Nearest Neighbor Recognition of Cucumber Disease Images Based on Kd-Tree

Gao Ronghua, Wu Huairui

Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing, 100097, China
National Engineering Research Center for Center for Information Technology in Agriculture, Beijing, 100097, China
Key Laboratory for Information Technologies in Agriculture, Ministry of Agriculture, Beijing, 100097, China

Abstract: Cucumber leaf images are collected during a nonstandard environment and color, shape, texture and surf characteristics are extracted by interactive segmentation. In order to organize the high dimensional characteristics of cucumber leaf image, the method of approximate nearest neighbor detection algorithm based on a Kd-tree is proposed and overcomes the inefficiency of the high dimensionality vector detection. Subspace data structure and data variance characteristics vector of cucumber disease image is calculated by the Kd-tree recursively generated and a number of pixel values closer to the adjacent area are detected because analyze connecting regional, so a nearest neighbor chain is constituted which can detect the diseases of image. Using cucumber leaves, stems, roots, fruits, seedling diseases image as experimental data sets, experimental results show that the nearest chain disease detection method based on Kd-tree is higher image precision and recall ratio.

Key words: Kd-tree, image proceeding, image segmentation, nearest neighbor link

INTRODUCTION

As a rich content, expression, intuitive multimedia information, image is an artificial visual perception enhancement form (He et al., 2011; Wu and Fu, 2011). But the research of image processing technology in the vegetables is later, agricultural technical staff break through the limitation of text description diagnosis and guide agricultural production by get growth crops images, such as leaf area, leaf perimeter, stem diameter, disease cordcrops types (Marino et al., 2008), crop nutrition deficiency category (Bokka et al., 2009) and so on, so as to avoid disaster loss because bungle the best time. Vatican (1985) analysis the infrared images of crop leaves with diseases, then confirmed pollution areas of leaves and proposed a method of diagnose damage by disease image. Machine vision theory are applied to image recognition of the vegetables pests by Pal and Pal (1993), which can improve the accuracy of disease image recognition. Then, on this basis, Panigrahi and Allen (1995) analyzed the insect pest corn fruit as an example and given accurately identify method for insect pest corn ear according to the gray change of defective parts. Sasai et al. (1999) established spectral reflection characteristics and shape characteristics using genetic algorithm and realized the automatic detection of cucumber anthroon but lack of the comprehensive utilization of color and texture information, then this method lead to low detection efficiency.

Research of vegetables disease diagnosis with image processing technology in China emerged very lately. Cui et al. (2005) compared two kinds of cucumber disease through the colorimetric system, which provides the main characteristic parameters for image processing technology in the application of cucumber disease. But the characteristics is too simplifies in this method. Tian and Li (2006) analyzed the texture image of cucumber disease and given a effective cucumber disease image classification result. Ji et al. (2001) established multi-layer fuzzy neural network and realized the plant disease species discrimination by integrated shape, texture and color information of disease plants. Conclusively, vegetables disease diagnosed automatically technology is not mature, such as cucumber diseases has diversity and cause to recognition difficultly.

In order to resolve the problem of cucumber disease automatic diagnosis, the method of nearest neighbor recognition of disease image based on Kd-tree is
proposed. All disease characterizes all are stored in a Kd-tree and detect a number of pixel values closer to the adjacent area according to connecting regional analysis, then a nearest neighbor chain is constituted which can detect the diseases of image. Using cucumber leaves, stems, roots, fruits, seedling diseases images as experimental data sets, experimental results show that the nearest chain disease detection method based on Kd-tree is higher image precision and recall ratio.

INTERACTIVE SEGMENTATION OF CUCUMBER DISEASE IMAGE

The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze (Von et al., 2013). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics (Ivan and Fernando, 2013). Let set I be a whole image region and image segmentation for I can be seen as I is divided into n number nonempty subsets I₁, I₂, ..., Iₙ which meet conditions as follows:

• \( \bigcup_{i=1}^{n} I_i = I \)

• I \( \cap I_j = \emptyset \) for all i, j, i \( \neq \) j, each area should not overlap with another after segmentation

• P (I₁) = TRUE for I = 1, 2, ..., N, each area uniqueness

• P (I₁ \( \cap \) I₂) = FALSE for i \( \neq \) j, different branch area with different characteristics

• Iᵢ be a connected region for i = 1, 2, ..., N, any two pixels is connected to each other in the same subdomain

Using Graph Cuts (Bae et al., 2011) method, then image segmentation problem is converted to the energy function which can measure front or background of node. Conversion formula is as follows:

\[
T(X) = \sum_{x \in I} T_i(x) + \lambda \sum_{x \in \partial I} T_j(x \cdot x_j) \tag{1}
\]

\( T_i(x) \) is a energy that measure node is the front or background, which calculation by front scene seed joint and background seed joint of user given some cucumber disease images in advance. If joint A (i) in the uncertain region \( \varphi \), then:

\[
\begin{align*}
T_i(x) &= \min_{\mathbf{x}} \| A(i) - K_n^i \| \\
T_i(x) &= \min_{\mathbf{x}} \| A(i) - K_n^i \|
\end{align*}
\]

Which \( m \)FIN is resolution of cucumber disease image. Thereupon, energy functions \( T_i \) is:

\[
\begin{align*}
T_i(x_i = 1) &= 0, T_i(x_i = 0) = \infty, \forall i \in F \\
T_i(x_i = 1) &= \infty, T_i(x_i = 0) = 0, \forall i \in B \\
T_i(x_i = 1) &= \frac{1}{l_i^0 + l_i^0}, T_i(x_i = 0) = \frac{1}{l_i^0 + l_i^0}, \forall i \in \varphi
\end{align*}
\]

Energy cost function:

\[
T_j(x \cdot x_j) = |x_i - x_j| g(A(i)) \tag{4}
\]

Which:

\[
g(x) = \frac{1}{1 + x}, A(i) = || A(i) - A(j) ||^2
\]

The minimum of the energy functions got by the Max Flow algorithm (Boykov and Jolly, 2001) and the cucumber disease image segmentation is realized by this method, as shown in Fig. 1, a interactive segmentation result of cucumber powdery mildew disease and bacterial angular leaf spot disease.

DETECTIONALGORITHM OF CUCUMBER DISEASE IMAGE BASED ON KD-TREE

Definition of Kd-tree: Every non-leaf node can be thought of as implicitly generating a splitting hyper plane that divides the space into two parts, known as half-spaces. So, for example, given a list of 6 points \((m_{1},n_{1}) (m_{2},n_{2}) (m_{3},n_{3}) (m_{4},n_{4}) (m_{5},n_{5}) (m_{6},n_{6})\), the space generate a Kd-tree six number by above six number points, as shown in Fig. 2.

Recent adjacent detection: Let M is a number of connected area on some cucumber disease images and the center coordinates is \((x_i, y_i)\) of connected areas \(K_i\), \(K_{i-1}\), \(K_M\) respectively, which i = 1, 2, ..., M and set up rectangular bounding box for every connected area, then the height and width of bounding box is \(h_i, w_i\) respectively and must meet the following conditions when \(K_i\) and \(K_j\) nearest:

\[
d(K_i, K_j) < (h_i + h_j) \\
h_i = h_j, w_i = w_j \\
x_i \leq x_j, \tan^{-1} \left( \frac{y_j - y_i}{x_j - x_i} \right) \leq \frac{\pi}{4}
\]

Which \( d(K_i, K_j) \) is an Euclidean distance of center points on two connected areas, \( \omega \) is an experience constant value. If there exists many multiple connected
Fig. 1(a-d): Interactive segmentation result of cucumber disease image. (a) Powdery mildew disease, (b) Segmentation result, (c) Bacterial angular leaf spot disease and (d) Segmentation result of bacterial angular leaf spot disease.

Fig. 2(a-c): Kd-tree and its space segmentation. (a) 2D Kd-tree, (b) Decomposition result and (c) 3-dimensional Kd-tree.
x pairs, then the closest connection region constitute the nearest pairs. If there not exists above conditions and the connected pairs, then the current connection regions is disturbing region and is rejected from disease image.

Extract the image feature vectors of diseases region in the nearest neighbor chain \( T = \{ t_i, i = 1, 2, \ldots, n \} \), which \( t_i \) is a point in the k-dimensional space and find m disease characteristics that have nearest distance of \( t \) in \( T \). Cucumber disease characteristics is by attribute weighted. Suppose a disease point on cucumber corresponding characteristic vector is:

\[
x = [t_1(x), t_2(x), \ldots, t_n(x)]
\]

Given a detection point and cucumber disease characteristics vector space \( O \) and defines the minimum distance "mindist" is the minimum distance from \( q \) to \( O \) as follows:

\[
|q, O| = \min \sum_{i=1}^{n} |q, x_i|^2, \forall X = (x_1, x_2, \ldots, x_n) \in O
\]

Define the function \( f \) which generate a Kd-tree from 1-k branches. Suppose \( s \) is a check point and \( L_1, L_2, \ldots, L_k \) are k number branches, which limit region of \( L_s \) is \( R_s \), then for \( \forall i, j, s \in R_i, s \not\in R_j \), have \( f(s, R_i, R_j, \ldots, R_k) \), so nearest neighbor query results are as follows:

- In each node, any \( R_i, R_j \) not intersect
- In each layer of the tree, all nodes in the collection of the boundaries of the region is a complete data space

**EXPERIMENTAL RESULTS AND ANALYSIS**

In this study, cucumber leaves, stems, roots, fruit, seedling are collected as the sample data. These image database including the typical cucumber disease of powdery mildew, downy mildew, blight and root rot disease which as analysis objects. The connection of position and disease species as show in Table 1.

Where, "\( \checkmark \)" is the disease may occur in different position and "\( \times \)" is impossible occur in this position, we can see from the Table 1, each disease are likely to be reflected in the different position. So, the description of disease characteristic is diversification and those lead to the dimension of characteristic vector is increased. Now, each disease species selected 10 images, which including leaf, stem, root, fruit and seedling. Those images were segmented firstly by the method in section 2, then time sequence from random number and query sequence which application the nearest neighbor detection algorithm based on the Kd-tree.

The time of establish Kd-tree (Ktime), the number of tree nodes (Knode), height of Kd-tree (Kheight), 20000 inquiries at the time (Qtime) in the tree as shown in Table 2.

Can be seen from Table 2, the less the number of nodes, the Kd-tree referred to in the longer linked lists, the search time increases. When the linked list of search time and overall seek time is reduced, contribution increases and the increase in the Kd-tree scale. From Fig.3, the nearest neighbor detection methods asked on Kd-tree to different cucumber disease images, which the precision rate in more than 90%, while the recall rate of 100%. Cucumber Disease image detection system interface shown in Fig. 4.
Fig. 4(a-b): System of cucumber disease detection. (a) Input interface of cucumber disease detection and (b) Results of disease detection

Table 2: Time of establish tree, No. of node, height of tree and inquiries

<table>
<thead>
<tr>
<th>Climate</th>
<th>Node</th>
<th>Height</th>
<th>Time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.202</td>
<td>520</td>
<td>19</td>
<td>0.526</td>
</tr>
<tr>
<td>1.235</td>
<td>745</td>
<td>21</td>
<td>0.048</td>
</tr>
<tr>
<td>1.356</td>
<td>4265</td>
<td>28</td>
<td>0.047</td>
</tr>
<tr>
<td>1.490</td>
<td>10240</td>
<td>32</td>
<td>0.061</td>
</tr>
<tr>
<td>1.581</td>
<td>21104</td>
<td>35</td>
<td>0.073</td>
</tr>
<tr>
<td>1.602</td>
<td>50689</td>
<td>39</td>
<td>0.061</td>
</tr>
<tr>
<td>1.697</td>
<td>53178</td>
<td>40</td>
<td>0.064</td>
</tr>
<tr>
<td>1.798</td>
<td>55890</td>
<td>42</td>
<td>0.067</td>
</tr>
</tbody>
</table>

**CONCLUSION**

Disease analysis and diagnosis by experience and pathology knowledge has been unable to solve practical problems caused by the disease type and lead to error judgment. Cucumber disease must be identified accurately in order to provide the reliable guidance for prevention.
In this study, image color, shape, texture and surf characteristics are extracted by interactive segmentation. And all disease characteristics are stored in a KD-tree and detect a number of pixel values closer to the adjacent area according to connecting regional analysis, then a nearest neighbor chain is constituted which can detect the diseases of image. Using cucumber leaves, stems, roots, fruits, seedling diseases image as experimental data sets, experimental results show that the nearest chain disease detection method based on KD-tree is higher image precision and recall ratio.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (61102126), National Key Technology Research and Development Program of the Ministry of Science and Technology of China (2011BAD21B02), Special Fund for Agro-scientific Research in the Public Interest (201303107).

REFERENCE


