

RESEARCH

Open Access



Development of an automatic sorting system for fresh ginsengs by image processing techniques

Seokhoon Jeong¹, Yong-Min Lee² and Sangjoon Lee^{1*} 

*Correspondence:

mcp94lee@sunmoon.ac.kr

¹ Division of Smart

Automotive Engineering,

SUN MOON University,

Chungnam, Asan, Republic

of Korea

Full list of author information
is available at the end of the
article

Abstract

This study was conducted with the objective of implementing a smart IoT (internet of things) factory consisting of an automatic 6-year-old fresh ginseng grade classification device. Conventionally, washed 6-year-old ginseng from farmlands is manually sorted into three grades using classification criteria such as weight and shape. However, the cost associated with this classification process has been on the increase. Consequently, to reduce this associated cost, we developed an automatic ginseng sorting device that classifies 6-year-old ginseng according to weight and shape via image processing and sends the classification results to a factory server over a network. Evaluations conducted of the performance of the developed machine using 100 units of 6-year-old ginseng showed that it has a high recognition rate, with an accuracy of 94% for Grade 1, 98% for Grade 2, and 90% for Grade 3.

Keywords: Agriculture classification, IoT factory, Image analysis

Introduction

Ginseng and red ginseng both originate from the Republic of Korea. Red ginseng products manufactured and sold by the Korea Tobacco & Ginseng Corporation are acknowledged around the world for their quality and reliability. Red ginseng has been peeled, heated through steaming at standard boiling temperatures of 100 °C (212 °F), and then dried or sun-dried. However, an objective quality classification is required to grade 6-year-old fresh ginseng for the manufacture of red ginseng rather than depending on the human eye or experience [1]. This grading has caused controversy with cultivators every year. Reportedly, hundreds of millions of dollars and 1530 skilled inspectors are required to grade raw ginseng each year. Automatic classification for some, if not all, 6-year-old fresh ginseng for the manufacture of red ginseng would significantly reduce inspection costs [2, 3]. To this end, in this study, an automatic 6-year-old ginseng sorting machine was developed. The developed machine sends the 6-year-old ginseng classification results to a factory server that stores the daily classification results, monitors the current classification status, and effectively performs production management.

Materials and methods

Design of automatic ginseng sorting machine

Figure 1 illustrates the operation of the automatic ginseng sorting machine. Among its components are an inlet conveyor belt as shown in Fig. 1a and a secondary classification conveyor belt as shown in Fig. 1b. The weight of the incoming ginseng is first estimated by a camera (Fig. 1(1)). Next, images of the ginseng are captured from four different directions via a 4-way camera, as shown in Fig. 1(2). Here, the 4-way camera module is a mechanism for determining whether three-dimensional image analysis is possible and will be used in future studies. The incoming ginseng is sorted via a weight-estimating algorithm and a shape analysis algorithm into Grades 1, 2, and 3, respectively. In addition, the sorting results are displayed on an liquid crystal display (LCD) monitor and transmitted to the factory server through an intranet. The ginseng is transferred from point A to point B at a speed of 0.5 m/s and image analysis execution speed is approximately 0.1 s/sample. Further, the developed machine can sort two to three units per second.

Figure 2 shows the actual developed system. The system comprises a control box for conveyor belt control, a display unit for checking the image processing results, and inlet and sorting conveyor belts.

Figure 3a shows the ginseng inlet conveyor belt, which is configured in black for efficient image processing. The four sorting actuators that control the module and the location of the classification conveyor belt are as shown in Fig. 3b.

The controller of the automatic ginseng sorting machine

The main control box provides efficient control of the ginseng sorting machine. Figure 4 is a block diagram of the control system box, which consists of a power controller, two alternative current (AC) induction motor drivers for the conveyor belt, four pneumatic actuator controllers, a ginseng position sensor on the conveyor, and a light controller

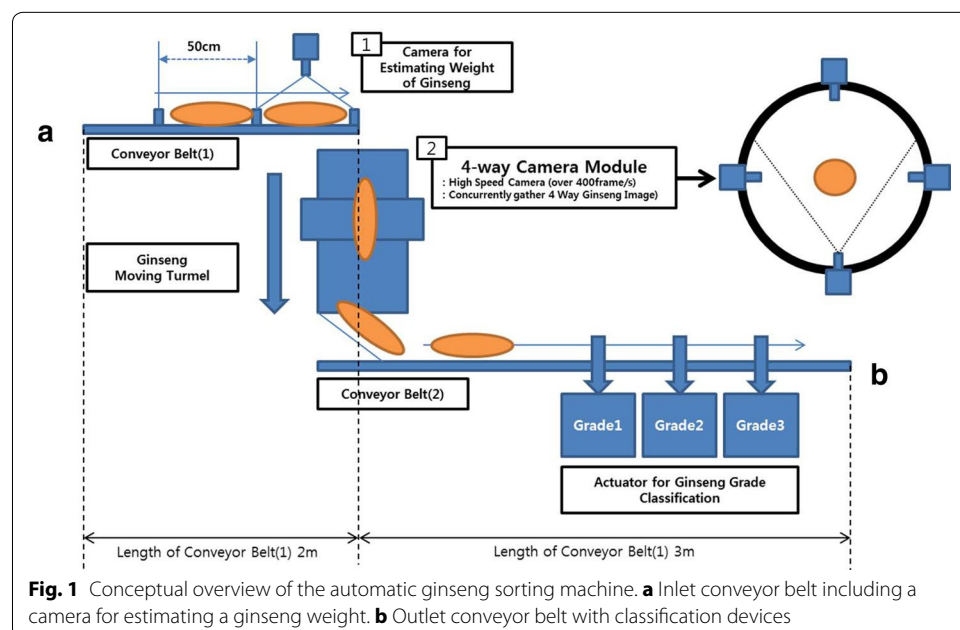
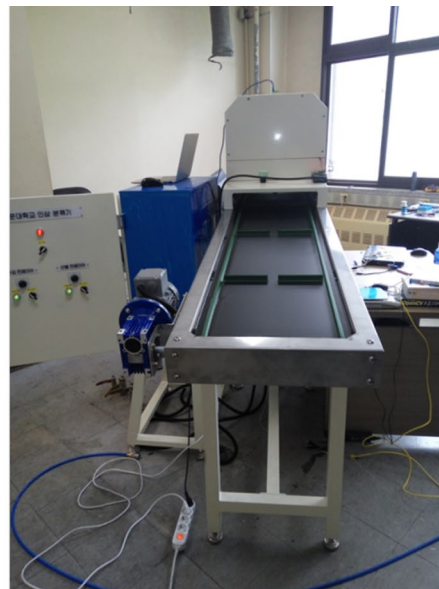




Fig. 2 The developed ginseng sorting machine



a

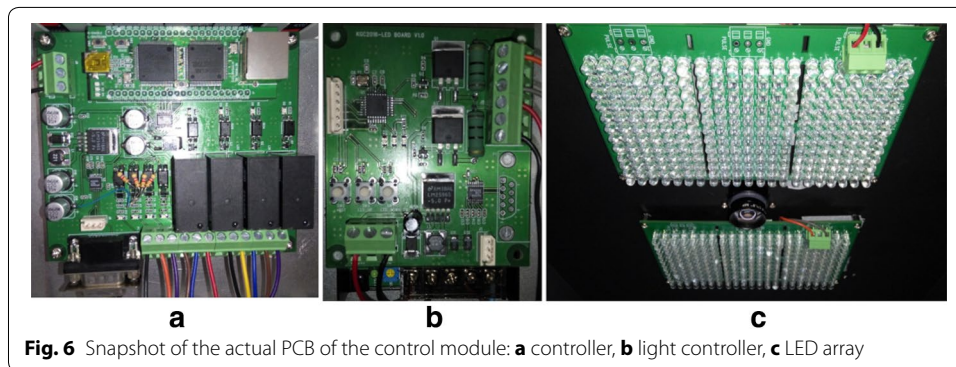
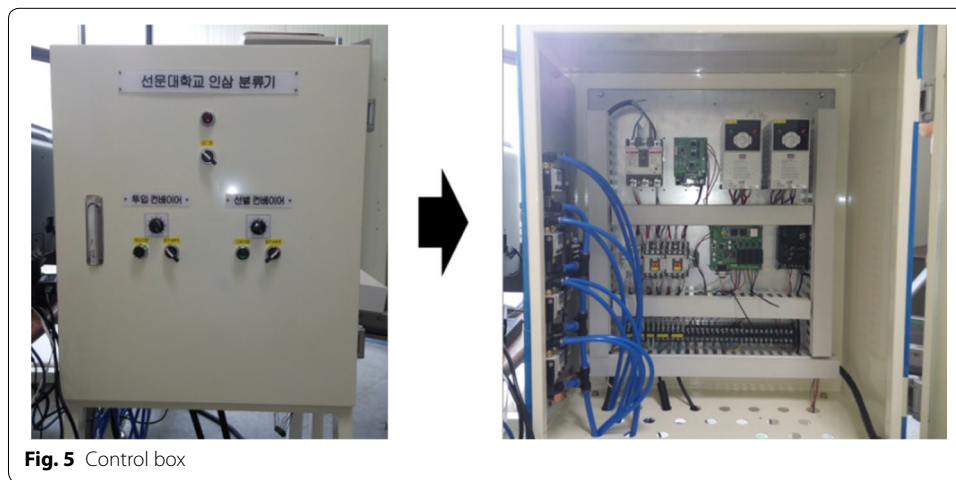
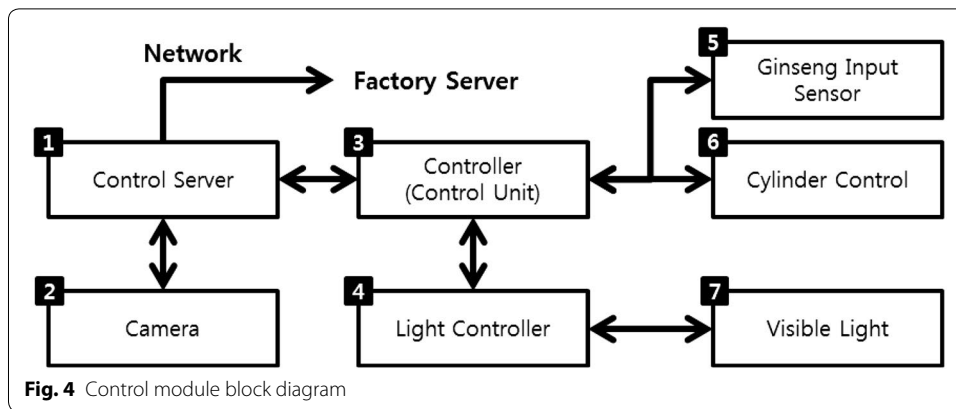


b

Fig. 3 The two conveyor belts in the automatic ginseng sorting machine: **a** ginseng inlet conveyor belt, **b** sorting conveyor belt

for the image-estimating camera. The five cameras are connected to a control server via universal serial bus (USB), as shown in Fig. 4(1). On completion of image analysis, the classification information is sent to the factory server and displayed on an LCD unit.

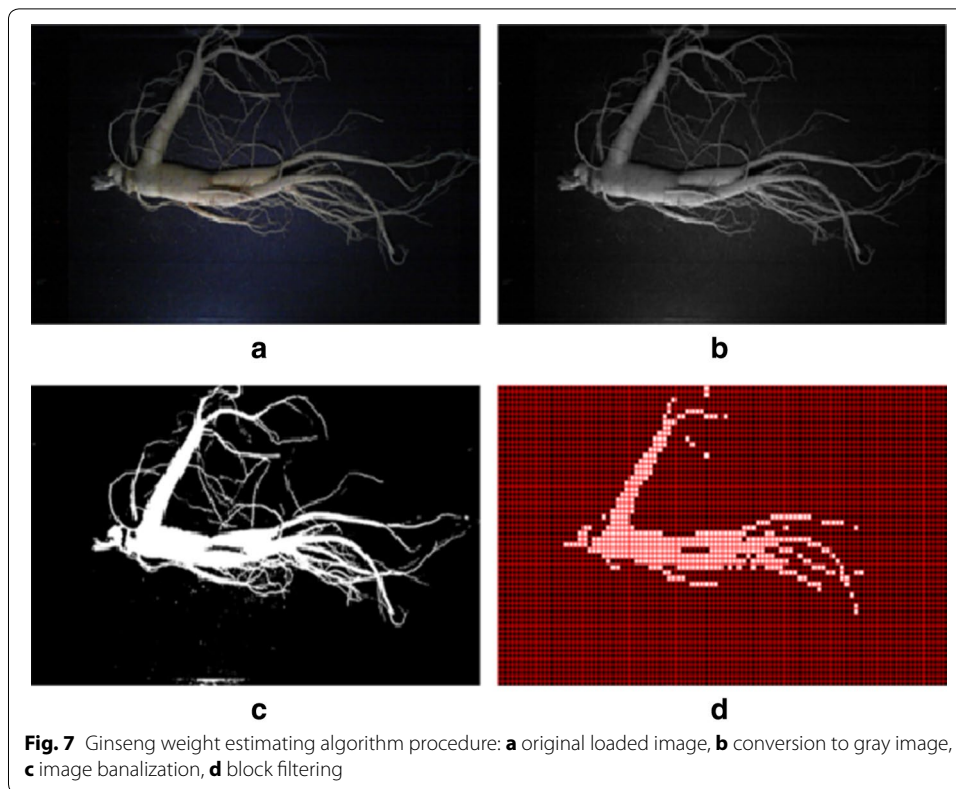
Figure 5 shows the control box for the complete ginseng sorting machine. All of the modules shown in Fig. 4 are located inside this control box. Figure 6a is a snapshot of the actual developed main control printed circuit board (PCB) and communicates with the server and receives the rating results to perform the cylinder control function. Figure 6b



shows a light controller for controlling the illumination brightness of a LED array, and Fig. 6c shows an array of light emitting diodes (LEDs) installed in a dark room as a light source of a camera.

The ginseng weight estimating algorithm

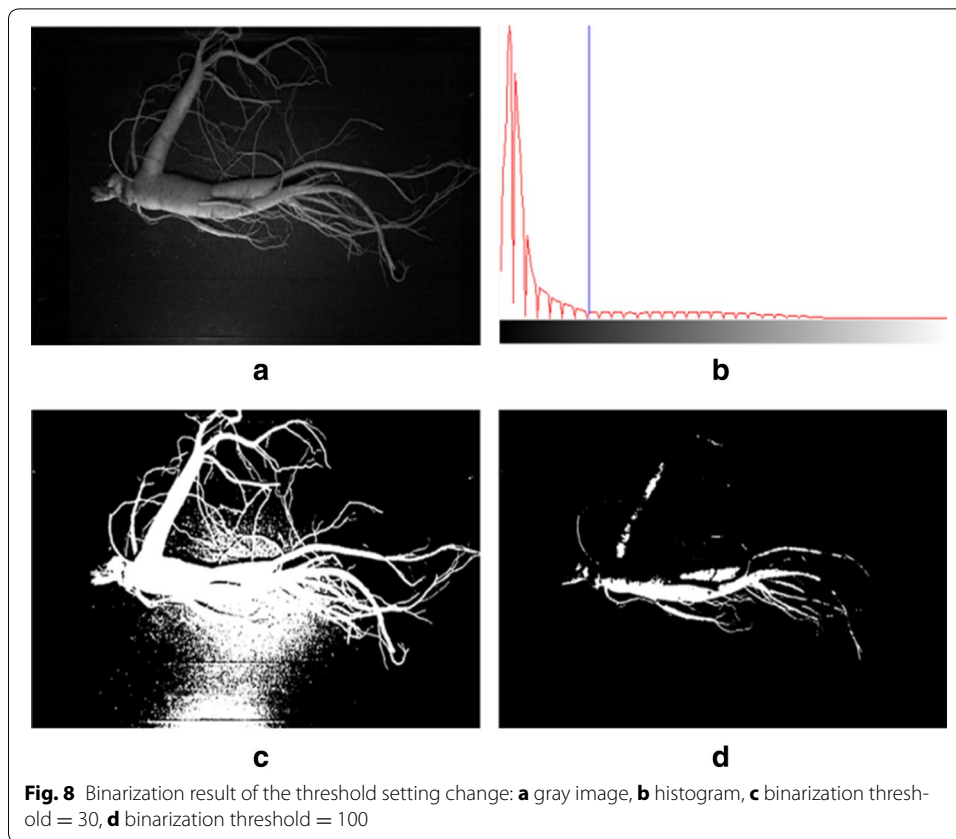
Figure 7 illustrates the ginseng weight estimation procedure. First, the original 1920×1080 pixel color image (Fig. 7a) is converted to a 256 level (8 bit) or zero (black)



to 255 (white) gray level value based on the differences in light and shade to analogize the correlation of the image with the fresh ginseng weight (Fig. 7b). Then, the gray image is filtered to extract the parameter most closely related to the weight information. The number of pixels ranging from zero to 255 is then counted in the converted gray image, and a histogram that shows the distribution developed. Next, the background region is eliminated from the fresh ginseng image and binarization performed to distinguish the body. Here, the binary image is subjected to partial bright lighting against the background to distinguish the fresh ginseng body part (Fig. 7c). Subsequently, block filtering is performed to eliminate the minute lateral roots that have little influence on the weight of the fresh ginseng, and the remaining blocks are used to calculate the parameters (Fig. 7d).

In image processing, binarization refers to the changing of all pixels higher than a given threshold to white and pixels below that threshold to black. Various binarization methods exist, including global fixed thresholding, locally adaptive thresholding, and hysteresis thresholding. As a method for determining the threshold value, there is a typical Otsu method, a histogram locating method, Huang and Wang et al. [4–8]. In this study, we used global fixed binarization because we used images acquired in the same environment with limited illumination and camera position through the developed ginseng image acquisition device.

Figure 8 shows the result of the binarization process according to the set threshold value. For the gray image (Fig. 8a), the histogram of the brightness values of the pixels (Fig. 8b) shows that the binomial distribution is evenly distributed from zero to 255 and

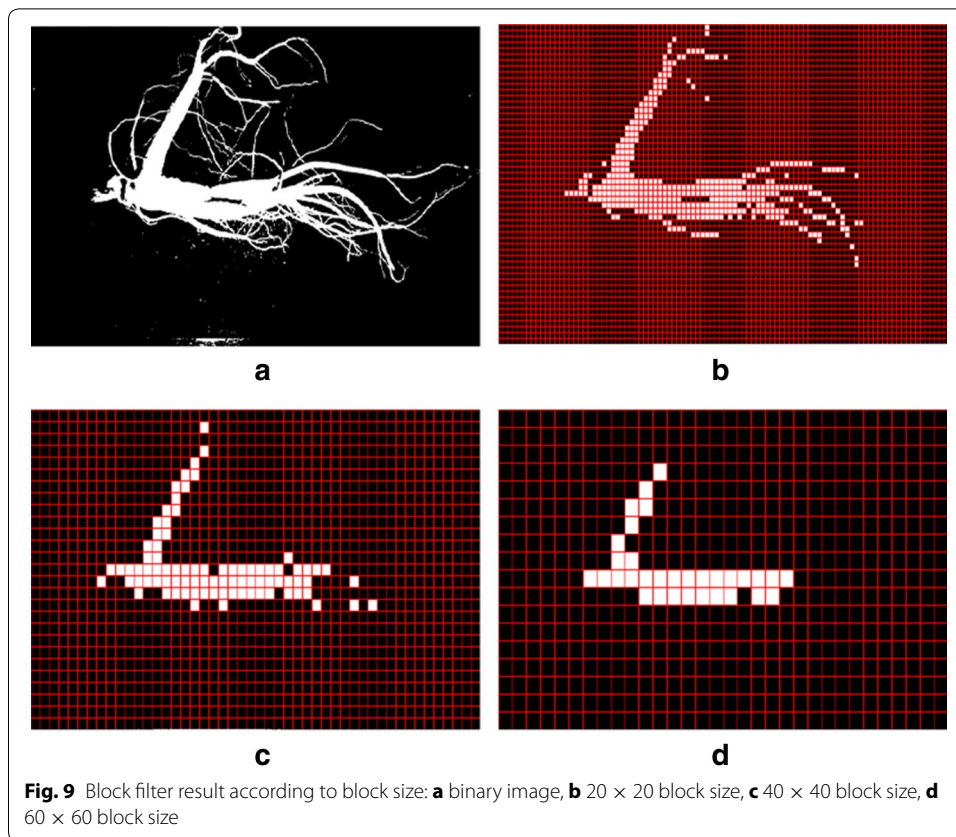


that the area between the ginseng and the background cannot be precisely divided. This is a difficult problem because, if the threshold is too low, as in Fig. 8c, the bright areas of the background are highlighted. Conversely, if it is too high, as shown in Fig. 8d, it will cause the black areas of the body to blend into the dark background, causing image loss. It is expected that it will be helpful to reduce errors by applying binarization method after detecting object using color of image [9].

A block filter divides regions of the 1920×1080 pixel images into blocks of certain sizes. The blocks are white if the distribution of white pixels is higher than a particular threshold (%) and black if it is lower than the threshold. This filter is used to remove fine roots and show the body region that provides much of the weight.

Figure 9 shows the observed results when the length \times width dimensions of the block filter were increased to 20×20 , 40×40 , and 60×60 while the threshold was fixed at 60%. The body shape was shown while the fine roots were removed with the 20×20 blocks (Fig. 9a), but losses occurred owing to the staircase phenomenon and resolution deterioration with increasing block size. Thus, it is important to set the proper block size and threshold.

Correlation analysis was also performed to investigate the relationship between the weight of the ginseng and the ginseng image pixel value. The correlation between the weight of the ginseng and the number of blocks remaining after the block filter was analyzed using linear regression analysis. Through this process, we examined the



optimal binarization threshold and the size of the block filter with the highest correlation coefficient.

First, the binarization threshold was changed from 41 to 50, and the value of the correlation coefficient observed to determine the threshold value with the highest correlation coefficient. Next, the change in the correlation coefficient was observed while varying the block size from 11 to 20 pixels (the width and length were the same) and the threshold value from 86 to 95%, respectively. Using the correlation coefficient used in this experiment, the t distribution was tested on both sides with a significance level of 0.05 for the number of hosiery relations. The test for the number of parental relationships of pixel values and weights of ginseng images in the parent group indicated that there is a correlation between the two variables in the population and a strong positive correlation was found as the correlation coefficient was closer to one.

Fresh ginseng is not processed after harvest in the field, it contains about 70% of water and it is difficult to store for a long time, so it is recommended to treat ginseng for ginseng production within 1 week. Therefore, the samples applied to the weight estimation algorithm used the samples within 1 week after cultivation to minimize errors.

The ginseng shape analysis algorithm

Fresh ginseng specimens have unique appearances. However, the names of certain parts are different. Figure 10 classifies the five main parts of fresh ginseng. Figure 10a shows the fibrous root, which is attached to the body or head region and is not present in every

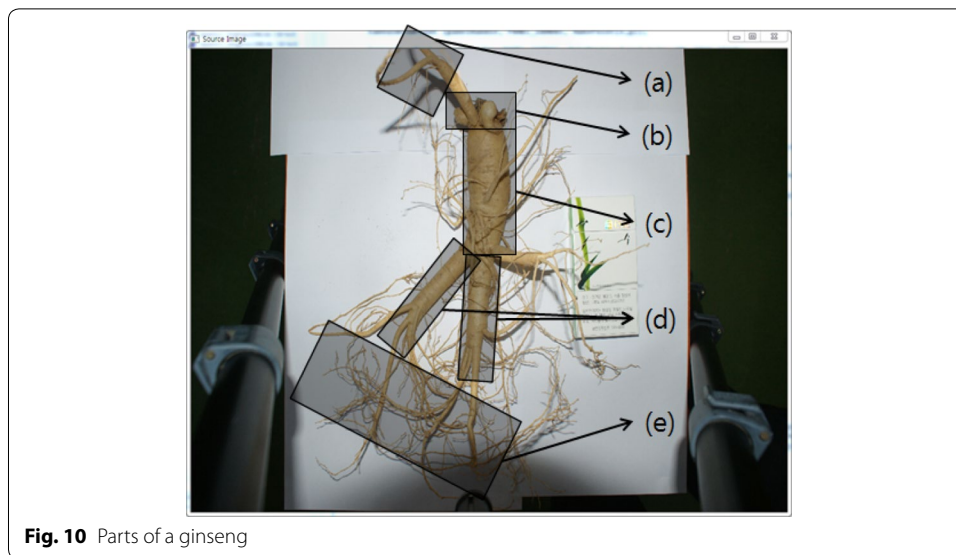


Fig. 10 Parts of a ginseng

fresh ginseng. Figure 10b shows the rhizome, which is present in every fresh ginseng at the head region. Figure 10c shows the body of fresh ginseng, which is called the taproot. This occupies most of the fresh ginseng and is the most important part when manufacturing red ginseng. Figure 10d shows the leg region, which is called the lateral root; each fresh ginseng has about one to five of these roots. Figure 10e shows the fine roots, which absorb the nutrients from the ground.

Table 1 presents the classification criteria used for Grades 1, 2, and 3 ginseng provided by the Korea ginseng corporation (KGC) ginseng research Institute. Decision parameters such as the fresh ginseng weight, taproot length, ratio, and number of lateral roots were used, and the fact that experts with long careers classify specimens with rough judgment without accurate measurement tools was examined. In this study, (1) the height of the taproot, (2) width of the taproot, (3) height/width ratio of the taproot, and (4) width/height ratio of the taproot were selected as the grading parameters of fresh ginseng for pattern recognition training and test data.

Experiments and results

Weight-estimation results

Regarding the threshold in the banalization in the gray image, the block filter settings were fixed (block size: 10×10 , threshold: 90%) and the correlation coefficient from the linear regression analysis was observed while changing the banalization threshold between 41 and 50. A threshold of 43 was determined to have the highest correlation (Table 1). Subsequently, the length \times width of the block filter and threshold in Tables 2 and 3 were fixed to the banalization threshold of 43. Based on increasing the block filter size from 11×11 to 20×20 pixels and increasing the threshold from 86 to 95% in the linear regression analysis, the highest settings for the correlation coefficient were set to 18×18 pixel block size and 90% threshold.

The correlation coefficient for the number of output pixels and fresh ginseng weight had a maximum value of 0.9162 when a block filter was applied with a banalization

Table 1 Grade classification criteria for 6-year-old ginseng from the KGC Ginseng Institute

Grade 1	Grade 2	Grade 3	Off-grade
Head			
Head balanced with body part	Head balanced with body part	Head balanced with body part	No limit
Body is partially fat or curve is not severe	Body is partially fat or curve is not severe	Body is partially fat or curve is not severe	No limit
Body and leg			
Body length is 6 cm or longer, more than two legs that are 5 cm or longer, but body radius must be 2/5 or less than the body length	Body length is 4 cm or longer, more than two legs that are 3 cm or longer, but body length is 8 cm or longer and is straight. The body radius must be half or less than the body length	Body length is 3 cm or longer, legs are balanced with the body, and body length of 5 cm or longer and is straight. Poor body shape with a weight of 50 g or higher is included	No limit

threshold of 43, size of 18×18 , and threshold of 90%. The estimated regression was $y = 2.517x + 45.82$, and analysis of variance was performed on the estimated regression equation to test the significance, as shown Fig. 11. It was considered to be significant with $p = 4.4652E-51$ based on an F test with a 0.05 significance level and decision coefficient of $r^2 = 0.839$. This means that 83% of the data are explained by the estimated regression equation.

Figure 11 shows the relationship between the weight of the fresh ginseng and the results after applying the block filter. In the figure, the causes of singular values such as points a and b are analyzed:

1. When the number of pixels is relatively large compared to the weight of the ginseng (point a in Fig. 11), the ginseng has many fine roots and a dense distribution, which occupies a relatively large volume.
2. When the number of pixels is relatively small compared to the weight of the ginseng (point b in Fig. 11), then the body is dark because of soil or foreign matter in the ginseng, or the root or fine roots of the ginseng cover the body, it takes up a small volume.

Figure 12 shows two examples of error occurrences. In the case where the fresh ginseng has many fine roots and their distribution is dense, as shown in the original image (Fig. 12a) and the filter image (Fig. 12b), most of the fine roots of the fresh ginseng remain untouched and the number of output pixels is relatively large. In addition, as shown in the original image (Fig. 12c) and the filter image (Fig. 12d), when the surface of the ginseng is covered with soil or foreign matter, it blends in with the background image and is consequently removed after filtering, resulting in only a relatively small number of pixels.

The result of shape analysis

We used a support vector machine (SVM) for shape pattern classification [10–12]. Figure 13 shows the grade classification method based on pattern recognition. The training data and test data for pattern recognition were designed such that they would not repeat. The correct match rate (%) of each grade, mean correct match rate, false rejection rate

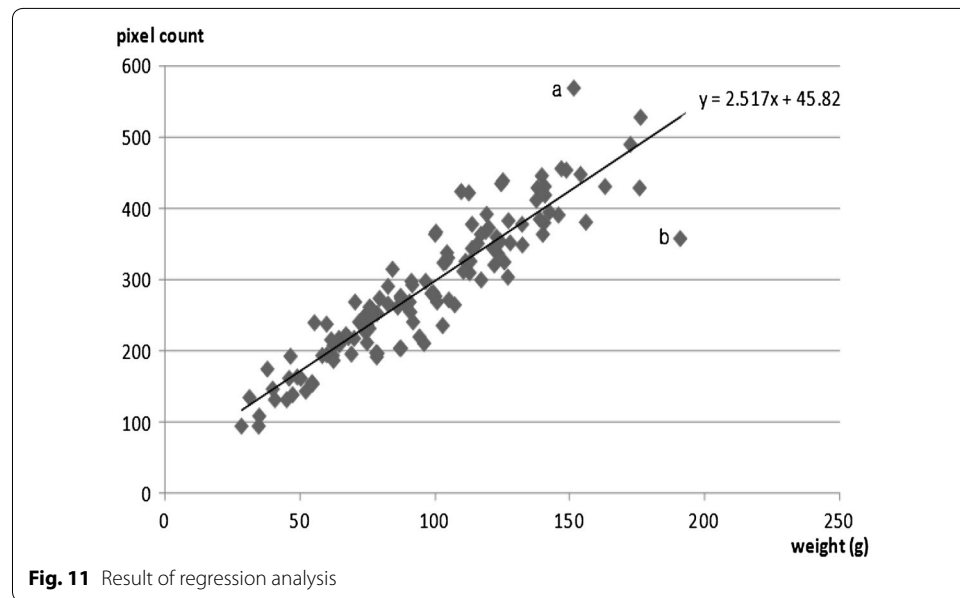
Table 2 Ginseng grade recognition results

Number of training data	Grade 1 recognition rate (%)	Grade 2 recognition rate (%)	Grade 3 recognition rate (%)	Average recognition rate (%)	FRR (%)	FAR (%)
10	94.0	98.0	90.0	94.0	6.0	5.4
15	93.3	97.8	88.9	93.3	6.6	5.9
20	96.6	100.0	80	92.2	7.7	6.3

Table 3 Comparison of ginseng classification algorithms

Algorithm	Accuracy (%)	Speed (s)	Acceptable
SVM	94	> 0.1	Good
MLP	30–40	< 0.1	Insufficient
Inception-v3 v3 image recognition model	Around 50	> 0.1	Insufficient

Accuracy depends on the number of iterations

**Fig. 11** Result of regression analysis

(FRR), and false acceptance rate (FAR) indices were used for performance evaluation. Table 2 presents the recognition rate and performance according to the amount of training data when each of the four parameters was used. The recognition rates were 94% for Grade 1, 98% for Grade 2, and 90% for Grade 3. A high recognition performance comprising 6.0% FRR and 5.4% FAR was obtained with four parameters and 10 training data points, as indicated in Table 2. The average recognition rate decreases to 95, 93.3 and 92.2% as training data increases to 10, 15, and 20, respectively. However, the detection rates in the first and second grades are highest at 96.6 and 100%, respectively, when the training data is 20 samples. As the number of samples increases in the training data selection process, data that are difficult to express characteristics of grade 3 will be selected.

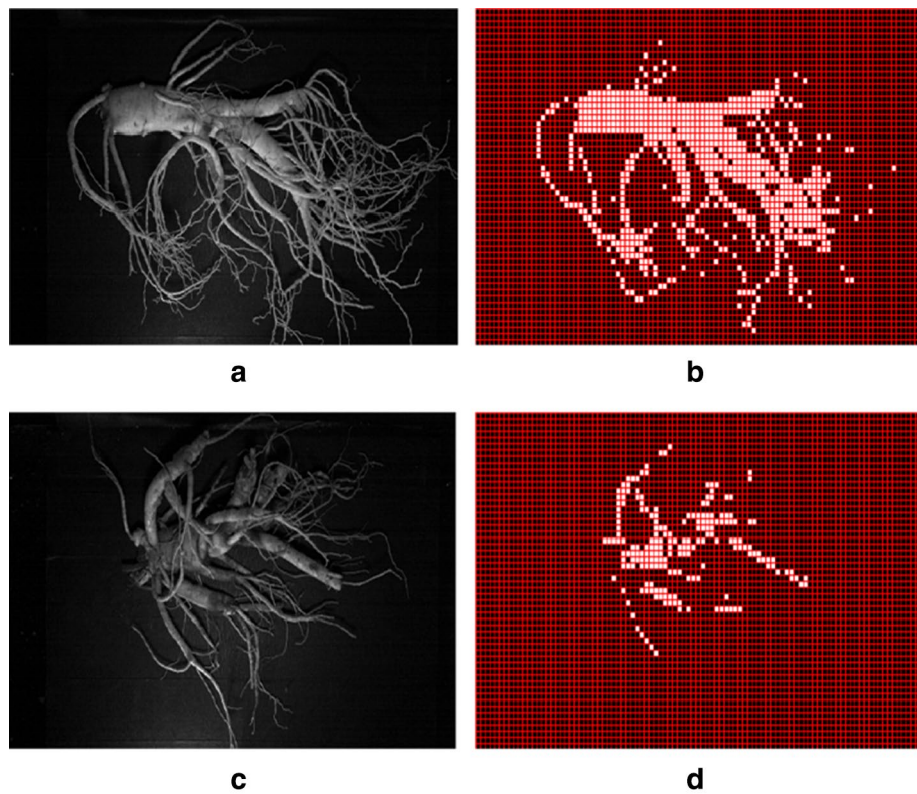
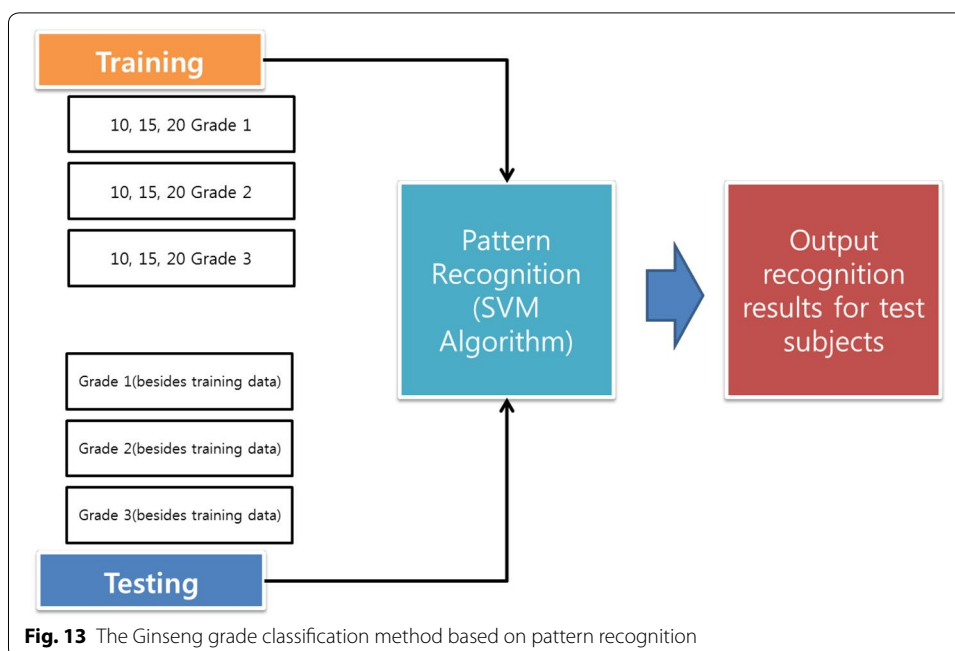


Fig. 12 Examples of causes of error: **a** gray image of error case 1, **b** filtered image of error case 1, **c** gray image of error case 2, **d** filtered image of error case 2

Conclusions

In this study, an automatic 6-year-old ginseng grade classification machine was developed. The developed machine sends the 6-year-old ginseng classification results to a factory server that stores the daily classification results, monitors the current classification status, and effectively performs production management. The ginseng classification is performed using image processing on the basis of criteria such as estimated weight and analyzed shape pattern. The various grades according to the weight of the ginseng are as follows: Grade 3: 30–40 g; Grade 2: 40–75 g; and Grade 1: 75 g and above. In this paper, the proposed algorithm segmented image sector to estimate the weight of a ginseng and experiments have shown that this method is effective for weight estimation. This image segmentation method is used to recognize the postures and gestures of the hand and body for using computer based vision system [13, 14]. Following the weight-estimation procedure, the group of candidates determined as being in Grade 1 was again evaluated using shape analysis and each candidate assigned a grade between one and three. Evaluations conducted of the performance of the developed machine using 100 units of 6-year-old ginseng showed that it has a high recognition rate, with an accuracy of 94% for Grade 1, 98% for Grade 2, and 90% for Grade 3.

We considered several applicable algorithms for ginseng classification, and compared the three algorithms SVM, MLP, and Inception-v3 image recognition model (created by Google) as shown Table 3, which are suitable for classification, and adopted the SVM



algorithm with the best performance among them. But SVM does not possible guarantee excellent recognition performance in general pattern recognition situations. Compared to the automatic classification time of imported ginseng, which is within 0.1 s/sample, there is a limit of the time required to sort the classification by 2–3 pcs/s using pneumatic cylinder. Shortening the overall classification time is a challenge in the future.

Authors' contributions

SL is responsible for the concept of the paper and writing, SL and SJ are responsible for the quantitative analysis of the presented results. Also, Y-ML contributed to the paper on the overall composition and advisory on the experiment. All authors read and approved the final manuscript.

Author details

¹ Division of Smart Automotive Engineering, SUN MOON University, Chungnam, Asan, Republic of Korea. ² School of Mechanical and ICT Convergence Engineering, SUN MOON University, Chungnam, Asan, Republic of Korea.

Acknowledgements

This paper is an extended version of a conference paper (CUTE2016).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea (NRF-2014H1C1A1066998) and supported by KGC ginseng research institute.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 9 March 2017 Accepted: 6 November 2017

Published online: 14 November 2017

References

- Kang JY, Lee MG, Kim YT (1996) Automatic decision-making on the grade of 6-year-old fresh ginseng (*Panax ginseng* C.A. Meyer) by and image analyzer-1. Shape and weight analyses according to the grade of fresh ginseng. *J Ginseng Res* 20(1):65–71

2. Jeong CM, Shin JS (2006) Comparison of grade of raw and red ginseng on each factor of quality in Korean and American ginseng. *Korean J Med Crop Sci* 14(4):229–233
3. Kang JY, Lee MG (2002) Automatic decision-making on the grade of 6-year-old fresh ginseng (*Panax ginseng* C.A. Meyer) by an Image Analyzer 1. Shape and weight analyses according to the grade of fresh ginseng. *J Ginseng Res* 26(1):6–9
4. Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybernet* 9(1):62–66
5. Chang CC, Wang LL (1997) A fast multilevel thresholding method based on lowpass and highpass filtering. *Pattern Recognit Lett* 18(14):1469–1478
6. Cheng HD, Chen YH, Jiang XH (2000) Thresholding using two-dimensional histogram and fuzzy entropy principle. *IEEE Trans Image Process* 9(4):732–735
7. Chi Z, Yan H, Pham T (1996) Fuzzy algorithms: with applications to image processing and pattern recognition. World Scientific
8. Huang LK, Wang MJ (1995) Image thresholding by minimizing the measure of fuzziness. *Pattern Recogn* 28:41–51
9. Truong MT, Kim S (2017) Parallel implementation of color-based particle filter for object tracking in embedded systems. *Hum Centric Comput Inf Sci* 7(1):2
10. Cortes C, Vapnik V (1995) Support-vector network. *Mach Learn* 20:273–297
11. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. *Data Min Knowl Discov* 2(2):121–167
12. Prabusankarlal KM, Thirumoorthy P, Manavalan R (2015) Assessment of combined textural and morphological features for diagnosis of breast masses in ultrasound. *Hum Centric Comput Inf Sci* 1(1):1–17
13. Marfia G, Rocchetti M (2017) A practical computer based vision system for posture and movement sensing in occupational medicine. *Multimed Tools Appl*. 76(6):8109–8129
14. Rocchetti M, Marfia G, Semeraro A (2012) Playing into the wild: a gesture-based interface for gaming in public spaces. *J Vis Commun Image Represent* 23(3):426–440

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
