

# Weed Seeds Identification based on Structure Elements' Descriptor

Minnan Tang and Cheng Cai \*

\*Department of Computer Science, College of Information Engineering, Northwest A&F University, Yangling, Shaanxi, China  
E-mail: cheney.chengcai@gmail.com Tel: +86-29-87091641

**Abstract**—the implementation of new methods for automatic, reliable identification and classification of seeds is of great technical and economic importance in agricultural industry. As in ocular inspection, the automatic classification of seeds should be based on knowledge of seed size, shape, color and texture. In this work, we assess the discriminating power of these characteristics for the unique identification of seeds of 216 weed species. We identified a nearly optimal set of 4 (three morphological and one color and textural) seed characteristics as classification parameters, using the performance of the Support Vector Machines as classifier. Among these characteristics, color and textural features are extracted and described by SED (structure elements' descriptor) simultaneously which proves to perform better than other image retrieval methods. The main findings of this paper are shown in the strong discrimination power of SED. Moreover, experimental results suggest that recognition rate reaches the peak with the combination of the morphological characteristics and SED.

## I. INTRODUCTION

The identification and classification of seeds are essential activities which make great contribution to the maximization of crop production. Of more than 30 million species of plants in the world, the total amount of weeds reaches up to 8,000[1]. Weeds bring serious harm to food productions which can result in enormous agricultural production losses every year. Weeds mostly spread and reproduce in the form of seeds mingling with crop seeds, grain and other plants. Therefore, the implementation of new methods for automatic, reliable and fast identification and classification of seeds is of major technical and economic importance in agricultural industry. For the identification of weed seeds, the classification process by professionals is extremely slow. At the same time, with regard to trade and technical application, the process processes poor reusability and the subjective result is hard to quantify and evaluate. Consequently, weed seeds identification by machine version has been a hot topic in the area of agricultural engineering in recent years. Compared with seeds identification within one species of a variety of categories, wild seeds identification is rather different. In general, the differences between weed seeds of the same species are obvious, but some of weed seeds in different species can also be similar because of correlation.

Most previous attempts by means of machine vision focused on seed size, shape, color and texture in tackling the identification problem. The previous researchers adopt

different classification methods in their study, such as Linear Discriminant Analysis(LDA)[2], Artificial neural networks (ANNs) analysis algorithm[3], Boosting analysis algorithm[4], and Naïve Bayes algorithm[5] etc.

Petersen and Krutz(1992) used color images, rather than black and white ones, to research, the result of which suggests color feature is of importance to improve classification accuracy[6].

More recently, Chtioui, Bertrand, Dattee and Devaux(1996) extracted parameters of seeds size, shape, and texture from four different species(yellow dock, wild oats, alfalfa, vetch). Then, they applied linear discriminant analysis and artificial neural networks(ANNs) respectively to identifying weed seeds. Their classification accuracy reached up to 92% with regarded size and shape as experimental parameters. Furthermore, when they combined size, shape and textural features, the accuracy reached nearly 99%. What is worth noting is that this research did not make full use of color features. What's more, it involved only four different species, thus not able to supply a good representation of inter-species seed otherness [2].

In 2002, Granitto and Navone proposed an identification method based on machine vision. By considering the features of size, shape, color and texture. They split 3163 color images of the 57 species, and identified an optimal set of 12 (six morphological, four color, and two textural) seed characteristics as classification parameters[3].

To improve the research conducted by Chtioui, Bertrand, Dattee and Devaux (1996), they presented an identification method based on machine vision with a large group of samples. The database they used contained 10310 images of 236 species. They extracted six morphological features, four color features and two textural features respectively, and then applied Linear Discriminant Analysis (LDA) and artificial neural networks (ANNs) to classification process. What's worth mentioning is that they researched identification only with morphological and textural features in the condition of dislodging the color feature parameter. The result indicated that the identification accuracy changed little when the color feature parameter was dislodged [5].

The methods mentioned above are almost based on feature extraction, such as size, shape, color and texture. In recent years, there have been a number of new ways that are not based on those features for weed seeds identification.

In 2009, Mengbo You and Cheng Cai proposed to classify weeds with their digital images by using PCA, 2DPCA, column-directional 2DPCA and  $(2D)^2$ PCA and compare their performances under different dimensions.[7] The same year, FengFu Zhao, Cheng Cai and JunPing Zhu presented a novel approach for the recognition of weed seeds, known as color Principal Component Analysis (PCA). Their experimental results demonstrated that the representation of color PCA gained much higher recognition rates than those of traditional PCA[8].

More recently, in 2010, Ming Zhang, Cheng Cai and JunPing Zhu used the compressive sensing theory, which has been applied to the field of machine learning, to do some dimension reduction treatment to avoid careful selection of the feature set. They also researched about the identification of weed seeds with corruption, which can help to deal with the real world recognition problem[9].

Also in 2010, Wencang Zhao and Junxin Wang presented a method to feature extraction based on visual invariance, aiming at the seeds of biological stability genetic character. Then they used Back Propagation (BP) Neural Network to identify weed seeds and analysis relationship between the changes of features dimension. The experimental results showed that the recognition rate of the 16 dimensions eigenvalue is up to 96%[10].

These previous attempts before 2005 mentioned above have a common disadvantage that color features and textual features have not been fully utilized during classification. However, color feature and textual feature often become the key factor in identification especially between weed seeds which morphological features are extremely similar.

In this paper, a novel color and texture descriptor SED is applied to weed seeds recognition, the color and textural features are described by SED simultaneously. The structure elements are defined by five structure elements denoting five directions, separately. Instead of RGB color space quantized into 256 colors, HSV color space quantized to 72 bins is used to extract the color feature and texture feature. In this way, the feature can be extracted more precisely. What's more, this method also has strong distinguishing power for image identification in large scale database[11]. Eventually, we combine it with the morphological features, identification accuracy reaches to the peak by this time.

The rest of the paper is organized as follows. First, we define the morphological parameters measured from weed seeds image, and choose the most relevant ones for classification. Then, we introduce how to extract the color feature and texture feature simultaneously by SED. Next, we present the experiment results obtained with the Support Vector Machines classifier. Finally, some conclusions are drawn in the last section.

## II. FEATURE EXTRACTION

### A. Weed Seeds Image Acquisition

We have built a database containing 9242 images of the 216 species considered which is supplied by the Seed

Analysis Laboratory at EEA Oliveros of INTA[5]. Images were taken with a  $768 \times 512$  pixel resolution on a black background. Through the rotation and translation adjustments, we make the tip of weed seeds in the same direction.

In order to illustrate the preliminary image processing result more effectively, in Fig.1, we show images from six different species selected to ensample how similar /different shapes, colors and textures could be.

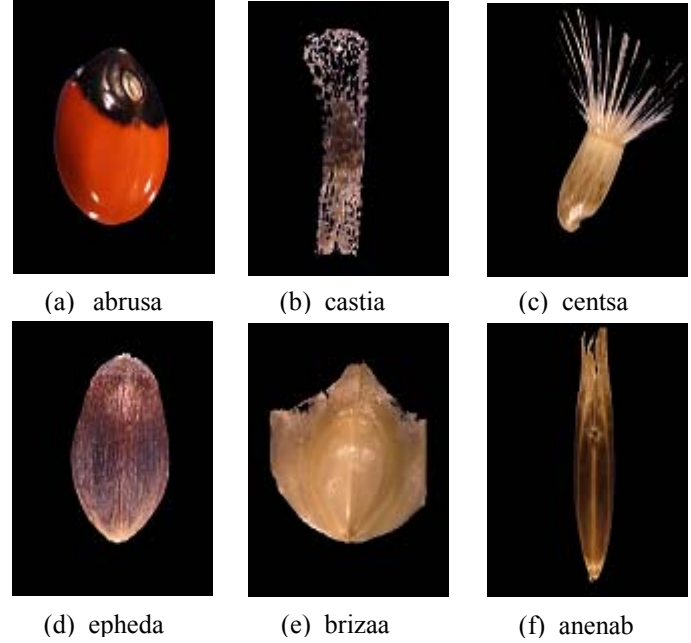


Fig. 1 Images of different seeds in the database. From top to column, first row: abrusa, castia, centsa; second row, epheda, brizaa, avenab.

### B. Morphology Features

Shape and size features of weed seeds could be easily acquired from gray level images. Finally, we selected 3 morphological features as parameters for classification, which are listed following (see Fig. 2) [3]:

- Square root of seed area [ $\text{SQRT}(A)$ ] where  $A$  denotes the seed area.
- Ratio of semi-axis lengths of main principal axis [ $h_1/h_2$ ].
- Ratio of seed and enclosing box areas [ $A/((h_1+h_2) \times (v_1+v_2))$ ].

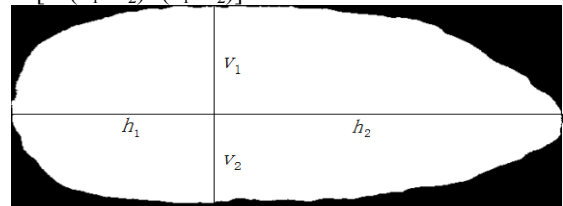


Fig. 2 Definition of quantities related to seed shape used to compute morphological features ( $v_1$  and  $v_2$  are respectively up and down semi-minor axis,  $h_1$  and  $h_2$  are respectively left and right semi-major axis)

### C. Color and Textural Features of SED

Color and textural features play an important role in weed seeds identification. We describe these characters by SED (structure elements' descriptor) [11] simultaneously which is

proved has a better performance than other image retrieval methods. This features extraction is consist of two major steps: color space conversion and structure elements' descriptor.

(a) Color space conversion

Instead of RGB color space, we use HSV color space which is widely used in color feature extraction. HSV color space has an advantage that it is in close proximity to human conceptual understanding of colors. We quantized the HSV color space to 72 bins in this paper in consideration of computing complexity.

According to the characteristic of HSV color space,  $H \in [0, 360], S \in [0, 1], V \in [0, 1]$ , the conversion process as follows:

Step1. Divide hue into 8 zones, saturation into 3 zones and values into 3 zones by follow equation:

$$H = \begin{cases} 0, & H \in [0, 24] \cup [345, 360] \\ 1, & H \in [25, 49] \\ 2, & H \in [50, 79] \\ 3, & H \in [80, 159] \\ 4, & H \in [160, 194] \\ 5, & H \in [195, 264] \\ 6, & H \in [265, 284] \\ 7, & H \in [285, 344] \end{cases} \quad (1)$$

$$S = \begin{cases} 0, & S \in [0, 0.15] \\ 1, & S \in (0.15, 0.8] \\ 2, & S \in (0.8, 1] \end{cases} \quad (2)$$

$$V = \begin{cases} 0, & V \in [0, 0.15] \\ 1, & V \in (0.15, 0.8] \\ 2, & V \in (0.8, 1] \end{cases} \quad (3)$$

Step2. Construct the color feature into one-dimension based on the following classification

$$P = 9H + 3S + V \quad (4)$$

Step 3. Compute the point set  $L_i$

We assume  $I$  is the result that has already been quantized by HSV color space. We use  $i$  to denote the value that has been quantized to 72 bins, respectively. And  $L_i$  denote the point set where the value is  $i$ , which can be computed by the following equation:

$$L_i = \{(x, y) | (x, y) \in I, I(x, y) = i, 0 \leq i \leq 71\} \quad (5)$$

In this way, convert three components of the HSV to a one dimensional vector, which has 72 main colors. Then we can use SED to represent the image on 72 kinds of main colors.

(b) The structure elements' descriptor

It is well known orientation is of great importance in image description. Thus, the structure elements are defined by five

elements denoting five directions, respectively. SED has  $2 \times 2$  matrixes and is showed in Fig. 3.

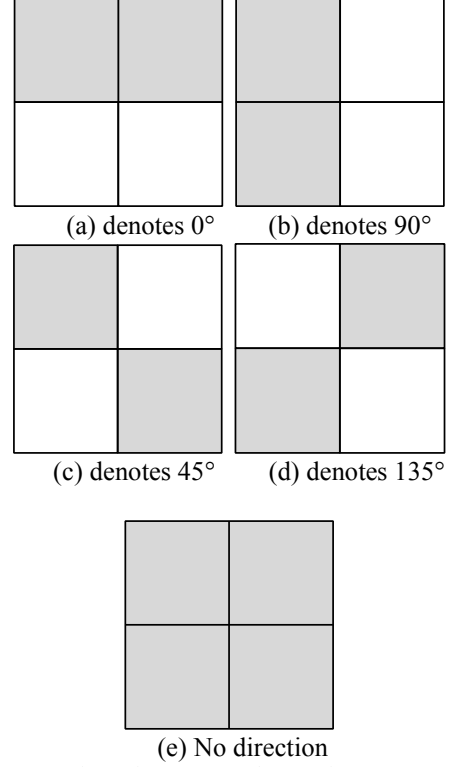


Fig. 3 Five structure elements in SED

We use the five element templates to detect the texture. Though the following three-step method can we acquire the final described image:

Step1. Starting from the base point  $(0,0)$ , move the  $2 \times 2$  SED from left to right and top to bottom with 2-step length.

Step2. If the structure element matches the value of the image, the value will be reserve, or not we will give up the value. Then we will obtain a SED map, denoted by  $S_i(x, y)$  ( $1 \leq i \leq 5$ ).

Step3. Suppose  $S(x, y)$  is the final map, which is obtained by fusing five SED maps based on the following rule:

$$S(x, y) = \{(x, y) | S_1(x, y) \cup S_2(x, y) \cup S_3(x, y) \cup S_4(x, y) \cup S_5(x, y)\} \quad (6)$$

There are five structure elements in SED, and the SED is extracted in every bins. Therefore, the feature vector is a 360-dimensional feature vector.

### III. EXPERIMENTS

#### A. Results

For the sake of comparing the discriminating power of the different sets of features, we conduct the experiment with LIBSVM(a library the Support Vector Machines, the default kernel -RBF is adopted)[12] built solely in terms of the three morphological feature, color and textural feature described by SED. In each experiment, we split the 9242 images of 216 species considered in training and test sets, randomly choosing, for each species, 75% of the images to build the classifier

and including the remaining 35% in the test. Table I gives the average performances and standard deviations over the 10 experiments.

TABLE I  
SUPPORT VECTOR MACHINES CLASSIFIER PERFORMANCES AS PERCENTAGE OF CORRECT SEED IDENTIFICATIONS

| Features | Morphology    | SED           | Combine       |
|----------|---------------|---------------|---------------|
| Accuracy | 0.2342±0.0085 | 0.6833±0.0074 | 0.9821±0.0090 |

MEAN VALUES AND STANDARD DEVIATIONS ARE ESTIMATED FROM 10 INDEPENDENT EXPERIMENTS, AS DESCRIBED IN THE MAIN TEXT.

The recognition rate for morphological features is only available at 23.42%. Color and textural feature described by SED achieves identification accuracy at 68.33%, which is obviously higher than morphological feature does. It is worthy of paying attention that the recognition rate reach up to 98.21% with the morphological feature and color and textural feature described by SED combined.

#### B. Discussion

For comparison, we also compute the recognition rate by other approaches mentioned in the introduction section. In the comparison experiment, we used the same ratio to segment the database so that make all the result with comparability. Table 2 gives the average performances of the different methods over 10 experiments.

TABLE II  
PERFORMANCE OF DIFFERENT OF METHODS AS PERCENTAGE OF CORRECT SEEDS IDENTIFICATIONS.

| Methods  | PCA[7] | (2D) <sup>2</sup> PCA[7] | Color PCA[8] | Our Method |
|----------|--------|--------------------------|--------------|------------|
| Accuracy | 0.9200 | 0.9200                   | 0.8080       | 0.98921    |

Table II shows that among all the identification methods, our framework obviously achieve the best performance.

#### IV. CONCLUSIONS

In this paper, a weed seeds identification method based on structure elements' descriptor is proposed. The experimental results show that the recognition rate is not satisfied when only make morphological feature or color and textural feature described by SED be classification parameter. Compare with these two features, SED obviously performed better than morphological feature. However, the identification accuracy would reach up to the peak with the combination of the morphological feature and SED.

Our experiment is based on the consideration of all the sample images without continuous occlusion and corruption which will surely exist in the real world. This is what needs to be improved further.

#### ACKNOWLEDGMENT

Project 61202188, 61303125 supported by National Natural Science Foundation of China.

Thanks the Seed Analysis Laboratory at EEA Oliveros of INTA for providing the weed seeds database

#### REFERENCES

- [1] Y. Li, "Weed Flora of China", *Chinese Agricultural Publishing House*, pp. 10-13.
- [2] Y. Chtioui, D. Bertrand, Y. Dattee, and M.-F. Devaux, "Identification of seeds by colour imaging: comparison of discriminant analysis and artificial neural network", *Journal of the Science of Food and Agriculture*, vol. 71, pp. 433-441, 1996.
- [3] P. M. Granitto, H. D. Navone, P. F. Verdes, and H. A. Ceccatto, "Weed seeds identification by mechine vision", *Computers and Electronics in Agriculture*, vol. 33, pp. 91-103, 2002.
- [4] P. M. Granitto, P. A. Garralda, P. F. Verdes, and H. A. Ceccatto, "Boosting Classifiers for Weed Seeds Identification", *Journal of Computer Science and Technology*, vol. 3, 2003.
- [5] P. M. Granitto, P. F. Verdes, and H. A. Ceccatto, "Large scale investigation of weed seed identification by mechine vision", *Computers and Electronics in Agriculture*, vol. 47, pp. 15-24, 2005.
- [6] P. E. H. Petersen and G. W. Krutz, "Automatic identification of weed seeds by color machine vision", *Seed Science and Technology*, vol. 20, pp. 193-208, 1992.
- [7] M. You and C. Cai, "Weed Seeds Classification Based on PCA, 2DPCA, Column-directional 2DPCA and (2D)<sup>2</sup> PCA," *International Asia Symposium on Intelligent Interaction and Affective Computing*, Wuhan, pp. 187-190, 2009.
- [8] F. Zhao, C. Cai, and J. Zhu, "Weed Seeds Recognition Using Color PCA", *the Second International Symposium on Knowledge Acquisition and Modeling*, Wuhan, pp. 375-378, 2009.
- [9] M. Zhang and C. Cai, "Sparse Representation for Weed Seeds Classification", *the International Conference on Green Circuits and Systems*, Shanghai, 2010.
- [10] W. Zhao and J. Wang, "Study of Feature Extraction Based Visual Invariance and Species Identification of Weed Seeds", *the International Conference on Natural Computation*, Yantai, Shandong, pp.631-635, 2010.
- [11] X. Wang and Z. Wang, "A novel method for image retrieval based on structure elements' descriptor", *Journal of Visual Communication and Image Representation*, vol. 24, pp. 63-74., 2013.
- [12] C.C. Chang and C.J. Lin, "A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1-27:27, 2011. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/>.