Multimodal Biometric Fusion using Facial Expressions and Voice Recognition

Jessica Teeslink (SDSM&T)

Faculty Advisor: Dr. Randy C. Hoover

SPACT — Center for Security Printing & Anti-Counterfeiting Technology

SOUTH DAKOTA SCHOOL OF MINES & TECHNOLOGY

Introduction:

Reliable end-user verification has become a drastic issue in not only the United States, but the ever changing world today. The ability to identify correctly users via log-in credentials, while rejecting attempts to spoof such credentials becomes more and more challenging each day.

- Reliable end-user verification generally relies on biometric markers
    - Face, voice, iris, fingerprint, etc.
- Combining such modalities has the potential to increase the robustness of such approaches.
    - The current work focuses on the fusion of two such modalities
    - facial region (mouth movements) and the corresponding voice from such movement.

## Procedure:
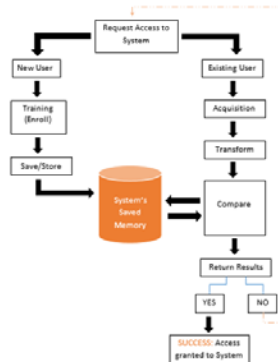
This system has two components: facial and vocal.

- Facial Expressions:
    - Focuses on user's particular lip movement during a specified phrase.
    - During the prompted time to recited the phrase a series of photos are captured of the users lip region.
- Voice Recognition:
    - Records prompted phrase
    - Transforms between time to frequency
- Data drawn is compared to the data stored in the specific user profile.
    - Produces a system that would require the user to pass both of the obstacles to be allowed access.

Voice Recognition:

When the new user begins training the system, enrolling as a user, the system will prompt the user to recite a phrase 3 times (shown on the right).

- Upon entering the system the user is prompted to recite a predetermined phrase
    - been previously saved in the system during the training phase.
- Acquision data captured from microphone
    - originally time domain function (Figure 1).
    - To create a more secure system, transformed into a more unique set of data by taking the fft (Fast Fourier Transform, Figure 2).
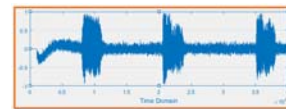
- Comparing between the saved data in the user's file and the proposed user's data, that was drawn upon entering the system
    - Cross-Correlation ran on the data.
    - Returns, with 95% confidence, the percentage of correlation between the two sets.
    - If the value returned falls above 80% the voice recognition half of the system verifies a success on authentication.

Facial Expressions:

Simultaneously, the camera takes multiple snapshots of the user. Images are:

- Transformed (row-scanned to an n-dimensional vector in image space).
- Concatenated temporally throughout the sequence.
- Principle Component Analysis (PCA) is then used to reduce the dimensionality of the data for later verification.
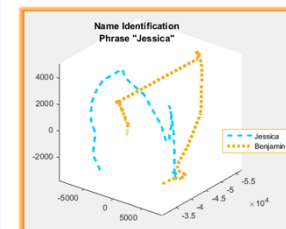
The image to the left show the characteristics of Jessica saying 'Jessica' and Benjamin saying 'Jessica'.
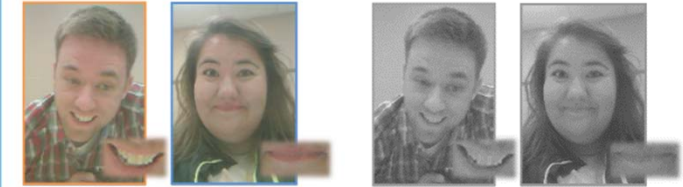
The camera on the system captures an image in color, shown on left, and *transforms* from RGB to grayscale. We then crop the image to keep only the lip region while the training phrase is recited.

Results:

The entire fusion algorithm was developed using MATLAB. The system is able to enroll new users, transform the data into a comparable file for both speech and video, and use these files for end-user verification in real-time.

Conclusion:

Voice recognition:
- program passes the user with success if 80% correlation between the two sets of data is calculated. That number, for a more secure system, would be higher.

Facial expression:
- While we're hopeful, the facial fusion is still in its infancy and will be the focus of our future work.

Future Work:

- Furthering the recognition on the facial aspect, as well as, the voice recognition to a higher percentage than 80% to authorize verification.
- Product development for an end-user platform.