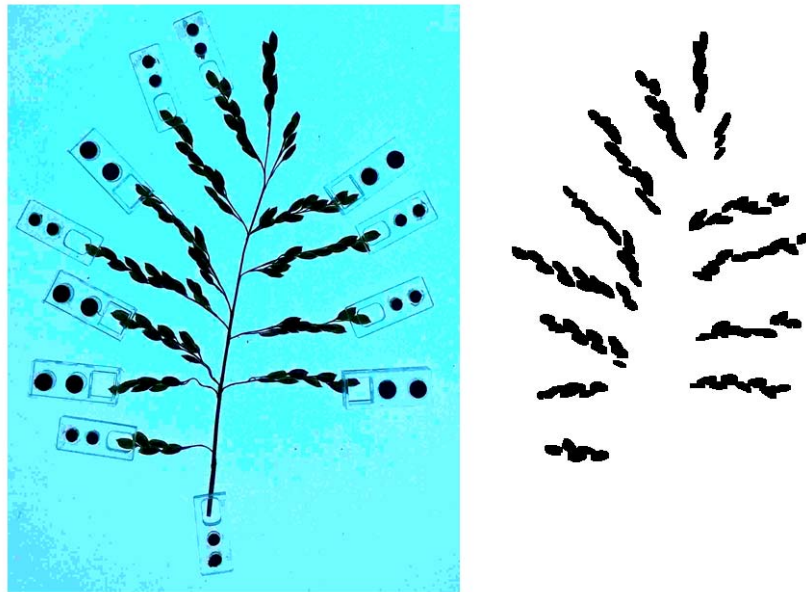
### 2.2.1. Materials

In this paper, 50 panicle samples from *Oryza sativa* ssp. *japonica* were collected from Jiangsu, China.

In addition, to verify the robustness of our method, we experimented with 5 panicle samples collected from rice recombinant inbred lines (RILs). This RIL population was derived by single seed descents from a cross between *O. sativa* spp. *indica* variety JP69 and *O. sativa* ssp. *japonica* variety Jiaoyuan5A, which show high yield heterosis in hybridized combination and are widely planted in Shanghai, China [22]. The plants were planted in a paddy field in Minhang (31.03° N, 121.45° E), Shanghai, during the summer season in 2017.

### 2.2.2. Image Preprocessing

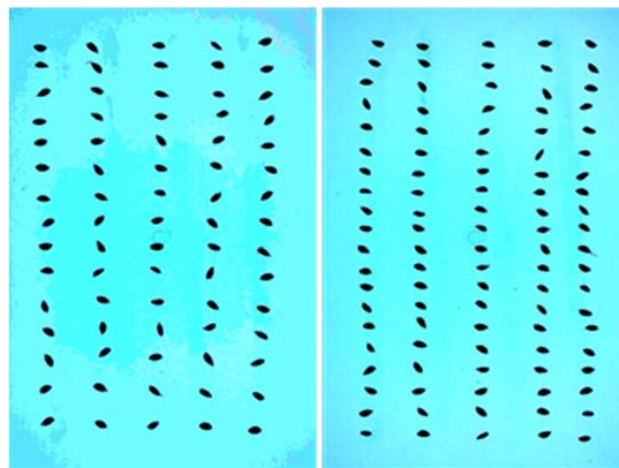
It is necessary to preprocess the image to obtain the morphological characteristics of the rice panicle image. Preprocessing operations include binarization, circle detection, corrosion and expansion and hole filling. Binarization could divide the rice panicle from the background, and circle detection was used to remove the black round magnets in the image. Corrosion and expansion operations were used to remove the stems of the rice panicle in the image. In this paper, the image processing environment was based on MATLAB R2016b (The MathWorks Inc., Natick, MA, USA). The preprocessed binary image is shown in Figure 2, which includes 23 connected regions.



**Figure 2.** Image preprocessing.

### 2.2.3. Calculating the Average Area of Rice Grains

Firstly, we randomly selected 175 rice grains and placed them neatly in the Seed & Panicle Phenotyping Instrument. Secondly, a picture of these rice grains was obtained, and we transformed it into a binary image using the Otsu method. Then, the non-target objects in the image were removed. Finally, we calculated the average number of pixels in the 175 connected regions, which was the average area of one rice grain. In our experiment, the average number of pixels of rice grains was  $m_0 = 2670.27$ . The sample image is shown in Figure 3.



**Figure 3.** Sample image used to calculate the average area of each rice grain.

## 2.3. Counting Algorithm with a Prior Correction Wavelet Model

### 2.3.1. Wavelet Frequency Ratio Method

First, the area of each connected region was calculated, and then it was divided by the average area of rice grains to get estimated value of grains in each connected region. Secondly, the average wavelet frequency of the edge contour sequence in each connected region was calculated, and the correction coefficient was obtained by dividing it by the average wavelet frequency without overlap.

Then, the estimate values were multiplied by the correction coefficient to obtain the corrected values. When the estimated value was less than 1, the estimated value was taken as 1, and the correction coefficient was also taken as 1. As such, we could obtain the number of rice grains in each connected region, and after summing them up we could obtain the total number of rice grains in the image.

### 2.3.2. Calculating the Main Axis of the Connected Regions

To obtain the fluctuation characteristics of each connected region’s edge of rice panicles by wavelet analysis, the 2-D coordinates of these edges needed to be transformed into a one-dimensional sequence. A workable way was to calculate the main axis  $l^a$  of each connected region  $a$ , and then calculated the projection of the edge coordinates on the main axis. Principal component analysis (PCA) was used to calculate the main axis  $l^a$ . Let the coordinates of the edge of each connected region  $a$  be  $X^a = \{x_1^a, x_2^a, \dots, x_m^a\}$ . The covariance matrix  $A$  is:

$$A = \frac{1}{m}(X^a - X(X^a))^T(X^a - X(X^a)) \tag{1}$$

The eigenvalue decomposition of the covariance matrix yields eigenvalues and eigenvectors:

$$A = Q\Sigma Q^{-1} \tag{2}$$

where  $Q$  is eigenvector matrix of  $A$  [23]. The first column of the matrix  $Q$  is the maximum variance projection direction of the edge coordinate in the connected region  $a$ , which is the direction of  $l^a$ . Then, the center point  $(x_c^a, y_c^a)$  of the region  $a$  can be calculated. So, the main axis  $l^a$  is a straight line passing through the center point  $(x_c^a, y_c^a)$ , whose direction is the first column of the matrix  $Q$ . Therefore, the expression of  $l^a$  is

$$Q_1(1) \times (y - y_c^a) = Q_1(2) \times (x - x_c^a) \tag{3}$$

where  $Q_1$  is the first column of the matrix  $Q$ . According to the above flow, the calculated main axis of each connected region is as shown in Figure 4.

We denoted the point coordinates of the contour in region  $a$  as  $X^a = \{(x_1^a, y_1^a), (x_2^a, y_2^a), \dots, (x_m^a, y_m^a)\}$ , and the distance from  $X^a$  to the corresponding main axis was denoted as  $s^a$ .



**Figure 4.** Main axis of each connected region.

### 2.3.3. A Prior Edge Average Wavelet Frequency Model without Overlapping

An ideal scenario without grain occlusions was first investigated to derive a prior model and a baseline frequency that could be used to describe and quantify the occlusion among grains in experimental samples.

An experiment was conducted to measure the average frequency of the edge of rice grains in a sample in which the grains do not overlap. First, we put 10 rice grains in a row without overlap, which is shown in Figure 5, and apply Otsu method to the image to obtain the binary image. Then, we calculated the main axis of the connected region, which is shown in Figure 6.

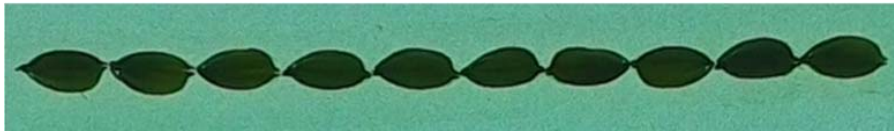


Figure 5. Image of rice grains without overlap.

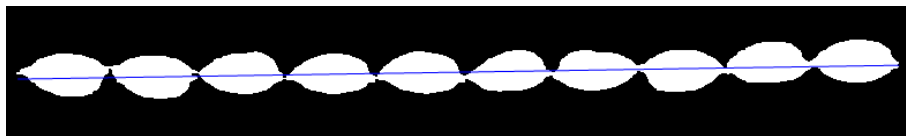


Figure 6. Binary image of rice grains without overlap and its main axis.

According to the main axis  $l^a$  and the edge coordinate sequence  $X^a$ , the distance from each edge point to  $l^a$  was calculated, which constituted a 1-D sequence  $s^a$ . The wave chart of  $s^a$  is shown in Figure 7. 1-D continuous wavelet transformation with 3rd order symmetric 'Morlet' wavelet was applied to  $s^a$ , whose sampling frequency was defined by 1000 Hz. The time-frequency diagram of  $s^a$  is shown in Figure 8. A wavelet coefficient matrix  $W(t, f)$  was obtained after continuous wavelet transformation to  $s^a$ . It could be observed that in the absence of overlap, the fluctuation frequency of the edge of rice grains was between 8 Hz and 16 Hz in our experimental conditions. Therefore, the wavelet reconstruction was performed on the sequence from 8 Hz to 16 Hz to obtain the main fluctuation information, and the reconstruction sequence was denoted as  $s'$ . The result is shown in Figure 7.

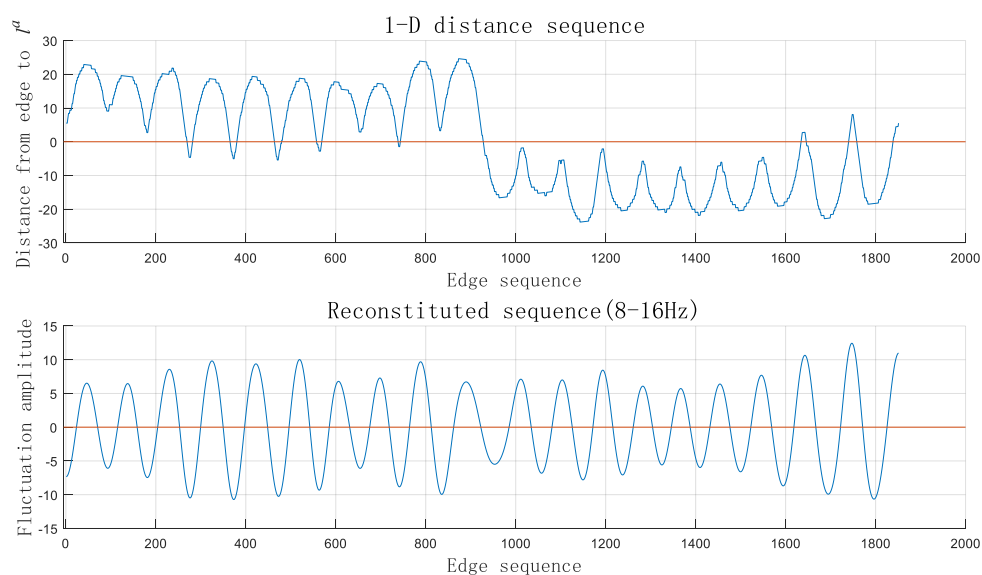
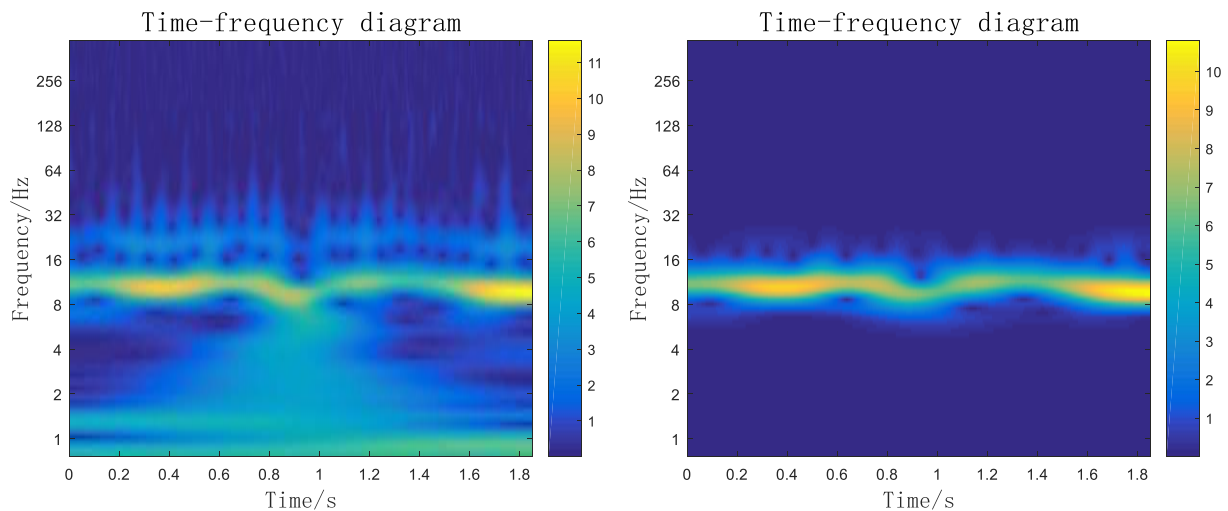


Figure 7. The waveform of the contour sequence and waveform after 8–16 Hz reconstruction.



**Figure 8.** Time-frequency image of sequences and TF image after wavelet reconstruction.

Continuous wavelet transformation was applied to  $s'$ , and we could obtain the new wavelet coefficient matrix  $W'(t, f)$  and corresponding frequency sequence  $F'$ . Thus, the average frequency of the edge of rice grains could be calculated by  $W'(t, f)$  and  $F'$ . First, the maximum frequency amplitude  $M(t)$  at each time  $t$  in  $W'(t, f)$  and the corresponding frequency  $F'_{idx}$  of  $M(t)$  could be found. We then defined the formulas below.

$$M(t) = \max W'(t, f) \quad (4)$$

$$f_0 = \frac{M^T F'_{idx}}{\sum_{t=1}^m M(t)} \quad (5)$$

According to the formulas above and  $s'$ , the average frequency of the edge of rice grains  $f_0$  was 10.5595 Hz.

#### 2.4. Hybrid On-Panicle Grain Counting Algorithm

In detail, after pre-processing the sample images, we could obtain the 1-D distance sequence of each connected region. We denoted it as  $s^a, a = 1, 2, \dots, P$ , where  $P$  was the number of connected regions in the image. 1-D continuous wavelet transformation was applied to each  $s^a$ . Considering the frequency of the edge sequence of rice grains with overlap will be larger than the frequency without overlap, we could take  $f_0$  to 64 Hz as the wavelet reconstruction range. The new wavelet reconstructed sequence  $s'^a$  was obtained. 1-D continuous wavelet transformation was applied to each  $s'^a$ , then we could obtain the new wavelet coefficient matrix  $W'(t, f)$  and corresponding frequency sequence  $F'$ . According to Formulas (4) and (5), we could obtain the average frequency  $f'^a$  of the edge of rice grains in the connected region  $a$ . Therefore, the hybrid model calculating the number of rice grains in the connected region  $a$  is

$$n(a) = \frac{m(a)}{m_0} \times \frac{f'^a}{f_0} \quad (6)$$

where,  $m_0$  is the average area of rice grains,  $m(a)$  is the area of the connected region of  $a$ , the first part of the right side is the area-based coarse estimation model, and the second part of the right side is the edge-based correction model. So, the total number of rice grains in the image is

$$N = \sum_{a=1}^P n(a) \quad (7)$$



### 3. Results

#### 3.1. Detailed Experimental Results of Two Samples

Figure 9 depicts two randomly selected samples, and in Figure 10 the panicle stems have been removed by image preprocessing. Sample A contains 23 connected regions and sample B contains 38 connected regions. The calculation results of these two samples are shown in Tables 1 and 2. For most regions, the wavelet frequency ratio is greater than 1, which is reasonable because compared with the edge contour of the non-overlapping rice grains, the edge contour frequency of the overlapping rice grains is greater. However, in a few cases, the connected region contains few grains and the length of the edge contour sequences is short, so the low frequency resolution of the sequence results in a less than 1 wavelet frequency ratio.

In the regions in which the counting result using the average area method is less than 1, the wavelet frequency ratio is not considered because of the lower frequency resolution of these regions, and the corrected number is defined as 1 in those regions. Tables 1 and 2 shows the estimated number from the coarse model and the corrected number from the hybrid model. The actual number of rice grains of sample A based on manual counting is 107, and the calculation number using wavelet frequency ratio method is 106. Compared to the coarse area algorithm without a correction model, the computation accuracy is improved from 73% to 99% by a percentage of 26%. In the sample B, the actual number is 129 and the calculation number is 121, with the computation accuracy improved from 71% to 94% by a percentage of 23%.

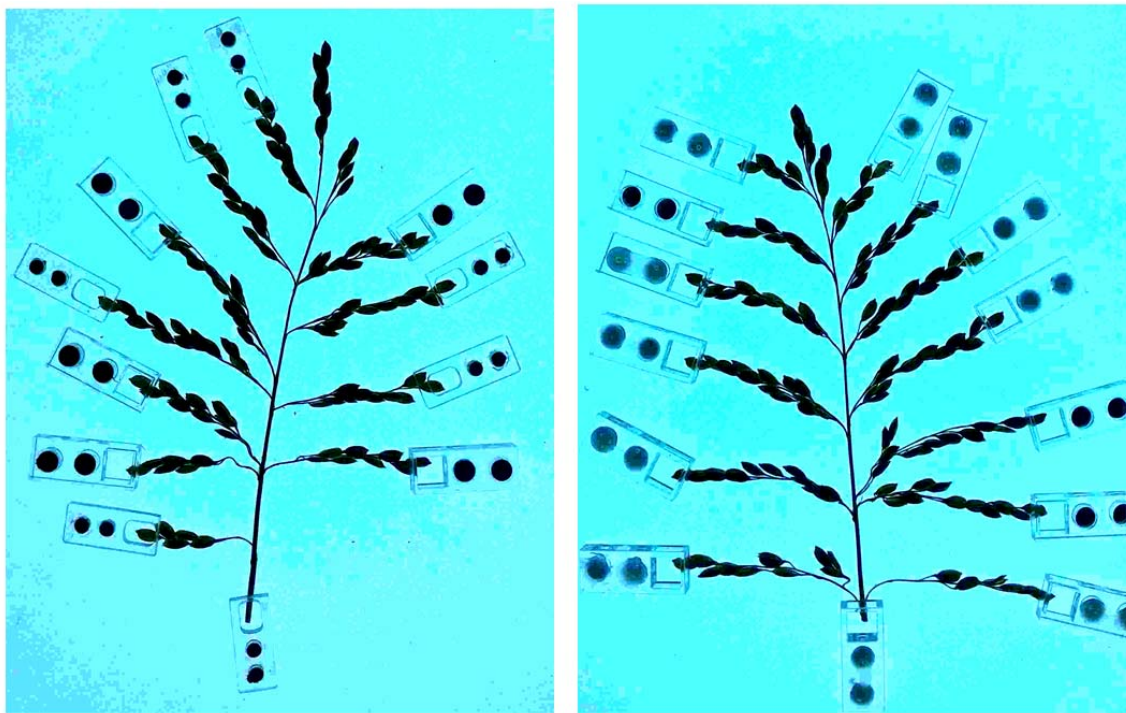


Figure 9. Original pictures of sample A (left) and sample B (right).

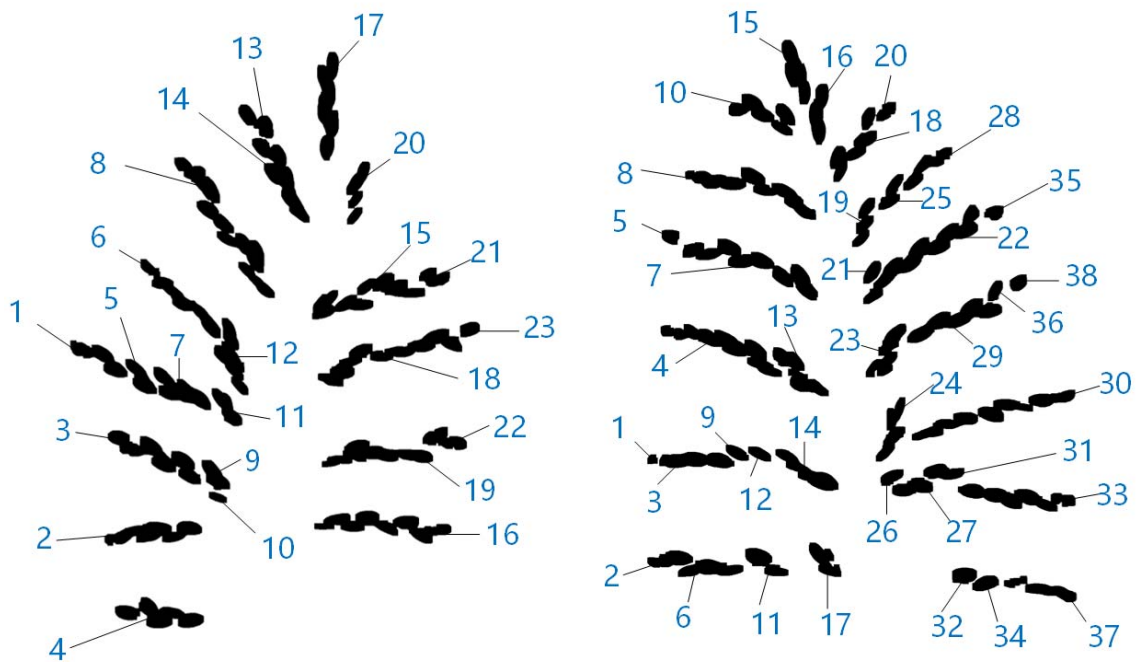


Figure 10. All connected regions of sample A (left) and sample B (right).

Table 1. Results for sample A \*.

Region No.	Wavelet Frequency Ratio	Estimated Number	Corrected Number	Region No.	Wavelet Frequency Ratio	Estimated Number	Corrected Number
1	1.255	2.500	3.139	13	1.299	1.760	2.286
2	1.299	4.122	5.355	14	1.257	4.681	5.884
3	1.289	5.122	6.602	15	1.440	5.634	8.111
4	1.369	4.150	5.680	16	1.405	5.536	7.780
5	1.372	1.621	2.225	17	1.199	4.770	5.721
6	1.208	3.679	4.445	18	1.493	6.347	9.474
7	1.395	3.215	4.484	19	1.490	4.321	6.438
8	1.330	7.871	10.471	20	1.256	2.171	2.727
9	1.692	1.472	2.491	21	1.498	1.233	1.847
10	1.082	0.412	1.000	22	1.268	1.818	2.305
11	1.211	1.606	1.946	23	1.379	0.722	1.000
12	1.221	3.463	4.229				

\* The actual number of grains in sample A based on manual counting is 107 while the summed result from average area method is 78, the calculated result from P-Trap is 91, and the summed result of the proposed wavelet frequency ratio method is 106. Computational accuracy is improved from 73% and 85% to 99%.

Table 2. Results for sample B \*.

Region No.	Wavelet Frequency Ratio	Estimated Number	Corrected Number	Region No.	Wavelet Frequency Ratio	Estimated Number	Corrected Number
1	1.438	0.258	1.000	21	0.870	0.801	1.000
2	1.252	1.637	2.049	22	1.080	6.651	7.183
3	1.310	2.840	3.721	23	1.534	2.760	4.234
4	1.358	5.596	7.599	24	1.356	2.431	3.295
5	0.932	0.647	1.000	25	1.098	1.347	1.479
6	1.637	2.266	3.709	26	1.383	0.770	1.066
7	1.281	6.584	8.437	27	1.178	1.578	1.859
8	1.445	5.692	8.224	28	1.222	2.101	2.568
9	0.914	0.710	1.000	29	0.993	4.355	4.326

Table 2. Cont.

Region No.	Wavelet Frequency Ratio	Estimated Number	Corrected Number	Region No.	Wavelet Frequency Ratio	Estimated Number	Corrected Number
10	1.301	3.645	4.744	30	1.405	5.769	8.106
11	1.015	1.819	1.847	31	1.245	1.418	1.766
12	0.948	0.671	1.000	32	1.351	1.012	1.367
13	1.639	3.120	5.115	33	1.302	4.676	6.087
14	1.581	2.679	4.236	34	1.348	0.929	1.253
15	1.179	2.933	3.457	35	0.870	0.667	1.000
16	1.132	2.326	2.633	36	1.116	0.655	1.000
17	1.618	1.765	2.857	37	1.243	2.159	2.684
18	1.162	2.810	3.266	38	0.987	0.670	1.000
19	1.419	1.790	2.540				
20	1.160	1.495	1.734				

\* The actual number of grains in sample B based on manual counting is 129 while the summed result from average area method is 92, the calculated result from P-Trap is 121, and the summed result of the proposed wavelet frequency ratio method is 121. The wavelet frequency ratio method and P-trap gave similar results with an accuracy of 94%, both of which were more accurate than the average area method whose accuracy is 71%.

### 3.2. Experimental Results of 55 Samples

Based on the wavelet frequency ratio method, we counted rice grains on 50 separate rice panicles from a single cultivar collected in Jiangsu (Table 3). Using our method, the average count accuracy for each panicle is  $94 \pm 4\%$ , compared to  $86 \pm 10\%$  for P-trap software. Additional 5 samples were used to validate the effectiveness of the proposed method (Table 4), in which the average count accuracy is 93%.

Table 3. Results for 50 panicle samples from a single *O. sativa* ssp. *japonica* cultivar \*.

No.	Actual Number	Calc. <sup>†</sup> Result	Acc. <sup>††</sup>	P-Trap Result	P-Trap Acc.	No.	Actual Number	Calc. <sup>†</sup> Result	Acc. <sup>††</sup>	P-Trap Result	P-Trap Acc.
1	57	48	84%	59	96%	26	88	81	92%	78	89%
2	75	79	95%	74	99%	27	87	76	87%	91	95%
3	74	74	100%	63	85%	28	113	109	96%	105	93%
4	108	96	89%	95	88%	29	99	89	90%	88	89%
5	76	68	89%	62	82%	30	105	97	92%	99	94%
6	129	121	94%	121	94%	31	77	70	91%	61	79%
7	61	58	95%	43	70%	32	79	76	96%	72	91%
8	76	72	95%	81	93%	33	120	121	99%	89	74%
9	79	75	95%	73	92%	34	110	106	96%	99	90%
10	105	104	99%	85	81%	35	81	71	88%	64	79%
11	74	76	97%	50	68%	36	107	106	99%	91	85%
12	112	108	96%	98	88%	37	108	98	91%	106	98%
13	87	76	87%	77	89%	38	87	78	90%	89	98%
14	113	122	92%	113	100%	39	109	120	90%	72	66%
15	67	61	91%	41	61%	40	64	63	98%	65	98%
16	97	97	100%	81	84%	41	69	68	99%	57	83%
17	119	120	99%	79	66%	42	119	111	93%	96	81%
18	126	117	93%	100	79%	43	112	111	99%	96	86%
19	102	100	98%	100	98%	44	110	104	95%	72	65%
20	134	147	90%	124	93%	45	103	97	94%	87	84%
21	83	69	83%	81	98%	46	109	108	99%	90	83%
22	95	96	99%	79	83%	47	142	135	95%	95	67%
23	121	112	93%	92	76%	48	51	53	96%	54	94%
24	80	81	99%	81	99%	49	59	57	97%	64	92%
25	91	78	86%	84	92%	50	90	85	94%	81	90%

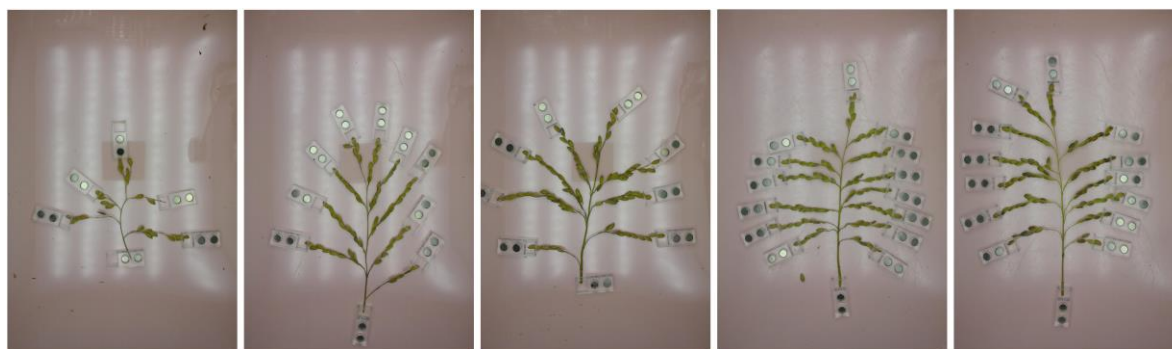
\* The average counting accuracy of proposed wavelet frequency ratio method is  $94 \pm 4\%$  while the average counting accuracy of P-Trap is  $86 \pm 10\%$ . <sup>†</sup> Calculation. <sup>††</sup> Accuracy.

To verify the robustness of our method, we also experimented with 5 panicles derived from a rice RIL population. This population had extensive diversity in panicle morphology because it was generated by single seed descents from a cross between *O. sativa* ssp. *indica* variety JP69 and *O. sativa* ssp. *japonica* variety Jiaoyuan5A, and these two varieties have dramatically different panicle types. These 5 samples are shown in Figure 11, and counting results for these samples are reported in Table 4. The average count accuracy for the wavelet frequency method was 93%.

**Table 4.** Results for 5 panicle samples from a RIL population \*.

No.	Actual Number	Calculation Result	Acc.
1	24	28	83%
2	70	69	99%
3	87	90	97%
4	111	102	92%
5	103	99	96%

\* The average counting accuracy is 93%.



**Figure 11.** 5 RIL population samples.

#### 4. Discussion

This study aimed to tackle the problem that overlap between the rice grains on a panicle interferes with the accurate automated counting of rice grains through image analysis. To address this problem, we proposed a wavelet analysis method of edge contour of the rice panicle. On the basis of the average area method, we used the wavelet frequency ratio between actual sequences and non-overlapping sequences as a prior correction factor to correct the number of rice grains. This method makes full use of the fluctuation information of the edge contour of each panicle to indicate the degree of overlap between the rice grains, which ameliorates the situation of low counting accuracy caused by the complex occlusion and overlapping scenarios in the image.

We randomly selected two samples and applied the method we proposed to them to calculate the number of rice grains. Compared with the count result of the P-Trap software [21], the result obtained by our method was comparable for sample B and more accurate for sample A. We propose that the accuracy of the P-Trap software was lower for sample A than for sample B because of a difference in the morphology of the two samples. The degree of overlap among the rice grains of sample A was more severe than the degree of overlap in sample B, and this is reflected in the higher average wavelet frequency ratio of the connected regions in sample A (1.335) compared to the average wavelet frequency ratio in sample B (1.245). An increase in the degree of overlap results a reduction in the area of the panicle, and so estimates of grain numbers that are based on the panicle area, such as the results obtained by the P-Trap software, are likely to be less accurate for densely-packed panicles than for panicles with low overlap between the rice grains.

We also compared the accuracy of our method to P-Trap software for 50 additional samples. Our method had a higher average counting accuracy and a lower standard deviation of accuracy compared to that of P-Trap. In many samples, the calculation result of the P-Trap software was comparable to our method. However, for approximately 10% of the samples, P-Trap dramatically underestimated the number of grains (sample numbers 15, 17, 39, 44 and 47).

The method we proposed in this paper based on wavelet analysis of the edge of panicles has the advantage that it makes use of the fluctuation information of the panicle edge to represent the degree of overlap between rice grains, which is used to correct the number of rice grains. This correction makes our method perform relatively well in samples with more severe overlap.

Our results reveal that area and edge contour wavelet frequency can effectively represent the number of rice grains in an image and improve the counting accuracy of image analysis software. The improvement of counting accuracy helps to estimate grain yields, which makes our improved software a valuable tool for rice breeders.

## 5. Conclusions

In summary, we proposed that there is a correlation between the number of grains in a rice panicle and the fluctuation frequency of its edge contours, and based on this assumption we developed a prior edge wavelet correction model for counting rice grains in a panicle based on image processing and wavelet analysis. In experiments with more than 50 panicle samples, the accuracy of this model was equal to or greater than 93%. This model may be particularly helpful when the degree of overlap between rice grains is relatively high.

**Author Contributions:** L.G., K.L. and T.W. conducted mathematical modeling and article writing. K.L. also completed the software development and experimental verification. C.L. and D.Z. supervised the whole project. Z.Y. and J.H. conducted the experimental verification.

**Funding:** This study was supported by the Natural Scientific Foundation of China under Grant NO. 51775333 and the Scientific Plan Project of Shanghai Science and Technology Committee under Grant NO. 16391903102.

**Acknowledgments:** The help of Zhihong Ma in sample collection and software development is gratefully acknowledged.

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

## References

1. Yang, W.; Duan, L.; Chen, G.; Xiong, L.; Liu, Q. Plant phenomics and high-throughput phenotyping: Accelerating rice functional genomics using multidisciplinary technologies. *Curr. Opin. Plant Biol.* **2013**, *16*, 180–187. [CrossRef] [PubMed]
2. Zhang, Q. Strategies for developing green super rice. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16402–16409. [CrossRef] [PubMed]
3. Mark, T.; Peter, L. Breeding Technologies to Increase Crop Production in a Changing World. *Science* **2010**, *327*, 818–822. [CrossRef]
4. Sun, J.; Rutkoski, J.E.; Poland, J.A. Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *Plant Genome* **2017**, *10*. [CrossRef] [PubMed]
5. Pratap, A.; Tomar, R.; Kumar, J. High-throughput plant phenotyping platforms. In *Phenomics in Crop Plants: Trends, Options and Limitations*; Jitendra, K., Aditya, P., Eds.; Springer: New Delhi, India, 2015; pp. 285–296. ISBN 978-81-322-2225-5.
6. Golzarian, M.R.; Frick, R.A.; Rajendran, K.; Berger, B.; Roy, S.; Tester, M.; Lun, D.S. Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods* **2011**, *7*, 2. [CrossRef] [PubMed]
7. PhenoFab: Digital Phenotyping. Available online: <http://www.keygene.com/products-tech/digital-phenotyping/> (accessed on 17 January 2018).
8. Tisne, S.; Serrand, Y.; Bach, L.; Gilbault, E.; Ben Ameer, R.; Balasse, H.; Voisin, R.; Bouchez, D.; Durand-Tardif, M.; Guerche, P.; et al. Phenoscope: An automated large-scale phenotyping platform offering high spatial homogeneity. *Plant J.* **2013**, *74*, 534–544. [CrossRef] [PubMed]

9. Pereyra-Irujo, G.A.; Gasco, E.D.; Peirone, L.S.; Aguirrezábal, L.A. GlyPh: A low-cost platform for phenotyping plant growth and water use. *Funct. Plant Biol.* **2012**, *39*, 905–913. [[CrossRef](#)]
10. Sadok, W.; Naudin, P.; Boussuge, B.; Muller, B.; Welcker, C.; Tardieu, F. Leaf growth rate per unit thermal time follows QTL-dependent daily patterns in hundreds of maize lines under naturally fluctuating conditions. *Plant Cell Environ.* **2007**, *30*, 135–146. [[CrossRef](#)] [[PubMed](#)]
11. Nagel, K.A.; Putz, A.; Gilmer, F.; Heinz, K.; Fischbach, A.; Pfeifer, J.; Faget, M.; Blossfeld, S.; Ernst, M.; Dimaki, C.; et al. GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons. *Funct. Plant Biol.* **2012**, *39*, 891–904. [[CrossRef](#)]
12. Jansen, M.; Gilmer, F.; Biskup, B.; Nagel, K.A.; Rascher, U.; Fischbach, A.; Briem, S.; Dreissen, G.; Tittmann, S.; Braun, S.; et al. Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in Arabidopsis thaliana and other rosette plants. *Funct. Plant Biol.* **2009**, *36*, 902–914. [[CrossRef](#)]
13. Walter, A.; Scharr, H.; Gilmer, F.; Zierer, R.; Nagel, K.A.; Ernst, M.; Wiese, A.; Virnich, O.; Christ, M.M.; Uhlig, B.; et al. Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: A setup and procedure designed for rapid optical phenotyping of different plant species. *New Phytol.* **2007**, *174*, 447–455. [[CrossRef](#)] [[PubMed](#)]
14. Granier, C.; Aguirrezabal, L.; Chenu, K.; Cookson, S.J.; Dauzat, M.; Hamard, P.; Thioux, J.J.; Rolland, G.; Bouchier-Combaud, S.; Lebaudy, A.; et al. PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol.* **2006**, *169*, 623–635. [[CrossRef](#)] [[PubMed](#)]
15. Zhao, S.; Gu, J.; Zhao, Y.; Hassan, M.; Li, Y.; Ding, W. A method for estimating spikelet number per panicle: Integrating image analysis and a 5-point calibration model. *Sci. Rep.* **2015**, *5*, 16241. [[CrossRef](#)] [[PubMed](#)]
16. Lingfeng, D.; Wanneng, Y.; Chenglong, H.; Qian, L. A novel machine-vision-based facility for the automatic evaluation of yield-related traits in rice. *Plant Methods* **2011**, *7*, 44. [[CrossRef](#)]
17. Crowell, S.; Korniliev, P.; Falcão, A.; Ismail, A.; Gregorio, G.; Mezey, J.; McCouch, S. Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nat. Commun.* **2016**, *7*, 10527. [[CrossRef](#)] [[PubMed](#)]
18. Lobet, G.; Draye, X.; Périlleux, C. An online database for plant image analysis software tools. *Plant Methods* **2013**, *9*, 38. [[CrossRef](#)] [[PubMed](#)]
19. Lou, J.; Zhou, M.; Li, Q.; Yuan, C.; Liu, H. An automatic red blood cell counting method based on spectral images. In Proceedings of the 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Datong, China, 15–17 October 2016.
20. Huang, X.; Jiang, Y.; Xu, H.; Liu, X.; Hou, H.W.; Yan, M.; Yu, H. A convolutional neural network based single-frame super-resolution for lensless blood cell counting. In Proceedings of the 2016 IEEE Biomedical Circuits and Systems Conference (BioCAS), Shanghai, China, 17–19 October 2016.
21. Faroq, A.T.; Adam, H.; Dos Anjos, A.; Lorieux, M.; Larmande, P.; Ghesquière, A.; Jouannic, S.; Shahbazkia, H.R. P-TRAP: A panicle trait phenotyping tool. *BMC Plant Biol.* **2013**, *13*, 122. [[CrossRef](#)]
22. Fan, W.; Zong, J.; Luo, Z.; Chen, M.; Zhao, X.; Zhang, D.; Qi, Y.; Yuan, Z. Development of a RAD-Seq based DNA polymorphism identification software, AgroMarker Finder, and its application in rice marker-assisted breeding. *PLoS ONE* **2016**, *11*, e0147187. [[CrossRef](#)] [[PubMed](#)]
23. Strang, G. *Introduction to Linear Algebra*, 4th ed.; Wellesley Cambridge Press: Wellesley, MA, USA, 1993; pp. 298–303. ISBN 978-0-9802327-1-4.

