Computers and Electronics in Agriculture 125 (2016) 48–55

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Local descriptors for soybean disease recognition

Rillian Diello Lucas Pires [a], Diogo Nunes Gonçalves [a], Jonatan Patrick Margarido Oruê [a], Wesley Eiji Sanches Kanashiro [a], Jose F. Rodrigues Jr. [b], Bruno Brandoli Machado [a,b], Wesley Nunes Gonçalves [a,*]

[a] Federal University of Mato Grosso do Sul, Rua Itibire Vieira, s/n, Ponta Pora, MS 79907-414, Brazil
[b] University of Sao Paulo, Av. Trabalhador Sancarlense, 400, Sao Carlos, SP 13566-590, Brazil

A R T I C L E   I N F O

Article history:
Received 29 September 2015
Received in revised form 28 April 2016
Accepted 29 April 2016
Available online 5 May 2016

Keywords:
Soybean disease recognition
Local descriptors
Bag-of-visual-words
Computer vision

A B S T R A C T

The detection of diseases is of vital importance to increase the productivity of soybean crops. The presence of the diseases is usually conducted visually, which is time-consuming and imprecise. To overcome these issues, there is a growing demand for technologies that aim at early and automated disease detection. In this line of work, we introduce an effective (over 98% of accuracy) and efficient (an average time of 0.1 s per image) method to computationally detect soybean diseases. Our method is based on image local descriptors and on the summarization technique Bag of Visual Words. We tested our approach on a dataset composed of 1200 scanned soybean leaves considering healthy samples, and samples with evidence of three diseases commonly observed in soybean crops – Mildew, Rust Tan, and Rust RB. The experimental results demonstrated the accuracy of the proposed approach and suggested that it can be easily applied to other kinds of crops.

1. Introduction

Soybean is one of the most important crops due to its beneficial effects on human health, to its role as a major nutrition source, and to its economic importance. It has been widely used in food and industrial applications because of its high protein and oil concentrations (Kumar et al., 2010). Soybean occupies very large crops in which the monocropping and conservation tillage are commonly used. Such cultivation systems, however, have favored the occurrence of a large number of diseases (Carmona et al., 2015) causing major economic losses. The solution is to apply preventive agrochemicals; but, because the identification of where the infestation took place is time-consuming, the usual practice is to use agrochemicals over the entire crop instead of only over specific subareas. This is an expensive practice that spreads unnecessary chemicals over terrain and air.

Accordingly, a more precise detection of the disease spots in the crop is an important step to decrease economic losses, to prevent the spread of diseases, and to reduce environmental pollution. Despite its importance, it is usually conducted visually by an expert (Moshou et al., 2004), an imprecise and time-consuming process, especially when carried out over large-scale farms. Alternatively, disease detection techniques based on chemical reactives are available, such as the ELISA (enzyme-linked immunosorbent assay) method and the PCR (polymerase chain reaction) method (Saponari et al., 2008; Yvon et al., 2009; Gutiérrez-Aguirre et al., 2009), however, they are expensive processes. Consequently, there is a demand for rapid and cheaper detection methods.

In this context, one active line of research is the use of image processing techniques. The idea is to have the computer analyze images of soybean leaves (and of other cultures) to detect diseases by means of pattern recognition methods. Gui et al. (2015), for example, proposed a method for soybean disease detection based on *salient regions* and *k-means clustering*. Shrivastava and Hooda (2014) proposed a method for detecting *brown spot* and *frog eye*, two common soybean diseases; they used shape features and *k-nearest neighbors classification*. Ma et al. (2014) proposed a technique for detecting insect-damaged vegetable soybean using *hyperspectral imaging*. A study to discriminate soybean leaflet shape using *neural networks* was proposed in the work of Oide and Ninomiya (2000). Yao et al. (2012) used *hyperspectral images* to study the damage caused by the herbicide glyphosate on soybean plants. Cui et al. (2010) reported image processing techniques for quantitatively detecting rust severity from soybean multispectral images.

Besides soybean, other cultures have been studied in the literature, such as the work performed by Rumpf et al. (2010), which presents an automatic system for classification of foliar sugar beet

* Corresponding author.
E-mail address: wesley.goncalves@ufms.br (W.N. Gonçalves).

http://dx.doi.org/10.1016/j.compag.2016.04.032
0168-1699/

diseases based on *Support Vector Machines* and *spectral vegetation indices*. Moshou et al. (2004) investigated the automatic recognition of yellow rust in wheat using *reflectance measurements* and *neural networks*. Liu et al. (2010) applied techniques *neural network* and *principal components analysis* for classifying fungal infection levels in rice panicles. Imaging techniques are also applied in the recognition of plant species (Gonçalves and Bruno, 2013; Casanova et al., 2012; Backes et al., 2010). A review of techniques for detecting plant diseases can be found in the work of Sankaran et al. (2010); a survey on methods that use digital image processing techniques to detect plant diseases is presented in the work of Barbedo (2013).

This paper proposes a novel approach for soybean disease recognition based on techniques *local descriptors* and *bag-of-visual words*. We experiment with five local descriptors (SURF, HOG, DSIFT, SIFT, and PHOW, as detailed in Section 3) applied over a large set of digital images (gray scale and colored) acquired from a real-world soybean plantation in Brazil. The proposed approach is applied to scanned images (visible spectrum to the human eye), which does not require hyperspectral images and, therefore, can be used with commodity hardware such as smartphones. From the extracted features, we calculate a summary (lower-dimensional) feature vector using technique *bag of visual words* (BOVW).

The use of local descriptors is attractive because they are distinctive, robust to occlusion, and do not require segmentation. Due to these advantages, several local descriptors have been proposed in the literature. Scale-invariant feature transform (SIFT) (Lowe, 2004) is one of the most used and known local descriptors. Due to its good performance, SIFT was later applied at dense grids, known as Dense SIFT (Vedaldi and Fulkerson, 2010; Liu et al., 2011), and at multiscale dense grids, known as Pyramid histograms of visual words (PHOW) (Bosch et al., 2007). SIFT also inspired other local descriptors such as Speeded-up Robust Features (SURF) (Bay et al., 2008), which uses some approximations to achieve better performance. Histogram of oriented gradients (HOG) (Dalal and Triggs, 2005) has also been widely used and has shown interesting results, especially in pedestrian recognition. The reader may refer to the work of Mikolajczyk and Schmid (2005) for a review of local descriptors applied over images in general.

For classification purposes – considering classes *disease* and *no disease*, we use the supervised machine learning technique Support Vector Machine (SVM) having as input the BOVW vectors. We evaluate our classification using classic ten-fold cross-validation and the metric *Correct Classification Rate* (CCR). In our experiments, descriptor PHOW, over colored images, achieved the highest performance. Therefore, we contribute by (i) introducing a systematic method for computational identification of diseases in soybean leaves; (ii) conducting an experiment over soybean that is unprecedented in its control, methodology, and scale; (iii) empirically comparing the main local descriptors found in the literature, providing guidance for future works on image-based classification.

The rest of this paper is organized as follows. Section 2 describes the fundamentals of bag-of-visual-words. The five local descriptors evaluated in this work are described in Section 3. Section 4 details the experimental design and the image dataset of soybean leaves, while Section 5 describes the results of the proposed approach. Finally, Section 6 concludes the paper and suggests future works.

2. Bag-of-visual-words – BOVW

The bag-of-visual-words (BOVW) (Csurka et al., 2004) is a popular model for image recognition inspired by the bag-of-words (BOW) used in natural language processing. According to BOVW, descriptors are extracted from images in order to build a vocabulary of visual words. Given the vocabulary, each descriptor of an image is assigned to one visual word and then a histogram of visual word occurrences is obtained to represent the image. Basically, this model can be described in the following steps (Fig. 1):

(1) *Feature extraction:* in the first step, a local descriptor (such as SIFT or SURF) is applied to the images to extract a set of descriptors that are important to recognition tasks. For an image I, a set $D_I = [d_1, \dots, d_N]$ is composed by N descriptor vectors $d_i \in \Re^M$, where M is the dimension that depends on the local descriptor method. The N vector descriptors can be extracted from salient regions by detecting keypoints or in a dense grid of the image. Keypoints are locations of interest in an image; they are invariant to linear transformations and are able to boost description techniques. We use the SIFT keypoint detection (Lowe, 2004). Fig. 1(a) illustrates the feature extraction step.

(2) *Vocabulary construction:* given P training images, we build a set with P descriptors $D = [D_{I_1}, \dots, D_{I_P}]$, so that D contains information from all training images. To build the visual vocabulary, a clustering algorithm is applied for grouping D such that descriptors in the same cluster are more similar to each other than to those in other clusters. Each cluster summarizes similar regions of the images by a single mean vector called centroid. In the bag-of-visual-words, the centroid is also called visual word. The most common clustering algorithm is the k-means (Kanungo et al., 2002) due to its simplicity and computational cost. The k-means partitions the descriptor set D into k groups by minimizing the Euclidean distance between descriptors. Thus, k visual words $C = [c_1, \dots, c_k], c_i \in \Re^M$ composes the visual vocabulary. This step is illustrated in Fig. 1(b).

(3) *Histogram of visual word occurrence:* given an image I, its descriptor set $D_I = [d_1, \dots, d_N]$, and the visual vocabulary $C = [c_1, \dots, c_k]$, we build a histogram $h_I \in \mathbb{N}^k$ with k bins, one for each visual word. Then, each descriptor $d_i \in D_I$ is assigned to the bin whose visual word is the closest to d_i according to the Euclidean distance. An example of this step can be seen in Fig. 1(c). An example of this step can be seen in Fig. 1(c).

3. Local descriptors

This section briefly describes the local descriptors used by our method for soybean disease recognition. We use gray-scale images and the local descriptors are applied in sub-regions of the leaf images, suggesting the use of local methods. Several methods for generating local descriptors have been reported in the literature and can be used as a previous step, such as selective search (Uijlings et al., 2013), objectness (Alexe et al., 2012), category-independent object proposals (Endres and Hoiem, 2014), constrained parametric min-cuts (CPMC) (Carreira and Sminchisescu, 2012), and multi-scale combinatorial grouping (Arbeláez et al., 2014).

3.1. Scale-invariant feature transform – SIFT

Scale-invariant feature transform – SIFT (Lowe, 2004) is one of the most important local descriptors. Basically, SIFT detects keypoints at multiple scales and extracts a local descriptor for each keypoint that is invariant to uniform scaling, translation, rotation and partially invariant to illumination changes. SIFT can be described in the three steps below.

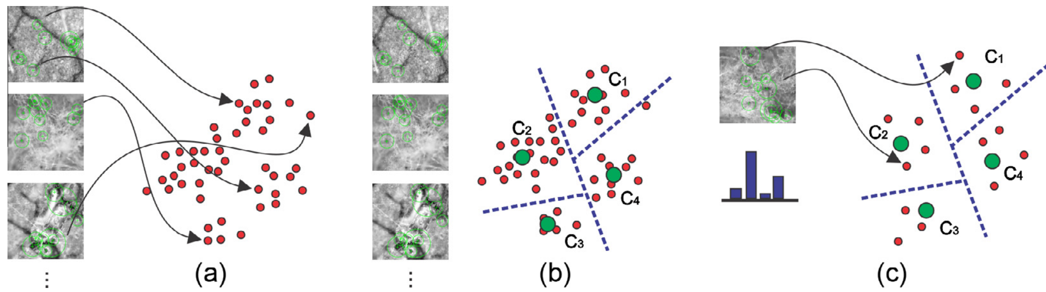


Fig. 1. Illustration of the bag-of-visual-words steps. (a) First, a local descriptor is applied to the images to identify keypoints and extract the descriptor vectors. (b) Then a clustering algorithm is applied to the descriptor vectors in order to obtain visual words. (c) Finally, a histogram of visual word occurrences is calculated by assigning each descriptor vector to one visual word of the vocabulary.

(1) *Scale-space and keypoint localization:* the first step is to convolve the image with Gaussian filters at different scales. Given an image I and a Gaussian filter G, the convolved image can be obtained by $L(x,y,\sigma) = I(x,y) \otimes G(x,y,\sigma)$, where $\otimes$ is the convolution operator and σ is the standard deviation related to the scale. In order to detect keypoints in the image, the difference of successive Gaussian-smoothed images separated by a constant k is calculated:

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma). \tag{1}$$

The Difference of Gaussians – DoG is performed in different octaves of the image in a Gaussian pyramid. Keypoints are identified as local maxima and minima in $D(x,y,\sigma)$ by comparing each point to its eight neighbors in the current image and nine neighbors in the scale above and below. If the point is larger or smaller than all of these neighbors, then it is selected as a keypoint. The keypoints are refined by rejecting points with low contrast or poorly located along an edge.

(2) *Orientation assignment:* to achieve invariance to rotation, it is assigned one or more orientations to each keypoint based on the local gradient directions. For the Gaussian-smoothed image $L(x,y,\sigma)$ at the keypoint scale of σ, considering gradient $m(x,y)$ (Eq. (2)) and orientation $\theta(x,y)$ (Eq. (3)), it is calculated a neighborhood around the keypoint location.

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \tag{2}$$

$$\theta(x,y) = \tan^{-1}\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)} \tag{3}$$

An orientation histogram with 36 bins covering 360° is built from the neighborhood gradient orientations. Each point added to the histogram is weighted by its gradient magnitude. The orientations of the keypoint are given by the peaks of this histogram.

(3) *Keypoint descriptor:* the two previous steps assign a location, scale and orientation to the keypoints. This step computes a descriptor vector for each keypoint to describe the local region. First, the image gradient magnitudes and orientations are calculated around the keypoint. To achieve rotation invariance, the gradient orientations and coordinates are rotated with respect to the keypoint orientation calculated in the previous step. Then a 4×4 grid centered at the keypoint is placed over the image (the size of the grid is related to the keypoint scale) and an orientation histogram with 8 directions is calculated for each of the 16 grids. Finally, the keypoint descriptor is a vector $d \in \Re^{128}$ that concatenates the orientation histogram of each of the 16 grids. The keypoint descriptor is normalized to unit length, then a threshold of 0.2 is applied to reduce the influence of large gradient magnitudes, and finally it is renormalized to unit length. Fig. 2(d) shows an example of SIFT applied in a soybean leaf.

3.2. Dense scale-invariant feature transform – DSIFT

SIFT is a sparse local descriptor since it consists of both feature detection and extraction. On the other hand, dense scale-invariant feature transform – DSIFT (Vedaldi and Fulkerson, 2010; Liu et al., 2011) extracts SIFT descriptors in a dense grid of the image (e.g., every 8 pixels). Thus, DSIFT assumes that all pixels or a grid of them are keypoints of the image. To describe the keypoints, DSIFT considers a fixed region around them, such as 16×16 pixels. Fig. 2 (a) and (b) shows an example of SIFT and DSIFT applied to a soybean leaf. SIFT detects and describes keypoints while DSIFT considers each pixel as a keypoint in order to describe them.

The advantage of SIFT compared to DSIFT is that the entire image is used for feature extraction, although the number of descriptors is generally greater than that of SIFT. On the other hand, the steps for detecting keypoints are not necessary according to the DSIFT method.

3.3. Pyramid histograms of visual words – PHOW

Pyramid histograms of visual words – PHOW (Bosch et al., 2007) is a variant that applies DSIFT at multiple scales. While DSIFT extracts descriptors using a fixed region around each keypoint on the grid, PHOW extracts multiple descriptors for a keypoint using increasingly larger square regions as shown in Fig. 2(c). In this figure, a keypoint is characterized by three descriptors, one for each scale. Here, we have used four scales as suggested by authors Bosch et al. (2007). Due to the multiple scales, PHOW describes images at different scales better than DSIFT.

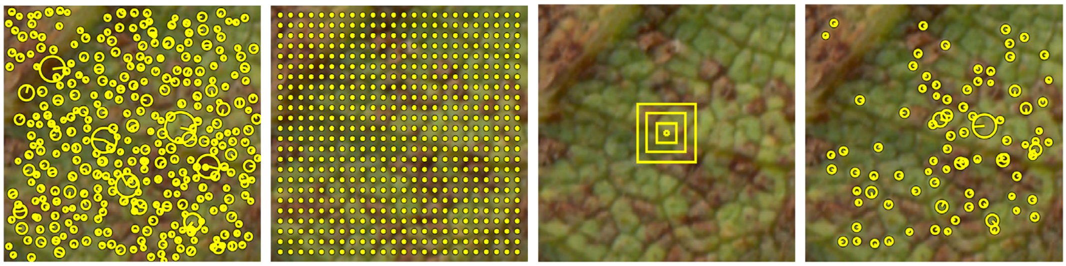
3.4. Speeded-Up robust features – SURF

Speeded-Up Robust Features – SURF (Bay et al., 2008) is a local descriptor partly inspired by SIFT. By using integral images and approximations, SURF can be computed faster than SIFT. The keypoints and descriptors can be obtained using the following steps.

(1) *Keypoint localization:* this step uses a basic Hessian-matrix approximation, which uses integral images to reduce computational time. Given a pixel (x,y) in an image I, the Hessian matrix $\mathcal{H}(x,y,\sigma)$ in (x,y) at scale σ is defined as

$$\mathcal{H}(x,y,\sigma) = \begin{bmatrix} L_{xx}(x,y,\sigma) & L_{xy}(x,y,\sigma) \\ L_{xy}(x,y,\sigma) & L_{yy}(x,y,\sigma) \end{bmatrix} \tag{4}$$

where $L_{xx}(x,y,\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\delta^2}{\delta x^2}g(\sigma)$ over a given image I at pixel (x,y);



(a) SIFT (b) DSIFT (c) PHOW (d) SURF

Fig. 2. Example of SIFT, DSIFT, PHOW and SURF applied to a soybean leaf. SIFT and SURF detect and describe keypoints, while DSIFT and PHOW extract a descriptor at each grid region of the image. Unlike DSIFT, PHOW uses multiple scales, which are related to the region around the keypoints.

similarly for $L_{xy}(x,y,\sigma)$ and $L_{yy}(x,y,\sigma)$. Since the convolutions L_{xx}, L_{xy}, L_{yy} are costly to calculate, SURF discretizes the Gaussian second order derivative $\frac{\delta^2}{\delta x^2}g(\sigma)$ using box filters. The box filters make efficient convolution through integral images (Bay et al., 2008).
The approximated determinant of the Hessian matrix represents the blob response at location (x,y) and scale σ. The greater the blob response, the greater the importance of the pixel in the image. The responses are calculated over several scales using boxes of different sizes, and a non-maximum suppression in a $3 \times 3 \times 3$ neighborhood is applied in order to localize keypoints.

(2) *Orientation assignment:* to be invariant to image rotation, SURF assigns a reproducible orientation for each keypoint. To this end, Haar wavelet responses are calculated in x- and y-directions within a circular neighborhood of radius $6s$, where s is the keypoint scale. Again, these responses can be calculated efficiently with the use of integral images. The wavelet responses are represented as points in a 2-dimensional space with the x-direction response in the abscissa and the y-direction in the ordinate. The horizontal and vertical responses within sliding windows of size $\frac{\pi}{3}$ are summed, forming a local orientation vector for each window. The largest vector of all windows defines the orientation of the keypoint.

(3) *Keypoint descriptor:* the keypoint descriptor is obtained from a square region with size $20s$ centered around the keypoint and oriented along the keypoint orientation. The square region is split into 4×4 sub-regions, and for each sub-region, a vector v with four values is calculated:

$$v = [\sum dx, \sum |dx|, \sum dy, \sum |dy|] \quad (5)$$

where dx and dy are, respectively, the wavelet response in the x- and y-directions computed at 5×5 regularly spaced sample points. Finally, the keypoint descriptor is composed by concatenating each of 16 sub-region vectors v, thereby forming a single vector with 64 values. Fig. 2(d) shows an example of the SURF applied to a leaf image.

3.5. Histogram of oriented gradients – HOG

The histogram of oriented gradients (HOG) (Dalal and Triggs, 2005) is a feature descriptor first proposed for pedestrian detection in images. Basically, it decomposes the image into a dense grid of cells, computes a histogram of oriented gradients in each cell, and normalizes the histogram using the overlapping local contrast of its cells. This descriptor can be described in three steps:

(1) *Gradient computation:* the first step is the computation of gradient values using a 1-D centered derivative mask in both horizontal and vertical directions using masks $[-1,0,1]$ and $[-1,0,1]^T$. The authors evaluated other complex masks (e.g., cubic-corrected, 3×3 Sobel and 2×2 diagonal masks), but they decrease performance in object detection tasks.

(2) *Spatial/orientation binning:* the second step divides the image into rectangular cells and, for each one, calculates an edge-oriented histogram. The bins of the histogram are evenly spaced over 0–180° and each pixel contributes with a weighted vote for the edge-oriented histogram based on its gradient. Dalal and Triggs (2005) found that 9 bins perform the best.

(3) *Normalization and descriptor block:* for better invariance to illumination and contrast, the gradient strengths are locally normalized by grouping cells into blocks. The blocks overlap, which means that a cell belongs to more than one block. Following the original work, we have used rectangular blocks (R-HOG blocks), 2×2 cell blocks of 8×8 pixels cells. To normalize each histogram, let v be a non-normalized vector composed by concatenating all histograms of a given block. Then, the concatenated histogram is normalized by $v' = v/\sqrt{||v||_2^2 + e^2}$, where e is a regularization term. The normalized vector v' is the descriptor calculated at the dense grid of the image.

4. Material and methods

4.1. Experimental design

The plant experiment was done in four fields of the Phytopathology Department of the Federal University of Grande Dourados (UFGD), Brazil. The crop evaluated was soybean [Glycine max (L.) Merr.] of BMX Potencia RR® (BRASMAX).

The density of the soybean fields was of about 300,000 plants ha^{-1}. For all fields, 320 kg ha^{-1} of N-P-K (02-23-23) were applied in-furrow immediately before sowing. No N-fertilizer was applied in any field. The experimental design was a completely randomized block with four replicates. Each plot had 50 rows, spaced by 0.5 m, with 50 m (width) × 25 m (length) (1250 m^2). Plots were separated by at least 10 m, where small terraces of approximately 5 m width were built to prevent contamination by superficial run-off containing bacteria or fertilizer, which could occur due to heavy rains that commonly occur in the summer season. We did not use herbicides in three fields out of four. For one field, herbicides were used in order to have samples with no disease, while insects were controlled with biological and chemical insecticides.

In Dourados (22°22′S and 54°80′W) the fields are at an altitude of 600 m and the soil is classified as Latossolo Vermelho Distrófico (Brazilian classification), or as Typic Haplustox (Soil Taxonomy, USDA). The climate is classified as tropical with dry winter and wet summer.

4.2. Image sampling

Plant leaves were randomly collected in three different stages: V4 – forth trifoliate, V5 – fifth trifoliate and R1 – blooming stage. At the V4 and V5 stages, nine plants were randomly collected per plot for evaluation of leaf diseases, especially those related to fungi. At stage R1, nine plants were collected for evaluation. Sampled material was split in the trifoliates of the growing stage. For this region of Brazil, two classes of diseases are commonly found: mildew and soybean rust. Soybean rust is caused by the fungus Phakopsora pachyrhizi Sydow & Sydow. During collection, we created groups of leaves classified according to two types of color lesions: Rust TAN and Rust RB. TAN lesions are tan in color, and RB refers to reddish-brown lesion color (Bonde et al., 2006). The RB lesion type is considered a resistant lesion type when compared to a fully susceptible TAN lesion (Miles et al., 2007). Furthermore, RB lesions are not sparsely sporulating uredinia.

After sampling the crops, the collected leaves went through a digital image acquisition, as illustrated in Fig. 3. The idea consists three parts: (1) scanning of leaves, (2) texture sample picking, and (3) set of texture samples. Digital images were acquired by scanning each leaf.

In this study, we used a scanner model HP Scanjet 1200, at 1200 dpi resolution, with images generated in tiff format. For each sampling of the plant stage, nine images were collected. From each leaf scanning, a specialist manually spotted 200×200 windows recognized as evidence of disease. We selected 300 windows of each disease (Mildew, Rust RB and Rust Tan), totalizing 900 sub images. Another 300 were collected from the healthy plants. Such sub image windows were saved without file compression. Therefore, the image dataset is composed of 1200 samples divided into four classes, three of them attacked by fungi and one without diseases.

5. Experiments and discussion

In this section, we describe experiments and results obtained with the use of local descriptors and BOVW. In the classification step, we have used the Support Vector Machine – SVM classifier using stratified 10-fold cross-validation. This methodology is well-accepted in machine learning and artificial intelligence. In the stratified 10-fold cross-validation, the images of the dataset are partitioned into 10 folds ensuring that each fold has the same proportion of each class. Then, one fold is used for testing while the remaining folds are used for training the SVM classifier. The process is repeated 10 times using each fold exactly once. Finally, the Correct Classification Rate (CCR) corresponds to the average of the 10-rounds execution.

5.1. Local descriptor comparison

We have compared the local descriptors using different numbers of k visual words, as presented in Fig. 4. As can be seen in the figure, local descriptor PHOW overcame the other local descriptors for all values of k. SIFT and DSIFT also performed well with a CCR nearly of 90%. On the other hand, HOG and SURF showed the worst performance with CCRs around 80% and 70%, respectively.

Table 1 shows the best CCR for each local descriptor and its respective number of visual words k. Comparing dense local descriptors, PHOW, DSIFT, and HOG provided, respectively, 96.25%(±1.53), 90.00%(±1.76), and 80.00%(±3.04) for 3500, 2000, and 1000 visual words. For sparse local descriptors, SIFT and SURF achieved 93.66%(±2.58) and 71.25%(±3.50) using, respectively, $k = 3500$ and 500.

To better visualize the performance of each descriptor, Fig. 5 presents the confusion matrices for the best configuration of visual words. We can observe that, for all the local descriptors, Rust RB and Rust TAN diseases are the most imprecise, while Mildew was classified more accurately. As described previously, PHOW has shown a good characterization of soybean diseases, especially for classes Mildew and Healthy. Despite the superiority to other descriptors, PHOW wrongly classified 13 images of Rust RB as Rust TAN and 21 images of Rust TAN as Rust RB.

5.2. Color local descriptor

In order to improve the characterization of soybean diseases, this section presents the results obtained by the application of PHOW to color images. We have chosen this local descriptor since it provided the best performance as shown in the previous section. For color images, PHOW is applied to each channel of the color space and the descriptors are concatenated to form a single descriptor with color information. Then, BOVW is applied in the same way as before.

Fig. 6 shows the results of descriptor PHOW applied to color spaces RGB, HSV, and Opponent color spaces (van de Sande et al., 2010) using different values of k. As it can be seen, the three PHOW color variants improved the CCR compared to the PHOW grayscale variant. Comparing only color variants, PHOW-HSV and PHOW-Opponent have shown similar results, and superior results if compared to PHOW-RGB. These results demonstrate that HSV and Opponent color spaces are better suited to represent color, which is important for the analysis of leaves. In addition, it is worth to note that, using only 50 visual words, PHOW-HSV provided 96.42 (±1.62), an excellent result with fewer features.

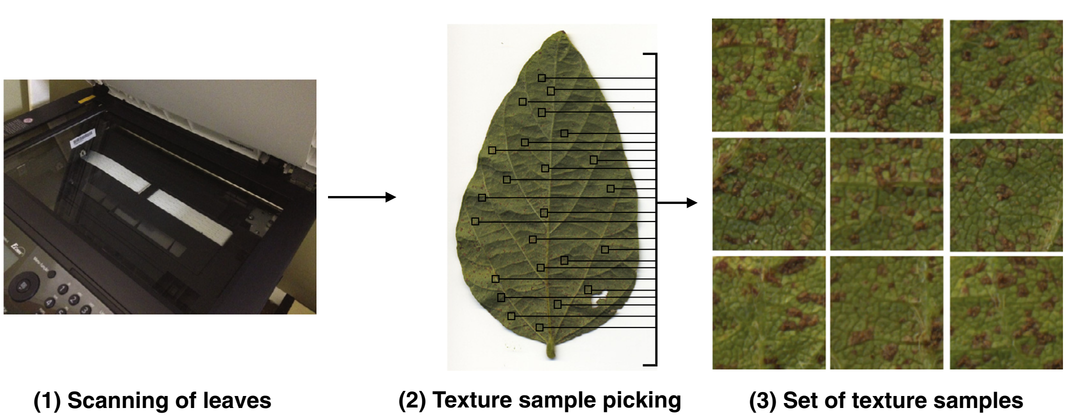


Fig. 3. Image acquisition procedure adopted in this study. Four classes compose our image dataset.

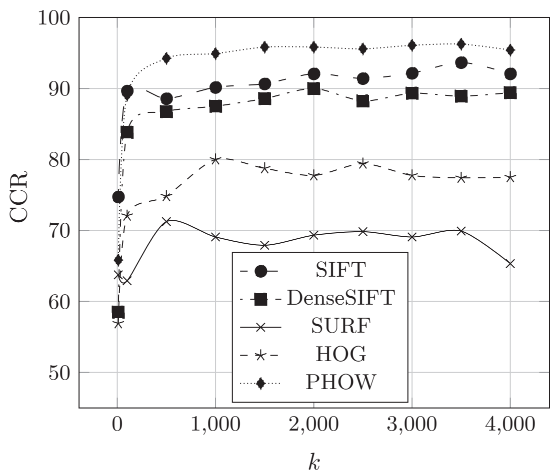


Fig. 4. CCR obtained by the five local descriptors using different values of *k*.

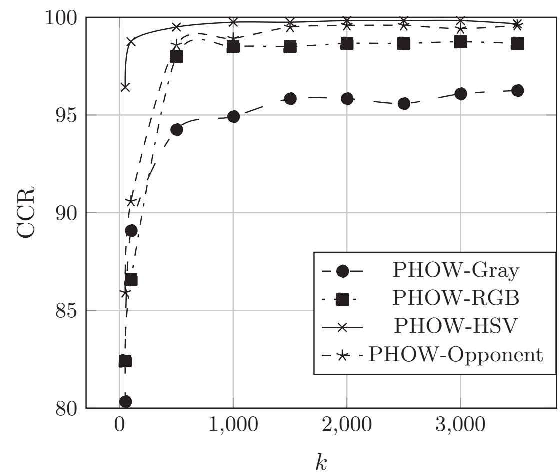


Fig. 6. CCR obtained by the three PHOW color variants using different values of *k*.

Table 1
Best CCRs obtained by each local descriptor and its respective number of visual words.

Local descriptor	*k*	CCR	F-measure
SURF	500	71.25(±3.50)	0.71(±0.03)
HOG	1000	80.00(±3.04)	0.80(±0.12)
DSIFT	2000	90.00(±1.76)	0.90(±0.07)
SIFT	3500	93.66(±2.58)	0.94(±0.05)
PHOW	3500	96.25(±1.53)	0.96(±0.03)

Table 2
Experimental results related to the application of PHOW to each channel of the RGB, HSV, and Opponent color spaces.

Local descriptor	*k*	CCR	F-measure
PHOW-Gray	3500	96.25(±1.53)	0.96(±0.03)
PHOW-RGB	3000	98.75(±0.98)	0.99(±0.01)
PHOW-HSV	2000	99.83(±0.35)	1.00(±0.00)
PHOW-Opponent	2000	99.58(±0.44)	1.00(±0.00)

Table 2 summarizes the best results obtained by each PHOW variant. The best performance of 99.83(±0.35) was obtained by PHOW-HSV followed by PHOW-Opponent, then PHOW-RGB, and PHOW-Gray with 99.58(±0.44), 98.75(±0.98), and 96.25(±1.53), respectively. We can also observe that PHOW-Gray requires more visual words than the color variants to achieve better accuracy. Using 2000 visual words, PHOW-HSV and PHOW-RGB achieved their best result while PHOW-Gray required 3500 visual words. These results corroborate the improvement to soybean disease characterization using color.

The confusion matrices for the PHOW color variants are shown in Fig. 7. As one can see, the three PHOW color variants improved the characterization of soybean diseases. PHOW-HSV and PHOW-Opponent missed only two and five samples of *Rust TAN* and *Rust RB* diseases, respectively. These excellent results are corroborated by the scatter plots of Fig. 8. In these plots, the features extracted using the best configuration of visual words are projected into a two-dimensional space using Principal Component Analysis (Fukunaga, 1990). We can observe that Healthy and Mildew are better represented by all PHOW color variants. Due to the similarity of the samples, Rust RB and Rust TAN are spatially close in the projected space, although they can be effectively discriminated using PHOW color variants.

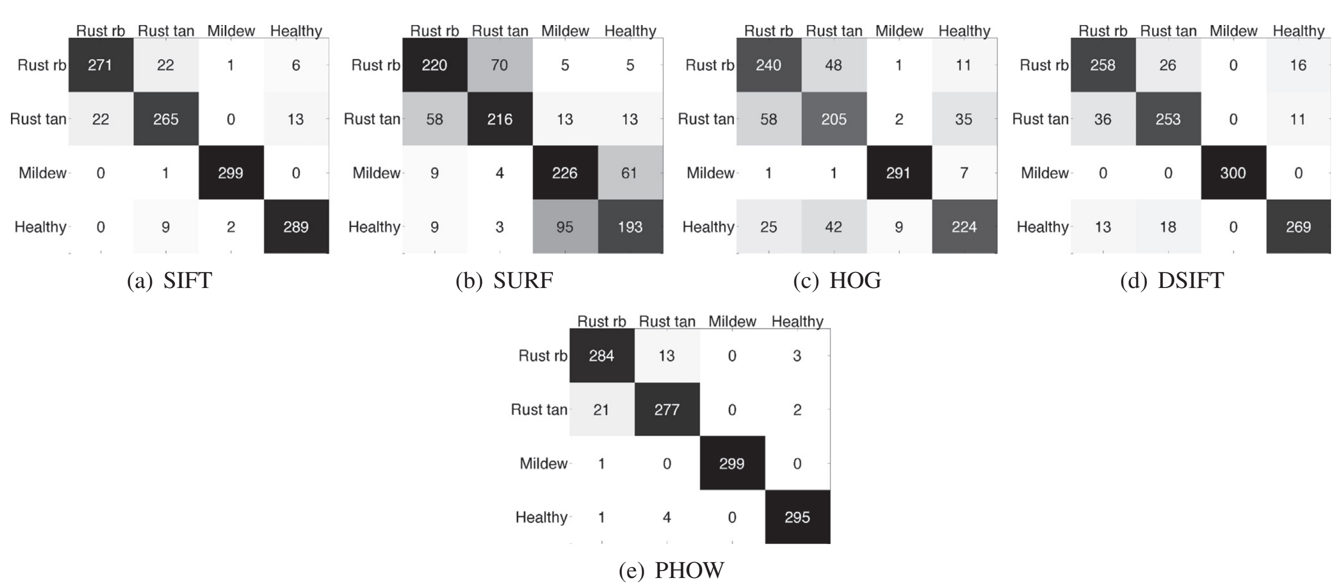
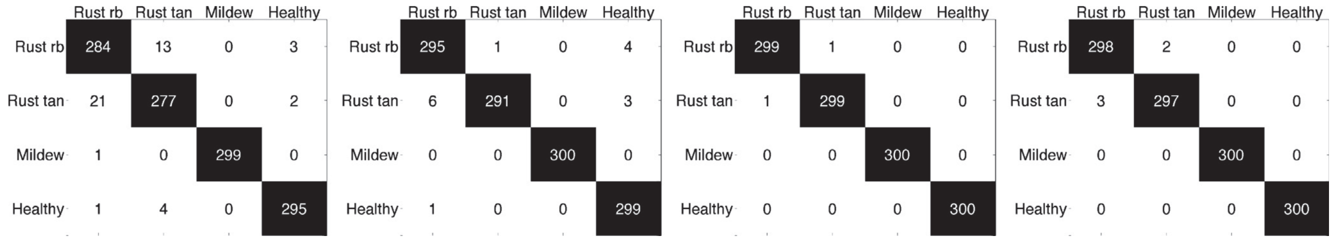
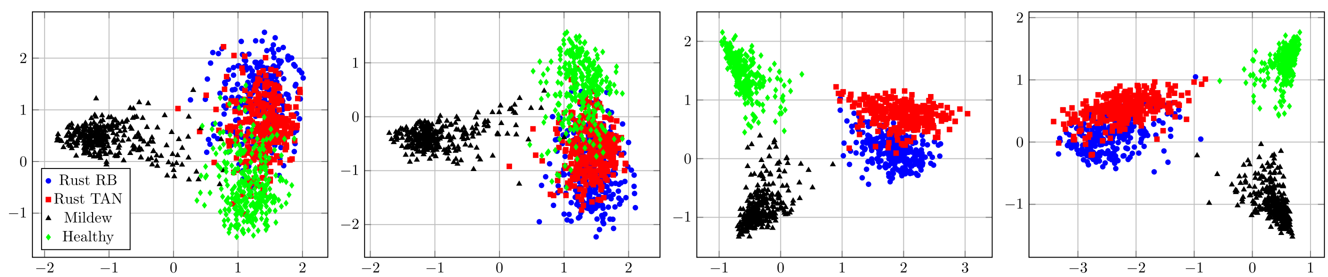


Fig. 5. Confusion matrices of each local descriptors using the best configuration of visual words. The main diagonal corresponds to the number of accurately classified leaves.



(a) PHOW-Gray (b) PHOW-RGB (c) PHOW-HSV (d) PHOW-Opponent

Fig. 7. Confusion matrices of each color variant of descriptor PHOW using the best configuration of visual words.



(a) PHOW-Gray (b) PHOW-RGB (c) PHOW-HSV (d) PHOW-Opponent

Fig. 8. Features of each PHOW color variant projected using Principal Component Analysis into a two-dimensional space.

5.3. Computational cost

In this section, we show the wall-clock time, in seconds, to compute the local descriptors. The experiments are performed on an Intel Core i5 1.3 GHz with 4 GB of RAM running OSX 10.9.5. The time was measured only for the stages of keypoint detection and extraction, excluding the time spent on secondary steps, such as image reading. The values correspond to the mean and standard deviation calculated for 1,200 images of 200×200 pixels.

In our experiments, HOG was the fastest, taking $0.0043(\pm0.001)$ s to process each image. As expected, SURF with an average time of $0.011(\pm0.016)$ s was faster than SIFT with as average of $0.064(\pm0.011)$. Regarding dense local descriptors, DSIFT and PHOW performed with an average time of $0.134(\pm0.001)$ and $0.143(\pm0.006)$ seconds per image, respectively. In general, all local descriptors perform in an acceptable running time.

6. Conclusion

Soybean disease detection is an important process to decrease economic losses in agriculture, and to reduce environmental pollution due to the excessive use of agrochemicals. Answering to this demand, we proposed a new approach for automated detection of soybean diseases. We used image local descriptors and Bag of Visual Words (BOVW) to define a methodology able to computationally represent images of soybean leaves, while maintaining visual information regarding potential diseases. Our experiments compared the use of 8 local descriptor variants using an image dataset composed of 1200 leaves. We considered healthy leaves and leaves infected by diseases Mildew, Rust Tan, and Rust RB.

The experimental results showed that the use of local descriptors, together with the BOVW technique, were effective (over 98% of accuracy) and efficient (an average time of 0.1 s per image) in the task of computationally detecting soybean diseases. These results showed that descriptor PHOW provided the best result, followed by descriptors SIFT and DSIFT. We found out that descriptor PHOW works better for color spaces if compared to the gray scale, reaching correct classification rates ranging from 96.25% (±1.53) to 99.83 (±0.35). Due to the generality of the method, we believe it can be applied to other crops, such as cotton or wheat.

The proposed approach can be employed as a tool to guide users, such as farmers and non-experts, so to identify diseases in soybean crops. In practice, the user can scan leaves with a commodity cell phone directly in the soybean crop and apply the proposed approach. In the future, the images may be automatically obtained with a scanner embedded in agricultural machinery, or in unmanned aerial vehicles.

As part of the future work, we plan to evaluate other steps of the BOVW, such as the vocabulary construction and the feature coding. We also consider investigating diseases in their early stages and measuring their severity.

Acknowledgment

The authors acknowledge support from PET-Fronteira, CAPES, CNPq and FUNDECT.

References

Alexe, B., Deselaers, T., Ferrari, V., 2012. Measuring the objectness of image windows. IEEE Trans. Pattern Anal. Mach. Intell. 34 (11), 2189–2202.

Arbeláez, P., Pont-Tuset, J., Barron, J., Marques, F., Malik, J., 2014. Multiscale combinatorial grouping. In: Computer Vision and Pattern Recognition, pp. 1–8.

Backes, A.R., Gonçalves, W.N., Martinez, A.S., Bruno, O.M., 2010. Texture analysis and classification using deterministic tourist walk. Pattern Recogn. 43 (3), 685–694.

Barbedo, J.G.A., 2013. Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus 2 (1). http://dx.doi.org/10.1186/2193-1801-2-660.

Bay, H., Ess, A., Tuytelaars, T., Gool, L.V., 2008. Speeded-up robust features (surf). Comput. Vis. Image Underst. 110 (3), 346–359.

Bonde, M.R., Nester, S.E., Austin, C.N., Stone, C.L., Hartman, G.L., Frederick, R.D., Miles, M.R., 2006. Evaluation of virulence of *Phakopsora pachyrhizi* and *P. meibomiae* isolates. Plant Dis. 90, 708–716.

Bosch, A., Zisserman, A., Muoz, X., 2007. Image classification using random forests and ferns. In: IEEE 11th International Conference on Computer Vision, pp. 1–8.

Carmona, M., Sautua, F., Perelman, S., Gally, M., Reis, E.M., 2015. Development and validation of a fungicide scoring system for management of late season soybean diseases in argentina. Crop Prot. 70, 83–91.

Carreira, J., Sminchisescu, C., 2012. Cpmc: Automatic object segmentation using constrained parametric min-cuts. IEEE Trans. Pattern Anal. Mach. Intell. 34 (7), 1312–1328.

Casanova, D., Florindo, Joao B., Gonçalves, W.N., Bruno, O.M., 2012. IFSC/USP at imageclef 2012: plant identification task. In: Working Notes for CLEF 2012 Conference, pp. 1–7.

Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C., 2004. Visual categorization with bags of keypoints. In: ECCV International Workshop on Statistical Learning in Computer Vision, pp. 1–22.

Cui, D., Zhang, Q., Li, M., Hartman, G.L., Zhao, Y., 2010. Image processing methods for quantitatively detecting soybean rust from multispectral images. Biosyst. Eng. 107 (3), 186–193.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: Proceedings of the Conference on Computer Vision and Pattern Recognition., CVPR '05. IEEE Computer Society, Washington, DC, USA, pp. 886–893.

Endres, I., Hoiem, D., 2014. Category-independent object proposals with diverse ranking. IEEE Trans. Pattern Anal. Mach. Intell. 36 (2), 222–234.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Academic Press, San Diego.

Gonçalves, W.N., Bruno, O.M., 2013. Combining fractal and deterministic walkers for texture analysis and classification. Pattern Recogn. 46 (11), 2953–2968.

Gui, J., Hao, L., Zhang, Q., Bao, X., 2015. A new method for soybean leaf disease detection based on modified salient regions. Int. J. Multimedia Ubiquitous Eng. 10 (6), 45–52.

Gutiérrez-Aguirre, I., Mehle, N., Delić, D., Gruden, K., Mumford, R., Ravnikar, M., 2009. Real-time quantitative {PCR} based sensitive detection and genotype discrimination of pepino mosaic virus. J. Virol. Methods 162 (1–2), 46–55.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y., 2002. An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. 24 (7), 881–892.

Kumar, V., Rani, A., Dixit, A.K., Pratap, D., Bhatnagar, D., 2010. A comparative assessment of total phenolic content, ferric reducing-anti-oxidative power, free radical-scavenging activity, vitamin c and isoflavones content in soybean with varying seed coat colour. Food Res. Int. 43 (1), 323–328.

Liu, C., Yuen, J., Torralba, A., 2011. Sift flow: dense correspondence across scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell. 33 (5), 978–994.

Liu, Z.-Y., Wu, H.-F., Huang, J.-F., 2010. Application of neural networks to discriminate fungal infection levels in rice panicles using hyperspectral reflectance and principal components analysis. Comput. Electron. Agric. 72 (2), 99–106.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110.

Ma, Y., Huang, M., Yang, B., Zhu, Q., 2014. Automatic threshold method and optimal wavelength selection for insect-damaged vegetable soybean detection using hyperspectral images. Comput. Electron. Agric. 106, 102–110.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27 (10), 1615–1630.

Miles, M.R., Pastor-Corrales, M.A., Hartman, G.L., Frederick, R.D., 2007. Differential response of common bean cultivars to *Phakopsora pachyrhizi*. Plant Dis. 91, 698–704.

Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., Ramon, H., 2004. Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. Comput. Electron. Agric. 44 (3), 173–188.

Oide, M., Ninomiya, S., 2000. Discrimination of soybean leaflet shape by neural networks with image input. Comput. Electron. Agric. 29 (1–2), 59–72.

Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., Plümer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Comput. Electron. Agric. 74 (1), 91–99.

Sankaran, S., Mishra, A., Ehsani, R., Davis, C., 2010. A review of advanced techniques for detecting plant diseases. Comput. Electron. Agric. 72 (1), 1–13.

Saponari, M., Manjunath, K., Yokomi, R.K., 2008. Quantitative detection of citrus tristeza virus in citrus and aphids by real-time reverse transcription-pcr (taqman). J. Virol. Methods 147 (1), 43–53.

Shrivastava, S., Hooda, D.S., 2014. Automatic brown spot and frog eye detection from the image captured in the field. Am. J. Intell. Syst. 4 (4), 131–134.

Uijlings, J., van de Sande, K., Gevers, T., Smeulders, A., 2013. Selective search for object recognition. Int. J. Comput. Vis. 104 (2), 154–171. http://dx.doi.org/10.1007/s11263-013-0620-5.

van de Sande, K., Gevers, T., Snoek, C., 2010. Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. 32 (9), 1582–1596.

Vedaldi, A., Fulkerson, B., 2010. Vlfeat: An open and portable library of computer vision algorithms. In: Proceedings of the International Conference on Multimedia. ACM, pp. 1469–1472.

Yao, H., Huang, Y., Hruska, Z., Thomson, S.J., Reddy, K.N., 2012. Using vegetation index and modified derivative for early detection of soybean plant injury from glyphosate. Comput. Electron. Agric. 89, 145–157.

Yvon, M., Thébaud, G., Alary, R., Labonne, G., 2009. Specific detection and quantification of the phytopathogenic agent 'candidatus phytoplasma prunorum'. Mol. Cell. Probes 23 (5), 227–234.