Identity Recognition Using One-Shot Learning and Model-Based Approach

Aravind Chandradoss
Ohio State University

Introduction:

In this report, the results of Identity (Facial) Recognition of 100 people with various occlusion and lighting using One-Shot learning and model-based (PCA) method is discussed. For this study, I used AR-Face dataset, which includes 26 images for each 50 male and 50 female identities with various lighting and occlusion taken at two different sessions (13 images each) as shown below.

I tried two approaches,

- one shot learning (Siamese network)
- model based (Correlation)

Note that the images in the dataset are already aligned. Nevertheless, If needed this pre processing step of aligning the faces can be done using MTCNN - face detection [2] as done in FaceNet [1].

In One shot learning approach, I tried Siamese network approach, and used FaceNet pre-trained on VGGFace2 and CASIA-Webface for embedding the input image. The embedded feature (512 feature vector) is used for classification. Since the model was trained on triplet loss, I used distance (absolute dist and square dist) as the metric for classification (one can easily infer this from T-SNE plot as shown in fig *).
For evaluation, I considered 4 cases, and each case is repeated 10 times with random test image for all 100 identity.

Note that, the image of format "*-001-*.bmp" is image without occlusion and lighting effects.

CASE A1: I used leave one out approach and used all the remaining 25 images for comparison. The test image is picked randomly with varying lights and occlusion. I got **100% classification accuracy** on all 100 identities. The reason for this can be inferred from T-SNE plot (Reduced PCA-2D plot is also shown).

CASE A2: I used only one image as reference (Infact, I used *-001-*.bmp) and did the same. I got **accuracy around (91-94%).** The confusion matrix is shown below.

CASE A3: Session-Based, I picked random test image with varying lighting and occlusion from 2nd session and used all the images from 1st session as reference. I got **87-95% accuracy.**

CASE A4: Session-Based, Similar to CASEA3, but here, I used only one image from the 1st session as reference. **I got accuracy of 88 - 93%**. plots are shown.

Model Based: In this approach, model based on correlations in the input faces is used the model and reconstruct error is used as the metric for classification. The model is generated for each identities with varying lighting and occlusion, using PCA. Then, the correlated features is used for classification (reconstruction error). For this part, I mainly focused on analyzing the effects over the two session. The model was generated using embedding features for the first session, and test image are taken from second session.

Similar to one-shot, in model based, I tried 4 cases.

CASE B1: With all eigen bases as the model for reconstruction. I got accuracy around **95-99%.**

CASE B2: With only the first 10, I got around **97-98%.** Confusion matrix and results are shown.

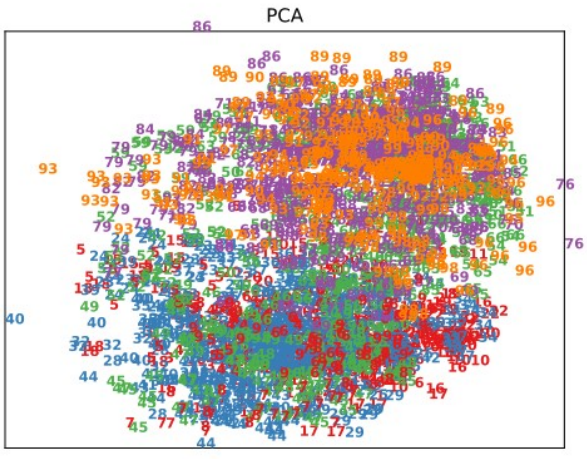
CASE B3: With first 5 vectors, I got around **95-98%** classification accuracy.

CASE B4: With first vector along, I got **94-97%** classification accuracy. The reason is because, the network is trained with triplet loss, there for, the embedded feature are can be well inferred with just one principle vector (linear combination)

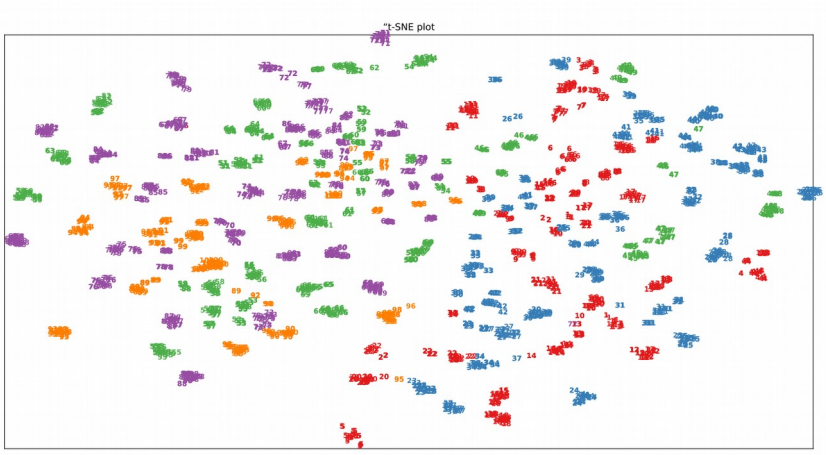
RESULTs:

Images are in svg format (feel free to look at attached file)

PCA 2D plot (less interpret able)



TSNE plot



From TSNE Plot, We can clearly see the each identity is well separated in high dimension (512 dimension). This is reason for high accuracy in CASE A1 (just the distance metric gave the closest match).

NOTE: Each cluster in the plot represents each identity.

RESULTS and INFERENCES:

```
Idx:  37  38
Idx:  32  43
Idx:  63  62
Idx:  73  72
```

Reason for CASE A1 to give high accuracy. ("Idx: test img idx, matched img idx"). We can see that algorithm picked the most similar image (in fact the adjacent image) for identification. This is not the case, when I used Single image (*-001-*.bmp) or Model for classification.

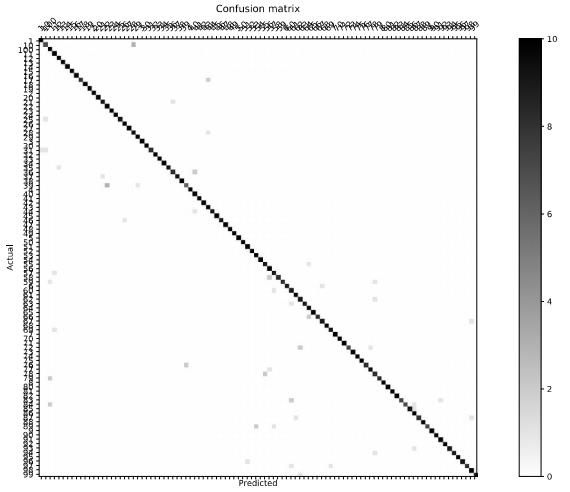
The results for each cases as confusion matrix (Images might be small (have to plot for 100 identities), feel free to refer the attached svg file for more details)

CASE A1: Not soo great plot (got 100% classification, refer TSNE plot)

CASE A2:

Test: Random (both sess)
Ref: Single image

```
[100 rows x 100 columns]
correct: 938 incorrect: 62 = 0.938 %
```

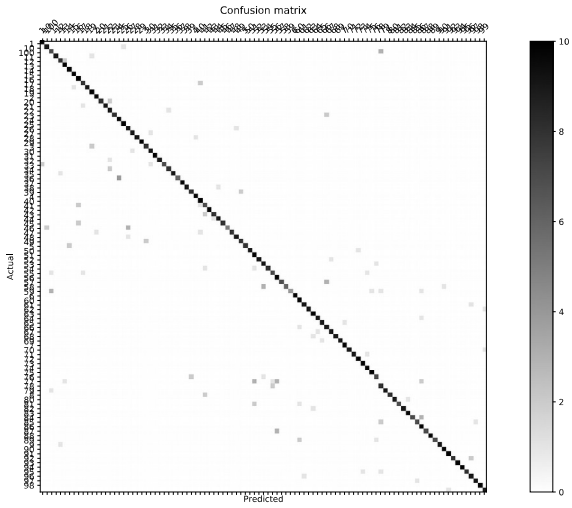


CASE A3:

Test: 2nd sess (random)
Ref: 1st sess (all)

Acc= 87 - 95 %

```
[100 rows x 99 columns]
correct: 870 incorrect: 130 = 0.87 %
```

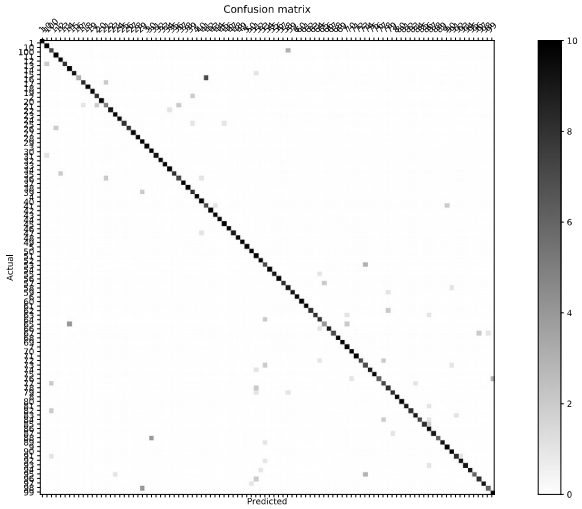


CASE A4:

Test: 2nd Sess (random)
Ref: 1st sess (single image)

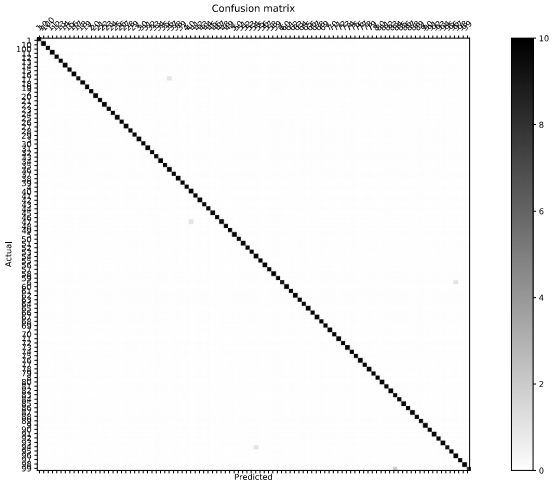
Acc = 88 – 93%

```
[100 rows x 100 columns]
correct: 888 incorrect: 112 = 0.888 %
```



CASE B1:

```
[100 rows x 100 columns]
correct: 994 incorrect: 6 = 0.994 %
```



Refer attachment to see the confusion matrix
Here, I have shown only the accuracy results

CASE B2:

```
[100 rows x 100 columns]
correct: 988 incorrect: 12 = 0.988 %
```

CASE B3

```
[100 rows x 100 columns]
correct: 959 incorrect: 41 = 0.959 %
```

CASE B4:

```
[100 rows x 100 columns]
correct: 941 incorrect: 59 = 0.941 %
```

References:

[1] Florian Schroff, Dmitry Kalenichenko, James Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustring https://arxiv.org/pdf/1503.03832.pdf”

[2] Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10):1499–1503.