

Automated Identification of Rice and Corn Pests(Order Hemiptera) through GLCM using K-Nearest Neighbors Algorithm

John Viscel M. Sangkal and Concepcion L. Khan

I. INTRODUCTION

A. Background of the Study

Philippines is an agricultural country. Environmental conditions such as soil and weather conditions make the Philippines suitable for planting crops as a source of living. Rice and Corn are two of the most important and primary source of food in the country, however these crops may lose its quantity and quality when attacked by different pests. Farmers lose an estimated average of 37% of their rice crop to pests and diseases every year. Therefore it is top priority to find an effective method to control the damage of the pests. In addition to good crop management, timely and accurate diagnosis can significantly reduce losses. [1]

B. Statement of the Problem

One vital issue with field crop is identification of pests. When symptoms of any pests take place on crops in fields, proper action should be taken to control it in order to minimize loss to farmers. Farmers use their experience or knowledge and reference guide books for the correct action to take place. If they do not have proper knowledge, then misidentification of any pests can take place and incorrect control measures like non-affecting pesticides may be used leading to waste of effort, time, and money and it may also lead to serious problems to crops

Farmers may also approach any agricultural experts to advise them regarding detection and identification of pests affecting their crops and how to manage them. However, Sometimes, they have to go long distances to consult an expert and may find themselves wasting their time and effort because the expert may not be available at the time. Also, the expert the farmer consults may not be in a position to advise.

In order to avoid wasting resources and to produce maximum yield of crops, an online library and system for identifying possible pests through analysis of pest images needs to be designed for farmers to easily identify pests. [2]

C. Significance of the Study

Enormous agricultural yield is lost every year, due to rapid infestation by pests and insects thus plant pest identification

is important for food security, quality of life and a stable agricultural economy. If farmers properly identify these pests then they can take preventive as well as control measures such as spraying pesticides to increase production. [3] By developing an online library and application that extracts observable features of a pest from an image through digital image processing techniques and classify the species of the pest, farmers who have few experience and knowledge about pest will have an initial diagnosis of the pest that infected the crops.

D. Objectives

Generally, this study aims to devise an online library system of pest and diseases for crops integrated with a system that will identify the possible species of the pest under the order Hemiptera in a pest image through digital image processing and provide all relevant information about the possible species of the pests. Specifically, it aims to:

- Develop a pests and diseases library for rice and corn
- Extract texture features such as energy, entropy, homogeneity, and correlation using GLCM
- Identify the possible pest species through analysis of extracted features
- Provide relevant information about the possible species of the pest such as destructive stage, plant part and growth stage damaged, nature of damage, management practices and environmental factors
- Find the percentage of the accuracy of the system

E. Scope and Limitations

The library of pests only consists of major pests and diseases for rice and corn.

This study only identifies pest under the order Hemiptera(True Bugs). The features of the pest that will be used to identify the species of the pest are texture features such as energy, entropy, homogeneity, and correlation which will be extracted using GLCM matrix.

The users of the system is assumed to capture the pest image using smart phone camera or digital camera.

II. REVIEW OF RELATED LITERATURE

A. Pest Order Hemiptera

Order Hemiptera are characterized by 2 pairs of wings, although some species may be wingless and others have

Presented to the Faculty of the Institute of Computer Science, University of the Philippines Los Baños in partial fulfillment of the requirements for the Degree of Bachelor of Science in Computer Science

only forewings. Wings are generally membranous but in some species the forewings may be hardened at the base. They also have piercing or sucking mouthparts appearing as a sharply pointed tube known as a proboscis or rostrum, which extends from the underside of the head. Their eyes may be compound of various forms. Hemipterans' antennae vary and may be either short, or long and conspicuous. The young of hemipterans look like small adults. Some bugs may be mistaken for beetles but can be distinguished by their mouthparts as beetles have mandibulate mouthparts while bugs have sucking/piercing mouthparts.

This order is divided into 3 suborders; true bugs (Heteroptera); hoppers (Auchenorrhyncha) and; aphids, scale insects, lerps and mealy bugs (Sternorrhyncha).

Figure 1 shows the general life cycle of Hemipterans. Hemipterans develop by incomplete metamorphosis and young (or nymphs) generally resemble small, wingless adults in general structure although their colour and markings may be very different. In some species such as cicadas, the nymphs who are specialised for burrowing appear quite different from the adults. Nymphs will moult 6 to 8 times depending on the species before they attain maturity.

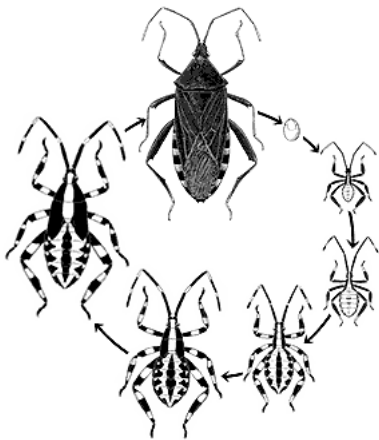


Fig. 1. Life Cycle of Order Hemiptera

Most species of Hemiptera are plant feeders, sucking sap with many causing considerable damage to crops, ornamental garden plants such as roses, shrubs and trees. Some species are bloodsuckers of mammals and birds while others are predators that feed on other invertebrates, including some pest species and are therefore beneficial to man.

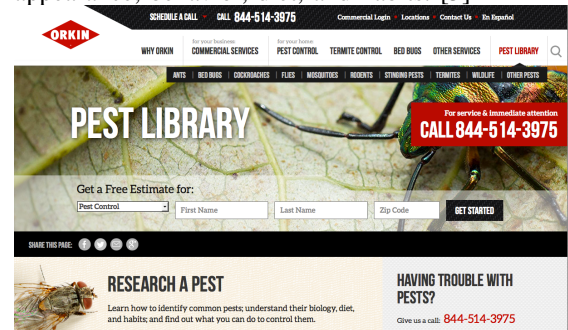
The proboscis of hemipterans contains cutting blades and a two-channelled tube. Hemipterans feed by cutting into a plant or animal and sending saliva down one of the tubes to begin digestion. The liquid food is then sucked up the other tube. [4]

B. Some Online Pest Library and Their Features

1) Orkin.com

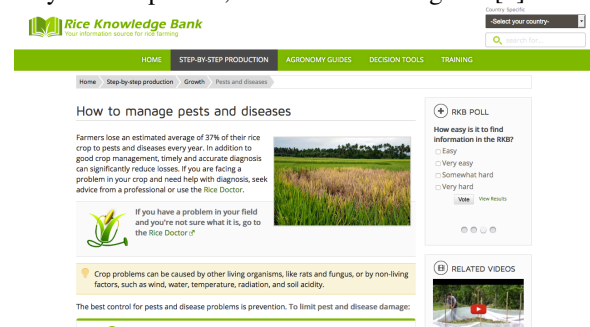
-Orkin.com is a commercial website which provides a library of pests not only found in field crops but also in the house and in the wild. It implements a system

for the user to schedule or contact with a specialist. It also has a search engine for pests and displays vital informations about them such as common name of pest, scientific name, facts, identification, control, appearance, behavior, diet, and habits. [5]



2) IRRI Rice Knowledge Bank

-IRRI's Rice Knowledge Bank provides a detailed step-by-step control and prevention of pest to limit the damage as well as a library of pests that provides information like common name of pest, scientific name, what it does, why and where it occurs, how to identify, why is it important, and how to manage it. [1]



C. Image Processing for Pest Species Identification

There have been many recent studies around the world about pest identification and detection through image processing. In 2014, [6] developed a decision support system that uses image segmentation to extract the pest from the image then extract the features of the pest such as color and shape. The features will be saved in a databased with the name of the pest. A Support Vector Machine classifier would be finding out the type of pest then the system will give preventive and control measures to the user.

In 2013, [7] used the same method as Kandalkar and Deorankar but the features they extracted are entropy, mean, standard deviation, contrast, energy, correlation, and eccentricity.

In 2010, [8] developed a novel method to classify insects by analyzing color histogram and GLCM (Gray-Level Co-occurrence Matrices) of wing images is proposed. The image is preprocessed to get the ROI(Region of Interest) then the color space is converted to HSI (Hue-Saturation-Intensity). Color histograms of the ROI are generated from hue and saturation distributions. Afterward, the color image is converted

to grayscale and their GLCM(Gray Level Co-occurrence Matrices) features are extracted. K-Nearest Neighbor algorithm is used for classification. A lepidopteran insect database with 100 species is used to test the method. With a recognition rate of 71.1% and an ideal time performance achieved, the experimental results testify the efficiency of proposed method.

[9] made an automatic identification system of agricultural pest based on machine vision. They extracted 12 features including color, area, perimeter, fractal dimension, and ridgelet edge moment. They implemented feature selection to select 7 features out of the 12. Afterwards, Back Propagation neural network classifier was used to identify the pests. The pest recognition rate for 8 species was 85.7% which shows that the system is efficient.

In 2011, [10] used template matching to recognize pecan weevils. Five recognition methods were implemented: Normalize cross-correlation, Fourier Descriptors, Zernike moments, String matching, and Regional properties. They collected 205 pecan weevils for the training set and randomly selected 30 pecan weevils and 74 other insects for the testing set. The results showed that Region-based methods were better in representing and recognizing biological objects such as pests. The optimum result among the tested order of Zernike moments was found to be at the order 3. When the Zernike moments and Region properties recognition methods were used in combination, it achieved a recognition rate of 100% and a processing time of 0.44 sec which proves the efficiency of the method.

[11] conducted a study for automatic identification of whiteflies, aphids, and thrips. The size and color components of the object were selected as features for identification because aphids have small variation in color information and the body size is substantially different from other species. The identification error for whiteflies and thrips was reduced when the data for aphids was analyzed first.

In 2013, [12] used k-means clustering algorithm and correspondence filters to automatically detect and recognize pest from an image. The detected the data set by partitioning the data space into Voronoi cells in order to find clusters of comparable spatial extents. With this, they separated the pest from the background. By extracting variant distinctive attributes between the pest and its habitat and using correspondence filters, they identified the plant pest and obtained correlation peak values for different datasets. The study showed that the recognition probability from the pest image is directly proportional to the height of the output signal and inversely proportional to the viewing angles. The correspondence filter can achieve rotational invariance of pests up to angles of 360 degrees which proves the effectiveness of the algorithm.

In 2014, [3] proposed a new approach. Saliency map based segmentation was used to separate the foreground from the background. Afterwards, features were extracted from the segmented image of the pest. One feature is energy which is computed via discrete wavelet transform. The features later on was stored with the database along with the name of the pest. With back propagation neural network, the pest is classified and control and preventive measures were provided.

The preceeding review shows that there have been many

methods and algorithms developed for automated pest identification and each has its own advantage and disadvantages. There has been no study on the method to automatically identify pest under order Hemiptera making this study relevant.

III. MATERIALS AND METHODS

A. Materials

Data for this study will be actual images of pests(Hemiptera) taken from IRRI and other fields. A digital camera and smart phone will be used to collect images. Images will be taken under daylight with no flash.

B. Development Tools

The system will be developed under Ubuntu Operating System. To be able to implement the system, it will be necessary to install and use the following:

- 1) Content Management System(CMS): A Content Management System allows publishing, editing and modifying content, organizing, deleting as well as maintenance from a central interface. It will be easier for non-developers to maintain a system. WordPress 4.0 shall be used for the pest and diseases library where the identification of pest will be integrated. WordPress is built on PHP and MySQL. [13]
- 2) Digital Image Processing Library: OpenCV is an open source library for image processing. It is optimized in C++ but it also has Python and Java Interfaces. For this development, the C++ Interface will be used. It will be integrated with WordPress. [14]
- 3) Database: MySQL 5.1.41 will be used as database for this application. It is an open-source database management system. This is where the information about the different pests and diseases shall be stored and the data for the pest and diseases identification.

C. System Implementation

The system consists of a pests and diseases library for crops where the system for pest identification will be integrated to. Figure 2 shows the phases undergone by each pest image. It is composed of five basic phases of digital image processing. The first phase is the collection of pest images of order Hemiptera in order to create a database. Using saliency map techniques, the image is segmented to separate the pest from the background. Features of the segmented image like energy of an image is extracted and these features are stored in the database with the name of the pest. Using K-nearest neighbors algorithm, the pest will be identified and the system will give relevant information to control it.

1) *Image Acquisition*: Images for the training and input data will be obtained using a digital camera or a scanner and will be uploaded to the system. For this study, a database of the features from the training data will be needed for classifying the pest to its proper species.

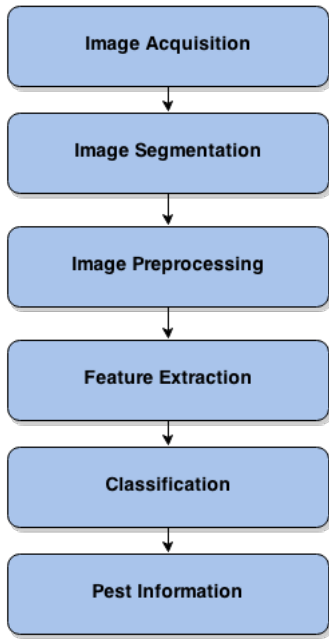


Fig. 2. Methodology Phases

2) *Image Segmentation*: Image Segmentation is the process of partitioning an image into homogenous groups. Its goal is to cluster pixels into noticeable image regions. [15] For this study, adaptive thresholding will be implemented to segment the image. Using segmentation, the pest image, which is the foreground will be extracted and will be given to the image preprocessing stage.

3) *Image Preprocessing*: Image Preprocessing involves removing unwanted distortions, degradation, and noise during the image segmentation. [16] Processes such as median filter, binary erosion, and binary dilation will be used to restore the pest image as close as possible to its original state. This process outputs a corrected image of the pest then it will be passed to the feature extraction phase.

4) *Feature Extraction*: A feature can be also viewed as a point of interest for image description. Properties of a good feature include consistency over several images of the same scene, invariant towards certain transformations, insensitive to noise, and salient or noticeable. A feature should be some quantifiable property of an object such that it quantifies significant characteristics of the object. Image features usually include color, shape and texture features. [17]

In this study, a gray-level co-occurrence matrix (GLCM) will be generated and normalized to compute the texture features. A gray-level co-occurrence matrix (GLCM) is essentially a two-dimensional histogram in which the (i,j) th element is the frequency of event i co-occurs with event j . A co-occurrence matrix is specified by the relative frequencies $P(i,j,d,\phi)$ in which two pixels, separated by distance d , occur in a direction specified by the angle ϕ , one with gray level i and the other with gray level j . A co-occurrence matrix is therefore a function of distance d , angle ϕ and grayscales i and j . [18]

For this study, the features to be computed from the GLCM for classifying the pest are Entropy, Homogeneity, Energy, and

Correlation.

1) Entropy in any system represents disorder, where in the case of texture analysis is a measure of its spatial disorder.

images/entropy.png

2) Homogeneity measures the uniformity of the non-zero entries in the GLCM.

$$Homogeneity = \sum_{i,j} \frac{1}{1 - (i - j)^2} p(i, j)$$

3) Energy is a measure of local homogeneity and therefore it represents the opposite of the Entropy. Basically this feature will tell us how uniform the texture is.

$$Energy = \sum_{i,j} p(i, j)^2$$

4) The Correlation texture measures the linear dependency of gray levels on those of neighboring pixels.

Correlation	$\frac{\sum_{m=0}^{G-1} \sum_{n=0}^{G-1} mnp(m, n) - \mu_x \mu_y}{\sigma_x \sigma_y}$ <p>where</p> $\mu_x = \sum_{m=0}^{G-1} m \sum_{n=0}^{G-1} p(m, n)$ $\mu_y = \sum_{n=0}^{G-1} n \sum_{m=0}^{G-1} p(m, n)$ $\sigma_x = \sum_{m=0}^{G-1} (m - \mu_x)^2 \sum_{n=0}^{G-1} p(m, n)$ $\sigma_y = \sum_{n=0}^{G-1} (n - \mu_y)^2 \sum_{m=0}^{G-1} p(m, n)$
-------------	--

The features that will be extracted will be the input for the classification stage.

5) *Classification*: With the help of K-nearest neighbors algorithm, the possible species and life stage of the pest will be identified and the system will give preventive as well as control measures to it.

K-nearest neighbors algorithm is a non parametric lazy learning algorithm. It means that it does not make any assumptions on the underlying data distribution. It is a lazy learning algorithm because it defers the decision to generalize beyond the training examples until a new query is encountered. stores all training samples and predicts the response by analyzing a certain number (K) of the nearest neighbors of the sample. [19]

Training samples are vectors in a multidimensional feature space, each with a class label. In the training phase, the feature vectors and class labels of the training samples will be stored in a database.

For the classification phase, a number k will be defined by the programmer. The unlabeled vector(a query) is classified by assigning the label which is the most frequent among the k training samples nearest to that query point.

To find the input data's K nearest neighbors in this study, Euclidean Distance will be used to calculate the distance.

REFERENCES

- [1] IRRI. How to manage pests and diseases. [Online]. Available: <http://www.knowledgebank.irri.org/step-by-step-production/growth/pests-and-diseases>
- [2] G. Kandalkar, A. V. Deorankar, and P. N. Chatur, "Classification of agricultural pests using dwf and back propagation neural networks," *International Journal of Computer Science and Information Technologies*, vol. 5, pp. 4034–4037, 2014.
- [3] —, "A review of literature on application of image processing for identification of agricultural pests on various crops," *International Journal of Engineering Science and Innovative Technology*, vol. 2, pp. 601–606, Nov. 2013.
- [4] Hemiptera: bugs, aphids and cicadas. [Online]. Available: <http://www.ento.csiro.au/education/insects/hemiptera.html>
- [5] Orkin. Orkin pest library. [Online]. Available: http://www.orkin.com/learningcenter/pest_library/
- [6] G. Y. Kandalkar and A. V. Deorankar, "Overview of image processing approach for identification of agricultural pest," *International Journal of Advanced Electronics and Communication Systems*, Feb. 2014.
- [7] 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, Mar./Oct. 2013.
- [8] L. Zhu and Z. Zhang, "Auto-classification of insect images based on color histogram and glcm," in *Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 2010*.
- [9] F. Zhang, G. Wang, J. Ye, and Q. Liu, "An automatic identification system of agricultural pests based on machine vision," *IOSR Journal of Electronics and Communication Engineering*, vol. 6, pp. 68–74, May/June 2013.
- [10] S. M. Al-Saqer, P. Weckler, J. Solie, M. Stone, and Wayadande, "Identification of pecan weevils through image processing," *American Journal of Agricultural and Biological Sciences*, vol. 6, pp. 69–79, 2011.
- [11] J. Cho, J. Choi, M. Qiao, C. Ji, H. Kim, K. Uhm, and T. Chon, "Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis," *International Journal of Mathematics and Computers in Simulation*, vol. 1, pp. 46–53, 2007.
- [12] F. Fina, P. Birch, R. Young, J. Obu, B. Faithpraise, and C. Chatwin, "Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters," *International Journal of Advanced Biotechnology and Research*, vol. 4, pp. 189–199, 2013.
- [13] WordPress.org. Wordpress.org. [Online]. Available: <https://wordpress.com/>
- [14] itseez. OpenCV: Open source computer vision. [Online]. Available: <http://opencv.org/>
- [15] Image segmentation. [Online]. Available: <http://www.cs.toronto.edu/~jepson/csc2503/segmentation.pdf>
- [16] U. of Delaware. Image preprocessing. [Online]. Available: http://climate.geog.udel.edu/~tracyd/geog474/geog474_preprocessing.html
- [17] A. Aichert. Feature extraction. [Online]. Available: <http://home.in.tum.de/~aichert/featurepres.pdf>
- [18] M. E. Manaa, N. T. Obies, and T. A. Al-Assadi, "Object classification using a neural networks with gray-co-occurrence matrices (glcm)."
- [19] S. Thirumuruganathan. A detailed introduction to k-nearest neighbor (knn) algorithm. [Online]. Available: <http://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/>