An Algorithm Based on Core Characteristic Extraction of Watermelon Seeds by Automated Separating System

YONG SUN[1] YUN BAI [2] LIHONG GANG[3] QIANGGUO PU[2] NIKOS MASTORAKIS[4]
[1]Computer Science & Technology Institute, Soochow University, Sushou, Jiangsu 215006, China
[2]Computer Center, University of Science and Technology of Suzhou, Sushou, Jiangsu 215006, China
[3]CCP School, Tancheng, Shangdong, China
[4]Hellenic Naval Academy, Terma Hatzikyriakou, 18539 Piraeus, Greece
Also: WSEAS, Ag.I.Theologou 17-23, 15773, Zografou, Athens, Greece

Abstract: The algorithm of extracting a series of characteristic values, taking the flat objects, seeds, as the experiment is applied into the auto-distinguishing seeds system software successfully. It fulfills the blank in this area. It shows it is practical for the machine to distinguish the flat objects with the equipment visual. It designs and realizes the new algorithm of the length of orthogonal long and minor axis, the degree of connecting superficial pattern and the area; brings forward the concept of the coverage degree of black region and the way to distinguish different colorful objects with the distribution difference of histogram gradation. The system software has a very bright future. It is meaningful to the distinguishing objects areas.

Key-words: long and short axis, the degree of connecting surface patterns, the coverage degree of black region, core characteristic extraction, watermelon seeds

1 Introduction

Recently, Image Recognition has been applied into every field, recognizing, dividing, picking and evaluating tiny objects such as screws[1], cereals and apples[2], etc. The application of new technology lowers the labor costs as well as raises the efficiency of dividing and picking. However, the research of dividing and picking the flat granulosis like watermelon seeds have not been done. As flat granulosis, watermelon seeds have their own characteristics: tiny, irregular shape, rough surface and uneven colors, which makes it difficult to divide and select. The way of dividing and picking watermelon seeds with the computer image control equipment has a bright future[3]. According to a survey, work manual sorting is very common in all of the seeds factories, which is very inefficient and inaccurate. So, this research is very practical and can be served as important reference for other relevant fields, its basic concept is applicable.

Taking watermelon seeds as the object, we design a series of Core Characteristic Extraction by studying its' characteristics, and apply it into the software system of separating watermelon seeds automatically. The algorithms includes the length of orthogonal long and minor axis of the watermelon seeds, the proportion between the length and width, the area, the degree of connecting surface patterns, the coverage degree of black region and the distribution difference of histogram gradation. Using these parameters, it can divide and sort watermelon successfully.

2 System Design

2.1 System Development

The software system is made of image-gathering equipment, image-processing equipment and information output equipment.

Image-gathering equipment is Sumsung SCC - C4203P low intensity color & black and white conversion camera, gathering seeds image. The configuration is: P-4 2.8G CPU, 512M Flash Memory, 80G hard disk, Microsoft Windows XP/2000, Microsoft Visual C++6.0.

2.2 The object of system design

The object of the research is normal watermelon seed. It has different qualities: the best ones, the bad ones, the tablets, the Bending Alice, the black ones, the tile flaps, the skewbald ones, the half flaps and the white skins(Figure 1, from left to right). The object of the system design is to divide and pick the best seeds according to the extracted characteristic values by processing the images with computer.

Fig. 1 Examples of different seeds

2.3 The System Design Flow Diagram

The system design flow is illustrated in Fig 2. Firstly, the hardware system lines the seeds as N * N, which is easy to take picture. Then, it de-noises the figure, converses it gray, makes it binary digit and extracts the outline[4]. It divides the figure equally according to the lattice to make each of the tiny figures fit one of the seeds figures. Then calculates every characteristic value of each seed and matches the results with the normal standard to locate the qualified seeds. Finally, the equipment will divide and pick according to the location.

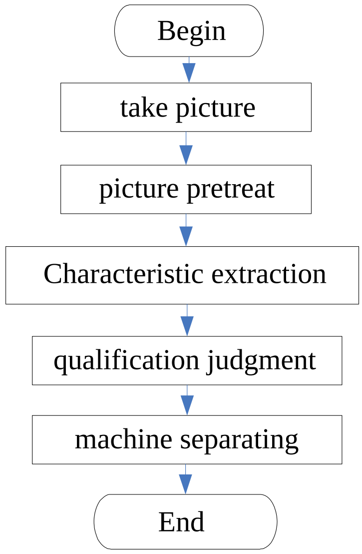


Fig. 2 Software System Design Flow

3 The Extraction of Characteristic Value and Qualification Judgment

3.1 Image Acquisition

Acquired image will directly affect the precision of the extraction of characteristic value as well as the correctness of Recognition Results. To acquire a clear and stable image, we design the following imaging equipment:

1. a square N * N seed lattice transparent template to make imaging easier.
2. a parallel support with grooves on each side to transport transparent template. The support will move with the conveyor. It stops at the pre-made location for a short while and starts taking pictures.
3. Stabilize the camera at some height. It will be controlled on and off by the software.
4. Seal the template and camera into an enclosed chamber. Put 5 light sources above the template and at each side of the template to make the image without shadow.

3.2 Image Pre-handling

Convert the 24bit colorful image to 8 bit gray image and maintain the copy of the 24bit color image. The 8 bit gray image meets the need of extraction of characteristic value and raises the image processing speed. Evaluate the image with the weighted average method. The weighted average method is: to put different weights to R, G and B according to its importance and index, then average the weights, as $R = G = B = (W_R R + W_G G + W_B B)/3$. W_R, W_G, W_B is the weight of R, G and B. As human eye is most sensitive to green, then red and blue lest, so it will get a better gray image when $W_G > W_B > W_R$. The final result is[5]: the best gray image will be acquired when W_R=0.299 , W_G=0.587 , W_B=0.114.

Convert it binary digital on the basis of 8 bit gray image. As the image is gray and different with the background, therefore global threshold segmentation algorithm is chosen. The use can adjust the threshold automatically, according to the actual situation to make the acquired image to be the most effective. To make it binary digital does not change the original image. Pixel value will be used for latter calculation.

To determine the length of orthogonal long and minor axis and the area of the watermelon seeds, the outline of seeds has to be extracted[6]. The accurateness of outline extraction will influence the accuracy of the resulting characteristic value. Through all kinds of comparisons and analyses, a method is chosen that combines the threshold outline extraction and image outline of the hull extract on the basis of Sobe[7]. The method first splits the basic outline and the background with construction act threshold then refines the outline further with outline extraction algorithm to make sure to get a more accurate reading

3.3 The extraction of characteristic value

The extraction of characteristic value is the core of the software system. The accuracy of the extraction of characteristic value will effects the correctness of judgment and is the basis for the possibility of replacing the manual sorting with automated sorting. This research designs the algorithms of the length of orthogonal long and minor axis of the watermelon seeds, the proportion between the length and width, the area, the degree of connecting superficial patterns, the coverage degree of black region and the distribution difference of histogram gradation according to its own characteristics. These algorithms can be applied to relevant areas. The following is the procedure of designing and realizing these algorithms. Pre-handling image and defining the entire figure are the basis of all these algorithms.

3.3.1 The design and characterization of the algorithms of the length of orthogonal long and minor axis and the ratio between the length and width.

The length of orthogonal long and minor axis and the ratio between the length and width is an important index of evaluating the seeds. The definitions are[8]: Long axis is the direct line between the 2 farthest points in the outline of the object image; Minor axis is the longest line among these in which fraction of all objects in a straight line with the long axis perpendicular to the image of the outline lines were cut by the length of the largest segments. The ratio between the length and width is the ratio between the long and minor axis. To raise the efficiency and accuracy, a new algorithm of the length of orthogonal long and minor axis is designed.

1. Take the small image, which was cut as the center to build an orthogonal Cartesian coordinates. To obtain the coordinates of the points of images at x axis , y axis , $y = x$ and $y = -x$. Locate them as 1 to 8 in clockwise direction. These 8 points divide the outline of image into 8 areas, named as 1, 2 … 8. Take the first point as starting point of the first area, etc.
2. First, to calculate the starting point of the first area with the distance between two points formula, i.e., the distance from Point 1 to the rest points. The rule is: First, the distance increase. When it reaches the maximum distance, it decreases. The farthest point is named “n”. To calculate the distance between “n” and the

starting point and all the points at the first area, locate the coordinate between the farthest points. Then start at the staring point at the outline of the second area. Follow the procedure above. If the distance between some point and the other 7 points show a longer distance than the maximum earlier defined, repeat the above procedure accordingly. Combine the 2 procedures to get the length of the long axis. The shapes of most of the seeds are oval. Usually, there are 2 maximum figures. If there are more, the efficiency of the algorithm will decrease to normal algorithm. But the efficiency will increase wholly, doing no harm to the accurateness of algorithm**.** The diagram of sub-solving of long axis is shown in fig. 3.

3. Y axis divides the image into 2 parts. Locate the longest coordinates to the Y axis on both sides. It will make a rectangle. Take the top point and bottom point as the starting and ending point and take pictures from left to right to get the length of the short axis.
4. To calculate the ratio between the length and width with the long and minor axis. It has been proven that we solve the algorithm of long axis with the formula $5n^2/16+2n+56$ as normal algorithm with $n(n-1)/2$. When the pixels at the outline of the image are over 25, the algorithm decreases the frequency of calculation, which increases the efficiency. The algorithm solves with the definition of long and minor axis and the result is real and accurate.

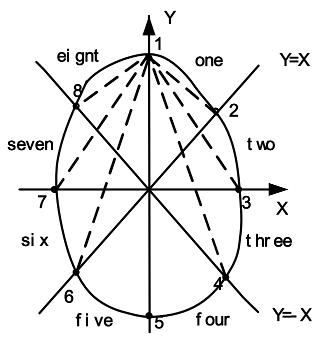
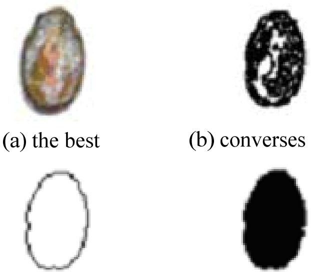


Fig. 3 Diagram of subdividing the field & determining the long axis

3.3.2 The design and realization of the square extraction algorithm

During the process of sorting, it can take out the tiny and immature seeds according to its square. To increase accuracy of the characteristic value, the method of combination of boundary tracking[9] and regional growth was created. Connectivity and better smooth image outline can be obtained with curve tracking extracting image outline[10]. Then do the regional growth with the principle of the difference of regional grayness to fulfill the whole inner part with the black pixels.

The pixels in the digital figures are taken as the center of little square grid. So the region area is equal to the area of all the small squares which are represented by all the pixels[11]. After processing, the black points in the areas are uniformly distributed, which are different with the background color. It is easy to count the black pixels to stand for the areas of the seeds. The result is shown as Fig. 4.



(a) the best (b) converses it gray

(c) boundary tracking (d) regionalgrowth

Fig. 4 The procedure of extraction of the seeds areas

3.3.3 The design and realization of the algorithm for coverage degree of black region

After a large number of trials, it was shown that the white skins, the tablet, the black seeds and the half flaps are either similar or opposite in their nature. The white skins are white; the tablets are similar to the white skins, but are black at the edge. The black seeds are basically black; the half flaps have 2 parts: one is black and the other is white. The best seeds are white at the center and brightly black at the rest areas. We can separate the four kinds of seeds and determine the percentages for which the black areas are covered. The binary digit figure shows the distribution of the black areas. We can calculate the number of black pixels. The ratio between the number and the area shows the distribution of black pixels, which is called the coverage degree of black region. It will be handled in latter part.

3.3.4 The design and realization of the distribution difference of histogram gradation

The general seeds include the bad seeds, which are rough, bleak, gray and white and therefore unqualified. Compared with the good seeds, the bad seeds are bleak and the percentage of gray and white is high. So it is suggested to draw a 24-bit color histogram to distinguish the good and the bad seeds.

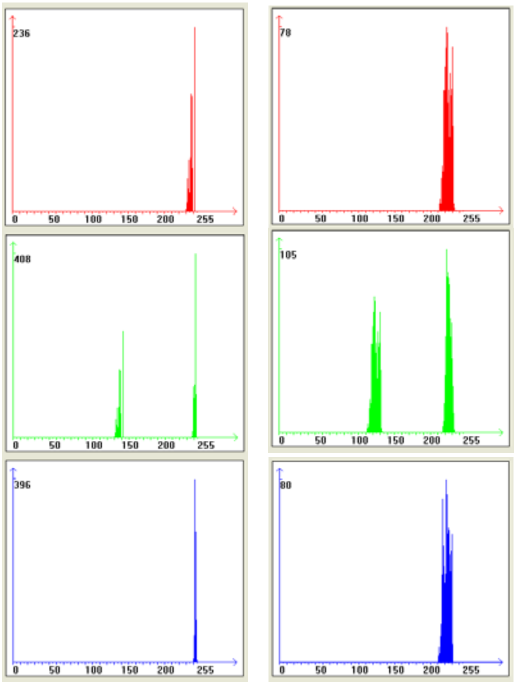


Fig 5 The comparison of gray distribution histogram between the good and the bad seeds

Fig 5's left shows the histogram of the good seeds and the right is the histogram of the bad ones. It is clear that the bad ones are mostly white. So in the gray area above 200, the pixels are frequently appeared at some degree. The maximum frequency is 236, 408 and 396 in the red, green and blue histogram, while it is 78, 105 and 80 for the best ones. The distribution of the gray, the best ones is a lot more than that of the bad ones. As part of this research paper, t was tested that this method can distinguish between the different seeds.

3.3.5 The design and realization of the degree of connecting superficial pattern

The skewbald (with patches of white and brown) is similar to the best seeds in every way. It could not be distinguished from the best seeds. But it is unique in that there are a lot of different cyclic patterns. So it is the basis of differentiation to calculate the number of the cyclic patterns, which leads to solve the degree of connecting superficial pattern.

To solve the degree of connecting superficial patterns, we suggest take the way of set merge on the basis of reducing shadows. We deducted the black pixels at the outline from the binary digital pixels to get the black areas in the new figures, which originally were the white areas. It is shown that the figure after reducing shadows has only one connecting area, which is dotted and includes not more than 3 connecting areas. While the connecting area of the skewbald ones after reducing shadows are at least 4 or more.. It shows clearly in Fig. 5.

The number of the connecting area after reducing shadows is the basis to evaluate the skewbald ones[12]. To reduce the influence of noises and improve the efficiency and accuracy of resolution, we designed the algorithm of binary image multi-connecting calculation resolution under the influence of noises. It divides the image into some grids. It can filter out the noises in the original images by judging the grid's color. In the process of searching, we found the grid image replaces the original image. Then combines the grids in the same connecting area into a set[13]. The method avoids over-scanning and noise images. It is simple and fast to calculate the number of all kinds of connecting areas.

3.4 Evaluation Criteria

According to the analysis and experiments, we extracted the length of orthogonal long and minor axis, the ratio between the length and width, the area, the degree of connecting superficial pattern the coverage degree and the average gray degree of pixels. With these values, we can distinguish the best seeds. We applied the following criteria:

1. The length of the long axis of the best seeds is at lease 35 pixels and the length of the short axis is minimum 22. The ratio between the long and the short axis is at least 1.20. Any seed below these figures are unqualified.
2. The area of the best one is at least 800 pixels. Otherwise it is unqualified.
3. The connecting degree of the best one is over or equal to 3. Any one over 3 is the skewbald one. The coverage degree of the best ones is between 0.75 to 0.90; for the white skins, it is under 0.30; for the tablets, it is between 0.30 to 0.75 and for the black ones, it is over 0.90.
4. After doing the scanning, divide the binary digit processing into 2 parts. Calculate the number of the black pixels. If the ratio between the 2 parts is under 50%, the seeds are classified as half flaps.
5. Prepare a histogram. When the gray degree is between 200 and 255, the seeds for which the highest frequency of the red, green and blue are over 200 are the bad seeds.

In the above procedure, it can separate the unqualified seeds from the good ones.

Compared with the manual handling, the software system is more accurate. It shows that the accuracy in identifying & separating seeds is over 90 %. Take 100 seeds as example. It will take 10 seconds. It would take 7 minutes to sort 1.0kg seeds. That is 200 Kg seeds a day, which is equal to 40 labor hours. This will save cost and time.

Conclusion

The algorithm of extracting a series of characteristic values, taking the flat objects, seeds, as the experiment is applied into the auto-distinguishing seeds system software successfully. It fulfills the blank in this area. It shows it is practical for the machine to distinguish the flat objects with the equipment visual. It designs and realizes the new algorithm of the length of orthogonal long and minor axis, the degree of connecting superficial pattern and the area; brings forward the concept of the coverage degree of black region and the way to distinguish different colorful objects with the distribution difference of histogram gradation. The system software has a very bright future. It is meaningful to the distinguishing objects areas.

References:

[1] Caiyun Hou, Huiyuan Li, Yanfen Shang, *The image distinguish and detection of the corn quality*, China Corn and Oil academy, 18 (2003) 80-83.

[2] Qingzhong Li, Yihua Wang, *Apple auto distinguishing system hardware development on the basis of computer visual*, Journal of agricultural machinery, 31(2), (2000) 56-59.

[3] Petc. *Catfish feature identification via computer vision*. ASAE Paper, 39(5), (1996) 1923-1931.

[4] Feng Li, Zhengkai Liu, Nenghai Yu, *The research on the extraction of human eyes edge on the basis of image division*, Computer Application Research, 8 (2000) 100-102.

[5] Rafael C. Gonzalez, Richard E. Woods. *Digital Image Processing*. Beijing: Electronics Industry Publishing House, (2003).

[6] Zugaj D Lattuati V. *A new approach of color images egmentation based on fusing region and edge segmentation outputs*. Pattern Recognition, 31(2), (1998) 105-113.

[7] Zongqi Lu, Cheng Liang, *Refine Edge with Sobel*, Chinese Journal of Image and Graphics, 5(6), (2000) 516-520.

[8] Jixiang Sun, *Image Analysis*, Beijing: Science Press, (2005).

[9] Tong Li, Qingyun Shi, *A marginal growth of gray and color image segmentation*, Chinese Journal of Image and Graphics, 5(11), (2000) 911-915.

[10] Wen W, Xia A. *Verifying edges of visual inspection purposes*, Pattern Recognition Letters, 20 (1999) 315-328.

[11] Qiudong Sun, Shunfu Gao, Yongping Qiu, *The area calculation of digital closed image*, Computer Application and Software, 22(7), (2005) 102-103.

[12] Zhilin Li, Yunjie Pang, *A fast algorithm of a binary image connected components labeled*, Engineering Graphics Journal, 3 (1998) 80-86.

[13] Shulong Zhu, *The relation between the area and space in the divided image*, The Journal of mapping Institute of the People's Liberation Army, 11(1), (1994) 17-20.

[14] Xuezhe Li, Qiangguo Pu, Nikos Mastorakis, *The immune algorithm and its application to blood pressure measuring*, WSEAS Transactions on Electronics, 3(9), (2006) 490-494

[15] Zhengqin Wu, Qiangguo Pu, Nikos Mastorakis, *The application of the system parameter fusion principle to assessing quality in digital reference services*, WSEAS Transactions on Information Science and Applications, 2(10), (2005) 1635-1640

Supported by: Suzhou Science And Technology Development Project for Industry (No：SG0518)

First Author: Yong Sun, Associate Professor, graduate students tutor, mainly engaged in software engineering, database and object oriented programming, etc.