Deepfake Video Detection with Real Time Scene Authentication

Kelly Yen, Hadleigh Schwartz, Dr. Xia Zhou
Department of Computer Science, **MobileX Lab,** Columbia University

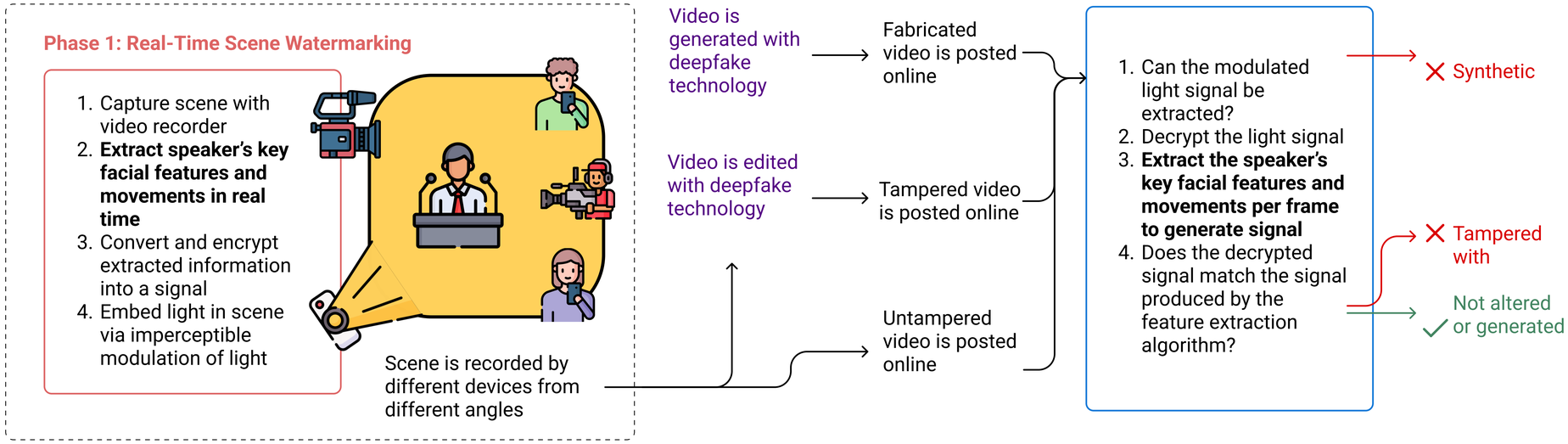
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Background

- "Deepfake" refers to media that has been generated by artificial intelligence, typically an artificial neural network.
- Recent deepfake technology is capable of generating video content that is indistinguishable from real media, introducing a new class of security threats.
- Efforts to distinguish deepfake videos from authentic videos typically involve training a secondary model to recognize specific artifacts or physiological cues.
- Many deepfake generating technology evolves quickly enough to circumvent traditional countermeasures
- To avoid a deepfake generation vs. detection arms race, we propose a detection solution based on scene authentication

Proposed System

Scene: A well Known individual delivers a speech



SURE Summer 2023 Project: Feature Extraction and Processing Pipeline

Goals

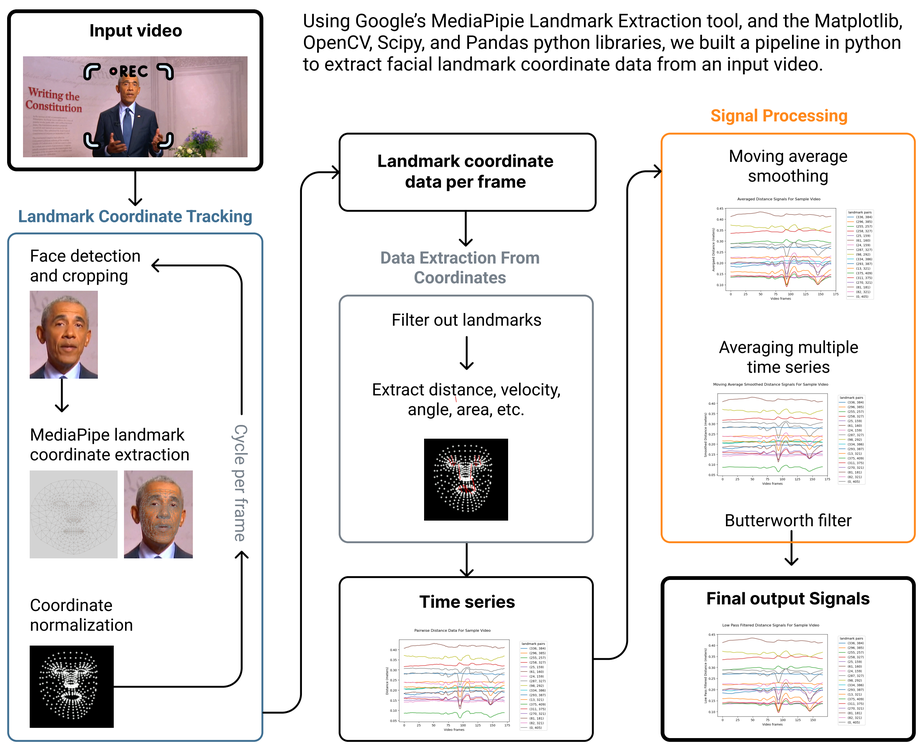
Develop a feature extraction and processing pipeline for use in both scene watermarking and video authentication phases.

The correlation table below demonstrates desired outcomes for different comparison scenarios:

	Correlation results
Person B vs. Person A saying "I am the president of Columbia University"	Low correlation: Differentiate between identities
Person A says "I am the president of Cucumber University" vs. "I am the president of Columbia University"	Low correlation: Differentiate between utterances
Person A says "I am the president of Columbia University" with an angry expression vs. with a happy expression	Low correlation: Differentiate between expressions
Person A says "I am the president of Columbia University" captured by camera 2 vs. captured by camera 1	High correlation: Stay consistent between cameras capturing the same scene

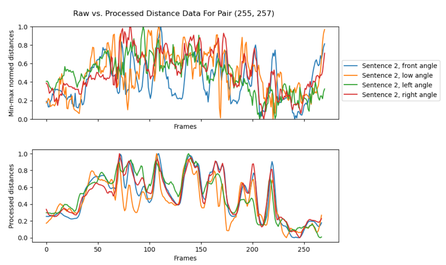
Method

Using Google's MediaPipie Landmark Extraction tool, and the Matplotlib, OpenCV, Scipy, and Pandas python libraries, we built a pipeline in python to extract facial landmark coordinate data from an input video.

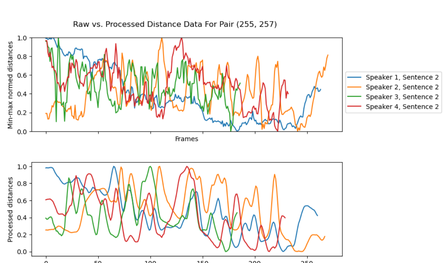


Results

Comparing distance signals of landmark pair 255 and 257 before and after pipeline processing.



Each line represents the same scene captured from a different camera angle. We expect the signals to look the same.



Each line represents a different speaker saying the same thing. We expect the signals to look different.

Future Work

1. **Collect and Run Test Data.** Test out pipeline on video samples from people of different ages, racial identities, and gender identities.
2. **Finalize Landmark Selection.** Find best performing landmarks for differentiating across utterances, identities, and expressions
3. **Finalize Pipeline.** Finalize optimizing the pipeline.
4. **Integrate With Entire System.** Integrate pipeline with the core unit and test out scene and video authentication components.

References

1) Danielle K. Citron & Robert Chesney, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security , in 107 California Law Review 1753 (2019).
2) Rahdari, F., Rashedi, E. & Eftekhari, M. A Multimodal Emotion Recognition System Using Facial Landmark Analysis. Iran J Sci Technol Trans Electr Eng 43 (Suppl 1), 171–189 (2019). doi: 10.1007/s40998-018-0142-9
3) F. Noroozi, M. Marjanovic, A. Njegus, S. Escalera and G. Anbarjafari, "Audio-Visual Emotion Recognition in Video Clips," in IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 60-75, 1 Jan.-March 2019, doi: 10.1109/TAFFC.2017.2713783.
4) Ryumina, E., & Karpov, A. (2020). Facial expression recognition using distance importance scores between facial landmarks. Proceedings of the 30th International Conference on Computer Graphics and Machine Vision (GraphiCon 2020). Part 2. https://doi.org/10.51130/graphicon-2020-2-3-32
5) Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes. ACM Computing Surveys, 54(1), 1–41. https://doi.org/10.1145/3425780