Face Image Quality Assessment for Face Selection in Surveillance Video using Convolutional Neural Networks

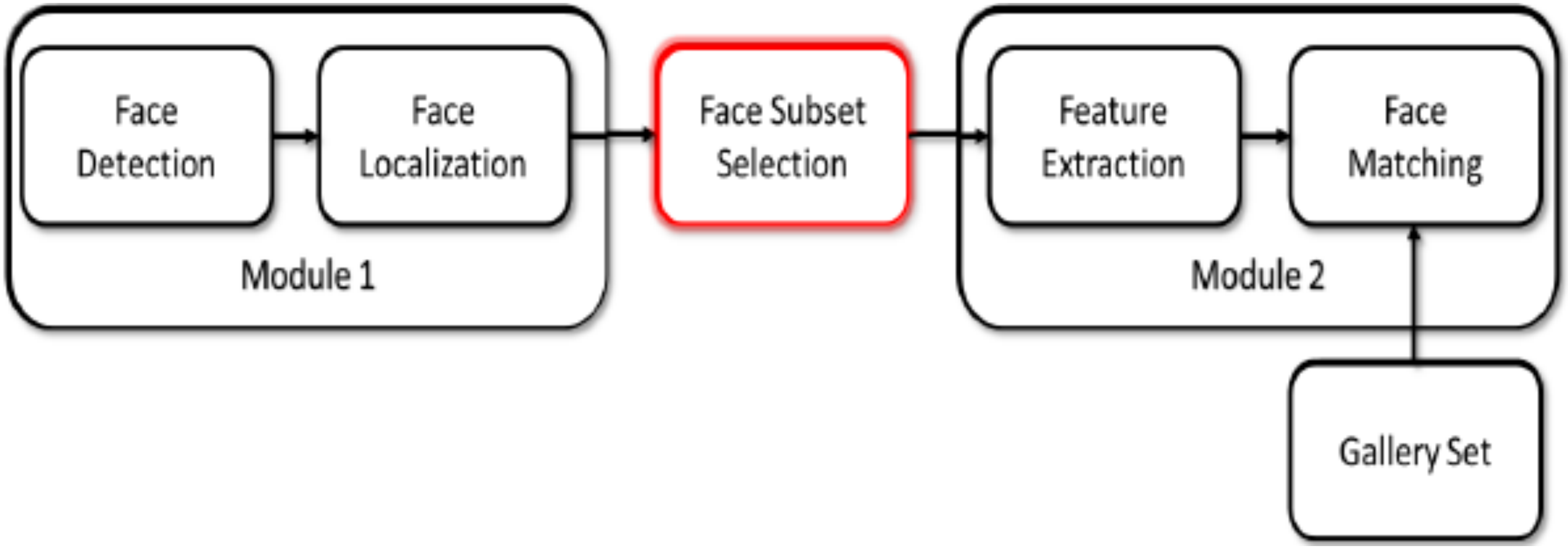
Vignesh Sankar, K. V. S. N. L. Manasa Priya, Sumohana Channappayya

Indian Institute of Technology Hyderabad

Literature Review

- Berrani et al - Statistical approaches
- Wong et al - “Standard” face
- Chen et al. - Multiple feature fusion and learning to rank

FACE RECOGNITION SYSTEM



Face Detection and Localization

- Viola Jones Haar feature based cascade classifier.
- Localization was done and resized to 64x64 pixels.

Face Feature Extraction

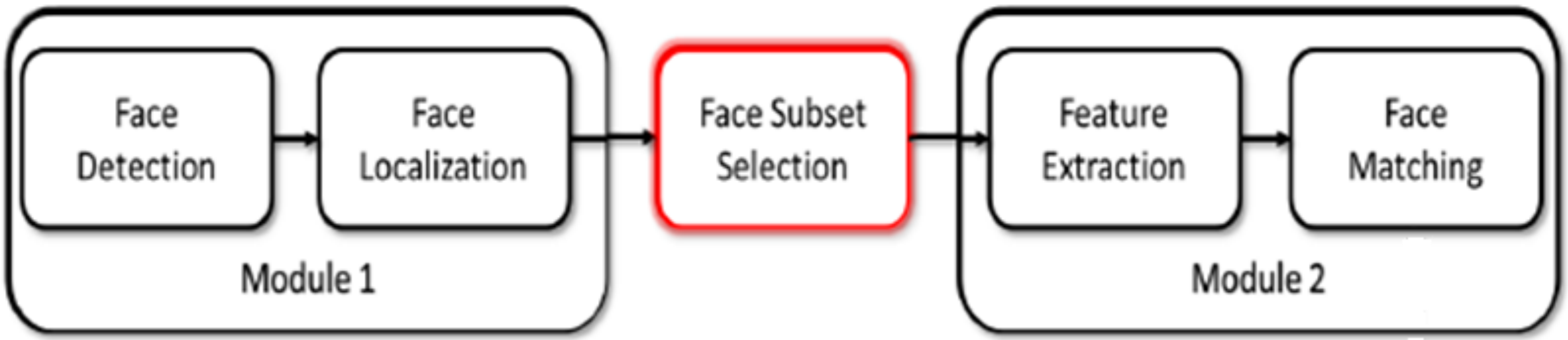
- Feature vectors are formed from the high quality face images given by subset selection.
- Two Feature Extraction techniques,
  - Local binary patterns (LBP)
  - Histogram of Oriented Gradients (HOG)
- LBP makes use of both shape and texture information.
- HOG makes use of the distribution of intensity gradient or edge directions to describe a image.
- These two form a reasonable subset to prove that the developed FQA algorithm works across the different FR algorithm

Face Matching

- Mutual Subspace Method (MSM) is used for image set matching.
- For each video sequence in the gallery set, feature vectors are calculated and compared with feature vectors of probe sequence using MSM.
- The two image sets are considered similar if the canonical angle between two image sets is within the threshold.
- This threshold is calculated from minimum error rate.

Face Subset Selection

- Different Face Recognition (FR) algorithms have different advantages.
- Fixed definition for determining the quality of face image doesn't take the full advantage of the given FR algorithm.



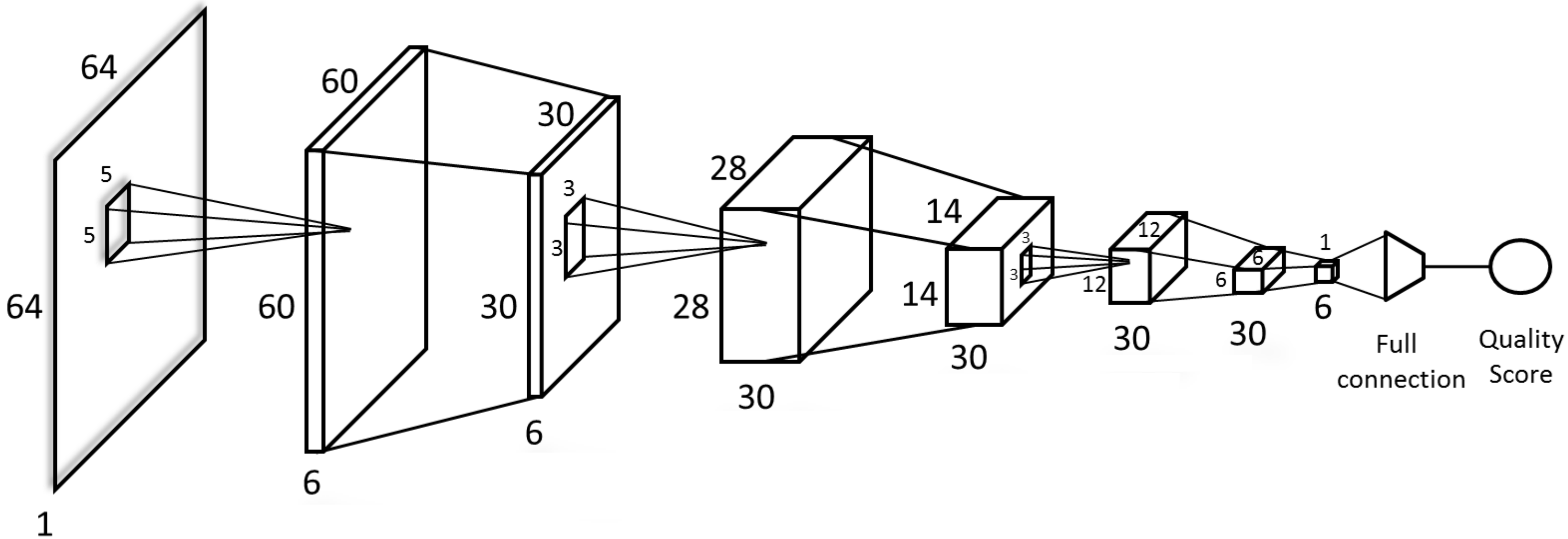
Face Subset Selection

- Goal is to **select** the faces images that perform best in the second module and are considered as high quality face images.
- Quality of the face image is defined with respect to the FR algorithm.
- Convolutional Neural Networks (CNN) is used to model the performance of the FR algorithm.

Why CNN ??

- CNN accepts the entire 2D image as input so there is no need of using explicit feature extraction.
- This is useful in our case where the definition of quality of face image is not fixed.
- CNN learns its parameters and defines the quality of the face image depending on the FR algorithm.

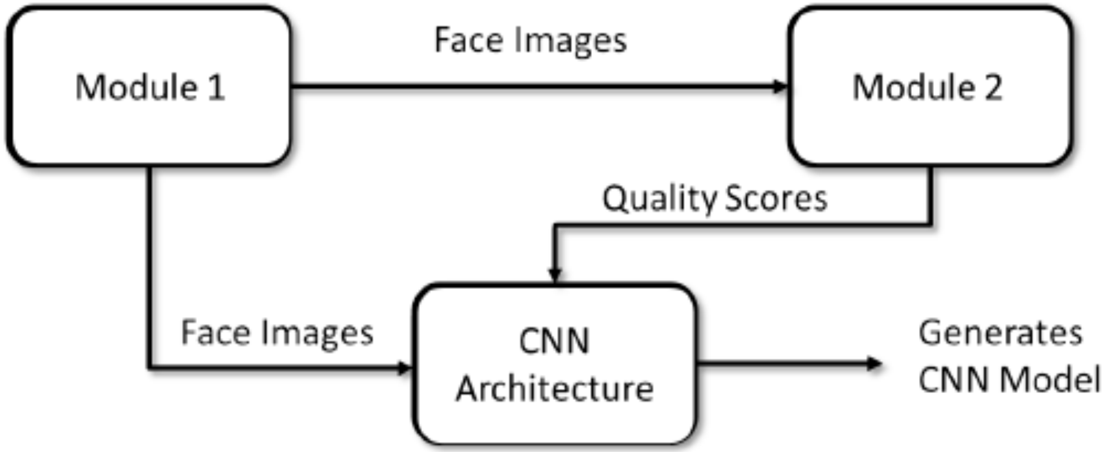
EXPERIMENTAL SETUP FOR FQA



EXPERIMENTAL SETUP FOR FQA

- ChokePoint dataset which is ideally suited for face recognition/ verification in the surveillance scenario is used.
- The dataset consists of 25 subjects (19 male and 6 female) with 64,204 face images
- We divide the images into training and testing sets.
  - Set 1 contains image sequences of 13 subjects for training the CNN
  - Set 2 contains the rest of the images sequences to evaluate the performance of the FR algorithm.
- PCA whitening is done to reduce the redundancy and correlation in the input face image.

Training



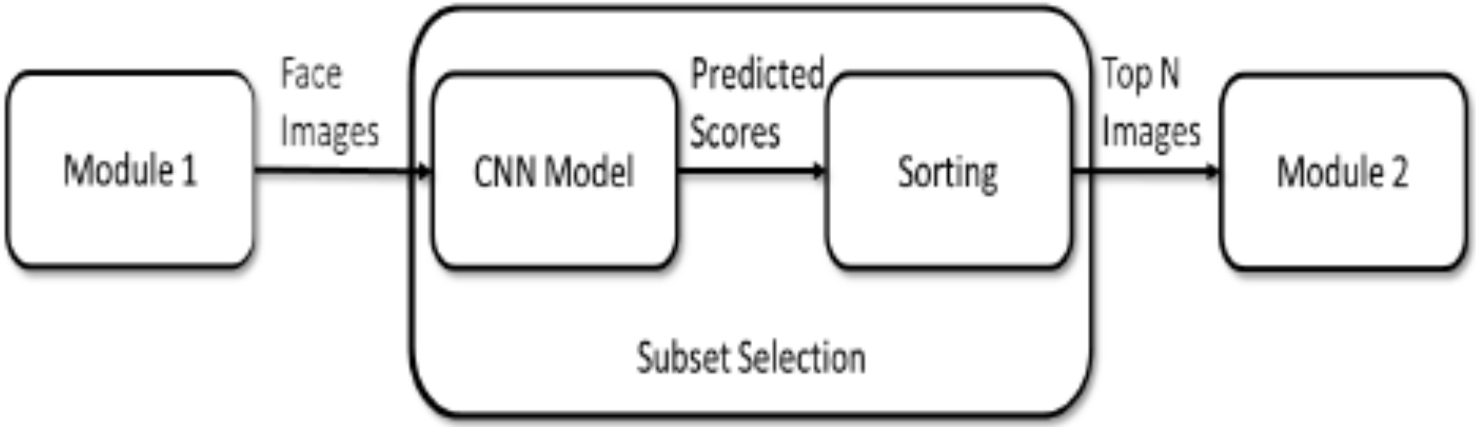
- Face image with the corresponding quality score (MSM score) is used to train the network.

$$S = \frac{1}{N}\sum_{n=1}^{N}\left\| f(I_n, W) - Q_n \right\|_2^2 \qquad W' = \arg\min_{W} S$$

where $\boldsymbol{I_n}$ be the pre processed input face image, $\boldsymbol{Q_n}$ be the quality/MSM score, $\boldsymbol{W}$ is the weight matrix of the network, $\boldsymbol{f(I_n, W)}$ be the predicted score of the input face image

Testing

- Pre processed face images in the probe sequence is given as the input to the trained CNN.
- Quality scores are predicted for each face image in the sequence and sorted to get **Top N** high quality images.



Experimental Setup OF FR

- Testing set is divided into G1 and G2 where each dataset plays the role of development and evaluation sets respectively.
- By considering one group as development set (labeled set), we calculated matched and mismatched scores.
- Threshold is where the sum of False Acceptance Rate (FAR) and False Rejection Rate (FRR) is minimum i.e., Minimum Error Rate.
- Applying this threshold on the scores of pairs of evaluation sets, recognition rate (R_{G2}) is calculated as follows

$$R_{G2} = 0.5 \times \left[(1 - FAR) + (1 - FRR) \right]$$

- Then roles of G1 and G2 are changed to calculate R_{G1}
- Final Recognition rate is the average of RG1 and RG2.

$$R_{avg} = 0.5 \times \left(R_{G1} + R_{G2} \right)$$

Results

Video-based face verification performance on the ChokePoint dateset, using LBP and MSM

Subset Selection Method	N = 4	N = 8	N = 16
Sequential	0.6114	0.6174	0.6278
Random	0.6825	0.6910	0.7040
Probabilistic Based	0.6995	0.7181	0.7252
Rank Based	0.7328	0.7511	0.7645
Proposed Method	0.7226	0.7564	0.7786

Video-based face verification performance on the ChokePoint dateset, using HOG and MSM

Subset Selection Method	N = 4	N = 8	N = 16
Sequential	0.6419	0.6504	0.6669
Random	0.7329	0.7552	0.7706
Probabilistic Based	0.7603	0.7753	0.7876
Rank Based	0.7843	0.7870	0.7857
Proposed Method	0.7589	0.7775	0.7917

Conclusion

- From the results, we can infer that high verification performance is achieved by the proposed method which in turn implies that it is able to select the best subset of faces from the sequence of faces.
- From the initial results, we strongly believe that the proposed algorithm is promising and has attractive features.
- As part of future work, we plan to improve the performance of the algorithm by fine-tuning the parameters.