

# Automated Identification of Moth Species

Y.M. Leong, Elham Y.K, Lim L.H.S, Sarinder Kaur Dhillon

**Abstract—** In this paper, Digital Automated Identification System (DAISY) was used to identify species of local moths. 210 species of superfamily *Bombycoidea* from Moth of Borneo (Part 3) were trained in DAISY. The overall identification of Moths gave a fairly accurate retrieval, with F1= 0.81.

**Keywords—** DAISY, automated identification, computer-assisted taxonomy, image classification

## I. Introduction

Malaysia has a National Biodiversity Index value of 0.809 (1.000 maximum) with 15,584km<sup>2</sup> of protected areas [1]. Being the 12<sup>th</sup> richest biodiversity community in the world, Malaysia's biodiversity is great opportunity for taxonomical investigations. Taxonomy is important in many biodiversity and biological applications [2]. On the other hand taxonomic impediments [3-5] made identification and classification of species hard task. Other options like for example molecular barcoding for genetic identification [6, 7], online accessibility to species' structure identification keys [8], and computer based techniques based on species image analysis [9], were tested as alternatives. Therefore, the ability of automating identification mechanism is a motivation for seeking these techniques.

The demand for increasing accuracy and speed of identification resulted recent developments in the computing world and introducing computing applications in the biological domains. One of the many crucial areas that computing applications introduced is the automated identification systems for species images, aimed to assist taxonomists in identifying organisms. Although, the role of experts and taxonomists is always important in identification but an advantage of automated identification systems is less dependency on experts. This motivates all researchers, either experts or not, who need taxonomy access [11, 5]. The automated identification systems by means of image analysis apply low level visual features from species images as meaningful attributes to classify them [10]. In biodiversity monitoring [12] and agriculture [13, 14], automated identification systems for insects received a huge interest. Manual identification of broad number of insect specimens is a big challenge for researchers. Many computational aided semi-automated systems have been implemented to aid classification [15-17]. Beside mentioned semi-automated systems, there are fully automated identification systems also developed for example Digital Automated Identification System (DAISY) [18], SPecies IDentification Automated (SPIDA) [25,26], Automated Bee Identification System (ABIS) [27], Automate Identification

Of Bees [19], Automatic Identification Of Whiteflies, Aphids And Thrips [20], Automatic Identification Of Live Moths [21], Automated Insect Identification [22, 23], Automated Real-Time Dynamic Identification Of Flying And Resting Butterfly [24]. DAISY was previously tested on various types of organism such as British bumble bees [28], Costa Rican Hawkmoths [8] and British Lepidoptera: Moths [29] while SPIDA was tested mainly on Australian ground spiders [25], ABIS was tested on bees [27]. DAISY, SPIDA, Automatic Identification of Whiteflies, Aphids and Thrips use artificial neural network while ABIS and Automated Insect Identification [22, 23] use support vector machine and Automated Real-Time Dynamic Identification of Flying and Resting Butterfly use random tree. Watson et al [18] had done work on training and testing of DAISY using live British Macrolepidoptera, which was light trapped in Treborth Botanical Garden, Gwynedd, UK. By using the right forewings of moths, the mean rank affinity of each species was computed as DAISY listed the possible species in order of decreasing likelihood when it makes identification. However, current DAISY only returns one species after identification and not a list of possible species. If DAISY was not able to identify a species, it returns "not classifiable" or "below activation threshold affinity". The local moths and the additional hindwings of moths remain untested. As such, DAISY was used in this study to identify the superfamily *Bombycoidea* moths which were located in Borneo, Malaysia. In this paper, automated identification of local moths using DAISY is done using:

- (1) The right forewing of moths
- (2) Additional hindwing for *Actias maenas* and *Actias selene* (these two species have significant hindwings. (Fig. 1)



Figure 1: On the left is an image of *Actias maenas* and on the right is an image of *Actias selene*. Both species have significant hindwings compared to other species.

## II. Methodology

### A. Collection of Images for Training Set

The moth images that were used as training images in this paper were scanned and digitized from the Moths of Borneo (Part 3) book [30].

HP Scanjet 5590 was used to digitize the images from the book while Adobe Photoshop CS5 was used to pre-process them. DAISY was installed on a HP computer with Intel core duo 2 to conduct this study.

The moth images consist of six families in the superfamily *Bombycoidea*. This superfamily was particularly chosen because they offer a variety of wing shapes that are significant. This is an important characteristic as the wing shapes were used by the identification system to help identify the species of moths. A total of 273 images consist of 20 images from family *Bombycidae* (15 species), 22 images from family *Eupterotidae* (15 species), 1 image from family *Brahmaeidae* (1 species), 38 images from family *Saturniidae* (22 species), 92 images from family *Lasiocampidae* (62 species) and 100 images from family *Sphingidae* (95 species) were obtained from the Moths of Borneo (Part 3) book [30]. These images, which are also known as training images, were used to train DAISY's artificial neural network (ANN). According to Hall and O'Neill [31], "DAISY recognizes objects by comparing their images to images that are already known by the system", therefore training set must be created in order for DAISY to identify an image.

### B. Processing of Moth Images for Training Set

Duplication of wings and standardization of the moth images were done using the Adobe Photoshop CS5. Duplication of wings ensures that the best side of the wing is duplicated so that the image is ready to be used as training image (Fig. 2). When building a training set, it is important to decide which side of the wing is to be used, and it should be standardized throughout the training sets to simplify the usage later on. DAISY does not limit the training to only one side of the wing. However, in this paper, only the right side of wing was chosen as the standard wing side for identification due to the time constrains and huge number of species used. *Actias maenas* and *Actias selene* contain long hindwings which is a significant characteristic in identifying them [30]. As such the hindwings of only these two species were used to train DAISY's ANN in this paper. While only the right side of wings were used, it is recommended to duplicate the better side of the wings to get nice, even and whole pictures of the moths to alleviate the future work for the left side of wings, and perhaps all the moths' hindwings too. All images were converted into *tif/tiff* files as DAISY only support images in these formats. The image size was also standardized to 600X480 pixels, as this size fits nicely on the DAISY windows.



Figure 2: Duplication of the wing of *Trabala ganesha* from *Lasiocampidae* family. This figure shows that the species above do not have a complete image, its right wing showing only half. As such, the left wing was duplicated, flipped vertically and attached to the right side to form an even and whole image.

### C. Building the Training Set

The training sets in DAISY are only as good as the images that were used to build them. To build the training set, the images were imported into DAISY. For every image, the wing shape was pinpointed by creating a polygonal region of interest (polyROI) around the image using polyROI function from DAISY's panel (Fig. 3). Each species needs at least 10 images with wings pinpointed to build the training set. The Build Tool function in DAISY was used to build the training sets of images uploaded into the system. DAISY creates a new folder (.ipm) for each of the training set built in DAISY. For example, the family *Saturniidae* will have its own training set named *Saturniidae.ipm*. This folder contains the key classifiers for moth identification. All the six families selected in this study have their own respective training sets (*familyname.ipm*). When combined, the six training sets form the training set for their superfamily *Bombycoidea* (*superBombycoidea.ipm*) (Fig. 4).

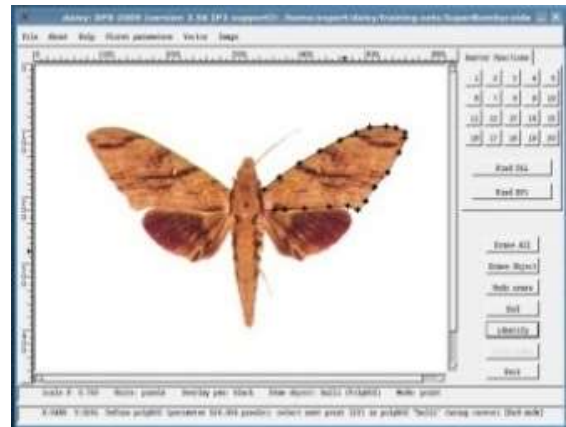


Figure 3: The wing shape of a moth was pinpointed using polyROI. Region of the wanted wing shape was highlighted by clicking around the edge of the wing. It is of extreme importance to ensure that the region was as close to the edge as possible because the accuracy of the identification will be affected by any additional part from the background which was also highlighted.

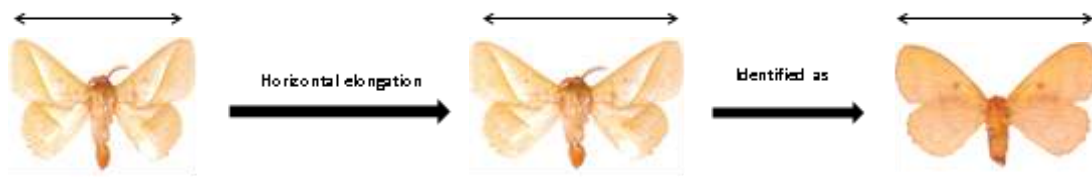


Figure 4: An example of horizontal elongation of *Trabala ganessa*. The width of the moth appeared to have been increased. On the far left is a training image of *Trabala ganessa* and in the middle is the image of *Trabala ganessa* after distortion (horizontal elongation). Notice that the width of the overall moth has change. This distorted image was identified as *Trabala rotundapex*.

#### D. Building the test data set

In this paper, the test data set consists of the following images.

- (i) **Training Images.** Images from Superfamily *Bombycoidea* which were used to build the training set (see above). It consists of 273 images from six families; *Bombycidae*, *Eupterotidae*, *Brahmaeidae*, *Saturniidae*, *Lasiocampidae* and *Sphingidae*.
- (ii) **Distorted Images.** The training images were distorted (enlargement, diminishment, horizontal & vertical elongation and horizontal & vertical compression). It consists of 273 images from six families; *Bombycidae*, *Eupterotidae*, *Brahmaeidae*, *Saturniidae*, *Lasiocampidae* and *Sphingidae*.
- (iii) **Web images of Superfamily *Bombycoidea*** consists of 246 images from six families; *Bombycidae*, *Eupterotidae*, *Brahmaeidae*, *Saturniidae*, *Lasiocampidae* and *Sphingidae*. The images were taken from the web.
- (iv) **Web images from other parts of Moths of Borneo.** It consists of 30 images from eight different family; *Notodontidae*, *Lymantriidae*, *Arctiidae*, *Drepaninae*, *Callidulidae*, *Geometridae*, *Notuidae* & *Noctuidae* [30].
- (v) **DAISY's built in training images.** There are two training sets available in DAISY which are UK Butterflies (207 images) and Belize (285 images). UK Butterflies is the entire UK Butterfly fauna (imaged from British Museum of Natural History collections) and Belize Sphingids dataset was collected at the Natural History Museum Field Station (Las Cuevas) [32].

#### E. Testing the accuracy of Moths identification in DAISY

To identify a moth, first the picture of that moth is imported and opened in DAISY. Next, polyROI function was selected and region of the wing was highlighted. Then

“identify” button was clicked on and a new window popped out for the result. If DAISY is able to identify the moth, the name of the species will appear as shown below (Fig. 5). Otherwise, it will return a wrong species name (wrong identification) or “not classifiable” or “below activation threshold affinity” instead.



Figure 5: An example of correct identification of *Macroglassum psedungus*.

#### F. Testing the accuracy of Moths identification in DAISY

Two types of test were conducted. In the first testing, all the images in the test data set (see above) was tested against the superfamily *Bombycoidea* training set and their respective family training set (Table 1 to 3 in the results section) and against the superfamily *Bombycoidea* training set only (Table 4, 5). Second testing was done using our training images (superfamily *Bombycoidea*) against DAISY's sample training sets (UK Butterflies & Belize) (Table 6, 7 in the results section). Precision (1), Recall (2) and F-measure (3) were reported accordingly.

$$\text{Precision} = \frac{R}{N} \quad (1)$$

$$\text{Recall} = \frac{R}{M} \quad (2)$$

$$\text{F-measure (F1)} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

### III. Results and Discussion

$R$  = Relevant images that are retrieved (correct)  
 $N$  = Retrieved images during identification (correct + wrong)  
 $M$  = Relevant images that are collected for each testing (total)

The results of the testing above are presented in Tables 1 to 7. In the tables, correct refers to correct identification, wrong refers to wrong identification and NotID refers to not classifiable/below activation threshold affinity. Precision & recall are only available for Table 1-3 and are only based of images that were tested against Superfamily *Bombycoidea* training set. Precision & recall are not available for Table 4-7 because all the images are not considered as relevant (these species are not trained in DAISY).

Table 1: Identification of superfamily *Bombycoidea* training images broken down to its six families (column) tested against Superfamily *Bombycoidea* training set and families' respective training set (row). Superfamily *Bombycoidea* = { $R = 273$ ,  $N = 273$  and  $M = 273$ . Precision = 1.00, Recall = 1.00,  $F1 = 1.00$ }. Respective families = { $R=273$ ,  $N= 273$  and  $M=273$ . Precision =1.00, Recall = 1.00,  $F1 =1.00$ }.

Training Set	Identification	Test Data of Training Images (families of superfamily <i>Bombycoidea</i> )							Accuracy of Retrieval (F1)
		Eupterotidae	Bombycidae	Brahmaeidae	Saturniidae	Lasiocampidae	Sphingidae	TOTAL	
Superfamily <i>Bombycoidea</i>	Correct	22	20	1	38	92	100	273	1
	Wrong	0	0	0	0	0	0	0	
	Not ID	0	0	0	0	0	0	0	
								273	1
Respective families	Correct	22	20	1	38	92	100	273	
	Wrong	0	0	0	0	0	0	0	
	Not ID	0	0	0	0	0	0	0	
								273	1

Table 2: Identification of **distorted training images** (enlargement, shrinking, horizontal & vertical compression and horizontal & vertical elongation) broken down to its six families (column) tested against Superfamily *Bombycoidea* training set and families' respective training set (row). Superfamily *Bombycoidea* = { $R = 257$ ,  $N = 270$  and  $M = 273$ . Precision = 0.95, Recall = 0.94,  $F1 = 0.94$ }. Respective families = { $R=263$ ,  $N= 270$  and  $M=273$ . Precision =0.97, Recall = 0.96,  $F1 =0.96$ }.

Training Set	Identification	Test Data of distorted training images (families of superfamily <i>Bombycoidea</i> )							Accuracy of Retrieval (F1)
		Eupterotidae	Bombycidae	Brahmaeidae	Saturniidae	Lasiocampidae	Sphingidae	TOTAL	
Superfamily <i>Bombycoidea</i>	Correct	20	18	1	38	87	93	257	0.94
	Wrong	1	1	0	0	5	6	13	
	Not ID	1	1	0	0	0	1	3	
								273	0.94
Respective families	Correct	22	20	1	37	90	93	263	
	Wrong	0	0	0	0	2	5	7	
	Not ID	0	0	0	1	0	2	3	
								273	0.96

Table 3: Identification of Web images of Superfamily *Bombycoidea* from six families (column) tested against Superfamily *Bombycoidea* training set and families' respective training set (row). Superfamily *Bombycoidea* = { $R = 97$ ,  $N = 217$  and  $M = 246$ . Precision = 0.45, Recall = 0.39,  $F1 = 0.42$ }. Respective families = { $R=95$ ,  $N= 214$  and  $M=246$ . Precision =0.44, Recall = 0.39,  $F1 =0.41$ }.

Training Set	Identification	Test Data of web images (families of Superfamily <i>Bombycoidea</i> )							Accuracy of Retrieval (F1)
		Eupterotidae	Bombycidae	Brahmaeidae	Saturniidae	Lasiocampidae	Sphingidae	TOTAL	
Superfamily <i>Bombycoidea</i>	Correct	2	2	2	10	5	76	97	0.42
	Wrong	3	2	0	8	25	82	120	
	Not ID	2	4	0	9	3	11	29	
								246	0.42
Respective families	Correct	1	3	2	13	7	69	95	
	Wrong	4	1	0	7	21	86	119	
	Not ID	2	4	0	7	5	14	32	
								246	0.41

Table 4: Identification of eight different families' images collected from Moths of Borneo (Part 4-16) tested against Superfamily *Bombycoidea* training set.



Training Set	Identification	Test Data of web images ( Families in Part 4-16 of Moths of Borneo)							
		Notodontidae	Lymantriidae	Arctiidae	Drepaninae	Callidulidae	Geometridae	Notuidae	Noctuidae
Superfamily	Correct	0	0	0	0	0	0	0	0
Bombycoidea	Wrong	3	5	5	0	1	2	3	4
	Not ID	1	0	0	2	0	0	3	1
<b>Total</b>		4	5	5	2	1	2	6	5

Table 5: Identification of DAISY’s sample training images (column) tested against Superfamily *Bombycoidea* training set (row).

Training Set	Identification	Test Data ( DAISY sample training set)		
		UK butterflies	Belize	Total
Superfamily Bombycoidea	Correct	0	0	0
	Wrong	114	185	299
	Not ID	171	21	192
	<b>Total</b>	285	206	<b>491</b>

Table 6: Identification of superfamily *Bombycoidea* training images broken down into details, to its six families (column) tested against UK Butterflies training set (row).

Training Set	Identification	Test Data consists of Training Images (families of superfamily <i>Bombycoidea</i> )						
		Eupterotidae	Bombycidae	Brahmaeidae	Saturniidae	Lasiocampidae	Sphingidae	TOTAL
UK butterflies	Correct	0	0	0	0	0	0	
	Wrong	4	2	0	7	17	6	36
	Not ID	18	18	1	31	75	94	237
							<b>273</b>	

Table 7: Identification of superfamily *Bombycoidea* training images broken down into details to its six families (column) tested against Belize training set (row).

Training Set	Identification	Test Data consists of Training Images (families of superfamily <i>Bombycoidea</i> )						
		Eupterotidae	Bombycidae	Brahmaeidae	Saturniidae	Lasiocampidae	Sphingidae	TOTAL
Belize	Correct	0	0	0	0	0	0	
	Wrong	2	1	0	3	13	27	46
	Not ID	20	19	1	35	79	73	227
							<b>273</b>	

In table 1, the training images, when identified against superfamily *Bombycoidea* training set and also against their respective family training sets, returned 100% accuracy (F1= 1). It shows that as long as the image was used to train DAISY, it will be able to recognize the same image if it was reused as test data. 1 for F1 indicates that DAISY was able to identify all the species correctly in term of exactness (precision) and completeness (recall).

In Table 2, distorted training images returns a higher accuracy when identified against their respective family training sets (F1=0.96) compared to superfamily *Bombycoidea* training (F1=0.94) (See above for distortion of images). At species level, the moths in the same family tend to be very similar to each other. Some are also similar across the superfamily [30]. Normally, in order to differentiate the species, either the genital area or wing venation will be used [30]. As such, when the size or shape was altered during distortion, one species may be confused as another species. From the distortions, it shows that the size and shape of moth image affects the identification whenever DAISY was used. As the superfamily *Bombycoidea* training set has a lot more species in it than the individual family training sets, the probability of confusing one moth to another was increased. This would explain why F1 value dropped by 0.06 to 0.94 for this table compared to Table 1.

In Table 3, however, identification of web images when tested against the superfamily *Bombycoidea* training set have slightly higher accuracy (F1= 0.42) compared to when tested against their respective family training sets (F1=0.41). Only test images; 2 live moth images from this family was taken from Baron G. [33] of the *Brahmaeidae* family were all identified correctly when identified against both superfamily *Bombycoidea* and its own training set. This could be because the image of the *Brahmaeidae* family has a very clear wing shape. If the shady area around the wings were mistaken as wing shape, it has higher chances of the species being wrongly identified or NotID by DAISY. The accuracy of pinpointing the wing shapes also affects the identification. As almost half of the web images were live moth images, it was harder to pinpoint the wing shapes accurately. This is because it was hard to differentiate the forewings from the hindwings during moths’ resting position in nature. Generally, the F1 values in Table 3 are low (0.42 and 0.41) compared to Table 1 (F1=1). This could be due to the differences in wing positions. As stated earlier and also in previous works [8], identification of DAISY will only be as good as the images that were provided to it as training images. Therefore, the higher the quality of images and the more images of a same species are used to train DAISY, the better the identification will be. The more variety of moth’s position was provided to DAISY, the more

accurate the identification will be. This is especially true for species with high similarity in their wing shapes. Thus, the F1 value can be increased if images of moth with different wing positions were added into the training set.

In Table 4, images from other volumes of Moth of Borneo (Part 4, 5, 7, 8, 9, 12, 15 & 16) were tested against superfamily Bombycoidea training set. Out of the 30 images that were identified, seven were identified as NotID, 23 returned wrong identification and none were identified correctly. None of these images are considered as “relevant images that are retrieved”. As a result, precision and recall value is not available for this table. It is quite odd that 76% wrong identification (23 images) were returned when they should be returned as NotID because DAISY was not trained to identify these families. This shows that DAISY tends to identify images to the nearest possible species (Fig. 6) according to the wing shape.

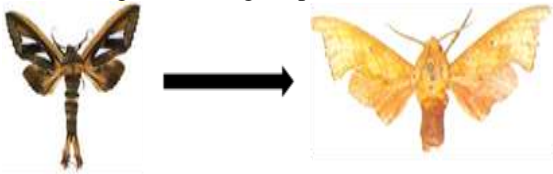


Figure 6: An example of wrong identification in Table 4 (other parts of Moths Of Borneo). DAISY wrongly identified *Tarsolepis sommeri* (left) as *Smerinthulus quadripunctatus*(right) because both these species have similar scaled wings.

Table 5 shows the identification of two DAISY’s sample training images tested against our superfamily Bombycoidea training set. In this test, DAISY returned 171 NotID for UK Butterflies and 21 NotID for Belize images. 114 images from UK Butterflies and 185 images from Belize were identified wrongly. Out of the 114 wrong identifications for UK Butterflies, most were identified as Lasiocampidae family (64 images), followed by 35 images as Eupterotidae family, 9 as Bombycidae family, 5 as Saturniidae family and 1 as Sphingidae family. Out of the 64 images of the Lasiocampidae family, 25 were identified as *Lebeda cognate* species. This high number of wrong identification was caused by high similarity of wing shapes of the UK Butterflies to wing shapes of Lasiocampidae family moths. Out of the 185 wrong identifications from Belize, 161



Figure 7: An example of two wrong identifications from Table 5 (Belize). On the far left is *Protambulyx strigilis* followed by *Pachyliodes resumens*. Both were wrongly identified as *Ambulyx obliterata*. Notice that the wing shapes of both species are very similar to *Ambulyx obliterata* from *Sphingidae* family.

images were identified as Sphingidae family, 12 images as Lasiocampidae family, 9 as Eupterotidae family and 3 as Bombycidae family. Out of the 161 images, 22 were identified as *Ambulyx obliterata* (Fig. 7). All the moths in Belize are of Sphingidae family too. However, the species from the Sphingidae family in Belize are different from the species from the Sphingidae family in local Moths of Borneo. As such they are not trained in superfamily

Bombycoidea training set. This is the reason why 161 of 185 images of Belize were identified wrongly as other species in Sphingidae family. This also proves that, Moths species in a particular family will have very similar wing shapes, although they belong to different species (Holloway J.D., 1998).

As for Table 6, identification of superfamily Bombycoidea training images which were tested against UK butterflies training set is displayed. Out of the 273 images, DAISY returned 36 wrong identifications (whereby highest of 9 images were identified as *Erebia epiphron* species) and 237 NotID. Table 7 displays identification of superfamily Bombycoidea training images tested against Belize training set, where 227 NotID were returned with the remaining 46 images as wrong identifications (highest of 9 images were identified as *Callionima parce* species). Again, the similarity of wing shapes caused the retrieval of wrong identification. By comparing the results in Table 5, 6 and 7, we can see that DAISY’s sample training sets (UK Butterflies and Belize) performed better than our superfamily Bombycoidea training set. This is because, when UK Butterflies and Belize were tested against superfamily Bombycoidea training set, it returned with 114 (40%) and 185 (89.8%) wrong identifications respectively. While when the superfamily Bombycoidea training images were tested against UK Butterflies and Belize training sets, it returned with only 36 (13.1%) and 46 (16.8%) wrong identification respectively. As UK Butterflies training sets used images with similar orientation (wings evenly spread out), this would explain why UK Butterflies returned with less wrong identification compared to Belize which used live moth images when used as training set as well as test data against our superfamily *Bombycoidea*. Another reason of why Belize returned with more wrong identification is because images of Belize are of moths from *Sphingidae* family, which is one of the families in superfamily *Bombycoidea*. On the other hand, UK Butterflies are of butterflies which are less similar (as in the wing shapes) to our superfamily *Bombycoidea* moths. The summary of the tests conducted above is presented in Table 8. Precision of the overall moths identified is 0.58 while the recall gives 0.79 and F1 value is 0.67 (R=627, N=1082, M=792).

As shown in Table 8 the precision of the overall moths identified (Table 1 to 3) is 0.83 while the recall gives 0.79 and F1 value is 0.81 (maximum 1.00). This shows that in this study, DAISY identified the Moths of Borneo (Part 3) more likely in terms of exactness (precision) and less likely as completeness (recall).

Table 8: Summary of identification of test data using DAISY  
 Superfamily Bombycoidea Training Set

Test Data Images	Returned results						
	Correct	Wrong	NotID	Total	Precision	Recall	F1
Moths of Borneo - Part3 (Table 1)	273	-	-	273	1.00	1.00	1.00
Distorted Moths of Borneo - Part 3 (Table 2)	257	13	3	273	0.95	0.94	0.94
Web (Table 3)	97	120	29	246	0.45	0.39	0.42
8 other families from Moths of Borneo - Part 4 to 16 (Table 4)	-	23	7	30	-	-	-
UK butterflies (Table 5)	-	114	171	285	-	-	-
Belize (Table 5)	-	185	21	206	-	-	-
<b>Total</b>	<b>627</b>	<b>455</b>	<b>231</b>	<b>1313</b>			

DAISY's Sample Training Set

Test Data Images	Returned results						
	Correct	Wrong	NotID	Total	Precision	Recall	F1
Superfamily Bombycoidea (Table 6)	-	36	237	273	-	-	-
Superfamily Bombycoidea (Table 7)	-	46	227	273	-	-	-
<b>Total</b>	<b>0</b>	<b>82</b>	<b>464</b>	<b>546</b>			

The value might increase if DAISY is trained further by adding the local live moths as well as images of 8 different families in Moths of Borneo (Part 4-16).

## IV. Conclusion

The overall identification of Moths in this paper gave a fairly accurate retrieval, with F1= 0.81. However, using wing shapes as the only feature in identification might not be sufficient, especially for species identification. The wing pattern and genitalia area are two other important characteristics which can be used for moth identification and this will be conducted in the future work. We believe that the routine identification burden of taxonomists can be reduced using image recognitions systems, such as DAISY which is used in this paper.

## Acknowledgement

We would like to thank Dato' Henry Barlow for contributing his expertise in moths and allowing us to use the moths of Borneo for this study. This study was supported by University of Malaya's Postgraduate Research Fund (PG092-2013B) and the University of Malaya's Research Grant (FP032-2014B) to the fourth author.

## References

- [1] SECRETARIAT, Global Biodiversity Outlook 1, Montreal, 2001.
- [2] A. Bortolus, "Error Cascades in the Biological Sciences: The Unwanted Consequences of Using Bad Taxonomy in Ecology"

- AMBIO: A Journal of the Human Environment, vol. 37, pp. 114-118, March 2008.
- [3] P. J. D. Weeks and K. J. Gaston, "Image analysis, neural networks, and the taxonomic impediment to biodiversity studies" *Biodiversity & Conservation*, vol. 6, pp. 263-274, February 1997.
- [4] H. C. J. Godfray, "Challenge for taxonomy: the discipline will have to reinvent itself if it is to survive and flourish" *Nature*, vol. 417, pp. 17-19, May 2002.
- [5] N. MacLeod, M. Benfield and P. Culverhouse, "Time to automate identification" *Nature*, vol. 467, pp. 154-155, September 2010.
- [6] P. D. N. Hebert, A. Cywinska, S. L. Ball and J. R. deWaard, "Biological identifications through DNA barcodes." *Proc Biol Sci.*, vol. 270, pp. 313-21, February 2003.
- [7] P. D. N. Hebert, T. R. Gregory, "The promise of DNA barcoding for taxonomy" *Syst. Biol.* vol. 54, pp. 852-859, 2005.
- [8] K. J. Gaston and M. A. O'Neill, "Automated species identification: why not?" *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* vol. 359, pp. 655-667, April 2004.
- [9] P. J. D. Weeks, M. A. O'Neill, K. J. Gaston and I. D. Gauld, "Automating insect identification: exploring the limitations of a prototype system" *Journal of Applied Entomology*, vol. 123, pp. 1-8, January 1999.
- [10] L. Feng, B. Bhanu, J. Heraty, "A software system for automated identification and retrieval of moth images based on wing attributes" *Pattern Recognition*, vol. 51, pp. 225-241, March 2016.
- [11] J. L. Salle, Q. Wheeler, P. Jackway, S. Winterton, D. Hobern, D. Lovell, "Accelerating taxonomic discovery through automated extraction." *Zootaxa*, vol. 2217, pp.43-55, September 2009.
- [12] H. M. Pereira, S. Ferrier, M. Walters, G. N. Geller, R. H. G. Jongman, R. J. Scholes, M. W. Bruford, N. Brummitt, S. H. M. Butchart, A. C. Cardoso, N. C. Coops, E. Dulloo, D. P. Faith, J. Freyhof, R. D. Gregory, C. Heip, R. Höft, G. Hurr, W. Jetz, D. S. Karp, M. A. McGeoch, D. Obura, Y. Onoda, N. Pettorelli, B.

- Reyers, R. Sayre, J. P. W. Scharlemann, S. N. Stuart, E. Turak, M. Walpole, M. Wegmann " Essential biodiversity variables", *Science*, vol.339, pp. 277-278, January 2013.
- [13] S. J. Bacon, S. Bacher, A. Aebi, "Gaps in border controls are related to quarantine alien insect invasions in Europe", *PLoS One*, vol. 7, e47689, October 2012.
- [14] S. Kumschick, S. Bacher, W. Dawson, J. Heikkilä, "A conceptual framework for prioritization of invasive alien species for management according to their impact" *Neo. Biota*. vol.15, pp. 69–100, December 2012.
- [15] C. Wen, D. E. Guyer, W. Li, "Local feature-based identification and classification for orchard insects" *Biosyst. Eng.* vol.104. pp.299–307, November 2009.
- [16] T. M. Franco, D. Wittmann, M. Drauschke, S. Müller, V. Steinhage, M. A. F. Bezerra-Laure, D. D. Jong, L. S. Goncalves, "Identification of Africanized honey bees through wing morphometrics: a fast and efficient procedure" *Apido-logie*, vol. 39, pp. 488–494, September 2008.
- [17] U. Jean, D. Arne, F. Simon, G. Malcolm, "A stilet project case study: the evaluation of a computer-based visual key for fossil identification" *Assoc. Learn. Technol. J.* vol.4, pp. 40–47, 1996.
- [18] M. A. O'Neill, I. D. Gauld, K. J. Gaston, P. Weeks, "Daisy: an automated invertebrate identification system using holistic vision techniques", In *Proc. Inaugural Meeting BioNET-INTERNATIONAL Group for Computer-Aided Taxonomy (BIGCAT)* (ed. D. Chesmore, L. Yorke, P. Bridge & S. Gallagher), pp. 13–22, 2000. Egham: BioNET-INTERNATIONAL Technical Secretariat.
- [19] S. Schroder, D. Wittmann, W. Drescher, V. Roth, V. Steinhage, A. B. Cremers, "The new key to bees: automated identification by image analysis of wings. Pollinating bees—the Conservation Link Between Agriculture and Nature" Ministry of Environment, Brasília, pp. 209-216.
- [20] L. Feng and B. Bhanu, "Automated identification and retrieval of moth images with semantically related visual attributes on the wings." in *20th IEEE International Conference on Image Processing (ICIP)*, pp. 2577 – 2581, September 2013.
- [21] M. Mayo and A. T. Watson, "Automatic species identification of live moths" *KNOWL-BASED. SYST.* vol.20, pp. 195-202, March 2007.
- [22] 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, T. G. Dietterich, "Automated insect identification through concatenated histograms of local appearance features: feature vector generation and region detection for deformable objects" *MACH. VISION. APPL.* vol. 19, pp. 105-123, March 2008.
- [23] Z. Le-Qing and Z. Zhen, "Automatic insect classification based on local mean colour feature and Supported Vector Machines" *ORIENT. INSECTS.* vol. 46, pp. 260-269, July 2012.
- [24] K. S. Loke, S. Egerton, D. Cristofaro, S. Clementson, "Automated real time dynamic identification of flying and resting butterfly species in the natural environment" in *International Conference on Environment Science and Engineering (ICESE 2011)*, pp. 179 – 183, April 2011.
- [25] K. N. Russell, M.T. Do, N. I. Platnick, "Introducing SPIDA-Web: An Automated Identification System for Biological Species" *Proceedings of Taxonomic Database Working Group Annual Meeting, St Petersburg, Russia, 11–18 September 2005*.
- [26] K. N. Russell, M.T. Do, N. I. Platnick, "Introducing Spida-Web: Wavelets, Neural Networks and Internet Accessibility in an Image-Based Automated Identification System." *Automated Taxon Identification in Systematics: Theory, Approaches and Applications*, pp. 131-152, 2008.
- [27] T. Arbuckle, S. Schröder, V. Steinhage, D. Wittmann, "Biodiversity Informatics in Action: Identification and Monitoring of Bee Species using ABIS" In: *Proc. 15th Int. Symp. Informatics for Environmental Protection, Zurich*, ISBN 3-89518-370-9, 2001.
- [28] M. Pajak, "Identification of British *Bombus* and *Megabombus* using DAISY", B.A. 3rd Year Honours Project, University of Oxford, United Kingdom, 2000.
- [29] A. T. WATSON, M. A. O'NEILL, I. Kitching, "Automated identification of live moths (Macrolepidoptera) using digital automated identification System (DAISY)." *Systematics and Biodiversity*, vol. 1, pp. 287-300, September 2003.
- [30] J. D. HOLLOWAY, "THE MOTHS OF BORNEO" PART 3 SUPERFAMILY BOMBYCOIDEA: FAMILIES LASIOCAMPIDAE, EUPTEROTIDAE, BOMBYCIDAE, BRAHMAEIDAE, SATURNIIDAE, SPHINGIDAE, Southdene, 1998.
- [31] Z.E. Hall, and M. A. O'Neill (2004). *Daisy User Manual. The DAISY Automated Insect Identification Project*, Tumbling Dice Ltd.
- [32] M. A. O'NEILL, "DAISY: A Practical Tool for Automated Species Identification." Retrieved June, 2010, from <http://www.tumblingdice.co.uk/daisy/doverview.pdf>