# Identification of crop weed based on image texture features

**Jie Zhang[1], Fuli Song[2], Jiali Tang[1*]**

[1]*School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, Jiangsu, China*

[2]*Personnel Division, Henan Radio & Television University, Zhengzhou 450008, Henan, China*

*Received 1 March 2014, www.cmnt.lv*

**Abstract**

By using computer image processing technology, texture features of weed in the corn seedling field are analysed, and then we present an algorithm combining Support Vector Machine (SVM) to form a classifier and promote an improved method of RBF network. The experimental results show that the proposed method is effective.

*Keywords:* crop, weed identification, image texture features, support vector machine (SVM)

## 1 Introduction

We know that weeds cause great impact on the growth and yield of crops. Along with the development of agricultural information, a new method called image-based machine recognition pattern was introduced for weed identification. Meanwhile the new method can realize directed drug spray, protect the environment and save money.

To separate the weeds from other crops precisely, we must be able to extract the characteristics of various types of weeds accurately. Many studies have confirmed that if we extract the colour feature, the shape feature and the texture feature of weeds, then put them together, we can achieve more accurate identification. However, most studies are just a combination of several types of features, so it takes a large amount of computation. To some extent, it affects the efficiency of identification. However the research results of single feature-based identification are less [1, 2].

In this article we will use single feature-based identification mode to only extract texture feature of weeds and then pre-process the images collected by us. On this basis we will use SVM classifier to identify them effectively and combine the RBF kernel function with the low-order polynomial kernel function to achieve optimization. Experimental results show that the proposed method is effective, and worth promotion.

## 2 Image acquisition and pre-processing

### 2.1 IMAGE ACQUISITION

In this study weed images will be collected from an experimental field. And we will use Nikon digital camera to capture images of the sample at three different times of sunny day (6:00, 10:00, and 16:00), and then store the original images. The common kinds of weeds include: cephalanoplos, pigweed, crabgrass and filed bindweed.

### 2.2 IMAGE PRE-PROCESSING

According to the requirement of image processing and analysis, there are three steps of image pre-processing operation: greying, enhancement and denoising and image segmentation.

#### *2.2.1 Greying*

There is quite a big difference between crop and weed images with soil images. The original images are of 24 bit true colour bitmap, which contains three components: R(red), G(green) and B(blue). And the green component in plant-growing period is most obvious and it has the largest difference with soil. In the processing of greying the original sample images, extra-green character is introduced. By combining the R, G, B three colour components, the greyscale image is generated. The specific approach [3] is shown in Equation (1).

$$E_G(\mathrm{x,y}) = 2G(x,y) - R(x,y) - B(x,y)\,. \tag{1}$$

In the above, $G(x,y)$, $R(x,y)$ and $B(x,y)$ stand for the R, G, B three colour components of the original image, and $E_G(x,y)$ is the extra-green value. Through this algorithm, we can obtain the grey image of the sample processing results.

#### *2.2.2 Enhancement and denoising*

In this processing the greyscale images can be improved by using filtering. In this study, neighbour average filtering and median filtering will be compared. There are some studies having shown that the two filtering methods have similar effects and most of the noise can be filtered [4]. In consideration of the smaller amount of calculation of median filtering and large number of image samples to be processed, we choose median filtering method to promote

[*] ***Corresponding author*** e-mail: tangjl@jsut.edu.cn

efficiency. At the same time, it can not only ensure the quality of processing, but also reduce the use of system resources. We divide the images to be processed into different S-windows, and take the pixel grey value in the windows as the grey value of each window. It is expressed in Equation (2).

$$g(x,y) = median\{\text{f(x-i,y-i)}\},\text{(i,j)} \in \text{S} \tag{2}$$

In the above the grey values of pixels in window are represented as $g(x,y)$, and $median()$ is the median value function. Take cephalanoplos for example, the following are the original image and processed image. Please see Figure 1.

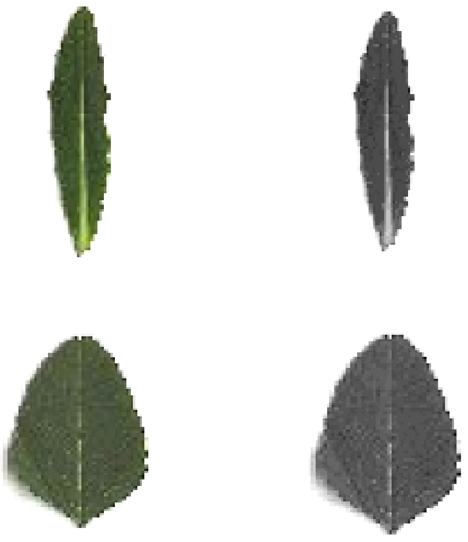

FIGURE 1 Comparison of the original image and processed image

*2.2.3 Image segmentation*

The image segmentation is implemented with threshold, which is selected with the iterative method [5, 6].

Firstly, traverse the pixel images to be processed to obtain the maximum and minimum values of the grey image, which are represented as $T_{max}$ and $T_{min}$ respectively;

Secondly, initialize the grey threshold value, and take the mid-value of maximum and minimum values, i.e. Mid-value $T0=(T_{max}+T_{min})/2$;

Then, take $T0$ as the threshold value, to divide the images to be processed into foreground and background regions, and get $M1$ the average grey value of foreground, and $M2$ the average grey value of background;

Lastly, combine $M1$ and $M2$ to obtain the new grey threshold value: $Tk+1=(M1+M2)/2$. And compare $Tk+1$ with last grey threshold value. If they are equal, $Tk+1$ is the final threshold value. If not, go on the iteration of step 2. Finally, we can obtain the binary image after segmentation.

## 3 Extraction of Texture Feature

In micro regularity, texture feature is often difficult to describe in quantitative way or qualitative way, but it has obvious macroscopic rules and the directional distribution regularity. The extracted texture features described herein are:

Average Value: $m = \sum_{i=0}^{L-1} z_i p(z_i)$;

Standard Deviation: $\sigma = \sqrt{\mu_2}$;

Energy: $E(\theta,d) = \sum_i^n \sum_j^n M_{(\theta,d)}(i,j)^2$;

Entropy: $H(\theta,d) = -\sum_i^n \sum_j^n M(\theta,d)(i,j)\log_2 M_{(\theta,d)}(i,j)$;

Moment of Inertia: $I(\theta,d) = \sum_i^n \sum_j^n (i-j)^2 M_{(\theta,d)}(i,j)$;

Dependency: $C(\theta,d) = \dfrac{\sum_i^n \sum_j^n i^* j^* M_{(\theta,d)}(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$;

Seven invariants of Hu invariant moment are introduced: Φ1—Φ7.

Part of the texture feature data is shown in the Table 1.

TABLE 1 Table of texture feature information

| Weed species | Cephalanoplos | Pigweed | Crabgrass | Filed bindweed |
|---|---|---|---|---|
| Average value | 91.939 | 96.547 | 93.612 | 112.334 |
| Standard deviation | 67.6944 | 66.6184 | 64.1488 | 76.97856 |
| Moment of inertia | 7.696 | 7.464 | 6.968 | 8.3616 |
| Energy | 0.1392 | 0.5512 | 0.4336 | 0.52032 |
| Dependency | 0.0848 | 0.0144 | 0.0144 | 0.01728 |
| Entropy | 5.0752 | 5.2128 | 5.212 | 6.2544 |

Part of the data from invariant moment is shown in the Table 2.

TABLE 2 Table of invariant moment information

| Invariant moment | Cephalanoplos | Pigweed | Crabgrass | Filed bindweed |
|---|---|---|---|---|
| 1 | 2.619 | 2.470 | 2.492 | 2.991 |
| 2 | 2.177 | 2.275 | 2.440 | 2.928 |
| 3 | 5.289 | 4.528 | 4.427 | 5.313 |
| 4 | 1.403 | 0.944 | 1.134 | 1.361 |
| 5 | 5.736 | 4.287 | 4.199 | 5.039 |
| 6 | 3.515 | 2.989 | 2.650 | 3.180 |
| 7 | 4.881 | 3.692 | 3.671 | 4.406 |

There are too many texture parameters of weeds, and to enhance efficiency, we select the SVM based on RBF kernel function:

Make $U \in R^n, V \in R^n, g \in R^+$, and the RBF kernel function is defined in Equation (3).

$$K(U,V) = \exp(-g \| U - V \|^2) \quad (3)$$

In the above, $R$ represents a vector space, and $g$ is the parameter of kernel function.

## 4 Weed identification

### 4.1 SVM-BASED METHOD

In this study SVM (support vector machine) model is used for recognition. When using SVM into various recognition patterns, it will be faced with the problem of determination of the specific function and method parameters. At present, there is no highly-common and efficient mode, and many studies are based on trial and error method or empirical method to select the parameters. In order to identify the various types of weeds accurately, the Stprtool toolbox in MATLAB will be introduced in this study, and the way of one-to-many will be used.

### 4.2 DATA STANDARDIZATION

A lot of weed image data in this study should be initialized before being imported into MATLAB. That means we should convert initial data format supported by the Stprtool toolbox of MATLAB. Standard format of various types of image data can be expressed as:

<value1><value2> … < valuen>

The training data and test data of image for SVM are all of $m \times 1$ matrix structure. By extracting the specific values of the training set and test set from weed images, the identification results (expressed as integer) is generated. This study focuses on common four classes among the corn crop weeds, so the target values fall in the interval of [1,4]. In consideration of the six texture features (Average Value, Standard Deviation, Moment of Inertia, Energy, Dependency, Entropy) and seven invariant moment (Invariant Moments 1-7), the value of <value> is determined to be a single-precision positive real number.

### 4.3 DATA NORMALIZATION

To improve the efficiency of operation, and avoid the impact from a too-large or too-small characteristic parameters, all of the data should be normalized. Another reason for normalization is to enhance the processing efficiency of the system. Due to that, the SVM algorithm involves core computing, and it will call EXP algorithm, only the normalized data will not consume too much system resources in these operations.

### 4.4 EXPERIMENTAL PROCESS AND RESULTS

There are 20 images of each weed, 80 experimental samples in total. 40 weed images are selected to build SVM test set randomly and the rest 40 images used as test set. 13 characteristics are used as SVM input. In order to select the best method, the present study compares the RBF kernel function, Polynomial kernel function and Linear kernel function, in which training and testing are conducted respectively.

#### *4.4.1 Several kinds of kernel functions performance analysis*

For the 80 image samples, we use RBF, Linear and Polynomial kernel functions to test them respectively. For Polynomial kernel function, the tests of from second order to fifth order are conducted. The Table 3 shows the results.

TABLE 3 Results of several kernel functions

| Function Name | Parameter | Support vector | Identification rate |
|---|---|---|---|
| RBF kernel function | Penalty factor=32 σ =0.5 | 189 | 90% |
| Linear kernel function | Penalty factor =1000 | 178 | 78.6% |
| Polynomial kernel function (Second order) | Penalty factor =1000 | 178 | 85% |
| Polynomial kernel function (Third order) | Penalty factor =1000 | 177 | 76.3% |
| Polynomial kernel function (Fourth order) | Penalty factor =1000 | 174 | 75% |
| Polynomial kernel function (Fifth order) | Penalty factor =1000 | 175 | 73.8% |

The test data shows that the best performance is RBF (radial basis function), and the rate of weed identification is 90%; the next is Polynomial kernel function (second-order), and the identification rate is 85%. The others are under 80%. It shows that the lower-order Polynomial kernel function is more appropriate in weed identification. The reason is that, although the learning ability of the system increases with the order, the complexity of the system also increases exponentially, and the "over learning" situation may happen.

#### *4.4.2 Optimization of the classifier*

In the process of SVM identification, each kernel function has their own characteristics and is suitable for different situations. From the perspective of bigger categories, kernel function includes global function and local function. The polynomial kernel function is a standard global function. It has a higher adaptability. Meanwhile, RBF belongs to standard local function, and its influence is limited to the test sample data, which makes its interpolation ability high and better than polynomial function.

In this study, we make use of the advantages of the above two types of kernel functions in an innovative way. They can be combined to achieve better identification results. This thinking also proves that the idea of SVM-based learning performance and generalization performance is great.

Based on the above research, RBF has a better interpolation ability and the Polynomial kernel function (second order) has a better extrapolating ability. The combined kernel function is constructed as follows. See Equation (4).

$$K(x,y)=\lambda_1\cdot(xy+1)^q+\lambda_2\cdot\exp\{-\frac{\|x-y\|^2}{2\sigma^2}\} \quad (4)$$

In the above, $\lambda_1$ and $\lambda_2$ stand for the weight of RBF and Polynomial kernel function (second order) in the combination kernel function, and they should be $\lambda_1+\lambda_2=1$. K(x,y) is kernel function, $x\in R^d$ and $y\in[-1,+1]$ are class identifiers; $(xy+1)^q$ is RBF; $\exp\{-\frac{\|x-y\|^2}{2\sigma^2}\}$ is Polynomial kernel function. Take different $\lambda_1$ and $\lambda_2$, and the simulation data is shown in the Table 4.

TABLE 4 Simulation data of optimized classifier

| λ1 | λ2 | Identification rate |
|---|---|---|
| 0.9 | 0.1 | 81.25% |
| 0.8 | 0.2 | 81.25% |
| 0.7 | 0.3 | 80% |
| 0.6 | 0.4 | 78.75% |
| 0.5 | 0.5 | 78.75% |
| 0.4 | 0.6 | 81.25% |
| 0.3 | 0.7 | 78.75% |
| 0.2 | 0.8 | 92.5% |
| 0.1 | 0.9 | 78.75% |

The data shows that the identification rate of the system vary greatly with the different λ1 and λ2 values. When λ1=0.2 and λ2=0.8, the combined function generates the highest identification rate 92.5%, higher than the conventional SVM kernel function result. Therefore, the performance of this kind of optimized classifier is better than single kernel function.

## 5 Conclusions

In this paper, a method of image processing is used to identify texture features of crop weeds. On the basis of traditional SVM algorithm, a combined identification mode is constructed. The experiment confirmed a better recognition result. The improved method helps reduce the amount of herbicide spraying, and realize intelligent management and operation of the weeding system.

## Acknowledgments

The authors are indebted to their colleagues of School of Computer Engineering, Jiangsu University of Technology, for their suggestive discussion and zealous help. This study is supported by the IUR Prospective Joint Project of Jiangsu Province (BY2014038-05), and the Youth Scientific Research Foundation of Jiangsu University of Technology (No. KYY13027).

## References

[1] Drăgut L, Tiede D, Levick S R 2010 ESP: a tool to estimate scale parameter for multi group resolution image segmentation of remotely sensed data *International Journal of Geographical Information* **24**(6) 859-71

[2] Han N, Wang K, Yu L, Zhang X 2012 Integration of texture and landscape features into object-based classification for delineating torreya using ikonos imagery *International Journal of Remote Sensing* **33**(7) 2003-33

[3] Jin H, Li P, Cheng T 2012 Land cover classification using CHRIS/PROBA images and multi-temporal texture *International Journal of Remote Sensing* **33**(1) 101-19

[4] Kabir S, He D C, Samusi MA 2010 Texture analysis of IKONOS satellite imagery for urban land use and land cover classification *The Imaging Science Journal* **58** (3) 163-70

[5] Ota T, Mizoue N, Yoshida S 2011 Influence of using texture information in remote sensed data on the accuracy of forest type classification at different levels of spatial resolution *Journal of Forest Research* **16**(6) 432-7

[6] Shen J, Liu LP, Lin X 2011 Cropland extraction from very high spatial resolution satellite imagery by object-based classification using improved mean shift and one-class support vector machines *Sensor Letters* **9**(3) 997-1005

## Authors

**Jie Zhang, born on March 18, 1983, China**

**Current position:** researcher at Jiangsu University of Technology, China.
**University studies:** Msc. degree in computer engineering from Yangzhou University, in 2012.
**Scientific interest:** software development and image processing.
**Publications:** 10 papers.

**Fuli Song, born on July 3, 1982, China**

**Current position:** lecture of the personnel division at Henan Radio & Television University, China.
**University studies:** Msc. Degree in software engineering from University of Electronic Science and Technology, in 2013.
**Scientific interest:** software engineering and image processing.

**Jiali Tang, born on May 10, 1980, China**

**Current position:** associate professor at Jiangsu University of Technology, China.
**University studies:** Msc. Degree in Material Processing Engineering from Nanjing University of Aeronautics and Astronautics, in 2005.
**Scientific interest:** industrial control software development and image super-resolution reconstruction.
**Publications:** 20 papers.