Soil & Tillage Research 162 (2016) 8–17

Contents lists available at ScienceDirect

Soil & Tillage Research

journal homepage: www.elsevier.com/locate/still

Application of machine vision for classification of soil aggregate size

Fatemeh Rahimi Ajdadi[a], Yousef Abbaspour Gilandeh[a,*], Kaveh Mollazade[b], Reza PR Hasanzadeh[c]

[a] Department of Biosystems Engineering, Faculty of Agricultural Technology and Natural Resources, University of Mohaghegh Ardabili, 56199-11367 Ardabil, Iran
[b] Department of Biosystems Engineering, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran
[c] Department of Electrical Engineering, Faculty of Engineering, University of Guilan, 41996-13769 Rasht, Iran

ARTICLE INFO

Article history:
Received 26 October 2015
Received in revised form 14 April 2016
Accepted 16 April 2016
Available online 30 April 2016

Keywords:
Precision agriculture
Image processing
Mean weight diameter
Texture feature
Tilled soil

ABSTRACT

Tillage operations demand more than half of the total energy consumed in mechanized agriculture. Simultaneous measurement of tillage quality during the operation, would present the possibility of real time adjustments of the tillage tool parameters. The development of such a method would result in a desirable plough with the least possible running cost. On that basis, the purpose of this study was to develop an algorithm that supplies the potential of real-time measurement of tillage quality using image processing. Photography was performed at three camera heights and covering nine different sizes of soil aggregates. Textural information from tilled soil images was extracted by four methods, including first order statistics of image histogram, gray level co-occurrence matrix; gray level run length matrix and local binary pattern. A data mining procedure by CfsSubsetEval was used for feature selection. Networks with topology of 19-19-1, 14-22-1, and 17-20-1 neurons represented the best classification performance for photography heights of 60, 80, and 100 cm, respectively. The best overall accuracy of the ANN classifier was obtained from images taken at the height of 60 cm (72.04%). Results indicated that the present approach for estimating mean weight diameter up to about 35 mm had the best performance with an accuracy of over 80%. The technique suggested in this study is feasible for implementation in variable rate secondary tillage machines.

1. Introduction

Soil is the main source of production in agriculture. It provides a medium for plant establishment, seed germination, root growth and development. The physical and structural properties of soil have a direct effect on rooting and seed germination. Fundamentally, granular soil structure is the most desirable because it allows for good water infiltration, maintains water properly, increases air capacity, facilitates soil conditioning and reduces resistance to rooting.

The purpose of tillage operation is to enhance the soil's tilth. Tilth is the physical condition of soil in relation to its aggregate size. Proper sizes of aggregates cause a good air to moisture ratio and consequently, produce a better yield. The most desirable tillage condition is one whereby the size of aggregate around a seedbed is changed according to seed requirements. The creation of smaller aggregates causes energy losses due to unneeded tillage as well as extra tillage, in turn, increases the potential for soil compaction. On the other hand, in case of larger aggregates, re-tillage operation will be required. As a result, if the operator is informed about the sizes of aggregates during tillage, he can achieve desirable tillage quality by adjusting parameters such as tractor forward speed and tillage depth. Knowing that the tillage operation expends more than half of the energy consumption of agricultural productions, this serves to keep production costs low.

This serves to keep production costs low. Braunack and Dexter (1989a,b) illustrated the importance and efficacy of different aggregate sizes in a seedbed on plant emergence and growth. Assessing the quality of a tillage operation is accomplished by a degree of soil pulverization and is defined in terms of aggregate size (Spoor et al., 1976). Researchers generally use MWD[1] of aggregates as the most important criterion for defining the degree of soil pulverization. Currently, sieving soil through a set of standard multiple sieves is the method used to measure the MWD index. In this method, sieves are arranged on top of each other and they gradually decrease in size from top to bottom. Obviously, extensive sampling is required for application of this method on

* Corresponding author.
E-mail address: abbaspour@uma.ac.ir (Y.A. Gilandeh).

[1] mean weight diameter

http://dx.doi.org/10.1016/j.still.2016.04.012
0167-1987/

large-scale tillage quality and tilth. Furthermore, some factors such as aggregate failure during sampling and transportation to the laboratory, air drying and sieving can produce results that deviate from actual readings. Static measurement is another disadvantage of this method that hampers 'on the go' applications of. Indeed, precision agricultural approaches and spatial variability are not considered. However, using a dynamic approach can lead to the use of simultaneous measurements for 'on the go' mechanisms to improve soil pulverization. In addition, it selects optimal sowing conditions in terms of the best size distribution of soil aggregates.

Several research groups attempted to develop new methods that were not limited by the process of sieve analysis. For instance, Olsen (1992) used a horizontal mini-penetrometer with a diameter of 5 mm with sensitivity to force only at its probe tip. The result indicated that direct sensing of individual aggregate size was impossible by this method. However, it was capable to derive a parameter from the penetration force signal to determine overall coarseness of an aggregate bed. Some researchers have suggested image analysis as an alternative to the laborious sieving method to evaluate seedbed roughness. Application of image analysis has increased rapidly in recent years due to advances in technology such as digital cameras, higher resolution, faster processors, digital image capturing and improved storage capacity (Atkinson, 2008).

Following this progress, there have been many applications of image processing in other fields of agricultural research, but the technique has had little application in clod analysis, mainly because of the difficulty of accurate determination because of very low-level color variation substrate between soil and substrate. Additionally, the nature of soil and overlapping of the aggregates makes it difficult to discriminate aggregates and to determine diameter. Research on image processing falls into two major groups: one group in which imaging was done in a laboratory, and individual aggregates were spread on a white background (Campbell, 1979; Sandri et al., 1998). In the other group of research studies, imaging was done in the field without eliminating overlap between aggregates. (Stafford and Ambler, 1988, 1990; Bogrekci and Godwin, 2007). The latter method highlights the potential for application of machine vision for variable rate tillage machines.

Bogrekci and Godwin (2007) applied morphological methods of image analysis to estimate distribution of clod size. The report concluded that the technique of contrast enhancement had greater precision compared with other techniques. Also, the value of MWD obtained by their method was 21% larger than standard sieving method. One disadvantage of approaches based on morphological features is that computations lead to an MWD value that is higher than the actual one due to omission of small aggregates.

Itoh et al. (2008) developed an algorithm to measure aggregate size using rock fragments as experimental material; The relationship between aggregate size and textural features was expressed by multiple regression equations. It should be noted that measuring the size of a rock fragment does not incur problems such as crunching and overlapping.

In recent years, several field studies on clod detection by DEM[2] images have been conducted (Vannier et al., 2009; Taconet et al., 2010,2013; Chimi-Chiadjeu et al., 2012, 2014). These images are taken by stereo photogrammetry systems and the value of every pixel represents the elevation or altitude (z coordinate) (Chimi-Chiadjeu et al., 2014). In stereo photogrammetry technique, to develope 3D simulation of clods, two photographic images of the same spot are taken by combining two cameras looking vertically on the soil (Taconet and Ciarletti, 2007). The elevation of each pixel is calculated by the information about image resolution, photography height and control points. The main drawback of the DEM method is that it cannot be used in real time applications such as variable rate tillage machines.

According to the literature, most of the research on determining soil clod size using image processing has been done in the three decades from 1970 to 2000. Subsequently, there has been no adequate research published on assessing tillage quality that has the potential for real time application. In addition, a review of the literature shows that in most previous studies, morphological features have only been used for determinations of clod size distribution. However, only recognizable clods were involved in MWD calculations and smaller or overlapped aggregates were overlooked. Overlapping between adjacent aggregates caused lower values for MWD than the actual value.

In the present study, statistical textural features extracted from soil images were correlated with soil aggregate size. This method was advantageous in that all components of soil were involved in aggregate size determination. Hence, one objective of this study was to develop an approach that enables assessment of soil tillage quality by image processing based on textural features extracted from soil images coupled with an artificial neural networks (ANNs) classifier. Another objective was to find the optimum distance of the camera from the ground for image analysis.

2. Materials and methods

2.1. Site soil preparation

Field experiments were done in the educational field of the Technical and Vocational Training Centre, Guilan Province, Rasht City. The site has a temperate climate and annual rainfall of 1359 mm. Soil texture in the test field is silty clay. The results of soil textural analysis are shown in Table 1. The purpose of preparing the test site was to achieve nine different aggregate size distributions of up to about 100 mm (MWD < 100 mm). This range of aggregate size distribution is common for agricultural applications including primary and secondary tillage. To meet this objective, firstly the whole area was prepared using a mould board plough to the depth of 35 cm. It was then divided into nine plots. Plots were tilled with a rotivator in order to obtain nine different aggregates size classes. Nine different aggregate sizes were supplied with three different types of parameters related to tillage operation; these were forward speed, cover position and frequency of tillage operation. Four forward speeds were obtained by changing four forward gears from 1 to 4 in heavy gears. Motor revolution was set at 1500 rpm. Cover positioning consisted of two levels of cover positioning (including down and top). Frequency of tillage for all classes of tilled soil was once except for class 1, in which tillage operation was performed twice by the rotavator.

[2] Digital Elevation Model

Table 1
Soil properties of the test site.

Soil textural fraction	Mass percent
sand	12
silt	60
clay	28

Table 2
Combination of different parameters used to obtain nine different clod sizes.

Parameter type		Class number of aggregate size distribution								
		1	2	3	4	5	6	7	8	9
Forward speed ($km\,h^{-1}$)	1.78	×	×				×			
	2.68			×				×		
	3.57				×				×	
	5.2					×				×
Cover position	down	×	×	×	×	×				
	top						×	×	×	×
Frequency of rotavator	1		×	×	×	×	×	×	×	×
	2	×								

The combination of different parameters to obtain nine different aggregate sizes is represented in Table 2. Classes 1 and 9 indicate the finest and coarsest aggregates, respectively. Fig. 1 shows sample images of different classes of tillage quality.

After preparation of the nine plots, they were left for two days to let the soil clods dry out. This preparation was done to facilitate sieving and to reduce possible errors that could be caused from the clods sticking together. Thereafter, ten samples of each plot were taken randomly for sieving analysis with flat-screened sieves of sizes: 1.58, 3.18, 6.35, 19.05, 38.1, 44.45, 63.5, 88.9, and 114.3 mm. The reciprocating motion of the sieves was provided by an electric shaker with amplitude of motion of 2.2 mm for 15 min. The MWD of the soil samples was calculated as follows:

$$D = \Sigma_{i=0}^{n} \overline{x_i} \times w_i \quad (1)$$

where:

D: Diameter of a clod in millimeters.

$\overline{x_i}$: Mean diameter of each size fraction in millimeters.

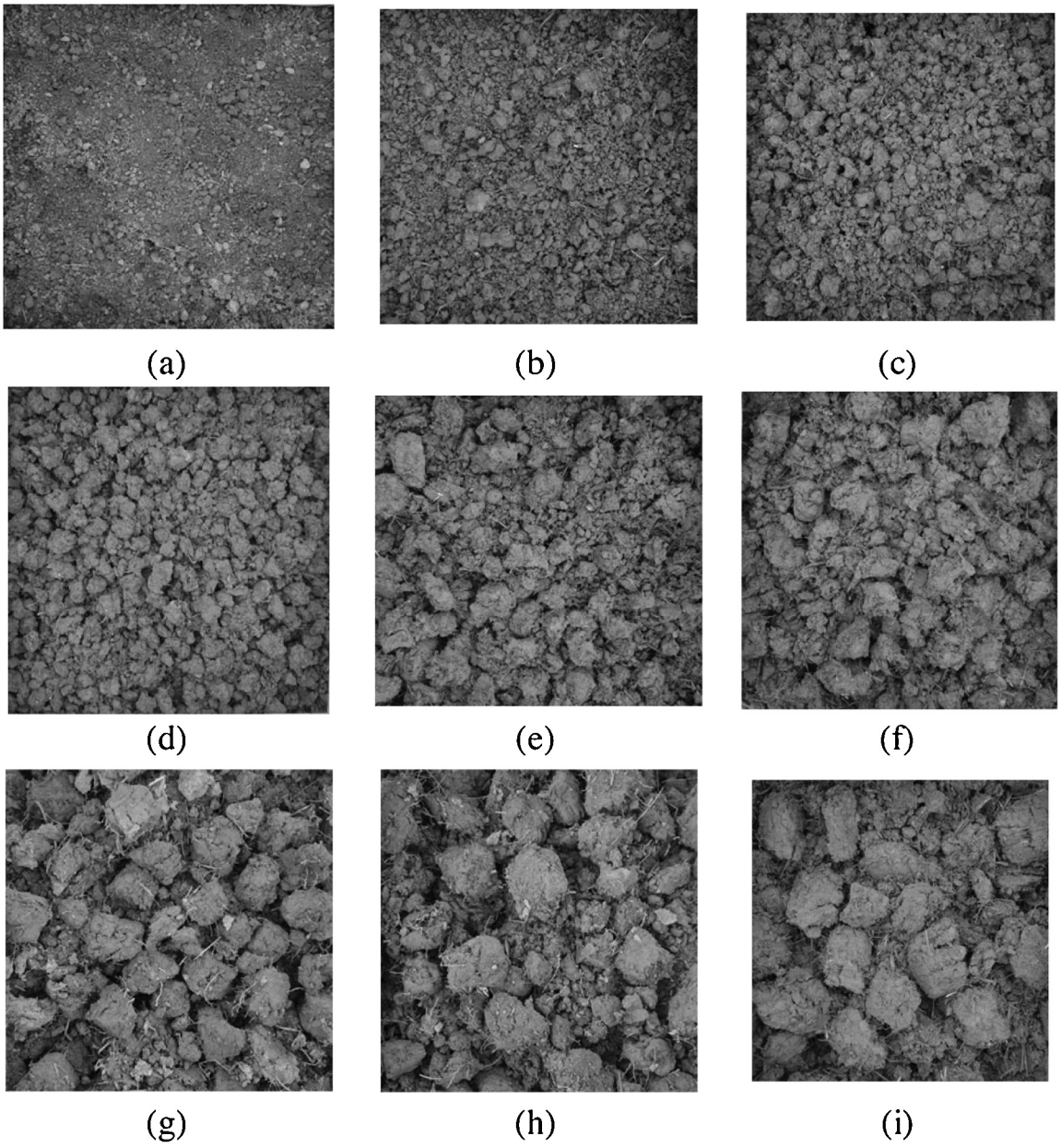

Fig. 1. The view of different classes of aggregate size distribution in the test site. Images of *a* to *i* represent finest to coarsest tilled soil, respectively.

Table 3
The MWD for each class of aggregate size obtained by the standard sieve (millimeter).

Class No.	Min	Max	Mean	Standard deviation
1	2.0	4.3	3.38	0.81
2	10.2	15.9	12.87	1.75
3	16.3	25.0	19.83	2.66
4	23.7	37.6	29.57	4.31
5	34.1	50.4	40.62	4.74
6	43.9	62.9	51.79	6.30
7	49.5	65.3	57.45	6.07
8	60.0	85.0	70.29	6.95
9	85.0	109.0	93.36	6.90

w_i: Proportion of total sample weight occurring in the corresponding size fraction.

n: Number of sieves.

Table 3 shows the MWDs for each class of aggregate sizes obtained by the standard method using sieves.

2.2. Image acquisition

Images were taken with a digital camcorder (Canon PC1586, Canon Inc., China, 12.1 Mega pixels, Canon Zoom Lens 4X, 5.0–20.0 mm 1:2.8-5.9). All images were acquired using the ambient light source in the field. First, a sample image was taken together with a wooden framed grid so that the frame was placed on the surface of the soil and then the image was recorded. This framed grid was used at the image registration stage to select the best transformation for sample images in order to reduce the effect of distortion. The frame was also applied for the subsequent stages of photography after determination of the best transformation.

As mentioned, the secondary objective of this study was to find the best height of photography for making determinations of soil tillage quality. Accordingly, photography was performed at three levels of camera height from the ground surface; 60, 80, and 100 cm. The camera was fixed perpendicularly on the top of a footstool that allowed adjustment of camera height in the range of 60–100 cm.

For each class of aggregate size, about 150 images were taken (the number of images in class 2 was less than 150).

2.3. Image preprocessing

The image registration technique was applied to eliminate distortion created in the images.

Table 4
Available texture features for each image measured by statistical techniques.

1st order statistics of image histogram and local binary pattern		Gray level co-occurrence matrix		Gray level run length matrix	
Feature name	Formula[a]	Feature name	Formula[b]	Feature name	Formula[c]
Mean gray level	$\mu = \Sigma_{i=0}^{L-1} z_i p(z_i)$	Contrast	$\sum_{i,j} \lvert i-j\rvert^2 p(i,j)$	Short run emphasis (SRE)	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{r(i,j)}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$
Standard deviation	$\sigma = \left(\Sigma_{i=0}^{L-1}(z_i-\mu)^2 p(z_i)\right)^{1/2}$	Correlation	$\sum_{i,j}\frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j}$	Long run emphasis (LRE)	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} j^2 r(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$
Smoothness	$1 - 1/(1+\sigma^2)$	Uniformity	$\sum_{i,j} p(i,j)^2$	Gray level non-uniformity (GLNU)	$\frac{\sum_{i=1}^{N_g}\left(\sum_{j=1}^{N_r} r(i,j)\right)^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$
Skewness	$^1/_{\sigma^3}\Sigma_{i=0}^{L-1}(z_i-\mu)^3 p(z_i)$	Homogeneity	$\sum_{i,j}\frac{p(i,j)}{1+\lvert i-j\rvert}$	Run length non-uniformity (RLNU)	$\frac{\sum_{i=1}^{N_r}\left(\sum_{j=1}^{N_g} r(i,j)\right)^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$
Uniformity (energy)	$\Sigma_{i=0}^{L-1} p^2(z_i)$	Entropy	$-\sum_{i,j} p(i,j) log p(i,j)$	Run percentage (RP)	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}{n_p}$
Entropy	$-\Sigma_{i=0}^{L-1} p(z_i) log_2 p(z_i)$	Maximum of probability	$Max(p(i,j))$	Low gray level run emphasis (LGRE)	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{r(i,j)}{i^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$
Kurtosis (4th moment)	$^1/_{\sigma^4}\Sigma_{i=0}^{L-1}(z_i-\mu)^4 p(z_i)$	Dissimilarity	$\sum_{i,j}\lvert i-j\rvert p(i,j)$	High gray level run emphasis (HGRE)	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j).i^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$
Coefficient of variation	$^\sigma/_\mu$	Cluster shade	$\sum_{i,j}((i-\mu i)+(j-\mu j))^3 p(i,j)$	Short run low gray level emphasis (SRLGE)	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{r(i,j)}{i^2 j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$
		Cluster prominence	$\sum_{i,j}((i-\mu i)+(j-\mu j))^4 p(i,j)$	Short run high gray level emphasis (SRHGE)	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{r(i,j).i^2}{j^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$
		Variance	$\Sigma_{i,j}(i-\mu i)^2 p(i,j)$	Long run low gray level emphasis (LRLGE)	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{r(i,j).j^2}{i^2}}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$
				Long run high gray level emphasis (LRHGE)	$\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j).i^2.j^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} r(i,j)}$

[a] μ_n, z_i, $p(z_i)$, and L represent mean nth moment, random variable related to intensity, normalized histogram of intensity levels, and number of possible intensity levels, respectively.

[b] σ is standard deviation, μ represents the mean, and $p(i, j)$ is gray level co-occurrence matrix.

[c] $r(i, j)$, N_g, N_r, and n_p are run length matrix, number of gray levels in an image, number of run lengths, and sum of image pixels, respectively (Mollazade et al., 2013).

This technique includes two steps: I. selection of the appropriate type of transformation and determination of its parameters, and II. selection of a proper re-sampling algorithm. Images were registered with the image processing toolbox in MATLABR2013a software facilities (The Mathworks, Inc., Natick, MA, USA).

Different transformations including affine, polynomial order 2, and polynomial order 3 were tested to achieve the best result for image registration. The minimum number of control points needed for operation was 3, 6, and 10 for affine, polynomial order 2, and polynomial order 3, respectively. Control points are recognizable points on the original images that are applied for registering the images. Conventionally, Root Mean Squared Error (RMSE) is calculated based on the method of residuals to find the best transformation (Nguyen, 2015).

ArcGIS software was used to evaluate accuracy related to each method of transformation. The nearest neighbor method was selected for re-sampling because of its speed.

In addition, it is based on real values existent in images so that it does not produce new digits that maybe unrealistic. This is especially useful when images are later used for classification.

2.4. Image processing

MATLAB R2013a and its image processing toolbox were used for processing soil images (The Mathworks, Inc., Natick, MA, USA).

2.4.1. Texture analysis

Texture is defined as a function of spatial variations of pixel intensity. Texture analysis has shown potential as an important method in many applications of computerized image analysis for classification of images based on local spatial variations of intensity or color. In the present study, unlike most previous studies, tilled soil surface was not considered as distinct aggregates or clods, rather it was studied in terms of texture. In this research, texture-based features were considered from four statistical techniques. These methods were a gray level co-occurrence matrix, a gray level run length matrix, first order statistics of image histogram, and a local binary pattern.

2.4.1.1. Gray level co-occurrence matrix. The tilled soil images were analyzed using the distance $d = 1$ pixel with angles 0, 45, 90, and 135° as suggested by Haralick (1979). Following that stage, GLCMs[3] were normalized and ten statistics were extracted from each normalized GLCM (Table 4) (Haralick et al., 1973). A total of 40 features were extracted from each image (4 orientations × 10 statistics).

2.4.1.2. Gray level run length matrix. A run-length matrix *p (i, j)* is defined as the number of runs with pixels of gray level *i* and run length *j*. Various texture features can then be derived from the GLRLM[4] (Tang, 1998). In this study, fourty-four features were obtained in the feature extraction stage from GLRLM as shown in Table 4 that exhibits four orientations with angles of 0, 45, 90, and 135° × 11 statistics.

2.4.1.3. First order statistics of image histogram (FOSIH). Assuming random variable, '*I*' is represented by gray levels of the image region. The first-order histogram *P (I)* is defined as below (Aggarwal and Agrawal, 2012):

$$P(I) = \frac{\text{Numberofpixelswithgraylevel I}}{\text{Totalnumberofpixelsintheregion}} \quad (2)$$

Eight statistical characteristics, detailed in Table 4, were worked out as suggested by Gonzalez et al. (2004).

2.4.1.4. Local binary pattern. The LBP[5] describes each pixel relative to the gray level of its neighboring pixels. If the gray level of a neighboring pixel is higher than or equal to the central pixel then the value is one, in other cases it is zero. LBP is robust to rotation by clockwise rotation of the obtained binary number and selection of the maximum possible value (Ojala et al., 2002).

In the present study, eight features were extracted from LBP, as shown in Table 4.

2.5. Data mining process for feature selection

The large size of input vector of ANN makes it more complicated and causes an adverse effect on performance of the classification process. Therefore, the procedure of data mining may be used. CfsSubsetEval evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features that are highly correlated with a class while having low intercorrelations, are preferred (Hall, 1998). In addition, Greedy Stepwise was applied as the search method. WEKA software was employed for the feature selection stage. The WEKA workbench is known as an easy-to-use and robust software for data mining.

2.6. ANN classifier

The multilayer perceptron artificial neural network was used for classifying tilled soil images. MLPs[6] are feed forward ANNs trained with the standard back propagation algorithm. In this study, the single-hidden layer architecture of the MLP was applied. The selected features were fed to the network as inputs. The number of output neurons was set to one. It should be noted that in this way, a set of thresholds was written in the related program to assign the network output to one of the nine classes of tillage quality.

Tangent sigmoid and linear transfer functions were used in developing the MLP models for hidden and output layers, respectively. Input data were normalized between −1 and 1.

Data were randomized in order to refuse over fitting and to ensure proper classification. Then, data for each camera height were divided into three groups: 60% for training, 15% for cross-validation and the remaining 25% for the testing stage. The reason for taking part of the data for cross-validation was to prevent overtraining of the network.

The optimum network structure was determined by trial and error. Performance of MLP models was compared by mean square error (*MSE*) of training and cross-validation stages to find different numbers of neurons in the hidden layer and their correlation coefficients (*r*).

Each topology was implemented 10 times. To select the best structure of a network, the minimum mean square error and maximum correlation coefficient in training and cross-validation stages were considered. The difference between MSEs[7] in training and cross-validation stages was also taken into account. If the difference was determined as high it indicated that the network was trapped in a local extremum. MATLAB codes were implemented in a laptop computer with the configuration: Core 2Duo CPU, 2 GHz, 2.99 GB RAM, Windows XP configuration.

In order to evaluate performance of classification, a confusion

[3] Gray Level Co-occurrence Matrices
[4] Gray level run length matrix
[5] Local binary pattern
[6] Multilayer Perceptrons
[7] Mean Square Error

matrix was formed using a test dataset on the network with the optimum topology. A confusion matrix is an array, which summarizes a comparison between two variables related to a dataset.

The classification performance was evaluated by statistical measures extracted from values of the confusion matrix. In this study, statistical measures of specificity, sensitivity, precision, accuracy, and *AUC* were calculated (Eqs. (3)–(7)).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Sensitivity(recall)} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{6}$$

$$AUC = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{7}$$

where:

T, *F*, *P* and *N* denote True, False, Positive, and Negative, respectively. In this way, *TP*, *TN*, *FP*, and *FN* were true positive, true negative, false positive, and false negative, respectively.

In general, sensitivity indicates how well a model identifies positive cases and specificity measures how well it identifies negative cases, whereas, accuracy is expected to measure how well it identifies both categories (Aggarwal and Agrawal, 2012). Nevertheless, effectiveness of ANNs classifier is defined by the accuracy measure as it cannot be a good performance criterion alone. In other words, if both sensitivity and specificity are high (low), accuracy will be high (low). However, if any one of the measures of either sensitivity or specificity is high and another is low, then accuracy will be biased towards one of them (Aggarwal and Agrawal, 2012).

Overall, precision assessed class agreement of data with positive labels defined by the ANN. AUC[8] is an effective criteria used to measure performance of a classifier. AUC shows the ability of a classifier to prevent false classification (Sokolova and Lapalme, 2009). It is the area under the curve of receiver operating characteristic. ROC[9] is the plot in which a true positive rate versus a false positive rate is drawn (Fawcett, 2006). All the mentioned statistical measures were calculated for each dataset obtained from different camera heights during photography.

3. Results and discussion

3.1. Image registration

The total RMSE[10] related to each tested transformation in each camera height is represented in Table 5. A lower value of total RMSE indicates better accuracy of the transformation. It can be inferred from Table 5 that the transformation of polynomial order 3 had the lowest total RMSE of the 3 tested camera heights during photography. The obtained values for total RMSE of the tested transformations were very low, so all of the tested transformations represented acceptable accuracy in terms of image registration. Finally, an affine transformation was selected for image registration because it needed only four control points (the fourth point was used to rectify the process so it was not essential). In fact, selections for transformation type considered the number of control points as well as the total RMSE value. The four control points required for affine transformation were selected from the four corners of the wooden frame without using grids. However, the other transformations (polynomial order 2 and polynomial order 3) needed the grid because further control points were required. The existence of grids in the images could have caused difficulty in the feature extraction process.

Table 5
The results of total RMSEs for tested transformations (meter).

	Transformations		
Camera height (cm)	Affine	Polynomial order 2	Polynomial order 3
60	0.00150	0.00054	0.00016
80	0.00130	0.00046	0.00025
100	0.00110	0.00033	0.00026

Fig. 2 presents a visual display of the results of image registration using affine transformation for each height of photography. The red borders were intentionally embedded in images to illustrate the relocation and rectification of images after image registration.

3.2. Feature selection

Table 6 indicates that 19, 14, and 17 features were selected for photography heights of 60, 80, and 100 cm, respectively as effective features on classifying tillage quality.

The results of feature selection show that for a photography height of 60 cm, the algorithms of GLCM, GLRLM, FOSIH and LBP participated in this stage with numbers of 2, 10, 5, and 2, respectively. For the height of 80 cm, participations of algorithms were 1, 6, 4, and 3 features, respectively. Also, for height of 100 cm, these algorithms involved 2, 8, 4, and 3 features, respectively. This indicates that the GLCM and GLRLM algorithms had the lowest and highest numbers, respectively, of effective features on classifying soil tillage quality.

3.3. ANN classifier

Table 7 shows the best structures for each photography height along with performance factors of MLP. According to Table 7, networks with number of neurons 26, 22, and 20 in the hidden layer represented the best performance for the photography heights of 60, 80, and 100 cm, respectively. The classifier was evaluated based on the test data set by confusion matrices related to ANN classifiers with the 19-19-1, 14-22-1, and 17-20-1 topology for the photography heights of 60, 80, and 100 cm, respectively (See Table 8).

The bold cases (diagonal elements in Table 8) represent the number of points for which the estimated class was equal to the true class, i.e., the number of correctly classified instances of each class.

The confusion matrix related to Camera height of 60 cm shows that non-zero elements were gathered around the major diagonal. This indicates proper performance of the classifier in identifying the different classes of tillage quality. In terms of True Positive Rate

[8] Area under curve
[9] Receiver Operating Characteristic
[10] Root Mean Square Error

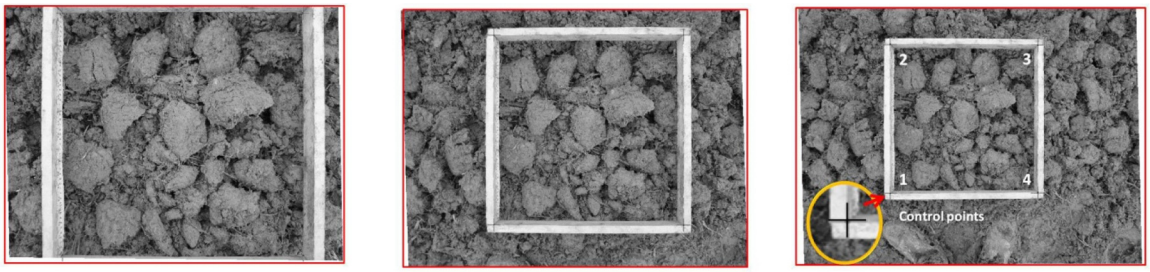

Fig. 2. The results of image registration using affine transformation for each height of photography; a, b and c are registered images acquired at photography heights of 60, 80 and 100 cm, respectively. Control points are illustrated in section c.

Table 6
Results of feature selection by the CfsSubsetEval method for three heights of photography.

No.	The height of photography (cm)		
	60	80	100
1	Contrast (GLCM-0)	Cluster prominence (GLCM-135)	Cluster shade (GLCM-45)
2	Cluster prominence (GLCM-135)	Long run emphasis(GLRLM-0)	Cluster prominence (GLCM-135)
3	Run length non-uniformity (GLRLM-0)	Run length non-uniformity (GLRLM-45)	Short run low gray level emphasis (GLRLM-0)
4	Long run emphasis (GLRLM-45)	Short run high gray level emphasis (GLRLM-45)	Short run high gray level emphasis (GLRLM-0)
5	Run length non-uniformity (GLRLM-45)	Run length non-uniformity (GLRLM-135)	Run length non-uniformity (GLRLM-45)
6	Short run high gray level emphasis (GLRLM-45)	Low gray level run emphasis (GLRLM-135)	Short run emphasis (GLRLM-90)
7	Run length non-uniformity (GLRLM-90)	Short run high gray level(GLRLM-135) emphasis	Long run emphasis (GLRLM-90)
8	Short run low gray level emphasis (GLRLM-90)	Mean gray level (h)	Short run high gray level emphasis (GLRLM-90)
9	Long run emphasis (GLRLM-135)	Third moment (h)	Run length non-uniformity (GLRLM-135)
10	Run length non-uniformity (GLRLM-135)	Measure of uniformity (h)	Long run high gray level emphasis (GLRLM-135)
11	High gray level run emphasis (GLRLM-135)	Coefficient of variation (h)	Average gray level (h)
12	Short run high gray level emphasis (GLRLM-135)	Average gray level (LBP)	Average contrast (h)
13	Mean gray level (h)	Third moment (LBP)	Third moment (h)
14	Third moment (h)	Measure of uniformity (LBP)	Coefficient of variation (h)
15	Entropy (h)		Average gray level (LBP)
16	Third moment (h)		Average contrast (LBP)
17	Coefficient of variation (h)		Third moment (LBP)
18	Average gray level (LBP)		
19	Third moment (LBP)		

GLCM: Gray level co-occurrence matrix, GLRLM: Gray level run length matrix, h: First order statistics of image histogram and LBP: Local binary pattern algorithms. 0, 45, 90 and 135 indicate the orientations in degree.

Table 7
Performance of MLP for the best topology of network at each camera height.

Camera height (cm)	The best topology	Training		Cross Validation	
		MSE	r	MSE	r
60	19-26–1	0.2809	0.9786	0.2591	0.9792
80	14-22–1	0.1948	0.9847	0.3531	0.9739
100	17-20–1	0.2181	0.9829	0.3544	0.9763

for individual classes, the model was able to obtain over 80% TPR[11] for classes of 1, 2, 4, and 9. However, for the classes of 1 and 9, TPR values were 95 and 100%, respectively. Conversely, there were only a few false negatives in the corresponding confusion matrix. This means that the classifier had good performance for identification of positive cases. For the photography height of 80 cm, the TPR value was more than 80% for the classes of 1, 2, 6, and 9. The highest value belonged to class 9 (93.10%) followed by classes 6 (87.1%) and 1 (85%). There was a minimum value of TPR in class 4 because of an increase of false negatives in this class. For the photography height of 100 cm, the highest value of TPR was obtained in classes 1 (96.77%) and 9 (95.45%). TPR in class 4 had the lowest value, as there was a relatively high interference with classes 3 and 5. Table 8 indicates that false negative values for class 9 were 0, 2, and 1 respectively for photography heights of 60, 80, and 100 cm. Considering MWDs of different classes of tillage quality (Table 3), there was larger interval between class 9 and 8 than for any other consecutive pair. For this reason, the false negatives in class 9 were very low.

4. Effect of photography height on determining accuracy of tillage quality

The result of performance evaluation of the proposed method for each class of tillage quality is summarized in Table 9 as an average. Performance comparison of the classifier between the three heights of photography demonstrated that the highest average accuracy per class was obtained at the height of 60 cm with values of 93.79% against 92.84%, and 92.57% for photography heights of 80 and 100 cm, respectively.

A maximum overall accuracy of 72.04% was obtained for the photography height of 60 cm compared to heights of 80 (67.88%) and 100 cm (66.67%). There was a significant improvement in overall accuracy with reduced camera height from 100 cm to 60 cm. This can be related to spatial resolution and existence of higher information from the soil surface because the camera was closer to the soil surface. In terms of other statistical measures, the best performance was achieved at the height of 60 cm, such that

[11] True Positive Rate

Table 8
Confusion matrices related to ANN classifier for nine classes of aggregate size distribution (tillage quality).

Estimated	Actual																									
	photography height of 60 cm									photography height of 80 cm									photography height of 100 cm								
	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
1	**35**	2	1	0	0	0	0	0	0	**34**	2	1	0	0	0	0	0	0	**30**	5	1	0	0	0	0	0	0
2	2	**21**	8	0	0	0	0	0	0	6	**16**	6	2	1	0	0	0	0	1	**19**	4	1	0	0	0	0	0
3	0	2	**25**	1	1	0	0	0	0	0	2	**21**	12	1	0	0	0	0	0	2	**22**	18	1	0	0	0	0
4	0	0	3	**29**	8	0	0	0	0	0	0	7	**21**	2	0	0	0	0	0	0	4	**30**	7	1	0	0	0
5	0	0	0	4	**21**	4	0	0	0	0	0	1	13	**15**	2	0	0	0	0	0	3	13	**24**	3	0	0	0
6	0	0	0	0	4	**14**	11	5	0	0	0	0	0	4	**27**	8	1	0	0	0	0	0	4	**24**	9	0	0
7	0	0	0	0	0	4	**32**	7	0	0	0	0	0	0	0	**39**	7	0	0	0	0	0	0	2	**26**	6	0
8	0	0	0	1	0	1	4	**37**	0	0	0	0	0	0	1	8	**24**	2	0	0	0	0	0	0	9	**24**	1
9	0	0	0	0	0	3	6	10	**23**	0	0	0	0	2	1	2	12	**27**	0	0	0	0	0	1	3	11	**21**

the average precision was 71.93% against 66.94, and 67.56% for the heights of 80 and 100 cm, respectively. In addition, at the height of 60 cm the measures of sensitivity and specificity had higher values at 74.19 and 96.54%.

Comparison of AUC values for the three photography heights of 60, 80, and 100 cm (Table 9) showed the highest AUC at the height of 60 cm (85.36%). A greater AUC of a classifier indicates better performance. This evaluation is related to the existence of more information gathered at a lower height of photography.

Based on these observations mentioned above, it can be inferred that information detail increases in relation to a lower height of photography. Hence, the dataset obtained for the camera height of 60 cm was determined as the most appropriate for better identification of tillage quality compared to the datasets relating to heights of 80 and 100 cm. In conclusion, the photography height of 60 cm is introduced as the optimum height of photography for determining tillage quality using image analysis.

4.1. Performance evaluation of the proposed method

The overall accuracy per each class for photography height of 60 cm, which was suggested as the optimum height of photography for classes of 1–9 were 98.48, 95.74, 95.14, 94.83, 93.62, 90.27, 90.27, 91.49, and 94.22%, respectively. In addition, averages for accuracy and precision per each class were 93.79 and 71.93%, respectively.

Overall accuracy of the ANN classifier obtained for the suggested approach was 72.01%. It should be noted that accuracy depends on values of sensitivity and specificity. Thereupon, if both sensitivity and specificity are high (low), accuracy will be high (low). However if any one of sensitivity or specificity is high and the other is low, then accuracy will be biased towards one of them (Aggarwal and Agrawal, 2012). Thus, considering the high level of difference between averages of sensitivity and specificity (74.19% against 96.54%), it can be concluded that the classification accuracy was biased. Therefore, another measure namely the area under the curve was calculated, too. In practice, the measure of AUC performs very well and is used when a general measure of predictability is desired (Marrocco et al., 2008). Some studies have been conducted on the relationship between AUC and accuracy. It has been established that AUC is statistically consistent and a more discriminating value than accuracy (Cortes and Mohri, 2003; Huang and Ling, 2005). On this basis, the average AUC was determined as 85.36%. The highest values of AUC regarding individual classes belonged to classes 9 (96.90%) and 1 (96.78%). Classes 2 and 4 with values of 90.36 and 89.56% ranked highest.

Considering the value of statistical measures detailed above, the method used in the present study demonstrated proper performance for determining soil tillage quality. Comparing average accuracy of discrimination between the classes, it was found that the ANN classifier represented the highest and lowest performances for classes 1 (98.48%) and 6 (90.27%), respectively. Overall, the majority of performance measures had higher values in class 1, except for sensitivity. Sensitivity had the highest value in class 9 with the value of 100% and after that in class 1 with the value of 94.59%. This proper performance continued with the slope slightly downward to class 5 (93.62%) and after that it had the minimum value of 90.27% in class 6 whereupon it continued with an ascending trend to class 9. Considering the related confusion matrix (Table 8), it was observed that in class 6 there was a relatively large interference with class 7 for classifier stratification. This can be attributed to closeness between MWDs related to these classes (obtained from sieve experiments represented in Table 3). As shown in Table 3, the MWDs related to classes 6 and 7 were

Table 9
The Performance measures of ANN classifier for determining class of tillage quality in three heights of photography.

Class	Camera height of 60 cm					Camera height of 80 cm					Camera height of 100 cm				
	Ac	Pr	Se	Sp	AUC	Ac	Pr	Se	Spe	AUC	Ac	Pr	Se	Sp	AUC
1	98.48	92.11	94.59	98.97	96.78	97.26	91.89	85.00	98.96	91.98	97.87	83.33	96.77	97.99	97.38
2	95.74	67.74	84.00	96.71	90.36	94.22	51.61	80.00	95.15	87.57	96.05	76.00	73.08	98.02	85.55
3	95.14	86.21	67.57	98.63	83.10	90.88	58.33	58.33	94.88	76.61	89.97	51.16	64.71	92.88	78.79
4	94.83	72.50	82.86	96.26	89.56	89.06	70.00	43.75	96.80	70.27	86.63	71.43	48.39	95.51	71.95
5	93.62	72.41	61.76	97.29	79.53	92.10	48.39	60.00	94.74	77.37	90.58	55.81	66.67	93.52	80.09
6	90.27	41.18	53.85	93.40	73.62	94.83	67.50	87.10	95.64	91.37	93.92	64.86	77.42	95.64	86.53
7	90.27	74.42	60.38	96.01	78.20	92.40	84.78	68.42	97.43	82.92	91.19	76.47	55.32	97.16	76.24
8	91.49	86.05	62.71	97.78	80.24	90.58	68.57	54.55	96.14	75.34	91.79	70.59	58.54	96.53	77.53
9	94.22	54.76	100.00	93.79	96.90	94.22	61.36	93.10	94.33	93.72	95.14	58.33	95.45	95.11	95.28
Average	93.79	71.93	74.19	96.54	85.36	92.84	66.94	70.03	96.01	83.02	92.57	67.56	70.70	95.82	83.26
Overall accuracy	72.04					67.88					66.67				

The Ac, Pr, Se and Sp are abbreviation for accuracy, precision, sensitivity and specificity, respectively. The values are in percent.

Table 10
The stepwise accuracy of classifier with addition of classes 2–9.

	The number of class involved in classification (from class 1)							
	Up to class 2	Up to class 3	Up to class 4	Up to class 5	Up to class 6	Up to class 7	Up to class 8	Up to class 9
Accuracy	90.32	81.82	82.09	77.98	74.74	71.66	69.93	72.04

51.79 and 57.45 mm, respectively. These evaluations are relatively closer than other pairs of classes. However, there was relatively high standard deviation between MWDs of two classes of 6 and 7 (obtained by standard sieve measurements detailed in Table 3) that caused their interference together. Furthermore, lower performance of the classifier in these classes can affect interference. The interference was also valid for class 9 in relation to class 8. In this case, it was observed that the maximum value of MWD in class 8 was equal to the minimum value related for class 9 that can affect interference (Table 3).

Table 10 shows stepwise overall accuracy of classification with adding the classes from 2 to 9. As deduced, adding subsequent classes from 2 to 9 to reduced accuracy of the classification. The reducing tendency of accuracy according to an increased number of classes was expected due to complexity of the classification between higher numbers in classes. Table 10 reveals that having only two classes, 1 and 2, a high accuracy of 90.32% was obtained. While, after adding class 3 to two previous classes, accuracy significantly reduced to 81.82%. However, adding class 4 showed no reduction in accuracy. The reason for this significant reduction in overall accuracy between classes 2 and 3 was the greater difference between mean of MWDs of the two classes that caused an increase of classification interference. This subject can be well realized from the related confusion matrix (Table 8) from high interference between class 3 and class 2 so that the FN of class 3 was relatively high (FN = 8).

After class 4, adding the next classes (5–9), the overall accuracy reached below 80% so that the overall accuracy for nine classes was 72.04%. The reason for this increase in overall accuracy from class 8–9 was higher difference between MWDs than in previous pairs that simplified distinction.

ANN classified the four first classes with the overall accuracy of 82%. The reason for reducing the majority of statistical measures between classes 1 and 9 was attributed to interactions of soil texture and gray level values with tillage intensity. Greater MWD for a class means a greater size (area) of aggregates. Aggregates with greater area composed from more pixels in an image. Thus, there was a wider range of gray level intensity in the larger aggregates. This means a greater number of gray intensity values between 0 and 255 in an aggregate and more proximity between gray intensity values. More diversity in gray intensity values can result in higher error in classification of textural classes, partly because of a greater need for comprehensive data in the training stage. Suppose one hundred aggregates relating to a special class (belong to greater sizes), the number of possible states for aggregate configuration (indentations and projections) and other effective factors on textural features is very different caused by a higher number of pixels. In other words, the texture features will have a wider range of values in larger aggregates than in smaller ones thus affecting classification performance. Furthermore, the classifier needs of more data in the training stage caused higher variation in texture features. However, for finer soil texture, each aggregate contains less pixels that in turn produces less diversity in gray level intensity and other effective factors.

Consequently, in a texture-based method, dissociation and resolution between textures with larger aggregate size confronts higher error and interference between the classes. Contrary to this, methods of texture analysis have higher power for recognizing smaller and finer aggregates. It is suggested that for larger clod sizes, other feature extraction techniques should be used and their performances evaluated.

Another point to consider is that the most important factor in proper classification is the discrimination power of the applied features. In the current study, due to proximity of textural features in the cloddiness texture (classes 5–9), identification of these classes was more complex. However, in the finer classes with a higher degree of pulverization, as clods were converted to aggregates, the textural features represented better performance in classification. Considering the objective of secondary tillage as preparation for an appropriate seedbed and fine aggregate, the presented approach is more useful in secondary tillage for identifying tillage quality than for primary tillage.

As a final point, the superiority of this approach is that it resembles an actual situation because it has comprehensively considered the soil surface and the structural components of the soil including clods and aggregates in an image that was then subjected to analysis.

5. Conclusions

The following conclusions were drawn from the results of this study to determine tillage quality using the textural analysis method at three heights of photography:

1. The photography height of 60 cm represented the best performance for determining soil tillage quality compared to the heights of 80 and 100 cm
2. The texture-based method had good performance for determining aggregate size distribution.
3. The present approach had better performance for classes 1–4 than for classes 5–9 for identification of soil tillage quality. Overall, the approach represented more accurate results for finer soil texture than coarser texture. Accuracy over 80% was obtained for tilled soil with MWDs lower than 35 mm.
4. The advantage of this approach is that it closely resembles an actual situation because the soil surface was completely involved in the analysis and all components of the soil including its clods and aggregates were investigated from a photographic image.
5. Considering the objective of secondary tillage (preparation of appropriate seedbed and fine aggregate), the approach presented in this study is more useful in secondary tillage for determining tillage quality than in primary tillage.

Acknowledgments

Authors are grateful to Technical and Vocational Training Centre, Guilan, Rasht, for providing the laboratory facilities for this project. Also, they thank A. Ahmadi Ara for his expertise and his help during the experiments.

References

Aggarwal, N., Agrawal, R.K., 2012. First and second order statistics features for classification of magnetic resonance brain images. J. Signal Inf. Process. 3, 146–153.

Atkinson, B.S., 2008. Identification of optimum seedbed preparation for establishment using soil structural visualization. PhD Summary Report No. 6. School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough EL12 5RD Project No. RD-2004-3031.

Bogrekci, I., Godwin, R.J., 2007. Development of an image processing technique for soil tilth sensing. Biosyst. Eng. 97, 323–331.

Braunack, M.V., Dexter, A.R., 1989a. Soil aggregation in the seedbed: a review: i. Properties of aggregates and beds of aggregates. Soil Till. Res. 14, 259–279.

Braunack, M.V., Dexter, A.R., 1989b. Soil aggregation in the seedbed: a review. II. Effect of aggregate sizes on plant growth. Soil Till. Res. 14, 281–298.

Campbell, D.J., 1979. Clod size distribution measurement of field samples by image analysing computer. Unpublished Paper SIN/274. Scottish Institute of Agricultural Engineering, UK.

Chimi-Chiadjeu, O., Le Hégarat-Mascle, S., Vannier, E., Dusséaux, R., Taconet, O., 2012. Segmentation of elevation images based on a morphology approach for agricultural clod detection. 5th International Congress on Image and Signal Processing (CISP'12), IEEE, pp. 701–705.

Chimi-Chiadjeu, O., Le Hégarat-Mascle, S., Vannier, E., Taconet, O., Dusséaux, R., 2014. Automatic clod detection and boundary estimation from Digital Elevation Model images using different approaches. Catena 118, 73–83.

Cortes, C., Mohri, M., 2003. Auc optimization vs. error rate minimization. Adv. Neur. In (NIPS 2003) .

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874.

Gonzalez, R.C., Woods, R.E., Eddins, S.L., 2004. Digital Image Processing Using MATLAB. Prentice-Hall Press, USA.

Hall, M.A., 1998. Correlation-based Feature Subset Selection for Machine Learning. Hamilton, New Zealand.

Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. Systems, Man and Cybernetics. IEEE Trans. 3, 610–621.

Haralick, R.M., 1979. Statistical and structural approaches to texture. Proc. IEEE 67, 786–804.

Huang, J., Ling, C.X., 2005. Using auc and accuracy in evaluating learning algorithms. IEEE Trans. Knowl. Data Eng. 17, 299–310.

Itoh, H., Matsuo, K., Oida, A., Nakashima, H., Miyasaka, J., Izumi, T., 2008. Aggregate size measurement by machine vision. J. Terramechanics. 45, 137–145.

Marrocco, C., Duin, R.P.W., Tortorella, F., 2008. Maximizing the area under the ROC curve by pairwise feature combination. Pattern Recogn. 41, 1961–1974.

Mollazade, K., Omid, M., Akhlaghian Tab, F., Rezaei Kalaj, Y., Mohtasebi, S.S., Zude, M., 2013. Analysis of texture-based features for predicting mechanical properties of horticultural products by laser light backscattering imaging. Comput. Electron Agr. 98, 34–45.

Nguyen, T., 2015. Optimal ground control points for geometric correction using genetic algorithm with global accuracy. Eur. J. Remote Sens. 48, 101–120.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002. Multiresolution grey-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. 24 (7), 971–987.

Olsen, H.J., 1992. Sensing of aggregate size by means of a horizontal mini-penetrometer. Soil Till. Res. 24, 79–94.

Sandri, R., Anken, T., Hilfiker, T., Sartori, L., Bollhalder, H., 1998. Comparison of methods for determining cloddiness in seedbed preparation. Soil Till. Res. 45, 75–90.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inform. Process. Manag. 45, 427–437.

Spoor, G., Godwin, R.J., Taylor, J.C., 1976. Evaluation of physical properties of cultivated layers for the comparison of different tillage treatments. ISTRO Seventh International Conference, Sweden.

Stafford, J.V., Ambler, B., 1988. Seedbed assessment using video image analysis. ASAE Paper. No. 88-3541 88-3541.

Stafford, J.V., Ambler, B., 1990. Computer vision as a sensing system for soil cultivator control. Proc. ImechE C419/041.

Taconet, O., Ciarletti, V., 2007. Estimating soil roughness indices on a ridge-and-furrow surface using stereo photogrammetry. Soil Till. Res. 93, 64–76.

Taconet, O., Vannier, E., Le Hégarat-Mascle, S., 2010. A contour-based approach for clods identification and characterization on a soil surface. Soil Till. Res. 109, 123–132.

Taconet, O., Dusseaux, R., Vannier, E., Chimi-Chiadjeu, O., 2013. Statistical description of seedbed cloddiness by structuring objects using digital elevation models. Comput. Geosci. 60, 117–125.

Tang, X., 1998. Texture information in run-length matrices. IEEE Trans. Image Process. 7, 1602–1609.

Vannier, E., Ciarletti, V., Darboux, F., 2009. Wavelet-based detection of clods on a soil surface. Comput. Geosci. 35, 2259–2267.