Register 6 (1) 2020 pp. 11-21
Jurnal Ilmiah Teknologi Sistem Informasi (Scientific Journal of Information System Technology)

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

Available online to www.journal.unipdu.ac.id

Unipdu

S2-Accredited – SK No. 34/E/KPT/2018
Journal page is available to www.journal.unipdu.ac.id:8080/index.php/register

Butterfly identification using gray level co-occurrence matrix (glcm) extraction feature and k-nearest neighbor (knn) classification

Rico Andrian [a], Devi Maharani [b], Meizano Ardhi Muhammad [c], Akmal Junaidi [d]

[a,b,c,d] *Department of Computer Science, Universitas Lampung, Bandar Lampung, Indonesia*

email: [a] rico.andrian@fmipa.unila.ac.id, [b] devi.maharani1582@students.unila.ac.id, [c] meizano@eng.unila.ac.id, [d] akmal.junaidi@fmipa.unila.ac.id

ARTICLE INFO

Article history:
Received 28 June 2019
Revised 8 September 2019
Accepted 30 October 2019
Published 4 December 2019

Keywords:
butterflies
glcm
knn
pattern recognition

IEEE style in citing this article:
R. Andrian, D. Maharani, M. A. Muhammad and A. Junaidi, "Butterfly identification using gray level co-occurrence matrix (glcm) extraction feature and k-nearest neighbor (knn) classification," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 6, no. 1, pp. 11-21, 2020.

ABSTRACT

Gita Persada Butterfly Park is the only breeding of engineered in situ butterflies in Indonesia. It is located in Lampung and has approximately 211 species of breeding butterflies. Each species of Butterflies has a different texture on its wings. The Limited ability of the human eye to distinguishing typical textures on butterfly species is the reason for conducting a research on butterfly identification based on pattern recognition. The dataset consists of 600 images of butterfly's upper wing from six species: Centhosia penthesilea, Papilio memnon, Papilio nephelus, Pachliopta aristolochiae, Papilio peranthus and Troides helena. The pre-processing stage is conducted using scaling, segmentation and grayscale methods. The GLCM method is used to recognize the characteristics of butterfly images using pixel distance (d) = 1 and Angular direction 0^o, 45^o, 90^o and 135^o. The features used is angular second moment, contrast, homogeneity and correlation. KNN classification method in this study uses k values=1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 and 23 based on the Rule of Thumb. The result of this study indicate that Centhosia penthesilea and Papilio nephelus classes can be classified properly compared to the other 4 classes and require a classification time of 2 seconds at each angular orientation. The highest accuracy is 91.1% with a value of k = 5 in the angle of 90o and error rate=8.9%. Classification error occured because the value of the test data features is more dominant with the value of the training image features in different classes than the supposed class. Another reason is because of imperfect test data.

1. Introduction

Gita Persada Butterfly Park is the only engineered in situ Butterfly Breeding in Indonesia with approximately 211 species of breeding butterflies. Butterflies belong to the insect class and have different venations in each species. Each butterfly's venation is given a name or description code to facilitate the classification of butterflies with different patterns of shapes and colour [1]. Butterfly identification was carried out to distinguish the butterfly species in the Gita Persada Butterfly Garden. Researchers usually only use their eyes as butterfly identification tools, the limitation of the human eye to distinguish unique textures in each butterfly species is the reason for conducting butterfly identification based on pattern recognition research.

This study uses the extraction of the Gray Level Co-occurrence Matrix (GLCM) feature and the K-Nearest Neighbor (KNN) classification in identifying butterfly images. The GLCM and KNN methods have been widely used in pattern recognition research, such as research Kulkarni et al [2] that recognizes human iris based on 3 types of models: Outer Circle Mask, Inner Circle Mask and ROI based Iris using GLCM texture features resulting in an accuracy rate of 96.3%. Kaushal and Bala [3] identified plant diseases based on GLCM texture features, K-Mean Clustering segmentation, and KNN

ISSN 2502-3357 (Online)
R. Andrian et al./Register 6 (1) 2020 pp. 11-21
ISSN 2503-0477 (Print)

classification producing an accuracy of 80% -90%, then research Indriani et al [4] in identifying maturity Tomato fruit based on GLCM texture features and HSV colour features resulting accuracy rate of 90% - 100%.

Relevant research, Purnomo [5] classified butterflies using Gray Level Co-occurrence Matrix (GLCM) feature extraction method and the K-Nearest Neighbor (KNN) classification method with a dataset of 832 images consisting of 10 species of butterflies. The accuracy rate reached 88% with parameters d = 1, θ = 45° and *k* = 3. Ertuğrul et al [6] classified butterflies based on their energy texture using Gray-Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), and Gabor Filtered (GF) extraction features, and K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Extreme Learning Machine (ELM) methods. The accuracy level is 99.26% for the KNN method, 98.16% for the SVM method, and 99.47% for the ELM method.

The purpose of this study is to identify butterfly images using feature extraction of the Gray Level Co-occurrence Matrix (GLCM) and K-Nearest Neighbor (KNN) classification on Cethosia penthesilea, Papilio memnon, Papilio nephelus, Pachliopta aristolochiae, Papilio peranthus, and Troides helena in order to facilitate the Gita Persada Butterfly Park in identifying butterfly species and to support developers in building butterfly identification applications.

2. Research Method

Research stages in the butterfly images identification using feature extraction of the Gray Level Co-occurrence Matrix (GLCM) and the classification of K-Nearest Neighbor (KNN) are shown in Figure 1.

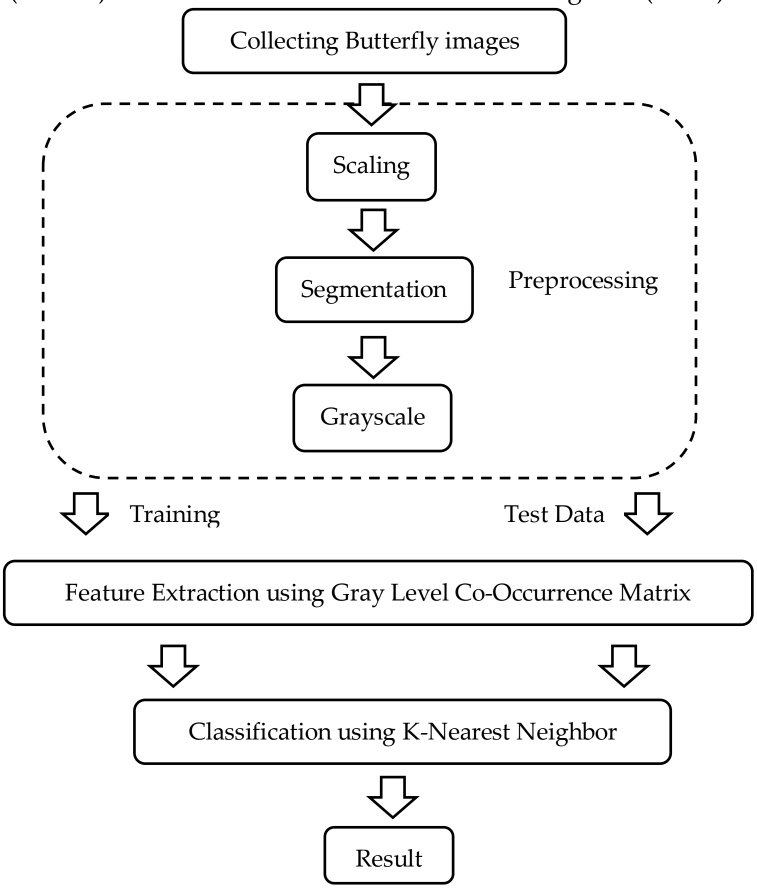


Figure 1. Research stages

2.1. Butterfly image collection

Collection of butterfly images obtained from Gita Persada Butterfly Park. The location where the data was taken is located on Wan Abdurrahman street, Hutan, Kecamatan Hutan, Lampung. The images of the butterflies used were Cethosia penthesilea (Figure 2), Papilio memnon (Figure 3), Papilio peranthus (Figure 4), Troides helena (Figure 5), Papilio nephelus (Figure 6), and Pachliopta aristolochiae (Figure 7). The images taken for each type are 100 images in JPG (Joint Photographic Group) format. The next image will be stored in a folder with the name of the dataset.

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)
R. Andrian et al./Register 6 (1) 2020 pp. 11-21

Figure 2. Image of Cethosia penthesilea

Figure 3. Image of Citra Papilio memnon

Figure 4. Image of Papilio peranthus

Figure 5. Image of Troides helena

Figure 6. Image of Papilio nephelus

Figure 7. Image of Pachliopta aristolochiae

2.2. Scaling

The image which was originally sized 6000 × 4000 pixels was cropped to 256 × 256 pixels in order to show its characteristics. The image pixel is reduced to limit the object value to compare so it will not hamper the classification process [7].

2.3. Segmentation

Segmentation is part of the preprocessing phase which aims to separate the foreground from the background [7]. The segmentation process is done manually using Adobe Photoshop software. The image that has gone through the segmentation process can be seen in Figure 8.

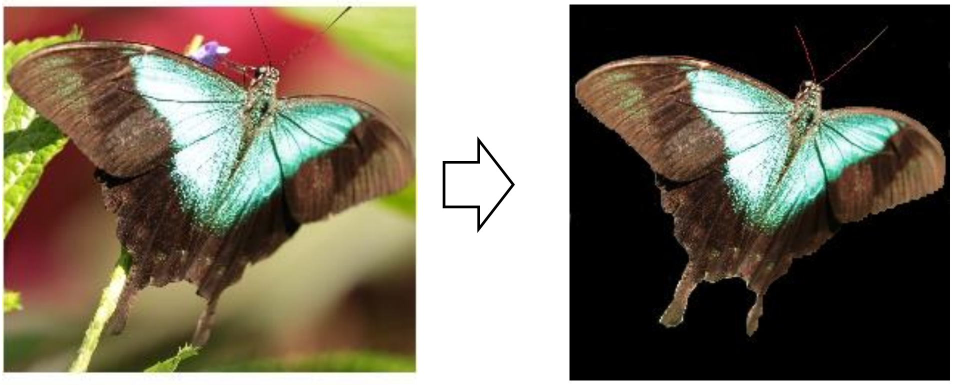


Figure 8. Segmentation stages

2.4. Grayscale

Grayscale is a process to change the color of an object into a grayish image which aims to convert all RGB (Red Green Blue) color information that has 3 image composing matrices converted to only 1 matrix [8]. This study uses 8 levels of grayish values. The gray level is simplified so that not much data is processed, because in a colour image, every pixel has three layers namely Red, Green and Blue while for the gray image, per pixel is only represented by one level of gray [9].

2.5. GLCM feature extraction

GLCM feature extraction begins by creating a co-occurrence matrix. A co-occurrence matrix represents neighborhood relations between two pixels in an image in various orientation directions and spatial distance [10]. This matrix is formed from an image by looking at the relationship between the two pixels at a certain distance and angular orientation [11]. This matrix is used to extract texture features from an

R. Andrian et al./Register 6 (1) 2020 pp. 11-21
ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

image with angular orientation 0° (horizontal), 45° (diagonal), 90° (vertical), and 135° (anti-diagonal) [12].

The features used in this study are:

1. Angular Second Moment (ASM)
   ASM is used to calculate the pair intensity concentration on matrices using Equation 1.
   $$\mathrm{ASM} = \sum_{i=i}^{Ng} \sum_{j=1}^{Ng} P_{ij}{}^{2} \quad (1)$$
2. Contrast
   Contrast is a measure of the distribution of matrix elements in an image, calculated using Equation 2.
   $$\mathrm{Contrast} = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (i-j)^2 \times P_{ij} \quad (2)$$
3. Homogeneity
   Homogeneity or Inverse Difference Moment (IDM) shows the homogeneity of an image that has a similar degree of gray, calculated using Equation 3.
   $$\mathrm{IDM} = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} \frac{1}{1+(i-j)^2} \times P_{ij} \quad (3)$$
4. Correlation
   Correlation is a measure of linear dependence between gray level values in an image, calculated using Equation 4.
   $$\mathrm{Correlation} = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} \frac{(i-\mu_i)(j-\mu_j) \times Pij}{1+(i-j)^2} \quad (4)$$
   Description:
   P_{ij} = Values in row *i* and column *j* in the GLCM matrix
   N_g = Gray level in the image
   i = Position or location (index) value on the row in the GLCM matrix
   j = Position or location (index) value in the column in the GLCM matrix
   μ_i, σ_i = Mean and Standard Deviation on the GLCM matrix
   $\boldsymbol{\mu_j}, \boldsymbol{\sigma_j}$ = Mean and Standard Deviation on the vertical GLCM matrix

2.6. KNN classification

Classification is the process of finding a model (or function) that describes and distinguishes classes of data or concepts that aim to be used to predict classes from objects whose class labels are unknown [13]. KNN is a method for identifying objects based on the most similar data (nearest neighbors) with the specified number of *k* and classifying them into new classes [14]. KNN has attributes that are initialized as *k*, the number of neighbor values referred to the KNN classification, the number k values are positive integers, small and odd numbers [11]. The KNN classification algorithm predicts the test sample category according to the training sample *k* which is the nearest neighbor to the test sample, and puts it into the category that has the largest probability category [15]. *k* value in the KNN classification is obtained using the Rule of Thumb rule in Equation 5, resulting *k* value 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23.

$$k = \sqrt{n} \quad (5)$$

Description:

$\boldsymbol{k}$ = Number of nearest neighbors to be selected

$\boldsymbol{n}$ = number of datasets used

The stages of classification using the KNN method in this study are as follows:

1. Determine the value of *k* (number of nearest neighbors to be selected)
2. Calculate the distance between classified data with all training data using euclidean distance (Equation 6),
   $$D(a,b) = \sqrt{\sum_{n}^{kn} (a_k - b_k)^2} \quad (6)$$
   Description:

$D(a, b)$ = Euclidean distance
a = Training data point
b = Testing data point

3. Sort the distance formed ascending
4. Determine the nearest distance k
5. Pair the appropriate class
6. Find class with the the highest k nearest neighbors and set the class as the data class to be evaluated.

2.7. Calculation of accuracy level

Calculation of accuracy is one of the important things in pattern recognition. This process is carried out as one of the evaluation benchmarks in a system. Accuracy level measurement can be used in various ways one of which is to use the Detection Rate. The Detection Rate Equation is represented as follows (Equation 7) [16],

$$\text{Detection Rate} = \frac{\text{TP}}{\text{TP+TN}} \times 100\% \quad (7)$$

Description:
TP (True Positive) = Data predicted is true

TN (True Negative) = Data predicted is wrong

3. Result and Discussion

Table 1. Test results with angular direction 0°, 45°, 90° dan 135°

Total value of k	**Accuracy at each angular direction**			
	0°	45°	90°	135°
1	88,3%	89,4%	90%	90,5%
3	88,3%	87,7%	90,5%	90,5%
5	88,3%	89,4%	91,1%	89,4%
7	87,2%	87,2%	90,5%	90%
9	86,6%	87,2%	89,4%	88,8%
11	85,5%	86,6%	90%	87,2%
13	83,8%	86,6%	89,4%	86,1%
15	83,8%	86,1%	89,4%	85,5%
17	83,8%	85%	88,3%	85,5%
19	83,8%	85,5%	88,8%	85,5%
21	83,8%	84,4%	87,7%	85%
23	83,8%	84,4%	87,7%	85%

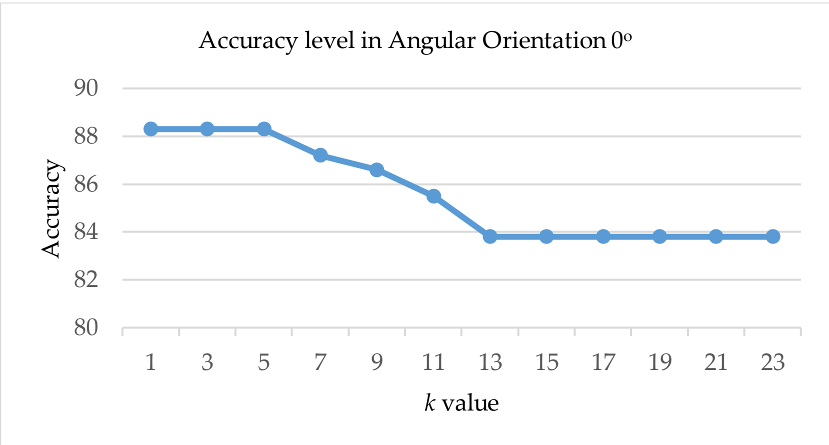


Figure 9. Graph of differences in accuracy level in the angular orientation 0°

This study uses a comparison of training data and test data of 70:30 (based on research conducted by Yodha and Kurniawan [17], so the training data used amounted to 420 images and test data of 180 images. The results of the GLCM feature extraction of the training data and test data are stored in the

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)
R. Andrian et al./Register 6 (1) 2020 pp. 11-21

48.9 KB file namely latih.mat and 21 KB file namely uji.mat, so that the system has good generalization capabilities in classifying data. The test is carried out at four orientation angles, 0°, 45°, 90°, dan 135°, the test results are displayed in Table 1 in the form of accuracy level or detection rate (Equation 7).

Table 1 shows the accuracy of each experimen. It can be seen from the table that the highest accuracy rate is 91.1% at an angular orientation of 90° with a value of k = 5. The lowest accuracy rate of 83.8% is obtained at an angular orientation of 0° with k value = 13, 15, 17, 19, 21, and 23. The difference in accuracy level at each angular orientation shown in Table 1 can be seen in the graphs in Figure 9, Figure 10, Figure 11 and Figure 12.

Figure 9 shows the accuracy level of each k value with the angular orientation of 0°. The highest accuracy is obtained at k = 1, 3, and 5 which is 88.3%, while the lowest accuracy is obtained at k = 13, 15, 17, 19, 21, and 23 which is 83.8%.

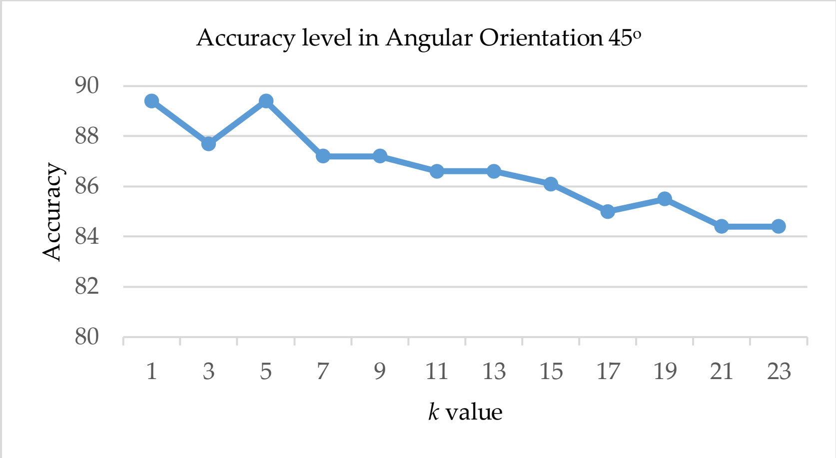


Figure 10. Graph of differences in the level of accuracy in the angular orientation 45°

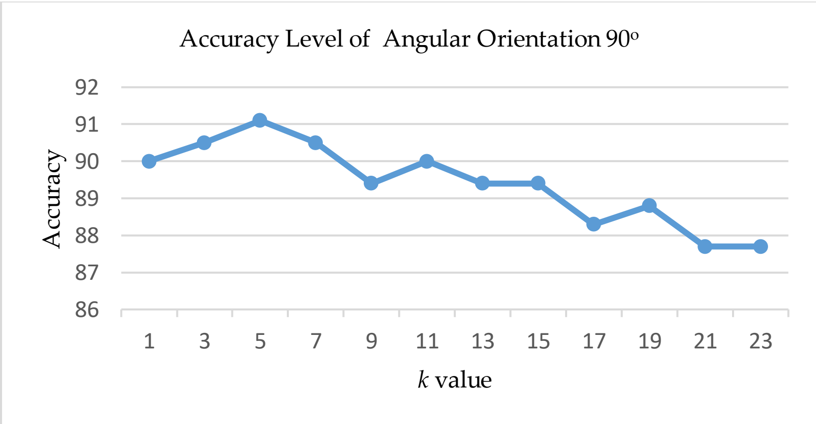


Figure 11. Graph of differences in the level of accuracy in the angular orientation 90°

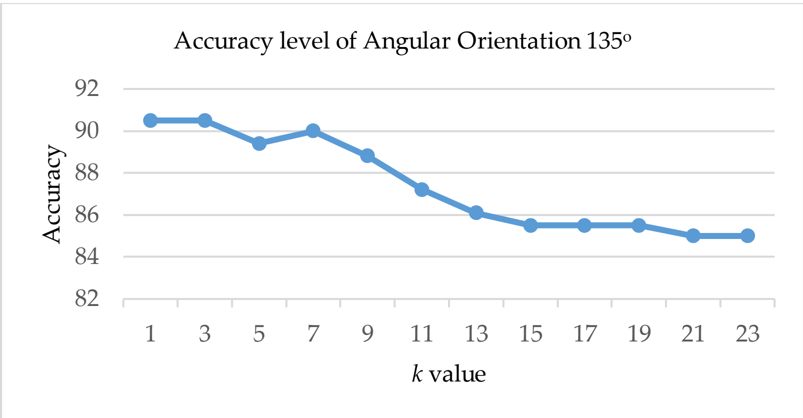


Figure 12. Graph of differences in the level of accuracy in the angular orientation of 135°

ISSN 2502-3357 (Online)
R. Andrian et al./Register 6 (1) 2020 pp. 11-21 ISSN 2503-0477 (Print)

Figure 10 shows the accuracy level of each *k* value with an angular orientation of 45°. The highest accuracy is obtained at *k* = 1 and 5 that is equal to 89.4%, while the lowest accuracy is obtained at *k* = 21 and 23 that is equal to 84.4%.

Figure 11 shows the accuracy level of each *k* value with angular orientation 90°. The highest accuracy is obtained at a value of *k* = 5 that is equal to 91.1%, while the lowest accuracy is obtained at a value of *k* = 21 and 23 that is equal to 87.7%.

Figure 12 shows the accuracy level of each *k* value with an angular orientation of 135°. The highest accuracy is obtained at values *k* = 1 and 3 that is equal to 90.5%, while the lowest accuracy is obtained at values *k* = 21 and 23 that is equal to 85%.

The butterfly identification system using GLCM feature extraction and KNN classification uses at the same time 420 butterfly images at the training data extraction stage, 180 butterfly images for the extraction stage of testing data, and 180 butterfly images at the classified testing data classification stage in each direction of the angle. The time required to run the process is shown in Table 2 and Table 3.

Table 2. Time required for the system at the extraction stage

	Training Data	**Testing Data**
Extraction stage	13 seconds	6 seconds

Table 3. The time needed for the system at the classification stage

	Time in each angular direction			
	0°	**45°**	**90°**	**135°**
Classification Stage	2 seconds	2 seconds	2 seconds	2 seconds

The test results at four accuracy level at each angular direction (Table 1) of the KNN classification results experience changes in the level of accuracy that tends to be unstable in each experiment due to the value of *k*. The third experiment had the best accuracy rate of 91.1% with *k* value = 5. The confusion matrix table was used to present the image test classified by the KNN method shown in Table 4.

Table 4. Confusion KNN classification results *k* = 5

	Cethosia penthesilea	**Papilio peranthus**	**Papilio nephelus**	**Pachliopta aristolochiae**	**Troides helena**	**Papilio memnon**
Cethosia penthesilea	30	0	0	0	0	0
Papilio peranthus	0	29	0	0	0	1
Papilio nephelus	0	0	30	0	0	0
Pachliopta aristolochiae	0	0	4	25	0	1
Troides helena	0	0	3	1	26	0
Papilio memnon	1	5	0	0	0	24

Confusion matrix data from Table 4 is used to measure the performance of butterfly identification systems, by calculating the amount of sensitivity (recall), precision, accuracy, dan error rate [11]. The calculation results of sensitivity (recall), precision, accuracy, dan error rate can be seen in Table 5.

1. Recall

   Recall is the level of success of the system in finding back information. The recall value can be calculated using Equation 8 [11],

   $$\text{Recall} = \frac{\sum_{i=1}^{l} \text{TP}_i}{\sum_{i=1}^{l} (\text{TP}_i + \text{FP}_i)} \times 100\% \quad (8)$$

   Description:
   TP_i = Positive data detected as correct in the *i*-class
   FN_i = Positive data detected as negative data in the *i*-class
   l = Number of the classes
2. Precision

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

Precision is the level of accuracy between the information requested by the user and the answers provided by the system. The precision value can be calculated using Equation 9 [11].

$$\text{Precision} = \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l} (TP_i + FP_i)} \times 100\% \quad (9)$$

Description:
TP_i = Positive data detected as correct in the *i*-class
FP_i = Negative data detected as positive in the *i*-class
l = Number of the classes

3. Accuracy

Accuracy is the level of closeness between the predicted value of the system and the actual value. Accuracy can be calculated using Equation 10 [11].

$$\text{Accuracy} = \frac{\sum_{i=1}^{l} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} \times 100\% \quad (10)$$

Description:
TP_i = Positive data detected as correct in the i-class
TN_i = Negative data detected as correct in the i-class
FP_i = Negative data detected as positive in the i-class
FN_i = Positive data detected as negative in the i-class
l = Number of the classes

4. Error Rate

Error Rate is the level of error that occurs in the classification of test data. Error rate can be calculated using Equation 11 [11].

$$\text{Error Rate} = 100\% - \text{Accuracy} \quad (11)$$

Table 5. Recall, precision, accuracy, dan error rate of butterfly classification results

Butterfly class	Result	
	Recall	**Precision**
Cethosia penthesilea	96.8%	100%
Papilio peranthus	85.3%	96.6%
Papilio nephelus	81.1%	100%
Pachliopta aristolochiae	96.2%	83.3%
Troides helena	100%	86.6%
Papilio memnon	92.3%	80%
Accuracy	91.1%	
Error Rate	8.9%	

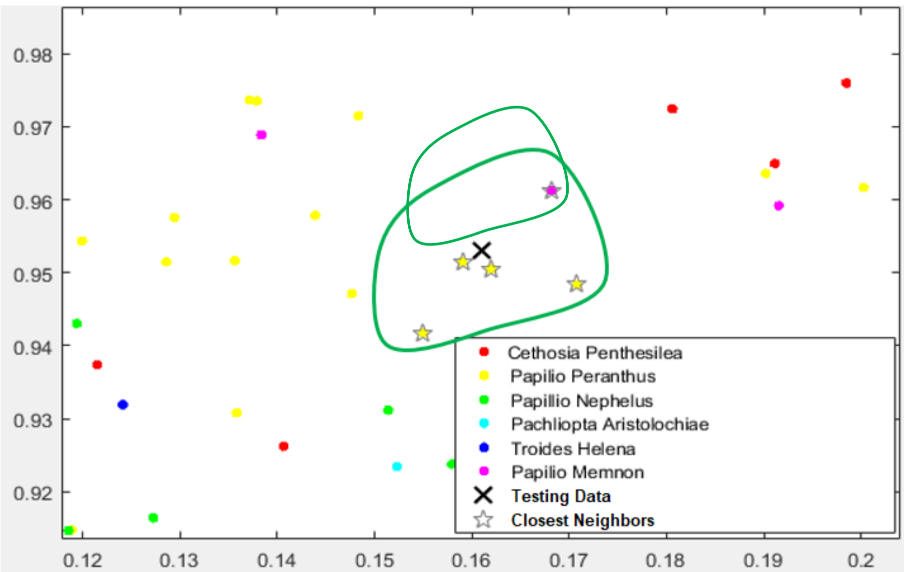
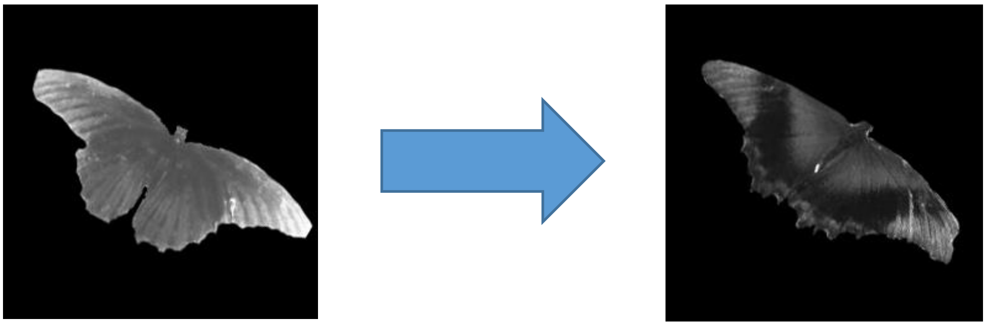


Figure 13. Distribution of test data for the of Papilio memnon class with KNN of nearest neighbors number is 5 ($k = 5$)

ISSN 2502-3357 (Online)
R. Andrian et al./Register 6 (1) 2020 pp. 11-21 ISSN 2503-0477 (Print)

The best level of search effectiveness was found in the Troides helena class with a recall value of 100%, while the lowest level of search effectiveness was in the Papilio nephelus class with a recall value of 81.1%. The high recall value indicates that the system is more effective in separating the butterfly classes properly in large numbers, that made Troides helena Class more effective in finding its class compared to the Papilio nephelus class.

The precision value is used to measure the accuracy of the system in displaying the information desired by the user according to the answers given. Based on Table 5, the precision value of the best butterfly class is in the class of Cethosia penthesilea and Papilio nephelus with an accuracy level of 100%, which means the testing data of the Cethosia penthesilea class and Papilio nephelus class were all correctly detected. Identification errors can occur because there are testing data that have features adjacent to features in other classes, shown in Figure 13.



(a) (b)

Figure 14. Images of Papilio memnon test data (a) classified as Papilio peranthus class (b)

Figure 13 shows that there is a test data distribution of Papilio memnon which was not identified correctly. Identification errors can occur because the image features value of the Papilio memnon test data to the number of nearest neighbors 5 (k = 5) is more dominant than it to the Papilio peranthus class, based on the results of the euclidean distance calculation on the Papilio memnon test data using Equation 6 shown in Table 6.

Table 6. Results of the euclidean distance calculation in the Papilio memnon test data with $k = 5$

Butterfly Class	ASM	Contrast	Homogeneity	Correlation	Euclidean Distance
Papilio memnon (test sample)	0.540980264	0.161090686	0.956637342	0.953063227	
Papilio peranthus	0.552544056	0.170815165	0.943716943	0.948376755	0.963022763
Papilio peranthus	0.573557773	0.147650194	0.949068354	0.947126024	0.99269613
Papilio memnon	0.56091588	0.179871588	0.960369968	0.92911441	0.966607848
Papilio peranthus	0.572224731	0.135628634	0.949444596	0.951653544	1.002047138

Table 6 shows the training data of the Papilio peranthus class yielding 4 neighbors closest to Papilio memnon Butterfly test data compared with the results of the euclidean distance calculation in the Papilio memnon class training data which yields only 1 neighbor nearest to Papilio memnon test data. The Papilio memnon test data (Figure 14) was not correctly identified, so it was classified as the Papilio peranthus class.

(a) (b) (c) (d) (e)

Figure 15. Imperfect test data of Papilio memnon class

ISSN 2502-3357 (Online)
ISSN 2503-0477 (Print)

Figure 14 shows that the image of the Papilio memnon test data (a) was incorrectly classified because the image of the Papilio memnon test data was more dominant to the Papilio peranthus class (b) (based on Table 6) than to the Papilio memnon class. Other Papilio memnon test data that were misclassified are shown in Figure 15.

Figure 15 shows that there are imperfect Papilio memnon class test data due to lack of lighting in Figure 15 (a), (b), (c), (d), and (e) resulting in Papilio memnon test data that is not properly classified.

4. Conclusions

The conclusion obtained based on research that has been done is the method of extracting the Gray Level Co-Occurrence Matrix (GLCM) feature and the K-Nearest Neighbor (KNN) classification has been successfully implemented in the butterfly pattern recognition process with the highest accuracy rate of 91.1% and errors classification of 8.9% and in the angular orientation testing of 90° with k = 5. Misclassification occurs because the value of the features in the test image is more dominant with the value of the training image features in a different class compared to the class that should be. Beside that, there are some imperfect test data. Other reason is that the system can recognize the Cethosia penthesilea and Papilio nephelus class with 100% accuracy, and it also can conduct data search in the Troides helena class more effectively with the highest recall value of 100% compare to the Cethosia penthesilea, Papilio memnon, Papilio nephelus, Pachliopta aristolochiae and Papilio peranthus Class.

For further research, there should be more number of butterfly image dataset used in order to improve accuracy. It is also important to increase the scope of the butterfly class so that the system can classify other butterfly types and develop a butterfly pattern recognition process using the feature extraction method and other classification methods. Taking pictures of butterflies also requires adequate lighting and uses different angles, such as using the lower wing angle and the angle when they had not flap their wings

5. References

[1] S. Herawati, A. Larasati, A. Djausal and M. Martinus, Kupu-Kupu Lampung Taman Kupu-Kupu Gita Persada, Bandar Lampung: Yayasan Sahabat Alam, 2016.

[2] S. B. Kulkarni, R. B. Kulkarni, U. P. Kulkarni and R. S. Hegadi, "GLCM-Based Multiclass Iris Recognition Using FKNN and KNN," *International Journal of Image and Graphics,* vol. 14, no. 3, 2014.

[3] G. Kausha and R. Bala, "GLCM and KNN based Algorithm for Plant Disease Detection," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,* vol. 6, no. 7, pp. 5845-5852, 2017.

[4] O. R. Indriani, E. J. Kusuma, C. A. Sari, E. H. Rachmawanto and D. R. I. M. Setiadi, "Tomatoes Classification Using K-NN Based on GLCM and HSV Color Space," in *International Conference on Innovative and Creative Information Technology (ICITech)*, Salatiga, 2017.

[5] G. A. Purnomo, "Klasifikasi Spesies Kupu-Kupu Menggunakan Ekstraksi GLCM dan Algoritma Klasifikasi K-NN," Universitas Dian Nuswantoro, Semarang, 2015.

[6] Ö. F. Ertuğrul, Y. Kaya, L. Kaycı and R. Tekin, "A Vision System for Classifying Butterfly Species by using Law's Texture Energy Measures," *International Journal on Computer Vision, Machine Learning and Data Mining,* vol. 1, pp. 16-24, 2015.

[7] S. Safinaz, "An Efficient Algorithm for Image Scaling with High Boost Filtering," *International Journal of Scientific and Research Publications,* vol. 4, no. 5, pp. 1-9, 2014.

[8] P. Kaler, "Study of grayscale image in image procesing," *International Journal of Recent and Innovation Trends in Computing and Communication,* vol. 4, no. 11, pp. 309-311, 2016.

[9] E. Budianita, J. Jasril and L. Handayani, "Implementasi Pengolahan Citra dan Klasifikasi K-Nearest Neighbour Untuk Membangun Aplikasi Pembeda Daging Sapi dan Babi," *Jurnal Sains, Teknologi dan Industri,* vol. 12, no. 2, pp. 242-247, 2015.

[10] M. A. Agmalaro, A. Kustiyo and A. R. Akbar, "Identifikasi Tanaman Buah Tropika Berdasarkan Tekstur Permukaan Daun Menggunakan Jaringan Syaraf Tiruan," *Jurnal Ilmu Komputer Agri-Informatika,* vol. 2, no. 2, pp. 73-82, 2013.

ISSN 2502-3357 (Online)
R. Andrian et al./Register 6 (1) 2020 pp. 11-21 ISSN 2503-0477 (Print)

[11] P. N. Andono, T.Sutojo and M. Muljono, Pengolahan Citra Digital, Yogyakarta: Andi, 2017.

[12] I. Zyout and I. Abdel-Qader, "Classification of Microcalcification Clusters via PSO-KNN Heuristic Parameter Selection and GLCM Features," *International Journal of Computer Applications,* vol. 31, no. 2, pp. 34-39, 2011.

[13] H. Leidiana, "Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor," *Jurnal Penelitian Ilmu Komputer,System Embedded & Logic,* vol. 1, no. 1, pp. 65-76, 2013.

[14] P. Harrington, Machine Learning in Action, Greenwich, CT, USA: Manning Publications Co, 2012.

[15] H. S. Khamis, K. W. Cheruiyot and S. Kimani, "Application of k- Nearest Neighbour Classification in Medical Data Mining," *International Journal of Information and Communication Technology Research,* vol. 4, no. 4, pp. 121-128, 2014.

[16] F. N. M. Sabri, N. M. Norwawi and K. Seman, "Identifying False Alarm Rates for Intrusion Detection System with Data Mining," *IJCSNS: International Journal of Computer Science and Network Security,* vol. 11, no. 4, pp. 95-99, 2011.

[17] J. W. Yodha and A. W. Kurniawan, "Pengenalan Motif Batik Menggunakan Deteksi Tepi Canny dan K-Nearest Neighbor," *Techno.COM,* vol. 13, no. 4, pp. 251-262, 2014.