ICIC Express Letters ICIC International ©2016 ISSN 1881-803X
Volume **10**, Number **5**, May **2016** pp. 1023–1028

WEED SEEDS IDENTIFICATION BASED ON INTERLACING PCA NET

Cailing Liu and Cheng Cai*

Department of Computer Science
College of Information Engineering
Northwest A&F University
No. 3, Taicheng Road, Yangling 712100, P. R. China
*Corresponding author: cheney.chengcai@gmail.com

Received October 2015; accepted January 2016

Abstract. *In this paper, an interlacing PCA net is proposed to extract weed seeds image features for automatic identification. Our interlacing PCA net has two layers: the first layer extracts 9 channel features through convolution with eigenvectors, and the second layer interlaces 9 channels into 3 subsets. In this model, the convolution filters of every stage are obtained by means of Principal Components Analysis (PCA) applied to the patches extracted from weed seed images. Interlacing combinations of the first stage output are the second stage input. Then a linear SVM is chosen as the classifier. Compared with the original tree-structure PCA net, our interlacing PCA net captures more information between channels. Experimental performance evaluation based on 211 species dataset shows that the recognition rate of our proposed interlacing PCA net arrives at 96.44%.*
Keywords: Seed identification, PCA network, Interlacing combinations, SVM classifier

1. **Introduction.** The manual identification of weed seeds by specialized technicians is rather slow and hard to quantify, both in its commercial value and in its technological implication. It is then of major technical and economic importance to implement computer vision-based methods for more reliable and faster classifications of weed seeds. The biggest challenge of computer vision-based methods is how to represent the difference of varied classes and the similarity within the classes. So how to represent images with more distinctive features is the key to classification.

Numerous experiments have been conducted on weed seeds for improving the classification accuracy by extracting features from images, but most of them are low-level features. Chtioui and Bertrand [1] extracted the size, shape and texture parameters from digitized color images of whole seed samples of rumex, wild oat, Lucerne and vetch in this investigation. Two pattern recognition approaches were attempted in the classification: stepwise discriminant analysis and artificial neural network. Granitto and Verdes [2,3] proposed a machine vision-based method to identify large weed seeds database. This investigation extracted 6 morphological, 4 color and 2 textural parameters from weed seed images and used naïve Bayes classifier and artificial neural network systems for classification. To perform the recognition of color weed seeds images, Zhao et al. [4] added color element to traditional PCA method. This method used the feature of 3D color tensor to generate vector spaces, and then color PCA was used to map the high dimensional space into low dimensional subspace, and extract features. Wafy et al. [5] extracted local features by using Scale-invariant feature transform (SIFT) [6] descriptor for weed seeds classification. Many other local descriptors could also be used to extract features, such as Binary robust invariant scalable key-points (BRISK) [7] and Fast retina key-point (FREAK) [8] descriptors. Many local features are concatenated together or add spatial feature pyramids [9,10] to improve the classification accuracy, but this will make the vector for image representation rather long. While the low-level features can be hand-crafted with great success for specific data and tasks, designing effective features for new data and tasks usually requires a new domain of knowledge since most hand-crafted features cannot be adopted directly in new conditions.

The idea of deep learning is to discover multiple levels of representation in the hope that higher-level features will represent more abstract semantics of data so that deep convolutional neural networks can draw significant attention because of its successful application in ImageNet classification [11]. A convolutional deep neural network architecture consists of multiple trainable stages stacked on top of each other, followed by a supervised classifier. Each stage generally comprises convolutional filter bank layer and feature pooling layer. Weights in each layer are usually initialized from a zero-mean Gaussian distribution, e.g., T.-H. Chan et al. [12] initialized the weights through PCA method and received good performance. K-nearest neighbor (KNN) [13], Bayes classifier [14] and SVM classifier [15] are the most commonly used classifiers in machine learning. The principle of Bayesian classifier calculates the posterior probability using prior probability of an object and Bayesian formula; the prior probability needs a large number of samples to get more accurate result. The idea of KNN is that if most of the k adjacent samples of the test sample belong to a certain category, then the whole test sample also belongs to this category. The KNN need compare the feature of the test sample with that of the training samples. The SVM classifier finds a hyperplane which separates two-class data with maximal margin so that the classification accuracy is high and the classification speed is independent of the number of the training samples. In this paper, we get the classification result through SVM classifier.

Inspired by the PCA Network of Tsung-Han Chan and the deep convolutional neural network, this paper designs a simple deep convolutional network which involves two stages for the classification of weed seeds, and the weights in every stage are initialized through PCA [16] method. Our experiments are conducted on a much larger database with 9,192 seed images of 211 common weed species, and some samples of this database are shown in Figure 1.

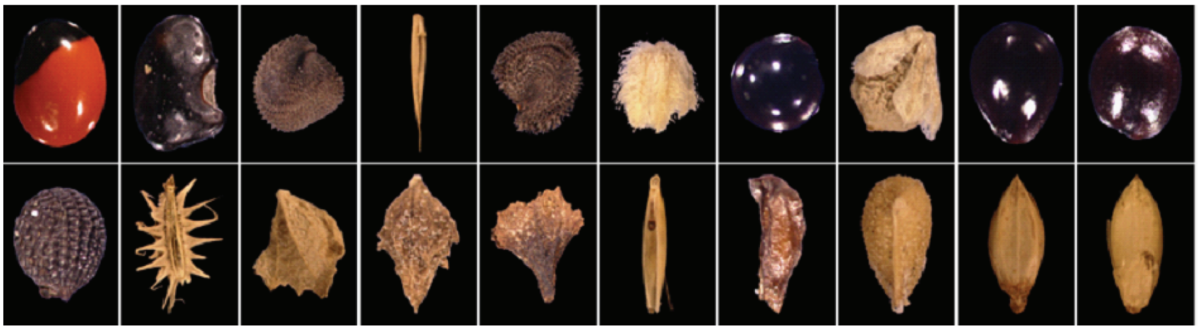


FIGURE 1. Weed seeds images

The remainder of this paper is organized as follows: Section 2 introduces how to build the interlacing PCA net and get the classification result through SVM classifier; Section 3 displays image classification results on the original test images and its artificially deformed images; and finally in Section 4, conclusions are drawn and some future research issues are also discussed.

2. **Interlacing PCA Net.** Suppose that we are given N input training images $\{I_i\}_{i=1}^{N}$ of size $m \times n$, and assume that the patch size is $k_p \times k_p$ at pth stages.

2.1. **Convolution filter extraction using PCA.** Around each pixel of input image in the first layer, we take a $k_1 \times k_1$ patch, and collect all overlapping patches of the ith image; i.e., $x_{i,1}, x_{i,2}, \ldots, x_{i,s_1s_2} \in R^{k_1k_1}$ where each $x_{i,j}$ denotes the jth vectorized patch in I_i and $s_1 = (m - k_1 + 1)$, $s_2 = (n - k_1 + 1)$. We then subtract the patch mean from each patch, and obtain $\overline{X}_i = [\overline{x}_{i,1}, \overline{x}_{i,2}, \ldots, \overline{x}_{i,s_1s_2}]$, where $\overline{x}_{i,j}$ is a mean-removed patch. After constructing the same matrix for all the input images and putting them together, we get

$$X = \left[\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_N\right]. \tag{1}$$

Assuming that the number of the filters in the first layer is L_1, PCA [16] minimizes the following reconstruction error within a family of orthonormal filters,

$$\min_{V \in R^{k_1k_1 \times L_1}} \|X - VV^TX\|_F^2, \quad \text{s.t.} \quad V^TV = I_{L_1}, \tag{2}$$

where I_{L_1} is the identity matrix of size $L_1 \times L_1$. The solution of Equation (2) is L_1 principal eigenvectors of XX^T. The PCA filters are then expressed as:

$$w_l^1 = mat_{k_1,k_1}\left(q_l\left(XX^T\right)\right) \in R^{k_1 \times k_1} \tag{3}$$

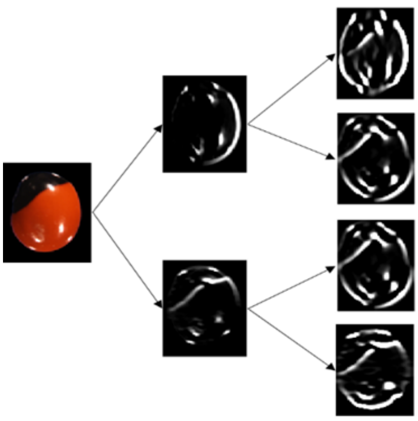
where $mat_{k_1,k_1}(v)$ is a function that maps a vector $v \in R^{k_1k_1}$ to a matrix $w \in R^{k_1 \times k_1}$, and $q_l(XX^T)$ denotes the lth principal eigenvector of XX^T. If the input image is colorized, we will take a $k_1 \times k_1 \times 3$ patch, and we map the eigenvector as a 3D matrix $w_l^1 \in R^{k_1 \times k_1 \times 3}$.

2.2. **Two-layer model.** In this paper, we have a 2 layer model which has good performance. Figure 2(a) shows how a two-stage PCA Network designed by Tsung-Han Chan et al. extracts features from an input image. We think that the model has three inputs in the first layer because of the colorized image, so every output of the first layer is formed by convolving three inputs with the corresponding filter and summing them up. For every input original image I_i and its three image channel $I_{i,c}$ $(c = 1, 2, 3)$, we will get L_1 outputs $Z_i^l \in R^{s_1 \times s_2}$ through the corresponding filter $w_{l,c}$:

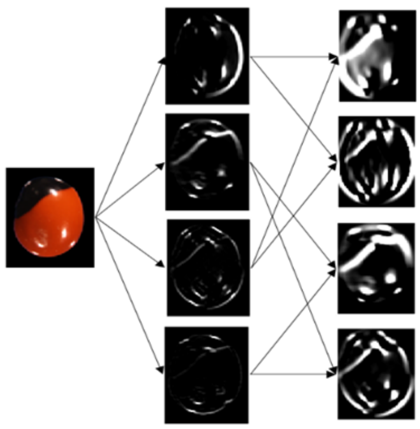
$$Z_i^l = \sum_{c=1}^{3} \sum_{l=1}^{L_1} I_{i,c} * w_{l,c}^1 \tag{4}$$

where $*$ is the 2D convolution operator and $w_{l,c}^1$ is the filter corresponding to $I_{i,c}$. However, the output of the second layer has only one input. Almost in every deep network, the output is formed by the sum of several or all inputs' convolution result and the higher abstract features are the combination of the lower features. For this reason, we modify the PCA Network by making several or all the first stage output as the input of the second stage output. The modified model is shown in Figure 2(b).

In the second stage, we assume that $A_i^1 \in R^{s_1 \times s_2 \times U_1}$ formed by U_1 outputs of the first stage as the input of one output of this stage. Like the first stage, we take a $k_2 \times k_2 \times U_1$



(a) Tree structure PCA net [12]



(b) Interlacing PCA net

Figure 2. Two-layer convolutional model

patch, and collect all patches of A_i^1, subtract patch mean from each patch, and form $\overline{Y}_i^1 = \left[\overline{y}_{i,1}^1, \ldots, \overline{y}_{i,s_3 s_4}^1\right]$, where $\overline{y}_{i,j}^1$ is a mean-removed patch and $s_3 = (s_1 - k_2 + 1)$, $s_4 = (s_2 - k_1 + 1)$. By constructing the matrix for all input images and putting them together, we get $Y^1 = \left[\overline{Y}_1^1, \ldots, \overline{Y}_N^1\right] \in R^{U_1 k_2 k_2 \times N s_3 s_4}$. Assuming the second stage has U groups of input, we get $Y = \left[Y^1, \ldots, Y^U\right] \in R^{U_1 k_2 k_2 \times U N s_3 s_4}$. The PCA filters of the second stage are obtained as

$$w_l^2 = mat_{k_2, k_2, U_1}(q_l(YY^T)) \in R^{k_2 \times k_2 \times U_1}, \quad l = 1, 2, \ldots, L_2. \tag{5}$$

For the input A_i^1, we will get L_2 outputs:

$$O_i^1 = \left\{A_i^1 * w_l^2\right\}_{l=1}^{L_2}. \tag{6}$$

For each input of the second stage, i.e., it has L_2 outputs. These outputs are binarized and marked as $H(O_i^1)$, where $H(\cdot)$ is a Heaviside step function whose value is one for positive entries and zero otherwise. For all the L_2 outputs, we view the vector of L_2 binary bits as a decimal number. This converts the L_2 outputs in O_i^1 to a single integer-valued image:

$$T_i^1 = \sum_{l=1}^{L_2} 2^{l-1} H\left(O_i^{1,l}\right) \tag{7}$$

where $O_i^{1,l}$ is the lth output in O_i^1. Every pixel in the image is an integer in the range of $[0, 2^{L_2} - 1]$.

For each of the U images T_i^u, $u = 1, \ldots, U$, we partition it into B blocks. We compute the histogram with 2^{L_2} bins of the decimal values in each block, and concatenate all B histograms into one vector and denote them as $\mathrm{Hist}(T_i^u)$. The feature of the input image I_i is defined as:

$$f_i = \left[\mathrm{Hist}\left(T_i^1\right), \ldots, \mathrm{Hist}\left(T_i^U\right)\right]^T \in R^{2^{L_2} UB}. \tag{8}$$

2.3. **Classifier.** This paper takes SVM as the classifier to get the classification results. As every sample in the training set has a class label, we get the feature f of the sample through the two-layer network so as to get $F = [f_1, \ldots, f_N] \in R^{2^{L_2} UB \times N}$ and $La = [la_1, \ldots, la_N] \in R^N$, where la_i represents the class label of the sample i. We build an SVM model for every two classes by using their features and class labels. For a test sample, we extract features through the interlacing PCA network and then get the possible class labels through all SVM models. The class label of the test sample is the one with the highest number of votes of all SVM models.

3. **Experiments.** Our experiment is conducted on a weed seeds database which has 9,192 images of 211 species. We split the database into a training set and a testing set. For weed seed identification, we randomly choose, for each species, 80% of the images as the training samples to build the classifier and the remaining 20% as the testing samples. This leaves a large database with 7,365 images for training and also a fairly large testing set with 1,827 images. The size of each image in this dataset is 110×80 and each image is colorized. All of the results quoted below correspond to an average over 10 independent experiments.

In this paper, we set $L_1 = 9$ and $L_2 = 9$ and the non-overlapping block as 19×21 in size. We vary the size of the patch of the two stages from 3 to 21. The results are shown in Figure 3, and we find that this experiment will have better performance when $k_1 = 5$ and $k_2 = 7$.

We improve the PCA net originally proposed by Tsung-Han Chan in that in the second stage, several or all the first stage outputs are the inputs jointly of every output. We make the comparison of these two networks with the original images and introduce artificial deformation to the image with a rotation, scaling or translation, as shown in Figure 4.

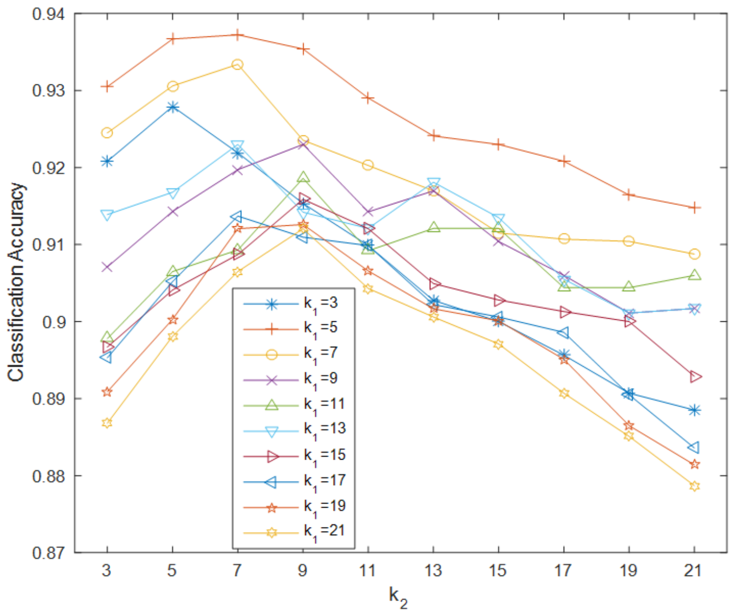


FIGURE 3. Classification accuracy with different patch sizes

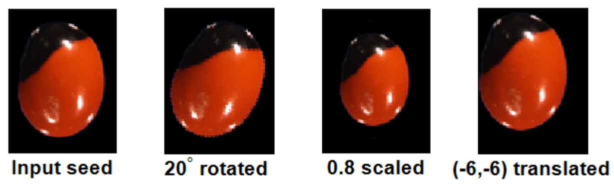


FIGURE 4. Original image and its artificially deformed images

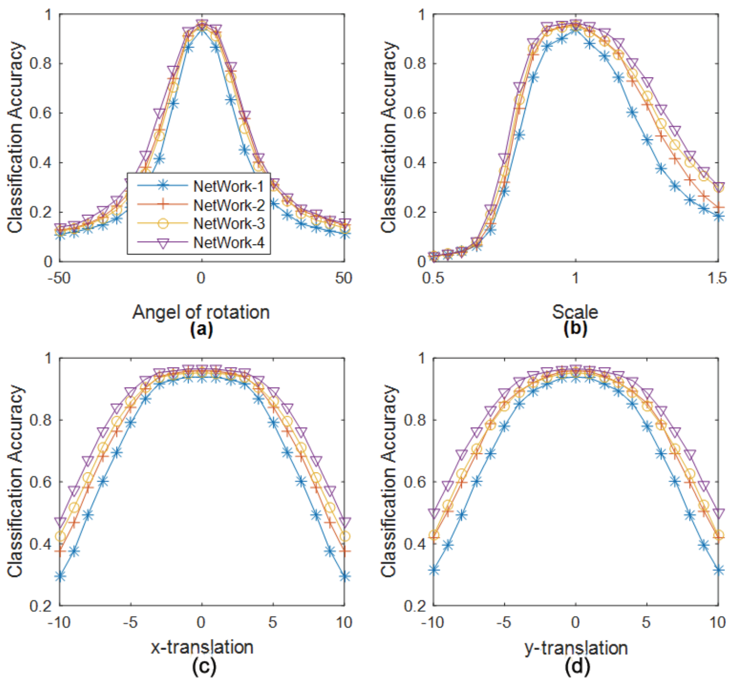


FIGURE 5. Classification accuracy of different networks on testing set

Figure 5 shows the classification accuracy with different networks on the original test images and its artificially deformed images, where NetWork-1 is proposed by Tsung-Han Chan, NetWork-2 represents the input groups $(1, 4, 7)$, $(2, 5, 8)$ and $(3, 6, 9)$ in the second stage formed by the output of the first stage, NetWork-3 represents the input groups $(1, 2, 3)$, $(4, 5, 6)$ and $(7, 8, 9)$, and NetWork-4 represents the combination groups in NetWork-2 and NetWork-3. It has been found that compared with the network proposed by Tsung-Han Chan, the modified network has improved the classification results and different groups performed differently due to the different distinction of higher abstract features in combination with the lower features. The network possesses robustness to image translation because the filter is shared in the whole image.

4. **Conclusion.** In this paper, we extract features from the construction of a 2 layer interlacing PCA net to represent weed seed images for classification performance. The best classification rate of the network is 96.44%. Compared with the tree-structure PCA net proposed by Tsung-Han Chan, ours performed better in classification and this indicated that the abstract features combined by the low level features can represent the image better. In the future, we can have more combinations to the prior outputs which are the inputs of this stage and build a deeper neural network to extract higher abstract features to represent images. We can also take layer by layer training for the filters.

Acknowledgements. This work has been supported by the National Natural Science Foundation of China (No. 61202188).

REFERENCES

[1] Y. Chtioui and D. Bertrand, Identification of seeds by colour imaging: Comparison of discriminant analysis and artificial neural network, *Journal of the Science of Food and Agriculture*, vol.71, no.4, pp.433-441, 1996.

[2] P. M. Granitto et al., Weed seeds identification by machine vision, *Computers and Electronics in Agriculture*, vol.33, no.2, pp.91-103, 2002.

[3] P. M. Granitto, P. F. Verdes and H. A. Ceccatto, Large-scale investigation of weed seed identification by machine vision, *Computers and Electronics in Agriculture*, vol.47, no.1, pp.15-24, 2005.

[4] F. Zhao, C. Cai and J. Zhu, Weed seeds recognition using color PCA, *International Symposium on Knowledge Acquisition and Modeling*, Wuhan, vol.1, pp.375-378, 2009.

[5] M. Wafy, H. Ibrahim and E. Kamel, Identification of weed seeds species in mixed sample with wheat grains using SIFT algorithm, *International Computer Engineering Conference*, Giza, pp.11-14, 2013.

[6] H. R. Kher and V. K. Thakar, Scale invariant feature transform based image matching and registration, *The 5th International Conference on Signal and Image Processing*, Jeju Island, pp.50-55, 2014.

[7] S. Leutenegger, M. Chli and R. Y. Siegwart, BRISK: Binary robust invariant scalable keypoints, *IEEE International Conference on Computer Vision*, Barcelona, vol.58, no.11, pp.2548-2555, 2011.

[8] A. Alahi, R. Ortiz and P. Vandergheynst, FREAK: Fast Retina keypoint, *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, pp.510-517, 2012.

[9] S. Lazebnik, C. Schmid and J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.2, pp.2169-2178, 2006.

[10] L. Seidenari et al., Local pyramidal descriptors for image recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.36, no.5, pp.1033-1040, 2014.

[11] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Neural Information Processing Systems*, 2012.

[12] T. H. Chan et al., PCANet: A simple deep learning baseline for image classification? *IEEE Trans. Image Processing*, vol.24, no.12, pp.5017-5032, 2014.

[13] J. Zhang, Y. Niu and H. Nie, Web document classification based on fuzzy k-NN algorithm, *International Conference on Computational Intelligence and Security*, vol.1, pp.193-196, 2009.

[14] T. M. Mitchell, *Machine Learning*, WCB/McGraww-Hill, Boston, 1997.

[15] J. Chen and L. Jiao, Classification mechanism of support vector machines, *International Conference on Signal Processing Proceedings*, Beijing, vol.3, pp.1556-1559, 2000.

[16] A. M. Martinez and A. C. Kak, PCA versus LDA, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.23, no.2, pp.228-233, 2001.