http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (23): 7385-7390, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.7385.7390

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

[1,2,3]Gao Ronghua, [1,2,3]Wu Huarui
[1]Beijing Research Center for Information Technology in Agriculture,
Beijing Academy of Agriculture and Forestry Sciences, Beijing, 100097, China
[2]National Engineering Research Center for Center for Information Technology in Agriculture,
Beijing, 100097, China
[3]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 cordceps 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 anthrax 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

Corresponding Author: Wu Huarui, Beijing Research Center for Information Technology in Agriculture,
Beijing Academy of Agriculture and Forestry Sciences, Beijing, 100097, China

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 c onnecting 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.

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 $I_1, I_2,\cdots, I_n$, which meet conditions as follows:

- $\bigcup_{i=1}^{n} I_i = I$
- $I_i \cap I_j = \varnothing$ for all i, j, i ≠ j, each area should not overlap with another after segmentation
- P (I_i) = TRUEfor I = 1, 2,···, N, each area uniqueness
- P ($I_i \cup I_j$) = FALSE for i ≠ j, different branch area with different characteristics
- I_i 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:

$$T(X) = \sum_{i \in p} T_1(x_i) + \lambda \sum_{i,j \in e} T_2(x_i, x_j) \quad (1)$$

T_1 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 φ, then:

$$\begin{cases} I_i^F = \min_n \| A(i) - K_n^F \| \\ I_i^B = \min_m \| A(i) - K_m^B \| \end{cases} \quad (2)$$

Which mHn is resolution of cucumber disease image. Thereupon, energy functions T_1 is:

$$\begin{cases} T_1(x_i = 1) = 0, T_1(x_i = 0) = \infty, \forall i \in F \\ T_1(x_i = 1) = \infty, T_1(x_i = 0) = 0, \forall i \in B \\ T_1(x_i = 1) = \dfrac{I_i^F}{I_i^F + I_i^B}, T_1(x_i = 0) = \dfrac{I_i^B}{I_i^F + I_i^B}, \forall i \in \varphi \end{cases} \quad (3)$$

Energy cost function:

$$T_2(x_i, x_j) = | x_i - x_j | g(A_{i,j}) \quad (4)$$

Which:

$$g(x) = \frac{1}{1+x}, A_{i,j} = \| A(i) - A(j) \|^2 \quad (5)$$

The minimum of the energy functionis got by the Max Flow algorithm (Boycov 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.

DETECTION ALGORITHM OF CUCUMBER DISEASE IMAGE BASED ON KD-TREE

Defination 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_0, n_0) (m_1, n_1) (m_2, n_2) (m_3, n_3) (m_4, n_4) (m_5, n_5), 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_1, $K_2,\cdots, K_M$, 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:

$$\begin{aligned} & d(K_i, K_j) < \omega \cdot \min(h_i, h_j) \\ & h_i \approx h_j, w_i \approx w_j \\ & x_i \le x_j, \tan^{-1}(\left| \frac{y_j - y_i}{x_j - x_i} \right|) \le \frac{\pi}{4} \end{aligned} \quad (6)$$

Which d (K_i, K_j) is an Euclidean distance of center points on two connected areas, ω is an experience constant value. If there exists many multiple connected

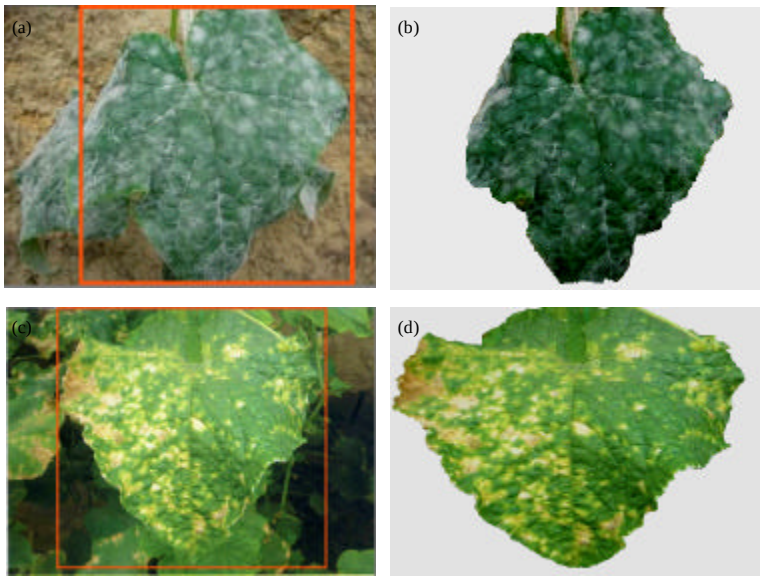

Fig. 1(a-d): Interactive segmentation result of cucumber disease image, (a) Powdery mildew disease, (b) Segmentation result, (c) Bacterial angular leaf spot disease and (d) Segmentation result of bacterial angular leaf spot disease

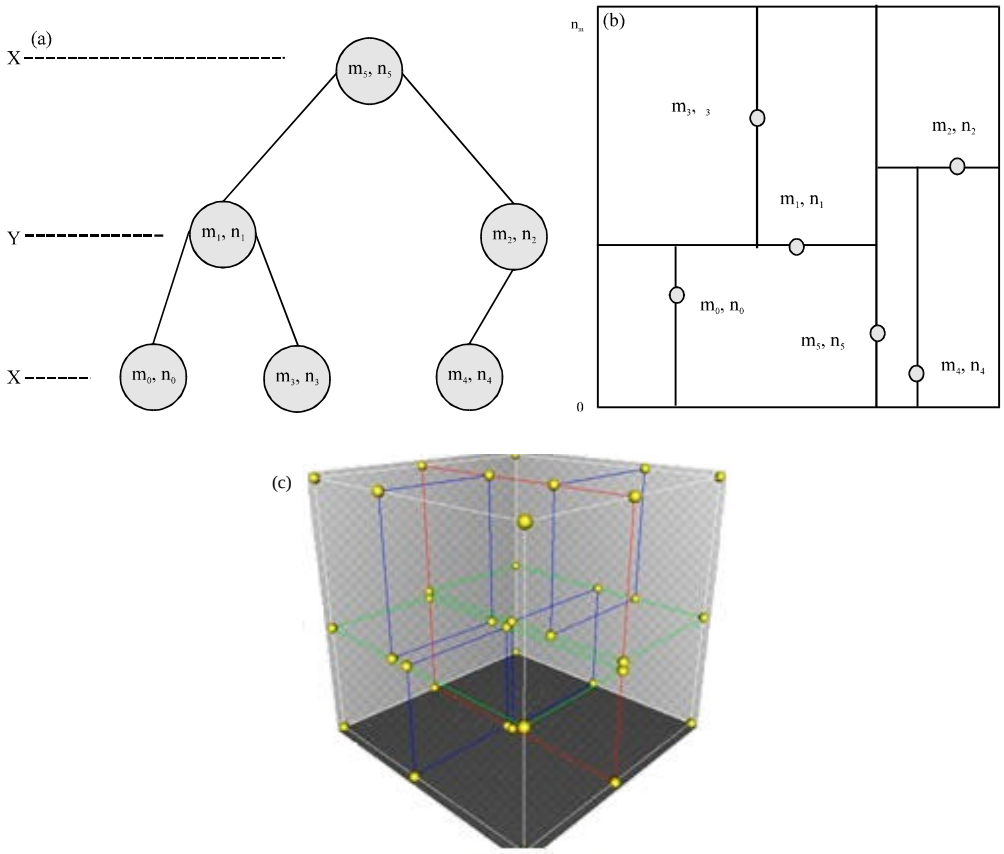

Fig. 2(a-c): Kd-tree and its space segmentation, (a) 2D Kd-tree, (b) Decomposition result and (c) 3-dimensional Kd-tree

xpairs, 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,···, 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:

$$x = [t_1(x), t_2(x), \cdots, t_n(x)] \tag{7}$$

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:

$$\| q, O \| = \min(\sum_{i=1}^{n} | q_i - x_i |^2, \forall X = \{x_1, x_2, \cdots, x_n\} \in O) \tag{8}$$

Define the function f which generate a Kd-tree from 1-k branches. Suppose s is a check point and L_1, L_2,···, L_k are k number branches, which limit region of L_i is R_i, then for ∀i, j, s∈R_i, s∉R_j, have f (s, R_1, R_2,···, 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, "√" is the disease may occur in different position and "×" 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.

Table 1: Position and disease species

	Position				
Disease	Leaves	Stems	Roots	Fruit	Seedling
Powdery mildew	√	√	×	×	√
Downy mildew	√	√	×	×	√
Blight	√	√	×	√	√
Root rot	√	√	√	×	×

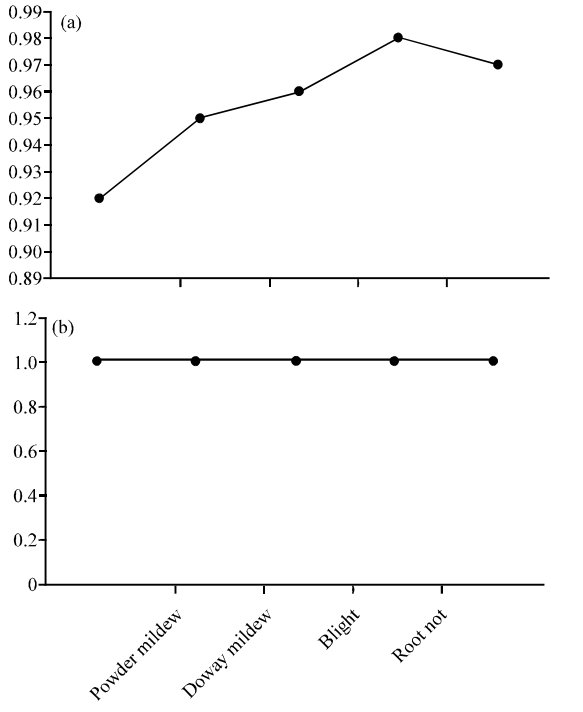

Fig. 3(a-b): Different diseases of cucumber image precision and recall, (a) Precision and (b) Recall

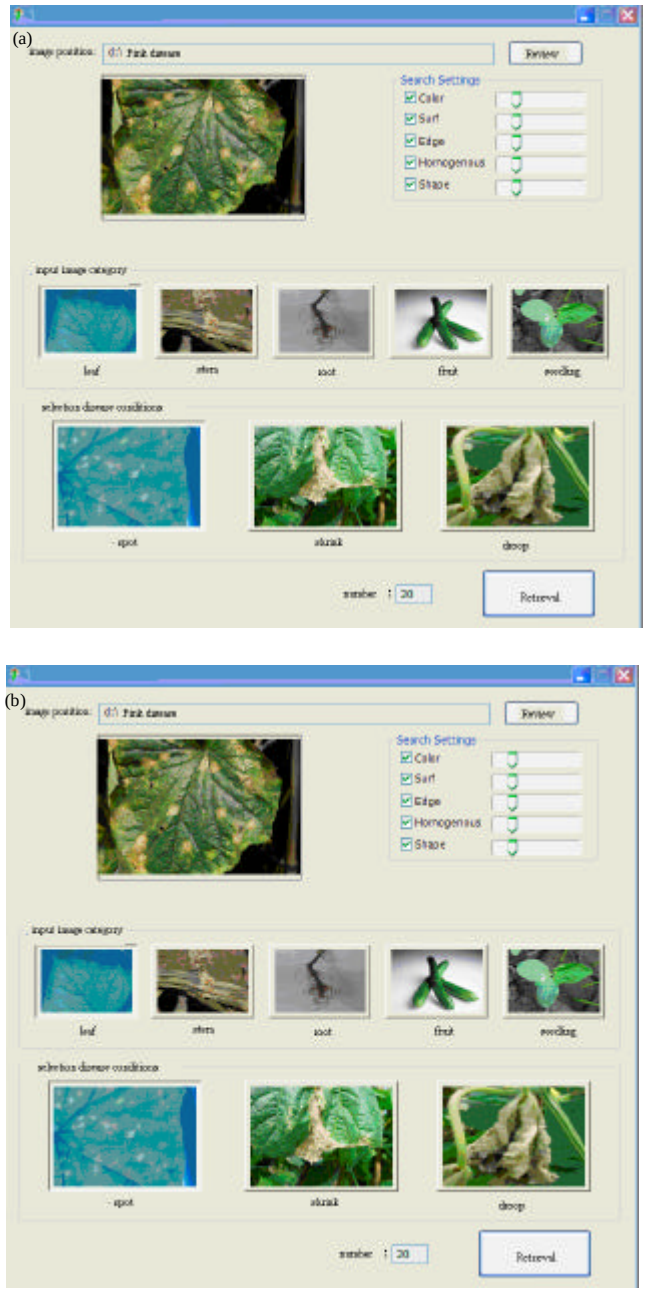

Fig. 4(a-b): System of cucumber disease detection, (a) Input interface of cucumber disease detection and (b) Results of disease detection

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

Ktime(s)	Knode	Kheight	Qtime(s)
1.202	520	19	0.526
1.235	745	21	0.048
1.356	4265	28	0.047
1.490	10240	32	0.061
1.581	21104	35	0.073
1.602	50689	39	0.061
1.697	53178	40	0.064
1.798	55890	42	0.067

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 characterizes 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

Bae, E., S. Juan and C.T. Xue, 2011. Graph cuts for curvature based image denoising. IEEE Trans. Image Process., 20: 1199-1210.

Bokka, R., L. Shaomao, W. Foshee and H. Kirkici, 2009. Pulsed electric field effects on the germination rate of yellow nutsedge seeds. Proceedings of the Pulsed Power Conference, June 28-2 July, 2009, Washington, DC., pp: 962-964.

Boycov, Y.Y. and M.P. Jolly, 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. Proceedings of the 8th IEEE International Conference on Computer Vision, July 7-14, 2001, Vancouver, BC., pp: 105-112.

Cui, Y.L., C.H.F., Peng and Dong X.Z., *et al*., 2005. Image processing and extracting color features of greenhouse diseased leaf. Trans. Chinese Soc. Agric. Eng., 21: 32-35.

He Y.F., Zhou L., Yu J.Q., T. Xu and T. Guan, 2011. Image retrieval based on locally features aggregating. Chinese J. Comput., 11: 2224-2233.

Ivan, G. D. and D.M. Fernando, 2013. A region-centered topic model for object discovery and category-based image segmentation. J. Pattern Recognit., 46: 2437-2449.

Ji S.W., R.B. Wang, J.J. Chen and X. Zhao, 2001. Research on recognizing weed from corn seedling by using computer image processing technology. Trans. Chinese Soc. Agric. Eng., 2: 154-156.

Marino, P., F.P. Fontan, M. A. Dominguez and S. Otero, 2008. Internetworking infrastructures for field sensors. Proceedings of the IEEE International Workshop on Factory Communication Systems, May 21-23, 2008, Dresden, pp: 113-116.

Pal, N.R. and S.K. Pal, 1993. A review on image segmentation techniques. Pattern Recog., 9: 1277-1294.

Panigrahi, D.A. and A. Allen, 1995. Telemedicine technology and clinical applications. JAMA, 273: 483-488.

Sasai, O., I. Kenji. K. Imou and T. Torii, 1999. Automatic diagnosis of plant disease recognition between healthy and diseased leaf. J. TSAM, 61: 119-126.

Tian, Y.W. and W.H. Li, 2006. Research on recognition of cucumber disease based on image procession in sunlight greenhouse. J. Agric. Mechanizat. Res., 2: 151-153.

Vatican, 1985. Application of image processing technology in the environment. Data of Electrical Engineers Magazine, pp: 455-458.

Von, L. T., S. Bremm, M. Kirschner, S. Wesargc and A. Kuijperd, 2013. Visual analytics for model-based medical image segmentation: opportunities and challenges. J. Exp. Sys. Appl., 40: 4934-4943.

Wu, C.H. and G.L. Fu, 2011. A stereo matching method based on k-means segmentation and neighborhood constraints relaxation. Chinese J. Comput., 4: 755-759.