Research on key technologies of medicine grain defect
detection system based on machine vision

Yueqiu Jiang*, Yujun Wang, Hongwei Gao, Shuang Ma

School of Information Science and Engineering, Shenyang Ligong University, NanPing Str. 6, 110159 Shenyang, China

Received 1 March 2014, www.cmnt.lv

Abstract

In the process of medicine grain production may generate many kinds of defects. If these unqualified medicine granules are not timely detected, it will not only affect the company's reputation but also the health of the patient. This paper mainly studied how to detect the unqualified medicine grain base on machine vision. It mainly consists of three kinds of common defects segmentation and defect area calculation. Firstly, preprocess the medicine grain image for the following procedures. Secondly, obtain the defect region by improved segmentation algorithm, in order to deal with three different drugs grain defects this paper improved three segmentation algorithms, for the damaged tablets propose a local edge detection algorithm that based on grey level difference, for the irregular-shaped tablets adopts the ellipse detection algorithm, which based on Hough transform technology, for the capsule has air uses the multi-scale Canny edge detection operator. Finally, adopt Chain code contour tracking algorithm and Three-point method calculate the area of the damaged tablets, determine whether the tablets meet the requirements. Experiments show that the system can detect unqualified medicine granule quickly and accurately, it is of great practical value.

Keywords: image processing, median filtering, defect detection, region segmentation, area calculation

1 Introduction

In recent years, many developed countries in Europe and America have carried out deeply research on product quality real time detection, and have made a great progress in the online quality detection and production control [1, 2]. But the technical details strictly confidential to the outside world. In china, due to the research starts late, the domestic pharmaceutical testing is mainly relay on manual that production efficiency relatively low, so there often comes out error detection or misses detection and other undesirable phenomena. It does not be able to satisfy the efficient industrial automation requirements. So study on the drug quality inspection has a very good social significance [3-5].

Three common unqualified medicine granules are: damaged tablets, irregular-shaped tablets, Capsule has air bubbles. The medicine grain defect detection is to detect the unqualified medicine grains on the production line. This task includes two main parts: the granule region segmentation and the damaged areas calculation. Image segmentation is dividing image into non-overlapping regions which based on the image grey, texture, shape and colour. According to the prior knowledge separate the target area and background. Through analysis three types of medicine grain image find that each type has their own characteristics. Because of the tablets interior region high connectivity features, it can use the SUSAN edge detection algorithm get edge information. But SUSAN operator is sensitive with threshold lead to edge detection unstable. Therefore, based on the grey level difference propose a local edge detection algorithm. This algorithm can effectively solve the problem. For the irregular-shaped tablets detection, it adopted the ellipse detection based on Hough transform technology. For the bubble capsule, this paper mainly uses the multi-scale Canny edge detection operator [6]. After obtained the defect region, in order to quantify the damage rate [7], this paper mainly uses a method based on Freeman chain code [8]. Because the medicine grains image has the connectivity characteristics, so adopt the algorithm that chain code combined with pixel count to calculate the area. In order to prevent misjudge, in this paper for first time proposed the “three point method” to calculate the area.

2 Image preprocessing

2.1 IMAGE GREY PROCESSING

Dealing with colour images is complicated and needs a long operation time. RGB image does not reflect the morphological characteristics. It is better to convert the colour image into grey image, which only contains brightness information with 256-level grey scales. The conversion Equation for grey image and colour image is:

\[ Y = 0.299R + 0.587G + 0.114B, \]

*Corresponding author e-mail: missjiangyuequ@sina.com
2.2 MEDIAN FILTER

In the process of collection and transportation, the image may be added noise, so there will be influence the image quality [9]. Median filter can smooth the image to reduce the noise effects.

When the noise is rarely and evenly distributed, median filter can better reduce noise in the image [10]. But the noise distribution in the image is relatively concentrated, traditional median filtering effect obviously cannot meet the requirements of the system [11, 12].

In order to meet the requirements that the median filter not only effectively remove the concentrated distribution noise, but also preferably preserve medicine grain image detail [13]. We propose an improved median filter algorithm. The main idea is to get rid of the maximum and minimum values in the sliding window, get the median of the two pixels, denoted $m_p$. Then get the difference between $m_p$ and the value of the corresponding pixel within the window, values $\Delta_p$. If $\Delta_p$ is within a predetermined range of threshold $\Delta TH$, use the mean value $m_p$ replace the grey value of the pixel point.

Let $W(i,j)$ window size is $n \times n$ its centre $(i,j)$. Improved filtering algorithm detailed steps are as follows:

Step 1: First assign $n = 3$, find out the maximum and minimum grey value $p_{\text{max}}$ and $p_{\text{min}}$ in window $W(i,j)$.

Step 2: Statistics the number of pixels value is not equal to $p_{\text{max}}$ or $p_{\text{min}}$ in the window and make it as $c_n$.

Step 3: If $c_1$ less than or equal to 1, then let $n = 5$ execution Step 2.

Step 4: If $c_1$ equal to 0, then let $n = 7$ execution Step 2.

Step 5: Within the sliding window, calculate the median of all elements pixel values:

$m_{p3} = \text{median (the number of pixels in the window } W_3(i,j) \text{ and the grey does not equal } p_{\text{max}} \text{ or } p_{\text{min}})$;

$m_{p5} = \text{median (the number of pixels in the window } W_5(i,j) \text{ and the grey does not equal } p_{\text{max}} \text{ or } p_{\text{min}})$;

$m_{p7} = \text{median (the number of pixels in the window } W_7(i,j) \text{ and the grey does not equal } p_{\text{max}} \text{ or } p_{\text{min}})$.

Step 6: Calculate the overall average $m_p$.

$$m_p = \left\{ \begin{array}{ll} m_{p3} & c_1 \geq 2 \\ m_{p5} & c_1 > 0, c_1 < 2 \\ m_{p7} & c_1 = 0 \end{array} \right.$$

(2)

Step 7: Filtering result $F(i,j)$ is calculated by the original pixel values and median values according to the Equation (3).

$$F(i,j) = a_1 \cdot E(i,j) + a_2 \cdot m_p,$$

(3)

Where $a_1$ and $a_2$ are the weighting coefficients satisfy the condition $a_1 + a_2 = 1$, wherein $a_2$ is determined by the Equation (4).

$$a_2 = \left\{ \begin{array}{ll} 0 & E(i,j) - m_p < TH_1 \\ \frac{E(i,j) - m_p - TH_1}{TH_1 - TH_1} & TH_1 \leq E(i,j) - m_p \leq TH_2, \\ 1 & E(i,j) - m_p > TH_2 \end{array} \right.$$

(4)

where parameters $TH_1$ and $TH_2$ are preset threshold.

Step 8: Go Step 1; processing the next target pixel.

3 Medicine grain of region segmentation algorithm research

Because there is no segmentation algorithm can adapt to all situations [14, 15]. According to the characteristics of each medicine grain image it proposes three different image defect detection segmentation algorithm based on machine vision.

3.1 REGIONAL SEGMENTATION FOR DEFECT TABLETS

Image edge is often accompanied by grey mutation that is to say the grey value of edge pixels satisfy discrete criteria [16, 17]. According to the principle, this paper put forward the improved edge detection algorithm based on local greyscale difference. The core of the algorithm includes computing greyscale differences, selecting candidate point, determining threshold value and refining edge. Details as follows:

Firstly, obtain candidate edge pixels. It can use the formula to quantify the grey step.

$$\delta = A - B.$$

(5)

Before determine $A$ and $B$, first define a constant metric $n$. When use a $3 \times 3$ template traverse the image, $n$ can measure the number of valid point in the template. According to the actual circumstance of medicine grain of image detection, here $n = 3$. Sort the pixels within the coverage area of the template, taking the top $n$ larger pixels summation assigned to $A$, taking $n$ smaller pixels summation assigned to $B$. For each pixel $x$, difference in local greyscale $\delta$ constantly changes with different $x$ position. When $x$ is located in the edge, $\delta$ is larger. So edge candidate can be selected according to the size of the value of $\delta$, the greater of the value, this point the more likely is edge pixel.

After traversing the whole image, local differences in per-pixel can be respectively calculated. According to the ideas of the SUSAN algorithm, here need to set a threshold to determine the candidate edge points. In order to avoid the shortcoming that SUSAN algorithm is affected by the threshold is too big. It can get the adaptive threshold according to the Equation (6):
Using gradient information, when \( dy/dx = \tan \theta \), for \( x \) do derivate, get the following equation:

\[
\frac{(x-a)^2}{m^2} + \frac{(y-b)^2}{n^2} = \tan\theta = 0 ,
\]

\[
a = x \pm \frac{\sqrt{m^2 \tan^2 \theta + n^2}}{m \tan \theta},
\]

\[
b = y \pm \frac{n^2}{\sqrt{m^2 \tan^2 \theta + n^2}}.
\]

Through comparing the Equations (10), (11) with (8) can found that the number of parameters reduced from the original four to three, so just set up two three-dimensional accumulation array \( A_x(a,m,n) \) and \( A_y(b,m,n) \) can meet the requirements. Every point in the image, in turn, change \( m \) and \( n \) calculate the corresponding \( a \) and \( b \), then respectively for \( m \), \( n \) accumulation, after the improved algorithm it obviously reduce amount of calculation.

3.3 THE SEGMENTATION ALGORITHM OF BUBBLES INSIDE THE CAPSULE

This section is mainly to detect capsules whether containing bubbles. Because of the bubbles within the capsule have a clear boundary and its colour significantly different with the surrounding, so the bubbles boundary can be clear detected. Canny operator detects the edge of the more complete and connection degree is also very good, but the result is sensitive to noise, so this paper proposed an adaptive multi-scale Canny edge detection algorithm. Specific steps are as follows:

1) Combined with capsule medicine grain images to determine the scale collection.

If \( \sigma \) represent one scale of the image, then \( 0 \leq \sigma \leq \sigma_{\text{max}} \), the scale of the image edge is always in a range, so assume medicine granule image edge contains \( K \) scale. It can use \( \{ \sigma_k \}_{k=0,1,\ldots,k-1} \) represent the scale collection, the collection can cover all of the image edges scale, it’s element is monotonic distribution and scale difference between the two adjacent elements within the set should be less than the scale interval \( \Delta \sigma \), generally \( 0 < \Delta \sigma < 0.5 \). \( \sigma_{\text{max}} \) is the maximum edge scale, \( \sigma_{\text{max}} = 2^{k-1} \sigma \). After sampling at equal intervals on the scale space \( 0 \sim \sigma_{\text{max}} \) can get the scale set.

2) Use Canny edge detection operator [11] to detect the capsules image edge information on multiple scales.

In the course of capsules edge detection, first with a Gaussian function for image smoothing, then derivative of capsules image get the original image edge information on various scales. The edge information at the scale \( \sigma_k \) can be expressed as

\[
f^*_k(r,c), k = 0,1,\ldots,k-1.
\]
3) Through fusion the information of scale edge to determine the edge.

After obtaining the information of each scale on the capsule image edge, complete the synthesis of various scale edge information to get the new multi-scale image edge information. In this paper, use the weighted summation algorithm.

\[
f_k'(r,c) = \sum_{i=0}^{\infty} w_i f_i'(r,c),
\]

where \( w_i \) represents the weight factor of each scale edge information. If ignore the noise it can use mean-weighted factor. However, due to uneven illumination bring significant noise effect to the capsule image, so the mean-weighted does not meet the requirements. The weights of different scales case determined by the following steps:

Step 1: On the scale \( \sigma_i \) use a Gaussian function which standard deviation is \( \sigma \) smooth the capsules image to reduce noise. \( f_i = f \cdot G(x,y,\sigma_i) \).

Step 2: In various scales of the capsule image, calculate the difference \( \Delta_1 = |f - f_1| \) and variance \( \Delta_2 = |f - f_2|^2 \).

Step 3: Calculates the difference and variance of image and the weight can be determined by the ratio of the two parameters.

4) The Non-maxima Suppression method to extract the edge points.

When using non-maxima suppression to get edge, mainly in order to find the ridge of the original image from the gradient image, that is to say our target point is maximum gradient, so discarded the edge is not a maximum, set \( a_1, a_2, a_3 \) are the 3x3 neighbourhood pixels of \( a \):

\[
a_1 \quad a_2 \quad a_3 \\
a_4 \quad a_5 \quad a_6 \\
a_7 \quad a_8
\]

If \( a \) satisfy one of the following conditions it could as the edge points, otherwise use \( \min(a_1,a_2,a_3,a_4,a_5,a_6,a_7,a_8) \) instead of \( a \).

1) \( a_1 + a_3 + a_6 < a_2 + a_4 + a_5 \) and \( a_2 + a_4 + a_6 < a_1 + a_3 + a_5 \), \( a \) is a vertical ridge.

2) \( a_1 + a_2 + a_5 < a_4 + a_6 + a_7 \), and \( a_4 + a_6 + a_7 < a_1 + a_2 + a_5 \), \( a \) is a horizontal ridge.

3) \( a_1 + a_2 + a_4 < a_3 + a_6 + a_5 \), and \( a_3 + a_6 + a_5 < a_1 + a_2 + a_4 \), \( a \) is a diagonal ridge.

4) \( a_2 + a_3 + a_7 < a_1 + a_4 + a_6 \), and \( a_1 + a_4 + a_6 < a_2 + a_3 + a_7 \), \( a \) is a 45-degree angle ridge.

5) \( a_2 + a_3 + a_7 < a_1 + a_4 + a_6 \), and \( a_1 + a_4 + a_6 < a_2 + a_3 + a_7 \), \( a \) is a 135-degree angle ridge.

4 The area calculation of the damaged tablets

Calculate the area of the damaged tablets is mainly used it quantify the damage rate. In the process of drug quality inspection need to pick out the tablet damage rate more than 3%. In this paper, the basic unit of area is pixel, by calculating the number of valid pixels in the target area to obtain the damaged area and other geometric characteristics. Since the area calculation methods all have their advantages and disadvantages and application scope. We combine with the actual situation of medicine grain of defects, proposed two methods to calculate area.

The first method: Combined chain code contour tracking algorithm [18], devised an effective multi-region contour area calculation method.

Chain code element of the contour is determined by a line segment which has both the length and direction, and each direction has a sequential number, shown in Figure 1.

**FIGURE 1 Chain code**

After medicine grain image segmentation and binarization, the target region and the background region were respectively marked as 0 and 1. According to the idea of chain code contour tracing, use eight to the chain code tracking medicine grain of image boundary specific steps are as follows [19]:

1) Set the starting point \( b_0 \) as the first marked black dots of picture. Use \( c_0 \) represents \( b_0 \) neighbour point on the left side. Figure 2b, \( c_0 \) always background points. Starting from the \( c_0 \) along the same sequence inspect \( b_0 \) eight adjacent points, find the first black adjacent points \( b_1 \) and set \( c_1 \) (background) as the point in front of the sequence of \( b_1 \). Save the location of the \( b_0 \) and \( c_0 \), in order to use in step 4.

2) Set \( b = b_0 \) and \( c = c_1 \). Figure 2c.

3) Starting from the \( c \) travels clockwise, set \( b \) eight adjacent points as \( n_1,n_2,...,n_8 \) respectively. Find the first \( n_i \) marked 1, set \( b = n_i \) and \( c = n_{i-1} \).

4) Repeat step 3, until \( b = b_0 \) and find the next boundary point \( b_1 \).

In Figure 2, the next point to be processed is marked in black, has been processed pixels marked grey, \( c \) is always background points in step 3, because \( n_i \) is the first point value is 1 when clockwise scanning. But the algorithm also can further improve in boundary description. After finding target area, the area is marked as grey, follow-up point may also be judged, which will be duplicate detection. Mark the point has been detected as a boundary pixel can reduce
unwanted detection can further accelerate the processing speed.

The second method: “three point method”

“Three point method” is mainly for round tablets, because its approximate circle area after image segmentation, regional boundary has good connectivity, traverse the medicine image, and select the outermost (uppermost, lowermost, leftmost and rightmost) “point”. Make straight line in the direction of maximum gradient points, it is vertical line. Calculate the intersection point of straight line, the intersection is the circular medicine circle, finally statistical the number of pixels that centre to each point (up, down, left and right edge point), calculated their average that is the drug tablet radius \( r \), according to the area \( S = \pi \cdot r^2 \), we can get the medicine grain area. Set four points \( p_1(x_1, y_1) \), \( p_2(x_2, y_2) \), \( p_3(x_3, y_3) \), \( p_4(x_4, y_4) \) as shown in Figure 3, after detection \( p_1(x_1, y_1) \) is the rightmost edge point, but in theory, the rightmost point should be \( p'_1(x_1, y_1) \). So according to the criterion, the point of \( p_1(x_1, y_1) \) should be discarded. Choose the remaining three points for subsequent calculation. Determining whether the detection point is an edge point by the following criterion as in Equation (13):

\[
\begin{cases}
|x_2 - x_1| < c \\
|y_1 - y_0| < c \\
|y_2 - y_4 + x_3 - x_1| < 2c \\
|y_1 - y_3 + x_4 - x_1| < 2c
\end{cases}
\]

where \( c \) is the preset threshold, it is determined whether the point is a valid edge point.

Based on the above ideas and combined with the specific circumstances of this topic, set target pixel is white, background pixels is black. Three point method measuring area details in the following five steps:

Set \( P_1, P_2 \) are two one-dimensional arrays.

Step 1: Detecting started from the top and left of the image and gradually went on towards the down and the right, until detect the first white pixels; this point is edge point record the pixel coordinates \( p_1(x, y) \).

Step 2: After obtain the first white pixel, scanning a new line and put the first met white point (the leftmost edge points) in the array \( P_1 \). At the same line when a black pixel is scanned again to put its front white pixel coordinates into the array \( P_2 \).

Step 3: When there is no white pixel to be detected in the whole row, means the end of detecting. Then the last element of the array \( P_2 \) is the most under the lowermost edge point.

Step 4: Extract the horizontal ordinate component of the each object in the array \( P_1, P_2 \) to sort. Determining the minimum and maximum of the horizontal component, thus the left-most, rightmost, uppermost, lowermost edge point coordinates are obtained.

Step 5: According to the Equation (13), determine whether the detection point is required edge point.

**5 Experimental results and analysis**

**5.1 REGIONAL SEGMENTATION RESULTS**

The improved segmentation algorithm was respectively applied to the medicine grain image, to test the effect of image segmentation and edge detection.
Figure 4 shows the damaged medicine grain region segmentation results obtained by the improved edge detection algorithm based on local greyscale difference.

Figure 4a shows the noise free image segmentation result. Figure 4b shows the image adding 3% of Gaussian noise and 2% salt and pepper noise.

Through the experimental results it can be seen that in the case of less noise, the algorithm can segment the region of medicine grain clearly. The second group is the result of edge detection after adding noise, common medicine grain image quality is much better than that situation. The improved algorithm detects the edge more complete and clear than SUSAN algorithm detected.

![Figure 4: Original and add noise image region segmentation results](image)

![Figure 5: Irregular-shaped and normal tablets region segmentation results](image)

Medicine grain qualified analysis: after dealing with the various steps above a single medicine area has to be designated. At the same time the elliptic equation is obtained by Equations (9-11), so it only needs to traverse each element of the \( A_n(a,m,n) \) and \( A_n(b,m,n) \), if get the value greater than 1 means that the oval medicine grain is qualified, or medicine granule irregular.

Figure 6 shows the bubble region segmentation in capsules results obtained by Multi-scale adaptive Canny edge detection algorithm.

![Figure 6: Multi-scale adaptive Canny edge detection](image)

5.2 MEDICINE GRAIN OF DEFECT QUANTITATIVE RESULTS

First, use the Freeman chain code area algorithm processing the image, for the damage medicine grain of defect areas can be detected, quantification of the defect region, obtained theoretical defect area and actual detection area. According to the drug quality specification, damage rate more than 3% the tablet is unqualified. So that can further calculate damage rate and judge whether the drug tablets are qualified. The experimental results are shown in Table 1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Theoretical area</th>
<th>Actual area</th>
<th>Damage rate</th>
<th>Qualified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>147730</td>
<td>137813</td>
<td>6.71%</td>
<td>no</td>
</tr>
<tr>
<td>2.</td>
<td>151327</td>
<td>147476</td>
<td>2.54%</td>
<td>yes</td>
</tr>
<tr>
<td>3.</td>
<td>150968</td>
<td>146682</td>
<td>2.84%</td>
<td>yes</td>
</tr>
<tr>
<td>4.</td>
<td>149978</td>
<td>136597</td>
<td>8.92%</td>
<td>no</td>
</tr>
<tr>
<td>5.</td>
<td>161468</td>
<td>154567</td>
<td>4.27%</td>
<td>no</td>
</tr>
<tr>
<td>6.</td>
<td>146798</td>
<td>136892</td>
<td>6.75%</td>
<td>no</td>
</tr>
</tbody>
</table>

In order to better compare the reliability and accuracy of the two algorithms, both methods deal with the same group of tablet images. The quantitative results are shown in Table 2.
6 Conclusions

The experimental results show that the segmentation result satisfies the requirement of experiment, in order to quantify the damage of drug, we use the Freeman chain code area algorithm and “three point method” respectively calculated the defect area. Although the algorithm implementation have differences lead to concrete numerical value different, it has same results. But this article is just study the area of damaged medicine grains, other types without quantitative treatment, might exist misjudgments, so the algorithms also need to be improved. The system cannot achieve complete automation, there still needs a further improvement and enhancement.

References

[1] Zhu M, Zou W 2010 Application of image processing in tablets defect inspection Computer Engineering and Design 31(23) 5151-4
[15] Chen Q 2011 Research of image segmentation algorithm based on mathematic Morphology Dissertation for the Master Degree in Engineering of Harbin University of Science and Technology
[16] Lim Y W, Lee S U 1990 On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques Pattern Recognition 9(23) 935-52