The Matching Research of Strawberry Diseases Image Features Based on KD-Tree Search Method

Chen Jianshu, Wang Jianlun, Wang Shuting, Liu Hao

The Matching Research of Strawberry Diseases Image Features Based On KD-Tree Search Method

Chen Jianshu[1][1※], Wang Jianlun[1],Wang Shuting[1], Liu Hao[1]

[1] College of Information and Electrical Engineering, China Agricultural University，Beijing 100083, P. R. China

Abstract. According to the problem of the low matching accuracy rate and the low speed in the matching technology for image search, we choose the image of the strawberry diseases as the object of the research, to extract the feature of strawberry diseases image and then do cluster analysis using SPSS, and choose the feature combination according to the clustering effect. Do search matching experiments on the feature combination by KD-Tree matching algorithm respectively, and determine the characteristics combination which search matching accuracy is higher by comparing the results. This feature combination of strawberry powdery mildew matching accuracy reached 83.3%, the deformity of strawberry matching accuracy reached 60.0%, the aim is to provide the research basis for rapid disease diagnosis on strawberry.

Keywords: Feature extraction; SPSS; KD-Tree; Search matching; Feature combination

1 Introduction

Corp diseases are easily happens and spreads quickly, and once the disaster occurs it will bring a great loss to farmers. Due to the traditional crop disease and pest database and diagnosis expert system using text description of shape, color, symptoms of crop diseases and insect pests, but the text description is inaccurate and subjective, and will lead to crop pests and diseases diagnosis result deviation and slowly.

With the popularity of hardware equipment and mature technology in image processing, automatic classification, growth monitoring in fruits and vegetables and plant diseases and insect pests prevention, agricultural engineering applications are using various methods based on computer vision technology.

Among them the search based on image feature point matching is one of the most important processes of image matching,and KD-Tree matching search is a commonly used method.For example,Du Zhenpeng image matching for feature detection and matching the search time is long, they do research on the image matching algorithm based on KD-Tree search method and SURF features.The algorithm firstly extracted from SURF image features and generate a feature vector,

[1]※Chen Jianshu, the first author

[1] Corresponding author, Wang Jianlun, Email: wangjianlun@cau.edu.cn

then the vector to construct KD-Tree index for the feature description, finally calculated the distance of the nearest neighbor for each feature point of a plurality of KD-Tree and finish the work of feature matching[1].Wu Han et.al for miniature in the design of the database engine to achieve a nested query and multi-table join queries introduced KD - Tree, and on the basis of B + Tree is improved, thus to speed up the query speed[2].J.J.Aguila et.al introduced a parallel implementation of fuzzy neural network method for distributed memory architecture based on the KD tree. This method has good precision and can better realize the fuzzy neural network method based on KD tree, the results of their implementation has good speed and efficiency, particularly in terms of speed,their method is 3~20 times faster than the sequential algorithm[3].

2 Material and Method

2.1 Material

The characteristic of image content mainly includes three types: color feature, shape feature and texture feature. In this paper we choose strawberry as the object of research, one reason is strawberry is a small berry with high economic value, it has a short growth period, high nutritional value and widely grown in the greenhouse. Another reason is strawberry both have clear characteristics of color and texture. The main object of study is normal strawberry, deformity strawberry and powdery mildew of strawberry. Due to the presence of powdery mildew of strawberry also covers abnormal and normal strawberry shape, therefore, in this study only do the color, texture features extracted and research.Fig.1 shows tree type of strawberry in the research.

(a) normal (b) deformity (c) powdery mildew

Fig. 1. Conventional and spectra after pretreatment

For color feature, it mainly contains the RGB and HSI color space[4], according to the analysis results, the mean and variance of the color components of RGB and HSI color space as a characteristic parameter under construction.

For texture feature, Haralick defined 14 gray co-occurrence matrix parameters for texture analysis. But Ulaby research found that Based on 14 texture features in CLCM, only 4 characteristic is not related[5]. These four characteristics is easy to calculate and can give a higher classification accuracy, so this paper uses the following four characteristics of the most commonly used to extract texture features of images: second moment, contrast, correlation, entropy.

2.2 SPSS

SPSS is a powerful data analysis software, it is the world's first statistical analysis software, its basic functions include data management, statistical analysis, chart analysis, clustering analysis and so on, in this paper we use the SPSS software of analysis functions.

Clustering is to get a collection of physical or abstract objects into a similar object composed of multiple classes of process, the clustering generated by the cluster is a collection of data objects, these objects with the same object in a cluster are similar to each other, and other objects in different clusters. Clustering of multi-dimensional image features in this paper, analysis that can make a normal and disease samples strawberry good clustering feature descriptor.

2.3 KD-Tree

KD-Tree (K-dimension tree functions) is a kind of data structure of data points in K dimensional space division, it was first proposed by Bentle [6], through by extended a binary search tree to high dimension space, realize retrieving multiple attribute data or multidimensional data, its purpose is to search the nearest data point and the query point the distance in the KD-tree, the most commonly used nearest node search algorithm is KNN nearest neighbor node search algorithm [7].

In this paper, characteristics of strawberry image is a set of multidimensional data, mainly through the combination of similarity matching between samples, that is to find the most adjacent nodes and feature point strawberry disease image to realize the rapid diagnosis of diseases of strawberry sample type, in this study using the KD-Tree search matching method. The structure of KD-Tree description as shown in table 1:

Table 1. Description of the KD-Tree data structure

Domain name	Data Type	Description of each domain name
Node-Data	Data vector	Data from a data point, is an n-dimensional vector
Range	Space vector	The space of the node represents
Split	Integer	The axis number perpendicular to the direction of super split plane
Left	KD-Tree	A KD-Tree Composed by all data points located in the super split plane of left sub-tree
Right	KD-Tree	A KD-Tree Composed by all data points located in the super split plane of right sub-tree
Parent	KD-Tree	Parent node

KD-Tree search in a small scale space has a very good performance, but when the number of nodes searches exponentially and will increase with dimension as space, especially the dimension is larger than 10 the KD-Tree search will become very slow. When carried out the KD-Tree nearest node searching and matching, the similarity evaluation of two feature vector can be adopted by a Euclidean distance of key point's characteristic as the evaluation standard [8].

3 Experiment and Results

In this paper we uses C1, C2, C3, C4, C5, C6 respectively represent the mean and variance of color components in RGB color space , C7, C8, C9, C10, C11, C12 respectively represent the mean and variance of color components in HSI color space for color feature. We uses T1, T2, T3, T4, T5, T6, T7, T8 to respectively represent the two order moment, contrast, correlation, entropy for texture feature.

3.1 SPSS Cluster Analysis

According to the color histogram and gray level co-occurrence matrix[9-10] and other methods of strawberry, powdery mildew of strawberry, deformity strawberry and normal strawberry image (for comparison) to extract color features and texture features, each feature clustering results(when the feature quantity of correct classification rate over 60 percent,we think it has a good performance in clustering) for each type of sample (in the divided into a kind of sample volume of the largest class definition for correct classification) statistical results as shown in Table 2, table 3, table 4:

Table 2. The clustering results of normal strawberry characteristic quantity

Feature quantity	Number of samples	Number of classification	Number of successful classification	Number of unsuccessful classification	Correct classification rate
C1	30	13	5	25	16.7%
C2	30	8	10	20	33.3%
C3	30	10	10	20	33.3%
C4	30	4	16	14	53.3%
C5	30	10	10	20	33.3%
C6	30	3	29	1	96.7%
C7	30	6	13	17	43.3%
C8	30	6	15	15	50.0%
C9	30	8	13	17	43.3%
C10	30	16	7	23	23.3%
C11	30	15	5	25	16.7%
C12	30	22	2	28	6.7%
T1	30	5	24	6	80.0%
T2	30	13	13	17	43.3%
T3	30	13	10	20	33.3%
T4	30	11	12	18	40.0%
T5	30	6	15	15	50.0%
T6	30	4	26	4	86.7%
T7	30	14	6	24	20.0%
T8	30	7	22	8	73.3%

From the table 2, it is clear that four feature quantities(Namely C6,T1,T6,T8) reach a good performance in clustering.

Tab 3. The clustering results of deformity strawberry characteristic quantity

Feature quantity	Number of samples	Number of classification	Number of successful classification	Number of unsuccessful classification	Correct classification rate
C1	30	5	16	14	53.3%
C2	30	17	6	24	20.0%
C3	30	15	6	24	20.0%
C4	30	10	8	22	26.7%
C5	30	13	7	23	23.3%
C6	30	13	7	23	23.3%
C7	30	2	29	1	96.7%
C8	30	9	18	12	60.0%
C9	30	10	8	22	26.7%
C10	30	16	2	25	6.7%
C11	30	17	3	27	10.0%
C12	30	3	27	3	90.0%
T1	30	15	7	23	23.3%
T2	30	11	8	22	26.7%
T3	30	9	8	22	26.7%
T4	30	15	9	21	30.0%
T5	30	5	16	14	53.3%
T6	30	11	12	18	40.0%
T7	30	16	3	27	10.0%
T8	30	2	28	2	93.3%

From the table 3, it is clear that four feature quantities(Namely C7,C8,C12,T8) reach a good performance in clustering.

Tab 4. The clustering results of powdery mildew of strawberry characteristic quantity

Feature quantity	Number of samples	Number of classification	Number of successful classification	Number of unsuccessful classification	Correct classification rate
C1	30	10	6	24	20.0%
C2	30	6	12	18	40.0%
C3	30	7	11	19	36.7%
C4	30	7	15	15	50.0%
C5	30	8	8	22	26.7%
C6	30	4	19	11	63.3%
C7	30	8	17	13	56.7%
C8	30	12	6	24	20.0%
C9	30	6	10	20	33.3%
C10	30	7	15	15	50.0%
C11	30	5	12	18	40.0%
C12	30	7	19	11	63.3%
T1	30	15	4	26	13.3%
T2	30	16	3	27	10.0%
T3	30	17	5	25	16.7%
T4	30	12	5	25	16.7%
T5	30	10	9	21	30.0%
T6	30	8	18	12	60.0%
T7	30	14	5	15	16.7%

T8	30	2	29	1	96.7%

From the table 4, it is clear that four feature quantities(Namely C6,C12,T6,T8) reach a good performance in clustering.

3.2 KD-Tree Search Matching

In this paper we uses 90 strawberry samples(including normal strawberry, strawberry of powdery mildew and deformity strawberry, each type samples includes 30 samples) selected three features to build 3D-Tree, three features to be matched samples of input KD-Tree search matching program. Here similarity evaluation for the two feature vector Euclidean distance metric is a key point features can be used as the evaluation standard. Because the feature vector consists of three parts, and can be set to the feature representing the two nodes. The Euclidean distance between them as follows:

$$d(x_i + x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + (x_{i3} - x_{j3})^2} \quad (15)$$

Matching results can be obtained by calculating the Euclidean distance nearest neighbor search for matching nodes, and analysis of a representative sample of the node type, whether right or wrong type of samples (mainly through the sample to be matching matches consistent criteria for judging), deformity strawberry and powdery mildew of strawberry search for matching results are shown in table 5, table 6 below:

Tab 5. The matching results of deformity strawberry

Sample	Nearest neighbor node coordinates			Euclidean distance	The Search nodes and sample type	
	X-Axis	Y- Axis	Z- Axis			
sample1	0.151583	0.114053	0.487964	0.0498854	62	Normal
sample2	0.16766	0.137424	0.872836	0.0424372	70	Deformity
sample 3	0.146195	0.117112	0.367566	0.0249852	57	Deformity
sample 4	0.276683	0.212421	0.234173	0.0446088	37	Powdery
sample 5	0.214051	0.169284	0.594228	0.118599	69	Normal
sample 6	0.107811	0.102575	0.164185	0.016096	44	Deformity
sample 7	0.152462	0.106678	0.450741	0.0140581	59	Deformity
sample 8	0.169123	0.135545	0.502643	0.0278196	63	Deformity
sample 9	0.082651	0.0960127	0.101104	0.00424334	36	Deformity
sample 10	0.180979	0.151676	0.526856	0.0674990	69	Powdery
sample 11	0.132939	0.101810	0.349519	0.0580199	49	Deformity
sample 12	0.142896	0.142075	0.496051	0.0278196	60	Deformity
sample 13	0.130797	0.113844	0.327825	0.014884	50	Normal
sample 14	0.157395	0.130031	0.409041	0.0270746	59	Normal
sample 15	0.130551	0.0964061	0.160626	0.0118761	55	Normal
sample 16	0.102352	0.090021	0.172653	0.0160960	56	Deformity
sample 17	0.16314	0.132223	0.830962	0.0424372	70	Deformity
sample 18	0.13495	0.125852	0.335578	0.0275788	61	Deformity
sample 19	0.143318	0.107794	0.390569	0.0249852	57	Deformity
sample 20	0.122062	0.107776	0.0855459	0.0253014	41	Powdery
sample 21	0.130797	0.113844	0.327825	0.0249000	51	Normal
sample 22	0.162125	0.123367	0.0777855	0.0301397	65	Normal

sample 23	0.146915	0.113122	0.0545784	0.0322891	52	Normal
sample 24	0.116402	0.10256	0.325141	0.00541093	45	Normal
sample 25	0.163326	0.113567	0.456412	0.0140571	53	Deformity
sample 26	0.0846707	0.0992257	0.103003	0.00424334	35	Deformity
sample 27	0.160056	0.131389	0.325597	0.0307228	74	Deformity
sample 28	0.0846707	0.0992257	0.103003	0.0306622	45	Deformity
sample 29	0.160056	0.131389	0.325597	0.0175819	62	Deformity
sample 30	0.180978	0.172669	0.537890	0.0237166	53	Deformity

Tab 6. The matching results of powdery mildew of strawberry

Sample	Nearest neighbor node coordinates			Euclidean distance	The Search nodes and sample type	
	X- Axis	Y- Axis	Z-Axis			
Sample1	0.150491	0.106692	0.0466092	0.00340927	42	Powdery
Sample 2	0.139989	0.109184	0.0182249	0.0186919	47	Powdery
Sample 3	0.133554	0.120663	0.246202	0.0125674	45	Powdery
Sample 4	0.203901	0.164683	0.381373	0.0211159	41	Normal
Sample 5	0.130551	0.0964061	0.160626	0.0189952	54	Normal
Sample 6	0.175663	0.166097	0.49239	0.0390932	60	Powdery
Sample 7	0.149334	0.109777	0.0474865	0.00820598	40	Powdery
Sample 8	0.121279	0.105738	0.0778598	0.00799019	37	Powdery
Sample 9	0.143153	0.119272	0.0277412	0.0106403	42	Powdery
Sample 10	0.153292	0.145267	0.282223	0.0284982	66	Normal
Sample 11	0.127353	0.112451	0.242179	0.0251362	58	Powdery
Sample 12	0.0584718	0.119704	0.0186351	0.0106403	46	Powdery
Sample 13	0.117785	0.095876	0.0486139	0.0069166	39	Powdery
Sample 14	0.170719	0.141748	0.0251527	0.0250481	43	Powdery
Sample 15	0.149337	0.109777	0.0474865	0.00340927	32	Powdery
Sample 16	0.20153	0.157729	0.492711	0.0271892	55	Normal
Sample 17	0.180979	0.151676	0.526856	0.0237166	55	Powdery
Sample 18	0.146915	0.113122	0.0545784	0.0115159	44	Powdery
Sample 19	0.171319	0.132066	0.0282659	0.0101883	42	Powdery
Sample 20	0.133554	0.120663	0.246202	0.0110485	41	Powdery
Sample 21	0.133554	0.120663	0.246202	0.0110485	41	Powdery
Sample 22	0.273167	0.201867	0.277682	0.0449088	36	Deformity
Sample 23	0.127353	0.112451	0.242179	0.0110485	32	Powdery
Sample 24	0.170719	0.141748	0.0251527	0.0101883	44	Powdery
Sample 25	0.120286	0.0922447	0.0432852	0.0069116	36	Powdery
Sample 26	0.14864	0.119704	0.0186351	0.0136266	47	Powdery
Sample 27	0.196121	0.169159	0.0556842	0.0141064	36	Powdery
Sample 28	0.122062	0.107776	0.0855459	0.00799019	38	Powdery
Sample 29	0.197011	0.15916	0.0457738	0.0141064	38	Powdery
Sample 30	0.121279	0.105738	0.0778598	0.0147406	54	Powdery

For the deformity strawberry image feature points to the adjacent node searching, the number of matching correct results is 18, incorrect is 12, the search matching accuracy is 60%; For each samples of powdery mildew of strawberry image feature points to the adjacent node searching ,the number of matching the correct result is 25, incorrect is 5,the search matching accuracy is 83.3%.

4 Discussion and Conclusion

4.1 Discussion

In this paper we uses KD-Tree search algorithm to carry on the adjacent node search for each samples of deformity strawberry and powdery mildew of strawberry image feature points, as it can be seen from the search nodes of the method has excellent speed and better matching accuracy. For higher classification accuracy characteristics to a large extent on behalf of the samples in the characteristic quantities have similar common, so a comprehensive analysis of the classification accuracy of each characteristic quantity, the final choice C6, C12, T8 is carried out as a combination of characteristics Search Match study.

However, we found that when searching neighboring nodes for most powdery mildew of strawberry image feature points, Euclidean distance are focused on the range of 0.00340927 to 0.0147406 in the correct match result and the average search node is 43.2,but when searching neighboring nodes for deformity strawberry image feature points, the Euclidean distance range are more dispersed in the correct match result and the average search node is 55.8.Through by study the size of characteristics we know that the reason why the above conditions are mainly attributed to different types of strawberries manifested differences of amount between the each type of feature.

4.2 Conclusion

From the above matching results, select three characteristic quantities(C6, C12, T8) as the searching and matching feature combination of strawberry powdery mildew achieve a better search result for matching accuracy, and the search matching accuracy for strawberry powdery mildew is 83.3%, the deformity of strawberry matching accuracy reached 60.0%,the number of nodes of powdery mildew strawberry the adjacent nodes search to be less than deformity strawberry, and it can be seen from the search nodes of the method has excellent speed and better matching accuracy but search for deformity strawberry rate has yet to be matched by choosing some other combination of improved features.

The reason for searching and matching of three-dimensional data instead of higher dimensional in this paper is for various characteristic quantities manifested classification characteristics by the cluster analysis through SPSS, but in order to improve search matching accuracy rate we can manually add a dimensional data and by changing the size of the data to improve search matching accuracy rate and the expense is to search for matching speed will be slower than before.

References

1. Du Zhengpeng, Li Dehua,et al. Research on image matching algorithm based on KD-Tree search and SURF [J].Computer and digital engineering , 2012, 40(2):296~298.
2. Wu Han,Yang Kejian et al. Based on KD tree multidimensional index in a database application [J]. Automation Technology and Application.2007,26(9):37~39.

3. J.J.Aguilaa, E.Ariasb, M.M.Artigaoc, J.J.Mirallesc. A distributed memory implementation of the False Nearest Neighbors method based on kd-tree applied to electrocardiography. Procedia Computer Science 1, 2012：2579~2587.
4. Xi Zhihua.Research and application of intelligent fault diagnosis system of wheat leaf disease[M].He Fei:China University of science and technology,2010.
5. Haralick.Statistical and Structural Approaches to Texture[J]. Proc. IEEE,1979,67:786~804.
6. J.L.Bentley.Multi dimensional binary search trees used for associative searching[J].Communications of the ACM, 1975:509~217.
7. Ping Xueliang,Xu Lirong. Huge amounts of data K neighboring new algorithm based on space partition[J]. Journal of South China University of Technology. 2007,(35):65~70.
8. Zhang Yane,Li Minzan,Zhang Xijie,et al. The greenhouse cucumber leaf nutrition information detection based on computer vision technology.Journal of Agricultural Engineering[J]. 2005,21(8):102~105.
9. Wang Juan,Kong Bing,Jia Qiaoli. Image retrieval techniques based on color feature[J]. 2011,20(7):160~164.
10. Gao Chengcheng,Hui Xiaowei. Texture feature extraction based on gray level co-occurrence matrix[J]. Computer System Application.2010,(19):195~198.
11. Xiong Yunyan,Mao Yijun, et al. Application of the ordered KD-Tree on the image features matching[J]. Chemical IndustryInstrument and Automation. 2010，37(10)：84~87.
12. Liu Zhongwei,Zhang Yujin. Comprehensive utilization of color and texture features for image retrieval. Journal of Communication. 1999,20(5):36~40.
13. Xiao Peng,Xu Jun,Chen Shaochong. Texture feature extraction method[J].Electronics technology. 2010,(6):49~52.
14. Wu Funing,Yang Zibiao,Zhu Hong,Zheng Limin, et al. Review of the application of crop image classification and recognition based on color features[J]. Review of China Agricultural Science and technology. 2003,5(2):76~79.
15. Zou Xiuguo. Crop diseases and insect pests recognition based on computer vision research status[J]. Computer Application. 2011,20(6):238~242.
16. Tan Feng,Ma Xiaodan. Identification method based on the leaves of plant pests and diseases[J]. Journal of Agricultural Mechanization Research. 2009,(6):41~43.