Assignment #3: Face Recognition Pipeline

Biljana Vitanova
IBB 2024/25 , FRI, UL
bv7063@student.uni-lj

***Abstract*—The goal of this assignment was to familiarize oneself with the face recognition pipeline, which included face detection, feature extraction, and feature matching, essential for the identification and verification of persons.**

I. Introduction

Face recognition is a biometric modality aimed at distinguishing between different faces. The process involves extracting distinctive features or key points from each face and using these to form descriptors. These descriptors are then compared using specific metrics to determine whether two faces belong to the same person or different individuals. Face recognition is widely used in applications such as security, identity verification, and surveillance.

II. Methodology

For the **face detection step**, I used the Viola-Jones algorithm, which is simple and fast. The algorithm relies on **Haar features**, which are rectangular patterns that compute the weighted sum of pixel intensities, and a **cascade of classifiers** to determine if a region contains a face by filtering out non-face objects through multiple stages.

The main parameters I optimized included the **scale factor**, which defines how much the image is scaled at each step, **min neighbors**, specifies the number of overlapping rectangles required to confirm a face, and **min size**, sets the dimensions of the smallest detectable face.

For **feature extraction**, I used three methods.

1) The **Histogram of Oriented Gradients** (HOG) divides the image into cells and creates a histogram based on gradient orientations.
2) **Dense SIFT** descriptors uniformly distribute keypoints across the image and generate descriptors for each. And feature desciptors generated with deep-learning model
3) **AlexNet** [1] deep learning model, is used for efficient and low-cost feature extraction.

For **feature matching**, I've used the **Hellinger distance** to measure similarity.

III. Experiments

For this assignment, I used the CelebA-HQ-small dataset [2], which consists of 475 training images and 412 test images, each of size 512×512 pixels.

Fig. 1: Sample face images from the dataset.

After experimenting with different parameters for face detection step, I used a scale factor of 1.05, a minimum of 3 neighbors, and a minimum face size of (20, 40) pixels for detection. The scale factor proved to be the most critical for performance, as increasing it resulted in worse accuracy.

During the feature extraction step, each image was preprocessed to a predefined size of 512×512 pixels. This size was chosen as a trade-off between computation time and performance, offering better results. Also, converting images to grayscale was required for both HOG and Dense SIFT descriptors.

For the HOG descriptor, I used 9 orientations per histogram. Increasing the number of orientations to 17 did not improve performance. Pixels per cell were set to (4, 4), and the cells per block parameter significantly improved recognition accuracy, though it resulted in slower computation.

For the Dense SIFT descriptor, I experimented with different step sizes for distributing keypoints, which determine how many keypoints are used. In the end, I decided on using 5 for the step size, resulting in a larger descriptor but better results.

For alexnet, any particular tuning wasn't necessary.

IV. Results and Discussion

The performance of the recognition pipeline was evaluated separately for the detection step, the feature extraction step, and the entire pipeline as a whole.

A. Results

TABLE I: IoU performance on the training and test sets.

Split	IoU[%]
Train set	69.72
Test set	69.32

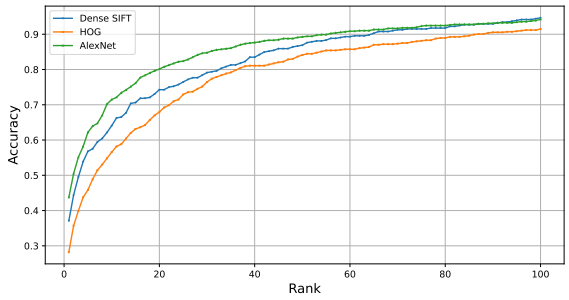

Fig. 2: **Cumulative Match Characteristic on whole images** (CMC) curve showing the retrieval accuracy for each descriptor, with accuracy displayed for ranks up to 100.

TABLE II: Rank-1 and Rank-5 on whole images

Descriptor	Rank-1 Accuracy[%]	Rank-5 Accuracy[%]
Dense SIFT	37.13	56.79
HOG	28.20	48.87
Alexnet	43.75	62.17

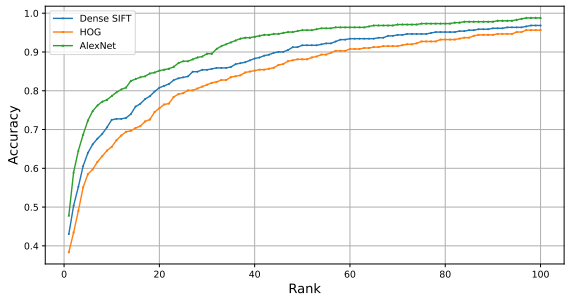

Fig. 3: **Cumulative Match Characteristic** (CMC) curve showing the retrieval accuracy of the complete face recognition pipeline for each descriptor, with accuracy displayed for ranks up to 100.

TABLE III: Rank-1 and Rank-5 on complete face recognition pipeline

Descriptor	Rank-1 Accuracy[%]	Rank-5 Accuracy[%]
Dense SIFT	43.06	63.99
HOG	38.34	58.49
Alexnet	47.81	72.33

B. Discussion

The IoU, when evaluated on the test set, showed similar results to those on the training set, likely because the test set was not significantly different from the training set and did not introduce new faces.

When evaluating the Rank-1 and Rank-5 accuracy on the entire dataset, AlexNet showed the best performance, followed by the SIFT descriptor, and then HOG. Dense SIFT and HOG demonstrated similar performance. The accuracy increased continuously with each rank.

When evaluating Rank-1 and Rank-5 accuracy across the whole recognition pipeline, the descriptors showed improved performance with a steeper increase in accuracy. However, there were no significant differences overall, with AlexNet again performing the best and HOG the worst.

Descriptors obtained with AlexNet were the most compact and required the least computation time for comparison. In contrast, HOG descriptors were significantly more computationally expensive, taking approximately four times longer to compare.

V. Conclusion

This assignment was a good starting point for understanding and building face recognition pipelines. It helped explore key steps like detection, feature extraction, and evaluation, while also identifying their strengths and weaknesses.

For future improvements, adding a face alignment step could help improve accuracy by making the inputs more consistent. Trying more advanced feature descriptors or deep learning methods might also lead to better results. To improve IoU performance, new evaluation metrics or better detection thresholds could be tested. Using larger and more varied datasets could make the pipeline work better in different situations.

References

[1] L. Ding, H. Li, C. Hu, W. Zhang, and S. Wang, "Alexnet feature extraction and multi-kernel learning for objectoriented classification," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-3, pp. 277–281, 04 2018.

[2] W. Xia, Y. Yang, J.-H. Xue, and B. Wu, "Towards open-world text-guided face image generation and manipulation," *arxiv preprint arxiv: 2104.08910*, 2021.