Tarjman: AI-Powered Lip Sync and Dubbing System

Graduation Project Proposal – Digital Egypt Pioneers Initiative

Abstract

Tarjman is an AI-powered system designed to automate the dubbing process for videos by synchronizing lip movements with translated or re-recorded speech. The system leverages computer vision and natural language processing techniques to produce realistic, high-quality dubbed videos across multiple languages. By combining lip synchronization, speech synthesis, and face tracking, Tarjman enables natural and expressive video translation that bridges communication barriers in media and education.

Objectives

- Develop an intelligent system capable of synchronizing lip movements with new audio tracks.
- Enable real-time or near-real-time dubbing for videos.
- Integrate text-to-speech and emotion-aware voice synthesis.
- Enhance multilingual accessibility in digital content.
- Provide a deployable system usable in media, education, and entertainment sectors.

Main Features

Lip Syncing:
Description: Synchronize the speaker's lip movements with the dubbed voice to ensure natural realism.
Tools/Technologies: Wav2Lip, GANs, OpenCV

Film Dubbing:
Description: Replace original speech with translated or enhanced audio tracks.
Tools/Technologies: Librosa, PyAudio

Image Processing:
Description: Preprocess video frames for face recognition and lip sync tasks (lighting adjustment, cropping, normalization).
Tools/Technologies: OpenCV, Pillow

Face Recognition:
Description: Detect and identify faces in each video frame.

Tools/Technologies: Dlib, Face Recognition, MTCNN

Face Tracking:
Description: Track the same face across consecutive frames to maintain processing consistency.
Tools/Technologies: OpenCV Trackers (CSRT, GOTURN), Dlib correlation tracker

Text-to-Speech (TTS):
Description: Generate human-like speech from text for dubbing.
Tools/Technologies: Tacotron 2, WaveNet, Google Cloud TTS , amazon polly

Audio Processing:
Description: Edit and align dubbed audio with the original tone, timing, and emotion.
Tools/Technologies: Librosa, PyAudioEffects, Audacity

Audio-Video Merging:
Description: Merge final processed audio with the video seamlessly.
Tools/Technologies: MoviePy, FFmpeg

Deep Learning & Neural Networks:
Description: Power the core AI models for lip sync and TTS.
Tools/Technologies: TensorFlow, PyTorch

Team Members and Roles

Member Name	Role
Yosra Naser Mansour	Image Processing
Aya Mahmoud Hussien	Lip Syncing
Yara Emad Eldien Sayed	NLP
Ali Ahmed	Face Recognition
Nada Haimn	NLP + Deployment
Menna Mostafa Salah	NLP

Technologies and Tools

- Programming Languages: Python, C#
- Frameworks:, TensorFlow
- Libraries: OpenCV, Dlib, Librosa, FFmpeg
- AI Models: Wav2Lip, Tacotron 2, WaveNet
- Deployment: Docker, AWS

Key Performance Indicators (KPIs)

- Lip synchronization accuracy ≥ 90%.
- Voice clarity and synchronization rated ≥ 4/5 by test users.
- Video processing time ≤ 1.5x real-time duration.
- Multilingual dubbing capability (at least 3 languages).

Innovation Value

Tarjman offers a new approach to automated dubbing by integrating computer vision and speech technologies for lifelike video translation. Unlike traditional dubbing systems, Tarjman provides realistic lip synchronization and emotional expression, creating a more immersive viewing experience.