A

PROJECT SYNOPSIS

ON

"Recognition of Dominant Colors in a Picture"

Submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

Session 2023-24

In

Computer Science and Engineering

Submitted by:

Sreethu K Binu (2100970100117)

Under the supervision of :

Dr. Aditya Dev Mishra

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS COLLEGE OF ENGINEERING AND TECHNOLGY
GREATER NOIDA

AFFILIATED TO
Dr. A.P.J. ABDUL KALAM TECHNICAL UNIVERSITY
LUCKNOW

INTRODUCTION

Extracting and analyzing dominant colors from images has become increasingly valuable. This project aims to develop an efficient method for identifying the most dominant colors in an image using the K-Means clustering algorithm.

K-Means clustering is a popular unsupervised machine learning technique that can effectively segment an image into 'k' clusters, where each cluster represents a dominant color in the image. By leveraging the power of K-Means clustering, we can automate the process of identifying prominent color palettes without the need for manual selection or analysis.

The project includes image preprocessing, K-means clustering, color frequency analysis, visualization, and parameter optimization. It will result in a tool that identifies the dominant colors and provides insights into their prevalence, allowing users to make informed decisions based on image aesthetics.

Our project will provide an efficient and practical solution for using color analysis in visual content. Whether you're a graphic designer or data scientist, our tools and knowledge will help you achieve your goals.

OBJECTIVES

1. **Image Preprocessing:** Create an image preprocessing pipeline that converts color space and flattens images for analysis.
2. **K-Means Clustering Implementation:** Use K-Means clustering with a specified 'k' value to find dominant colors in preprocessed image data. Optimize the model for performance and accuracy.
3. **Color Frequency Analysis:** Count pixels for each cluster center to analyze dominant color frequency and determine relative dominance.
4. **Visualization and Reporting:** Create a visual representation of dominant colors and generate a report with information about their prevalence.
5. **Parameter Optimization and Scalability:** Optimize K-Means clustering for different image sizes and qualities.

METHODOLOGY

1. **Image Preprocessing:** Load the image using a library like OpenCV, convert it to RGB if necessary, and flatten it into a 2D array.
2. **Determine the Number of Clusters (k):** Choose the number of dominant colors based on project requirements.
3. **K-Means Clustering:** Use scikit-learn to apply the K-Means clustering algorithm to the image data. Fit the model with the chosen number of clusters (k), which will group pixels based on their similar color values to the cluster centers.
4. **Extract Dominant Colors:** Retrieve the cluster centers, which represent the dominant colors in the image. These centers are in the RGB color space.
5. **Analyze Color Frequency:** Count pixels per cluster to determine color prevalence and calculate dominant color percentage/ratio.
6. **Sort Dominant Colors:** Optionally, sort the dominant colors by their frequency to identify the most dominant colors.
7. **Visualization:** Visualize the dominant colors - using a color palette, pie chart, or image with highlighted colors.
8. **Documentation and Testing:** Document code and methodology clearly. Test on diverse images for reliability and accuracy.

TOOLS/TECHNOLOGY USED

Hardware requirements:

1. Computer: A standard desktop or laptop computer is sufficient for smaller-scale image processing tasks.
2. Processor: Use a multi-core processor for faster image processing. A modern CPU with 2-4 cores is adequate for small to medium-sized images.
3. RAM: 4-8 GB RAM is recommended for image processing.
4. Storage: Use SSDs for faster data access and better system responsiveness while standard HDDs are adequate for storage.
5. Operating System: A modern operating system such as Windows, macOS, or Linux that supports the required software and libraries.
6. Software: Install required software - Python, OpenCV for image processing, and scikit-learn for machine learning.

Software requirements:

1. **Python:** Python is the primary programming language for this project.
2. **Python Libraries:** Install Python libraries using pip for image processing, K-means clustering, and data visualization.
    - **OpenCV:** For image loading, manipulation, and conversion.
    - **scikit-learn:** For K-Means clustering and data analysis.
    - **NumPy:** For numerical operations.
    - **Matplotlib:** For data visualization.
3. **Image Data:** Gather JPEG or PNG images compatible with OpenCV library for color analysis.

REFERENCES

1. https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning
2. https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1
3. https://pypi.org/project/scikit-learn/
4. https://www.youtube.com/watch?v=4b5d3muPQmA