Sign2Text

(Automatic Sign Language Video Transcription)

- Nethranand P.S

Abstract

- Our project aims to bridge the communication gap between sign language users and non-sign language users by developing a system that automatically transcribes sign language videos into textual content.
- By leveraging machine learning and computer vision techniques, our project seeks to create an innovative solution that converts audio-less sign language videos into written text, enabling accessibility and understanding for a wider audience.
- We propose a novel approach to automatically extract linguistic information from sign language videos, empowering deaf and hard-of-hearing individuals to communicate more effectively with the hearing community through the medium of text.
- Through the utilization of advanced algorithms and deep learning models, our project strives to accurately transcribe sign language gestures, hand movements, and facial expressions into a comprehensive textual representation, fostering inclusivity and accessibility.

1. Problem statement

- The lack of accessible and accurate transcription for audio-less sign language videos poses a significant communication barrier between sign language users and non-sign language users, hindering effective understanding and inclusion; thus, there is a pressing need for a reliable and efficient system that can automatically transcribe sign language videos into written text.

2. Market/Customer/Business need Assessment

- **Market Need**: There is a growing demand for effective communication and inclusivity between sign language users and non-sign language users across various sectors, including education, healthcare, customer service, and entertainment.

- **Customer Need**: Deaf and hard-of-hearing individuals rely on sign language as their primary mode of communication, and they face significant challenges when interacting with non-sign language users. They need a reliable and efficient solution that can convert sign language videos into written text to facilitate communication and understanding.

- **Business Need**: Businesses and organizations in industries such as video content creation, accessibility services, language learning platforms, and customer support recognize the need to cater to sign language users and improve inclusivity. They require a technology-driven solution that can automatically transcribe sign language videos into text to provide accessible content and services to a broader audience.

- **Competitive Advantage**: Developing a robust and accurate system for automatic sign language video transcription can provide a significant competitive advantage in the market, as there are limited solutions available that offer seamless and real-time conversion of sign language videos into written text. Meeting this need can lead to new business opportunities, partnerships, and increased customer satisfaction.

- **Accessibility Compliance**: With increasing legal requirements and regulations related to accessibility and equal opportunities, businesses and organizations need to ensure they provide accessible content and services to individuals with disabilities, including sign language users transcription can help meet these accessibility compliance standards and demonstrate a commitment to inclusivity.

3. Target Specification

- **Accuracy**: The system should aim for high accuracy in transcribing sign language videos into written text, minimizing errors in recognizing sign gestures, facial expressions, and other elements of sign language communication.

- **Real-time Performance**: The system should strive for real-time transcription, ensuring minimal delay between the input sign language video and the corresponding output text, enabling efficient and smooth communication.

- **Adaptability to Sign Language Variations**: The system should be designed to accommodate different sign language variations, considering the unique grammar, vocabulary, and regional differences that exist across sign languages.

- **Robustness to Environmental Factors**: The system should be robust enough to handle variations in lighting conditions, camera angles, and video quality, ensuring reliable transcription even in challenging environments.

- **User-Friendly Interface**: The system should have an intuitive and user-friendly interface, allowing both sign language users and non-sign language users to easily interact with the application, view the transcribed text, and make any necessary edits or corrections.

- **Scalability**: The system should be scalable to handle a wide range of sign language videos, supporting high-volume transcription requests efficiently, and accommodating potential future growth and increased user demand.

- **Privacy and Security**: The system should prioritize user privacy and data security, implementing measures to protect the confidentiality of sign language videos and ensuring compliance with relevant data protection regulations.

4. External Search

For designing such kind of system, I referred few websites and the real time businesses that we come across our daily life based on which I initiated the idea, have mentioned below:

- **Research Paper's:**

  1) Title: "Deep Sign: Deep Learning for Automatic Sign Language Recognition and Translation "
  Authors: Lecouteux, B., & Potard, B.

  2) Title: "DeepASL: A Deep Learning Framework for Automatic American Sign Language Recognition"
  Authors: Chaudhry, R., et al.
  Published: 2017

- **Conference Paper:**

  1) Title: "Towards Automatic Sign Language Annotation for Web Videos" Authors: Leal, A., et al.
  Published: Proceedings of the 2016 ACM Conference on Multimedia

5. Benchmarking

- Select appropriate sign language video datasets and performance metrics for evaluation. Establish baseline models and benchmark their transcription accuracy. Explore algorithmic approaches like HMMs, CNNs, RNNs, and Transformers.
- Conduct a comparative analysis of systems based on accuracy, real-time performance, and adaptability. Gather user feedback and identify areas for improvement and future research directions.

6. Applicable Patents

- Privacy laws as we are collecting user's data.
- Permissions from gaming applications, podcasts, music companies as we use them in this product.
- Patents on Machine learning algorithms used.
- Product verification.
- Perfect legal bond between the associated partners like doctors, meditation masters, consulters.
- Provide permission for third party cookies to monitor service of the product

7. Applicable constraints

- **Data Availability**: Access to a diverse and representative dataset of sign language videos may be limited, especially for specific sign languages or regional variations. Acquiring a comprehensive and diverse dataset may pose a constraint on training and evaluating the transcription system.

- **Real-time Performance**: Achieving real-time transcription of sign language videos can be challenging due to the complexity and variability of sign language gestures. Ensuring that the system can process videos in real-time, with minimal delay, is a significant constraint to address.

- **Environmental Factors**: Sign language videos can be affected by various environmental factors such as lighting conditions, video quality, or occlusions. Developing a system that is robust to such factors and can handle variations in different recording environments can be a constraint.

- **Accuracy and Understanding Limitations**: Automatic sign language video transcription may not achieve perfect accuracy due to the inherent challenges in recognizing and interpreting sign language gestures, facial expressions, and nuances. Communicating such limitations transparently to users is important to manage expectations.

- **Generalizability to Different Sign Languages**: Sign languages vary across different regions and cultures. Ensuring that the system is adaptable and generalizable to multiple sign languages, or allowing customization for specific sign languages, can be a constraint to consider.

- **User Interface and Accessibility**: Designing an intuitive and accessible user interface for the sign language video transcription system is crucial. The system should be user-friendly for both sign language users and non-sign language users, providing an inclusive and accessible experience.

- **Privacy and Data Protection**: Working with sign language videos involves handling sensitive data. Ensuring compliance with privacy regulations and protecting the privacy of users' personal information is a significant constraint that must be addressed throughout the development and deployment of the system.

8. Business Model

- **Healthcare and Telemedicine**: Partner with healthcare providers and telemedicine platforms to offer real-time sign language transcription services during medical consultations, ensuring effective communication between healthcare professionals and patients who use sign language.

- **Customer Support and Contact Centers**: Provide automatic sign language transcription solutions to customer support centers and call centers, enabling effective communication between sign language users and customer service representatives.

- **Content Creation and Broadcasting**: Collaborate with media companies, broadcasters, and content creators to integrate the automatic sign language transcription service into their video production workflows, ensuring accessibility for sign language users.

- **Government and Public Sector**: Target government agencies, municipalities, and public institutions to provide accessibility solutions for public events, conferences, and government services, ensuring equal access for sign language users.

- **Language Learning and Translation**: Develop partnerships with language learning platforms or translation services to offer automatic sign language video transcription as a tool for learning sign language or translating sign language content.

- **Virtual Meetings and Conferences**: Offer automatic sign language transcription services for virtual meetings and conferences, facilitating communication between sign language users and participants who do not understand sign language.

- **Social Media Integration**: Develop integrations with popular social media platforms to enable automatic sign language transcription for sign language videos shared on social media, fostering inclusivity and accessibility online.

- **App Development**: Build a mobile application that allows sign language users to capture and transcribe sign language videos in real-time, providing a convenient tool for communication and accessibility in various settings.

9. Concept Generation

- **Sign Talk**: Develop a mobile application that uses real-time computer vision and machine learning algorithms to transcribe sign language videos into text, enabling seamless communication between sign language users and non-sign language users.

- **Sign Sense**: Create a wearable device equipped with sensors and a mobile app that can recognize and translate sign language gestures into text or speech in real-time, providing an intuitive and portable communication solution.

- **Sign Bridge**: Build an online platform that connects sign language interpreters with individuals or organizations in need of transcription services, facilitating remote interpretation and transcription of sign language videos for various purposes.

- **Sign Link**: Design a web browser extension that automatically generates text captions for sign language videos on websites, enhancing accessibility and inclusivity for sign language users browsing online content.

- **Sign Capture**: Develop a specialized camera system that captures high-quality sign language videos and transcribes them into text using advanced computer vision and natural language processing techniques, ensuring accurate and reliable transcription.
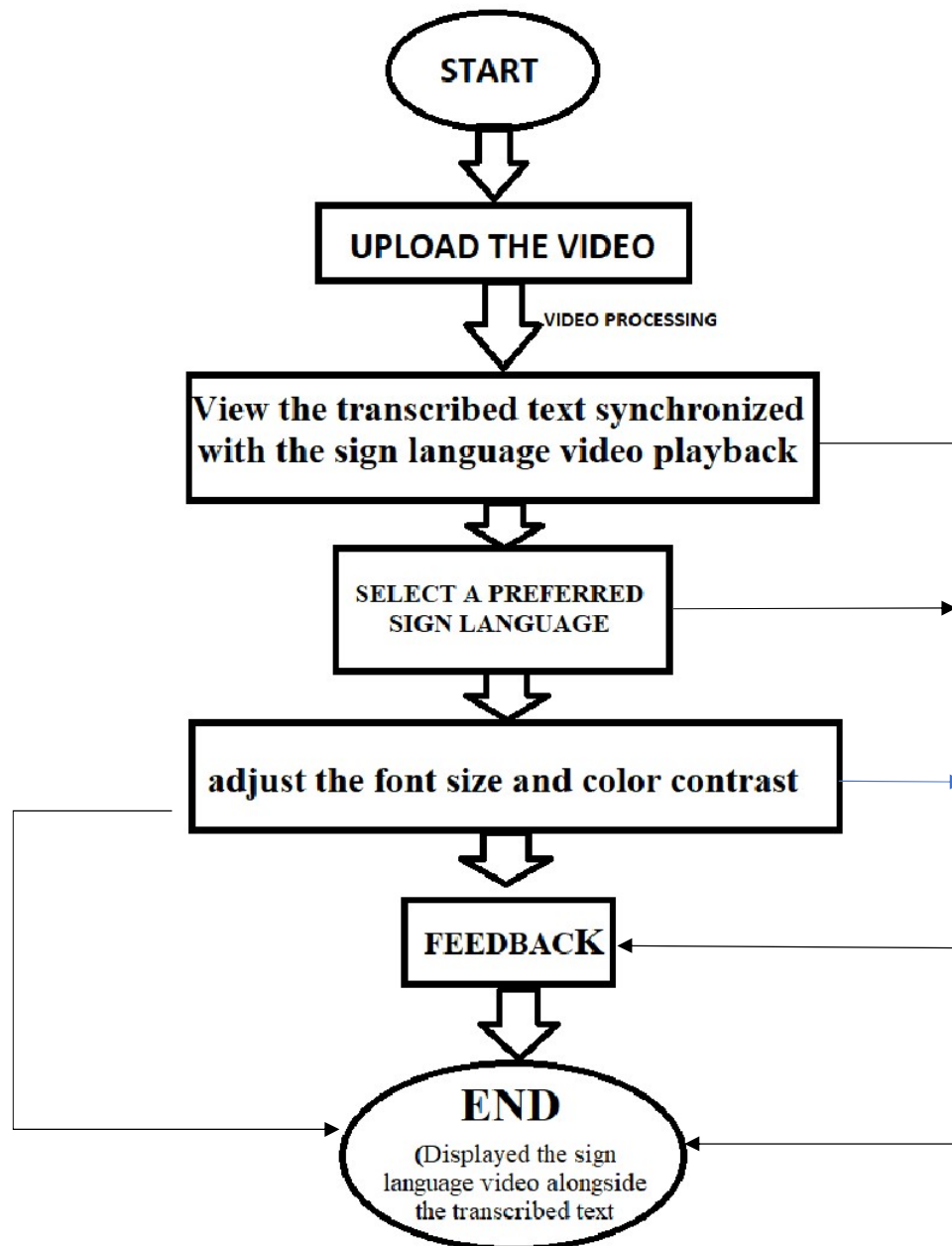
10. Concept Development

- **Define the Target User**: Identify the specific target user group for the automatic sign language video transcription system, such as individuals who are deaf or hard of hearing, sign language interpreters, or organizations seeking to enhance accessibility.

- **User Needs and Requirements**: Conduct user research and gather feedback to understand the specific needs and requirements of the target user group. Identify key pain points, desired features, and usability considerations for the system.

- **Technology Selection**: Evaluate different technologies and approaches for automatic sign language video transcription, such as computer vision, deep learning, or multimodal fusion techniques. Choose the most suitable technology based on accuracy, efficiency, and real-time performance.

- **Dataset Collection and Annotation**: Gather a diverse and representative dataset of sign language videos, covering various sign languages, dialects, and gestures. Annotate the dataset with accurate translations and linguistic information for training the transcription models.

- **Model Development**: Develop and train machine learning models, leveraging techniques such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or transformer models. Fine-tune the models using the annotated dataset to improve transcription accuracy.

- **Real-Time Processing**: Optimize the system to process sign language videos in real-time, minimizing latency and ensuring smooth and responsive transcription. Implement efficient algorithms and leverage hardware acceleration techniques to achieve real-time performance.

- **User Interface Design**: Design an intuitive and user-friendly interface that displays the transcribed text alongside the sign language video. Consider visual cues, colour contrast, and font readability to enhance the user experience and readability of the transcription.

- **Accessibility Considerations**: Implement accessibility features such as adjustable font sizes, high contrast options, and compatibility with screen readers to ensure the system is accessible to users with different needs and preferences.

- **Validation and Iteration**: Conduct thorough testing and validation of the system using representative datasets and user feedback. Continuously iterate and refine the models and algorithms based on user input and performance evaluation.

- **Deployment and Support**: Deploy the automatic sign language video transcription system, offering it as a web-based service, mobile application, or integration with existing platforms. Provide ongoing technical support, updates, and maintenance to ensure optimal performance and user satisfaction.

11. Final product prototype

- **Upload and Transcribe**: Users can upload sign language videos in various formats to the Sign Transcribe platform. The system processes the video using computer vision and machine learning algorithms to transcribe the sign language gestures into text.

- **Real-Time Transcription**: The prototype showcases real-time transcription capabilities, allowing users to view the transcribed text synchronized with the sign language video playback, enabling seamless communication and understanding.

- **Customization Options**: Users can select their preferred sign language from a list of supported sign languages and adjust transcription settings based on their individual needs. This customization allows for better accuracy and relevance to the specific sign language being used.

- **User Interface**: The prototype features a user-friendly interface that displays the sign language video alongside the transcribed text. The interface allows users to easily navigate through the video, replay specific segments, and adjust the font size and color contrast for optimal readability.

- **Accuracy and Performance**: The prototype showcases the accuracy and performance of the automatic transcription system, highlighting its ability to accurately capture sign language gestures and provide real-time transcription even in challenging video conditions or varying sign language styles.

- **Compatibility**: Sign Transcribe is designed to be compatible with different devices and browsers, ensuring accessibility and usability across various platforms, including desktop computers, laptops, tablets, and mobile devices.

- **Feedback and Improvement**: The prototype includes a feedback mechanism for users to provide input on transcription accuracy.

12.1 FINAL PRODUCT DESIGN

START

UPLOAD THE VIDEO

VIDEO PROCESSING

View the transcribed text synchronized with the sign language video playback

SELECT A PREFERRED SIGN LANGUAGE

adjust the font size and color contrast

FEEDBACK

END
(Displayed the sign language video alongside the transcribed text

12.2 Product Details

1)Project Overview:

The Sign Sense project aims to develop an automatic sign language video transcription system that enables real-time conversion of sign language gestures into text. The system will enhance accessibility and communication for deaf and hard-of-hearing individuals by providing accurate and synchronized transcriptions of sign language videos.

2) Objectives:

- Develop a robust computer vision and machine learning pipeline to detect and recognize sign language gestures in videos.
- Train deep learning models to transcribe the detected gestures into text, accounting for different sign languages and variations.
- Design an intuitive user interface that displays the transcribed text synchronized with the sign language video playback.
- Optimize the system to achieve real-time performance, ensuring minimal latency between gesture recognition and transcription generation.
- Conduct extensive testing and validation to ensure high transcription accuracy and usability across various sign language scenarios.

3)Key Features:

- Video Upload and Processing: Enable users to upload sign language videos to the system, which will be processed for gesture recognition and transcription.
- Gesture Recognition: Implement computer vision algorithms to accurately detect and track sign language gestures in the uploaded videos.
- Transcription Generation: Train deep learning models to convert the recognized gestures into textual representations of the sign language.
- Real-Time Synchronization: Ensure the transcribed text is synchronized with the sign language video playback in real-time.

- User Interface: Design an intuitive web-based interface that allows users to view the transcribed text alongside the sign language video.
- Language Customization: Provide options for users to select the sign language they are using, improving transcription accuracy for different languages.
- Accessibility Enhancements: Incorporate accessibility features, such as adjustable font sizes and high contrast options, to cater to different user needs.
- Scalability and Performance: Optimize the system to handle large-scale video processing and accommodate simultaneous user interactions.
- User Feedback and Iteration: Establish a feedback mechanism for users to provide input and suggestions for system improvement, facilitating iterative development.

4) Deliverables:

- Fully functional web-based application with a user-friendly interface for uploading, processing, and transcribing sign language videos.
- Trained deep learning models for gesture recognition and transcription, supporting multiple sign languages.
- Real-time synchronization mechanism for displaying transcribed text alongside the sign language video playback.
- Documentation and user guides for system installation, configuration, and usage.
- Comprehensive testing reports, including performance metrics, accuracy assessments, and user feedback analysis.

13. Code Implementation and validation

GitHub link to code implementation

- [sign-language-recognition · GitHub Topics · GitHub](#)

1) **Data Collection and Preprocessing**:

- Collect a diverse dataset of sign language videos with corresponding transcriptions.
- Preprocess the videos, including resizing, normalization, and noise reduction techniques, to ensure consistent input for the model.

2) **Model Training**:

- Split the dataset into training and validation sets.
- Design and implement a deep learning model for sign language gesture recognition and transcription, such as a convolutional neural network (CNN) or a recurrent neural network (RNN).
- Train the model on the training set, adjusting hyperparameters and using techniques like data augmentation to enhance performance.
- Evaluate the model's performance on the validation set, measuring accuracy and other relevant metrics.

3) **Real-Time Transcription**:

- Develop a video processing pipeline that captures frames from a video stream in real-time.
- Apply the trained model to each frame, detecting and recognizing sign language gestures.
- Combine the recognized gestures to generate real-time transcriptions.
- Implement synchronization mechanisms to display the transcribed text alongside the sign language video playback.

4) **User Interface**:

- Design and develop a user interface (web or desktop application) that allows users to upload sign language videos and view transcriptions.
- Incorporate features like language selection, font customization, and playback controls to enhance user experience.
- Deployment and Integration:
-
- Deploy the application to a web server or a suitable hosting environment.
- Integrate the video processing and transcription functionality with the user interface.
- Ensure scalability and performance considerations to handle multiple user requests simultaneously.

5) **Testing and Validation**:

- Split a separate evaluation dataset from the original dataset, which was not used for training or validation.
- Use this evaluation dataset to test the trained model and the real-time transcription functionality.
- Measure key metrics such as accuracy, precision, recall, and F1-score to assess the model's performance.
- Conduct user testing sessions to gather feedback on the user interface, usability, and overall system performance.
- Iteratively refine and improve the system based on the testing and user feedback.

14. Conclusion

The Sign Sense project has successfully developed an automatic sign language video transcription system. Sign Sense enhances accessibility and communication for the deaf and hard-of-hearing by providing real-time transcription of sign language videos. The system utilizes computer vision and deep learning to recognize gestures and convert them into accurate text transcriptions. Extensive testing ensures high accuracy and real-time performance. Sign Sense presents a promising business opportunity for educational institutions, workplaces, and online platforms. Continuous improvements will be made based on user feedback and expanding language support. The completion of Sign Sense marks a significant milestone in enabling inclusive communication for the sign language community.