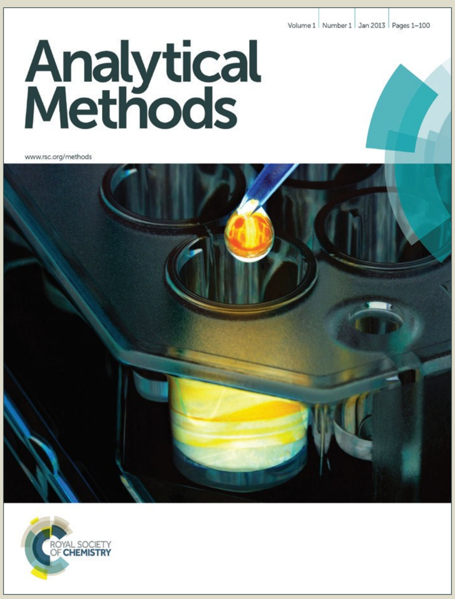
Analytical Methods

Accepted Manuscript



Analytical Methods

ARTICLE

A method for discrimination of processed ginger based on image color feature and support vector machine model

Received 08th December 2015,
Accepted 00th February 2016

DOI: 10.1039/x0xx00000x

www.rsc.org/methods

Sujuan Zhou,[a] Jiang Meng,[b] Zhanpeng Huang [a],Shizhong Jiang [a],Yongqiu Tu [a]

The discrimination of Chinese herbal medicine contains many appearance characters. Of which, color is one of the most important attributes. But all along color expressions in pharmacopoeia were vague. It's difficult to unify cognition by artificial and subjective discrimination method. In this topic, digital image processing and machine learning technology were introduced to help solving these problems. By extracting the numerical of quantified color from images of processed ginger, experiments demonstrated that the changing rule of HSV was consistent with the true color and can effectively discriminate the three processed ginger. After describing colors in terms of HSV, a SVM model for discrimination of processed ginger was constructed. An accurate rate of 98.0277% was acquired to identify the unknown samples. Therefore, the proposed discrimination method based on image color feature and SVM model is well able to quantify the color of processed ginger， and to evaluate the quality of appearance characters of Chinese herbal pieces objectively.

1. Introduction

The quality control and evaluation model of Chinese herbal medicine is one of the most important issues in the modern herbs development and has become a hotspot in processing industry. For a long time, the discrimination of processed herbal medicine quality mostly depends on related literature or experience of Chinese medical pharmacist.[1] Discriminated by this way, the results may be subjective and controversial. The discriminations among Chinese herbal pieces contain feature of color, shape, texture, etc. It's essential to find the objective expression of appearance characters using modern science and technology.

In recent years, scholars have studied about the objective quantification and discrimination of herbal medicine.[2-5] Huang Xuesi used color meter and electronic nose to acquire the sensor and color values of different processed products of betelnut, and then identified the optimal fire of betelnut.[6] Yang Tianjun used electronic nose and machine vision to identify different processed products of hawthorn.[7] Meng Qing'an proposed that objective expression of color was the basic of automatic discrimination of herbal medicine.[8] However, to our knowledge, no investigation has reported on the objective quantification for the discrimination of processed ginger.

Processed ginger are commonly used herbal medicines.[9] Different processed products of ginger have many differences in appearance. Chinese Pharmacopoeia 2010 edition[10] describe the characters of processed ginger like this, "the surface of Zingiberis Rhizoma (Ganjiang, ZR) appears brownish yellow or grayish brown, coarse, with longitudinal wrinkles","the surface of Zingiberis Rhizoma Preparatum (Paojiang, ZRP) appears brownish black or brown, irregular blocks with finger shaped branch", "the surface of Zingiberis Rhizoma Carbonisata(Jiangtan, ZRC) appears burned black". From these descriptions we can see color is one of the most important characters of identifying processed ginger. But these ambiguous descriptions of color are unfavorable for the standardization of processing technology. It's necessary to quantify the color of different processed products of herbal medicine. Digital image processing and machine learning technology is introduced to help the objective quantification of herbal medicine color, which can enhance the objectivity of identifying the color of herbal medicine and produce progress in the internationalization of herbal medicine.[11]

2. Materials and methods

2.1. Experimental materials

The samples were collected from SiChuan, GuiZhou, YunNan and other provinces, including three species of processed ginger. All these samples were identified by Professor Jizhu Liu(College of Traditional Chinese Medicine, Guangdong Pharmaceutical University).Part of each batch were respectively processed to ZR, ZRP and ZRC according to Chinese Pharmacopoeia (2010 edition). Each sample was taken photo from both sides. More than 150 pieces of images were taken from the sample to form an image database. For the difference of light condition and camera angle, black box with a fixed light source was needed to reduce difference produced

[a.] *College of Medical Information Engineering, Guangdong Pharmaceutical University, Guangzhou, 510006, China. E-mail: susona2002@163.com; Tel: +86 020 39352207.*
[b.] *College of Traditional Chinese Medicine, Guangdong Pharmaceutical University, Guangzhou, 510006, China. E-mail: jiangmeng666@126.com; Tel: +86 020 39352169.*

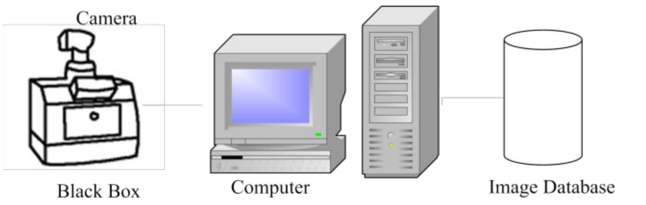


Fig.1 Images acquisition equipment

by image acquisition. Here used a UV analyser with jacklight to avoid backscattering effects from other light sources. Each sample was placed manually on a pure background at a distance of 30 cm to the camera.[12] The images acquisition system was shown as Fig.1.

The image acquisition system used in this research consists of three components: RGB color camera (CASIO COMPUTER CO.,LTD., Japan) at resolution of 4608*3456 pixels; a UV analyser (WD-9403C) used as dark box; and a computer installed with software for image processing. A number of these images are stored as image database for further processing.

Each sample image was saved as the format of JPG. Fig.2 showed the images of some samples.

2.2. Experimental methods

In order to quantize the color of processed ginger, object images must be firstly segmented from original image. And then color feature can be extracted subsequently. Lastly, a classification model can be constructed based on quantification of color. The steps were shown as Fig.3.

2.2.1 Pre-processing of image. Image noise reduction and image segmentation are the essential prerequisites of image analysis.

Image noise is random variation of brightness or color information in images[13], which leads to a breakdown of image quality. Many algorithms are used for noise reduction, such as mean filter, Gaussian filter and mathematical morphology.[14] Open-Close operation in mathematical morphology was used in this topic to reduce image noise.

Image segmentation serves as a key and fundamental step to high-level tasks such as image retrieval, object recognition and image understanding[15]. It is the process of pixels classification. The classification is mainly based on the two characters of grey level: discontinuity between regions and similarity among regions.[16] Threshold segmentation method is widely used in image segmentation because of its simplicity and high efficiency. In eqn(1), threshold T is used to segment image f(x,y) into foreground and background parts.

$$g(x,y) = \begin{cases} 1 & f(x,y) > T \\ 0 & f(x,y) <= T \end{cases} \quad \cdots\cdots\cdots\cdots\cdots\cdots (1)$$

For the object images of herbal medicine differ greatly from background images in color, threshold segmentation may be suitable. But this method led to unfavourable result after

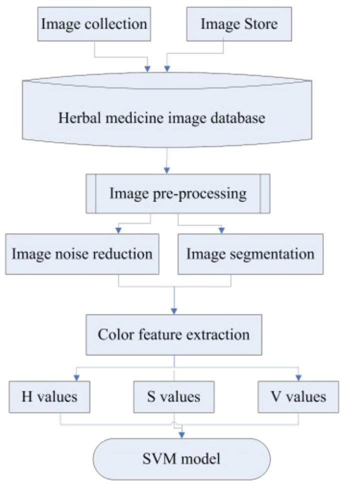
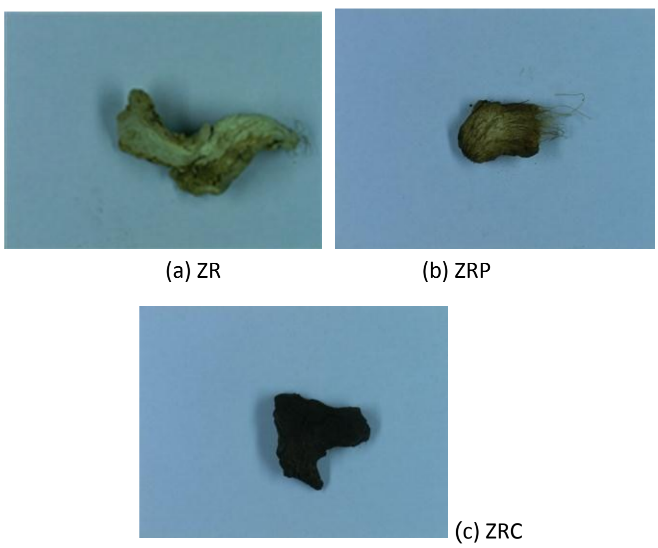


Fig.3 Image processing steps



(a) ZR (b) ZRP

(c) ZRC

Fig.2 Images of processed ginger

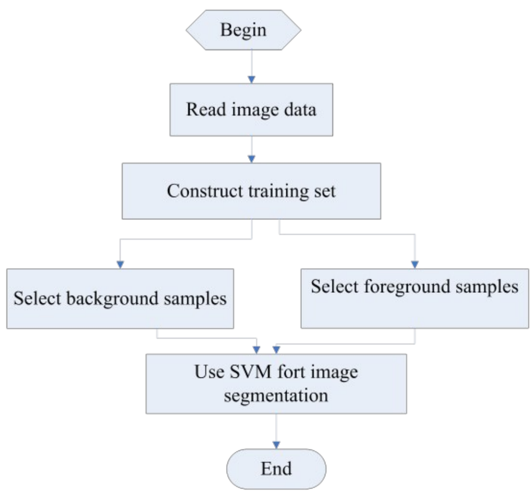


Fig.4 Image segmentation based on SVM

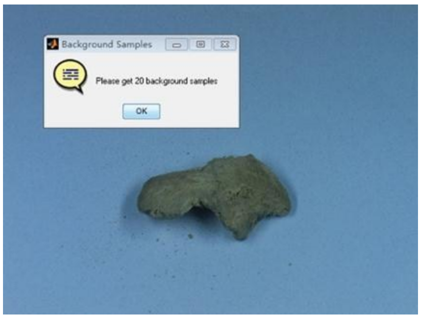


Fig.5 Images of processed ginger

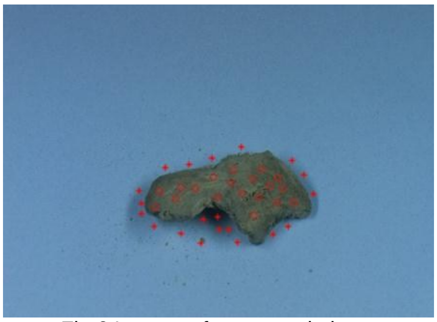


Fig.6 Images of processed ginger

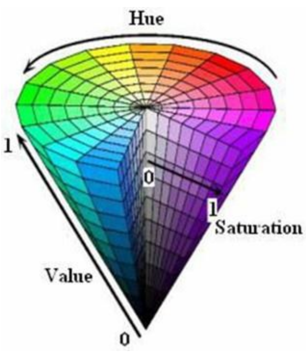


Fig.7 HSV color space

experimentation. Reasons might be that shadow and noise produced during image acquisition had affected the effect of segmentation. A new image segmentation method based on Support Vector Machine (SVM)[17] was introduced to improve the segmentation effect. In this method, image segmentation can be viewed as a classification problem, which means labelling each pixel according to certain essential characteristics[18]. As a classical classification method, SVM has many attractive features compared with other machine learning systems, including the absence of local minima, its speed and scalability and its ability to condense information contained in the training sets. As Fig.4 showed, considering segmenting a given image into foreground and background. Samples were selected from foreground and background image as training sets to train SVM and get accurate classification of pixels[19]. The steps were shown as Fig.5 and Fig.6. In Fig.5, after clicking "OK", you can use mouse to get 20 background samples (signed as red "*"in Fig.6) and 20 foreground samples (signed as red "o" in Fig.6) to be segmented.

2.2.2 Color feature extraction. Images of different processed products of ginger were saved as RGB (Red, Green, Black) color space. RGB color space describes color in terms of the amount of red, green, and blue present. Because the R, G, and B components of an object's color in a digital image are all correlated with the amount of light hitting the object, and therefore with each other, image descriptions in terms of those components make object discrimination difficult. HSV[20] color space(shown as Fig.7) describes colours in terms of Hue, Saturation, and Value. The HSV model describes colours similarly to how the human eyes tend to perceive color. In situations where color description plays an integral role, the HSV color model is often preferred over the RGB model.

Eqn(2) is a simple RGB-to-HSV color converter function, which H represents the color type, S represents the vibrancy of the color and V represents the brightness of the color.

$$H = \begin{cases} \arccos \dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}} & \cdots\cdots B \le G \\ 2\pi - \arccos \dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}} & \cdots\cdots B > G \end{cases}$$

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (2)$$

$$V = \frac{\max(R,G,B)}{255}$$

2.2.3 Construction classification model of SVM. In machine learning, SVM is supervised learning model with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis.

The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes. SVM was firstly proposed by Cortes and Vapnik[17] and has been widely applied in the biological and other sciences. Here SVM was used to train model of identifying processed ginger based on image color feature. RBF kernel function is used in SVM classification. The RBF kernel on two samples x and x', represented as feature vectors in some input space, is defined as[21]

$$K(x, x') = \exp\left(-\frac{\left\|x - x'\right\|^2}{2\sigma^2}\right)$$

$\left\|x - x'\right\|$ may be recognized as the squared Euclidean distance between the two feature vectors. σ is a free parameter.

There are two parameters for an RBF kernel: C and γ. It is not known beforehand which C and γ are best for a given problem; cross-validation[22] is used to identify good (C, γ) so that the classifier can accurately predict unknown data. In v-fold cross-validation, we first divide the training sets into v subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining v − 1 subsets. Thus, each instance of the whole training sets is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified.

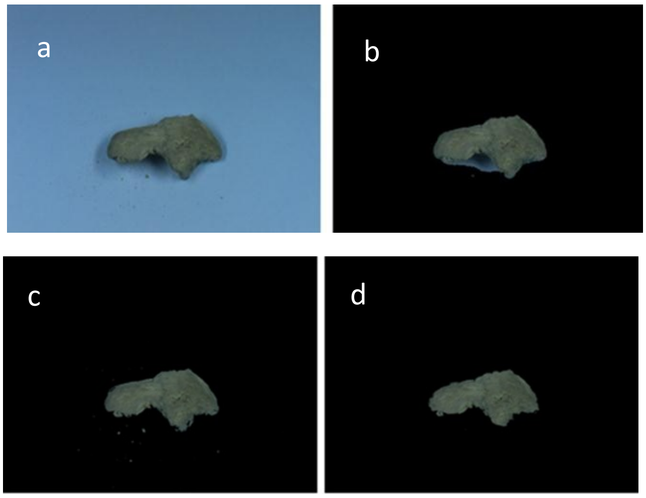


Fig.8 Segmentation result contrast

LIBSVM[23] is integrated software for support vector classification, developed at the National Taiwan University. Here used LIBSVM to construct SVM model.

3. Results and discussion

3.1 Target image extraction

Traditional method of threshold segmentation and new segmentation method based on SVM were both used for contrast. Fig.8a was the original image of one ZR sample. Fig.8b was the segmentation result of traditional method of threshold segmentation. From Fig.8b we can find that shadow was remained in the object image as noise. Fig.8c and Fig.8d were the results of segmentation method based on SVM. Fig.8c was the result of selecting 10 feature points as training sets and Fig.8d was the result of selecting 20 feature points. From Fig.8c and Fig.8d We can see the more feature points selected for training, the better effect of segmentation. But the increase of feature points will bring burden to machine learning and lead to down of speed. So the selection number of feature points is very crucial. 20 feature points were suitably selected for training in this topic through simulation and experiment.

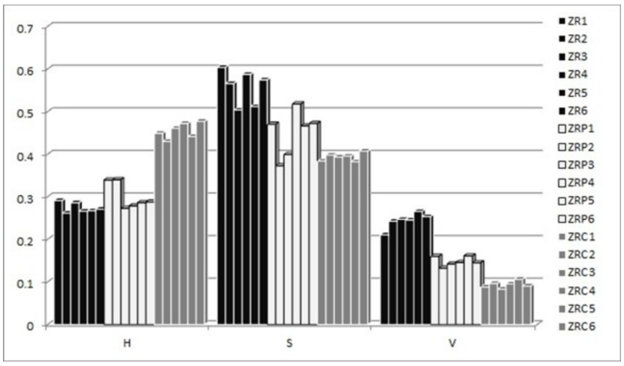


Fig.9 Distribution of HSV value of different processed ginger

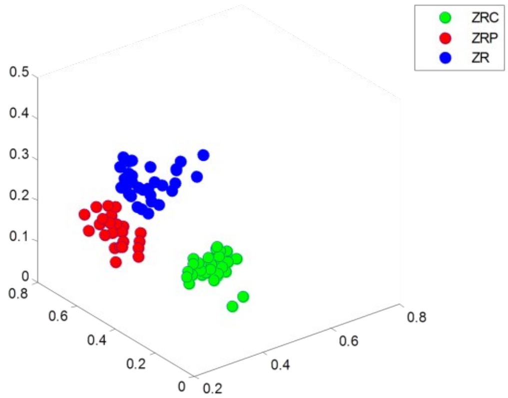


Fig.10 FCM result

3.2 Results of color feature extraction and analysis

After changing from RGB color space to HSV space, HSV color space describes colours in terms of Hue, Saturation, and Value. From Fig.9 we can observe the regular distribution of Hue, Saturation, and Value from different processed ginger. The values of H gradually increased from ZR to ZRC, which accorded with the variation trend of appearance color from ZR to ZRC. On the contrary, the values of S and V decreased from ZR to ZRC, which meant chroma and lightness lead to dark and black.

Cluster analysis was done to verify the above change rule of color. Data clustering is the process of dividing data elements into classes or clusters so that items in the same class are as similar as possible, and items in different classes are as dissimilar as possible.

One of the most widely used fuzzy clustering algorithms is the Fuzzy C-Means (FCM) Algorithm[24]. The FCM algorithm attempts to partition a finite collection of elements into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of cluster centres and a partition matrix, where each element

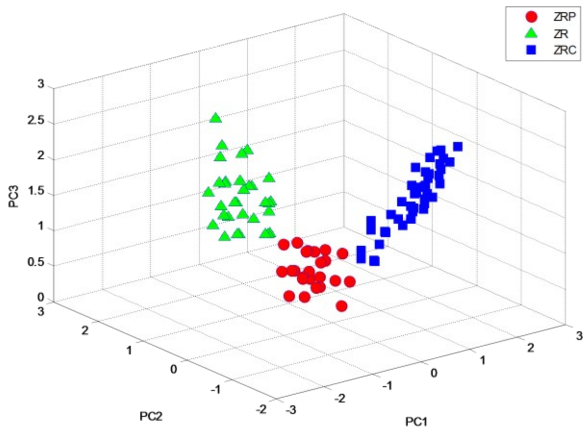


Fig.11 PCA score plots for discriminating processed ginger according to HSV

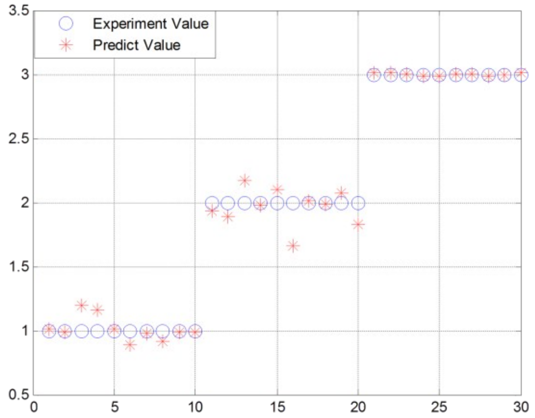


Fig.12 SVM model prediction result

tells the degree to which element belongs to cluster.

Here used [H, S, V] as X for input. Fig.10 was the FCM result. We can see data elements were assigned to three clusters, corresponding to ZR, ZRP and ZRC. This indicated that values of HSV can be used to identify different processed ginger from color feature.

In addition, the 3 parameters (H, S,V) were used as the inputs of PCA(Principal Component Analysis). PCA is a chemometric, linear, unsupervised, and pattern recognition technique. It's used for analysing, classifying, and reducing the dimensionality of numerical datasets in a multivariate problem.[25] Fig.11 showed a three-dimensional scores plot of the first three principal components (PC1 = 89.099%;PC2 = 7.659%; PC3 = 3.242%). As can be seen, the samples representing three groups can be discriminated clearly.

3.3 SVM classification model based on color feature

After extracting color feature from different processed ginger, 150 groups of values expressed as (H, S, V) were selected as datasets. 120 groups from each kind of processed products were taken randomly as the training sets, and the rest 30 groups were taken as the validation sets. The values of (H, S, V) were used as the inputs of the SVM, and the expected outputs were the classification of different processed ginger (ZR as 1, ZRP as 2 and ZRC as 3). The publicly available LIBSVM software was used to process the SVM computation with the radial basis function as the kernel. After experimenting many times, the predict result was shown in Fig.12, with accurate rate of 98.0277% by average. The high accuration indicated that the SVM model could well identify different processed products of ginger in color feature.

4. Conclusions

Currently, the common methods for discrimination of Chinese herbal medicine mostly depend on their appearance characters. Color is one of the most important appearance characters of Chinese herbal medicine. But traditional artificial methods of identifying color mostly rely on subjective experiences or obscure expressions, which is unfavourable for the standardization of processing technology. This research sought to develop a rapid, objective, and accurate detection method to discriminate different products of processed ginger.

In this paper, digital image processing and pattern-classification technology were introduced into the objective quantification of herbal medicine color. After extracting the color feature and describing colours in terms of HSV, a SVM model for discrimination of processed ginger was constructed. Experiments showed the prediction results were accurate and reliable. The proposed method is promising for the objective discrimination of herbal medicinal material and will achieve the controllability and stability of processing quality at the same time.

What's more, this paper mainly discussed about how to discriminate different processed ginger. But it's not rich enough to discriminate processed ginger from other herbs, because different herbs not only differ in color, but also in other characters, such as odor and taste. To discriminate processed ginger from other herbs, other technology such as electronic nose and electronic tongue technology should be included in. This is the following study direction in the future.

Acknowledgements

This work was financially supported by the Project of the National Natural Science Foundation of China (No. 81102809). We are also grateful to many of our colleagues for supporting this work.

References

1. S.L.Yang, S.P.Xie, M.Xu, C.Zhang, N.Wu, J.Yang, L.Zhang, D.Y.Zhang, Y.Jiang and C.J.Wu, *Analytical Methods*, 2015, **7**, 943-952.
2. C.Apetreia, I.M.Apetreib, S.Villanuevab, J. A. d. Sajab, F.Gutierrez-Rosalesc and M.L.Rodriguez-Mendezd, *Analytica Chimica Acta*, 2010, **663**, 91-97.
3. H. Hong, X. Yang, Z. You and F. Cheng, *Aquacultural Engineering*, 2014, **63**, 62-71.
4. S.Q.Cui, J.Wang, L.C.Yang, J.F.Wu and X.L.Wang, *Journal of Pharmaceutical and Biomedical Analysis*, 2015, **102**, 64-77.
5. M.Pein, D.Kirsanov, P.Ciosek, M.D.Valle and I.Yaroshenko, *Journal of Pharmaceutical and Biomedical Analysis*, 2015, **114**, 321-329.
6. X.S.Huang, W.M.Li, X.L.Zhang, J.Jia, H.H.Zhang and C.J.Wu, *China Journal of Chinese Materia Medica*, 2009, **34**, 1786-1791.
7. T.J.Yang, S.L.Yang, L.Liang, C.J.Wu and A.L.Li, *Lishizhen Medicine and Materia Medica Research*, 2014, **25**, 2399-2402.
8. Q.A.Meng and E.S.Liu, *Tianjing Journal of Chinese Tradictional Medicine*, 2014, **31**, 696-699.
9. X.M.Deng, J.Y.Yu, M.Zhao, B.Zhao, X.Y.Xue, C.T.Che, J.Meng and S.M.Wang, *Journal of Separation Science*, 2015, **38**, 2945-2952.
10. C. P. Commission, *Pharmacopoeia of the People's Republic of China 2010 English Edition*, China Medical Science and Technology Press 2011.

11. Q.Liu, X.P.Liu, Z.X.Wen and R.T.Wang, *Automation & Instrumentation*, 2012, **1**, 158-160.
12. M.Xu, S.L.Yang, W.Peng, Y.J.Liu, D.S.Xie, X.Y.Li and C.J.Wu, *Evidence-Based Complementary and Alternative Medicine*, 2015, **2015**, 1-10.
13. J.Ellenberge and CE.Chow, presented in part at the Multimedia Computing and Communications, 2010.
14. Serra.J, *Image Analysis and Mathematical Morphology*, Academic Press, New York, 1988.
15. H.D.Cheng, X.H.Jiang, Y.Sun and J.L.Wang, *Pattern Recognition*, 2001, **34**, 2259-2281.
16. R. C.Gonzalez and R. E.Woods, *Digital Image Processing*, Electronic Industry Press, America, Third edn., 2012.
17. C.Cortes and V.Vapnik, *Machine Learning*, 1995, **20**, 273-297.
18. X.F.Bai and W.J.Wang, *Neurocomputing*, 2014, **136**, 243-255.
19. X.C.Wang, F.Shi, L.Yu and Y.Li, *Beihang University Press*, 2014, **2**, 153-159.
20. Petrou.M, *Image processing*, Machine Press, 2005.
21. Jean-Philippe.Vert, Koji.Tsuda and Bernhard.Schölkopf, *Kernel Methods in Computational Biology*, 2004, **7**, 1-42.
22. G.C.Cawley, presented in part at the International Joint Conference on Neural Networks, 2006.
23. C.C.Chang and C.J.Lin, A Practical Guide to Support Vector Classification, http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html.
24. Bezdek and James.C, *Pattern Recognition with Fuzzy Objective Function Algoritms*, Plenum Press, New York, 1981.
25. R. V. Altena, S. Duggirala, M. I. P. Groschel and T. S. v. d. Werf, *Current Pharmaceutical Design*, 2011, **17**, 2853-2862.

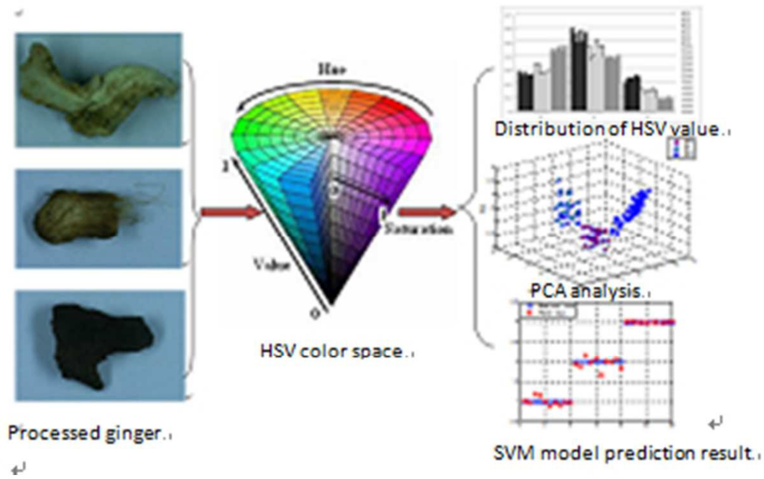


Image color feature with SVM and PCA were employed to discriminate different processed products of ginger.
102x62mm (96 x 96 DPI)