Image Classification of Paddy Field Insect Pests Using Gradient-Based Features

Kanesh Venugoban and Amirthalingam Ramanan

***Abstract*—Agriculture is one of the principal economic activities of the Jaffna peninsula in the Northern Province of Sri Lanka. Over 60% of the work force in the peninsula depends on agriculture for their livelihood. Paddy cultivation in the peninsula contributes substantially to the gross national income of the country. Such Paddy crops are affected by the attack of insect pests. Therefore paddy field insect pest identification is an important task to the sustainable agricultural development in the Jaffna peninsula. This paper offers a framework to classify images of paddy field insect pests using gradient-based features through the bag-of-words approach. Images of twenty classes of paddy field insect pests were obtained from Google Images and photographs taken by the Faculty of Agriculture, University of Jaffna, Sri Lanka. The images were then classified through the system that involves identification of regions of interest and representation of those regions as scale-invariant feature transform (SIFT) or speeded-up robust features (SURF) descriptors, construction of codebooks which provides a way to map the descriptors into a fixed-length vector in histogram space, and the multi-class classification of the feature histograms using support vector machines (SVMs). Furthermore, the histograms of oriented gradient (HOG) descriptors were applied in classification. As a baseline classifier the nearest neighbour approach was used and compared with SVM-based classifiers. Testing results show that HOG descriptors significantly outperform existing local-invariant features: SIFT and SURF in paddy field insect pests classification. HOG descriptors when combined with SURF features yield around 90% accuracy in classification. For simplicity and speed, linear SVM was used as a classifier throughout the study.**

***Index Terms*—Bag-of-words, paddy field insect pests, hog, sift, surf.**

I. Introduction

Rice is one of the main crops in Jaffna, but great economic loss does occur for paddy farmers because of plant diseases and insect pests every year. Jaffna is located in the Northern tip of Sri Lanka at a longitude of $79^0 45' – 80^0 20'$ and latitude of $9^0 30' – 9^0 50'$. The population of Jaffna peninsula is around 0.7 million. Agriculture and fisheries sectors play a crucial role to the gross production of Jaffna. Therefore, the development of an automated system for paddy field insect pest identification is of great significance. Computer vision techniques have great significance on the automatic identification of the images of insect pests. Those techniques not only can decrease the labour, but also can improve the speed and precision of the identification and diagnosis, when compared to manual method. In this regard, recognition of paddy field insect pests is challenging because the insect pests are highly articulated, they exhibit a high degree of intra-pest variation in size and colour, and some insect pests are difficult to distinguish visually, despite prominent dorsal patterning. The manual classification of such insect pests in paddy fields can be time consuming and requires substantial technical expertise. The task becomes more challenging when insect pests are to be recognised from still images using an automated system. Images of one insect pest may be taken from different viewpoints, cluttered background, or may suffer transformation such as rotation, noise, etc. So it is likely that two images of the same insect pest will be different. To address these challenges, we have adopted the gradient-based features in classifying images of paddy field insect pests. The primary advantage of this approach is that it is invariant to changes in pose and scale as long as the features can be reliably detected. Furthermore, with an appropriate choice of classifier, not all features need to be detected in order to achieve high classification accuracy. Hence, even if some features are occluded or fail to be detected, the method can still succeed.

In this paper, twenty species of paddy field insect pests are considered that are prominently found in Jaffna paddy fields. We compared our system with SIFT [1] and SURF [2] descriptors using the bag-of-words approach [3] and the HOG [4] features. Our testing results show that HOG features perform better than SIFT and SURF. Moreover, the performance of the system was checked when features are concatenated. The concatenation of HOG and SURF yields 8% of improved performance. The system achieved around 90% average accurate rate in classifying test images of those twenty species of paddy field insect pests.

The rest of this paper is organised as follows. In Section II, we summarise different techniques that are closely related to insect classification and recently published. Section III summarises gradient-based features: SIFT, SURF and HOG. Section IV provides an overview to the bag-of-words approach. In Section V, the components of this approach (vocabulary construction and classification) are described in detail. This section also describes the empirical evaluation with the experimental setup, a brief description of the dataset, and the testing results. Finally, Section VI concludes the paper with a discussion of the findings towards future extensions.

II. Previous Work

In [5], the authors have used a bag-of-features approach to

Manuscript received July 9, 2013; revised September 10, 2013.

K. Venugoban is with the Faculty of Agriculture, University of Jaffna, Jaffna, Sri Lanka (e-mail: venugobank@ jfn.ac.lk).

A. Ramanan is with the Department of Computer Science, Faculty of Science, University of Jaffna, Sri Lanka (e-mail: a.ramanan@jfn.ac.lk).

automate rapid-throughput taxonomic identification of stonefly larvae. 263 stonefly larvae were collected of four stonefly taxa from freshwater streams in the mid-Willamette valley and Cascade Range of Oregon. Approximately ten photos were obtained of each specimen, which yields 20 individual images. These were then manually examined, and all images that gave a dorsal view within 30 degrees of vertical were selected for analysis. The images were then classified through a process that involves: Identification of regions of interest, representation of those regions as SIFT vectors, classification of the SIFT vectors into a histogram of detected features, and classification of the histogram by an ensemble of logistic model trees. In their work, they have applied three region detectors: Hessian-affine detector and the Kadir entropy detector, including a newly developed principal curvature-based region (PCBR) detector. The construction of a codebook was performed by a Gaussian mixture model (GMM). The authors claim that their PCBR detector outperforms the other two detectors while showing a classification accuracy of 82% for four classes and 95% for three classes.

In [6], the authors have proposed a system to detect whiteflies, aphids and thrips on the infected crops in greenhouse. Images of the infected leaf are captured by a camera and pre-processed using image processing techniques such as converting images from RGB to gray scales and filtering in order to obtain an enhanced image set of pests. In feature extraction, some properties of the image are considered. A variety of region properties and gray covariance matrix properties such as entropy, mean, standard deviation, contrast, energy, correlation and eccentricity are extracted from those images. The classification was performed by the use of support vector machines. The authors claim that the prototype system proved rapid detection of pests and exhibits the same performance level as a classical manual approach.

In [7], the authors have proposed a system for tea insect pests classification using correlation-based feature selection (CFS) and incremental back propagation learning network (IBPLN). The authors have created a database concentrating on eight major insect pests from the records of different tea gardens of North-Bengal districts of India. The database consists of 609 instances belonging to eight classes described by 11 attributes (signs and symptoms); all of which are nominal. The classification was performed using artificial neural networks. The classification results were compared with the original feature set and reduced feature set. Their study demonstrates that CFS can be used for reducing the feature vector and CFS+IBPLN combination can be used for other classification problems.

We demonstrate good performance on ‘real world’ images of paddy field insect pests. Though some artificial systems for identifying insect pests exist, to the best of our knowledge, there is no system in the literature for classifying paddy field insect pests.

III. Gradient-Based Features

The introduction of powerful patch-based Scale-Invariant Feature Transform (SIFT) descriptors [1] had a significant impact on the popularity of local features. Interest points combined with local descriptors were started to be used as a black box providing reliable and repeatable measurements from images for a wide range of applications. The assumption is, in different image classes, the statistical distribution of the patches was different. More recently, Speeded-up Robust Features (SURF) [2] are also becoming popular due to their faster performance with less number of interest points and dimension when compared to SIFT. Furthermore, Histogram of Oriented Gradients (HOG) has also become one of the most popular low-level image representations mainly on the problem of pedestrian detection in static images [4]. The difference is that, SIFT or SURF describes the features at the candidate location (i.e., keypoint), while HOG describes the feature over the given region. This means that HOG can roughly represent the shape of an interest object.

A. *Scale-Invariant Feature Transform*

SIFT is a method to extract distinctive features from gray-level images, by filtering images at multiple scales and patches of interest that have sharp changes in local image intensities. The SIFT algorithm consists of four major stages: Scale-space extrema detection, keypoint localisation, orientation assignment, and representation of a keypoint descriptor. The features are located at maxima and minima of a difference of Gaussian (DoG) functions applied in scale space. Next, the descriptors are computed as a set of orientation histograms on 4×4 pixel neighbourhoods, and each histogram contains 8 bins. This leads to a SIFT feature vector with $4\times4\times8$ (i.e., 128) dimensions on each patch.

B. *Speeded-up Robust Features*

SURF is partly inspired by SIFT that makes use of integral images. The scale space is analysed by up-scaling the integral image-based filter sizes in combination with a fast Hessian matrix-based approach. SURF features can be extracted faster than SIFT using the gain of integral images and yields a lower dimensional feature descriptor (i.e., 64 dimensions) resulting in faster matching and less storage space. The detection of interest points is selected by relying on the determinant of the Hessian matrix where the determinant is maximum. Next, the descriptors are computed based on orientation using 2D Haar wavelet responses calculated in a 4×4 sub region around each interest point, resulting in a 32 dimensional vector. When information about the polarity of the intensity changes is considered, this in turn results in a 64 dimensional vector. The extended version of SURF has the same dimension as SIFT.

C. *Histogram of Oriented Gradients*

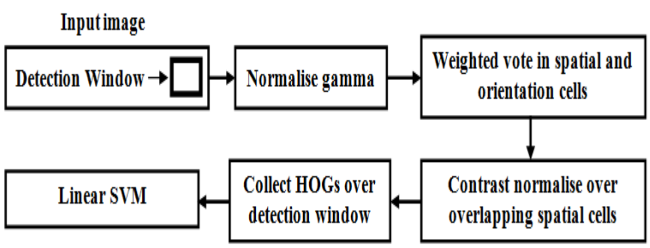

Fig. 1. An overview of HOG feature extraction. The detection window is scanned across the image at all positions and scales.

HOG is similar to that of edge orientation histograms, scale-invariant feature transform descriptors, and shape contexts, but differs in the way as it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalisation for improved accuracy. Fig. 1 summarises the HOG feature extraction process.

IV. Bag-of-Words Approach

The bag-of-words approach was originally used in text mining and is now widely used in visual object recognition [3], [8], [9], robot navigation, visual data mining and classification problems in cell biology. In the bag-of-words approach, invariant-features are first extracted from local regions on images and a visual codebook is constructed by applying a clustering algorithm on a subset of the features where the cluster centres are considered as “visual words” or ‘codewords’ in the codebook. Each feature in an image is then quantised to the closest word in the codebook, and an entire image is represented as a global histogram counting the number of occurrences of each word in the codebook. The size of the resulting histogram is equal to the number of words in the codebook and hence the number of clusters obtained from the clustering algorithm. The codebook is usually constructed by applying the traditional *K*-means clustering algorithm or other hierarchical algorithms. This approach is shown to be robust to distortions in images. One potential drawback of this approach is that the inter-patch relationships and global image structure are ignored. This, however, can be partially compensated by sampling dense and redundant features from the images. The basic idea behind the bag-of-words approach is illustrated in Fig. 2.

V. Experimental Setup

We present our experimental results under three categories. In the first category, the performances of the bag-of-words approach using SIFT and SURF descriptors were compared in classifying the paddy field insect pests. In the second category, the performance of HOG features was tested in classifying the insect pests. Finally, the performance of feature concatenation of SIFT and SURF with HOG was compared.

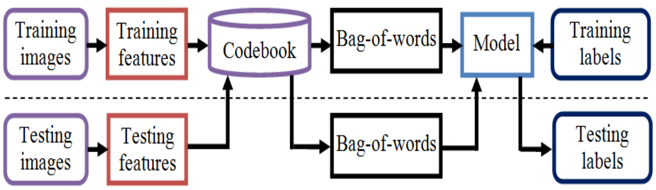

Fig. 2. Bag-of-words approach.

Initially a visual codebook is constructed by clustering a subset of the local features extracted from training images. The centre of each cluster is referred to as codeword in the learnt codebook. Each local feature in a test image is then mapped to the closest codeword and each test image is represented as a histogram of visual words.

A. Dataset

We obtained images of twenty species of paddy field insect pests with 10 images per species from Google Images and photographs taken by the Department of Agricultural Biology, University of Jaffna, Sri Lanka. Fig. 3 shows some example images of the selected twenty species of insect pests that are mostly found in Jaffna paddy fields. This image set has significant viewpoint changes, different backgrounds, arbitrary rotations, and scale differences within each class. Fig. 4 shows some of the intra-class variations that are found in the image set.

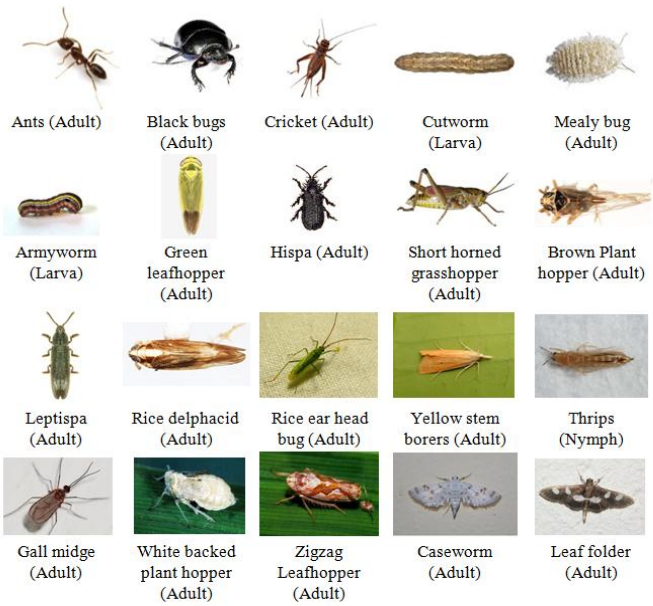

Fig. 3. One image from each of the twenty species of paddy field insect pests.

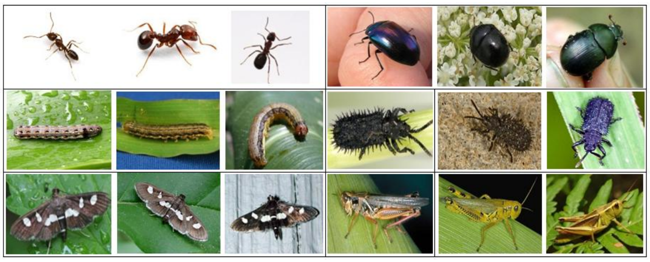

Fig. 4. Example images of paddy field insect pests having intra-class variations.

B. Feature Extraction

In our first approach, SIFT or SURF features were extracted from local patches on each image of the same class. These features were then quantised into visual codebook in which the descriptors of the same class were represented as a bag-of-words. Thus, our approach could take advantage of the class membership information of images. In our second approach, we maintained all the HOG parameters to be constant across different classes of insect pests. The 128×128 pixel detection window was divided into 15 blocks horizontally and 15 blocks vertically, for a total of 225 blocks. Each block contains 4 cells with an 8-bin histogram for each cell, for a total of 32 values per block. This brings the final vector size to 15 blocks across × 15 blocks vertically × 4 cells per block × 8-bins per histogram = 7200 values.

C. Codebook Construction

The SIFT or SURF features computed from regular patches on the codebook images are then clustered using the traditional K-means algorithm. Since this algorithm depends on the initial centres, we repeated the algorithm with ten random initialisations from which the one resulting in the smallest summed within-cluster distance was selected. We also studied the effect of cluster size (i.e., the size of the codebook) on the classification rate by setting the number K of K-means to 25, 50, 75, 100, 150, 200 and 250. We constructed separate codebooks of learnt features for each insect pest category and concatenated these codebooks into a global codebook of size $20\times K$. Testing results show that the global codebook of size 1000 (i.e., K=50) gave us better performance in classification and was fixed throughout the entire study.

D. Feature Representation

After the codebooks are constructed, the images in each group are quantised against the locally merged global codebook. Features computed on regular patches on images were compared with the visual words in the global codebook, and the word closest to the feature in terms of Euclidean distance was used to represent it. Then the entire image group was represented as multiple bags of words. Since the order of the words in the bag was irrelevant as long as it was fixed, the bag could be represented as a vector counting the number of occurrences of each word in the image group. Feature vectors of HOG+SURF each of size 8200 were fed into a multi-class SVM classifier with linear kernel function.

E. Multi-Class Classification

SVM is a supervised learning technique based on a statistical learning theory that can be used for pattern classification [10]. In general SVMs outperform other classifiers in their generalisation performance. A linear SVM finds the hyperplane leaving the largest possible fraction of points of the same class on the same side, while maximising the distance of either class from the hyperplane. SVMs were originally developed for solving binary classification problems and then binary SVMs have also been extended to solve the problem of multi-class pattern classification.

In multi-class classification each training point belongs to exactly one of the different classes. The goal is to construct a function which, given a new data point, will correctly predict the class to which the new data point belongs. There are four standard techniques frequently employed by SVMs to tackle multiclass problems, namely One-Versus-One (OVO) [11], One-Versus-All (OVA) [12], Directed Acyclic Graph (DAG) [13], and Unbalanced Decision Tree (UDT) [14]. We used the one-versus-all (OVA) linear SVMs. The implementation of multi-class classifiers was performed using the SVM^{light} package [15]. The regularisation parameter C of linear SVM was tuned with a range of values [$2^{-2}, 2^{-1}, ..., 2^{11}, 2^{12}$] by means of cross-validation on the training set.

F. Testing Results

We reported the classification rate as follows:

$$Rate=\frac{Number\ of\ correctly\ classified\ images}{Total\ number\ of\ testing\ images}\times 100\%$$

In Table I, we reported the mean classification accuracy, together with the standard deviation, over the two-fold cross-validation.

TABLE I: PERCENTAGE OF IMAGES CORRECTLY CLASSIFIED BY THE SYSTEM AS MEAN AVERAGE WITH STANDARD DEVIATION. THREE GRADIENT-BASED FEATURES AND THEIR CONCATENATION ARE COMPARED WITH TWO DIFFERENT CLASSIFIERS. THE CONCATENATED FEATURE VECTOR IS EACH OF SIZE 8200 VALUES

Feature \ **Classifier**	**Nearest-Neighbour**	**SVM**
SIFT	46.5 ± 0.6074	63.5 ± 0.1213
SURF	53.0 ± 1.4142	72.5 ± 0.7721
HOG	73.5 ± 0.7071	81.0 ± 0.4142
HOG+SIFT	75.0 ± 0.4142	84.0 ± 0.0000
HOG+SURF	76.5 ± 0.7071	**89.5** ± 0.7071

Even though overlapping blocks in extracting HOG features improve performance, the size of descriptors increases. We have also tested the 256×256 pixel detection window which was divided into 31 blocks across and 31 blocks vertically, for a total of 961 blocks. Each block contained 4 cells with an 8-bin histogram for each cell, for a total of 32 values per block. This brought the final vector size to 30752 values. The higher dimension feature vector (i.e. 31752) yielded only a performance increase of 1% in the classification (see Table II).

TABLE II: MEAN RATE OF CLASSIFICATION USING A 30752 HOG FEATURE VECTOR AND ITS CONCATENATION WITH SIFT OR SURF HISTOGRAMS OF SIZE 1000 WHEN COMPARED WITH NEAREST NEIGHBOUR AND SVM CLASSIFIERS

Feature \ **Classifier**	**Nearest-Neighbour**	**SVM**
HOG	77.0 ± 0.4142	86.5 ± 0.7071
HOG+SIFT	80.5 ± 0.7071	89.0 ± 0.4142
HOG+SURF	81.0 ± 0.0000	**90.5** ± 0.7071

All of our experiments were implemented in MATLAB and executed on a desktop computer with an Intel Core 2 running at 2.4GHz and 4GB of RAM.

VI. DISCUSSION AND CONCLUSION

In this paper we have illustrated a framework to classify twenty species of paddy field insect pests that are prominently found in Jaffna paddy fields using gradient-based features. SIFT and SURF descriptors are invariant to common image transformations, such as scale changes, image rotation, and small changes in illumination. These descriptors are also invariant to translations as from the use of local features. SURF features can be extracted faster than SIFT using the gain of integral images and yield a lower dimensional feature descriptor resulting in faster matching and less storage space. One of the important factors affecting the performance of local feature methods is the image resolution, since keypoint extraction tends not to work well on low-resolution images.

The HOG+SURF representation has several advantages. It captures edge or gradient structure that is very characteristic of local shape and it does so in a local representation with an easily controllable degree of invariance to local geometric and photometric transformations: translations or rotations make little difference if they are much smaller that the local spatial or orientation bin size. HOG features vector is usually high dimension as the gradients are computed over the entire image. The dimension of HOG features can be reduced by carrying out PCA technique. Overall the method used in this paper appear promising, both based on results and on the simplicity of its implementation.

This framework can be further integrated into mobile phones with a slight modification. This can further help farmers when an unknown image of a paddy field insect pest is captured via a phone, the model can predict the species and help farmers for further action. In future research, this system could be extended to automate annotation of paddy field insect pest's life-cycle from pattern images.

ACKNOWLEDGMENT

The authors would like to thank the Department of Agricultural Biology, University of Jaffna for providing information and images of paddy field insect pests.

REFERENCES

[1] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.
[2] A. Bay, Ess T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," *Computer Vision and Image Understanding*, vol. 110, pp. 346–359, 2008.
[3] C. Csurka, R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Workshop on Statistical Learning in Computer Vision*, ECCV, pp. 2004, 1–22.
[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc.* the *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 886–893.
[5] N. Larios, H. Deng, W. Zhang, M. Sarpola, J. Yuen, R. Paasch, A. Moldenke, D. A. Lytle, S. R. Correa, E. N. Mortensen, L. G. Shapiro, and T. G. Dietterich, "Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects," *Machine Vision and Applications*, Springer, vol. 19, issue. 2, pp. 105-123, 2008.
[6] R. G. Mundada and V. V. Gohokar, "Detection and classification of pests in greenhouse using image processing," *IOSR Journal of Electronics and Communication Engineering*, vol. 5, pp. 57-63, 2013.
[7] R. K. Samanta and I. Ghosh, "Tea insect pests classification based on artificial neural networks," *International Journal of Computer Engineering Science (IJCES)*, vol. 2, issue. 6, 2012.
[8] M. Everingham, L. Van-Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge," *International Journal of Computer Vision*, pp. 303–338, 2010.
[9] A. Ramanan and M. Niranjan, "A review of codebook models in patch-based visual object recognition," *Journal of Signal Processing Systems*, Springer New York, vol. 68, no. 3, pp. 333-352, 2011.
[10] N. Cristianini and J. S. Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.
[11] R. Debnath, N. Takahide, and H. Takahashi, "A decision based one-against-one method for multiclass support vector machine," *Pattern Analysis Application*, vol. 7, pp. 164–175, 2004.
[12] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *Journal of Machine Learning Research*, vol. 5, pp. 101–141, 2004.
[13] J. C. Platt, N. Cristianini, and J. S. Taylor, "Large margin dags for multiclass classification," *Advances in Neural Information Processing Systems (NIPS'00)*, vol. 12, pp. 547–553, 2000.
[14] A. Ramanan, S. Suppharangsan, and M. Niranjan, "Unbalanced decision trees for multi-class classification," in *Proc. the IEEE International Conference on Industrial and Information Systems (ICIIS'07)*, 2007, pp. 291–294.
[15] T. Joachims, B. Schlkopf, C. Burges, and A. Smola, "Making large-scale svm learning practical," *Advances in Kernel Methods-Support Vector Learning*, The MIT Press, Cambridge, MA, USA, 1999.

Kanesh Venugoban is an instructor in computer technology at the Faculty of Agriculture, University of Jaffna, Sri Lanka. He received his B.Sc. Honours in 2005 from the University of Jaffna, and received his M.Sc. in computer science in 2013 from the University of Peradeniya, Sri Lanka. His research interest is in image processing.

Amirthalingam Ramanan is a senior lecturer at the Department of Computer Science, University of Jaffna, Sri Lanka. He received his B.Sc. Honours in computer science from the University of Jaffna, Sri Lanka in 2002 and Ph.D. in computer science from the University of Southampton, England in 2010. His research interests are in the algorithmic and applied aspects of Machine Learning and Computer Vision.