Using Partial Least Squares in Butterfly Species Identification

Alexandre Silva, Sincler Meireles, Samira Silva
Universidade do Estado de Minas Gerais (UEMG), Divinópolis, Brazil
alexandre.silva10@hotmail.com, sincler.meireles@uemg.br, samira.silva@uemg.br

***Resumo*—Butterflies are important insects in nature, and along with moths constitute the Lepidoptera order. At the global level, the number of existing butterfly species is approximately 16,000. Therefore, the identification of their species in images by humans consists in a laborious task. In this paper, we propose a novel approach to recognize butterfly species in images by combining handcrafted descriptors and the Partial Last Squares (PLS) algorithm. A set of PLS models are trained using an one-against-all protocol. The test phase consists in presenting images to all classifiers and the one which provides the highest response value contains in the positive set the predicted class. The performance of the proposed approach is evaluated on the Leeds Butterfly dataset. Experiments were conducted using HOG and LBP descriptors, separately and combined. The approach using HOG singly reported an accuracy rate of 68.72%, while using only LBP resulted in an accuracy rate of 77.33%. Combining both descriptors this value changes to 76.27%. The proposed approach achieves the best results in all three versions when compared to state-of-the-art approaches. Experiments have shown that describing images with LBP provides the highest accuracy values since it extracts texture information, what is an important characteristic to distinguish butterflies. However, information of color and shape, added by HOG, appears to make different species confused.**

I. Introduction

Butterflies are important insects in nature, and along with moths constitute the Lepidoptera order. The Lepidoptera are found in virtually all regions of the world, but especially in tropical countries. In Brazil, for example, there are more than 3,500 species of butterflies registered [1]. At the global level, this number increases to about 16,000 species [2], resulting in numerous variations such as different coloring, textures and patterns, which attracts the attention not only of biologists but also of the general public around the world.

Butterflies are species of living beings and therefore can be classified taxonomically in descending order into the following taxons: kingdom, phylum, class, order, family, genus and species [3]. Butterflies belong to the *Animalia* kingdom, to the *Arthropoda* phylum, to the *Insecta* class, and to the *Lepidoptera* order. They differ in terms of family, gender, and species. Therefore, when noticing an unknown butterfly, there is much work to be done to determine its classification, which can be an extensive and laborious task.

In order to identify species in a traditional manner, researchers usually begin by predicting the most general taxon, the family. Then, they seek for determining their genus, to finally predict the most specific taxon, the species. Each of the taxons is defined according to characteristics that living beings have in common [4]. Therefore, predicting taxons at higher levels, such as kingdom, phylum, and class, for a living being, is easier than specifying its family, genus, and species.

Strategies for identifying butterfly species in images manually may sometimes not be efficient [5]. The first reason is that they require a lot of time, since it is necessary a comparison between the image (or its characteristics) containing an unknown butterfly with an arsenal of registered characteristics which are associated to different species. Another reason is the fact that identifying butterfly species manually can lead to errors due to human limitations [6]. The professionals who perform these comparisons can often confuse color-related characteristics if they have some degree of color blindness, for example. Moreover, the patterns on butterfly wings can also be confusing because they resemble for different species.

In the last few years, automatic identification of butterfly species have been the aim of several works [7]–[9]. This task, performed by an automatic system, comprises the analysis of an image portraying a butterfly in order to determine its species. These systems are based on Machine Learning [10]. In automatic butterfly identification, a system are supposed to learn characteristics which define each species. For this purpose, it looks for finding patterns in image samples containing a butterfly. After the learning process, the system can identify an image containing an unknown species of butterfly basing on the learned characteristics for each species.

There are several advantages in performing the identification of butterfly species using an automatic system [11]. The computer can learn a number of different species far superior to what is possible to a human. In addition, a system with this purpose could recognize several patterns that would go unnoticed by professionals, and are important to describe a specific species. Finally, an efficient system in this sense can provide more accurate predictions in a shorter time when compared to the manual identification.

In view of this, this work aims at performing automatic identification of butterfly species in images. For this purpose, features from a butterfly dataset are extracted by using the Local Binary Patterns (LBP) [12] and Histogram of Oriented Gradients (HOG) [13] descriptors. Then, Partial Least Squares (PLS) [14] models are trained basing on features of half of images, using the one-against-all protocol. Finally, the remainder of images are presented to trained models in order to determine their species.

II. Related Work

A vast number of challenges are posed by the automatic identification of butterfly species. Many of them are associated to the similarity of design patterns on butterfly wings of different species. Figure 1 depicts two butterfly species: *Heliconius Wallacei* (left) e *Heliconius Sara* (right).

Figura 1. *Heliconius Wallacei* (left) and *Heliconius Sara* (right) species [15].

In the figure above, we can note a huge similarity in the patterns of texture, colors and shape of wings presented by the two species. The blue color is predominant in the central region of butterfly's wings and two white spots stand out in regions almost equivalent in both butterflies.

The Figure 2 below also portrays two different species of butterflies: *Limenitis Archippus* (left) and *Danaus Plexippus* (right). The two species also have several common characteristics in their color and texture patterns. However, there is a small difference in behavior between them that causes one to cause problems and the other does not: the species *Danaus Plexippus* feeds on milkweed plants, which makes its body highly toxic, leading to the poisoning of its predators, like birds. For this reason, in certain places the presence of this type of species may not be so interesting.

Figura 2. *Limenitis Archippus* (left) and *Danaus Plexippus* (right) species [15].

Challenges, such as those illustrated in the previous examples, make researchers feel increasingly motivated to develop methods for butterflies identification which address the difficulties imposed by the similarities between species. Despite the fact that more visible characteristics such as color, texture and wing shape are easier to observe, in very similar species (such as those mentioned above) it may be necessary to examine the external structural characteristics of the genitals in order to distinguish them [16].

In recent years, DNA has been used as one of the characteristics that distinguishes species not only of butterflies, but also of different types of living beings [17]. However, extracting information about species using DNA is not a simple task. The use of an automatic butterfly identification system is not only a faster strategy, but also a cheaper one for this task. Despite this, there are currently few intelligent systems for identifying butterflies that are actually used.

The work in [8] focus on the development of a content-based image recovery system aiming at classifying butterfly images according to characteristics such as: texture, color and shape. Concentrating in only four species of lepidopteras that are considered as rice pests, the work in [18] proposes an automatic identification system able to recognize the *Chilo Suppressalis*, *Sesamia Inferens*, *Cnaphalocrocis Medinalis* e *Pararaguttata Bremeret* species. They extract the same information of the work in [8] but the classification stage is performed using the SVM (*Support Vector Machine*) algorithm.

Focusing on correctly classifying recurrent insects in orchard regions, the approach described in [19] uses different local descriptors that are combined with six classifiers to carry out the recognition of its species. With the goal of recognizing butterfly models in images with disorder in the background, the work in [20] employs a model that is based on characteristics associated with the location of butterflies.

Finally, more recent approaches in literature address the recognition of butterfly species in images problem using Neural Networks. The work presented in [21] explores characteristics of five textures (homogeneity, contrast, energy, correlation and entropy) and three colors (the average gray level of layers R, G and B) in order to identify and classify butterfly images by using a Artificial Neural Network, a model based on the biological neural system. Considering a wider scope, the approach in [22] aims to automatically recognize species of fish, plants and butterflies in images.The extraction of features has the objective of obtaining the geometry, morphology and texture of images. Then, an Artificial Neural Network is used as a pattern recognition method.

All the above methods propose strategies for the automatic recognition of lepidoptera species that are not necessarily the existing species in Brazil. And, more specifically, they do not consider the identification of butterfly species that cause most of problems in several areas, such as agriculture, for example. Therefore, it is necessary to carry out more researches and propose new methodologies, so that species that exist in Brazil, especially those that result in pests, can be identified more precisely and in an adequate time.

III. Proposed Approach

In this section, a novel methodology to perform butterfly species identification is detailed. The sequence of stages which are comprised by the proposed approach is described as follows.

A. Cropping and Normalization

The first step in the proposed approach is cropping and normalizing images. To crop butterfly images we use a mask image which suggests whether a pixel from the original image belongs to the foreground (butterfly area) or to the background. Cropping images is necessary since their background may hinder the identification process. In some images, the background and the region occupied by the butterfly may not be distinguished and the descriptor may extract features also from the background. Therefore, in this step we somehow remove the background influence. Then, the normalization task is executed. This task consists in making all the images have the same resolution, and it is important in order to guarantee that the feature extraction stage performs correctly.

B. Feature Extraction

Feature extraction is the process of representing images using a vector which describes their attributes [23]. The method to represent an image varies according to the information you want to capture. Considering the characteristics of butterflies, this work concentrates in the extraction from images of three characteristics: color, shape and texture. Therefore, two descriptors which handle this information are employed: the Histogram of Oriented Gradients (HOG) and the Local Binary Patterns (LBP).

The HOG descriptor [13] extracts color and shape information. The main idea of this visual descriptor is that the appearance and shape of objects in images can be represented by distributing pixel intensity gradients or based on edges directions. The descriptor generation process consists of four phases: computing the gradient in each pixel, organizing sets of pixels in cells, organizing cells into blocks and obtaining the descriptor. A histogram is generated for each block and their concatenation results in the HOG descriptor. HOG is computed for the three image channels, R (Red), G (Green) and B (Blue), and the resulting three descriptors are concatenated to form a single vector.

The second image descriptor employed, which handles texture information, is the LBP [12]. Based on a 3x3 neighborhood around each pixel, this approach compares the central pixel with its neighbors and assigns 1 to them if they have a value greater than or equal to the central pixel and 0 otherwise. An 8-bit vector is then generated for each pixel, describing it. Each of these vectors are used to generate a histogram that will represent the image texture.

In this work, each cropped and normalized image is described by both descriptors, HOG and LBP. Figure 3 summarizes the cropping, normalization and feature extraction processes.

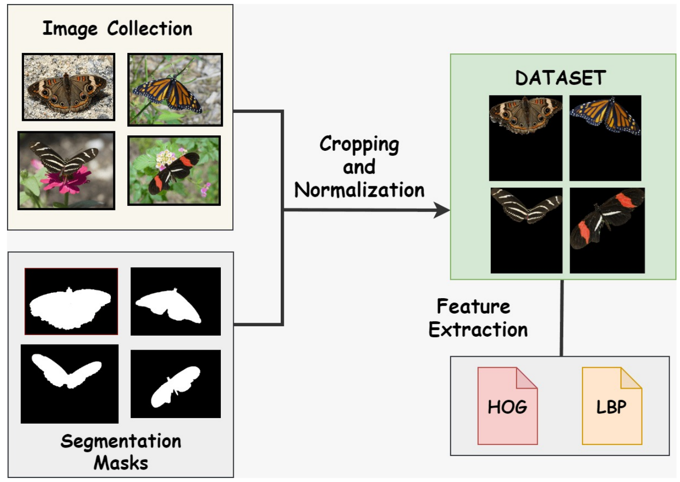


Figura 3. Cropping, Normalization and Feature Extraction Processes.

C. Classification

The classification stage consists in predicting species of butterfly images. For each species, sample images are divided into training (50%) and test (50%) sets. Training and test phases, using the features of both sets, respectively, are presented as follows.

1) Training Phase: In this work, the training phase considers the one-against-all (OAA) protocol. Applying this protocol, a classifier is trained for each species using as positive class a determined species and the remainder as negative. The number of trained models is equal to the number of species/classes in the dataset. The classifier considered in this work is the Partial Least Squares(PLS) [14] algorithm.

PLS consists of a regression that models the relation between characteristics through latent variables. It has been used in the literature in several approaches, but its use for the recognition of butterfly species has not yet been tested. It was chosen for this task since it can handle high dimension feature vectors well. In addition, it can also provide good results in classification when there are few samples for training.

The use of PLS is given as follows: with the features extracted from training set (containing images and their labels), a PLS model is created for each class by considering in training stage features from images of each species against the remaining. Thus, we have a PLS model specialized in each species.

2) Test Phase: This stage aims at predicting the labels (species) of each butterfly image in test set. Therefore, features extracted from each image of this set are presented to each classifier which returns a response value. This value denotes how similar to the positive images of each classifier is the test image. The classifier whose response value is the highest was trained using as positive set samples of the most likely class. Thus, this is the predicted class/species to that image. We perform the same to all images in test set. The proposed approach is summarized in Figure 4 as follows.

IV. Experimental Results

This section describes the experimental results regarding the proposed approach to identify butterfly species in images. We detail the experiment configuration, comprising a description of the dataset employed and the descriptor parameters values, a discussion about the main results and a comparison between the proposed approach and literature methods.

A. Experiment Configuration

As aforementioned, in our experiments 50% of samples of each species are used to training and 50% to test. This distribution was chosen since they have proved to be a balanced partitioning. Information about the dataset and descriptor parameters chosen are presented as follows.

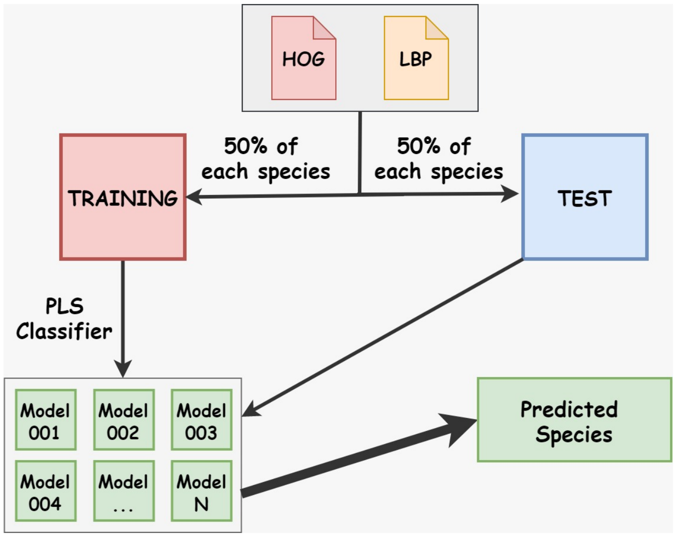


Figura 4. A summarization of the proposed approach considering all the stages.

1) Dataset: Images of Leeds Butterfly dataset [24] are used in order to perform experiments with the proposed approach. This dataset comprises 832 images from ten butterfly species, ranging from 55 to 100 images per species. Besides, it provides masks which delimitate butterfly region in images. Figure 5 retracts a butterfly image (left) and its mask (right).

Figura 5. *Danaus Plexippus* (left) species and its segmentation mask (right) [24].

In Table I, we present Leeds Butterfly Dataset species and their corresponding numbers used to report the results in Section IV-B.

Number	Species
001	*Danaus plexippus*
002	*Heliconius charitonius*
003	*Heliconius erato*
004	*Junonia coenia*
005	*Lycaena phlaeas*
006	*Nymphalis antiopa*
007	*Papilio cresphontes*
008	*Pieris rapae*
009	*Vanessa atalanta*
010	*Vanessa cardui*

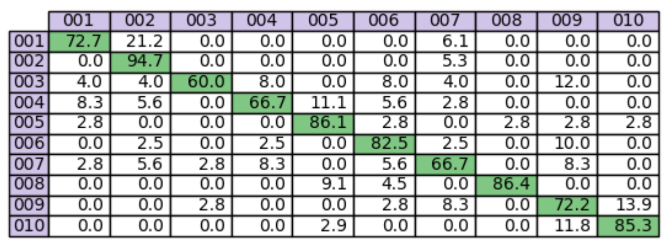
Tabela I
SPECIES OF LEEDS BUTTERFLY DATASET AND THEIR CORRESPONDING NUMBERS IN THE EXPERIMENTS.

2) Descriptor Parameters: The parameters of experiments using the LBP descriptor were defined according to the images resolution, which was set to 640x480 pixels. In our experiments, the highest performance occurred using radius of 24 pixels and 72 neighboring pixels of sampling. Some tests were performed varying the parameters in order to validate the choice. However, it was found that these parameters generate a vector that best describes the texture in LBP.

Similarly, tests were carried out in order to measure the impact in prediction for each parameter of HOG descriptor. For the number of orientations, experiments have shown that 8 was an adequate value, since its variation reduces the accuracy value. However, regarding the cells size, it was noticed that when it increases the accuracy rate also increases. This trend follows until the dimension of 20x20 pixels, in which the highest level of accuracy rate was reached. Finally, for the dimension of the blocks, it was noted that best performance is achieved when its value is 1x1.

B. Results and Discussion

In this work, three different experiments were performed: using LBP and HOG separately, and combining them into a single vector by concatenating them. Next, we report the achieved results using confusion matrices. Rows represent the correct class and columns represent the prediction by the proposed approach. Figure 6 contains the confusion matrix of experiments using LBP to describe butterfly images. In other words, only texture information is considered. The value of average accuracy is 77.33% for this matrix.



	001	002	003	004	005	006	007	008	009	010
001	72.7	21.2	0.0	0.0	0.0	0.0	6.1	0.0	0.0	0.0
002	0.0	94.7	0.0	0.0	0.0	0.0	5.3	0.0	0.0	0.0
003	4.0	4.0	60.0	8.0	0.0	8.0	4.0	0.0	12.0	0.0
004	8.3	5.6	0.0	66.7	11.1	5.6	2.8	0.0	0.0	0.0
005	2.8	0.0	0.0	0.0	86.1	2.8	0.0	2.8	2.8	2.8
006	0.0	2.5	0.0	2.5	0.0	82.5	2.5	0.0	10.0	0.0
007	2.8	5.6	2.8	8.3	0.0	5.6	66.7	0.0	8.3	0.0
008	0.0	0.0	0.0	0.0	9.1	4.5	0.0	86.4	0.0	0.0
009	0.0	0.0	2.8	0.0	0.0	2.8	8.3	0.0	72.2	13.9
010	0.0	0.0	0.0	0.0	2.9	0.0	0.0	0.0	11.8	85.3

Figura 6. Confusion matrix using LBP descriptor. Average accuracy: 77.33%.

Figure 7 portrays the confusion matrix of experiments using HOG to describe butterfly images. In other words, color and shape information is considered in this experiment. The value of average accuracy is 68.72% for this matrix.

Analyzing the results provided in Figures 6 and 7 it is possible to notice that the LBP descriptor performed approximately 10% better than HOG, in terms of average accuracy, when evaluating the use of the descriptors separately. With this information, we can conclude that, considering the extracted characteristics, texture seems to be the feature that best distinguishes butterfly species. In same way, the patterns found in butterflies are unique but they have similar colors and shapes, what makes unfeasible the use of descriptors which captures this information, such as HOG.

	001	002	003	004	005	006	007	008	009	010
001	90.9	0.0	3.0	0.0	0.0	0.0	3.0	3.0	0.0	0.0
002	0.0	76.3	15.8	2.6	0.0	2.6	0.0	0.0	0.0	2.6
003	4.0	16.0	68.0	8.0	0.0	0.0	0.0	0.0	4.0	0.0
004	2.8	0.0	0.0	58.3	8.3	22.2	0.0	0.0	0.0	8.3
005	2.8	0.0	2.8	2.8	69.4	2.8	0.0	8.3	5.6	5.6
006	0.0	5.0	0.0	2.5	2.5	80.0	0.0	0.0	10.0	0.0
007	2.8	0.0	8.3	0.0	5.6	5.6	77.8	0.0	0.0	0.0
008	4.5	0.0	0.0	0.0	4.5	4.5	9.1	72.7	4.5	0.0
009	2.8	0.0	2.8	2.8	0.0	8.3	2.8	0.0	55.6	25.0
010	0.0	2.9	5.9	11.8	14.7	5.9	0.0	2.9	17.6	38.2

Figura 7. Confusion matrix using HOG descriptor. Average accuracy: 68.72%.

For example, in the confusion matrix of Figure 7 we can note that the 009 (*Vanessa Atalanta*) and 010 (*Vanessa Cardui*) species are confused. A quick search by images from these species on the Internet shows that they have characteristics in common, mainly in color and shapes. Therefore, a color or shape descriptor may not be the best choice to represent them. Figure 8 presents samples of images from boths species extracted from Leeds Butterfly Dataset [24]. When we look at the confusion matrix in Figure 6, it is possible to notice that texture information extracted by LBP seems to improve the accuracy of these species.

Figura 8. *Vanessa Atalanta* (left) and *Vanessa Cardui* (right) species [24].

Figure 9 depicts the confusion matrix using a concatenation of LBP and HOG to describe butterfly images. In other words, texture, color and shape information is considered in the experiments. The value of average accuracy is 76.27% for this matrix.

	001	002	003	004	005	006	007	008	009	010
001	90.9	0.0	3.0	0.0	0.0	0.0	6.1	0.0	0.0	0.0
002	0.0	81.6	10.5	2.6	0.0	2.6	0.0	0.0	0.0	2.6
003	4.0	8.0	68.0	8.0	0.0	8.0	0.0	0.0	4.0	0.0
004	2.8	0.0	0.0	61.1	8.3	19.4	5.6	0.0	0.0	2.8
005	2.8	0.0	2.8	8.3	75.0	0.0	0.0	0.0	5.6	5.6
006	0.0	2.5	0.0	2.5	2.5	80.0	0.0	0.0	12.5	0.0
007	2.8	0.0	5.6	0.0	2.8	8.3	80.6	0.0	0.0	0.0
008	0.0	0.0	0.0	0.0	0.0	4.5	0.0	90.9	4.5	0.0
009	2.8	0.0	2.8	0.0	0.0	13.9	0.0	0.0	61.1	19.4
010	0.0	0.0	5.9	2.9	0.0	8.8	0.0	0.0	8.8	73.5

Figura 9. Confusion matrix combining HOG and LBP descriptors. Average accuracy: 76.27%.

Analyzing the results provided in Figure 9, it is possible to notice that the concatenation of HOG and LBP, in comparison to them separately, improves the accuracy of some species in which the three information (color, shape and texture) are important. However, in other species the average accuracy decreases since adding more information can make the species being confused. This situation occurs when butterfly species have different textures but similar colors, for example. The same behavior occurs when different species have different colors but similar textures.

C. Literature Comparison

In this section, we compare the proposed approach to literature methods. Table II shows the comparison between the three proposals of image description and available results in literature regarding Leeds Butterfly Dataset, using the value of average accuracy in percentage.

Method	Average Accuracy
Ground Truth Templates [24]	56.3%
Automatically Learnt Templates [24]	54.4%
Proposed Approach 1 - HOG	68.72%
Proposed Approach 2 - LBP	**77.33%**
Proposed Approach 3 - HOG + LBP	76.27%

Tabela II
COMPARISON OF AVERAGE ACCURACY WITH LITERATURE APPROACHES.

As shown in the table above, the three ways of describing images outperforms the results available in literature considering the Leeds Butterfly Dataset. The difference achieves approximately 21% when comparing the "Proposed Approach 2", which provides the best average accuracy, to the Ground Truth Templates approach. Therefore, the combination of LBP descriptor and PLS classifiers trained using the OAA protocol seems to be a promising approach to identify butterfly species.

V. CONCLUSIONS

The automatic recognition of butterfly species in images is important in several fields. However, it still represents an unsolved problem of Computer Vision since their appearance is very similar and their patterns are difficult to distinguish. Currently, approaches based on Deep Learning seem to be very promising to solve a countless number of problems. However, in a condition in which few samples are available to the training phase, they fail. Therefore, approaches using handcrafted descriptors are still necessary.

In this work, we have proposed a novel approach to identify species of butterflies in images using LBP and HOG descriptors in combination with the PLS classifier in an one-against-all protocol. Results have shown that the proposed method outperformed state-of-the-art approaches in approximately 21%, considering the average accuracy. Besides, it is also possible to conclude that LBP seems to better describe butterfly images since they differ more in texture than in color or shape. The combination of the aforementioned descriptors with the PLS classifier seems to be propitious in the recognition of butterfly species when compared to results available in literature.

Finally, as future works we intend to perform the experiments using a larger dataset which recently has been collected by the authors. Besides, we also consider performing experiments regarding other descriptors of literature since they can capture different information, and also modifying the classifier employed in order to evaluate its influence in average accuracy.

REFERÊNCIAS

[1] "Borboletas e Mariposas - Agencia de Informacao Embrapa," <http://www.agencia.cnptia.embrapa.br/Agencia16/AG01/arvore/AG01_81_911200585235.html>, accessed in: 2019-03-13.

[2] C. Io, *Monographia rhopalocerorum sinensium*. Henan Scientific and Technological Publishing House, 1994.

[3] T. M. Porter, J. F. Gibson, S. Shokralla, D. J. Baird, G. B. Golding, and M. Hajibabaei, "Rapid and accurate taxonomic classification of insect (class insecta) cytochrome c oxidase subunit 1 (coi) dna barcode sequences using a naïve bayesian classifier," *Molecular Ecology Resources*, vol. 14, no. 5, pp. 929–942, 2014.

[4] R. Field, *Butterflies: identification and life history*. Museum Victoria, 2013.

[5] K. J. Gaston and M. A. O'Neill, "Automated species identification: why not?" *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 359, no. 1444, pp. 655–667, 2004.

[6] E. Y. Kalafi, C. Town, and S. K. Dhillon, "How automated image analysis techniques help scientists in species identification and classification?" *Folia morphologica*, vol. 77, no. 2, pp. 179–193, 2018.

[7] L. K. Yılmaz Kaya and R. Tekin, "A computer vision system for the automatic identification of butterfly species via gabor-filter-based texture features and extreme learning machine: GF+ ELM," *Tem Journal*, p. 13, 2013.

[8] J. Wang, L. Ji, A. Liang, and D. Yuan, "The identification of butterfly families using content-based image retrieval," *Biosystems engineering*, vol. 111, no. 1, pp. 24–32, 2012.

[9] Y. Kaya, L. Kayci, and M. Uyar, "Automatic identification of butterfly species based on local binary patterns and artificial neural network," *Applied Soft Computing*, vol. 28, pp. 132–137, 2015.

[10] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[11] Y. Kaya, L. Kayci, R. Tekin, and Ö. Faruk Ertuğrul, "Evaluation of texture features for automatic detecting butterfly species using extreme learning machine," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 26, no. 2, pp. 267–281, 2014.

[12] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, ser. CVPR '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 886–893. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2005.177

[14] H. Wold, "Partial least squares," *Encyclopedia of Statistical Sciences*, pp. 581–591, 1985.

[15] Wikipedia contributors, "Wikipedia, the free encyclopedia," 2020, [Accessed in 30-May-2020]. [Online]. Available: https://en.wikipedia.org/

[16] L. Kayci, "Erek dağı (van) papilionoidea ve hesperioidea ekolojisi ve faunası üzerine araştırmalar (lepidoptera)," *Priamus Suppl*, vol. 6, pp. 1–47, 2007.

[17] P. D. Hebert and T. R. Gregory, "The promise of dna barcoding for taxonomy," *Systematic biology*, vol. 54, no. 5, pp. 852–859, 2005.

[18] Y. Qing, L. Jun, Q.-j. Liu, G.-q. Diao, B.-j. Yang, H.-m. Chen, and T. Jian, "An insect imaging system to automate rice light-trap pest identification," *Journal of Integrative Agriculture*, vol. 11, no. 6, pp. 978–985, 2012.

[19] C. Wen, D. E. Guyer, and W. Li, "Local feature-based identification and classification for orchard insects," *Biosystems engineering*, vol. 104, no. 3, pp. 299–307, 2009.

[20] C. Schmid, G. Dorkó, S. Lazebnik, K. Mikolajczyk, and J. Ponce, "Pattern recognition with local invariant features," in *Handbook of Pattern recognition and computer vision*. World Scientific, 2005, pp. 71–92.

[21] Y. Kaya and L. Kayci, "Application of artificial neural network for automatic detection of butterfly species using color and texture features," *The visual computer*, vol. 30, no. 1, pp. 71–79, 2014.

[22] A. Hernandez-Serna and L. F. Jimenez-Segura, "Automatic identification of species with neural networks," *PeerJ*, vol. 2, p. e563, 2014.

[23] M. Nixon and A. S. Aguado, *Feature extraction and image processing for computer vision*. Academic Press, 2012.

[24] J. Wang, K. Markert, and M. Everingham, "Learning models for object recognition from natural language descriptions." in *BMVC*, vol. 1, no. 2009, 2009, p. 2.