VIDEO FACIAL RECOGNITION: SPEED MATTERS

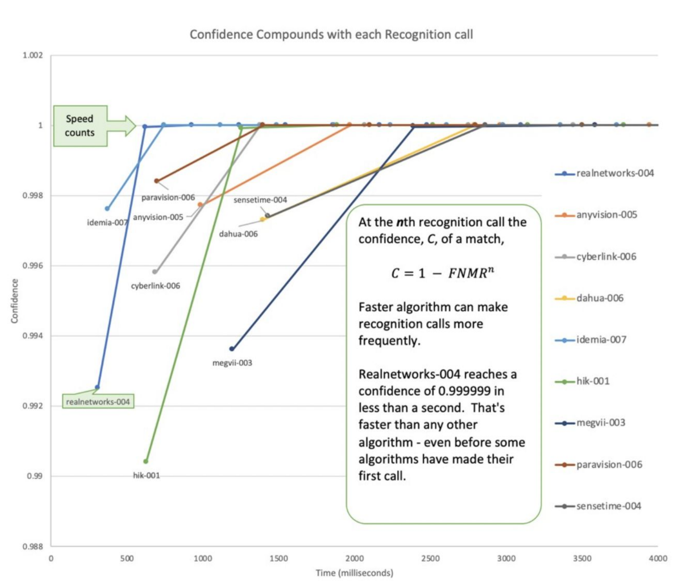
STUDENTS: SUNGCHAN PARK, LEONARD SHIN, VICTOR LI

realnetworks

25

Introduction

- Face Recognition has progressed from high-powered off-line search to ubiquitous IoT devices.
- Current FR evaluations (from NIST) measure the ability to match a single image against a reference database.
- This project aims to redefine FR accuracy for real-time video and prove the hypothesis that accuracy increases with successive frames of a live video feed.
- The two hypotheses we aimed to test were 1) That frames can support each other regardless of ordering. (This is the hypothesis displayed on this poster), and 2) That a frame can exclude members, reducing the search pool for subsequent frames.



Project Requirements

- The code consists of an input module to input a video frame by frame, and a comparison module to execute the main function of the code which is to aggregate the recognition scores found over the sequential video frames
- The input identity database for testing should exceed 10,000 unique identities.
- Number of testing videos: Estimated around 1000. Majority of them are human facial images, including multiple faces presented in the same pictures
- The team should utilize the RealNetworks SAFR SDK and consumer algorithm to test this hypothesis and substantiate their claims
- The team should measure the similarity score, confidence, and FNMR to analyze results.
- Maximum Process time: 5 images per second for application

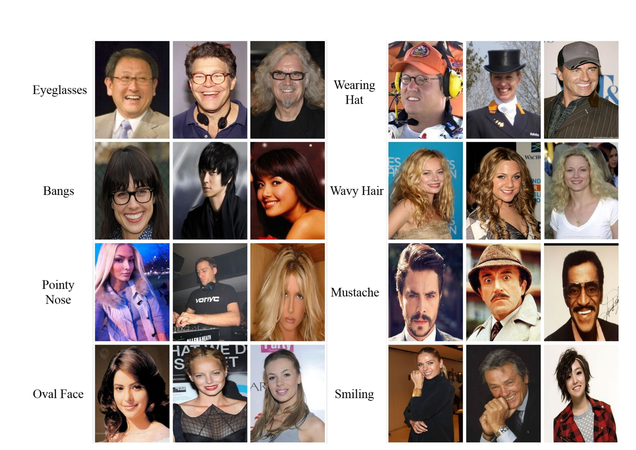
Datasets

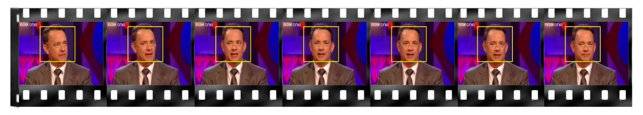
Identity Datasets

- Combined dataset of over 15,000 unique individual faces from the following databases
- 10k US Adult Faces Database [1]
- Chicago Faces Database [2]
- FACES [3]
- CelebA [4]
- Labeled Faces in the Wild [5]

Video Testing Dataset

- Original Dataset is the YouTube Faces Database [6] labeled through 2D and 3D keypoints [7]
- Consists of 2200 test videos of 800 unique individuals





Same / **Not Same** ?

System Implementation

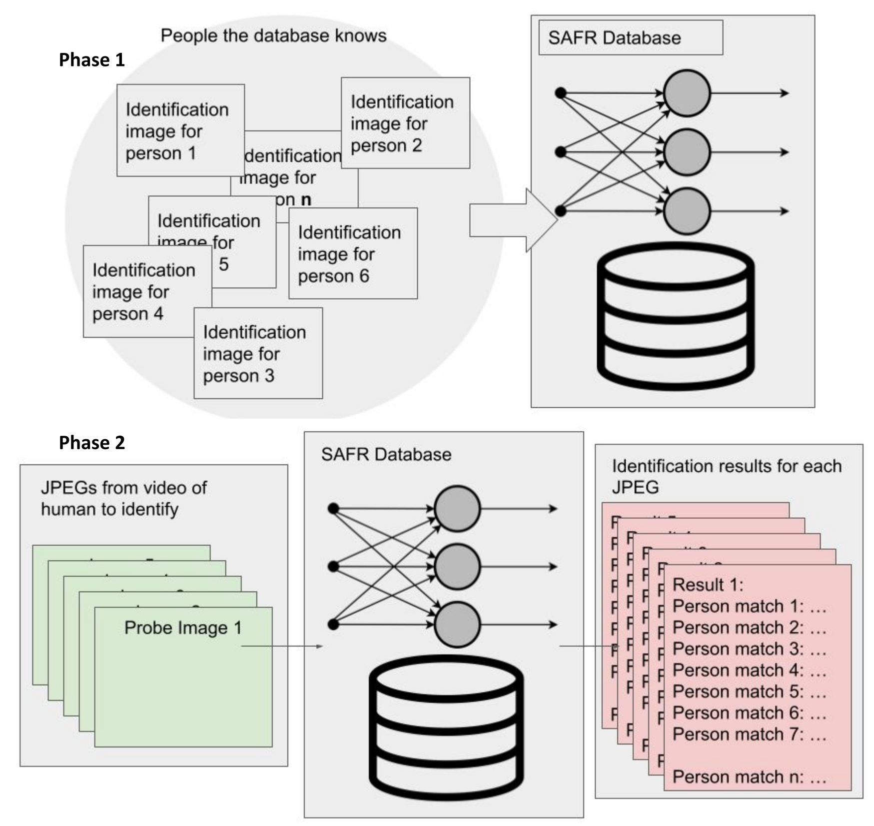
- *Speed matters* aims to determine if recognition values from previous video frames can increase confidence for identification of people in future frames.

Phase 1:

- Identity images are placed into the SAFR database to be used as reference images. The SAFR database quantifies these identification frames as vectors.

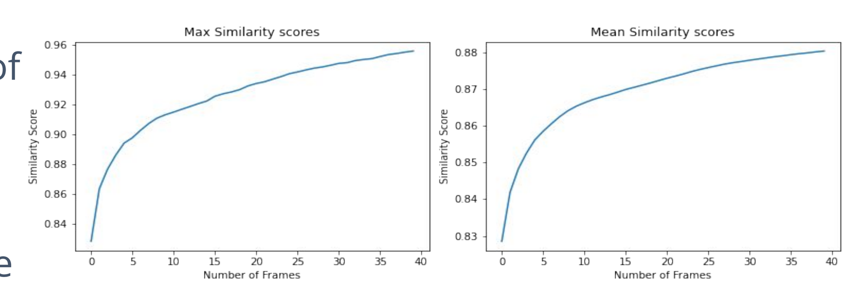
Phase 2:

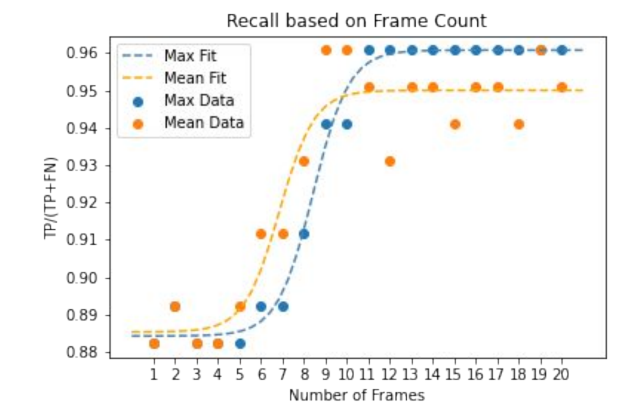
- Video is decomposed into sequences of probe JPEG frames.
- Probe JPEG frames are provided one at a time to the SAFR database for identification and the identification results, such as the best **n** face matches are recorded.
- The SAFR database creates a machine learned vector quantifying the face in the probe image and compares this vector against the identity vectors
- These results are then analyzed to see if information from successive frames can improve recognition accuracy.

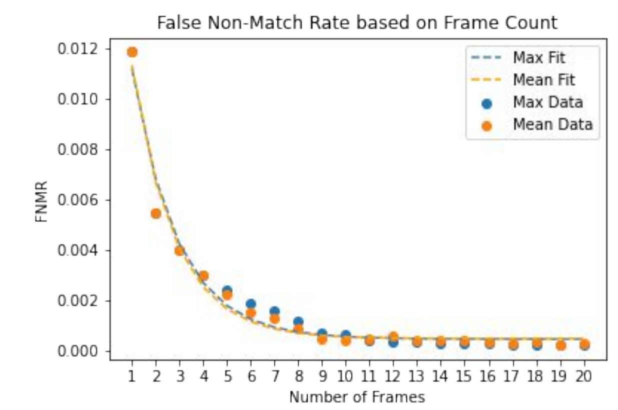


Results

- Hypothesis: Accuracy of the model would increase proportionally with the number of frames utilized.
- The result selection was determined through the usage of selecting the image with the maximum similarity score and the identity with the greatest mean similarity score over that frame set.
- While individual runs produced different results for the two methods, on average the pattern was similar between the two.
- The middle graph shows the increase in recall based on frame count.
- The recall (or identification percentage) was a calculation showing the ratio of correct outcomes over trials run: True positives / (True positives + False negatives)
- The precision data seemed to approximately fit a sigmoid curve.
- A graph showing the decreasing false non-match rate based on the frame count
- Both methods produced an approximate negative exponential relationship of FNMR = 0.018 * $e^{(-0.5 * x)}$ + 0.00044







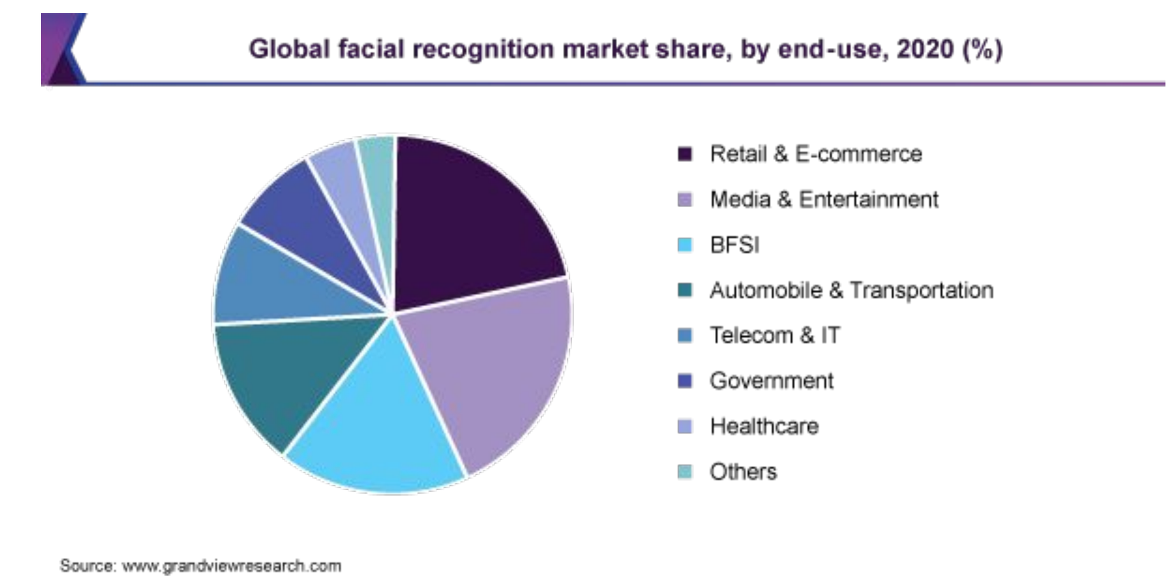
Conclusion and Project Future

Our hypothesis is confirmed that accuracy of the model increases with the number of the frame utilized. The precision of the model appeared to have a sigmoid relation. Furthermore, the False Non-Match Rate (FNMR) had also decreased with the increasing number of frames, while both max and mean similarity score increases. These appeared to decrease in an exponential manner. In short, our experiment results show that our hypothesis of increasing accuracy with increase in number of frames stands true .

Video Facial Recognition allows user identification to be faster and more efficient. There are fields where facial recognition shows promising future and prospect, such as security verification, banking and retailing, resolving queuing problems in hospitals and public areas, and more.

Our future objectives and milestones are listed below:

- In the future of the project development, we will expand the size of the image gallery to 100K images, and then 1 million images.
- We plan to open the application to 3rd party in order to test their own algorithm, as well as giving beta test to the algorithms we had developed.
- In addition, we will develop a strategy to interlace the face detection and recognition that yields the lowest FNMR value.
- We also plan to develop a gallery of database that does various strategies to speed up identity matching, such as sorting, indexing, and sharding
- We will also be experimenting with other potential approaches to speed up the recognition process, and finding the optimal facial recognition strategy based on the the FNMR values each strategy delivers.



Acknowledgments and References

We would like to say a special thanks to the following people
Faculty/TA: Arka Majumdar, Payman Arabshahi, Jill Kaatz, Kelly Ho
Industry Mentors: Reza Rassool, Milko Boic, Steven McMillen

[1].Minear, M. & Park, D.C.(2004). A lifespan database of adult facial stimuli. Behavior Research Methods, Instruments, & Computers, 36, 630-633.
[2].CFD: Ma, Correll, & Wittenbrink (2015). The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data. Behavior Research Methods, 47, 1122-1135. https://doi.org/10.3758/s13428-014-0532-5.
[3] Holland, C. A. C., Ebner, N. C., Lin, T., & Samanez-Larkin, G. R. (2019). Emotion identification across adulthood using the Dynamic FACES database of emotional expressions in younger, middle aged, and older adults. Cognition and Emotion, 33, 245-257. doi:10.1080/02699931.2018.1445981.
[4]Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. 2015 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2015.425
[5] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. University of Massachusetts, Amherst, Technical Report 07-49, October, 2007.
[6] Lior Wolf, Tal Hassner and Itay Maoz. Face Recognition in Unconstrained Videos with Matched Background Similarity.IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2011.
[7] Bulat, A., & Tzimiropoulos, G. (2017). How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks). 2017 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2017.116

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY of WASHINGTON

ADVISORS: KELLY HO, ARKA MAJUMDAR, REZA RASSOOL, MILKO BOIC, STEVEN MCMILLEN

SPONSORS: REALNETWORKS