Image Classification Using Global Descriptors

MIRAHSANI Sana

February 7, 2026

1 Introduction

In this TP, we investigate classification pipelines using global descriptors, focusing on color and texture features. Specifically, we implement classification based on color histograms, Local Binary Pattern (LBP) histograms, their fusion, and finally, descriptors enriched with spatial information via geometric subdivision. The objective is to evaluate the discriminative power of these descriptors and understand how combining complementary visual cues and incorporating spatial structure can improve recognition performance. Experiments are conducted on a subset of the Caltech-101 dataset, providing insights into the strengths and limitations of each descriptor type.

2 Dataset

For this TP, I used a subset of the Caltech-101 dataset, which contains 1000 images distributed equally across 10 classes. The classes are: Faces_easy, Leopards, Motorbikes, Airplanes, Bonsai, Car_side, Chandelier, Hawksbill, Ketch, and Watch. All images are in JPG format and were read using OpenCV (cv2.imread). No additional preprocessing was required because the images are already clean and correctly labeled, allowing us to focus directly on descriptor extraction and the classification pipeline.

3 Classification Using Color Histograms

3.1 Methodology

In this experiment, image classification is based on global color histograms used as visual descriptors. For each image, a joint color histogram is computed by quantizing each RGB channel into 8 bins, resulting in a descriptor of size 512. The quantization is performed by mapping pixel intensity values to discrete bins, after which the three color channels are combined into a single flattened index representing each color configuration. A histogram is then built by counting the occurrences of these indices and is normalized. To process the full dataset, all images listed in the label file are read using OpenCV, converted from BGR to RGB format, and passed to the histogram computation function. The resulting feature vectors are aggregated into a feature matrix X, while the corresponding class labels are stored in a label vector y. These descriptors are subsequently used within a Jupyter notebook to implement the full classification pipeline, including data splitting, classifier training, prediction, and evaluation.

3.2 Recognition Rate

The overall recognition rate is 0.545 (54.5This is likely due to the joint color histogram being a naive descriptor: it captures only the frequency of colors, ignoring spatial relationships. Classes with similar colors are often confused, while distinct-color classes are recognized well.

3.3 Confusion Matrix Analysis

Below is the confusion matrix by global color histogram.

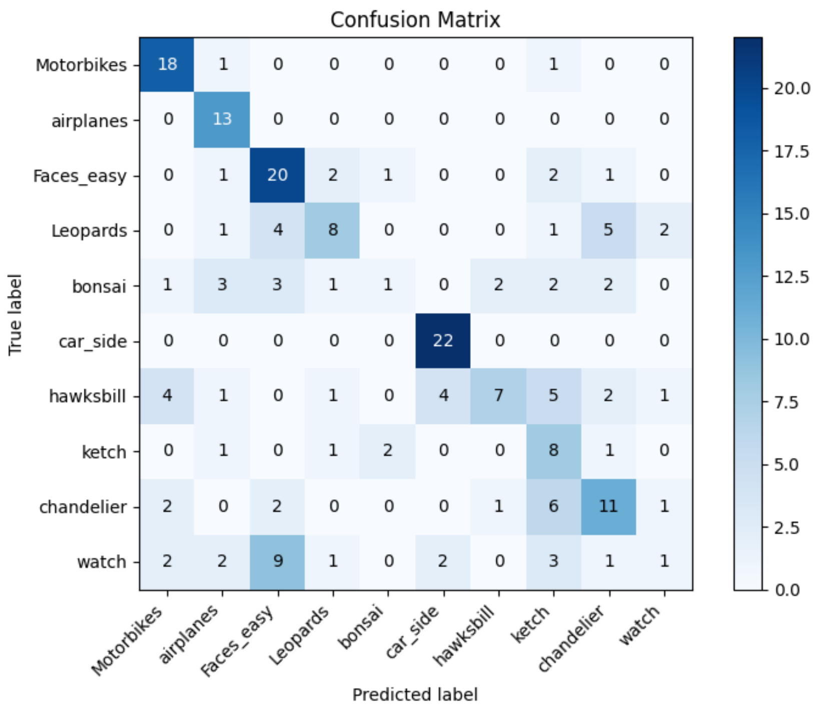

Figure 1: A comparison of key evaluation metrics

1. **Best-recognized classes:** Leopards, car_side
2. **Worst-recognized classes:** bonsai, watch
3. **Most confused classes:** watch, bonsai, chandelier, airplanes

Interpretation: The results show that color alone is insufficient to distinguish all classes. Classes with unique colors are classified correctly, but overlapping colors lead to misclassification. Incorporating texture or shape descriptors (e.g., LBP, HOG) could improve performance.

4 Classification Using LBP Histograms

4.1 Methodology

In this experiment, image classification is performed using Local Binary Pattern (LBP) histograms as texture-based descriptors. For each image, the RGB image is first converted to grayscale, after which LBP codes are computed at each pixel using 8 neighboring points and a radius of 1, with the uniform LBP variant. The resulting LBP image is then summarized by a histogram that counts the occurrences of each LBP code, where the number of bins is automatically determined from the maximum LBP value. This histogram is normalized to represent a probability distribution, ensuring invariance to image size. To extract descriptors for the entire dataset, the same function used previously for color histograms is reused to apply the LBP histogram computation to all images, producing a feature matrix and the corresponding label vector. These descriptors are then employed within the same classification pipeline as before, allowing a direct comparison with color-based descriptors.

4.2 Recognition Rate

Using LBP histograms as descriptors, the classification system achieves a recognition rate of 34.5% on the test set. This result is relatively low compared to color-based descriptors, indicating that texture information alone is not sufficient to effectively discriminate between all classes in the dataset. This can be explained by the fact that several object categories are primarily characterized by global shape or color rather than local texture patterns.

4.3 Confusion Matrix Analysis

The confusion matrix obtained with LBP histograms highlights a strong variability in recognition performance across classes. Classes such as Leopards and Car_side are among the best recognized, which can be attributed to their distinctive and repetitive texture patterns that are well captured by LBP descriptors. In contrast, classes like Bonsai and Watch exhibit very low recognition rates and are frequently misclassified, indicating that their visual characteristics are not well represented by local texture information alone. Additionally, significant confusion is observed between visually diverse classes such as Watch, Bonsai, Chandelier, and Airplanes, suggesting that LBP histograms struggle to discriminate objects whose appearance relies more on global structure or shape rather than fine-grained texture.

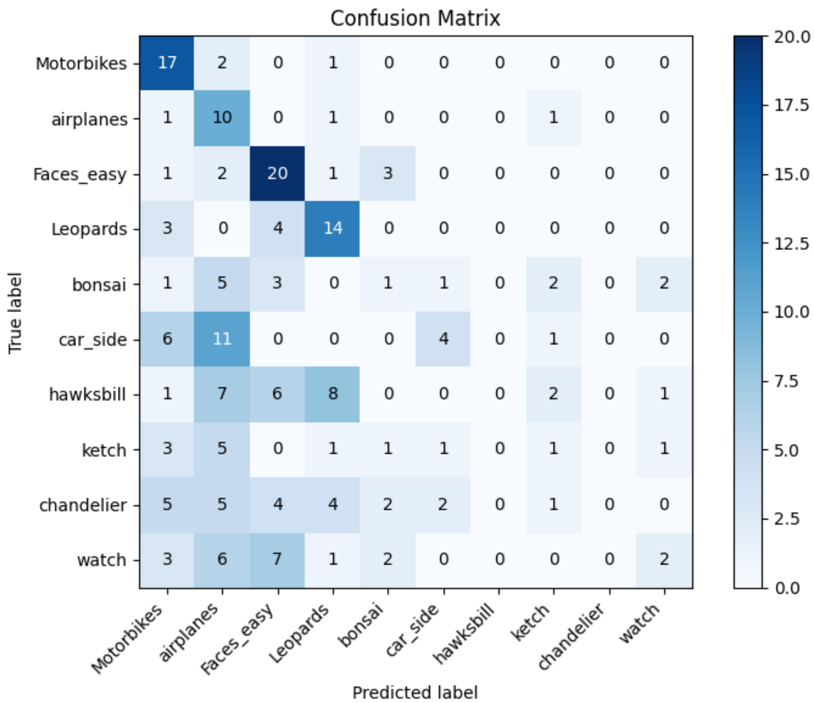

Figure 2: A comparison of key evaluation metrics

5 Fusion of Descriptors

5.1 Methodology

To combine color and LBP descriptors, an early fusion strategy is adopted by concatenating the two feature vectors into a single descriptor. This fusion is implemented using the np.hstack function, which horizontally stacks the color histogram and LBP histogram corresponding to the same image. The resulting fused descriptor integrates both color and texture information and is then used within the same classification pipeline to compute the confusion matrix and evaluate the classification performance.

5.2 Recognition Rate

Using the fused descriptor obtained by concatenating color and LBP histograms, the classification system achieves a recognition rate of 57% on the test set. This result represents a significant improvement compared to using each descriptor independently, showing that combining complementary visual information leads to better discrimination between classes.

5.3 Confusion Matrix Analysis

The confusion matrix obtained with the fused descriptor shows a clear improvement in classification performance compared to using color or LBP histograms alone. Most classes are now better recognized, with strong diagonal values indicating correct predictions, particularly for Faces_easy, Motorbikes, and Car_side. Misclassifications are reduced, though some confusion remains in visually similar classes such as Ketch, Chandelier, and Watch. Overall, combining color and texture information allows the classifier to capture complementary visual cues, leading to higher discrimination between classes and a significant increase in recognition rate.

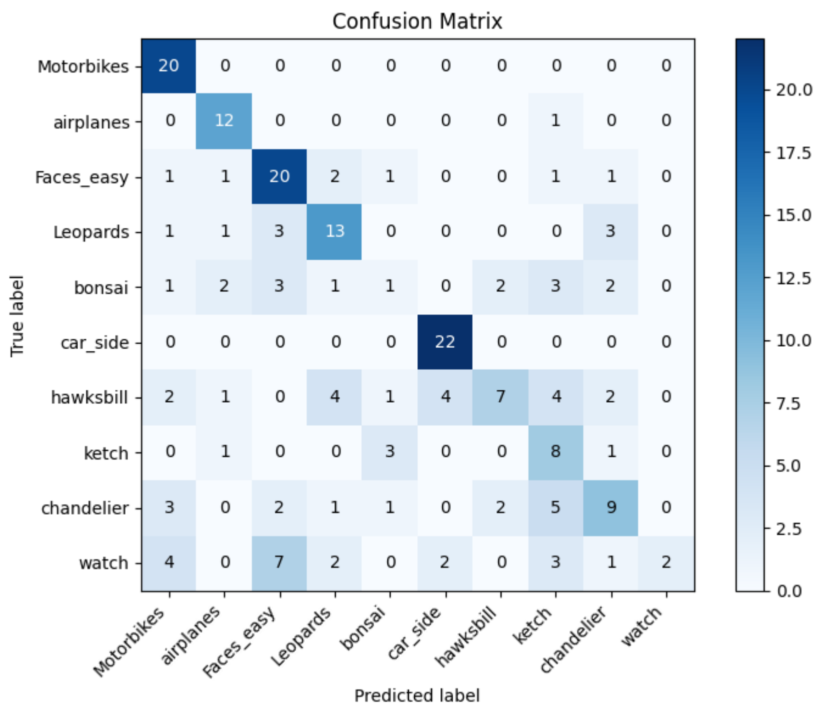

Figure 3: A comparison of key evaluation metrics

6 Geometric Subdivision of Images

6.1 Methodology

To incorporate spatial information into the descriptors, each image is subdivided into a 5×5 grid of regions, resulting in 25 equally sized sub-images. For each region, a color histogram is computed using the same quantization strategy as for the global descriptor. The histograms of all regions are then concatenated to form a single feature vector representing the entire image, effectively capturing both color and spatial distribution information. This method allows the classifier to leverage local patterns within different parts of the image, improving discrimination between classes that may have similar global color distributions but differ in spatial arrangement.

6.2 Recognition Rate

Using the 5×5 geometric subdivision of images, the classification system achieves a recognition rate of 75% on the test set. This represents a substantial improvement over global descriptors and fused descriptors, highlighting the importance of spatial information. By capturing local color distributions within each region, the descriptor is better able to discriminate between classes with similar overall appearance but different spatial layouts, leading to more accurate predictions.

6.3 Confusion Matrix Analysis

The confusion matrix obtained with the 5×5 geometric subdivision highlights a marked improvement in class discrimination compared to previous descriptors. Most classes, including Faces_easy, Motorbikes, and Car_side, are now recognized with high accuracy, as shown by the strong diagonal values. Misclassifications are mainly limited to a few visually similar classes, such as Ketch, Chandelier, and Watch, which still exhibit some confusion. Overall, incorporating spatial information through regional histograms allows the classifier to distinguish classes that are difficult to separate based solely on global color or texture, explaining the significant increase in recognition rate to 75

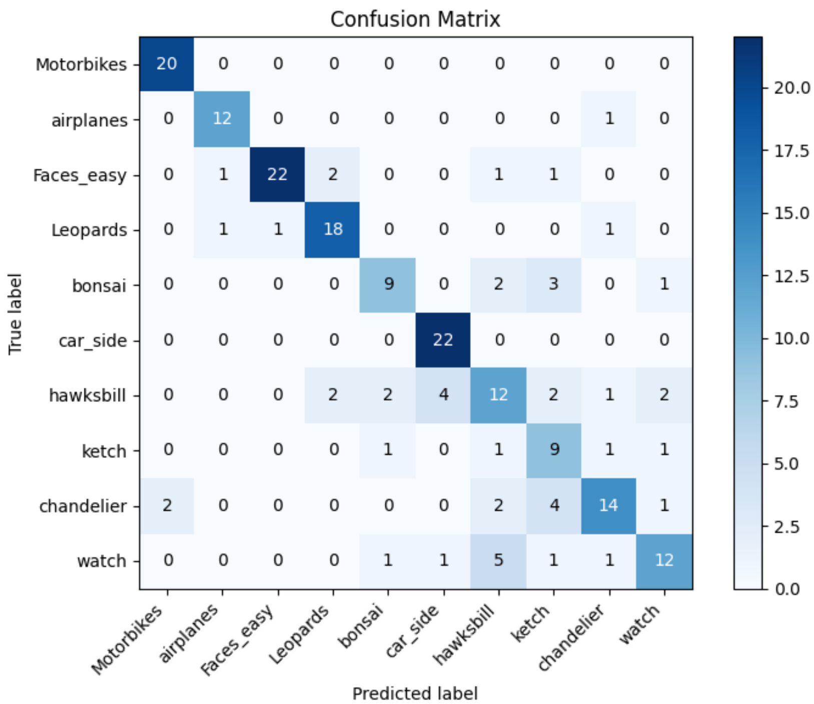

Figure 4: A comparison of key evaluation metrics

7 Conclusion

In this TP, we explored several approaches for image classification using global descriptors on a subset of Caltech-101. Classification based on color histograms achieved moderate performance by capturing global color distributions, while LBP histograms provided complementary texture information but were less effective alone. Combining these descriptors through early fusion improved recognition, demonstrating the benefit of integrating color and texture features. Finally, incorporating spatial information via a 5×5 geometric subdivision significantly enhanced performance, achieving the highest recognition rate of 75%, as it allows the classifier to differentiate classes with similar global appearance but distinct local patterns. Overall, these experiments highlight the importance of selecting descriptors that capture complementary visual cues and spatial structure for effective image classification.