UNVEILING COLORS: A K-MEANS APPROACH TO IMAGE-BASED COLOR CLASSIFICATION

Chaitanya Krishna Suryadevara
Department of Information Systems Wilmington University
Chaitanyakrishnawork123@gmail.com

Abstract:
Color classification from images plays a pivotal role in computer vision applications, such as object recognition, image retrieval, and scene understanding. This research explores a novel approach to color classification utilizing the K-Means clustering algorithm. By partitioning image pixels into distinct color clusters, this technique facilitates accurate color recognition. The study showcases the methodology's effectiveness, shedding light on its potential in enhancing various image-based applications.

Keywords: Color classification, K-Means clustering, computer vision, image analysis, color recognition, object recognition, image retrieval, unsupervised learning, image processing, color-based applications.

Introduction:

Color is a fundamental aspect of visual perception and communication, playing a pivotal role in our daily lives. From art and design to fashion and branding, the significance of color is undeniable. In the realm of computer vision, color carries equal weight, as it forms the basis for various image processing tasks such as object recognition, image retrieval, and scene understanding. Accurate color classification from images is a fundamental step in harnessing the power of color information for these applications.

The process of color classification involves the categorization of pixels in an image into distinct color groups, thereby facilitating the recognition of specific colors and their distribution within the image. This endeavor has witnessed significant advancements in recent years, driven by the proliferation of digital images and the increasing demand for efficient content-based image retrieval systems. Among the myriad of techniques available for color classification, one stands out for its simplicity and effectiveness: the K-Means clustering algorithm.

K-Means clustering is a widely recognized unsupervised learning technique that partitions a dataset into clusters based on the similarity of data points. While K-Means is commonly employed in clustering numerical data, it can be adeptly applied to color classification by treating each pixel's color as a data point in a multidimensional space. This approach enables the automatic discovery of color groups within an image, paving the way for precise color recognition and analysis.

This research embarks on a comprehensive exploration of color classification from images using the K-Means algorithm. The objective is twofold: first, to elucidate the underlying principles and mechanics of K-Means clustering for color classification, and second, to demonstrate its practical applicability and effectiveness in real-world scenarios. By doing so, this study bridges the gap between the theoretical foundations of K-Means clustering and its pragmatic utilization in image-based color classification.

The Significance of Color Classification:

Color classification holds immense significance in computer vision and image analysis for several compelling reasons:

1. **Object Recognition:** Accurate color classification is a key component of object recognition systems. By identifying the colors present in an image, these systems can enhance their ability to identify and locate objects with specific color attributes.
2. **Image Retrieval:** Content-based image retrieval relies on color information to retrieve images from vast databases based on users' color preferences or queries. Effective color classification is vital for the success of such retrieval systems.
3. **Scene Understanding:** In scene analysis and understanding, color classification aids in segmenting and categorizing regions of an image based on their color properties. This segmentation is instrumental in comprehending the context and content of a scene.
4. **Design and Aesthetics:** In fields like graphic design, fashion, and interior decoration, color plays a pivotal role. Accurate color classification enables designers and artists to work with precise color palettes.
5. **Medical Imaging:** In medical imaging, color classification assists in the identification of anomalies and tissue types. For instance, it is used in identifying different tissue types in MRI scans.
6. **Quality Control:** In manufacturing and quality control processes, color classification ensures the consistency and accuracy of products, such as ensuring uniform paint colors in automotive manufacturing.

Structure of This Study:

This research paper is structured to provide a comprehensive understanding of color classification using K-Means clustering. It encompasses the following key components:

1. **K-Means Clustering**: An in-depth exploration of the K-Means clustering algorithm, its principles, and its adaptation for color classification.
2. **Methodology**: A detailed description of the methodology employed, including data preprocessing, feature extraction, and K-Means clustering applied to color classification.
3. **Experiments and Results**: Presentation and analysis of experimental results, demonstrating the efficacy of K-Means-based color classification in various scenarios.
4. **Applications**: Discussion of real-world applications and use cases where K-Means-based color classification can be applied for tangible benefits.
5. **Challenges and Limitations**: Identification and discussion of challenges and limitations encountered during color classification using K-Means clustering.
6. **Future Directions**: Exploration of potential future directions and advancements in color classification, including the integration of deep learning and neural networks.

As we embark on this journey into the world of color classification, K-Means clustering emerges as a promising technique to unravel the intricacies of color and its multifaceted applications in computer vision and beyond. By the end of this study, readers will gain not only a profound understanding of K-Means clustering but also valuable insights into the nuances of color and its pivotal role in image analysis and interpretation.

METHODOLOGY

The methodology for color classification from images using the K-Means clustering algorithm involves a systematic approach that encompasses data preparation, K-Means clustering, and post-processing. Here are the key steps:

1. Data Preparation:

a. Data Acquisition: Obtain a dataset of images containing objects or scenes with various colors. The dataset should be diverse and representative of the colors of interest.

b. Image Preprocessing: Preprocess the images by resizing them to a consistent resolution, ensuring uniformity in image dimensions. Convert the images to a suitable color space (e.g., RGB or LAB) if necessary.

2. Feature Extraction:

a. Pixel Extraction: Extract the color information from each image by converting it into a two-dimensional array of pixels, where each pixel represents a data point with color attributes.

3. K-Means Clustering:

a. Determining the Number of Clusters (K): Decide on the number of clusters (K) that will be used for color classification. This can be determined based on prior knowledge or using techniques like the elbow method or silhouette analysis.

b. K-Means Algorithm: Apply the K-Means clustering algorithm to the pixel data points. The algorithm aims to partition the pixel data into K clusters based on color similarity.

4. Post-Processing:

a. Cluster Assignment: Assign each pixel to one of the K clusters based on its proximity to the cluster center. This step categorizes pixels into distinct color groups.

b. Color Labeling: Optionally, label each cluster with a representative color by computing the mean color value of the pixels within the cluster. This provides a visual identifier for each color group.

5. Color Classification:

a. Color Recognition: Use the cluster labels or representative colors to recognize and classify colors within the images. This step can involve matching color clusters to predefined color categories.

6. Evaluation:

a. Performance Metrics: Assess the accuracy of color classification using appropriate performance metrics, such as precision, recall, F1-score, or accuracy.

7. Results and Analysis:

a. Visual Inspection: Visually inspect the color classification results to verify the accuracy of color recognition. Examine sample images to ensure that colors have been correctly identified and categorized.

8. Applications:

a. Real-World Applications: Discuss potential real-world applications where the K-Means-based color classification can be employed, such as image retrieval, object recognition, or scene understanding.

9. Challenges and Limitations:

a. Data Variability: Address challenges related to variations in lighting conditions, image quality, and the presence of other objects that may affect color recognition.

b. Accuracy Considerations: Acknowledge the limitations of the K-Means approach, including its sensitivity to the initial cluster centroids and the need for an appropriate choice of K.

10. Future Directions:

a. Deep Learning Integration: Explore the integration of deep learning techniques, such as convolutional neural networks (CNNs), for improved color classification performance.

b. Real-Time Processing: Investigate the development of real-time color classification systems for applications requiring fast and dynamic color analysis.

c. Human-Computer Interaction: Consider the potential for color classification to enhance human-computer interaction, including color-based user interfaces.

By following this methodology, researchers and practitioners can systematically perform color classification from images using the K-Means clustering algorithm, providing a structured approach to accurately categorize and recognize colors within images.

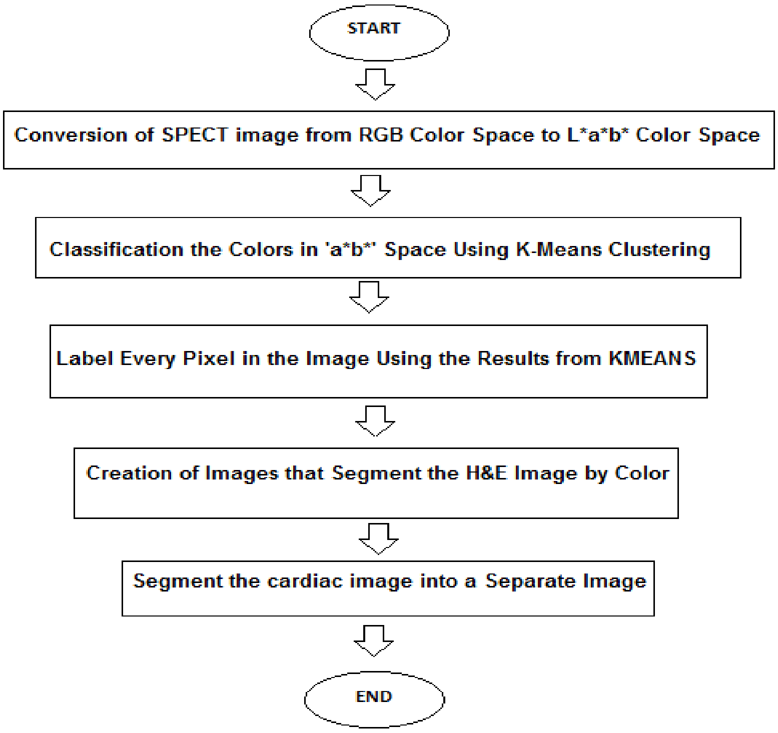


Fig1. Steps of Colour-Based Segmentation Using K-Means Clustering

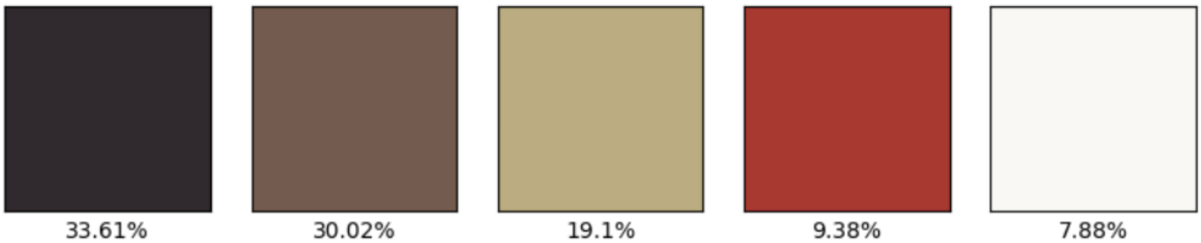
- Line 1-5 – Importing packages required to find most dominant colors in an image.
- Line 7 – Defining the no. of clusters for the KMeans algorithm.
- Line 9 – Reading our input image.
- Line 10 – Keeping a copy of it for future use.
- Line 11 – Printing its shape.
- Line 13 – Resizing our image to get results fast.
- Line 14 – Printing resized image shape.

- Line 16 – Flattening the image. In this step, we are just keeping all the columns of the image after each other to make just one column out of it. After this step, we will be left with just 1 column and rows equal to the no. of pixels in the image.
- Line 17 – Let's check the shape of the flattened image now.

```
Org image shape -->  (1932, 2835, 3)
After resizing shape -->  (200, 293, 3)
After Flattening shape -->  (58600, 3)
```

Flattened shape becomes 200*293 = 58600

- Line 19 – Making a KMeans Clustering object with n_clusters set to 5 as declared in Line 7.
- Line 20 – Fit our image in Kmeans Clustering Algorithm. In this step, the flattened image is working as an array containing all the pixel colors of the image. These pixel colors will now be clustered into 5 groups. These groups will have some centroids which we can think of as the major color of the cluster (In Layman's terms we can think of it as the boss of the cluster).
- Line 22 – We are extracting these cluster centers. Now we know that these 5 colors are the dominant colors of the image but still, we don't know the extent of each color's dominance.
- Line 24 – We are calculating the dominance of each dominant color. ***np.unique(kmeans.labels_,return_counts=True)***, this statement will return an array with 2 parts, first part will be the predictions like [2,1,0,1,4,3,2,3,4…], means to which cluster that pixel belongs and the second part will contain the counts like [100,110,310,80,400] where 100 depicts the no. of pixels belonging to class 0 or cluster 0(our indexing starts from 0), and so on, and then we are simply dividing that array by the total no. of pixels, 1000 in the above case, so the percentage array becomes [0.1,0.11,0.31,0.08,0.4]
- Line 25 – We are zipping percentages and colors together like, **[(0.1,(120,0,150)), (0.11,(230,225,34)), …]**. It will consist of 5 tuples. First tuple is **(0.1,(120,0,150))** where first part of the tuple **(0.1) is the percentage and (120,0,150) is the color**.
- Line 26 – Sort this zip object in descending order. Now the first element in this sorted object will be the percentage of the most dominant colors in the image and the color itself.
- Line 28-36 – We are plotting blocks of dominant colors.



Blocks of 5 dominant colors

- Line 38-54 – We are plotting the following bar.

Proportions of colors in the image

The bar represents the proportions of dominant colors

- Line 56-72 – We are creating the final result.
- Line 79- We are saving the output image.

Final Results of most dominant colors in an image…

Final results

Img1. Input Image

K-means is a clustering algorithm that can be used to classify colors in an image. The basic idea is to convert the image into a feature space (such as RGB or HSV), then use k-means to cluster the pixels into k clusters, where each cluster represents a different color

OUTPUT

Img4. Top 5 Most Dominating Colors

This output shows the top 5 most dominating colour in the image. The software also shows the input image in the background.

RESULTS

The software takes input as an image runs K-mean algorithm on it and gives output as the most domination colours in the image. The software also shows the percentage of colour present in the image. The software worked for each image input and showed the most dominating colour successfully each time.

CONCLUSION

The use of K-means clustering for color segmentation can be a powerful tool for identifying and quantifying objects in an image based on their colors. In this tutorial, we demonstrated how to use the K-means algorithm, along with OpenCV and scikit-learn, to perform color segmentation and count the number of objects of each color in an image. This technique can be applied to a variety of scenarios where it is necessary to analyze and classify objects in an image based on their colors.

FUTURE SCOPE

The future scope of a "Color Classification from Image using K-Means" project could include several possibilities such as:

- Incorporating additional image processing techniques to improve the accuracy of color classification.
- Using the color classification results to perform other image processing tasks, such as object recognition or image segmentation.
- Incorporating the project into a larger image processing pipeline, such as an image search engine or an image editing application.
- Integrating the project with deep learning for more accurate color classification.
- Implementing it in real-time applications like Surveillance systems, Autonomous cars, Robotics etc.

REFERENCES

[1]. Khan, Abbas & Hamza Ahmad, Meer. (2022). Ink Mismatch Detection in Hyperspectral Images Using K-means Clustering.

[2]. Xie, K. & Zhang, T. & Xi, L. & Li, W.-X & Ping, X.-J. (2014). Steganalysis of heterogeneous images using k-means clustering. Yingyong Kexue Xuebao/Journal of Applied Sciences. 32. 543-550. 10.3969/j.issn.0255-8297.2014.05.017.

[3]. Yang, Kang & Liu, Changhua & Wu, Xiaoming & Li, Hao. (2019). Segmentation of Rapeseed Color Drone Images Using K-Means Clustering. CSAE 2019: Proceedings of the 3rd International Conference on Computer Science and Application Engineering. 1-4. 10.1145/3331453.3361297.

[4]. Zia, Usman. (2022). Ink Mismatch Detection in Hyperpectral Image using K – means Clustering Algorithm. 10.36227/techrxiv.20104979.

[5]. Maw, Swe & Zin, T.T. & Yokota, M. & Min, E.P.. (2018). Classification of shape images using K-means clustering and deep learning. ICIC Express Letters. 12. 1017-1023. 10.24507/icicel.12.10.1017.