Butterfly Species Recognition Using Artificial Neural Network

S. S. N. Alhady and Xin Yong Kai

Abstract In 2017, there are about 20,000 species of butterfly has been discovered all over the world. Butterfly is well known because of its beautiful wings pattern and its benefits to the environment. In this research, butterfly species recognition is automated using artificial intelligence. Pattern on the butterfly wings is used as a parameter to determine the species of the butterfly. The butterfly image is captured and the background of the image is removed to make the recognition process easier. Local binary pattern (LBP) descriptor is then applied to the processed image and a histogram consist of image information is computed. Artificial Neural Network (ANN) is used to classify the image. Two types of butterfly species were selected namely *ideopsis vulgaris* and *hypolimnas bolina*. Both of the species have been correctly identify with accuracy of 90% (for *ideopsis vulgaris*) and 100% (for *hypolimnas bolina*).

Keywords Pattern recognition · Artificial intelligence · Neural network

1 Introduction

Species recognition has been one of the research carried out by researchers in near decades, especially when the conservation of endangered species become more common and urgent nowadays. There are a few methods to carry out species identification for all sorts of living organism on earth. Some methods need the individual to be caught then collect sample and statistics to do further investigation. Individual with distinct size, shape and colour can be recognized effortlessly. When it comes to insects, they are small and they may look similar but are from different species. Locust and grasshopper made a good example in this case. Both have very similar appearance such as features and shape. These two-species behaved very

S. S. N. Alhady (✉) · X. Y. Kai
School of Electrical & Electronic Engineering, Universiti Sains Malaysia,
George Town, Malaysia
e-mail: sahal@usm.my

M. H. A. Hassan (ed.), *Intelligent Manufacturing & Mechatronics*,
Lecture Notes in Mechanical Engineering,
https://doi.org/10.1007/978-981-10-8788-2_40

differently after study has been carried out. Same goes to butterfly, all the butterflies on earth have similar appearance. They are classified under the category of insects. Scientists have classified the butterfly under different categories based on the butterfly wing shape and behavior. The butterflies have been divided into approximately 135 families [1–3].

Previously, identification of butterfly species was done by first glance observation. The first glace observation found not working properly. The genital characteristics analysis have to be carry out by preparing genetal slides of the collected butterflies through some spesific processing using various chemicals. This method become less popular due to argument and misinterpretation of butterfly species [2].

This research will focus on the recognition of butterfly species using artificial neural network (ANN). In this research, *ideopsis vulgaris* (commonly known as Blue Glassy Tiger butterfly) and 39 images of *hypolimnas bolina* (commonly known as Bluemoon butterfly) are being chosen. The butterflies that need to be recognise will be captured using any devices. Image processing technique which is local binary pattern (LBP) is used to process the pictures of butterflies [4, 5]. A single layer artificial neural network is used to match the processed data to the database and found the correct species [6].

2 Methodology

This section explains the overall project flow of automatic identification of butterflies' species. The overall project flow will be visualized using flow chart. The process to identify butterfly is separated into five phases. Each phase will be discussed in detail. The five phases are data collection, data pre-processing, feature extraction, classification and performance evaluation.

2.1 Data Collection

The total number of images collected is tabulated according to butterfly species and usage of the images.

2.2 Data Pre-processing

Two butterfly species were selected to be recognized in this project. Hence, images of butterfly are collected to carry out this project. Images collected were taken at the Penang Butterfly Farm (Entopia). After the photo had been taken, volunteer from Penang Butterfly Farm helped to identify the species of the butterfly. This was to ensure the images used in this project are unambiguous. Especially the butterfly species that have similar wings pattern, extra attention is needed to observe the details on the wings to ensure the butterfly species are the one that was chosen to be studied. The total butterfly images collected were 78 images. These images were made up of 39 images of *ideopsis vulgaris* (commonly known as Blue Glassy Tiger butterfly and 39 images of *hypolimnas bolina* (commonly known as Bluemoon butterfly).

2.3 Feature Extraction

Local binary pattern (LBP) descriptor was used to do feature extraction in this project [7, 8]. The RGB image was converted to grey scale image and use the with descriptor. In LBP descriptor, the image was divided into cells. Each cell contained R radius and N number of neighbours as shown in Fig. 1.

The centre pixel's value was compared with each of the neighbour pixel's value. Below showed the formula of local binary pattern for a pixel:

$$LBP(\mathrm{x}) = \sum_{n=0}^{p} \mathrm{S}(\mathrm{G}(\mathrm{x}_\mathrm{n})\mathrm{G}(\mathrm{x}))2^{\mathrm{n}-1}$$

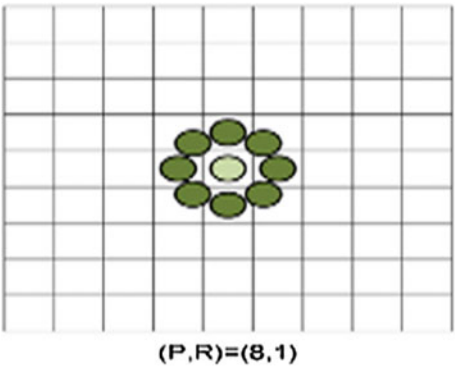
$$\mathrm{S}(\mathrm{t}) = \begin{cases} 1, t \geq 0 \\ 0, t < 0 \end{cases}$$

where

x location of centre pixel
x_i location of ith neighbouring pixel

When the neighbour pixel's value was smaller than the center pixel's value, the neighbour pixel's value was replaced by binary number "0". When the neighbour pixel's value was greater than the center pixel's value, the neighbour pixel's value

Fig. 1 Example of dividing image into cells using different values of LBP descriptor



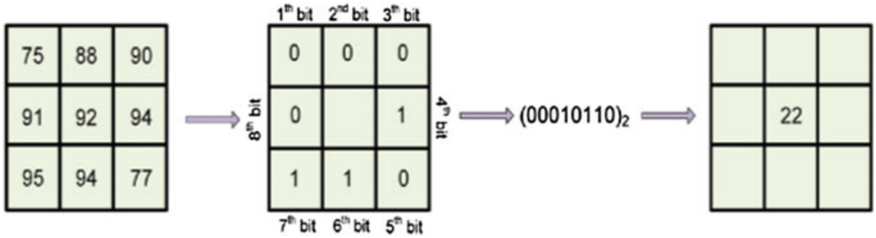


Fig. 2 Process of computing decimal representation of a cell using LBP descriptor

was replaced by binary number "1". This resulted in a series of binary number representing the cell as shown in Fig. 2.

A histogram containing components equivalent to the number of cells was computed. The average, deviation, energy, entropy and correlation of the histogram was calculated using the following formula.

$$average = \frac{1}{AB}\sum_{x=0}^{A-1}\sum_{y=0}^{B-1} f(x,y)$$

$$deviation = \sqrt{\frac{1}{AB}\sum_{x=0}^{A-1}\sum_{y=0}^{B-1} (f(x,y) - average)^2}$$

$$energy = \frac{1}{AB}\sum_{x=0}^{A-1}\sum_{y=0}^{B-1} f^2(x,y)$$

where

μ mean
σ standard deviation

2.4 Classification

An Artificial neural network was a model that designed based on the structure and functions of the human brain neural networks. Neural network can be trained to perform different tasks such as computer vision, speech recognition and pattern recognition. There were a few advantages in using neural network which are ability to learn based on data given, suitable for real-time application and high fault tolerance using redundant information coding.

In this project, the average, deviation, energy, entropy and correlation of the histogram calculated from previous stage were used as input to the artificial neural network (ANN). The output of this stage were the butterfly species. In artificial neural network (ANN), there were a few strategies for learning to be carried out which were supervised learning, unsupervised learning and reinforcement learning. The learning strategy used in this project was supervised learning.

In supervised learning, a set of known input data and known output data was taken and the artificial neural network model was trained to give reasonable response of new data. The goal of supervised learning was to use existing data to do prediction for a new data. Since the purpose of this project was to do classification, the responses of the model were categorical variables [9].

There were a few steps involved in supervised learning which were data preparation, choosing algorithm, fitting a model, choosing validation method, examine fitting parameters performance, update fitting parameters and use fitted model for prediction.

2.5 Performance Evaluation

The classification results were tabulated into a confusion matrix. This made the process of analyzing the results easier as it was summarized in a table. The accuracy of the designed system was obtained by calculating the percentage of successful recognized image over the total image selected. The formula is presented as below.

$$Accuracy = \frac{number\ of\ sucessful\ recognised\ image}{total\ image\ selected} \times 100\%$$

3 Results

This section described the outcome of this research. Firstly, the total number of pictures collected is tabulated according to butterfly species and usage of the pictures. Next, the pictures are pre-processed. Feature extraction has been done on the edited pictures to get in formation of the pictures to do further classification of butterfly species. Finally, the classification step carried out using Artificial Neural Network is done and results are recorded.

3.1 Data Collection

Table below shows the number of pictures collected for each chosen species of butterflies. The number of pictures used for training purpose and testing purpose are listed in the Table 1.

Table 1 Number of pictures collected

Species	Species name	Number of pictures for training	Number of pictures for testing
1	*Ideopsis vulgaris*	29	29
2	*Hypolimnas bolina*	10	10
	Total	39	39

3.2 Data Pre-processing

Figure 3 shows the coloured image of the butterfly. The back ground of the image will be removed and replaced with white colour background.

3.3 Feature Extraction

Figure 4 shows the grayscale image that has been converted from the white colored background butterfly image. Local Binary Pattern (LBP) descriptor is used to convert the gray scaled image into LBP image. The histogram shows the summed up binary bits of range 0–255.

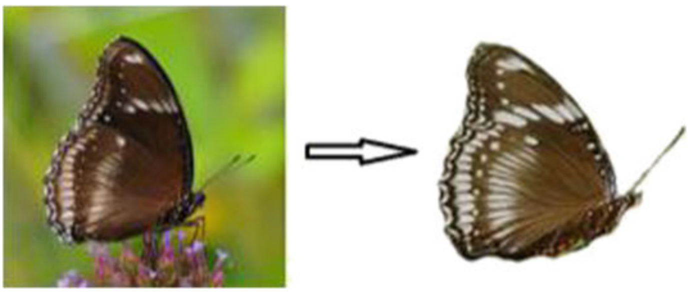


Fig. 3 Original and processed butterfly images

Fig. 4 Gray scaled image being converted to LBP image and LBP histogram

Table 2 ANN architecture and training parameters

Architecture	Training
Number of layers = 3	Performance function, MSE
Number of neuron on layers	Learning rule used Levenberg-Marquardt
Input = 3	
Hidden = 105	
Output = 2	
Initial weight = 0.01	Learning rate 0.7
Initial bias = 0.01	Momentum constant 0.97
Activation function = sigmoid for hidden and output layer	Performance goal = 2E−10

Table 3 Confusion matrix of butterfly species recognition

Butterfly species	Butterfly species 1	Butterfly species 2	Accuracy (%)
Butterfly species 1	9	1	90
Butterfly species 2	0	10	100

3.4 Classification

Table 2 shows the shows the Artificial Neural Network (ANN) architecture and training parameters.

Table 3 summarized the results of the butterfly species recognition results from Table 2. The accuracy of species recognition *for ideopsis vulgaris* butterfly species and *hypolimnas bolina* butterfly species are computed and recorded. It can be observed that recognition *of ideopsis vulgaris* butterfly species has the accuracy of 90% while recognition *of hypolimnas bolina* butterfly species 100% accurate.

4 Discussion

In this research, images of chosen butterfly species are recognized automatically using Artificial Neural Network (ANN). The research has been carried out in five phases which are data collection, image pre-processing, features extraction, classification and performance evaluation.

In data collection phase, the butterfly images are taken at the Penang Butterfly Farm. Two type of butterfly images are taken which are butterfly with opened wings and butterfly with closed wings. There are total of 78 images collected in this phase. The images include 48 images of *Ideopsis vulgaris* butterfly species and 49 images of *hypolimnas bolina* butterfly species. In the 49 images of each butterfly species, 39 images are used for training purpose while 10 images are used for testing. The images for training and testing are selected randomly.

In image pre-processing phase, the images background is removed. Also, the image is resized to same pixels. The images are adjusted so the size of the butterfly is roughly the same. The orientation of the butterfly is not fixed since the orientation will not affect the results of the butterfly species recognition.

In the feature extraction phase, the Local Binary Pattern (LBP) descriptor is used to do feature extraction based on the butterfly wings pattern. The images are divide into cells and each cell contained 9 pixels. A binary number will be computed for each cell. The binary numbers of all the cells are summarized into a histogram. In the histogram, it contains components equivalent to the number of cells was computed. The average, deviation, energy, entropy and correlation of the histogram was calculated. These data represent the pattern of the butterfly.

In classification phase, the average, deviation and energy are used as inputs for the neural network. The neural network consists of 3 layers which are input layer, hidden layer and output layer. There are 3 inputs in input layer and 2 outputs in output layer. The number of neurons in hidden layer is decided by training neural network with different number of neurons. The accuracies of the trained neural networks are recorded. Then, the results are compared to find the suitable number of neuron in hidden layer. After doing comparison among the accuracies of the trained neural network, 105 neurons in hidden layer of the neural network has the best performance. After the number of neurons in hidden layer has been decided, the performance of the neural network has been evaluated by looking at the performance plot and regression plot of the trained network. From the plots, it can be concluded that the training has been carried out correctly and the neural network is well trained. This can be told by the decreasing training mean squared error (MSE) over epochs, similar shape of validation mean squared error (MSE) and testing mean squared error (MSE) and R values that are very close to 1. There is no over fitting parameter occurs in the training. The testing data set is being used to test the trained neural network to identify selected butterfly species based on the average, deviation and energy calculated.

In performance evaluation phase, the results from classification stage is summarized into a confusion matric. This is to ensure the results can be read easily, the accuracy to recognize each species is calculated accordingly. As conclusion, ideopsis vulgaris butterfly species has lower accuracy as compared to hypolimnas bolina butterfly species. The accuracies for both the butterfly species are 90% and 100% respectively.

5 Conclusion

This research aimed to recognize butterfly species automatically using Artificial Neural Network. Two different type of butterfly have been chosen in this research. The chosen butterfly species are ideopsis vulgaris butterfly species and hypolimnas bolina butterfly species. The images of the mentioned butterfly species are collected and edited. Edited images will undergo feature extraction and classification.

The final results are recorded and the performance of butterfly species recognition is evaluated.

The key feature used to recognize the butterfly species in this research is the pattern on the butterfly's wings. The pattern of butterfly wings is digitalized into information using Local Binary Pattern (LBP) descriptor. After LBP descriptor, the pattern that can be visualized by human eyes is transposed to numbers. The values are the average, deviation and energy of the Local Binary Pattern (LBP) histogram. These values will be used as input for the neural network to be used as variables for species recognition.

At the end of this research, it can be concluded that the selected butterfly species can be recognized automatically. The accuracy of both the butterfly species are 90% and 100% respectively.

References

1. Kaya, Y., Kayci, L., Uyar, M.: Automatic identification of butterfly species based on local binary patterns and artificial neural network. Appl. Soft Comput. **28**, 132–137 (2015)
2. Kang, S.-H., Cho, J.-H., Lee, S.-H.: Identification of butterfly based on their shapes when viewed from different angles using an artificial neural network. J. Asia-Pacific Entomol. **17**, 143–149 (2014)
3. Kang, S.-H., Song, S.-H., Lee, S.-H.: Identification of butterfly species with a single neural network system. J. Asia-Pacific Entomol. **15**, 431–435 (2012)
4. Liu, L., Fieguth, P., Zhao, G., Pietikäinen, M., Hu, D.: Extended local binary patterns for face recognition. Inf. Sci. **358–359**, 56–72 (2016)
5. Liu, L., Fieguth, P., Guo, Y., Wang, X., Pietikäinen, M.: Local binary features for texture classification: taxonomy and experimental study. Pattern Recogn. **62**, 135–160 (2017)
6. Banerjee, A.K., Kiran, K., Murty, U.S.N., Venkateswarlu, C.: Classification and identification of mosquito species using artificial neural networks. Comput. Biol. Chem. **32**, 442–447 (2008)
7. Abbasi, S., Tajeripour, F.: Detection of brain tumor in 3D MRI images using local binary patterns and histogram orientation gradient. Neurocomputing **219**, 526–535 (2017)
8. Wang, X., Liang, J., Guo, F.: Feature extraction algorithm based on dual-scale decomposition and local binary descriptors for plant leaf recognition. Digit. Signal Proc. **34**, 101–107 (2014)
9. Żbikowski, R.: Use of artificial neural networks to identify the origin of green macroalgae. Estuar. Coast. Shelf Sci. **94**, 138–143 (2011)