Edge Geometric Measurement Based Principal Component Analysis in Strawberry Leaf Images

Jianlun Wang, Yu Han, Zetian Fu, Daoliang Li, Jianshu Chen, Shuting Wang

Edge Geometric Measurement Based Principal Component Analysis in Strawberry Leaf Images

Jianlun Wang[1]*, Yu Han*, Zetian Fu, Daoliang Li, Jianshu Chen, Shuting Wang

(College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, P.R.China)

Abstract. Edge geometric measurement analysis is an important method of image understanding and portraying the target feature. In this paper, we compress 17 interrelated shape descriptors which are based on edge geometric measure into 6 independent components, and discuss their meanings by using principal component analysis. The analyses in this article provide guidance for the shape feature optimization and accurate identification for greenhouse strawberry leaves images successfully.

Key word: Greenhouse strawberry leaves; Shape feature description; Principal Component Analysis; Edge geometric measure

1. Introductions

The performance comparison of several common description of shapes which are used in identify the leaves of strawberry grown in the greenhouse。Strawberries have a short growth cycle, less disease, easy to manage, high nutritional value, economic value advantages, widely grown in the greenhouse [1]。With the popularity of hardware and image processing technology matures, many field of agricultural engineering applications[6] have used a variety of techniques based on computer vision, such as automatic classification of the fruits and vegetables[2], quality traceability[3], the robot picking[4], growth status monitoring, and the early warning of pest and disease[5]. Among them, the target recognition are facing with many problems of different varieties of agricultural crops, complexity of the imaging

* These authors contributed equally to the article

1 Corresponding author. Tel: +86 13691008152. E-mail address: wangjianlun@cau.edu.cn (J. Wang).

background, range of issues such as characterization as a vital process in image processing, so it is difficult to form a high recognition rate of the general algorithm. Therefore, this paper according to greenhouse strawberry leaves, summarized 17 ways of shape description. On the high dimensional feature dimension reduction is compressed into 6 comprehensive index by using the principal component analysis , and has carried on the detailed explanation to provide guidance in getting more effective characteristics selection and comprehensive evaluation in practical application.

2. The shape descriptors in target region

The main difficulties in shape quantitative description is the lack precise and uniform definition. ASM think that shape parameter which are used to describe the micro-structure have some common in no dimension, quantitative description and sensitive to shape change[7]. This paper will divide geometry characteristic which are commonly used into the following several forms：Geometric description of the target boundary（Including the ratio between perimeter and area, fineness ratio, eccentricity, etc）、Measure of the target area and its surround polygon（Including Surrounded by rectangular compact, surrounded by polygonal compact degree, etc）、Geometric description of the target area（Including Centroid offset rate, Area symmetry, Circumference symmetry, etc）、Invariant moments（Hu Invariant moments）。We will give explanation about it one by one below:（Among them，A represents the area， represents the circumference，L represents the length of bounding rectangle，W represents the width of bounding rectangle，Aconvex is reprensents target convex hull area）：

（1） Perimeter area ratio：*P/A*，Range in (0,1)。

（2） Fineness ratio：*4*Pi*A/(P*P)*，Range in (0,1)。Get the maximum values(1) while contiguous area is round。

（3） Exterior ratio：*W/L*，Used to describe the shape of the target after plastic deformation, Take the maximum value(1) of a square and circular target and it will be reduced with slender degree about slender target。

（4） Eccentricity：The eccentricity of target which is equivalent to an ellipse(E), and the equivalent ellipse is defined as a voal which has the same centroid and secondary center moment with target。

（5） Skeleton eccentricity：Eccentricity of skeleton equivalent ellipse (E)，and skeleton equivalent ellipse is defined as a fitting ellipse which has the same skeleton with target。Generally,In order to avoid the influence of finely branch in skeleton, the first to do is deburring skeleton and then fitting ellipse。

（6） Compactness of bounding rectangle：*A/(L*W)*，Reflects the bump on the bounding rectangle of the target area.

（7） Compactness of bounding polygon：*A/Aconvex*，Reflects the bump on the minimum convex hull of the target area。

（8） Centroid offset rate：The ratio of the distance between centroid to boundary in the upper-left corner and lower right corner to the upper left corner,it reflects the relationship of the centroid position in the target area。

（9） Area symmetrical degrees：*min(Aleft, Aright)/max(Aleft, Aright)*, Aleft and Aright are represents the area size of left and right side of symmetry axis , regional symmetry in the area. It's the performance of regional symmetry in the area.If the region is completely symmetrical ,the maximum value will be 1.

（10） Perimeter symmetric degrees ：*min(Pleft,Pright)/max(Pleft,Pright)*,Pleft and Prightare represents the perimeter size of left and right side of symmetry axis , regional symmetry in the area. It's the performance of regional symmetry in the perimeter.If the region is completely symmetrical ,the maximum value will be 1.

（11） Hu Invariant moments：Ming-Kuei Hu apply algebraic invariant moment to pattern recognition and regard seven moment features about rotation and scaling translation invariant in the target area as the target shape characteristics。

3. Statistical analysis about the descriptor of shape

Statistical frequency of the above shape descriptor as follows to show：

Table 1. Statistical description of shape feature

Shape descriptors	Mean	Standard Error	Standard deviation	Coefficient of variation
Perimeter area ratio	0.009293	0.000254	0.002365	0.25451
Fineness ratio	0.654973	0.005491	0.051218	0.078199
Appearance ratio	0.755404	0.010079	0.094009	0.124449
Eccentricity	0.637541	0.012837	0.119733	0.187805

Skeleton eccentricity	0.664242	0.01463	0.136462	0.20544
Bounding rectangle compactness	0.774389	0.000705	0.006572	0.008487
Bounding polygon compactness	0.957814	0.001167	0.010888	0.011367
Centroid offset rate	0.521226	0.006561	0.061194	0.117403
Area symmetric rate	0.963479	0.011393	0.106268	0.110296
Perimeter symmetric rate	0.961412	0.002981	0.027803	0.028919
Hu1 Moment	0.206821	0.001691	0.015774	0.076269
Hu2 Moment	0.003439	0.000557	0.005196	1.511056
Hu3 Moment	9.22E-06	3.03E-06	2.83E-05	3.069955
Hu4 Moment	8.95E-06	3.02E-06	2.82E-05	3.147928
Hu5 Moment	8.67E-10	5.99E-10	5.59E-09	6.445866
Hu6 Moment	5.25E-07	2.61E-07	2.43E-06	4.631451
Hu7 Moment	6.24E-09	3.06E-09	2.86E-08	4.579208

4. Principal component analysis of the shape factor

Principal component analysis is a statistical method which is first proposed by Pearson in 1901, and then developed by Hotelling (1933). Its main purpose is that to reduce the number of variables, and make change for the formation of a linear combination of a few independent variables (principal components), but by the difference of the linear combination of ingredients into the largest, makingthe original multi-dimensional characteristics of these principal components show the greatest individual differences. In short, Principal component analysis is a number of variables into a few principal components (i.e. integrated variable) dimension reduction statistical method.

The general steps are as follows:

(1) The standardization of original data

First of all to the standardization of original data, each index data of the sample is mean 0, variance 1.

$$\chi_{ij} = \left(\chi_{ij} - \overline{\chi_j}\right) / \sqrt{\sigma_j} \quad (i = 1,2,\cdots n,\ j = 1,2,\cdots 17)$$

χ_{ij} as the i sample of the j shape descriptor value, $\overline{\chi_j}$ as mean value of the j shape descriptor value, σ_j as variance of the j shape descriptor value. In the principal component analysis of SPSS statistical software, standardization is automatically executed.

(2) Calculate the indicators of the correlation matrix R

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)^2 \sum_{k=1}^{n}(x_{kj} - \bar{x}_j)^2}} \quad (i = 1,2,\cdots,n,\ j = 1,2,\cdots 17)$$

r_{ij} as matrix of I rows and j columns representing related elements, namely the index I and index J correlation coefficient.

(3) Seek the correlation matrix R eigenvalue and characteristic vector

The characteristic equation is expressed as $|\lambda I - R| = 0$, Commonly used Jacobi method to calculate the eigenvalues and the corresponding eigenvectors, and the size of the eigenvalues in descending order. The correlation matrix are given below:

Table 2. The correlation matrix of shape characteristics about SPSS

	P/Area	Fineness ratio	Appearance ratio	Eccentricity	Skeleton eccentricity	Bounding rectangle compactness	Bounding polygon compactness	Centroid offset rate	Area symmetric rate	Perimeter symmetric rate	Hu1 invariant moment	Hu2 invariant moment	Hu3 invariant moment	Hu4 invariant moment	Hu5 invariant moment	Hu6 invariant moment	Hu7 invariant moment
P/Area	1.000	-0.390	-0.493	0.484	0.394	-0.457	-0.603	0.047	-0.246	-0.087	0.453	0.297	0.191	0.197	0.121	0.156	0.162
Fineness ratio	-0.390	1.000	0.586	-0.558	-0.339	0.377	0.685	0.084	0.095	0.347	-0.357	-0.212	-0.296	-0.294	-0.188	-0.255	-0.243
Appearance ratio	-0.493	0.586	1.000	-0.983	-0.643	0.315	0.404	0.109	0.114	0.041	-0.540	-0.394	-0.306	-0.310	-0.238	-0.291	-0.280
Eccentricity	0.484	-0.558	-0.983	1.000	0.591	-0.309	-0.400	-0.130	-0.113	-0.045	0.524	0.326	0.261	0.264	0.197	0.245	0.235
Skeleton eccentricity	0.394	-0.339	-0.643	0.591	1.000	-0.240	-0.388	-0.012	-0.154	-0.104	0.479	0.424	0.247	0.253	0.211	0.261	0.237
Bounding rectangle compactness	-0.457	0.377	0.315	-0.309	-0.240	1.000	0.536	-0.282	0.540	0.157	-0.547	-0.368	-0.604	-0.606	-0.461	-0.532	-0.540
Bounding polygon compactness	-0.603	0.685	0.404	-0.400	-0.388	0.536	1.000	-0.022	0.320	0.338	-0.438	-0.216	-0.226	-0.222	-0.106	-0.161	-0.168
Centroid offset rate	0.047	0.084	0.109	-0.130	-0.012	-0.282	-0.022	1.000	0.057	-0.011	0.084	-0.023	0.294	0.295	0.275	0.283	0.289
Area symmetric rate	-0.246	0.095	0.114	-0.113	-0.154	0.540	0.320	0.057	1.000	-0.016	-0.222	-0.282	-0.096	-0.097	-0.071	-0.096	-0.084
Perimeter symmetric rate	-0.087	0.347	0.041	-0.045	-0.104	0.157	0.338	-0.011	-0.016	1.000	-0.040	0.085	-0.291	-0.289	-0.267	-0.272	-0.282
Hu1 invariant moment	0.453	-0.357	-0.540	0.524	0.479	-0.547	-0.438	0.084	-0.222	-0.040	1.000	0.779	0.331	0.337	0.274	0.329	0.317
Hu2 invariant moment	0.297	-0.212	-0.349	0.326	0.424	-0.368	-0.216	-0.023	-0.282	0.085	0.779	1.000	0.074	0.078	0.110	0.160	0.100
Hu3 invariant moment	0.191	-0.296	-0.306	0.261	0.247	-0.604	-0.226	0.294	-0.096	-0.291	0.331	0.074	1.000	0.999	0.926	0.960	0.979
Hu4 invariant moment	0.197	-0.294	-0.310	0.264	0.253	-0.606	-0.222	0.295	-0.097	-0.289	0.337	0.078	0.999	1.000	0.927	0.961	0.980
Hu5 invariant moment	0.121	-0.188	-0.238	0.197	0.211	-0.461	-0.106	0.275	-0.071	-0.267	0.274	0.110	0.926	0.927	1.000	0.985	0.981
Hu6 invariant moment	0.156	-0.255	-0.291	0.245	0.261	-0.532	-0.161	0.283	-0.096	-0.272	0.329	0.160	0.960	0.961	0.985	1.000	0.989
Hu7 invariant moment	0.162	-0.243	-0.280	0.235	0.237	-0.540	-0.168	0.289	-0.084	-0.282	0.317	0.100	0.979	0.980	0.981	0.989	1.000

(4) Calculate the variance contribution rate and cumulative variance contribution rate

Contribution rate:

$$\frac{\lambda_i}{\sum_{k=1}^{p}\lambda_k}(i=1,2,\cdots,p)$$

The cumulative variance contribution rate:

$$\frac{\sum_{k=1}^{i}\lambda_k}{\sum_{k=1}^{p}\lambda_k}(i=1,2,\cdots,p)$$

Generally it takes the cumulative contribution rate of 85% ~ 95% m eigenvalues corresponding to the former m variables as main components, number of principal components for M.

Variance and the cumulative variance contribution rate as shown in the following table:

Table 3. Variance contribution rate and cumulative variance contribution rate of shape features about SPSS.

Total Variance Explained		
Component	Initial Eigenvalues	Extraction Sums of Squared Loadings

	Total	%of Variance	Cumulative %	Total	%of Variance	Cumulative %
1	6.783	39.901	39.901	6.783	39.901	39.901
2	3.334	19.614	59.515	3.334	19.614	59.515
3	1.504	8.847	68.362	1.504	8.847	68.362
4	1.405	8.264	76.626	1.405	8.264	76.626
5	0.933	5.489	82.115	0.933	5.489	82.115
6	0.848	4.985	87.100	0.848	4.985	87.100
7	0.608	3.578	90.679	0.608	3.578	90.679
8	0.535	3.145	93.824	0.535	3.145	93.824
9	0.409	2.408	96.232	0.409	2.408	96.232
10	0.263	1.545	97.777	0.263	1.545	97.777
11	0.191	1.123	98.901	0.191	1.123	98.901
12	0.114	0.674	99.574	0.114	0.674	99.574
13	0.052	0.308	99.882	0.052	0.308	99.882
14	0.011	0.065	99.947	0.011	0.065	99.947
15	0.007	0.042	99.989	0.007	0.042	99.989
16	0.001	0.007	99.996	0.001	0.007	99.996
17	0.001	0.004	100.00			

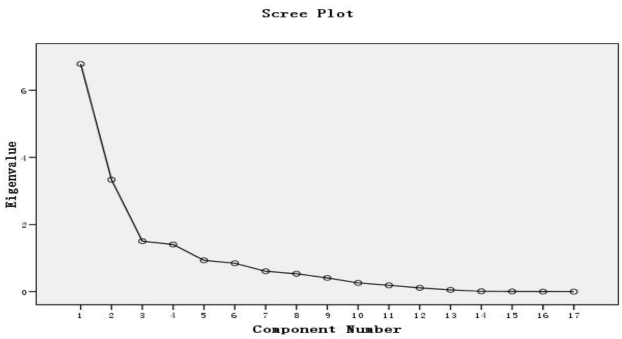

Figure 1. The SPSS components eigenvalue scatter diagram

(5) Determine the principal components

From cumulative variance contribution ratio of table 3 and each component eigenvalue scatter diagram of figure 1 show that 1-6 cumulative variance contribution rate has reached 87.1 percent, After the sixth component feature value is going to be more smaller, and the transform slow. Therefore, here take first 6 comprehensive variables as main component.

(6) Determine the principal component loading matrix L

$$l_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} e_{ij} (i, j = 1,2,\cdots, p)$$

In this paper data loading matrix results are as follows:

Table 4. Principal component loading matrix L about SPSS

Component Matrix(a)						
Shape descriptor	Component					
	F1	F2	F3	F4	F5	F6
Perimeter area ratio	0.502	-0.492	0.070	-0.208	0.225	-0.243
Fineness ratio	-0.560	0.410	0.387	0.314	-0.026	-0.061
Appearance ratio	-0.633	0.550	0.320	-0.314	0.003	0.208
Eccentricity	0.594	-0.568	-0.322	0.295	-0.011	-0.221
Skeleton eccentricity	0.536	-0.449	-0.113	0.272	0.055	0.071
Bounding rectangle compactness	-0.745	0.059	-0.430	0.302	0.018	0.125
Bounding polygon compactness	-0.536	0.476	0.055	0.564	-0.120	0.004
Centroid offset rate	0.223	0.357	0.318	-0.070	0.772	-0.103
Area symmetric rate	-0.274	0.243	-0.569	0.334	0.481	0.257
Perimeter symmetric rate	-0.314	-0.121	0.406	0.548	0.033	-0.499
Hu1 Moment	0.643	-0.430	0.328	0.215	0.102	0.338
Hu2 Moment	0.395	-0.464	0.487	0.308	-0.041	0.475
Hu3 Moment	0.860	0.479	-0.048	0.024	-0.047	-0.049
Hu4 Moment	0.863	0.476	-0.045	0.030	-0.045	-0.050
Hu5 Moment	0.793	0.545	-0.035	0.123	-0.093	0.019
Hu6 Moment	0.846	0.495	-0.021	0.110	-0.080	0.022
Hu7 Moment	0.843	0.518	-0.039	0.082	-0.069	-0.011

(7) Calculate the feature vectors to determine the linear expression of the principal component index of the original data.

Table 5. Characteristic vector of correlation matrix about SPSS

	Eigenvectors					
	F1	F2	F3	F4	F5	F6
P/Area	0.19	-0.27	0.06	-0.18	0.23	-0.26
Fineness ratio	-0.22	0.22	0.32	0.26	-0.03	-0.07
Min/MaxLen	-0.24	0.30	0.26	-0.26	0.00	0.23
Eccentricity	0.23	-0.31	-0.26	0.25	-0.01	-0.24
Skeleton eccentricity	0.21	-0.25	-0.09	0.23	0.06	0.08
Bounding rectangle compactness	-0.29	0.03	-0.35	0.25	0.02	0.14
Bounding polygon compactness	-0.21	0.26	0.04	0.48	-0.12	0.00
Centroid offset rate	0.09	0.20	0.26	-0.06	0.80	-0.11
Area symmetric rate	-0.11	0.13	-0.46	0.28	0.50	0.28
Perimeter symmetric rate	-0.12	-0.07	0.33	0.46	0.03	-0.54
Hu1 Moment	0.25	-0.24	0.27	0.18	0.11	0.37
Hu2 Moment	0.15	-0.25	0.40	0.26	-0.04	0.52
Hu3 Moment	0.33	0.26	-0.04	0.02	-0.05	-0.05
Hu4 Moment	0.33	0.26	-0.04	0.03	-0.05	-0.05
Hu5 Moment	0.30	0.30	-0.03	0.10	-0.10	0.02
Hu6 Moment	0.32	0.2 7	-0.02	0.09	-0.08	0.02
Hu7 Moment	0.32	0.28	-0.03	0.07	-0.07	-0.01

(8) $F_{1*6}=X_{1*17}* E_{17*6}$

E_{17*6} as eigenvector matrix in table 5，F_{1*6} as the principal component vector，X_{1*17} as the original shape descriptor vector

5. Experiments and results

The first principal component eigenvalue is 6.783, the variance contribution rate is 39.9%, the entire data of standard variant is 39.9%, only relying on the first principal component can not reflect most information of original data. The first principal component in geometric invariant feature has high positive loads, in Appearance ratio and bounding rectangle compactness has high negative loads, the centroid offset rate has low positive loads, in the area and perimeter symmetric rate has a low negative loads, in other variables have similar secondary loads. That is to say, large (small) to the first principal components tend to be large (small) geometric moment invariants, then appearance ratio and bounding rectangle compactness tend to have smaller values (large). As a result of invariant moments are based on the regional global statistical feature, with similar but different sets of images (like the fruit with similar to leaves)has better identification of. Therefore, the first principal component can be used to characterize the morphology of the more slender, not completely symmetrical, with concave boundary, and the global features of difference image sets with strong discernment.

The second principal component feature value is 3.334, the variance contribution rate of 19.614%, the entire data of standard variation is 19.614%.Through the observation to load factor matrix in table 4, we find that the second principal component in most variables are approximately equal to the load, so can be used as a comprehensive metric variables, reflect almost all the features of the original nature.In addition, because the main components in the bounding rectangle of compactness and area symmetrical rate have a low normal load and the perimeter symmetric rate with a low negative load. Therefore, while the second principal component analysis in comprehensive measure all the features, it slightly weak impact of foliage bending on region side.

The third principal component feature value is 1.504, showing that the entire data of standard variation is 8.847% separately. The third principal component in the area of symmetric rate ratio has high loads, in circumference symmetrical ratio, fineness rHu1 and Hu2 moments have high positive load.To the leaf image, if the edge has some curl, it may not affect the perimet circumference, but it changed the blade on the imaging surface of two-dimensional area, causing the image has high perimeter symmetric rate, and low rate of area of symmetry. Therefore, the higher of the third principal component values tend to express has an elongated oval blades, and the blade edge may appear in curling.

The fourth principal component feature value is 1.405, showing the entire data of standard variation is 8.264% separately. The fourth principal component in encircle polygon compactness and perimeter symmetric rate have high positive loads, in appearance ratio and the circumference area ratio have high negative loads, in the centroid offset ratio with low negative loads,at the moment invariant part has lower positive load. High bounding polygon compactness and perimeter symmetric ratio both show that the region is filling full, is more symmetrical convex polygon, and the appearance ratio and perimeter area ratio with higher load, the principal component has higher differentiate different width of the ability to target.To a certain extent, the fourth principal components can be associated with the first principal component is complementary, it weakens the global moment invariant statistical characteristics influence, but strengthen the shape complexity and symmetry as well as the appearance of the elongated convex characteristics.

The fifth principal component feature value is 0.933, showing the entire data of standard variation is 5.489% separately. On the center of mass migration rate of fifth principal components have significantly high positive loads the area of symmetric rate has high positive load, in addition, the other variables with have smaller load value.In general, low centroid rate figure centroid bias top right (according to the centroid offset rate calculation formula can be seen), high centroid rate figure centroid bias bottom left. Value of the larger of the fifth principal components tend to say mass at the bottom left of the symmetrical region.

The sixth principal component feature value is 0.848, showing the entire data of standard variation is 4.985% separately. In the perimeter symmetric rates on sixth principal components have higher negative loads, and in Hu1 and Hu2 with higher positive load, in other variables are not high load. Because Hu1 and Hu2 are calculated by using image moment of order two , and the image of two order moment is not included details of the general information,in the geometry mean variance, if we consider only the order for 2 sets of the moment,the original image is completely equivalent to the image centroid as the center and has the same two order moments of the ellipse. Therefore, the higher of the sixth principal component values tend to express elliptical rather than circular, may have side edges curled area,but the direction and size of the show is weak.

From the above analysis we can see that the first six principal components of the cumulative variance contribution ratio reach to 87.1%, it can be very good summary of the data group. And the six main component covers the area of concave and convex of elongated,the complexity symmetry of shape ,the centroid position and direction, as well as statistics based on geometric moment invariants and other aspects of information, a reasonable description of the region shape characteristic, the original 17 features integrated into 6, while retaining the original variables most of the information under the premise, realize dimension compression of the data.

6. Conclusion

This article from the 17 dimensions of the quantitative describe the single plane shape characteristics, including the target boundary geometric description (perimeter area ratio, fineness ratio, eccentricity and so on),the target area and its surrounding polygon metric (bounding rectangle compactness, encircle polygon compact degree and so on), the target area (centroid offset ratio, area of symmetry,perimeter symmetry degree and so on) invariant moment (Hu), covering graphics elongated, convexity, complexity, symmetry, centroid direction, ellipse, compact plumpness, number of shape, geometric invariant moment etc.. Then through by the method of principal component analysis to do variable compression on these characteristics, explaining the he first six principal components which cumulative variance contribution rate reach to 87.1%, in which elongated, convexity and geometric invariant moment contributed to the first principal component is larger; the second principal component can be used as a comprehensive measure factor, each characteristics of second main components in the contribution rate is almost equal; the third principal component more reflect the region of the ellipse and elongated; the fourth principal component in global invariant moment statistical properties of the performance is not outstanding, but focused on the expression of slim and compact plumpness; the higher of the fifth principal component values tend to areas near left lower direction of the center of mass, which leaves the image on the right is fine, but left is wide, image centroid shift; sixth main components on the ellipse has stronger descriptive power, but the direction and size of the description of is a little weak.

References

[1]. SuHai Sun, Heng Chen, Strawberry cultivation technology in greenhourse. Rural science and technology, 2010(8): Page 74-75.

[2]. YangLong Liu etc, Fruit shape classification based on wavelet descriptors. Journal of Zhejiang University (Agriculture and life science edition), 2010(3): Page 322-328.

[3]. GuoQing Zhang, Forest products traceability system research. The modern agricultural science and technology, 2011(22): Page 224, 228.

[4]. HongGuang Cui, Image processing technology in the application of agricultural robot. Journal of Agricultural Mechanization Research, 2008(1): Page 168-170.

[5]. Ying Xiang, Agricultural pest image remote transmission system research and design, 2006, Chinese Academy of Agricultural Sciences.

[6]. ShuHua Jiang, HaiBo Sun, The computer image technology in Agricultural Engineering . Journal of Agricultural Mechanization Research, 2006(11): Page 177-178.

[7]. ShuJin Zhang, Image engineering (Book) - image analysis Second Edition. 2005: Tsinghua University press. Page 311-342.