
AUTOMATED STEP-BY-STEP VISUAL PERFORMANCE GUIDE GENERATION FROM SINHALA DEMONSTRATION VIDEOS

Supervisor: Dr. L. Ranathunga

Group Members: 204104H

204137K

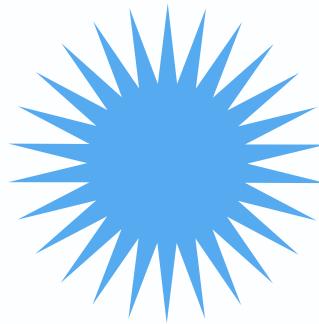
204041K

Kularathna M.D.S.A.

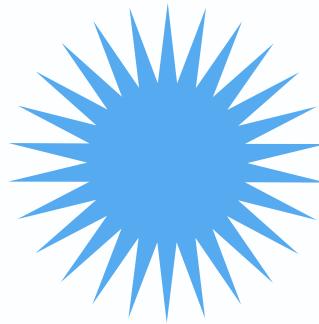
Nethmini S.A.R.

Dilshan K.G.A.P.

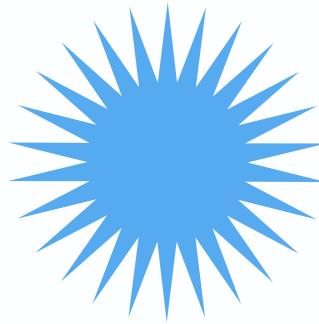
01. INTRODUCTION



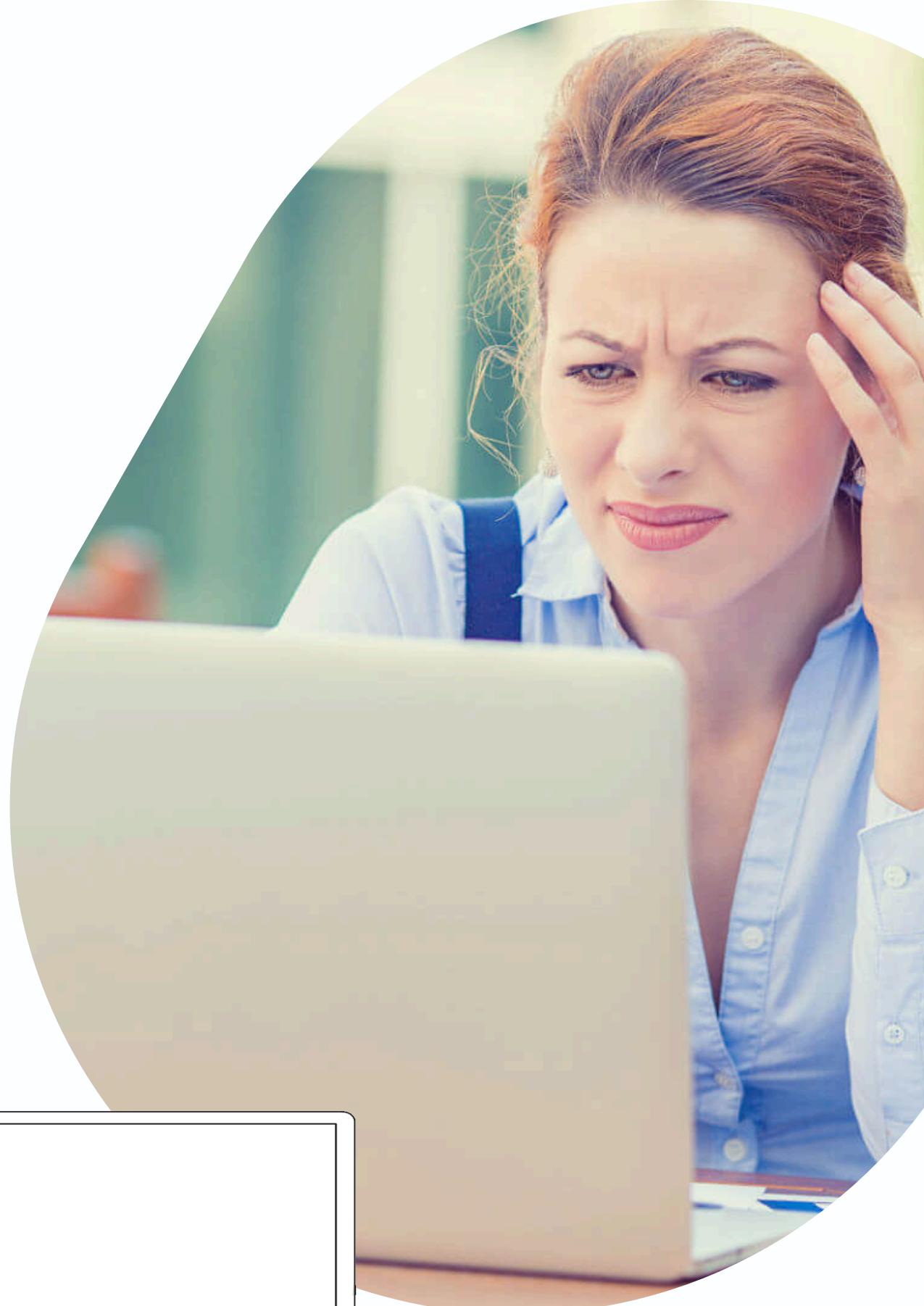
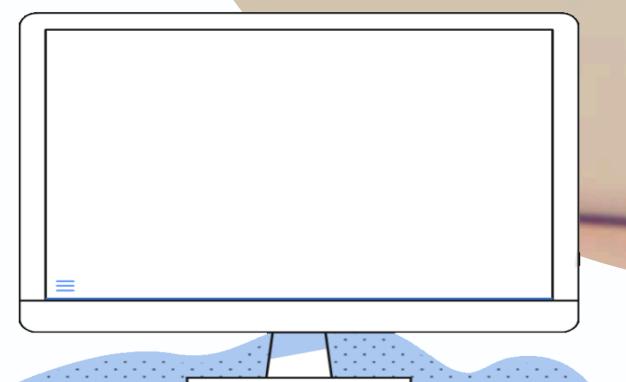
Many Sinhala demonstration videos lack performance guides, making it challenging for viewers to understand key details.



Our project focuses on solving the above problem by generating visual performance guides from demonstration videos.



Key domain areas include Natural Language Processing, Machine Learning and Image/Video Processing.





02. PROBLEM IN BRIEF

Viewer Frustration: Users have to rewind and skip forward repeatedly to locate important details in the video.

Accessibility Issues: Difficulty in quickly finding and understanding key information due to excessive filler content.

No Extraction Tools : There are no options to extract key information from Sinhala demonstration videos.

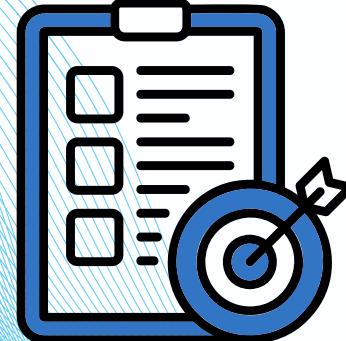
03. AIM & OBJECTIVES

AIM

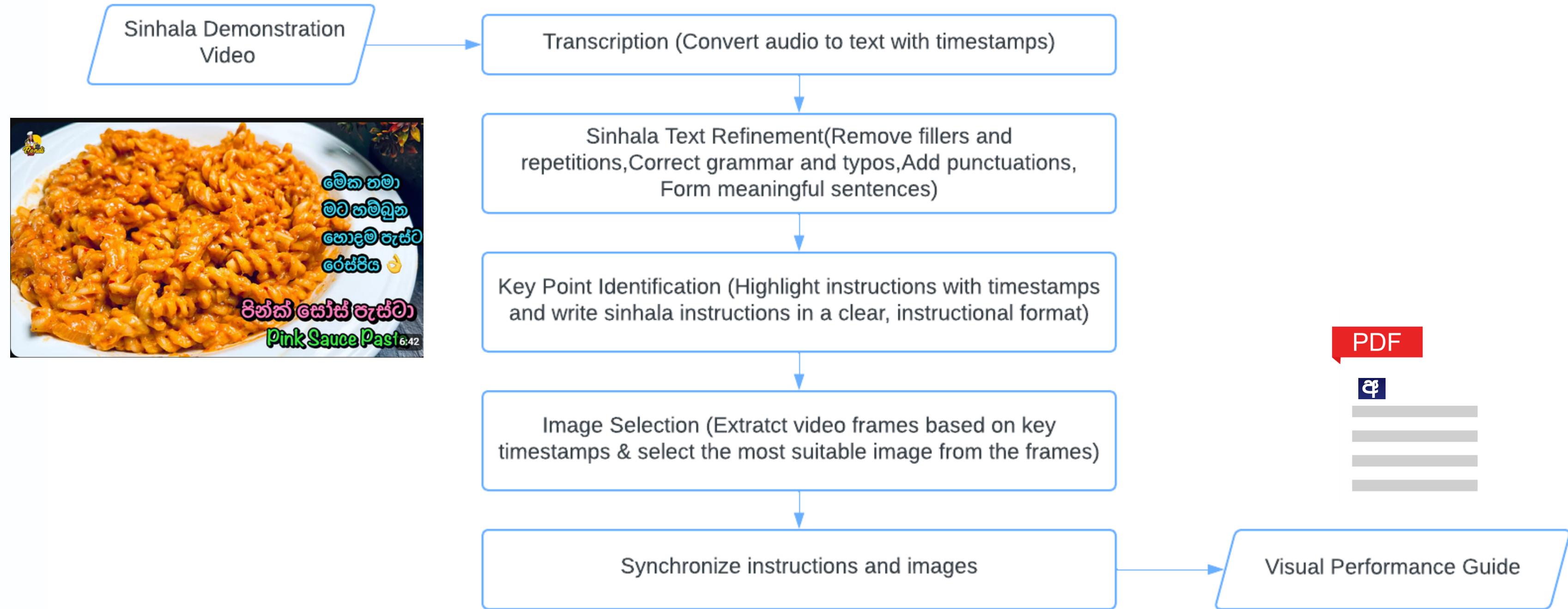
To automatically generate visual performance guides from Sinhala demonstration videos, allowing viewers to quickly grasp key information through synchronized text and visuals.

OBJECTIVES

- 01 Convert Sinhala audio to text, correct it and generate meaningful content.
- 02 Identify key points and create step-by-step instructions.
- 03 Capture and enhance key visuals to focus on important steps.
- 04 Generate a document by combining relevant visuals and instructions.
- 05 Enhance text accuracy, image selection, and performance guide quality using machine learning techniques.
- 06 Evaluate and benchmark the success factor for the algorithms used for the above operations.

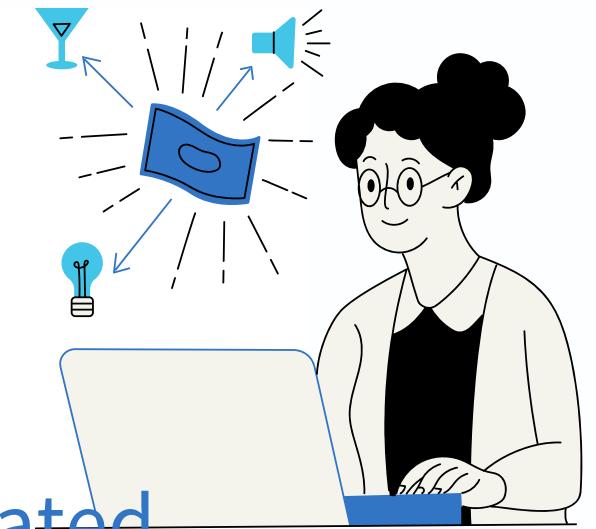


04. CONCEPT MAP OF THE PROJECT



05. SOLUTION IN DETAIL

This solution will focus on demonstration videos in the cooking and electronics-related domains. Videos considered will have a maximum duration of 10 minutes.



Correction of Transcriptions to enhance Clarity and Accuracy:

- Automatically transcribe the Sinhala audio while recording timestamps.
- The Transcribed text has lot of errors which may interrupt the process of extracting key information from it. Therefore we consider the following to solve this problem.

Eliminate filler words.

Remove repetitive words and phrases.

Correct grammar and spelling errors.

Insert punctuation to define sentence breaks.

05. SOLUTION IN DETAIL



Key Point Identification and Instruction Generation:

- The corrected transcribed text may have non-essential information that don't directly contribute to the main instructions .
- Key points must be identified along with their relevant timestamps in the text.
- Present key points as instructions in proper writing language.

05. SOLUTION IN DETAIL



Image Selection Aligned with Instructions:

- Video processing is used to extract frames from the video clip corresponding to the key point timestamps.
- A suitable image for each instruction is selected from the above mentioned frames using machine learning techniques.
- Synchronize the images and instructions to generate the visual performance guide.

06. PROJECT APPROACH

Technologies:

- Transcription: Use open-source API's for transcription with timestamps.
- Natural Language Processing libraries
- Video Processing: OpenCV for frame extraction and quality enhancement.
- Machine Learning frameworks for training models to improve output quality.



07.

PROJECT REQUIREMENTS

- **Data set**

Visual Performances guides we make manually.

- **Hardware requirements**

Laptop: HP Omen -16 GB RAM ,2.6 GHz ,500 GB SSD storage

Graphics card: NVIDIA GeForce GTX 1660 Ti

- **Software requirements**

Operating System: Windows

IDE/Code Editors: VS Code, Visual Studio, PyCharm

Programming Languages: Python

Libraries/Frameworks: Machine Learning Libraries

Natural Language Toolkit

Image/Video Processing Libraries

Speech-to-Text API (with Sinhala language support)





08. CONCLUSION

- Our project generates automated, step-by-step visual performance guides from Sinhala demonstration videos.
- We use speech-to-text, natural language processing, machine learning and image/video processing for this purpose.
- This solution helps Sinhala-speaking users quickly access key information, improving content accessibility and efficiency without having to watch full-length videos.

09. LIST OF REFERENCES

- [1] M. Punchimudiyanse and R.G.N. Meegama, "Web based Automated Speech-to-Text Translator for the Sinhala Language," 2016.
- [2] B. Kang, "A Review on Image & Video Processing," Int. J. Multimed. Ubiquitous Eng., vol. 2, May 2007.
- [3] M. A. Jahan and K. Wijesekara, "Automated text summarization of Sinhala online articles," 2023.



THANK YOU