Introduction

This code is designed to analyze and process a dataset containing images categorized by age groups. The dataset is structured such that each image's filename starts with the respective age of the person in the image. The code collects and processes these images to perform Exploratory Data Analysis (EDA) and analyze aspects such as age distribution, brightness characteristics, and histograms of pixel intensities across different age groups.

Methodology

Data Collection and Preprocessing:

The code iterates through a directory ('dataset') containing image files, extracting age information from filenames and appending the images to XData after resizing and normalizing them.

Simultaneously, it creates a list yData to store the age values and a dictionary age_list to count the occurrences of each age.

Exploratory Data Analysis (EDA):

Descriptive Statistics: The script generates descriptive statistics for age values using Pandas Series.describe() method to understand the central tendency, dispersion, and distribution of ages.

Data Visualization: Utilizing Matplotlib and Seaborn, the code creates a histogram to visualize the distribution of age values.

Histogram Equalization and Brightness Analysis:

Histogram Equalization: Defines a function calculate_histogram_and_cdf to compute the normalized histogram and Cumulative Distribution Function (CDF) for a given age value.

Analysis of Pixel Intensity Histograms: The code defines functions to calculate normalized histograms and Cumulative Distribution Functions (CDFs) for two specific age values ('1' and '100') and visualizes the differences between their pixel intensity distributions.

Average Brightness Calculation: The script computes the average pixel intensity (brightness) for images corresponding to age groups '10' and '60' and generates histograms to compare their brightness distributions.

Explanation of Methods

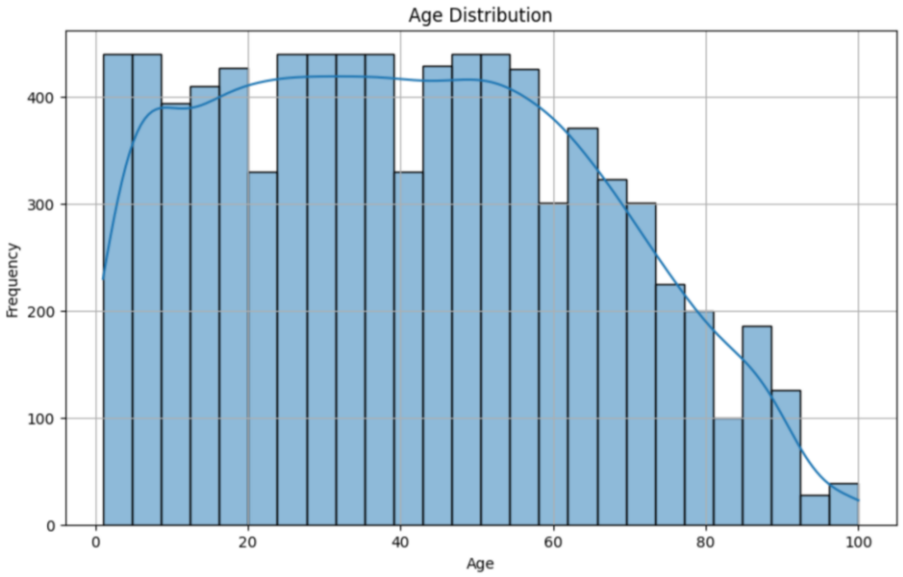
calculate_histogram_and_cdf(age_value): Calculates normalized histograms and CDFs for a specified age value by processing grayscale images from the dataset directory.

calculate_average_brightness(age_value): Computes the average pixel intensity (brightness) for a given age value by analyzing grayscale images from the dataset directory.

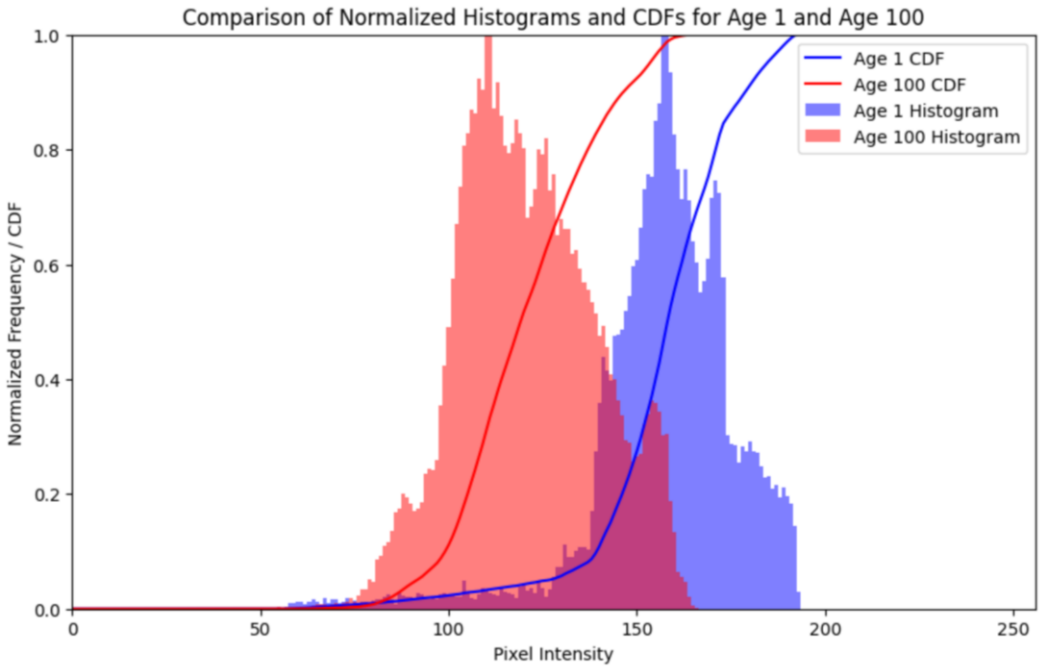
perform_edge_detection(age_value): Conducts Canny edge detection on images corresponding to a specific age value. It computes the count of edge pixels and returns a list of these counts.

Conclusion

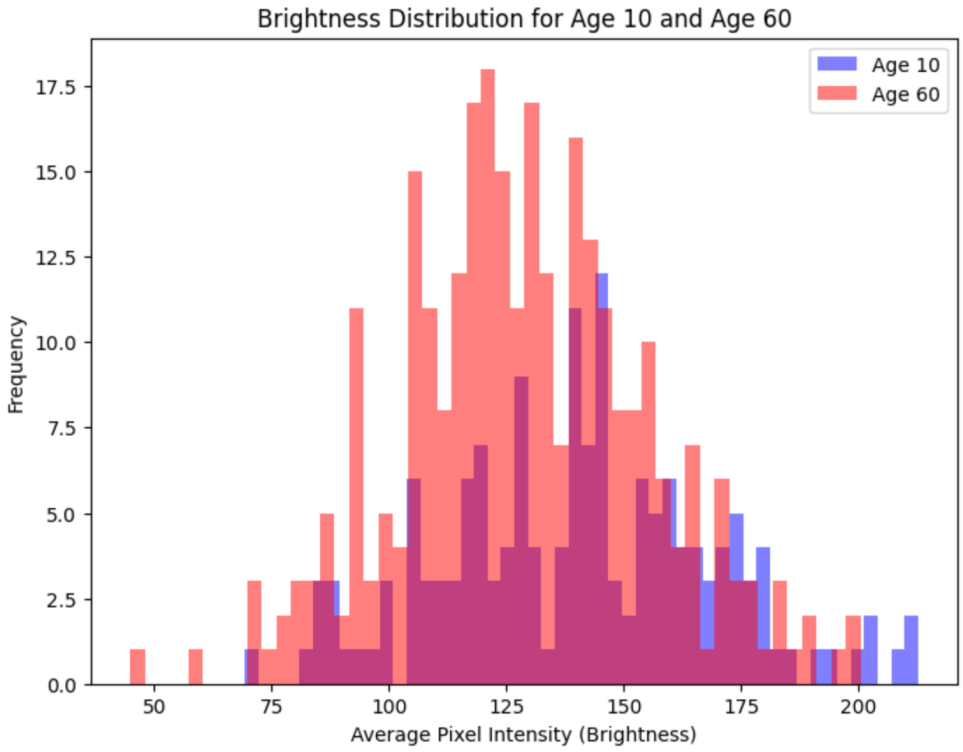
Descriptive statistics and visualization of the age distribution.



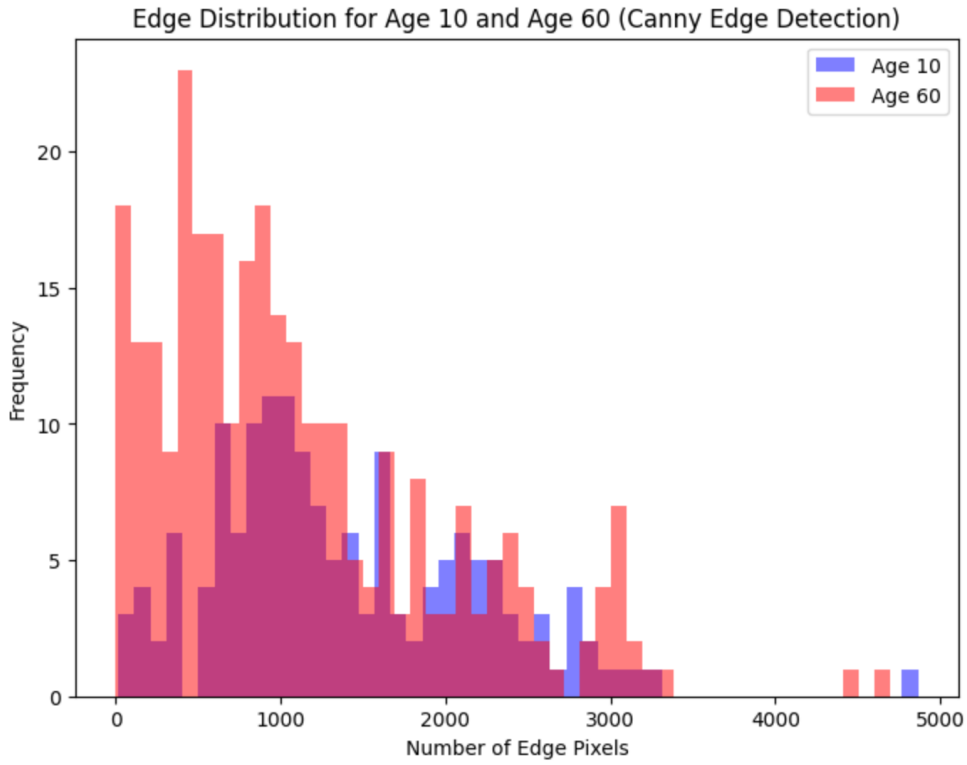
Comparison of normalized histograms and CDFs for age groups '1' and '100'.



Analysis and comparison of brightness distributions for age groups '10' and '60'.



Edge detection analysis using the Canny edge detection algorithm.



Through the performed EDA and image processing techniques, this code effectively demonstrates the distributional aspects of age in the dataset, showcases variations in pixel intensity histograms and brightness distributions across different age groups.

We observe a tail on the normalized histograms, indicating that contrast is a significant feature within the dataset. Additionally, it's evident that data pertaining to younger individuals tends to exhibit higher brightness levels. Moreover, there's a noticeable increase in edge distribution frequency for older individuals. This suggests that both brightness and edge distribution could be considered essential features to further investigate in the age detection process.

This analysis provide insights into age-related characteristics in the dataset and can potentially inform further image processing or machine learning tasks related to age estimation or facial recognition based on age.