Global Journal of Engineering and Technology [GJET].
ISSN: 2583-3359 (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journal-gjet-home/

Color-Based Image Segmentation Using K-Means Clustering Approach

BY

Jennifer Chi[1]*, Yeong Nain Chi[2]

School of Behavioral and Brain Sciences The University of Texas at Dallas
Department of Agriculture, Food and Resource Sciences University of Maryland Eastern Shore

Article History

Received: 05/02/2024
Accepted: 15/02/2024
Published: 17/02/2024

Vol –3 Issue –2

PP: - 11-15

Abstract:

*This study aims to develop and implement image processing functions using the histogram of a single, two-dimensional butterfly image. The specific objective is to classify the color feature sets of image pixels. The focus of the study is on color-based image segmentation, which is based on the assumption that regions of homogeneous colors within an image correspond to distinct clusters representing meaningful objects. The initial approach of the study involves the developing and implementing image processing functions based on the histogram of the butterfly image. These functions aim to extract essential color features from the image pixels and subsequently classify the image based on these features. To achieve this, the study employs a k-means clustering approach, resulting in a three-cluster solution: Cream Color Cluster, Yellow Color Cluster, and Orange Color Cluster. The analysis reveals significant dissimilarities among the three clusters obtained from the butterfly image in terms of histograms and visual features derived from the L*a*b* color space. Through this approach, valuable insights into the intricate color patterns exhibited by butterfly wings can be uncovered.*

Keywords: *Image segmentation, color-based, histogram, k-means clustering, MATLAB.*

1. Introduction

Butterflies are adored insects, celebrated for their captivating features. Their wings exhibit vivid colors and intricate patterns, which originate from the presence of minuscule scales. These scales house the genetic information for melanin, the pigment responsible for governing the colors and structure of butterfly wings. Collectively, these scales resemble pixels in an image, with each scale individually attached to the wing. This distinct attribute enables butterflies to shed scales, aiding their escape from spider webs. Scales are a defining feature of butterflies and moths, evident in the scientific name of their order, Lepidoptera, derived from the Greek words *lepidos* and *ptera*, meaning "scale" and "wing" (Dinwiddie, 2013).

The main objective of this study is to map and analyze the diverse range of color variations present within a single, two-dimensional butterfly image. To achieve this, the initial approach involves the developing and implementing image processing functions based on the histogram of the butterfly image. These functions aim to extract essential color features from the image pixels and subsequently classify the image based on these features. Through this approach, valuable insights into the intricate color patterns exhibited by butterfly wings can be uncovered.

The image can be represented as numeric data arrays, including a data matrix and a color map matrix, facilitating further image analysis using tools such as MATLAB. Image processing encompasses the use of computer algorithms to generate, manipulate, transmit, display, and analyze images, encompassing a wide range of techniques applicable across various domains. Among these techniques, segmentation plays a crucial role as a fundamental step towards optimal image performance (MathWorks, 2022). By effectively dividing an image into distinct segments, segmentation enables precise analysis and interpretation of individual objects, enhancing overall image comprehension and subsequent processing tasks.

In context of image segmentation, one approach is visual interpretation; however, it can be resource-intensive due to the large number of pixels and variations in human judgment. To overcome these challenges, automated algorithms can be employed using non-supervised or supervised methods, producing results consistent with human interpreters when assigning pixels to specific images.

Image segmentation holds a critical position in image processing as it enables the separation of objects within an image, facilitating the analysis of individual objects. This process involves dividing the image into multiple segments and assigning labels to each pixel (image object) based on

*Corresponding Author: Jennifer Chi

shared visual characteristics. By doing so, image segmentation simplifies subsequent image analysis tasks (MathWorks, 2022).

In this study, the focus is on color-based image segmentation, which relies on the color features of image pixels. This approach assumes that regions with homogeneous colors correspond to distinct clusters and meaningful objects within the image. Each cluster represents a group of pixels sharing similar color properties. It is important to note that the choice of color space significantly influences the segmentation results, as there is no single color space that can universally yield acceptable outcomes for all types of images (MathWorks, 2022).

2. Materials

A butterfly image (Figure 1) was obtained from Unsplash, an online repository for visuals, (https://unsplash.com/photos/TUTI-349CAM). The Image Processing Toolbox of MATLAB (R2022a) was utilized for the processing of the butterfly image. This included a histogram (Figure 2) and examining the L*a*b* Color Space (L*: lightness between black and white, a*: green and red, and b*: blue and yellow) representation of the butterfly image (Figure 3).

Figure 1: Butterfly Image

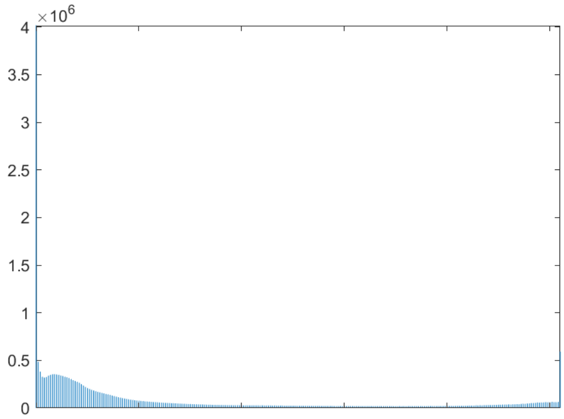

Figure 2: Histogram of Butterfly Image

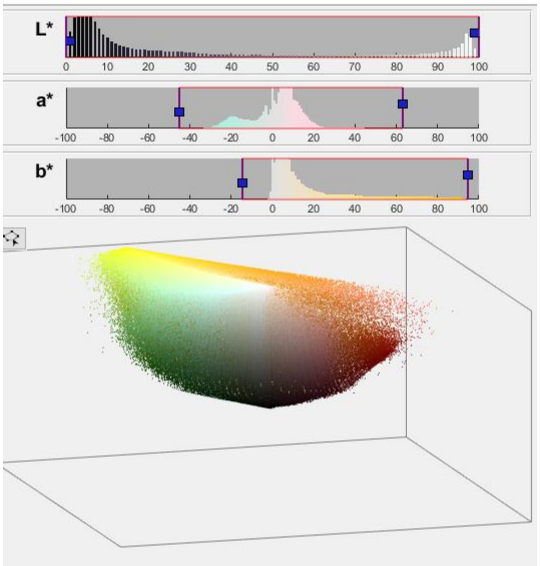

Figure 3: L*a*b* Color Space of Butterfly Image

3. Methods

3.1 Histogram Approach

The analysis of images involved extracting meaningful information, such as color identification, shape detection, object counting, property measurement, edge detection, noise removal, and statistical calculations to assess image quality. A fundamental tool in image processing was the histogram, which provided a comprehensive description of an image. In image processing, the histogram represented the distribution of pixel values or indices for indexed color images. It served as a crucial component for various tasks, such as image equalization, where the pixel intensities were adjusted to achieve the desirable contrast (MathWorks, 2022).

Histogram modeling, including techniques like histogram equalization, offered a sophisticated approach to modifying the dynamic range and contrast of an image by manipulating its intensity to achieve a desired shape. Histogram equalization, in particular, employed a non-linear mapping that reassigned intensity values of pixels, resulting in an output image with a uniform distribution of intensities and a flat histogram. This technique found applications in image comparison processes and the correction of non-linear effects caused by display systems (MathWorks, 2022).

The histogram approach proved highly efficient as it typically required only single pass through the pixels. This technique involved computing a histogram based on all pixels in the image, with the peaks and valleys of the histogram assisting in the identification of clusters within the image (Shapiro and Stockman, 2001). Additionally, the histogram approach could be applied on a per-pixel basis, where the resulting information determined the most prevalent color for each pixel location. Histograms demonstrated effectiveness in image classification and object class recognition tasks, often accomplished through *k*-means clustering analysis (Schroff et al., 2006).

*Corresponding Author: Jennifer Chi

3.2 K-Means Clustering Analysis

Clustering is a fundamental process that involved grouping object attributes and features to ensure that data objects within one group are more like each other than those in other groups. Clustering is found in various fields, including engineering, biology, medicine, psychology, economics, as well as in data mining, search engines, recommendation systems, knowledge discovery, bioinformatics, information retrieval, computer vision, pattern recognition, and image processing.

Among the various clustering algorithms, *k*-means clustering stands out as the most popular and widely used due to its simplicity, speed, and efficiency. This algorithm effectively classifies data into distinct groups based on input parameters and convergence patterns. It belongs to the category of unsupervised learning techniques and is commonly used for clustering unlabeled data. The objective of *k*-means clustering is to minimize a squared error function, known as the objective function (Kanungo et al., 2002), given by:

$$\mathbf{J = \Sigma^{k}_{j=1} \Sigma^{n}_{i=1} ||X_i - C_j||^2}$$

Where J represents the objective function, n denotes the number of objects, k denotes the number of clusters, x_i represents object i, c_j represents the centroid for cluster j, and $\mathbf{||X_i - C_j||}$ denotes the Euclidean distance between x_i and c_j.

The goal of *k*-means clustering was to partition n objects into k clusters, ensuring that each object belongs to the cluster with the closest mean. This iterative method yielded exactly k distinct clusters with the greatest possible dissimilarity. It assigned each object to the cluster with the nearest centroid and recalculates the centroid by averaging the objects within the cluster (Huang, 1998).

4. Results

MATLAB is a high-performance language renowned for its powerful commands and syntax, specifically designed for technical computing. Its Image Processing Toolbox offers a comprehensive suite of algorithms and visualization functions, catering to various image analysis tasks, including image processing, histogram modeling, and pattern recognition (MathWorks, 2022).

In studies by Wang et al. (2014), image classification using the L*a*b* color space consistently demonstrated promising results, often outperforming other color spaces. The L*a*b* color space comprises the 'L*' luminosity layer and the chromaticity layers ‘a*' and 'b*', representing color positioning along the red-green and blue-yellow axes, respectively. These layers encapsulate all essential color information and enable both visual differentiation and quantification of color differences using the Euclidean distance metric (Baldevbhai1 and Anand, 2012).

Building upon these considerations, this study adopted the L*a*b* color space for color-based butterfly image segmentation. Given that the color information existed in the 'a*b*' color space, the targeted objects were pixels with ‘a*’ and ‘b*’ values. MATLAB provided the function “imsegkmeans” to cluster the objects based on assigned clusters. For clustering the objects based on assigned clusters, the “imsegkmeans” function provided by MATLAB was utilized. This function assigned an index or label to each input object, facilitating the leveling of every pixel in the image (MathWorks, 2022) (Figure 4).

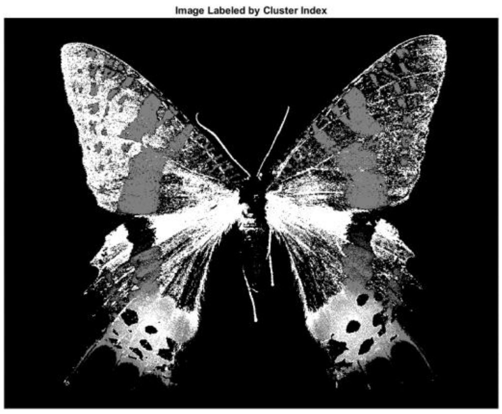

Figure 4: Image Labeled by Cluster Index

The *k*-means clustering analysis is employed to identify a solution with a specific number of clusters, resulting in a three-cluster solution: the *Cream Color Cluster*, *Yellow Color Cluster*, and *Orange Color Cluster*. Objects were assigned to clusters based on their distances from the cluster centers, computed using the Euclidean distance metric. Substantial differences were observed among the butterfly image clusters, both visually and in terms of histograms and L*a*b* color space features.

Cluster 1, referred to as the *Cream Color Cluster*, exhibited a cream color along the wing (Figure 5). The histogram of this cluster (Figure 6) displayed an initial steep drop, reaching a minimum at 0.5, and subsequently skewed towards zero. In the L*a*b* color space, the ‘a*’ layer ranged approximately between -20 and 10 (red-green axis), while the 'b*' layer ranged roughly from 0 to 20 (blue-yellow axis) (Figure 7).

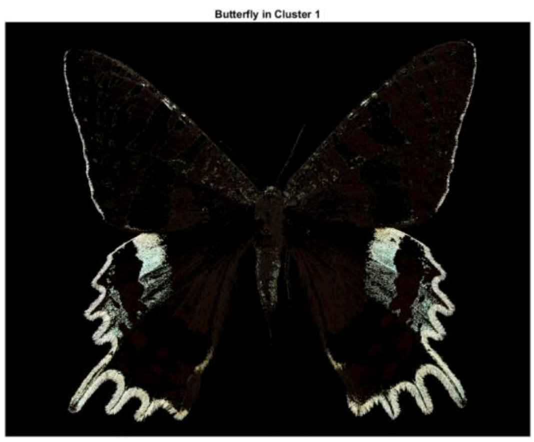

Figure 5: Butterfly Image Cluster 1

*Corresponding Author: Jennifer Chi

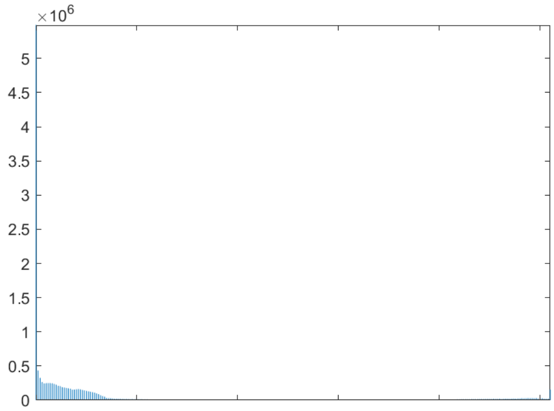

Figure 6: Histogram of Butterfly Image Cluster 1

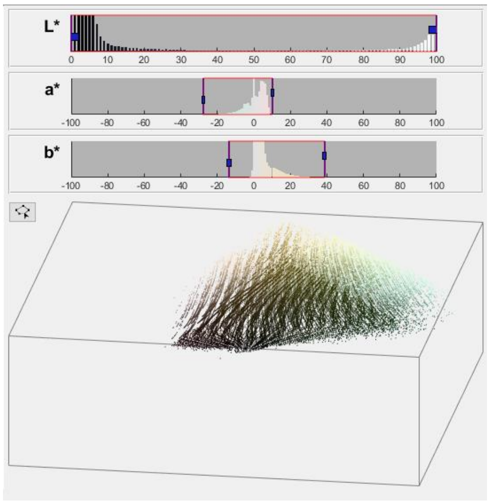

Figure 7: L*a*b* Color Space of Butterfly Image Cluster 1

Cluster 2, labelled as the *Yellow Color Cluster*, exhibited yellow color arranged symmetrically on both sides within this image cluster (Figure 8). The histogram of the *Yellow Color Cluster* (Figure 9) appeared flat and even, making it difficult to discern distinct peaks. In the L*a*b* color space, the 'a*' layer (representing the red-green axis) ranged approximately between -30 to 10, while the 'b*' layer (representing the blue-yellow axis) ranged roughly from 10 to 95 (Figure 10).

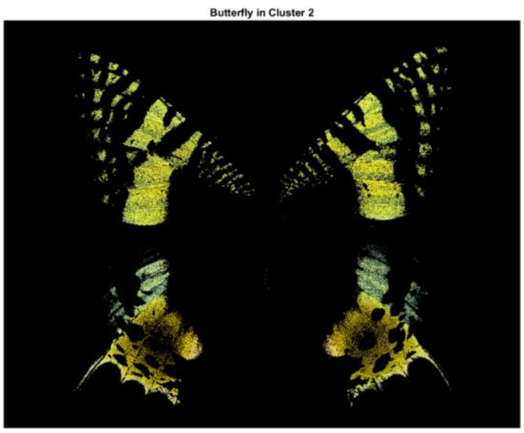

Figure 8: Butterfly Image Cluster 2

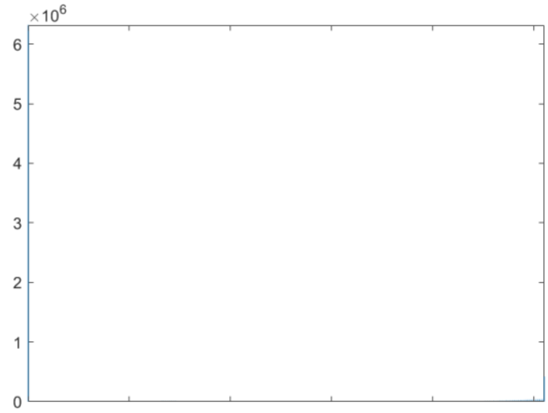

Figure 9: Histogram of Butterfly Image Cluster 2

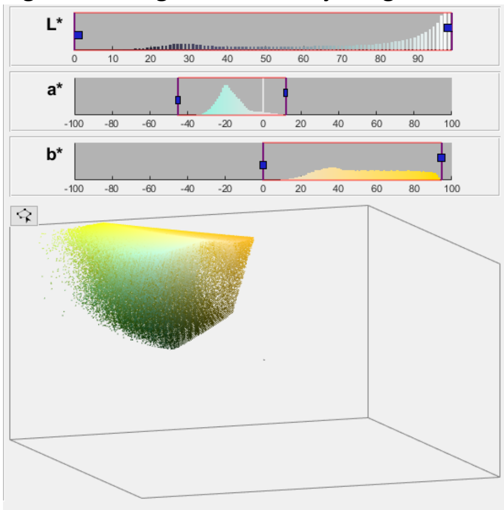

Figure 10: L*a*b* Color Space of Butterfly Image Cluster 2

Cluster 3, designated as the *Orange Color Cluster*, exclusively exhibited orange color, forming the background in this image cluster (Figure 11). The histogram of the *Orange Color Cluster* (Figure 12) displayed a relatively wave-like pattern. The color information in the 'a*' layer (red ~ green) lay between 0 to 40, while in the 'b*' layer (blue ~ yellow), it ranged approximately from 0 to 60 (Figure 13).

Figure 11: Butterfly Image Cluster 3

*Corresponding Author: Jennifer Chi

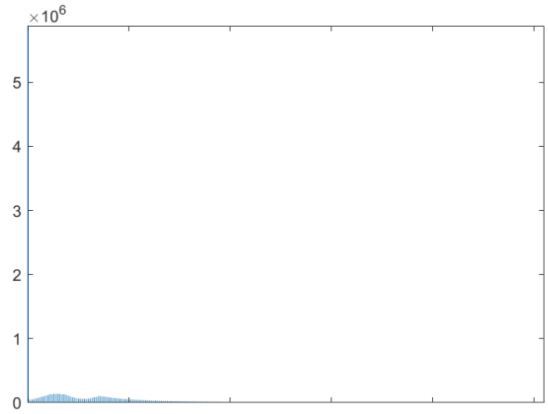

Figure 12: Histogram of Butterfly Image Cluster 2

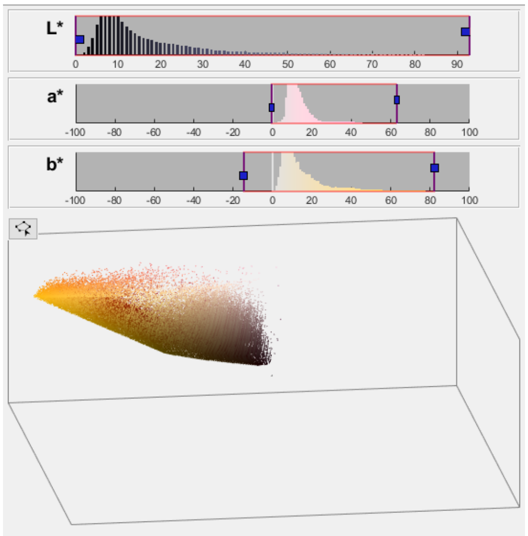

Figure 13: L*a*b* Color Space of Butterfly Image Cluster 3

5. Conclusion

The automatic identification of insects from images is a topic garnering increasing interest and popularity. This study focuses on utilizing a butterfly color image to provide hands-on experience in classifying a singly butterfly image using image processing tools such as MATLAB. The insights gained from this study can serve as a foundation for further research on butterfly identification and recognition, exploring the intricate composition of butterfly color and morphology.

Through the implementation of *k*-means clustering analysis, we achieved a three-cluster solution, labeled as *Cream Color Cluster*, *Yellow Color Cluster*, and *Orange Color Cluster*. The distinct differences observed among these butterfly image clusters, both in terms of histograms and L*a*b* color space features, underscored the effectiveness of this clustering approach.

Moving forward, the research can be extended by incorporating other feature extraction methods and classification techniques. Expanding the dataset will enhance the robustness and generalizability of the classification process. Additionally, considering the implementation of convolutional neural networks to achieve even more precise and accurate butterfly image classification will take the research to a more advanced level. Embracing these avenues for further investigation will undoubtedly contribute to advancing the field of image analysis and its practical applications.

Conflicts of Interest

All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publications.

References

1. Baldevbhai1, P. J. & Anand, R. S. 2012. Color image segmentation for medical images using L*a*b* color space. IOSR Journal of Electronics and Communication Engineering, 1(2), 24-45. DOI: 10.9790/2834-0122445
2. Dinwiddie, April. 2013. Mapping a morphometric analysis of butterfly and moth wing scale shape to a family-level phylogeny. Retrieved from: https://www.researchgate.net/publication/267905301_Mapping_a_morphometric_analysis_of_butterfly_and_moth_wing_scale_shape_to_a_family_level_phylogeny.
3. Huang, Z. 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery, 2, 283–304. Retrieved from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.15.4028&rep=rep1&type=pdf
4. Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silvermank, R. & Wu, A. Y. 2002. An Efficient k-Means Clustering Algorithm: Analysis and Implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7), 881-892. https://doi.org/10.1109/TPAMI.2002.1017616
5. MathWorks. 2022. MATLAB (R2022b) Image Processing Toolbox™ User's Guide. Natick, MA: The MathWorks, Inc.
6. Schroff, F., Criminisi, A. & Zisserman, A. 2006. Single-histogram class models for image segmentation. In: Kalra, P. K. and Peleg, S. (eds.) Computer Vision, Graphics and Image Processing. Lecture Notes in Computer Science, vol. 4338. Springer, Berlin, Heidelberg. Retrieved from: https://www.robots.ox.ac.uk/~vgg/publications/2006/Schroff06/schroff06.pdf
7. Shapiro, L. G. & Stockman, G. C. 2001. Computer Vision. pp 279–325. New Jersey: Prentice-Hall.
8. Wang, X., Hänsch, R., Ma, L. & Hellwich, O. 2014. Comparison of different color spaces for image segmentation using graph-cut. Proceedings of the 9th International Conference on Computer Vision Theory and Applications (VISAPP-2014), 301-308. Retrieved from: https://www.cv.tu-berlin.de/fileadmin/fg140/VISAPP_2014_127_CR.pdf

*Corresponding Author: Jennifer Chi