CSP 502
Computer Vision

Assignment 1

A comparative study of Face Recognition algorithms

Group: 6

Chintan Gandhi (201501019)

Anshul Jethvani (201501037)

Jay Modi (201501059)

Darshan Shah (201501094)

Harsh Shah (201501096)

Hardik Udeshi (201501113)

Problem Statement

- Robust Face Recognition using SIFT and HOG feature extraction in different classifier models
- Compare results of different approaches

Algorithms for Feature Extraction

- SIFT and HOG feature extractors have been the seminal algorithms in the space of this problem.
- Face images can be scaled, rotated and may have been taken under different illumination or from different viewpoints.
- The features of the test image need to be extracted and matched with those of training images.
- CNN can also be used for feature extraction and subsequent face recognition.

Approaches for Face Classification

A. Euclidean based Matching

- SIFT and HOG algorithms are used for feature extraction.
- We compare each feature vector of test image to the feature vectors of train image and compute their Euclidean distance.
- The overall euclidean distance, which we call score, between the test and train images is the normalised sum of euclidean distances of matched feature vectors.
- We repeat this matching of test images with all the training images and the label for test image is the label of the train image where the test-train image pair has minimum score (rank-1 match).

Approaches for Face Classification

B. Support Vector Machine (SVM)

- The feature vectors of the image obtained from SIFT/HOG feature extractor are flattened into a long one-dimensional vector. The long feature vector for an image is provided a corresponding label.
- This vector-label pair is generated for all training images and fed into the SVM. This is a case of multi-class SVM classification.
- The test images are passed as input to the SVM and the output of SVM is the predicted labels assigned to the test images.

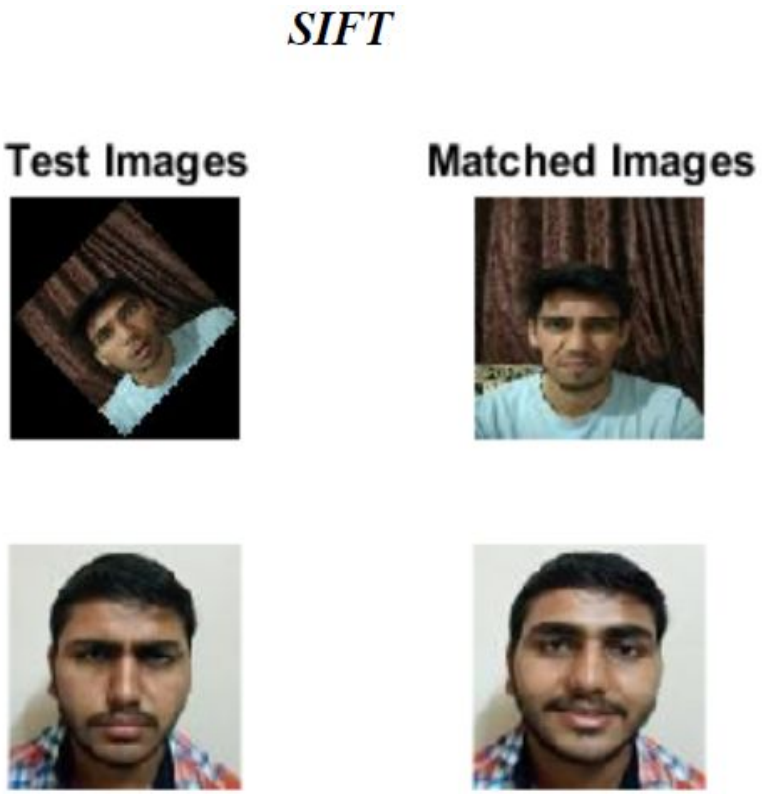
Approaches for Face Classification

C. Convolutional Neural Network (CNN)

For this, we have used two approaches:

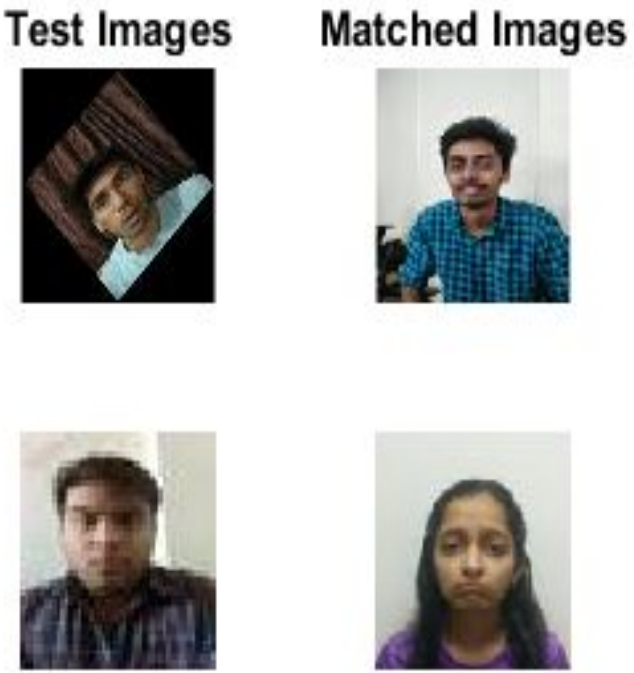
1. We have used a pre-trained network of Inception v2. The input image size is 96x96. The parameters of the network are the same as that of Inception v2 ConvNet model.
2. We have implemented a not so deep Scratch CNN (image input size: 300x300) which consists of five layers:- Convolution layer, Max Pooling layer, Fully Connected layer and two Dense Layers (one with Relu Activation and other with Sigmoid Activation(for just 40 classes available in the dataset)). We have used Adam Optimizer and Cross Entropy Loss Function.

Results: SIFT with Euclidean




```
Accuracy =

    0.5465

>> [precision,recall]

ans =

    1.0000    1.0000
    1.0000    0.5000
    0.7500    1.0000
    0.1429    0.5000
    0.6667    0.6667
    0.6667    0.6667
    0.3333    0.5000
    1.0000    0.2000
    0.2000    0.5000
    0.3333    1.0000
         0         0
       NaN         0
    0.5000    0.6667
       NaN       NaN
    0.6667    0.5000
    1.0000    0.4000
```

Results: HOG with Euclidean




```
Accuracy =

    0.4884

>> [precision,recall]

ans =

    0.5000    0.6667
    0.5000    0.5000
    1.0000    0.6667
    0.5000    1.0000
    0.3333    0.3333
    1.0000    0.3333
    0.2500    0.5000
         0         0
    1.0000    1.0000
    1.0000    1.0000
    0.5000    0.5000
    0.5000    0.5000
    0.3333    0.3333
         0       NaN
    0.1429    0.2500
         0         0
```

Results: SIFT & HOG with SVM classifier

```
../Dataset/orl/201501097/test/201501097_fear_0.jpg - Real label: 201501097 - Pred label:201501097
../Dataset/orl/201501101/test/201501101_anger_0.jpg - Real label: 201501101 - Pred label:201501101
../Dataset/orl/201501101/test/201501101_cap_0.jpg - Real label: 201501101 - Pred label:201501101
../Dataset/orl/201501104/test/2015011040_anger.jpg - Real label: 201501104 - Pred label:201501104
../Dataset/orl/201501104/test/2015011040_fear.jpg - Real label: 201501104 - Pred label:201501104
../Dataset/orl/201501109/test/201501109_anger_o.jpg - Real label: 201501109 - Pred label:201501109
../Dataset/orl/201501109/test/201501109_cap_o.jpg - Real label: 201501109 - Pred label:201501109
../Dataset/orl/201501109/test/201501109_fear_o.jpg - Real label: 201501109 - Pred label:201501109
../Dataset/orl/201501112/test/201501112_anger_0.jpg - Real label: 201501112 - Pred label:201501112
../Dataset/orl/201501121/test/201501121_anger_0.jpg - Real label: 201501121 - Pred label:201501121
../Dataset/orl/201501121/test/201501121_fear_0.jpg - Real label: 201501121 - Pred label:201501121
Accuracy obtained using SIFT is: 85.71%
```

SIFT with SVM Classifier

```
In [78]: runfile('E:/Important/hardik_sem7/CV/CV_HOG/CV_Assignment_HOG/src/main_svm.py', wdir='E:/Important/hardik_sem7/CV/
CV_HOG/CV_Assignment_HOG/src')
Reloaded modules: DataSet, extract_descriptors, SVM, image_processing
Getting the data ...
Applying Hog ...
Training the classifyer ...
(370, 3200)
(370, 1)


Final HOG Accuracy:
Score of Training:1.0
Score:0.7927927927927928
```

HOG with SVM Classifier

Results: CNN

```
it's 2015010210_fear, the distance is 0.400509 201501021_fear.jpg
*** 201501021 201501021
it's 201501034_Glasses_0, the distance is 0.433599 2015010250_goggle.jpg
it's 2015010110_happy, the distance is 0.393 2015010250_joy.jpg
it's 2015010250_disgust, the distance is 0.330322 2015010250_sadness.jpg
*** 201501025 201501025
it's 201501025_surprise, the distance is 0.328189 2015010250_surprise.jpg
*** 201501025 201501025
it's 17440010_googles, the distance is 0.383159 201501025_fear.jpg
it's 201501028_Joy, the distance is 0.302503 201501028o_joy.jpg
*** 201501028 201501028
it's 2015010280_fear, the distance is 0.330815 201501028o_sad.jpg
*** 201501028 201501028
it's 2015010280_fear, the distance is 0.270616 201501028o_surprise.jpg
*** 201501028 201501028
total correctly identified: 59
total incorrectly identified: 38
total accuracy : 60.824742268041234
```

CNN Pre-trained Inception Network (Input: 96x96)

```
Epoch 15/20
100/100 [==============================] - 68s 684ms/step - loss: 0.3108 - acc: 0.9150 - val_loss: 0.3618 - val_acc: 0.9394
Epoch 16/20
100/100 [==============================] - 68s 683ms/step - loss: 0.0641 - acc: 0.9900 - val_loss: 0.2327 - val_acc: 0.9697
Epoch 17/20
100/100 [==============================] - 68s 684ms/step - loss: 0.3476 - acc: 0.9150 - val_loss: 0.4085 - val_acc: 0.9091
Epoch 18/20
100/100 [==============================] - 68s 683ms/step - loss: 0.1466 - acc: 0.9700 - val_loss: 0.2452 - val_acc: 0.9697
Epoch 19/20
100/100 [==============================] - 68s 684ms/step - loss: 0.1267 - acc: 0.9750 - val_loss: 0.1815 - val_acc: 0.9848
Epoch 20/20
100/100 [==============================] - 68s 682ms/step - loss: 0.1354 - acc: 0.9750 - val_loss: 0.2514 - val_acc: 0.9697
```

Scratch CNN (Input: 300x300) (20 Epochs)

Observation and Analysis

- Euclidean distance approach gives less accuracy compared to SVM and CNN approaches since euclidean distance approach is a very crude approach. When image descriptors are obtained using SIFT, accuracy is better in comparison to image descriptors obtained by HOG.
- Multi-Class SVM classifier with SIFT & HOG gives better accuracy in comparison to Euclidean based approach.

Observation and Analysis contd...

- We have got poor accuracy when we used pre-trained CNN classifier for Face image classification compared to SIFT feature descriptor with SVM classification. The reason behind that is the number of images available in the dataset was only 570, while CNN requires the size of datasets to be in thousands. So, to tackle such problem, we need to augment the data.
- Also, we computed accuracy by taking boundary box on faces and we observed that the accuracy was improved in all the cases. The reason for this is because many of the images had high Background-Face ratio.
- Scratch CNN gives the highest accuracy among all three approaches. Also, keras' data augmentation class is used so the relationship is best learnt (comparatively) by the CNN.

References

- Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60, no. 2 (2004): 91-110.
- Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, pp. 886-893. IEEE, 2005.
- Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. "Rethinking the inception architecture for computer vision." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818-2826. 2016.

THANK YOU