

UNIVERSIDADE DE TAUBATÉ

Sistemas de Informação

André Lucca Oliveira Martins

Julia Kelly Silva Kishi

SINALIZE: Protótipo de sistema de tradução simultânea
de libras para Português.

TAUBATÉ – SP

2023

UNIVERSIDADE DE TAUBATÉ

Sistemas de Informação

André Lucca Oliveira Martins

Julia Kelly Silva Kishi

SINALIZE: Protótipo de sistema de tradução simultânea
de libras para Português.

Trabalho de Conclusão de Curso
apresentado para Graduação pelo
Curso de Sistemas de Informação
do Departamento de Informática da
Universidade de Taubaté, sob orien-
tação do Prof. Dawilmar

TAUBATÉ – SP

2023

**Grupo Especial de Tratamento da Informação - GETI
Sistema Integrado de Bibliotecas – SIBi
Universidade de Taubaté - Unitau**

M386s Martins, André Lucca Oliveira
 Sinalize: protótipo de sistema de tradução simultânea de libras para
 Português / André Lucca Oliveira Martins, Julia Kelly Silva Kishi. -- 2023.
 29 f. : il.

 Monografia (graduação) – Universidade de Taubaté, Departamento de
 Informática, 2023.
 Orientação: Prof. Me. Dawilmar Guimarães de Araújo, Departamento de
 Informática.

 1. Libras. 2. Tradução de sinais. 3. Reconhecimento de sinais. 4.
 Língua de sinais. I. Universidade de Taubaté. Departamento de Informática.
 Graduação em Sistemas de Informação. II. Kishi, Julia Kelly Silva.

CDD – 005.1

Dedicamos esta conquista aos nossos pais e familiares, que estiveram sempre presentes nos momentos importantes, nos apoiando; também aos nossos amigos e colegas da faculdade; e, por fim, à comunidade surda.

AGRADECIMENTOS

Agradecemos a todo o corpo docente da Universidade de Taubaté, em especial àqueles ligados ao curso de Sistemas de Informação. Destacamos o apoio fundamental do orientador, Prof. Dawilmar Guimarães de Araujo, que nos acompanhou durante todo o desenvolvimento do trabalho de graduação. Agradecemos também aos nossos pais e familiares por todo o apoio ao longo do curso, especialmente nesta fase final. Expressamos a nossa gratidão aos nossos amigos, por auxiliarem não apenas no âmbito pessoal, mas também durante a jornada acadêmica.

RESUMO

Martins, André. Kishi, Julia. **SINALIZE: PROTÓTIPO DE SISTEMA DE TRADUÇÃO SIMULTÂNEA DE LIBRAS PARA PORTUGUÊS**. 2023. Trabalho de Graduação, modalidade Trabalho de Conclusão de Curso, apresentado para obtenção do Certificado do Título de Graduação em Sistemas de Informação do Departamento de Informática da Universidade de Taubaté, Taubaté.

No Brasil mais de 10 milhões de pessoas têm algum problema relacionado à surdez, o que corresponde a 5% da população. No entanto, o ensino da língua de sinais no país é deficiente, o que resulta na falta de comunicação entre surdos e ouvintes. O intuito do desenvolvimento do projeto Sinalize é melhorar a comunicação entre pessoas surdas e ouvintes, funcionando como um intérprete básico entre os dois grupos. O projeto utiliza tecnologias como OpenCV, MediaPipe e TensorFlow, para coletar, armazenar dados sobre os sinais da LIBRAS, e treinar uma máquina para reconhecer esses sinais. Resultando em um protótipo de sistema desktop, que utilizara a webcam do usuário para fazer este trabalho de tradução.

Palavras-chave: Libras, Tradução de Sinais, Reconhecimento de Sinais, Língua de Sinais

ABSTRACT

Martins, André. Kishi, Julia. **SINALIZE: PROTÓTIPO DE SISTEMA DE TRADUÇÃO SIMULTÂNEA DE LIBRAS PARA PORTUGUÊS**. 2023. Trabalho de Graduação, modalidade Trabalho de Conclusão de Curso, apresentado para obtenção do Certificado do Título de Graduação em Sistemas de Informação do Departamento de Informática da Universidade de Taubaté, Taubaté.

In Brazil, more than 10 million people have some kind of hearing-related issue, which corresponds to 5% of the population. However, sign language education in the country is deficient, leading to a lack of communication between the deaf and hearing individuals. The purpose of developing the Sinalize project is to improve communication between deaf and hearing people, serving as a basic interpreter between the two groups. The project employs technologies such as OpenCV, MediaPipe, and TensorFlow to collect and store data on Brazilian Sign Language (LIBRAS) signs and train a machine to recognize these signs. This results in a desktop system prototype that utilizes the user's webcam to perform this translation work.

Keywords: Libras, Sign Translation, Sign Recognition, Sign Language

LISTA DE FIGURAS

Figura 1 - Exemplo de configurações de mão	15
Figura 2 - Exemplo MediaPipe.....	18
Figura 3 - Treinamento da LSTM	21
Figura 4 - Tela final apresentada ao usuário	22
Figura 5 - Gravação do sinal da letra B	23
Figura 6 - Fluxograma do protótipo.....	24
Figura 7 - Detecção do sinal da letra B.....	25
Figura 8 - Demonstração do protótipo em execução	26

SUMÁRIO

1 INTRODUÇÃO.....	10
1.1 TRABALHOS CORRELATADOS.....	12
1.1.1 Detecção de Imagem para Reconhecimento de Sinais de Libras	12
1.2 OBJETIVO DO TRABALHO	12
1.2.1 Objetivo Geral.....	12
1.2.2 Objetivo Específico	12
1.3 Estrutura do Trabalho.....	13
2 FUNDAMENTAÇÃO TEÓRICA.....	14
2.1 LINGUAGENS EM SINAIS: LIBRAS	14
2.2 TRADUÇÃO E INTERPRETAÇÃO.....	15
2.3 RECONHECIMENTO FACIAL.....	16
3 DESENVOLVIMENTO DO PROTOTIPO	17
3.1 FERRAMENTAS UTILIZADAS	17
3.1.1 Linguagens de Programação.....	17
3.1.2 Componentes adicionais.....	17
3.1.2.1 Mediapipe	18
3.1.2.2 OpenCV	18
3.1.2.3 TensorFlow	19
3.1.2.3.1 LSTM	19

3.1.2.4 Customtkinter.....	21
3.2 ARQUITETURA DO PROTOTIPO.....	22
4 TESTES E RESULTADOS.....	25
5 CONCLUSÕES.....	27
5.1 TRABALHOS FUTUROS	27
5.1.1. AUMENTO DA BIBLIOTECA DE SINAIS DE LIBRAS	27
5.1.2 AMPLIAÇÃO DO TREINAMENTO DA MÁQUINA (LSTM).....	28
5.1.3 PROCESSAMENTO DE IMAGENS E ARMAZENAMENTO EM SERVIDOR NA NUVEM.....	28
5.1.4 IMPLEMENTAÇÃO DE FEEDBACK INTERATIVO.....	28
6 REFERENCIAS	29

1 INTRODUÇÃO

Em fevereiro de 2020, o Instituto Brasileiro de Geografia Estatística (IBGE) divulgou estatísticas sobre os surdos no Brasil, uma pesquisa informando que mais de 10 milhões de pessoas tem algum problema relacionado a surdez, ou seja, 5% da população é surda no Brasil. Destes, 2,7 milhões não ouvem nada.

A definição de surdez (hipoacusia), em termos médicos, é categorizada de acordo com o nível da perda auditiva: audição normal, perda auditiva leve, perda auditiva leve, perda auditiva moderada, perda auditiva severa e perda auditiva profunda.

De acordo com o decreto 5.626, de 22 de dezembro de 2005, Art. 2º:

[...] considera-se pessoa surda àquela que, por ter perda auditiva, compreende e interage com o mundo por meio de experiências visuais, manifestando sua cultura principalmente pelo uso da Língua Brasileira de Sinais - Libras.

Assim, Língua Brasileira de Sinais ou Libras é uma língua de modalidade gestual-visual, ou seja, de sinais ou gestos. É considerada uma língua oficial do Brasil desde 24 de abril de 2002, através da Lei nº 10.436.

“Entende-se como Libras a forma de comunicação e expressão, em que o sistema linguístico de natureza visual-motora, com estrutura gramatical própria, constitui um sistema linguístico de transmissão de ideias e fatos, oriundos de comunidades de pessoas surdas do Brasil”.

Entretanto, no Brasil, ainda há certa dificuldade no aprendizado e no ensino da Libras, onde é apresentado de maneira superficial durante o período fundamental e as vezes durante o ensino médio. Estudantes saem das escolas sem saberem comunicar-se com integrantes da comunidade Surda, formada por 5% da população do Brasil, sendo assim se tornando incapazes de conversar, atender e até mesmo vender produtos para uma taxa significativa da população brasileira.

Há a proposta da Lei 2403/22: “Altera a Lei nº 9.394, de 20 de dezembro de 1996 (Lei de Diretrizes e Bases da Educação Nacional), para incluir conteúdo da Língua Brasileira de Sinais - LIBRAS nos currículos da educação básica” (BRASIL, 2023), que visa promover a inclusão e acessibilidade para pessoas com deficiência, representa um marco importante para garantir direitos fundamentais. Neste contexto, ressalta-se a importância complementar do Projeto de Tradução de Libras para o Português, em sintonia com essa iniciativa legislativa.

Enquanto aguardamos o desenrolar do trâmite da lei, é essencial reconhecer que a tradução de Libras para o português desempenha um papel significativo na efetivação da inclusão e no acesso à informação para a comunidade surda. Este projeto não apenas busca suprir lacunas de comunicação, mas também reforça os princípios de igualdade e oportunidade para todos os cidadãos.

O impacto positivo do projeto de tradução é evidente nos diversos aspectos da vida cotidiana. Facilitar a compreensão entre surdos e ouvintes não é apenas uma questão de comunicação, mas sim um passo crucial para garantir a participação plena e efetiva desses indivíduos em todas as esferas da sociedade.

À medida que a proposta da Lei 2403/22 avança em seu processo legislativo, é fundamental reconhecer a complementaridade e sinergia entre a legislação em trâmite e iniciativas práticas como o Projeto de Tradução de Libras para o Português. Ambas têm o objetivo comum de assegurar direitos e promover a inclusão, cada qual desempenhando um papel crucial nesse propósito.

A idealização deste projeto, ora denominado Sinalize, veio com a intenção de auxiliar na comunicação entre pessoas Ouvintes, Surda e Surdo-Cegas, onde a própria comunicação é precária e acaba sendo falha, realizando o intermédio entre o Falante e o Ouvinte, traduzindo sinais de libras para palavras em português, cumprindo o papel semelhante à de um intérprete, porém de forma mais básica.

1.1 TRABALHOS CORRELATADOS

1.1.1 Detecção de Imagem para Reconhecimento de Sinais de Libras

Em ÂNGELO (2023), explica que a tecnologia tem o potencial de beneficiar enormemente os indivíduos com deficiência em suas vidas diárias, especialmente aqueles com deficiência auditiva. A comunidade com deficiência auditiva enfrenta barreiras de comunicação, o que dificulta a sua integração na sociedade. Para resolver esta questão, estão a ser desenvolvidas tecnologias que promovam a reintegração e a acessibilidade. O foco do trabalho daquele autor é utilizar imagem e inteligência artificial para detectar sinais gestuais e auxiliar na educação de pessoas que desejam aprender língua de sinais, especificamente a língua de sinais oficial do Brasil chamada Libras. O objetivo é promover um maior uso da língua de sinais entre a população.

1.2 OBJETIVO DO TRABALHO

1.2.1 Objetivo Geral

Apresentar o protótipo de uma aplicação que facilite a comunicação em locais onde o receptor não saiba a Língua Brasileira de Sinais (LIBRAS), traduzindo sinais para a forma escrita do português.

1.2.2 Objetivo Específico

No entendimento do objetivo geral, pode-se identificar como objetivos específicos:

- Contextualizar sobre a situação da LIBRAS no Brasil;

- Conhecer brevemente a estrutura da LIBRAS e o seus sinais;
- Apresentar ferramentas que possibilitem o desenvolvimento da aplicação;
- Conhecer detalhadamente como cada ferramenta funciona dentro da aplicação;

1.3 Estrutura do Trabalho

O presente trabalho encontra-se dividido em 6 capítulos, incluindo esta introdução, a qual descreve a origem, a importância, os objetivos e escopo dele.

O segundo capítulo apresenta a fundamentação teórica sobre a Libras, o processo de Interpretação e Tradução.

O terceiro capítulo descreve a arquitetura do protótipo implementado com o método de detecção de sinais. Demonstra também a interface gráfica desenvolvida.

No quarto capítulo é apresentado os testes realizados, juntamente com as figuras ilustrando esse processo, também o resultado que irá aparecer ao usuário.

O quinto capítulo apresenta possíveis melhorias futuras ao protótipo, que podem ser desenvolvidos para um desenvolvimento mais efetivo do sistema.

E por último, no sexto capítulo, é disponibilizado as referências que foram utilizadas para o desenvolvimento deste trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 LINGUAGENS EM SINAIS: LIBRAS

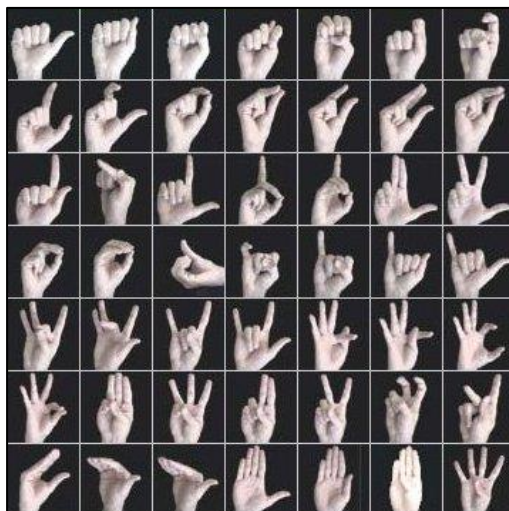
Os sinais, na LIBRAS, são compostos por 5 parâmetros, sendo eles configuração de mão, ponto de articulação, movimento, direção, expressões não-manuais. Brevemente, configuração de mão é forma e a orientação da mão usada para produzir sinais. Ponto de articulação é o local específico no corpo ou próximo a ele onde os sinais são produzidos. O movimento das mãos, dedos e braços usados para produzir sinais, a direção, velocidade e trajetória do movimento podem impactar o significado de um sinal. A posição e orientação da palma da mão em relação ao corpo ou outros objetos. A orientação da palma da mão pode modificar o significado de um sinal ou indicar aspectos como direcionalidade ou concordância (Figura 1) a seguir.

Expressões faciais, postura corporal e outros marcadores não manuais desempenham um papel importante na transmissão de emoção, intensidade e informações gramaticais em LIBRAS.

Como cita SOFIATO (2014), podemos afirmar que:

O nível fonológico da Libras compreende os seguintes parâmetros, de acordo com Fernandes (2003): configuração das mãos (forma que a mão assume ao realizar o sinal); localização do sinal; movimento das mãos e orientação da(s) palma(s) da(s) mão(s). Ao ser representado pictoricamente, esse parâmetro precisa ser o mais preciso possível para garantir a boa compreensão do sinal. Dessa forma, presume-se que um nativo na língua tenha melhores condições de fazê-lo do que um ouvinte que aprende Libras. (SOFIATO, 2014)

Figura 1 - Exemplo de configurações de mão



Fonte: <https://izabelapce.wordpress.com/2011/03/15/configuracoes-de-mao/>

2.2 TRADUÇÃO E INTERPRETAÇÃO

A tradução em língua de sinais envolve a conversão de texto escrito ou impresso, como livros, documentos ou legendas, em língua de sinais. Isso permite que indivíduos surdos acessem informações escritas em seu modo preferido de comunicação.

A interpretação em língua de sinais refere-se ao processo de transmissão de mensagens faladas em língua de sinais para indivíduos surdos ou vice-versa. Requer uma compreensão profunda da linguagem de sinais e da língua falada que está sendo interpretada.

Tanto a tradução como a interpretação em língua de sinais exigem indivíduos qualificados que sejam fluentes tanto na língua de sinais quanto na língua falada ou escrita utilizada. Esses profissionais devem compreender as nuances e o contexto cultural de ambas as línguas para transmitir com precisão o significado pretendido.

Como atividade comunicativa, a tradução é um ato de comunicação complexo que envolve duas situações comunicativas distintas, a da produção e recepção do texto original e a da produção e recepção do texto traduzido. Os dois

contextos condicionam as escolhas do tradutor, sempre levando em conta a função ou finalidade do texto traduzido no seu contexto de chegada (BEVILACQUA, 2017).

2.3 RECONHECIMENTO FACIAL

O reconhecimento facial tem aplicações significativas, como verificação de identidade, processamento de imagens, segurança de dados e detecção de fraudes. CASTRO (2022) destaca técnicas como aprendizado de máquina e redes neurais artificiais para aprimorar essa habilidade. A hipótese apresentada sugere que o Deep Learning é a tecnologia mais eficaz para o reconhecimento facial na atualidade.

É destacado o papel crucial do Deep Learning, especialmente em sistemas de inteligência artificial, com ênfase nas redes neurais convulsionais. (CASTRO, 2022)

3 DESENVOLVIMENTO DO PROTOTIPO

Nesse capítulo são descritas as ferramentas utilizadas para o desenvolvimento do Protótipo de Sistema de Tradução Simultânea de Libras para Português: Sinalize e como está arquitetado a sua estrutura em relação ao código.

3.1 FERRAMENTAS UTILIZADAS

3.1.1 Linguagens de Programação

As decisões de escolha baseiam-se em que linguagem de programação é um elemento essencial na solução de problemas através do desenvolvimento de software. A escolha correta da linguagem pode impactar a legibilidade, a expressividade, a eficiência e a escalabilidade da solução, além de influenciar o ecossistema de bibliotecas e frameworks disponíveis. Portanto, é fundamental selecionar a linguagem apropriada para resolver o problema de forma eficaz e obter os melhores resultados.

O software foi desenvolvido em Python versão 3.8.10, que é uma linguagem de programação interativa, de alto nível e orientada a objetos. Python é amplamente recomendado para aplicações de inteligência artificial e big data.

3.1.2 Componentes adicionais

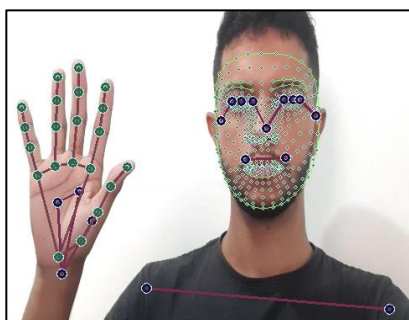
A importância de usar bibliotecas na programação está na facilidade de acesso a funcionalidades já implementadas e testadas por outros desenvolvedores. As bibliotecas oferecem diversas vantagens, tais como: Reutilização, Eficiência, Maior produtividade, Suporte comunitário etc., permitindo economia de tempo e código mais eficiente. Além disso, as bibliotecas contribuem para um ecossistema colaborativo, onde programadores podem compartilhar soluções e

beneficiar-se do conhecimento coletivo da comunidade. Isso possibilita o desenvolvimento de aplicações mais robustas e avançadas de maneira mais ágil. Em resumo, o uso de bibliotecas em Python pode economizar tempo, melhorar a qualidade do código, aumentar a produtividade e permitir que você aproveite o conhecimento e os padrões existentes em uma comunidade de suporte.

3.1.2.1 Mediapipe

O Mediapipe fornece um modelo holístico sobre cada parte do corpo (cabeça, mão, tronco etc.), conectado por pontos em determinadas áreas identificadas pela webcam, sendo utilizado, na versão 0.8.9.1, para a base do reconhecimento dos sinais. A exemplo da interação dos recursos da Mediapipe, Figura 2.

Figura 2 - Exemplo MediaPipe



Fonte: Autor

3.1.2.2 OpenCV

É utilizado a biblioteca OpenCV, na versão 4.7.0.72, para acionar a webcam, responsável pela captura de vídeo, em tempo real, para o reconhecimento dos sinais realizados.

3.1.2.3 TensorFlow

O TensorFlow é uma biblioteca de código aberto, para aprendizado de máquina em diversas áreas, no protótipo está sendo utilizado, na versão 2.4.1. Ele possibilita a criação e treinamento de redes neurais, que são capazes de identificar e interpretar padrões e correlações, utilizado no protótipo para a detecção dos sinais realizados.

Foi realizado um treinamento com a máquina para a identificação dos sinais, com 2000 testes, visando o aprimoramento do reconhecimento, alcançando uma precisão próxima a 1 (em uma escala de 0 a 1, onde 1 é 100% de precisão), como mostra a Figura 3 a seguir.

3.1.2.3.1 LSTM

O LSTM (*Long Short-Term Memory*) é um tipo de rede neural recorrente que é especialmente eficiente em lidar com dados sequenciais, como séries temporais, texto, entre outros. Ele é capaz de lembrar informações de longo prazo, superando alguns problemas de outras redes neurais recorrentes.

O LSTM possui uma arquitetura composta por células de memória que podem armazenar e atualizar informações ao longo do tempo. Cada célula possui três portões: o portão de entrada (*input gate*), o portão de esquecimento (*forget gate*) e o portão de saída (*output gate*).

- O portão de entrada decide quais informações serão atualizadas na célula de memória.

- O portão de esquecimento decide quais informações antigas serão mantidas ou esquecidas.

- O portão de saída decide quais informações serão transmitidas para a próxima célula ou como resultado.

Esses portões são controlados por funções de ativação e multiplicações de matrizes que aprendem a partir dos dados durante o treino do modelo LSTM.

Para treinar um modelo LSTM em Python, geralmente usamos uma divisão dos dados em conjuntos de treinamento (*train*) e teste (*test*). O conjunto de treinamento consiste em pares de sequências de entrada (*train-x*) e sequências de saída correspondentes (*train-y*). O modelo é treinado usando o *train-x* para prever o *train-y*.

O conjunto de teste (*test*) é usado para avaliar o desempenho do modelo em dados que ele nunca viu antes. Assim, após treinar o modelo com o conjunto de treinamento, podemos usar o conjunto de teste para ver quão bem o modelo generaliza para novos dados, medindo a precisão das previsões geradas pelo modelo.

A divisão dos dados em *train* e *test* é uma etapa importante no processo de treinamento e avaliação de um modelo de aprendizado de máquina. Uma proporção comum é usar cerca de 70-80% dos dados para treinamento e 20-30% para teste.

Em resumo, o processo de treinamento e teste de um modelo LSTM em Python envolve as seguintes etapas:

1. Dividir os dados em conjuntos de treinamento, validação e teste (*train*, *validation* e *test*).
2. Pré-processar os dados, como normalização, remoção de ruídos, tratamento de valores ausentes, codificação de sequências, etc.
3. Construir e compilar o modelo LSTM usando uma biblioteca como Keras ou TensorFlow.
4. Treinar o modelo usando o conjunto de treinamento (*train-x*, *train-y*), ajustando os hiperparâmetros com o conjunto de validação, se necessário.
5. Avaliar o desempenho do modelo usando o conjunto de teste (*test*) e métricas de avaliação adequadas, como acurácia, precisão, *recall*, etc.
6. Ajustar e otimizar o modelo, se necessário, com base nos resultados do passo anterior.

7. Fazer previsões usando o modelo treinado em novos dados ou em dados de teste adicionais.

Figura 3 - Treinamento da LSTM



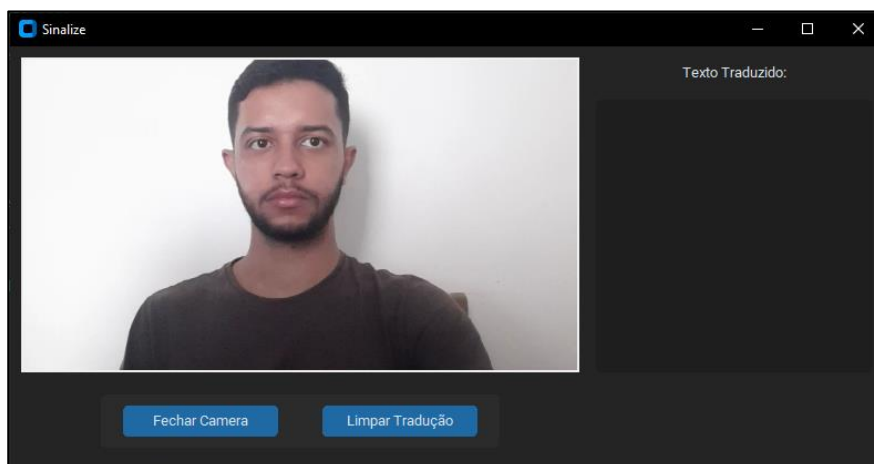
Fonte: Autor

A figura 3, mostrada anteriormente, é um gráfico detalhando o treinamento da precisão do sistema em relação a detecção dos sinais. Ao lado esquerdo do gráfico, no eixo Y, está a taxa precisão do protótipo, começando no 0.0 e indo até 1.0, em que 0.0 é uma taxa de 0% de precisão (uma detecção totalmente errada), e 1.0 é uma taxa de 100% de precisão (uma detecção correta sobre o sinal). Na parte inferior do gráfico, no eixo X, é identificado a quantidade de testes que foram realizados para o treinamento da máquina, com 2000 (dois mil) testes. Cada posição XY no gráfico, representa a acurácia do teste em determinado tempo de treinamento da máquina. Por exemplo, no intervalo entre 600 e 1000 testes realizados, foi obtido uma acuarria de 100%.

3.1.2.4 Customtkinter

Biblioteca responsável pela GUI (*Graphical User Interface*, ou em português: Interface Gráfica do Usuário), interface que será apresentada ao usuário final, que o utilizara para exibir o significado dos sinais captados pela webcam. Mostra de interação da Customtkinter na Figura 4 a seguir.

Figura 4 - Tela final apresentada ao usuário



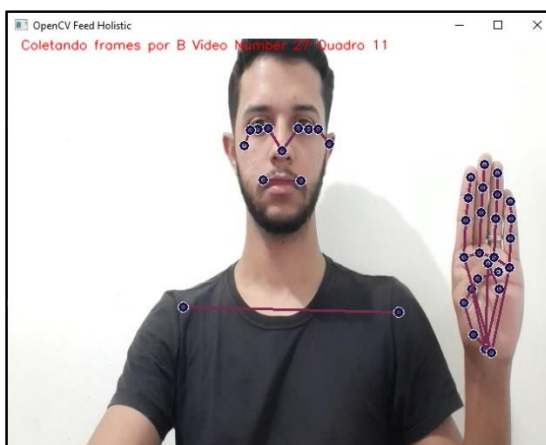
Fonte: Autor

3.2 ARQUITETURA DO PROTOTIPO

O protótipo é separado em algumas partes, primeiramente é necessário criar e organizar uma base de dados com os elementos coletados sobre cada sinal, da libras, que deseja fazer o reconhecimento, nessa etapa, é utilizado o OpenCV mais o MediaPipe para a extração das informações, sobre cada sinal, coletando os pontos X, Y e Z, em relação a câmera, de cada sinal salvo.

A coleta e o armazenamento são realizados da seguinte forma: é utilizada uma webcam para o processamento do MediaPipe, e após estar sendo processado por cima do vídeo, o MediaPipe, fornece coordenadas sobre cada ponto no corpo e na mão visualizados no vídeo. O vídeo é separado em 30 *frames* (quadros), e as coordenadas obtidas, dos pontos, de cada frame são extraídas e armazenadas em grupos, para cada sinal sendo realizado (Figura 5 a seguir).

Figura 5 - Gravação do sinal da letra B



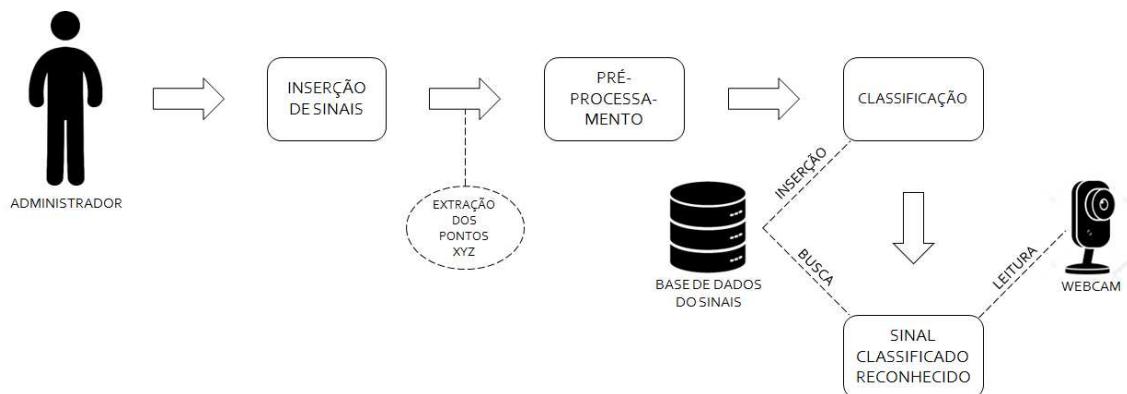
Fonte: Autor

Após a coleta e armazenamento, é utilizado o TensorFlow para realizar o treinamento da máquina para reconhecer os sinais, através do Keras. O treinamento é feito com as coordenadas obtidas, de cada sinal, do processo anterior. A máquina irá juntar uma sequência de 30 quadros (cada quadro possui um conjunto de coordenadas X, Y e Z, de cada ponto visualizado na câmera) e realizar a identificação do sinal demonstrado.

Porém todo esse processo citado anteriormente não é apresentado para o usuário final, eram etapas que rodavam previamente para poder haver a base de dados e o treinamento da máquina, utilizados para reconhecer finalmente as palavras sinalizadas pelo usuário final.

Finalmente é apresentado uma interface ao usuário (conforme a Figura 4 exibida anteriormente), que acessara a webcam do computador e irá capturar o vídeo dele sinalizando. Na interface é mostrado as imagens capturadas pela webcam, simultaneamente, ao lado direito do vídeo, é inserido uma caixa de texto, onde aparecerá o significado dos sinais reconhecidos pelo sistema. É disponibilizado ao usuário dois botões na interface, um botão responsável por abrir e fechar a captura de vídeo pela webcam, e o outro para limpar o texto que foi traduzido através do reconhecimento do sistema.

Figura 6 - Fluxograma do protótipo



Fonte: Autor

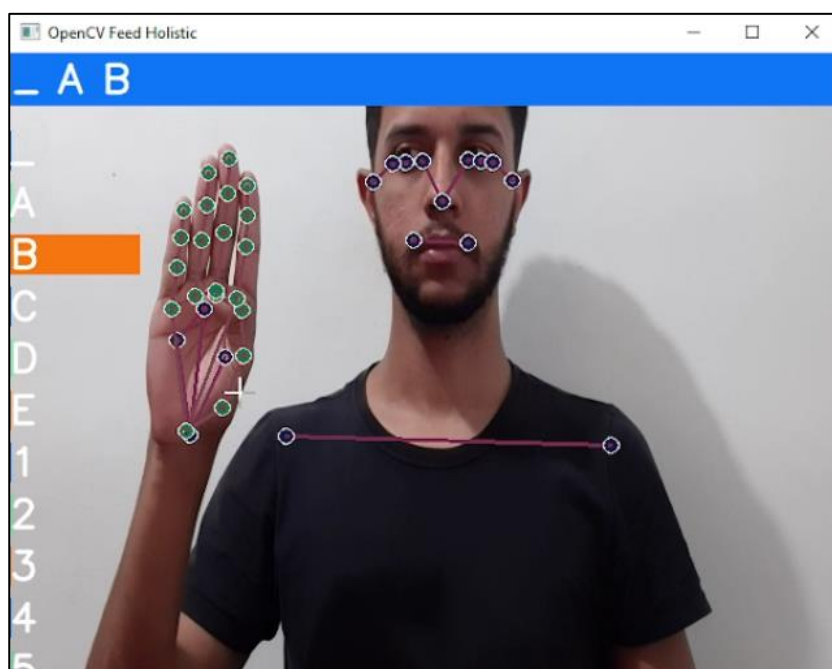
A figura 6 detalha o fluxograma do protótipo, começando pelo Administrador inserindo os sinais em libras, enquanto esses sinais são inseridos, simultaneamente o sistema já realiza a extração dos pontos XYZ de cada parte do sistema holístico, que foi gerado pelo MediaPipe. Após a etapa de inserção e extração, o algoritmo classifica as informações coletadas em grupos separados, para cada sinal que deseja realizar a tradução, em seguida é realizado o treinamento da máquina com os dados obtidos.

As informações geradas no processo de classificação são inseridas e projetadas no Banco de Dados, esse será utilizado para estabelecer se determinado sinal está classificado e é reconhecido pelo sistema, através da captura de vídeo pela Webcam, que será utilizada pelo usuário final.

4 TESTES E RESULTADOS

Para realizar os testes, foi desenvolvido uma interface específica, utilizando o OpenCV, este gera e fornece uma imagem em tempo real da webcam, possibilitando a verificação da câmera. Na imagem fornecida, é realizada a sobreposição do MediaPipe, na captura de vídeo ativa, para a detecção dos sinais, juntamente com a biblioteca de sinais previamente coletada.

Figura 7 - Detecção do sinal da letra B



Fonte: Autor

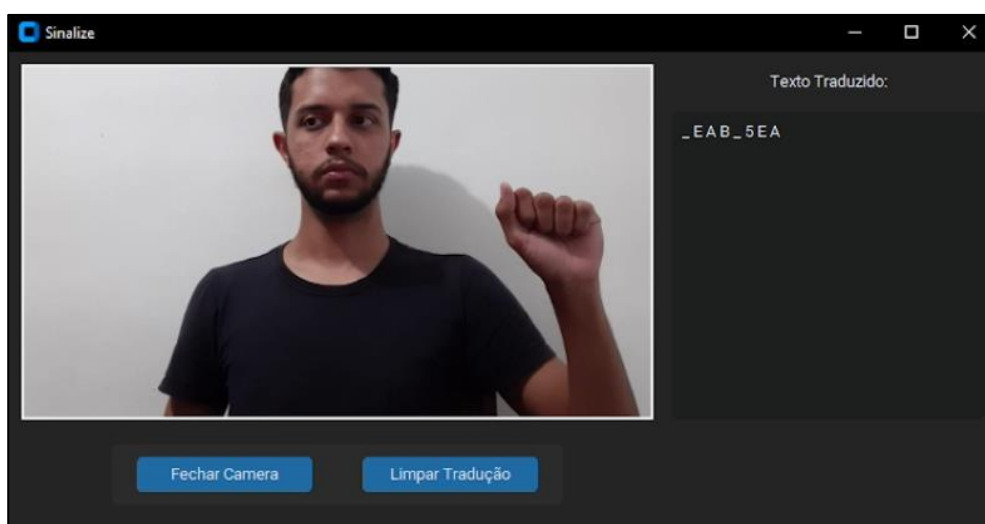
Na imagem gerada do OpenCV mais o MediaPipe, é acrescentada etiquetas ao canto esquerdo da tela, identificando alguns sinais que possivelmente podem ser detectados pelo protótipo. Cada etiqueta, correspondente por cada sinal, mostra uma barra horizontal expondo a taxa de precisão sobre a detecção do sinal visualizado. A barra representa uma taxa de 0% a 100% de precisão por cada sinal.

Na parte superior da tela, em uma faixa azul, é escrito, em sequência, os sinais que foram detectados. Os sinais que possuem uma taxa de precisão acima de 80%, e que mantêm essa precisão por uma determinada quantidade de quadros seguidos, são reconhecidos como a tradução para o sinal que está sendo sinalizado na câmera. Esse resultado, após a precisão, é que será entregue ao usuário final do protótipo.

Conforme o exemplo exibido anteriormente na figura 7, está sendo sinalizado e indicado a detecção e reconhecimento do sinal da letra B, com uma taxa de precisão acima do mínimo de 80%, sendo transcrito na barra horizontal na parte de cima da tela.

Após os testes realizados, foi desenvolvido uma GUI mais polida ao usuário final, em que aparecerá apenas a imagem que está sendo captura em tempo real, sem a sobreposição do MediaPipe por cima do usuário. Conforme a imagem captura, em tempo real, será disposta ao lado esquerdo, enquanto os sinais traduzidos aparecerão ao lado direito, como mostra a figura 8.

Figura 8 - Demonstração do protótipo em execução



Fonte: Autor

5 CONCLUSÕES

Este trabalho propôs uma solução de intermediação entre partes emissoras e receptoras na comunicação, onde há uma barreira linguística gerada pela dissonância de linguagens, em que uma parte se comunica pela Língua Portuguesa e a outra pela LIBRAS.

Utilizando ferramentas como Python e suas bibliotecas, tais como MediaPipe, para utilização do modelo holístico, OpenCV, TensorFlow, para criação e treinamento da rede neural LSTM, e CustomTKInter, para o desenvolvimento da interface final.

Resultando em um protótipo de sistema intermediador, com uma acurácia precisa, facilitando a comunicação, previamente citada, entre as duas partes, com a utilização de um computador com acesso à webcam.

5.1 TRABALHOS FUTUROS

Algumas melhorias, no sistema, podem ser realizadas ao longo do desenvolvimento do protótipo, a fim de maximizar a sua utilidade e eficiência em intermediar a comunicação em lugares necessários.

5.1.1. AUMENTO DA BIBLIOTECA DE SINAIS DE LIBRAS

Atualmente, nosso sistema tem uma limitação da nossa biblioteca com apenas 17 sinais de Libras é evidente, ressaltando a urgência de expandir essa base. A inclusão de novos gestos, expressões e variações regionais é crucial para uma tradução mais precisa e abrangente.

5.1.2 AMPLIAÇÃO DO TREINAMENTO DA MÁQUINA (LSTM)

Para aprimorar a precisão na tradução, é essencial intensificar o treinamento da máquina, utilizando técnicas como *Long Short-Term Memory* (LSTM). Isso implica em otimizar algoritmos, incluir uma gama mais ampla de dados para diversificar os padrões reconhecidos e melhorar a interpretação contextual dos sinais. A expansão do número de testes por sinal, de, por exemplo, 2000 para um número mais amplo, como 5000 ou 10000 testes, poderia proporcionar uma compreensão mais detalhada das variações e nuances dos gestos, resultando em previsões mais precisas e confiáveis.

5.1.3 PROCESSAMENTO DE IMAGENS E ARMAZENAMENTO EM SERVIDOR NA NUVEM

O aprimoramento do processamento de imagens é crucial para garantir uma identificação precisa dos sinais de Libras captados pela câmera. Isso requer a implementação de técnicas avançadas, como a detecção de movimento e análise detalhada de gestos. Além disso, armazenar essas imagens em servidores na nuvem é fundamental para assegurar acessibilidade e disponibilidade em tempo real, viabilizando o acesso remoto e a utilização em diferentes dispositivos.

5.1.4 IMPLEMENTAÇÃO DE FEEDBACK INTERATIVO

Integrar um sistema de *feedback* interativo permitiria que os usuários fornecessem informações sobre a precisão das traduções. Esse sistema de avaliação poderia ser utilizado para aprimorar continuamente o reconhecimento de sinais e a qualidade geral da tradução. A coleta de *feedback* dos usuários é crucial para identificar áreas de melhoria e aperfeiçoar constantemente o desempenho do sistema, tornando-o mais preciso e útil para a comunidade surda.

6 REFERENCIAS

ÂNGELO, Gabriel; PINA, Guilherme; ***Detecção De Imagem Para Reconhecimento De Sinais De Libras***. 2023. Disponível em: https://sites.uel.br/dc/wp-content/uploads/2022/09/PROJETO_TCC_GABRIEL_ANGELO_PEREZ_GASPARINI_SABAUDO.pdf. Acesso em: 6 out. 2023.

BEVILACQUA, Cleci Regina; KILIAN, Cristiane Krause. ***Tradução e Terminologia: relações necessárias e a formação do tradutor***, 2017. Disponível em: <https://seer.ufu.br/index.php/dominiosdelinguagem/article/view/37409/21495>. Acesso em: 23 set. 2023.

BRASIL. ***Projeto de Lei PL 2403/2022. Altera a Lei nº 9.394, de 20 de dezembro de 1996 (Lei de Diretrizes e Bases da Educação Nacional), para incluir conteúdo da Língua Brasileira de Sinais - LIBRAS nos currículos da educação básica***. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2334854#CE>. Acesso em: 10 out. 2023.

CASTRO, Ana Paula G; ACRAINI, Bruno M.; SILVA, Brenda S. A; FERNANDES, Leandro C.; ***Levantamento Da Literatura De Reconhecimento Facial***. 2022.

FERNANDES, Eulália. ***Linguagem e surdez*** Porto Alegre: Artmed, 2003.

MARQUEZI, Luana. ***LITERATURA SURDA: O PROCESSO DA TRADUÇÃO E TRANSCRIÇÃO EM SIGNWRITING***. 2018. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/210366/PGET0405-D.pdf?sequence=-1>. Acesso em: 19 set. 2023.

SOFIATO, Cássia Geciauskas; REILY, Lucia Helena; ***Dicionarização da língua brasileira de sinais: estudo comparativo iconográfico e lexical***. 2014. Disponível em: <https://www.scielo.br/j/ep/a/vY3XRbKqCzKG6kLpQdhd3dN/#>. Acesso em: 11 set. 2023.