ALYSSON CARLOS DE CASTRO CORDEIRO

Simultaneous Libras Translation System based on Computer Vision and Artificial Intelligence: An assistive technology for the inclusion of deaf and hearing students in Brazilian high schools

SÃO PAULO
2025

ALYSSON CARLOS DE CASTRO CORDEIRO

Simultaneous Libras Translation System based on Computer Vision and Artificial Intelligence: An assistive technology for the inclusion of deaf and hearing students in Brazilian high schools

Final Course Project submitted to the Institute of Technology and Leadership (INTELI), to obtain a bachelor's degree in Computer Engineering.

Advisor: Prof. Guilherme Henrique de Oliveira Cestari

SÃO PAULO
2025

Cataloging in Publication
Library and Documentation Service
Instituto de Tecnologia e Liderança (INTELI)
Data entered by the author.
(Cataloging record with international cataloging data, according to NBR 14724. The record will be completed later, after approval and before the final version is deposited. The completion of the cataloging record is the responsibility of the institution's library.)

Cordeiro, Alysson Carlos de Castro
Simultaneous Libras Translation System based on Computer Vision and Artificial Intelligence : An assistive technology for the inclusion of deaf and hearing students in Brazilian high schools / Alysson Carlos de Castro Cordeiro; advisor: Guilherme Henrique de Oliveira Cestari. – São Paulo, 2025. 25 p. : il

Final Course Project (Bachelor's degree) – Computer Science Course / Instituto de Tecnologia e Liderança. Bibliography

1. Brazilian Sign Language. 2. Computer Vision. 3. Artificial Intelligence. 4. Assistive Technology. 5. School Inclusion.

CDD. 23. ed..

Resumo

Referência: CORDEIRO, Alysson Carlos de Castro. **Sistema de tradução simultânea de Libras baseado em visão computacional e inteligência artificial**: uma tecnologia assistiva para a inclusão de estudantes surdos e ouvintes no ensino médio brasileiro. 2025. 23 f. TCC (Graduação) – Curso de Ciência da Computação, Instituto de Tecnologia e Liderança, São Paulo, 2025.

Resumo: A inclusão de estudantes surdos no ensino médio brasileiro enfrenta barreiras comunicacionais significativas devido à escassez de intérpretes e à baixa disseminação da Língua Brasileira de Sinais (Libras). Este trabalho apresenta o desenvolvimento de um sistema de tradução automática de Libras para texto, utilizando Visão Computacional e Inteligência Artificial, com foco na acessibilidade e no baixo custo para o ambiente escolar. A metodologia experimental envolveu a criação iterativa de modelos utilizando o framework MediaPipe para extração de marcos (landmarks), redes de Long Short-Term Memory (LSTM) para o processamento de gestos dinâmicos e Redes Neurais Convolucionais (CNN) para a classificação de sinais estáticos, como o alfabeto e numerais. Embora os modelos tenham alcançado alta acurácia em ambiente de validação controlado, superando os 90%, os testes práticos em tempo real revelaram desafios críticos de latência, variando entre 2 a 3 segundos, e sensibilidade a condições externas de iluminação e enquadramento. Os resultados indicam que, embora a arquitetura atual seja promissora, ela apresenta limitações para uma tradução estritamente simultânea. Como conclusão e direcionamento futuro, propõe-se a migração para arquiteturas híbridas de Long-term Recurrent Convolutional Networks (LRCN) e a reorientação do sistema para uma plataforma de auxílio ao aprendizado da Libras, priorizando a precisão pedagógica e a inclusão educacional efetiva.

Palavras-Chave: língua brasileira de sinais; visão computacional; inteligência artificial; tecnologia assistiva; inclusão escolar.

Abstract

Reference: CORDEIRO, Alysson Carlos de Castro. **Simultaneous Libras translation system based on computer vision and artificial intelligence**: an assistive technology for the inclusion of deaf and hearing students in Brazilian high schools. 2025. 23 p. Final Course Project (Bachelor) – Computer Science Course, Institute of Technology and Leadership, São Paulo, 2025.

Abstract: The inclusion of deaf students in Brazilian high schools faces significant communication barriers due to the shortage of interpreters and the limited dissemination of Brazilian Sign Language (Libras). This paper presents the development of an automatic Libras-to-text translation system using Computer Vision and Artificial Intelligence, focusing on accessibility and low cost for the school environment. The experimental methodology involved the iterative creation of models using the MediaPipe framework for landmark extraction, Long Short-Term Memory (LSTM) networks for dynamic gesture processing, and Convolutional Neural Networks (CNN) for static sign classification, such as the alphabet and numerals. Although the models achieved high accuracy in a controlled validation environment, exceeding 90%, real-time testing revealed critical latency challenges, ranging from 2 to 3 seconds, and sensitivity to external lighting and framing conditions. The results indicate that, while the current architecture is promising, it presents limitations for strictly simultaneous translation. As a conclusion and future direction, the study proposes migrating to hybrid Long-term Recurrent Convolutional Network (LRCN) architectures and reorienting the system toward a Libras learning support platform, prioritizing pedagogical precision and effective educational inclusion.

Key words: brazilian sign language; computer vision; artificial intelligence; assistive technology; school inclusion.

Summary

1 INTRODUCTION

Communication is a fundamental right and a basic necessity for full social participation. In Brazil, the Brazilian Sign Language (Libras) is recognized as the country's second official language and constitutes the primary means of communication for the deaf community (BRASIL, 2002). However, the low dissemination of Libras knowledge among hearing individuals and the shortage of qualified interpreters create significant barriers, particularly in educational contexts such as Brazilian high schools. These barriers limit inclusion, autonomy, and equitable access to education for deaf students, impacting their academic performance and future opportunities, such as university admission.

In this context, the development of Assistive Technologies (AT), such as automatic translation systems from Libras to text or speech, emerges as a promising solution to reduce these barriers. Currently, communication between deaf and hearing students in the school environment relies heavily on human interpreters, whose availability is limited, especially in public schools. Brazil has approximately 2.3 million people with severe or total hearing impairment (IBGE, 2024). Although previous estimates reached more than 10 million people—approximately 5% of the population (AGÊNCIA BRASIL, 2019)—less than 1% of the hearing population is estimated to have proficiency in Libras (SENADO FEDERAL, 2019). This reality leads to challenges such as school dropout and social exclusion.

Despite existing research in sign recognition, such as the use of Convolutional Neural Networks (CNNs) for static gestures and Recurrent Neural Networks (RNNs) for dynamic ones, significant gaps remain. Many studies fail to address the complexity of Libras, including regionalisms and facial expressions. Furthermore, few solutions are optimized for the Brazilian educational context or are financially accessible for public schools. Projects like *Hand Talk* offer translation but are not fully adapted for direct communication performance in classrooms.

Therefore, this project seeks to fill these gaps by developing a simultaneous Libras translation system that is affordable and culturally sensitive, focused on secondary education. The central research question is: *"How to develop a simultaneous Libras-to-text translation system using computer vision and artificial intelligence that* *is financially accessible and culturally sensitive for Brazilian high school students?"*. The hypothesis is that it is feasible to develop such a system using deep learning techniques capable of operating in real-time.

The general objective is to develop this translation system using Computer Vision and AI. Specific objectives include: Developing the computer vision model using the MediaPipe framework; utilizing Long Short-Term Memory (LSTM) to process dynamic signs for real-time translation; ensuring privacy, security, and scalability while promoting awareness about the deaf community.

Initially, the solution aims to capture alphabet letters and numerals from 1 to 9. This work is structured as follows: **Section 1** presents the introduction and objectives; **Section 2** covers the development, including the theoretical framework, state of the art, methodology, and results ; and **Section 3** concludes the study with final considerations.

2 DEVELOPMENT

2.1 Theoretical Framework and State of the Art

The development of an automatic translation system for Brazilian Sign Language (Libras) is an interdisciplinary field that combines linguistics, Computer Vision (CV), Artificial Intelligence (AI), and social inclusion. Libras is a visual-spatial language with its own grammatical structure, distinct from oral languages like Portuguese. It utilizes manual signs, facial expressions, and body movements to convey meaning, including elements such as hand configuration, articulation points, movement patterns, and non-manual markers.

Computer Vision provides the tools to capture and interpret these gestures. Convolutional Neural Networks (CNNs) are particularly effective at recognizing visual patterns in static gestures, such as hand shapes for letters and numbers. For dynamic signs involving movement over time, Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks and Transformers, are used to process sequential data. Frameworks like Google MediaPipe and OpenCV are essential for detecting key hand and facial landmarks in real-time.

The state of the art reveals significant progress, but challenges remain regarding accuracy, latency, and regional variations. Academic projects such as *Sign Language Detection* and *Sign Language Interpreter using Deep Learning* have shown high accuracy (above 95%) but often lack support for regionalisms or facial expressions. Commercially, applications like *Hand Talk* offer translations but are not optimized for the real-time, high-performance needs of a classroom environment. Furthermore, none of the analyzed solutions successfully combine a focus on Libras complexity with low-cost accessibility specifically for the Brazilian educational context.

2.2 Methodology

The experimental methodology followed an iterative development process. Initially, the study focused on the letter "H," which involves a rotational gesture. A binary LSTM model was developed using 63 features extracted from hand landmarks via MediaPipe across 30-frame sequences. This initial model achieved a validation accuracy of 95.00%.

However, expanding the system to include the letter "K" (an upward movement) revealed limitations in binary classification, leading to a transition to a multiclass model distinguishing "H," "K," and "others". The final dataset for this phase included over 2,000 valid sequences with variations in speed, angle, and lighting. The optimized architecture featured two LSTM layers (256 and 128 units), Batch Normalization for stability, and a softmax output layer, reaching a validation accuracy of 90.22%.

For static signs (letters "A," "B," and numbers 1-9), a CNN architecture was implemented. This model processed 64x64 grayscale images and utilized three convolutional layers and data augmentation to mitigate overfitting. The integration of both models was managed by a decision logic that prioritizes the highest prediction confidence above 80%.

2.3 Results

The initial phase of the project focused on a binary model for the letter "H," which initially showed 100% accuracy, indicating significant overfitting due to the small dataset. After improvements with an LSTM architecture and a larger dataset, the model achieved a more robust 98.75% training accuracy and 96.67% validation accuracy. However, the expansion to include the letter "K" revealed that the binary model could not effectively distinguish between the two signs, necessitating a transition to a multiclass model.

The multiclass training (H, K, and "Others") faced challenges with data imbalance, as the "Others" class represented 48.2% of the dataset. Initial results showed a discrepancy between training (60.86%) and validation accuracy (95.71%), confirming persistent overfitting. Real-time testing further highlighted usability issues: 1. Frequent confusion between "H" and "K" signs; 2. High sensitivity to external conditions such as lighting and distance (optimal at 30-50 cm); 3. A critical latency of 2–3 seconds, caused by the 30-frame processing window required for temporal information.

Usability tests with eight participants showed that while the response speed was perceived as efficient, the actual technical accuracy was insufficient for functional simultaneous translation.

2.4 Analysis or Discussion of Results

The results exposed fundamental limitations in the LSTM-only architecture. Analysis indicates that using LSTMs for static gestures is computationally inefficient, as the model is forced to learn "lack of change" rather than spatial patterns. Consequently, this study proposes a strategic pivot to a hybrid architecture known as Long-term Recurrent Convolutional Network (LRCN). The LRCN model unifies spatial and temporal analysis:

- **CNN Component**: Acts as a spatial feature extractor, processing individual frames to capture hand shapes and orientations.
- **LSTM Component**: Acts as a temporal modeler, analyzing the sequence of features over time to interpret movement.

This hybrid approach is identified as the superior solution for a unified recognition system, as it provides a single framework for both static and dynamic signs while potentially reducing latency through smaller, optimized processing windows.

Beyond technical considerations, the potential social impact is significant. Such technology could reduce linguistic barriers in Brazilian schools caused by the shortage of interpreters, improving educational access and promoting workplace inclusion. However, risks such as over-reliance on technology and the difficulty of capturing regional nuances must be considered. This research provides the roadmap for transitioning from a simple translator to a robust Libras learning support platform.

3 CONCLUSION

The primary goal of this research was to develop a real-time Libras-to-Portuguese recognition and translation system utilizing computer vision and deep learning techniques, with a specific focus on accessibility within the Brazilian secondary education system. Although the initial binary and multiclass models achieved promising accuracy levels in controlled validation environments—exceeding 90% in several instances—a detailed analysis of practical results and usability tests revealed critical limitations.

The identified challenges, such as a latency of 2 to 3 seconds, high sensitivity to lighting and distance variations, and difficulties in generalizing to real-world scenarios, demonstrated that the current model is insufficient for the objective of truly "simultaneous translation". Consequently, the research diagnosed the inadequacy of an LSTM-only architecture for a unified system of static and dynamic gestures, justifying a strategic pivot for the project. The technical conclusion points to the necessity of migrating to a hybrid Long-term Recurrent Convolutional Network (LRCN) architecture, which processes raw frames and unifies spatial (CNN) and temporal (LSTM) analysis into a single model to mitigate identified latency and efficiency issues.

Beyond the technical findings, the project highlights significant social impacts. The development of such technology has the potential to improve educational access for deaf students by reducing linguistic barriers caused by the shortage of qualified interpreters in Brazil. In professional environments, the system could facilitate daily communication, promoting inclusion in the labor market and valuing Libras as the country's second official language. However, the analysis of potential negative impacts notes risks such as over-reliance on technology, the difficulty of capturing regional nuances and facial expressions, and implementation costs that could limit access in underprivileged regions.

In final considerations, while the initial goal of a low-latency simultaneous translation system was not fully achieved with the current architecture, the research provided valuable insights that direct the project's future toward a more robust and inclusive Libras learning platform. The future roadmap includes implementing the LRCN architecture and ongoing validation with the deaf community to ensure the technology meets real communication and accessibility needs, fostering ethical and effective inclusion in Brazilian society.

REFERENCES

AGÊNCIA BRASIL. **País tem 10,7 milhões de pessoas com deficiência auditiva, diz estudo**. EBC, Brasília, DF, Oct. 13, 2019. Available at: https://agenciabrasil.ebc.com.br/geral/noticia/2019-10/brasil-tem-107-milhoes-de-deficientes-auditivos-diz-estudo. Accessed on: Nov. 11, 2025.

AVASTHI, Somyansh. **Sign Language Detection using CNN Architecture**. GitHub, 2021. Available at: https://github.com/SomyanshAvasthi/Sign-Language-Detection-using-CNN-Architecture. Accessed on: Apr. 30, 2025.

BORGES, Lucas Alves. **Convolutional Neural Networks - LIBRAS**. GitHub, 2020. Available at: https://github.com/lucaaslb/cnn-libras. Accessed on: Apr. 30, 2025.

BRADSKI, Gary. **OpenCV Library**. Open Source Computer Vision Library, 2000. Available at: https://opencv.org/. Accessed on: Apr. 29, 2025.

BRASIL. **Lei nº 10.436, de 24 de abril de 2002**. Dispõe sobre a Língua Brasileira de Sinais - Libras e dá outras providências. Diário Oficial da União, Brasília, Apr. 25, 2002.

CALIMAN, Pedro Ferreira. **Desenvolvimento de um sistema de reconhecimento de sinais do alfabeto manual de Libras utilizando MediaPipe Hands e rede LSTM**. 2024. Final Course Project (Bachelor in Computer Science) – Faculdade de Ciências, Universidade Estadual Paulista (UNESP), Bauru, 2024. Available at: https://repositorio.unesp.br/server/api/core/bitstreams/ff3139f3-4497-4fe4-808b-45abfa1388a0/content. Accessed on: Apr. 29, 2025.

DA, Luc. **IA na Tradução de Língua de Sinais: Uma Revolução**. Meliva.ai, Jan. 1, 2025. Available at: https://meliva.ai/artificial-intelligence/ia-traducao-lingua-de-sinais/. Accessed on: Apr. 30, 2025.

GOOGLE. **MediaPipe**. Google Developers, 2023. Available at: https://developers.google.com/mediapipe. Accessed on: Apr. 29, 2025.

GUPTA, Harsh. **Sign Language Interpreter using Deep Learning**. GitHub, 2021. Available at: https://github.com/harshbg/Sign-Language-Interpreter-using-Deep-Learning. Accessed on: Apr. 29, 2025.

HAND TALK. **Aplicativo Hand Talk**. 2024a. Available at: https://www.handtalk.me/br/. Accessed on: Apr. 29, 2025.

HAND TALK. **Os 5 parâmetros da Libras: quais são eles e sua importância**. 2024b. Available at: https://www.handtalk.me/br/blog/parametros-da-libras/. Accessed on: Apr. 29, 2025.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Censo Demográfico 2022: características gerais dos moradores, domicílios e moradores com deficiência.** Rio de Janeiro: IBGE, 2024. Available at: https://censo2022.ibge.gov.br/panorama/. Accessed on: Nov. 11, 2025.

IOFFE, Sergey; SZEGEDY, Christian. **Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift**. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML), 32., 2015, Lille, France. Proceedings [online]. 2015. Available at: https://arxiv.org/abs/1502.03167. Accessed on: Jul. 20, 2025.

MODOBILINGUE. **O que é Libras?**. Dec. 12, 2022. Available at: https://www.modobilingue.com.br/post/o-que-é-libras. Accessed on: Apr. 29, 2025.

PRAXEDES, Anderson. **Leitor de Libras**. GitHub, 2020. Available at: https://github.com/andersonprax/Leitor-de-Libras. Accessed on: Apr. 30, 2025.

QUADROS, Ronice Müller de; KARNOPP, Lodenir Becker. **Língua de sinais brasileira: estudos linguísticos**. Porto Alegre: Artmed, 2004.

REZENDE, Tamires Martins. **Aplicação de técnicas de inteligência computacional para análise da expressão facial em reconhecimento de sinais de Libras**. 2016. Dissertation (Master) - Universidade Federal de Minas Gerais, Belo Horizonte, 2016. Available at: https://www.researchgate.net/publication/344904811_Analise_da_expressao_facial_em_reconhecimento_de_sinais_de_Libras. Accessed on: Apr. 29, 2025.

SENADO FEDERAL. **Baixo alcance da língua de sinais leva surdos ao isolamento**. Agência Senado, Brasília, DF, Apr. 25, 2019. Available at: https://www12.senado.leg.br/noticias/especiais/especial-cidadania/baixo-alcance-da-lingua-de-sinais-leva-surdos-ao-isolamento. Accessed on: Nov. 11, 2025.

SERRANO, Lucas. **RAS-Libras: Reconhecimento do Alfabeto de Sinais**. GitHub, 2020. Available at: https://github.com/lucas-serrano/Projeto-LibRAS. Accessed on: Apr. 30, 2025.

THOMAS, Kevin Jose. **Sign Language Processing (ASL English Translation)**. GitHub, 2021. Available at: https://github.com/kevinjosethomas/sign-language-processing. Accessed on: Apr. 30, 2025.