Enabling Multilingual Communication: Automated Lip-synchronization Dubbing for Albanian Videos

ALDO DIKU

UNIVERSITY OF NEW YORK TIRANA

JULY 2025

This is to certify that I have read this project and that, in my opinion, it is fully adequate, in scope and quality, as a thesis for the degree of Bachelor of Arts in Computer Science.

(Title and Name)
(Project Advisor)   _____   _____

This is to confirm that this thesis complies with all the standards set by the Department of Computer Science of University of New York Tirana.

Date: Seal/Signature:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

First Name, Last Name:

Signature:

ABSTRACT

ENABLING MULTILINGUAL COMMUNICATION: AUTOMATED
LIP-SYNCHRONIZATION DUBBING FOR ALBANIAN VIDEOS

Diku, Aldo.
BA. in Computer Science
Thesis Advisor: Prof. Miralda Çuka
July 2025, 183 pages

Using available methods and tools (which i will mention when i gather all the tools and methods) to create lip-synced dubbed videos from an albanian video input and outputting a video of the same speaker in another language with the lips moving according to that languages movements.

Keywords: Lip-sync, dubbing, albanian language

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Miralda Çuka, for their invaluable guidance, unwavering support, and insightful feedback throughout this project. Their expertise and encouragement were instrumental in shaping this work.

Finally, I am grateful for the support of my family and friends. Something something a little bit longer and better made.

Contents

3   Introduction

3.1   Task description

Option 1

The increasing globalisation of online content consumption presents a significant challange for video creators aiming to reach diverse audiences, thereby creating the need to produce multilingual versions of their work. This often involves an increased workload, time and other resources, a perfect example of this is a German content creator in YouTube called der8auer, who for every english video has an identical german one. This project addresses the challange of automatic translation and lip synchronization of video content into multiple languages. The main focus of the project will be the Albanian language as an underrepresented and low resource language. To bridge a gap in current research and practical applications of automated video translation and lip synchronization.

Option 2

Being able to make only one video and have it in multiple languages will help creators but also the target audience as well. Cutting down on the time and money it takes to make two or more videos with the same subject but in different languages. One such case is a youtube creator named der8auer from Germany, for the same topic he creates two videos, one in english and one in german. Sometimes his videos are quite long and having to do them again in another language will tire you out. Having a model that can take your video, translate it into another language and synchronize your lips to the movement of the language of your choosing would be a tool in your arsenal.

This project focuses more on the Albanian language as a low resource and underrepresented language, addressing a significant gap in current research and practical applications.

3.2   Separation of work

The first step for this project was the data creation, seeing as there is not a lot of free and publicly available free-use content I can use we had to create the data myself.

While there may a lot of such videos availabe in YouTube, we need to ask for permission first before having to use their videos, having to wait and then we might face rejection.

After we have the video data we need to get the transcription of the videos, for this to be done automatically we cannot create the videos basing them on a transcription. Several models (citations needed) were available to try for albanian transcription although with different accuracy. Automatic Speech Recognition (ASR) models are measured using Word Error Rate (WER) where a score of 0% means perfect translation and 100% is the worst case, working on the albanian language I focused only on those models that had support for it. Whisper by OpenAI was one such model, availabe on HuggingFace. To make such model have a better performance (need some results of the model transcribing an audio file) we need to fine-tune it. Here lie another issue with working with a low resource language, not only are there low resources for video, but also for audio which we could use to fine-tune the model.