

Sistema Móvel para Reconhecimento Automático de Gestos Estáticos da Linguagem Brasileira de Sinais

Francisco Mendes da C. Neto¹, Rodrigo de M. S. Veras¹, Pedro H. P. Albuquerque²

¹Departamento de Computação – Universidade Federal do Piauí (UFPI)
Teresina – PI – Brazil

fmendesneto@gmail.com¹, rveras@ufpi.edu.br¹, albpedro897@hotmail.com²

Abstract. *The Brazilian Sign Language (LIBRAS) is the official language in the communication of the deaf in Brazil. Although it is the primary means of communication for the hearing impaired, LIBRAS is still not very widespread in the rest of society. In this way, sign language recognition is a research area of high relevance, since it aims to include deaf people in the rest of community. Based on this, we propose in this article a mobile system, compatible with any Android device with a camera, for the recognition of static gestures, that is, gestures that do not need the movement of hands, in LIBRAS. The proposed system uses a skin detection algorithm to segment the region of interest, then seeks to identify the hand of the individual and extract some characteristics regarding their position and finally recognize the gesture signaled through the information extracted using a classifier Support Vector Machine. The developed system obtained a rate of 86.5% accuracy in recognition of the evaluated gestures*

Resumo. *A Língua Brasileira de Sinais (LIBRAS) é a língua oficial na comunicação dos surdos no Brasil. Embora seja o principal meio de comunicação dos deficientes auditivos, a LIBRAS ainda é pouco difundida no resto da sociedade. Dessa forma, o reconhecimento de língua de sinais é uma área de pesquisa de grande relevância, pois tem o objetivo de incluir o surdo no resto da sociedade. Baseado nisso, propomos nesse artigo um sistema móvel, compatível com qualquer dispositivo Android com câmera, para o reconhecimento de gestos estáticos, ou seja, gestos que não necessitem do movimento das mãos, em LIBRAS. O sistema proposto utiliza um algoritmo de detecção de pele a fim de segmentar a região de interesse, em seguida procura-se identificar a mão do indivíduo e extrair algumas características referente ao seu posicionamento e finalmente reconhecer o gesto sinalizado através das informações extraídas usando um classificador Máquina de Vetor de Suporte. O sistema desenvolvido obteve uma taxa de 86,5% de acerto no reconhecimento dos gestos avaliados.*

1. Introdução

Segundo o censo do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2010 havia no Brasil uma população de mais de 190 milhões de pessoas, dessas 45,6 milhões (23,9%) apresentaram algum tipo de deficiência, sendo 9,7 milhões (5,1%) portadoras de deficiência auditiva. Algumas das dificuldades encontradas pelos deficientes podem ser contornadas por meio do uso de Tecnologias de Informação e Comunicação (TICs).

A apropriação de recursos tecnológicos modifica significativamente o estilo de vida, as interações e a conduta social ao inovar hábitos e atitudes em relação à educação, ao lazer e ao trabalho, à vida familiar e comunitária.

A dificuldade de se comunicar com a sociedade predominantemente ouvinte é um grande obstáculo no dia a dia para a comunidade surda. Isso ocorre principalmente pela falta de conhecimento dos ouvintes sobre a linguagem de sinais. Com isso, tarefas comuns e simples para a maioria da população, como fazer compras, realizar uma operação bancária e ir a uma consulta médica, podem ser um grande desafio para os deficientes auditivos, uma vez que a Língua Brasileira de Sinais (LIBRAS) é a única língua utilizada por eles para interagir com a sociedade.

Com base nas dificuldades apresentadas acima, propõe-se um sistema de Reconhecimento Automático de gestos estáticos em LIBRAS disponível para dispositivos que possuam câmera e o sistema operacional Android. Escolha é justificada por este ser o sistema mais utilizado no Brasil de acordo com os números divulgados em 2016 pela consultoria IDC, sendo que cerca de 93% dos aparelhos vendidos no Brasil possuem o Android.

O objetivo do sistema proposto é identificar um sinal em LIBRAS a partir da entrada fornecida pela câmera do dispositivo.

O restante do texto está dividido da seguinte forma: trabalhos relacionados são apresentados na Seção II. A metodologia proposta é apresentada na Seção III. Apresentamos os resultados e experimentos na Seção IV. Finalmente, temos as conclusões e os trabalhos futuros na Seção V.

2. Trabalhos Relacionados

Foi realizado uma busca por softwares que possuam o mesmo ou um objetivo semelhante ao sistema proposto. Na busca, foram encontrados sistemas, não necessariamente *mobile*, que fazem o reconhecimento de gestos de LIBRAS e o traduzem para português. No entanto, a pesquisa por sistemas móveis retornou apenas alguns aplicativos que fazem a conversão do português para a LIBRAS, ou seja, o inverso do objetivo mostrado nesse trabalho.

Tal busca também se estendeu para aplicações *Desktop*, agora retornando alguns programas listados na Tabela 1

Tabela 1. Resultado da pesquisa de anterioridade no INPI.

| Pedidos | Depósitos | Título |
|---------------------|------------|---|
| BR 51 2015 000885 4 | 04/08/2015 | TRADUTOR DE LIBRAS |
| BR 51 2014 000097 4 | 04/02/2014 | LOQUI-LIBRAS |
| 10695-4 | 19/05/2010 | SENSOR LIBRAS-SOFTWARE TRADUTOR DE LINGUA BRASILEIRA DE SINAIS (LIBRAS) |

Entre os programas listados na Tabela 1, apenas o terceiro realiza a tradução de LIBRAS para o português. Os demais também possuem o objetivo inverso. O programa Sensor Libras, no entanto, não pode ser comparado ao sistema proposto, visto que ele utiliza sensores que devem ser colocados nos pulsos do usuário, diminuindo consideravelmente a sua usabilidade, pois além de necessitar tal acessório, trata-se de uma aplicação que funciona apenas em desktop.

Além disso, foi feita uma busca científica para se ter conhecimento dos principais métodos proposto existente. Com isso, foi analisado o retorno da pesquisa, onde foram classificados como sendo de maior relevância para o desenvolvimento do sistema os seguintes trabalhos: [SIOLA 2010], [Koroishi 2016],[Brandt 2015], [Teodoro 2015] e [Souza et al. 2007]

Em [SIOLA 2010] o método possui diversas restrições entre elas: o processo deve ser realizado em um ambiente controlado, com um fundo de cor branca ou preta, somente o interprete da linguagem deve ser visível a câmera, o interprete deve utilizar roupas brancas ou pretas de manga comprida, uma luva de cor azul na mão direita deve ser usada para sinalizar os gestos e outra de cor laranja na mão esquerda, no início do vídeo o interprete deve estar com suas mãos posicionadas para cima e ao final do vídeo suas mãos deveram estar para baixo.

Ja em [Koroishi 2016] e [Brandt 2015], os autores estudaram a viabilidade do uso de sensores 3d, como o Kinect, para o reconhecimento da LIBRAS baseando-se nessa classificação para uma abordagem probabilística e com uso de redes neurais.

No trabalho de [Souza et al. 2007] o reconhecimento de um sinal em LIBRAS é feito através de uma sequência de várias imagens que são consideradas marcantes para o gesto representado, o seu trabalho dá ênfase no uso de Modelos de Markov Oculto (HMMs) para a classificação. Para testar o desempenho do método proposto foi construído um banco com amostras de imagem correspondente a 47 sinais diferentes, sendo que todas as imagens foram capturadas com o fundo estático e uniforme.

O último trabalho citado foi o de [Teodoro 2015] em seu projeto a autora desenvolveu um sistema de reconhecimento de algumas palavras em LIBRAS para desktop, que processa vídeos gravados e armazenados em um banco construído pelo próprio autor. Porém, apresenta uma usabilidade melhor do que os trabalhos anteriores, pois seu sistema não necessita do uso de luvas, nem outro tipo de acessório para o funcionamento. Nesse trabalho o sistema analisa a sequência de imagens gravadas que correspondem ao sinal, realiza uma etapa de pré-processamento para melhorar a imagem, em seguida segmenta a região de interesse através do método *Kovac*, faz o rastreamento do movimento da mão para identificar sinais dinâmicos, em seguida extrai as características utilizando o descritor *Random Forest* e reconhece o sinal.

Além dos trabalhos citados, o sistema desenvolvido por [Qin 2015] foi de grande relevância para esse projeto, nele foi desenvolvido um aplicativo que é capaz de abrir outros aplicativos instalados no celular através do reconhecimento dos dedos das mãos, porém seu sistema necessita de uma etapa de pré-treinamento sempre que é iniciado, precisando ser informado onde está o fundo e a mão do usuário.

3. Metodologia

Nesse trabalho, utilizamos o sistema operacional Android e a linguagem de programação Java. O sistema proposto é compatível com qualquer dispositivo Android, tendo apenas a câmera como pré-requisito funcional. O funcionamento do sistema ocorre em três etapas: detecção da mão, reconhecimento da mão e o reconhecimento do sinal. O Funcionamento do sistema está resumido no fluxograma apresentado na Figura 1 e será detalhado nas próximas seções.

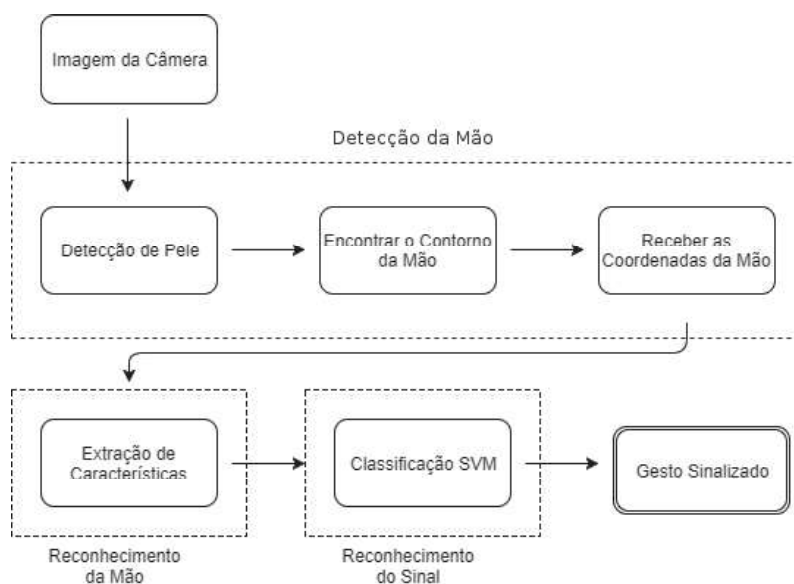


Figura 1. Fluxograma do sistema proposto.

3.1. Detecção da mão

Para a primeira etapa utilizamos um algoritmo de detecção de pele desenvolvido na linguagem Java, próprio para ser utilizado em Android. Esse algoritmo utiliza a câmera do dispositivo que o está executando e processa toda a imagem capturada, em seguida classifica cada pixel como sendo pele ou não pele, os pixels que forem classificados como não pele são preenchidos de preto, já os classificados como pele são preenchidos com um determinado tom de cinza, que varia de acordo com a exposição à luz, os pixels mais expostos tendem a ter um tom de cinza mais claro, já os que estiverem menos expostos terão um tom de cinza mais escuro. Com isso, obtém-se como saída do algoritmo uma matriz contendo somente a região da pele, como mostra a Figura 2.

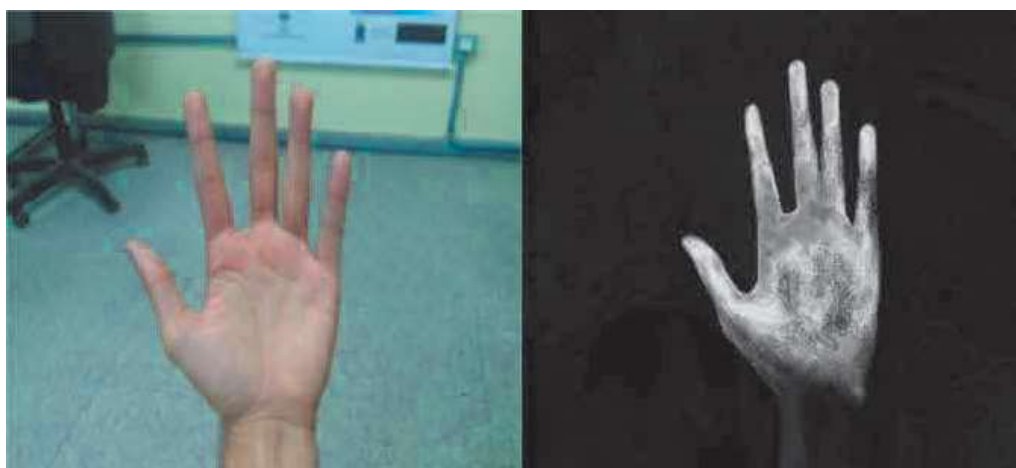


Figura 2. Matriz resultante da detecção de pele.

Após obter a matriz resultante com a região da pele detectada, o sistema aplica algumas técnicas na matriz a fim de melhorar a imagem com a região da mão. Para isso, inicialmente é aplicado o filtro da mediana com o objetivo de equilibrar a imagem que

possui pixels mais claros e mais escuros. Em seguida, a imagem passa por um processo de limiarização *threshold* disponível na biblioteca OpenCv, isso faz com que ela seja segmentada em todos os pixels que tiverem o valor inferior a taxa de *threshold*, no sistema proposto esta taxa foi definida com o valor 20, tal escolha foi feita devido a esse valor ter realizado a limiarização com mais eficiência, sem grandes perdas na imagem. Com a área de interesse (pele) segmentada, usamos a função da biblioteca OpenCv *findContours*, que nos retorna os contornos encontrados a partir da imagem passada.

3.2. Reconhecimento da mão

Como na etapa anterior a matriz foi restrita a conter apenas a região da mão, a função irá retornar somente os contornos correspondente aos da mão. A partir daí é criada uma caixa, chamada *boundingbox* na qual é limitada somente à região da mão, é a partir das informações representadas dentro do *boundingbox* que um sinal é reconhecido, como mostra a Figura 3.

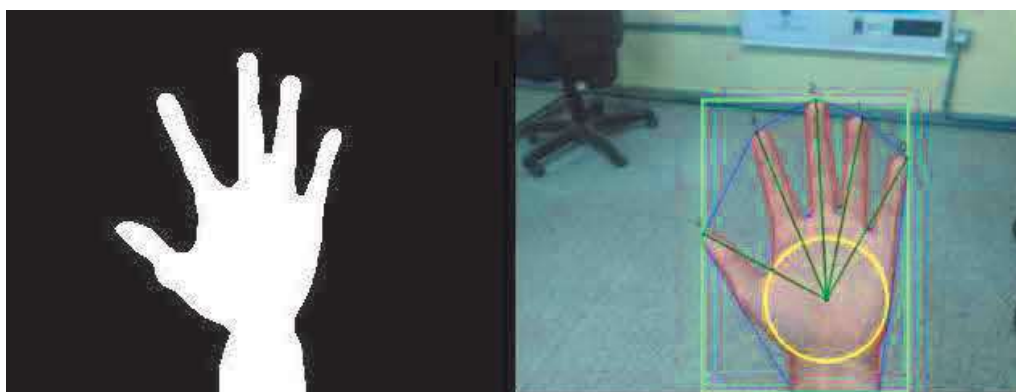


Figura 3. Representação do conjunto de informações.

As informações referentes a um sinal estão representadas da seguinte forma: o *boundingbox* está representado na Figura 3 pela caixa verde, as linhas em azul representa o menor polígono criado contendo todos pixels correspondente a mão, chamado de *convex hull*. As linhas em vermelho delimitam o contorno das mãos, já os pontos de cor azul são as áreas de *defect points* usados para dividir a área de contorno do final de um dedo e o início do outro. O círculo em amarelo representa a região da palma da mão, e finalmente as linhas em verde escuro são os vetores que representam os dedos e a ordem é indicada pelos números.

Os recursos utilizados para representar a configuração de mão usado em um sinal são basicamente os vetores de dedos. Esses vetores são calculados a partir do centro da palma da mão e a ponta dos dedos. O centro da palma da mão é calculado com o contorno e os pontos fornecidos pelo *convexhull* juntamente com o *boundingbox*. Já a ponta dos dedos é obtida utilizando somente as informações dos *defect points*. Com isso, os vetores de dedos correspondem ao valor da distância entre o centro da palma da mão até cada ponta dos dedos dividido pelo raio da palma da mão. Caso seja sinalizado um gesto que não exponha nenhum dedo, as características usadas para representar o sinal é o tamanho do formato que o contorno da mão irá assumir.

3.3. Reconhecimento do Gesto

A última etapa consiste no reconhecimento do gesto, nessa etapa utiliza-se o algoritmo de aprendizado de máquina SVM multiclases para treinar o sistema com as imagens cadastradas previamente que contêm as características na qual os gestos sinalizados devem conter para serem identificados corretamente.

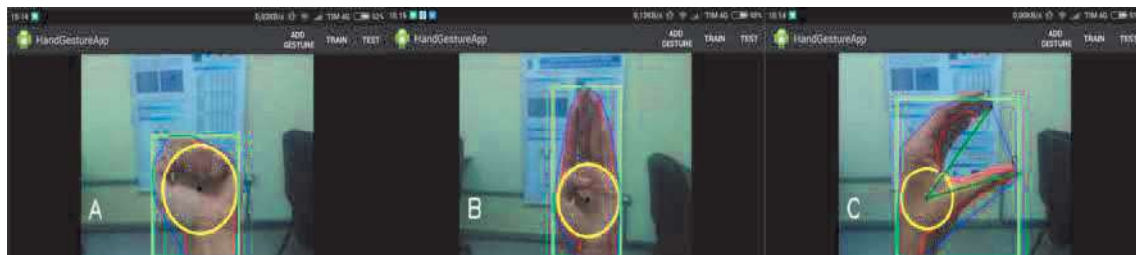


Figura 4. Reconhecimento das letras do alfabeto.

4. Resultados e Experimentos

Foram realizados testes para validar a eficiência do sistema proposto. Esses testes foram realizados em um smartphone Xiaomi Redmi 2 que utiliza o sistema operacional Android na versão 4.4 com as seguintes configurações:

- Modelo: Redmi 2;
- Marca: Xiaomi;
- Versão do Android: 4.4.4 KTU84P;
- Chipset: Qualcomm MSM8916 Snapdragon 410;
- CPU: Quad-core Max 1,2 GHz;
- Memória RAM: 1GB;
- GPU: Qualcomm Adreno 306;
- Câmera: 8 Mp.

Para testar a aplicação foram convidadas 5 pessoas que se disponibilizaram para sinalizar os gestos selecionados para o teste. Os voluntários escolhidos possuem diferentes tonalidades de cor de pele para evidenciar o desempenho do algoritmo desenvolvido com o maior leque de características possíveis, além disso, nenhum dos participantes possui conhecimentos avançados em LIBRAS.

Os gestos escolhidos para serem testados foram as 10 primeiras letras do alfabeto que possuam gesto estático, ou seja, que não seja necessário realizar nenhum movimento com a mão para o sinal ser executado. O sistema foi treinado com 10 amostras de gestos de cada letra, com isso as letras a serem analisadas são: “A”, “B”, “C”, “D”, “E”, “F”, “G”, “I”, “L” e “M”.

Os testes foram executados em ambientes com o fundo heterogêneo e não controlado. Foi criado um pequeno algoritmo para calcular o resultado dos testes, onde é contabilizado a quantidade de vezes que o sistema classificou a letra sinalizada corretamente dividido pelo total, ou seja, quantidade de vezes que o sistema classificou o sinal de forma correta ou incorreta durante os 50 primeiros frames da câmera, o que corresponde a mais ou menos de 7 a 10 segundos para cada letra. Os resultados dos testes variam de 0

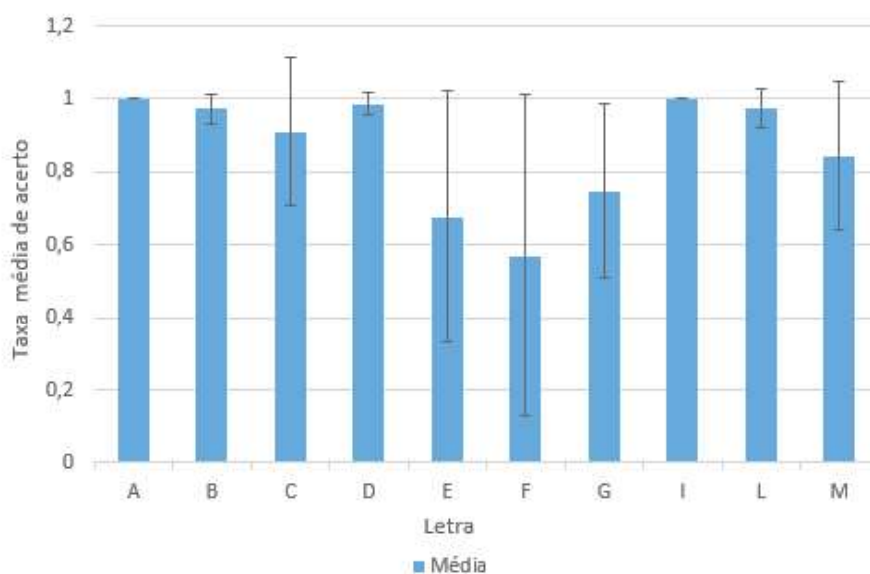
Tabela 2. Resultado dos testes de cada letra analisada.

| Letra | Voluntário 1 | Voluntário 2 | Voluntário 3 | Voluntário 4 | Voluntário 5 |
|-------|--------------|--------------|--------------|--------------|--------------|
| A | 1,0 | 1,0 | 1,0 | 1,0 | 1,0 |
| B | 0,91 | 1,0 | 0,95 | 1,0 | 1,0 |
| C | 1,0 | 1,0 | 1,0 | 0,55 | 1,0 |
| D | 1,0 | 1,0 | 0,93 | 1,0 | 1,0 |
| E | 0,79 | 0,34 | 1,0 | 0,97 | 0,28 |
| F | 1,0 | 0,61 | 0,16 | 0,08 | 1,0 |
| G | 1,0 | 0,71 | 0,97 | 0,43 | 0,63 |
| I | 1,0 | 1,0 | 1,0 | 1,0 | 1,0 |
| L | 1,0 | 1,0 | 1,0 | 0,88 | 1,0 |
| M | 0,69 | 1,0 | 0,97 | 1,0 | 0,56 |

(caso o sistema em nenhum momento tenha reconhecido a letra corretamente) a 1 (caso o sistema tenha classificado corretamente durante todo o período de análise), os resultados podem ser conferidos na Tabela 2.

Os resultados mostram que o sistema apresentou um desempenho satisfatório, tendo em vista que os voluntários não possuíam conhecimento em LIBRAS, consequentemente não apresentando habilidades para sinalizar os gestos com perfeição, isso fica evidente ao perceber que os gestos considerados mais fáceis obtiveram maiores taxas de acerto, enquanto os mais difíceis como a letra “F” obteve uma taxa menor.

Além disso, o conjunto de gestos usado nos testes contém gestos que se assemelham, como a das das letras “A” e “E” que não apresentam os dedos expostos o que seria determinante para diferenciá-los com mais precisão. A Figura 5 mostra a média das porcentagens exibidas anteriormente para cada letra.

**Figura 5. Desempenho do sistema para cada letra.**

Baseado nos resultados analisados na Figura 5 ao fazer uma média dos valores de cada letra podemos afirmar que o sistema faz o reconhecimento de gestos com uma taxa

de 86,5 % de acerto. Esse número foi significativamente afetado pelas baixas taxas de acerto das letras “E”, “F” e “G” já citadas anteriormente.

5. Conclusão e Trabalhos Futuros

O Trabalho proposto mostra-se uma ferramenta bastante relevante para facilitar a comunicação e inclusão entre surdos que utilizam a Língua Brasileira de Sinais com a população ouvinte que não conhece LIBRAS. Um dos principais objetivos desse trabalho é construir uma aplicação com uma boa usabilidade que possa ser utilizada facilmente no dia a dia. Apesar do sistema ter sido testado reconhecendo letras, pode ser usado também para reconhecer qualquer gesto estático. Os testes mostraram uma taxa de 86,5 % de eficiência, esse valor pode ser melhorado aumentando a quantidade de amostras para o treinamento dos gestos, pois assim o sistema será capaz de reconhecer os sinais quando executados com pequenas variações.

Como trabalhos futuros, pretende-se expandir o sistema para o reconhecimento de palavras que possuem sinais dinâmicos, ou seja, gestos sinalizados com o movimento da mão, isso pode ser feito usando o rastreamento da mão para mapear o percurso que ela irá percorrer no momento da sinalização.

Referências

- Brandt, A. H. (2015). Lirann: sistema de reconhecimento de libras baseado em redes neurais artificiais com kinect. *Trabalho de Conclusão de Curso, Universidade de Santa Cruz do Sul, Santa Cruz do Sul*, pages 30–38.
- Koroishi, G. O. (2016). Reconhecimento de sinais da libras por visão computacional. *Mecatrone*, 1(1):2–7.
- Qin, Y. (2015). Color-based hand gesture recognition on android.
- SIOLA, F. (2010). Desenvolvimento de um software para reconhecimento de sinais libras através de vídeo. *Trabalho de Conclusão de Curso, Universidade Federal do ABC, Santo André*, pages 32–41.
- Souza, K. P., Dias, J. B., and Pistori, H. (2007). Reconhecimento automático de gestos da lingua brasileira de sinais utilizando visao computacional. In *III Workshop de Visao Computacional*, pages 3–5.
- Teodoro, B. T. (2015). *Sistema de reconhecimento automático de Língua Brasileira de Sinais*. PhD thesis, Universidade de São Paulo.