

**BRUNA VIEIRA LOUZADA SILVA
GIOVANNA ONO KOROISHI**

**RECONHECIMENTO DE SINAIS DA
LIBRAS POR VISÃO COMPUTACIONAL**

Texto apresentado à Escola Politécnica da
Universidade de São Paulo como requisito
para a conclusão do curso de graduação
em Engenharia Mecatrônica, junto ao
Departamento de Engenharia Mecatrônica
e de Sistemas Mecânicos (PMR).

São Paulo
2014

**BRUNA VIEIRA LOUZADA SILVA
GIOVANNA ONO KOROISHI**

**RECONHECIMENTO DE SINAIS DA
LIBRAS POR VISÃO COMPUTACIONAL**

Texto apresentado à Escola Politécnica da
Universidade de São Paulo como requisito
para a conclusão do curso de graduação
em Engenharia Mecatrônica, junto ao
Departamento de Engenharia Mecatrônica
e de Sistemas Mecânicos (PMR).

Orientador:

Prof. Dr. Thiago de Castro Martins

São Paulo
2014

Catálogo-na-publicação

Koroishi, Giovanna Ono

**Reconhecimento de sinais da LIBRAS por visão computacional / G.O. Koroishi; B.V.L. Silva. – São Paulo, 2014.
p.**

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos.

**1.Reconhecimento de imagem 2.Visão computacional
3.Língua Brasileira de Sinais I.Silva, Bruna Vieira Louzada
II.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos III.t.**

AGRADECIMENTOS

Aos nossos pais e irmãos, pelo amor e carinho incondicionais e todos os ensinamentos transmitidos ao longo da jornada que nos trouxe até aqui.

A Pedro e Renato, pela paciência e apoio durante os momentos mais difíceis.

Ao nosso orientador, por insistir que sempre há uma luz no fim do túnel, mesmo quando pensávamos que era um trem vindo em nossa direção.

A Amanda e Stephanie, pela amizade e colaboração com a revisão e os desenhos.

Aos professores, pelo tempo dedicado à nossa formação, incluindo as dúvidas respondidas durante a madrugada.

E a todos aqueles que nos acompanharam nos últimos anos, sempre torcendo para que alcançássemos o tão sonhado diploma, o nosso muito obrigada.

RESUMO

O desenvolvimento da tecnologia em prol da acessibilidade de deficientes é um assunto que desperta o interesse de pesquisadores ao redor do mundo. Em particular à comunicação entre surdos e ouvintes, é possível notar grande desenvolvimento na via de tradução linguagem falada para linguagem de sinais. Entretanto, a via contrária possui o desafio técnico de visualização e reconhecimento de sinais, o qual é o mais complexo no caso brasileiro, já que a LIBRAS utiliza dedos largamente para expressar os sinais. Por outro lado, os sinais da LIBRAS possuem uma classificação Semântica que permitiu o desmembramento do problema em partes mais simples. Este projeto é um estudo da viabilidade do uso de sensores 3D para o reconhecimento automático de um conjunto de sinais da LIBRAS, baseando-se nessa classificação para uma abordagem probabilística. Durante o projeto, foi implementada uma prova de conceito funcional de um identificador automático de sinais da LIBRAS. O protótipo desenvolvido para essa verificação captura sinais através do sensor Kinect, segmenta o objeto de interesse (a mão direita) e calcula a probabilidade dos sinais. O principal desafio do trabalho foi reconhecer as configurações de mão e para solucioná-lo utilizou-se modelos estruturados em nuvens de pontos e o algoritmo de ICP. O projeto mostrou que essa abordagem torna viável o reconhecimento automático, pois obteve-se 65% de acerto entre 48 testes envolvendo 12 sinais diferentes, mesmo com as limitações de recursos e tempo existentes.

Palavras-chaves: LIBRAS, reconhecimento automático, visão computacional, Kinect, nuvem de pontos

ABSTRACT

Development of technologies towards the accessibility of the disabled is a subject that stimulates researchers all over the world. Specifically in the field of communication between deaf and listeners, is notable the development of tools to translate from the spoken language to the sign language, while the opposite direction is more technically challenging, due to the visualization and sign recognition. It is even more complex to automatically recognize the Brazilian signs, since LIBRAS widely uses the fingers to express the signs. On the other hand, the LIBRAS signs have a chermic classification that allowed to split the problem into simpler ones. This project is a study of the feasibility of using 3D sensors to automatically recognize a set of LIBRAS signs, based on this classification to a probabilistic approach. During the project, a proof of concept of an automatic LIBRAS sign recognizer was implemented. The developed prototype for this verification records the signs by the Kinect sensor, segments the object of interest (the right hand) and calculates the sign probabilities. The main challenge of the work was to recognize the hand's configuration, models structured in cloud points and the ICP algorithm were used to solve them. The project showed that this approach makes automatic recognition feasible, as it reached the level of 65% of correct signs in 48 tests with 12 different signs, even with the limited resources and time.

Keywords: LIBRAS, automatic recognition, computer vision, Kinect, cloud points

SUMÁRIO

Lista de Ilustrações

Lista de Tabelas

Lista de Abreviaturas e Siglas

1	Introdução	12
1.1	Tema	12
1.2	Justificativa da Escolha do Tema	12
1.3	Estado da Arte	13
2	Linguagem Brasileira de Sinais	17
2.1	Introdução à LIBRAS	17
2.2	Conceitos importantes	17
2.3	Classificação dos sinais	18
3	Requisitos do Projeto	20
3.0.1	Requisitos Funcionais	20
3.0.2	Requisitos Não-Funcionais	21
3.0.2.1	Requisitos do produto	21
3.0.2.2	Requisitos organizacionais	22
4	Análise e Design	23
4.1	Arquitetura	23
4.1.1	Diagrama de Sequência	25
4.2	Sensor Utilizado	26

5	Captação e segmentação	28
5.1	Captação da imagem	28
5.2	Segmentação da mão	28
6	Reconhecimento SematosÊmico	29
6.1	Reconhecimento dos SematosEmas de movimento	29
6.2	Reconhecimento dos SematosEmas de local de articulação	30
6.3	Reconhecimento do SematosEma de configuração de mão	32
6.3.1	Point Cloud Library	32
6.3.2	Banco de dados de modelos	32
6.3.3	Iterative Closest Point	33
6.3.4	Cálculo das probabilidades <i>a posteriori</i> dos SematosEmas de articulação de mão	34
7	Reconhecimento do MorfEma	38
7.1	Banco de dados de MorfEmas	38
7.1.1	Coordenar	39
7.1.2	Depressa	39
7.1.3	Deus	40
7.1.4	Elevar	40
7.1.5	Em cima	40
7.1.6	Gêmeo	41
7.1.7	Genro	41
7.1.8	Lado direito	41
7.1.9	Nora	42
7.1.10	Norte	42
7.1.11	Para cima	42

7.1.12 Teto	43
7.2 Paralelização	43
8 Implementação	44
8.1 Interface gráfica	44
9 Resultados	48
9.1 Limitações do sistema	49
10 Conclusão	50
11 Sugestões para Trabalhos Futuros	51
Referências	52
Apêndice A – Modelos	55

LISTA DE ILUSTRAÇÕES

1	Ilustrações dos sinais utilizados como exemplo.	18
2	Diagrama de caso de uso do protótipo.	20
3	SematosEmas escolhidos como requisitos funcionais do projeto. . .	21
4	Arquitetura do protótipo.	23
5	Detalhamento do processo de reconhecimento do MorfEma.	24
6	Diagrama de sequência do sistema.	25
7	Componentes do sensor Kinect. Adaptado de: Microsoft Developer Network.	26
8	Imagem de profundidade e esqueleto capturada pelo Kinect. . . .	27
9	Representação dos ângulos em relação aos seis semi-eixos.	29
10	Regiões dos SematosEmas de local de articulação. Adaptado de Microsoft Developer Network.	31
11	Comparação da largura de dois modelos.	34
12	Frames alinhados com seus respectivos modelos.	34
13	Coordenar: Mão em C, palma para a esquerda, tocando o peito. Movê-la para a frente.	39
14	Depressa: Mão em C, palma para a esquerda. Movê-la rapidamente para a esquerda e para a direita, diante da boca.	39
15	Deus: Mão em D, palma para a esquerda. Movê-la para cima, acima da cabeça.	40
16	Elevar: Mão aberta com os dedos flexionados. Movê-la do peito até a altura da cabeça.	40
17	Em cima: Mão em 1, palma para frente, na altura do ombro. Mover a mão para cima.	40

18	Gêmeo: Mão em G, palma para a frente. Mover a mão para a esquerda e para a direita.	41
19	Genro: Mão em G, palma para frente. Movê-la para a direita. . .	41
20	Lado direito: Mão em D, palma para frente, ao lado direito do corpo. Mover a mão para baixo.	41
21	Nora: Mão em N. Movê-la para a direita.	42
22	Norte: Mão em N. Movê-la para cima.	42
23	Para cima: Mão em 1, palma para a frente, na altura da cabeça. Movê-la ligeiramente para cima.	42
24	Teto: Mão aberta e dedos flexionados, palma para a esquerda, em cima da cabeça. Mover a mão para frente.	43
25	Interface Homem Máquina (IHM) ao iniciar o programa.	45
26	Sequencia da IHM ao analisar um sinal.	46
27	Interface Homem Máquina (IHM) após reconhecimento do sinal. .	47
28	Resultados dos testes: posições da resposta correta na resposta do sistema	48
29	Gráfico da porcentagem de acerto de MorfEmas em primeiro lugar pela quantidade.	49

LISTA DE TABELAS

1	Especificações do sensor Kinect	26
2	Probabilidades do posicionamento da mão dada a região identificada pelo sensor.	31
3	Resultados do ICP para dois testes com o SematosEma <i>C</i>	37
4	Especificações do computador utilizado nos testes	44

LISTA DE ABREVIATURAS E SIGLAS

2D	Bidimensional
3D	Tridimensional
GPU	<i>Graphic Processor Unit</i>
HMM	<i>Hidden Markov Models</i>
IHM	Interface Homem Máquina
ICP	<i>Iterative Closest Point</i>
LGP	Língua Gestual Portuguesa
LIBRAS	Linguagem Brasileira de Sinais
PCL	<i>Point Cloud Library</i>
RGB	<i>Red Blue Green</i>
RGBD	<i>Red Blue Green Depth</i>
SDK	<i>Software Development Kit</i>
USB	<i>Universal Serial Bus</i>
USP	Universidade de São Paulo

1 INTRODUÇÃO

1.1 Tema

O projeto teve como objetivo estudar a viabilidade do uso de sensores 3D para o reconhecimento automático de sinais da LIBRAS por visão computacional.

1.2 Justificativa da Escolha do Tema

De acordo com a Organização Mundial da Saúde (1), cerca de 360 milhões de pessoas possuem surdez incapacitante. Isso significa que mais de 5% da população mundial possui, no ouvido mais aguçado, perda auditiva maior que 40dB em adultos e 30dB em crianças, englobando tipos de suave a severo. Entre essa minoria, há aqueles cuja principal consequência da surdez é o impacto na comunicação verbal. Nesses casos, a comunicação pode ser feita de diferentes formas, tais como a leitura labial, a escrita, a leitura e a linguagem de sinais. Entretanto, as três primeiras formas são intrinsecamente ligadas à linguagem falada, enquanto a última possui suas próprias regras e estrutura gramatical (2), mostrando que tanto a linguagem de sinais quanto a falada são independentes e passíveis de tradução.

Há um grande desenvolvimento de tecnologias de tradução no sentido da linguagem falada para linguagem de sinais, existindo, por exemplo, aplicativos para celular que recebem frases digitadas e apresentam uma simulação dos sinais. Já o sentido contrário possui o desafio técnico da visualização e do reconhecimento dos sinais. No caso brasileiro, esse desafio se mostra ainda mais complexo, já que a LIBRAS (Linguagem Brasileira de Sinais) possui rica diversidade dos SematosEmas (conceito detalhado na seção 2.2) de articulação de mão, sendo muito importante distinguir a configuração das mãos e dos dedos, o que torna mais complexo o reconhecimento dos MorfEmas (conceito detalhado na seção 2.2) (3). Sendo assim, este trabalho é um estudo da viabilidade do uso de sensores 3D

para o reconhecimento automático de um conjunto de MorfEmas da LIBRAS.

1.3 Estado da Arte

Abordagens diversas já foram apresentadas para a identificação de sinais (da LIBRAS ou outras línguas) por meio de software. Há trabalhos que empregam apenas câmeras de vídeo RGB (Red, Green and Blue: sistema de representação de cores em que cada cor é representada com três valores, um para cada cor primária, vermelho, verde e azul), outros sensores RGBD (Red, Green, Blue and Depth: imagem que além da cor de cada pixel, possui a informação da distância do pixel à câmera) e outros sinais de sensores inerciais ou elétricos. Quanto ao uso de acessórios também há variações, alguns não os utilizam, enquanto outros tornam o contexto da utilização mais restrito ao exigir que o usuário faça uso de luvas coloridas ou com sensores de distâncias, acelerômetros ou eletrodos.

Dois trabalhos utilizam apenas dados de vídeos de pessoas com luvas coloridas: no trabalho de Dias e Peres(4) a trajetória do movimento é identificada pelo cálculo do centróide da mão segmentada e o padrão é reconhecido por Redes Neurais Artificiais. Digiampietri et al.(5) apresentam um sistema de informação para o reconhecimento automático da LIBRAS, composto por dois projetos. O primeiro são módulos para reconhecimento automático e o segundo é um ambiente para gerenciar experimentos científicos. Para processar o vídeo, a segmentação de imagens foi desenvolvida de três modos: mapa de cores feito pelo usuário manualmente, inteligência artificial de agrupamento com informação do usuário sobre qual segmento o grupo pertence, e algoritmo de classificação do Weka, software que agrupa algoritmos de aprendizado de máquina para extração de dados (6). A extração de características pode ser feita por três extratores: extrator de forma, área proporcional de cada segmento ou posição relativa de cada segmento.

Uma abordagem diferente é apresentada por Li et al.(7), que criam um sistema portátil utilizando um acelerômetro e quatro sensores eletromiográficos em cada antebraço para fazer o reconhecimento da língua de sinais chinesa pelos parâmetros: configuração das mãos (acelerômetros), orientação (eletrodos) e movimento (ambos). Os sinais elétricos dos eletrodos entre o fim de um sinal e o início do outro são baixos e isso é usado para fazer a separação entre dois sinais. O reconhecimento dos movimentos é feito com uma abordagem multi-stream do

Modelo Oculto de Markov (Hidden Markov Models - HMM), em que os modelos dos sinais dos acelerômetros e dos eletrodos são avaliados separadamente e os resultados são combinados depois. A média de acertos foi de 96,5% para 120 sinais e 86,7% para 200 sentenças. Os resultados são bons e o artigo explica cada fase do reconhecimento, porém não contempla todas as variáveis necessárias para reconhecimento da LIBRAS.

O uso de acessórios é pouco prático para o usuário, e por isso este projeto deseja fazer o reconhecimento sem que o usuário precise vesti-los. Muitos trabalhos que seguem esta linha utilizam-se ou de câmeras de vídeo ou de sensores RGBD para obter informações 3D dos sinais.

Bons resultados (média de acertos de 95,5%) foram obtidos por Yang(8), porém apenas para sinais estáticos (letras). O trabalho utiliza visão computacional para reconhecimento de 30 letras da língua chinesa de sinais a partir de vídeos de mãos apenas. As informações extraídas dos vídeos são: a diferença de frames para detectar regiões em movimento, detecção de pele e cinco descritores visuais e geométricos. O reconhecimento da letra é feito através de Support Vector Machine, aprendizado supervisionado.

Vogler e Metaxas(9) é um dos primeiros trabalhos que utiliza modelos 3D e Modelos Ocultos de Markov para fazer o reconhecimento de sinais. Neste artigo é descrito o reconhecimento da Língua Americana de Sinais a partir do uso de 3 câmeras que fornecem imagens para a criação do modelo 3D. É reconhecido o movimento de apenas um braço do usuário, sem as mãos. Os resultados mostram que os modelos 3D são melhores do que os 2D. Também são feitos experimentos com sinais isolados e frases com mais de uma palavra. O mesmo autor, em (10), estuda cada fonema como um estado do HMM, mas com uma abordagem paralela, ou seja, cada canal determina independentemente a probabilidade de um parâmetro (configuração da mão, orientação, posição e movimento), depois as probabilidades são multiplicadas para que se obtenha o fonema mais provável. São feitos experimentos também com pessoas caminhando, uma vez que a proposta é estudar reconhecimento de sinais da língua, já que esta possui classificação e configurações conhecidas, para depois estender o sistema para reconhecimento de outros movimentos humanos. O trabalho mostra que a abordagem paralela é mais eficiente para o reconhecimento de sentenças do que a regular (81 x 85% de acurácia). Apesar de abordar o reconhecimento de sinais da língua mais como

um meio para o estudo do que como finalidade, têm características interessantes, como o método paralelo do HMM.

Porfírio(11) grava com o Kinect da Microsoft as visões frontal e lateral da mão para reconhecer as 61 configurações de mão da LIBRAS com a criação de malhas 3D . De cada gravação são extraídos manualmente os quadros desejados e seus ruídos são corrigidos. As malhas 3D são geradas por reconstrução de silhueta e passam por suavização. O método de comparação usado é o descritor Spherical Harmonics, por ser invariante à rotação. Finalmente, a classificação da malha é feita com Support Vector Machine. Os melhores resultados foram de 98,52%. O banco de dados utilizado é próprio e armazena, além das malhas 3D, informações 2D das imagens captadas. Os resultados obtidos são bons, porém o método utilizado, como a escolha de quadros manualmente e tratamento de ruído das imagens, impossibilita que a classificação seja empregada automaticamente.

Uma equipe formada por pesquisadores da Chinese Academy of Sciences, Beijing Union University e da Microsoft Research Asia estão desenvolvendo um protótipo de um tradutor de linguagem de sinais com o uso do sensor Kinect (12). O objetivo é criar um sistema que traduza linguagens de sinais chinesa e norte-americana, das linguagens de sinais para as faladas, e vice-versa, com baixo custo. A tradução da língua falada para a de sinais será auxiliada por um avatar, que fará movimentos; na tradução oposta, serão gerados texto e simulação de voz. Esse é um esforço conjunto para dar acessibilidade aos deficientes auditivos, potencialmente podendo ser traduzido para outras linguagens de sinais.

O Kinect também pode ser uma ferramenta de ensino de linguagens de sinal. Zafrulla et al.(13) descrevem o desenvolvimento de um jogo chamado CopyCat para crianças surdas treinarem a linguagem norte-americana de sinais. O jogo consiste em apresentar imagens e a criança deve acertar os sinais que a descrevem. O usuário clica para começar os sinais e clica novamente para dizer que encerrou, além de vestir luvas coloridas que possuem também acelerômetros nos punhos. Como o objetivo é verificar se a frase está correta, a verificação é feita do fim da frase para o começo, ignorando possíveis palavras erradas feitas no início da frase.

Para o ensino da LIBRAS especificamente, Santos et al.(14) construíram um protótipo chamado KINLIB. O Kinect é utilizado para captura dos vídeos e a biblioteca Candescent para o reconhecimento de gestos. O trabalho concentra-se

em relatar a escolha do hardware e do software e os testes registrados são apenas para o reconhecimento de gestos estáticos. Este trabalho levou a dupla a consultar a biblioteca citada, porém esta não se mostrou satisfatória aos propósitos do projeto.

Um trabalho de destaque, porém não aplicado ao reconhecimento de linguagem de sinais é o trabalho de Oikonomidis, Kyriazis e Argyros(15), que faz o rastreamento de duas mãos interagindo de maneira complexa. Os dados são obtidos com um sensor RGBD (Kinect), a uma taxa de 4 Hz. Um modelo 3D de cada mão é feito com 26 graus de liberdade e juntas possuem 54 parâmetros. Simuladas as possíveis posições das duas mãos, é utilizado o método de otimização de enxame de partículas para obter qual configuração corresponde à imagem captada, tais cálculos são feitos em GPU. Os erros são da ordem de 6mm para mãos a 0,5m do sensor.

2 LINGUAGEM BRASILEIRA DE SINAIS

2.1 Introdução à LIBRAS

Como referência teórica da LIBRAS, além de um curso básico online (16), adotou-se o livro Novo Deit-LIBRAS (3). A primeira parte do livro contém introdução, breve histórico da LIBRAS e estudos sobre a educação de surdos no Brasil. Há uma descrição sistemática e classificatória dos elementos que compõem um sinal da LIBRAS. O corpo principal do Novo Deit-LIBRAS é um dicionário Português-LIBRAS-Inglês, com ilustrações e explicações sobre as configurações e movimentos dos sinais da LIBRAS, além de escrita visual direta por meio do SignWriting (sistema de escrita usado nas línguas gestuais). O livro é uma referência em LIBRAS, seus autores possuem muitas publicações na área e o livro foi distribuído aos alunos surdos das escolas da rede pública de ensino.

As linguagens de sinais possuem vocabulários, gramáticas, semânticas e sintaxes próprias, não são meros gestos mímicos das línguas faladas, por isso, cada país possui sua própria linguagem de sinais. Mesmo em países que falam a mesma língua, a língua de sinais é diferente, um exemplo é Portugal, que utiliza a Língua Gestual Portuguesa (LGP). Na LIBRAS também há regionalismos, alguns estados representam uma mesma palavra por meio de sinais diferentes, além de existirem sinais que são usados apenas em certos grupos.

2.2 Conceitos importantes

Este trabalho utiliza dois conceitos da LIBRAS muito importantes: SematosEma e MorfEma. SematosEmas são as menores unidades sublexicais da sinalização capazes de distinguir um sinal do outro, ou seja, é o detalhamento de como articular o sinal de acordo com determinados elementos; enquanto MorfEmas são as menores unidades sublexicais que codificam significado (3). Na língua portuguesa,

os SematosEmas poderiam ser confundidos com letras, e os MorfEmas, com palavras. Entretanto, o conceito se mostra mais complexo, pois um único sinal da LIBRAS pode compreender dois MorfEmas.

Para ilustrar esses conceitos, observamos os seguintes sinais: estudar (figura 1a) e universidade (figura 1b). O sinal *estudar* é, por si só, um MorfEma composto pelos SematosEmas: mãos abertas, palmas para cima, bater duas vezes o dorso dos dedos direitos sobre a palma dos dedos esquerdos (3). Percebe-se, por esse exemplo, que a descrição SematosÊmica é suficiente para a realização de um sinal. O sinal *universidade*, por sua vez, é composto pelo sinal *estudar*, seguido dos SematosEmas: mão em U, palma para frente, movê-la em um círculo vertical para a esquerda no sentido anti-horário (3). Como o sinal *universidade* é composto pelo MorfEma *estudar*, e, portanto, não é uma unidade mínima de significado, não pode ser classificado como MorfEma.

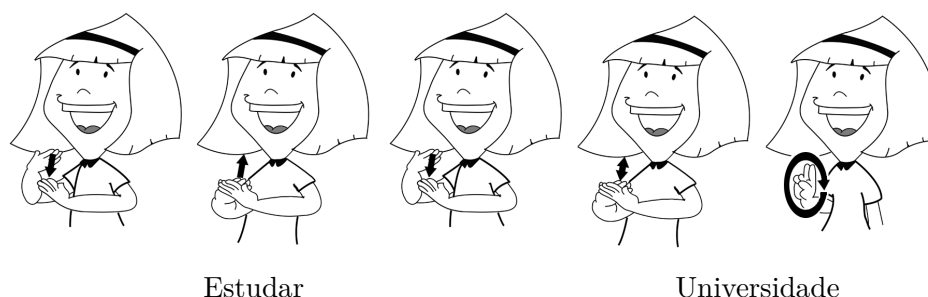


Figura 1: Ilustrações dos sinais utilizados como exemplo.

O reconhecimento de sinais compostos é um problema relacionado ao processamento de linguagem natural, o qual não seria possível ser abordado dado o prazo de um ano do projeto. Por isso, o protótipo foi limitado ao reconhecimento de sinais constituídos de apenas um MorfEma.

2.3 Classificação dos sinais

O Deit-LIBRAS classifica em quatro os elementos SematosÊmicos que compõem um sinal: articulação da mão e do braço e orientação da palma da mão; local da articulação; movimento e expressão facial (quando aplicável). Essa classificação foi utilizada como base para o sistema de reconhecimento de MorfEmas a partir dos SematosEmas, pois indica quais variáveis devem ser analisadas e identificadas a partir da aquisição dos vídeos.

Nos trabalhos de Guimarães (2), (17), (18), embora utilize outra nomenclatura, a classificação dos sinais também é relevante. O autor descreve em seus artigos o processo de avaliação de um modelo computacional descritivo para o reconhecimento da estrutura de sinais da LIBRAS, contando com exemplos de palavras diferenciadas por pequenas alterações de movimentos e referências de posição, além de código em XML para descrever a estrutura do sinal. Os resultados demonstram que o modelo contempla aspectos importantes para o reconhecimento correto de um sinal por computadores.

A importância dessa classificação pode ser observada também em Duduchi e Capovilla(19), os quais utilizam os SematosEmas para a organização dos sinais, uma vez que a ordem alfabética existente na língua portuguesa não se aplica à LIBRAS. O trabalho descreve um protótipo de dicionário digital que não exige conhecimento da escrita portuguesa da palavra procurada, o usuário faz a busca do sinal por meio da seleção de SematosEmas que o compõem.

3 REQUISITOS DO PROJETO

3.0.1 Requisitos Funcionais

O processamento do reconhecimento dos sinais deve ser realizado em um desktop com configurações padrões e deve respeitar o diagrama de caso de uso observado na figura 2.

De forma a tornar o uso do protótipo o mais natural possível, não deve ser necessário utilizar acessórios junto ao corpo para operar o sistema. Por isso, os dados utilizados para análise do sinal devem ser captados por um sensor que capture imagens.

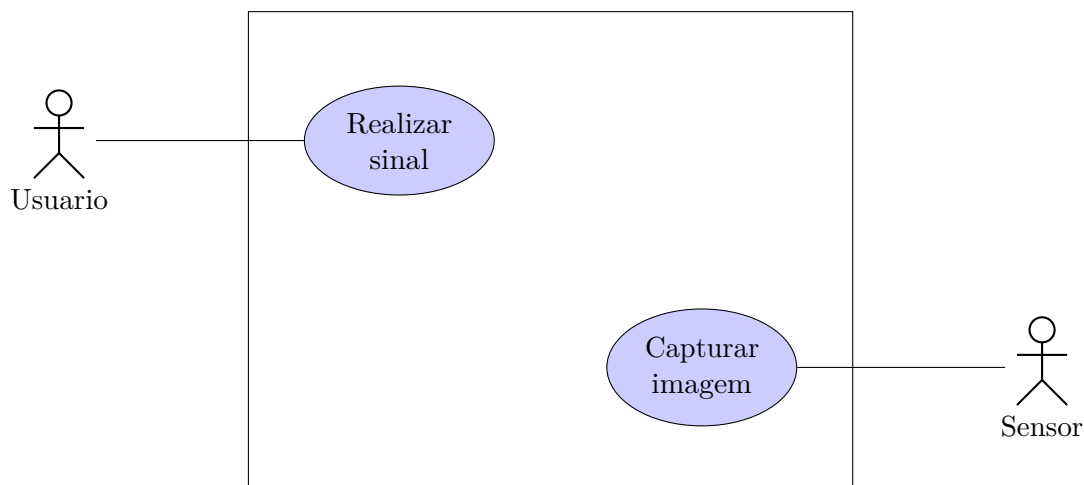


Figura 2: Diagrama de caso de uso do protótipo.

Para reconhecer o sinal, o protótipo deve identificar os SematosEmas que compõem o MorfEma e entregar as probabilidades de cada SematosEma presente para, então, calcular a probabilidade do MorfEma. A partir da classificação de SematosEmas dada pelo Novo Deit-LIBRAS (3), decidiu-se identificar apenas SematosEmas dos tipos articulação da mão e do braço e orientação da palma, local da articulação e movimento. Os SematosEmas selecionados dentre todas as possibilidades estão listados na figura 3.

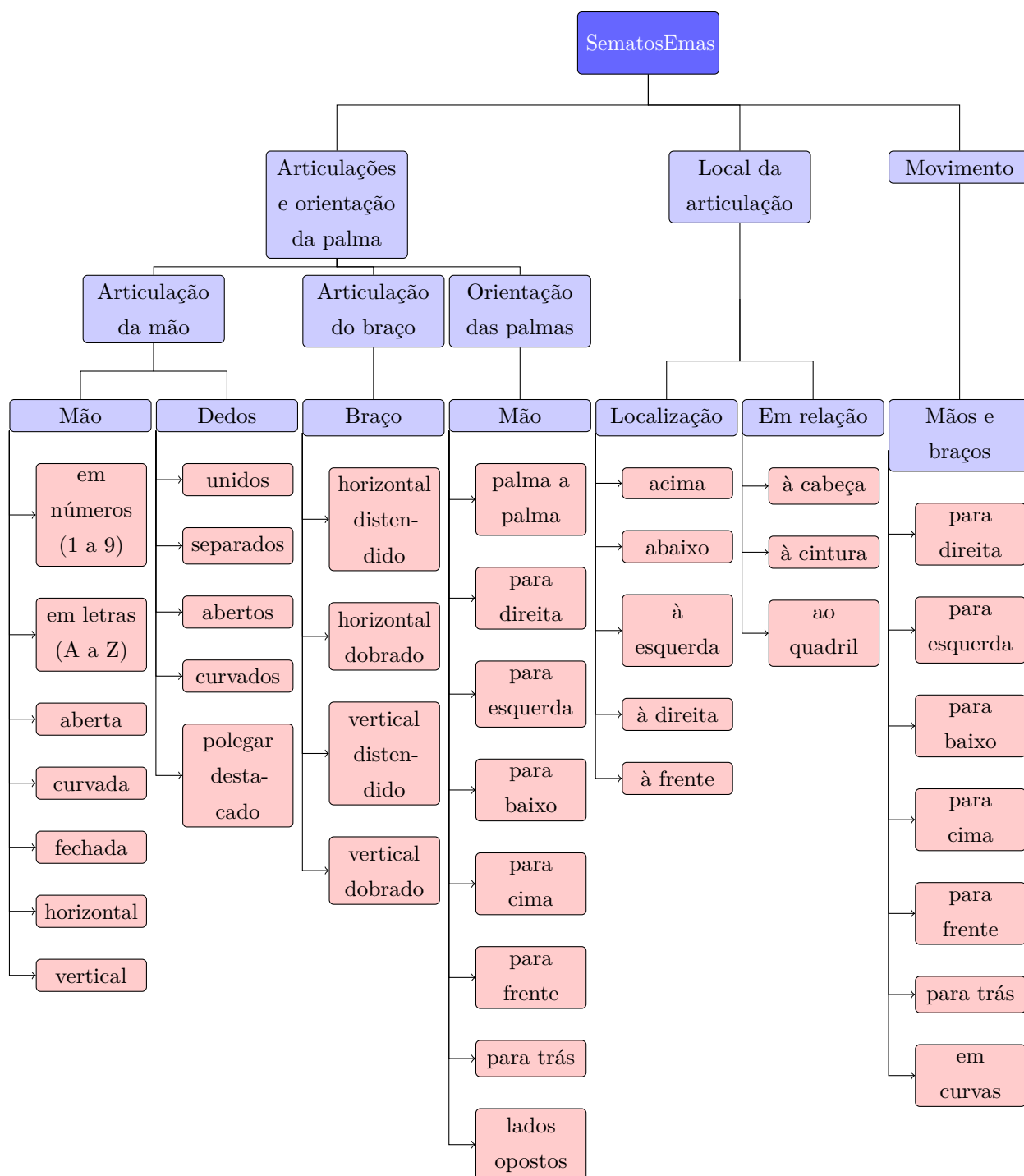


Figura 3: SematosEmas escolhidos como requisitos funcionais do projeto.

3.0.2 Requisitos Não-Funcionais

3.0.2.1 Requisitos do produto

O sistema tem como requisitos de produto os seguintes itens:

- ser facilmente operado por um não ouvinte, ou seja, outras linguagens além da LIBRAS devem ser minimamente necessárias para utilizar o sistema.
- possível de ser utilizado em um desktop com configurações padrões.
- possuir sistema de segmentação suficientemente eficiente para, uma vez identificadas as probabilidades dos SematosEmas, permitir que o sistema de reconhecimento identifique o MorfEma correto.

3.0.2.2 Requisitos organizacionais

- entregar o protótipo ao fim do ano letivo de 2014 para que possa ser avaliado por uma banca como trabalho de conclusão de curso de engenharia mecatrônica da Escola Politécnica (USP).
- o sistema deve ser uma plataforma de pesquisa para avaliar a viabilidade do reconhecimento de sinais da LIBRAS por visão computacional.

4 ANÁLISE E DESIGN

4.1 Arquitetura

O estudo dos trabalhos já realizados e de sinais da LIBRAS mostrou que o reconhecimento de determinados SematosEmas é complexo, pois alguns são bastante parecidos e podem haver imprecisões praticadas pelo próprio usuário. A sensibilidade dos sensores e as técnicas de segmentação conhecidas podem levar à não correspondência ou correspondência errada entre os SematosEmas reconhecidos e os SematosEmas realizados. Para minimizar essas distorções, propôs-se uma abordagem probabilística para os sistemas.

O protótipo foi, então, dividido em três sistemas independentes, de modo que o reconhecimento do sinal pudesse ser feito como mostrado na figura 4.

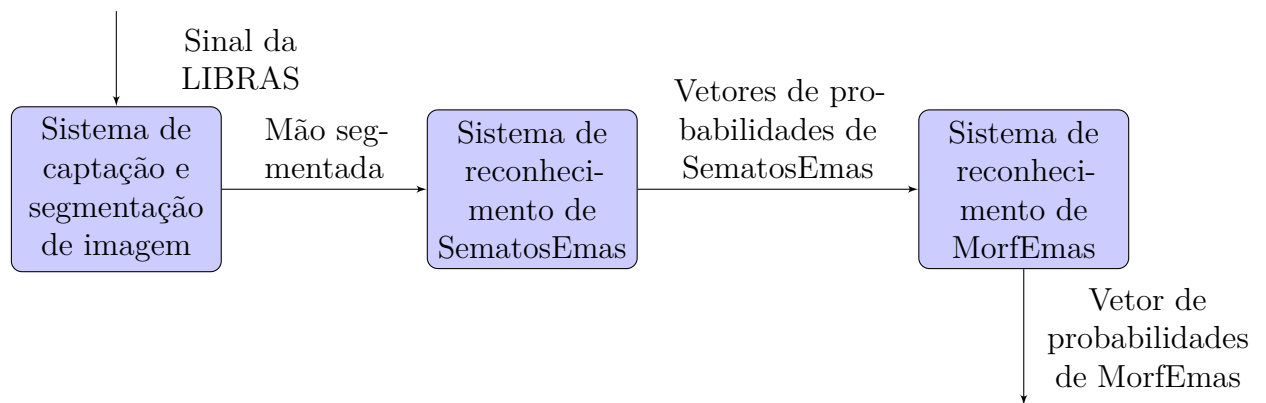


Figura 4: Arquitetura do protótipo.

Após a realização de um movimento que caracteriza um sinal da LIBRAS, o primeiro sistema capta do vídeo e segmenta a mão direita. O segundo sistema recebe a mão segmentada e, a cada frame, tem como saída três vetores: probabilidade de similaridade com cada SematosEma, probabilidade de movimentos, probabilidade de localização da mão. Cada SematosEma que compõe esses vetores foram escolhidos a partir de uma lista previamente determinada e detalhada nos

requisitos (seção 3.0.1).

Essas informações são reconhecidas e avaliadas pelo terceiro sistema, o qual classificará os vetores de probabilidades de acordo com as estruturas descritas por Capovilla, Raphael e Mauricio(3), determinando assim a probabilidade de um MorfEma ter sido realizado. A figura 5 utiliza uma amostra simplificada de MorfEmas para ilustrar detalhadamente como ocorre esse processo, tendo o sinal *Cachorro* como entrada. Esse exemplo supõe um banco de dados composto por este sinal e o de *Biscoito*, os quais se diferenciam apenas pela orientação da mão e frequência do movimento.

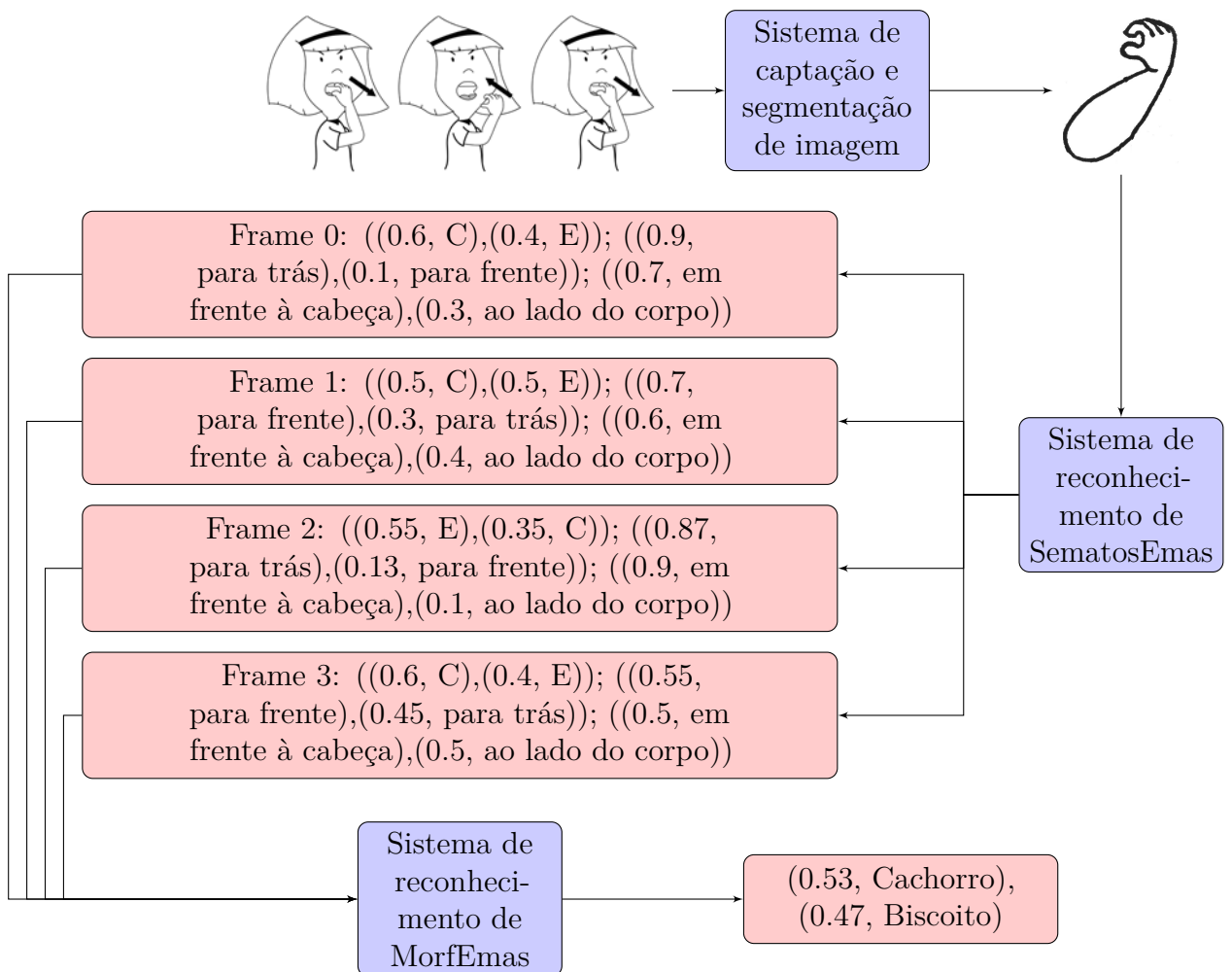


Figura 5: Detalhamento do processo de reconhecimento do MorfEma.

É importante observar que devido ao prazo de dois semestres do trabalho, nem todos os SematosEmas existentes foram adicionados ao banco de dados, isto é, nem todos os SematosEmas serão reconhecidos, e não é necessário que o sistema seja capaz de aprender novos sinais através dos sinais já realizados.

4.1.1 Diagrama de Sequência

O diagrama de sequência das principais etapas do código implementado e que representa o processo descrito na figura 5 encontra-se na figura 6.

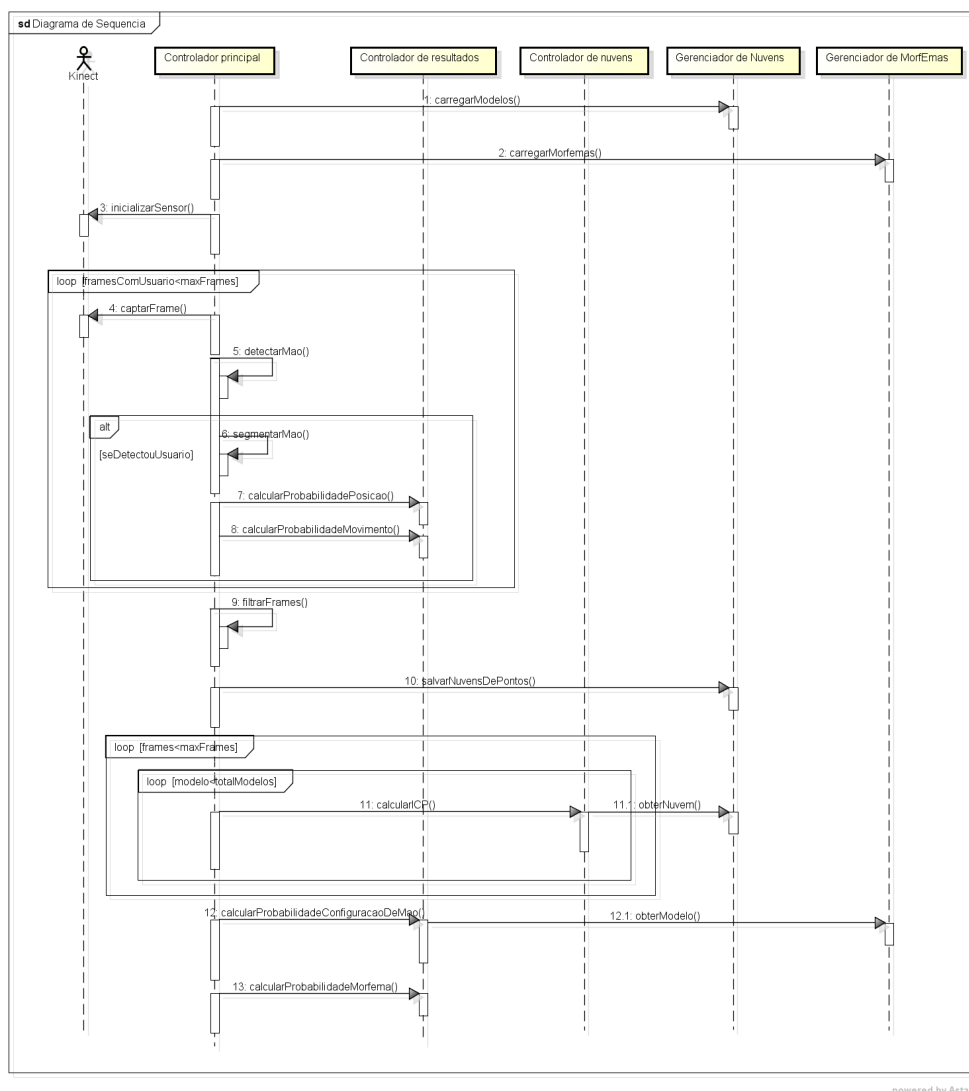


Figura 6: Diagrama de sequência do sistema.

O sensor inicia a captura do vídeo e da profundidade frame a frame e, a partir desses dados, procura detectar um usuário na cena. Assim que essa detecção ocorre, inicia-se a segmentação da região da mão direita e os cálculos das probabilidades da posição e do movimento da mão. Quando o número de frames após a detecção atinge um número máximo pré-determinado, os frames são filtrados, eliminando-se aqueles em que o movimento é de transição, pois estes causam ruído no reconhecimento.

Após o filtro, o cálculo de alinhamento é feito entre cada frame e modelo do banco, para, em seguida, determinar a probabilidade de cada modelo ter sido realizado. Por fim, as três probabilidades são combinadas no cálculo da probabilidade de cada MorfEma armazenado em um segundo banco de dados.

4.2 Sensor Utilizado

O Kinect (20), lançado pela Microsoft em 2010 para controle de vídeo games, tem sido aplicado para outros fins, entre eles o reconhecimento de sinais. Conforme pode ser visto na figura 7, o sensor é composto de:

1. Emissor infravermelho
2. Câmera de vídeo RGB
3. Sensor de profundidade, que capta os raios infravermelhos refletidos pelos objetos à frente do sensor
4. Motor de inclinação
5. Microfones

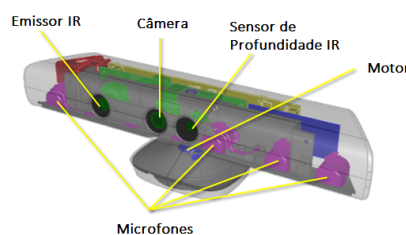


Figura 7: Componentes do sensor Kinect. Adaptado de: Microsoft Developer Network.

As especificações contidas em (21) e (22) estão resumidas na tabela 1.

Tabela 1: Especificações do sensor Kinect

Resolução do vídeo	640x480 pixels
Taxa de captura	30 Hz
Distância do objeto à câmera	de 0.8 a 3m
Abertura do sensor a 0.8m	87x63cm
Transmissão de dados	USB

A profundidade é bastante importante no reconhecimento de sinais da LIBRAS, pois um sinal tem como principais características a configuração e movimentação da mão e posição relativa à outras partes do corpo (como visto na seção 3.0.1). Isto porque a visão humana, para a qual foi criada a LIBRAS, é estéreo, ou seja, percebe naturalmente a profundidade. O Kinect, por sua vez, capta frames RGB com a câmera de vídeo e as profundidades com o sensor infravermelho.

A Microsoft possui um kit de desenvolvimento de software (SDK) para o Kinect com funções que simplificam o tratamento dos dados obtidos pelo sensor e facilita o armazenamento desses dados em uma nuvem de pontos. Além disso, ele não só adquire uma nuvem de pontos com a configuração da mão, como também armazena a posição frame a frame da posição da mão no espaço, permitindo a análise do movimento.

O sensor também é capaz de detectar o corpo inteiro, além de possuir em seu SDK funções de fácil manipulação que reconhecem posições chaves do corpo, como cabeça, ombros, cotovelos, pulsos e mãos. Um exemplo desse reconhecimento pode ser observado na figura 8.

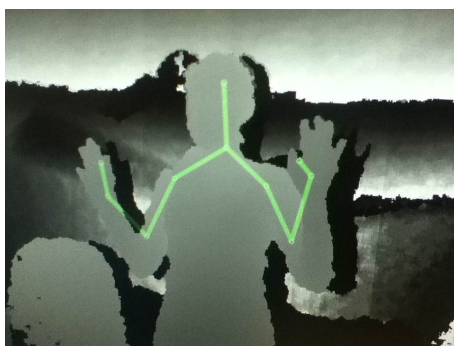


Figura 8: Imagem de profundidade e esqueleto capturada pelo Kinect.

5 CAPTAÇÃO E SEGMENTAÇÃO

5.1 Captação da imagem

A captura do vídeo RGBD do sinal da LIBRAS é realizada com o Kinect, cuja taxa de captura é de até 30 frames/s. Como esta taxa produz diferenças mínimas entre os frames e o tempo de processamento do sinal é diretamente relacionado à quantidade de frames, este projeto trabalha com taxa de 8 frames/s. Os frames são processados com a utilização do próprio SDK da Microsoft para o Kinect. Durante a captura é feita a segmentação das mãos e o reconhecimento dos SematosEmas de movimento e de local de articulação.

5.2 Segmentação da mão

O processo de segmentação consiste em definir uma região de interesse do frame completo e selecionar apenas os dados dos pontos dessa região. O SDK do Kinect já implementa funções de localização de juntas de um corpo humano, que representam, por exemplo, as coordenadas no espaço do ombro ou do cotovelo. Assim, obtém-se as coordenadas da junta "mão direita" no espaço, segmenta-se a região da mão definindo um paralelepípedo ao redor da junta e seleciona-se os pontos nessa região.

6 RECONHECIMENTO SEMATOSÊMICO

6.1 Reconhecimento dos SematosEmas de movimento

O movimento é avaliado através da diferença da posição da mão direita entre um frame e o seu anterior. Calcula-se o cosseno entre o vetor de deslocamento e os seis semi-eixos que representam os movimentos para cima, baixo, direita, esquerda, frente e trás, como na figura 9, sendo o menor ângulo corresponde ao movimento mais provável.

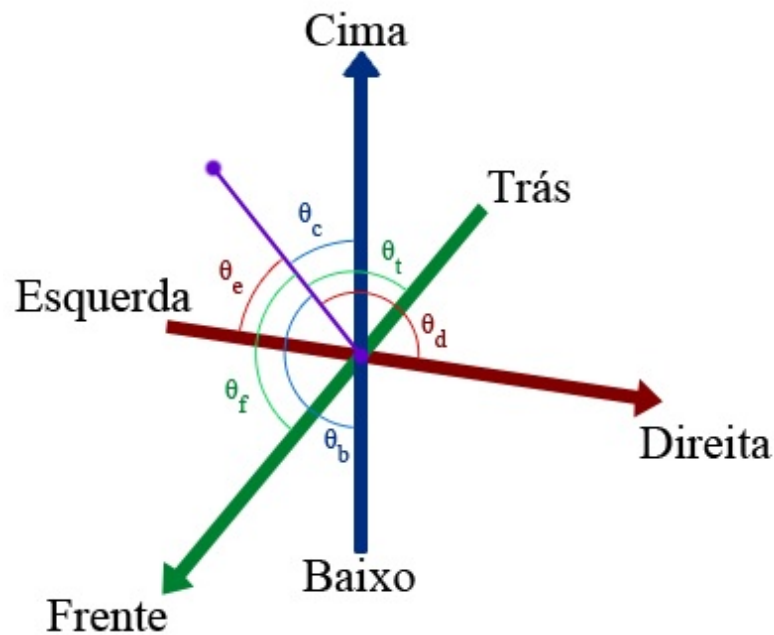


Figura 9: Representação dos ângulos em relação aos seis semi-eixos.

Para garantir que todos os frames consecutivos se movimentando na mesma direção sejam considerados, a probabilidade de cada movimento ter sido realizado é calculada após a mudança de direção da mão. Ou seja, se a mão se movimentar por 5 frames consecutivos para a direita e, em seguida, mudar a direção para baixo, inicia-se o cálculo da probabilidade do movimento entre a posição da mão

no quinto e primeiro frames.

A probabilidade da mão não ter se movido é sempre calculada pela função sigmóide:

$$1/(1 + e^{(d-t)/k})$$

Onde d é a amplitude detectada do movimento, t é a distância mínima para que seja considerado movimento e k é a distância a mais em relação a t em que ainda há 25% de probabilidade de não ter ocorrido o movimento. A função sigmóide foi escolhida pois é uma distribuição estatística em que as distâncias citadas são influentes.

Já a probabilidade de movimentação em cada uma das direções é inversamente proporcional ao ângulo formado entre o vetor de deslocamento e o semi-eixo correspondente àquela direção.

6.2 Reconhecimento dos SematosEma de local de articulação

O SematosEma de local de articulação é reconhecido através das coordenadas das juntas. O espaço de reconhecimento foi dividido nas seguintes regiões:

1. acima da cabeça
2. à direita da cabeça
3. em frente ao rosto
4. à esquerda da cabeça
5. à direita da cintura
6. em frente à cintura
7. à esquerda da cintura
8. abaixo da cintura

As coordenadas da mão direita são avaliadas em referência a outras juntas, conforme a figura 10.

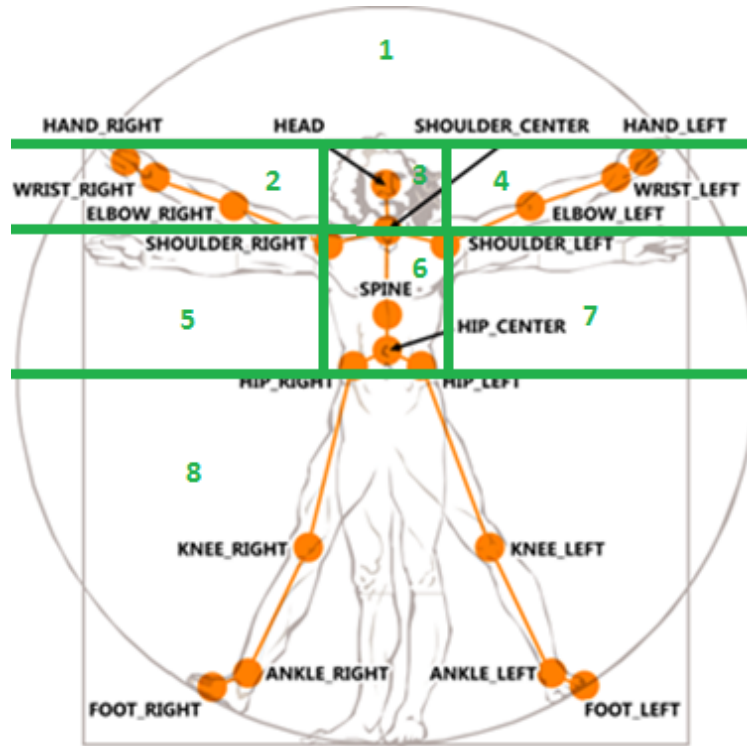


Figura 10: Regiões dos SematosEmas de local de articulação. Adaptado de Microsoft Developer Network.

Testes mostraram que o sensor e a função de identificação de juntas são precisos ao determinar a posição relativa da mão. Por isso, a probabilidade de que a mão esteja na região avaliada é considerada alta; para as regiões fronteiriças, baixa; e para as regiões mais distantes, menor ainda, conforme a tabela 2.

Tabela 2: Probabilidades do posicionamento da mão dada a região identificada pelo sensor.

$r \backslash P(r)$	1	2	3	4	5	6	7	8
1	0.8	0.05	0.05	0.05	0.0125	0.0125	0.0125	0.0125
2	0.05	0.8	0.05	0.0125	0.05	0.0125	0.0125	0.0125
3	0.04	0.04	0.8	0.04	0.0133	0.04	0.0133	0.0133
4	0.05	0.0125	0.05	0.8	0.0125	0.0125	0.05	0.0125
5	0.0125	0.05	0.0125	0.0125	0.8	0.05	0.0125	0.05
6	0.0133	0.0133	0.04	0.0133	0.04	0.8	0.04	0.04
7	0.0125	0.0125	0.0125	0.05	0.0125	0.05	0.8	0.05
8	0.0125	0.0125	0.0125	0.0125	0.05	0.05	0.05	0.8

6.3 Reconhecimento do SematosEma de configuração de mão

O reconhecimento da articulação da mão e dos dedos é o processo mais complexo e computacionalmente custoso no reconhecimento dos sinais. Após estudar os métodos utilizados na literatura, optou-se pelo reconhecimento por modelo, ou seja, o frame capturado é comparado com modelos dos SematosEmas pré-definidos e salvos em um banco de dados. Essa comparação é feita entre a nuvem de pontos do modelo e do sinal capturado.

6.3.1 Point Cloud Library

Uma nuvem de pontos é uma estrutura de dados que armazena as coordenadas dos pontos que a compõe. No contexto desse projeto, as duas nuvens de pontos de interesse (frame capturado e modelo) foram manipuladas com o auxílio da Point Cloud Library (PCL) (23), uma biblioteca de código aberto escrita em C++ que processa nuvens de pontos 3D. PCL é composta por módulos que podem ser compilados separadamente e todos possuem testes unitários e de regressão, os quais garantem seu comportamento correto. Com a biblioteca é possível filtrar imagens, estimar features, segmentar imagens, reconstruir superfícies, entre outras funções. A PCL foi escolhida para manusear as nuvens de pontos obtidas com o sensor por ser uma ferramenta de fácil manipulação e de implementação eficiente.

6.3.2 Banco de dados de modelos

Os modelos de SematosEmas de articulação de mão foram capturados com o próprio Kinect e o programa KinectFusion (24)(25). O KinectFusion permite duas abordagens para escanear um objeto: rotacionar o objeto na frente do sensor ou movimentar o sensor em torno do objeto, sendo que este último foi o procedimento usado no trabalho. As configurações de mãos escolhidas para constituir o banco de dados do projeto podem ser vistas no apêndice A.

É necessário notar, porém, que as configurações de mãos, além das descritas como letras e números, precisam ser complementadas pelas configurações dos dedos. Posto isso, utilizou-se configurações descritas por Capovilla, Raphael e Mauricio(3) como “formas adicionais de mão usadas em LIBRAS”, as quais

compõem os SematosEmas, como, por exemplo, *mão* aberta e *dedos* unidos; *mão* aberta e *dedos* separados; e *mão* aberta e *dedos* unidos e curvados.

6.3.3 Iterative Closest Point

Iterative Closest Point (ICP) é um algoritmo iterativo que visa minimizar a distancia entre duas nuvens de pontos através de movimentos de corpo rígido, translações e rotações. Este projeto utilizou a implementação do ICP feita pelo PCL.

A nuvem de pontos obtida pela segmentação da mão é comparada com os modelos do banco de dados através do ICP. Inicialmente, a nuvem capturada é rotacionada de forma a alinhar seu eixo de coordenadas com o do modelo. Em seguida, calcula-se as normais dessa nuvem, utilizando métodos da própria PCL e informações do sensor. Esse processo não é necessário para o tratamento do modelo, pois ele foi previamente salvo com essas informações. Isto é feito pois, durante o projeto, testes mostraram a importância da orientação das superfícies das nuvens comparadas, ou seja, a existência dos vetores normais dos pontos. A ausência de normais muitas vezes levava ao posicionamento errado das superfícies, como palma da mão posicionada frente-a-frente à palma do modelo.

Por fim, a nuvem passa pelo processo de ICP, no qual é transladada e rotacionada ate que fique o mais próxima possível do modelo. Quando a transformação converge, obtém-se a pontuação do alinhamento, que consiste na soma do quadrado das distâncias entre as duas nuvens. Entretanto, o ICP admite que as nuvens a serem comparadas são diferenciadas apenas por uma transformação de corpo rígido, o que não é verdadeiro nesse projeto. Isto porque, além de compararmos o mesmo frame com diferentes modelos, a nuvem do modelo é uma representação completa em 3D do SemastosEma, enquanto a do frame capturado é a visão de um SemastosEma a partir de uma posição fixa. Além disso, o volume dos modelos afeta o cálculo das distâncias, por exemplo, como o modelo *A* possui maior volume do que o modelo *Mão aberta dedos unidos*, a pontuação de alinhamento com o frame *A* se mostra maior, portanto pior, quando comparada ao modelo *A*. A figura 11 ilustra o problema.

Sendo assim, para utilizar a pontuação obtida pelo processo, duas considerações foram feitas: calcular a pontuação apenas entre pontos a 3mm da nuvem do frame e

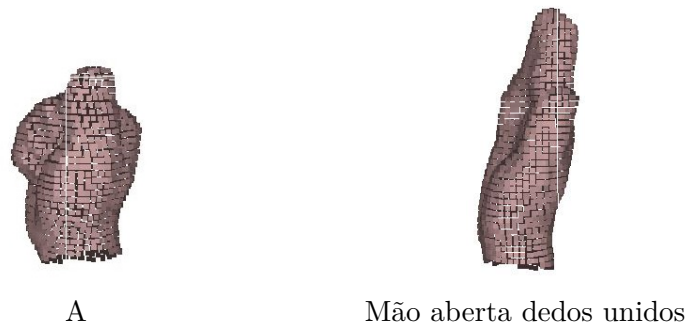


Figura 11: Comparação da largura de dois modelos.

utilizar uma pontuação relativa para o reconhecimento do SemastosEma. Eliminar todos os pontos a mais de 3mm de distância do frame reduziu consideravelmente as pontuações obtidas pelo ICP, porém, não se mostrou o suficiente para resolver o problema. Adicionalmente à essa solução, decidiu-se determinar as pontuações nas quais é sabida que o frame e o modelo estão bem alinhados, como visto na figura 12.



Figura 12: Frames alinhados com seus respectivos modelos.

Foram feitos 15 testes para cada modelo, obtendo-se, assim, a média de pontuação para o bom alinhamento entre o frame e seu respectivo modelo, já considerando os 3mm limites para o cálculo das distâncias. A pontuação utilizada é, então, a subtração entre o valor obtido pelo ICP e a média de alinhamento para o modelo a ser comparado.

6.3.4 Cálculo das probabilidades *a posteriori* dos SemastosEmas de articulação de mão

Para reconhecer o MorfEma é necessário saber qual a probabilidade de um SematosEma de articulação de mão ter sido realmente realizado, tendo os resul-

tados do ICP para os dados capturados com o sensor. Tal problema é bastante complexo, podendo ser considerado como um refinamento do reconhecimento de membros do corpo realizado pelo Kinect. O reconhecimento de poses com o Kinect é um problema que foi solucionado pela Microsoft por meio de *machine learning*, utilizando muitos recursos computacionais e um extenso e diversificado banco de dados (26).

Como este trabalho possui um problema semelhante, mas sem dispor de tempo e de recursos suficientes como o projeto do Kinect, o cálculo das probabilidades dos SematosEmas de articulação de mão se baseou em um procedimento bem mais simples de aprendizado de máquina. Realizaram-se três testes para cada articulação de mão disponível no banco de modelos e com estes dados, observou-se a correspondência entre SematosEma realizado e as respostas dadas pelo ICP.

A tabela 3 mostra alguns resultados do ICP para o sinal de C , quanto menor o valor, melhor foi alinhamento.

Como pode ser observado, a correlação entre o resultado do ICP e a configuração de mão real não é tão boa. Portanto buscou-se inferir a resposta correta com base nas respostas amostradas nos testes. Pressupôs-se, também, que o usuário realizou, necessariamente, algum dos SematosEmas presentes no banco. Para limitar o escopo da resolução do problema, para cada frame, os resultados do ICP foram ordenados do melhor modelo alinhado para o pior e foram utilizados apenas os cinco primeiros lugares, de forma qualitativa (a colocação do modelo é utilizada e não o valor bruto do alinhamento).

Além disso, para incorporar os erros não-modelados do sensor e do algoritmo de alinhamento, foi considerado um termo de distribuição homogênea de probabilidades para todos os SematosEmas.

Quando o usuário realiza um sinal com a mão na configuração $C1$, o frame será comparado com todos os modelos e os cinco mais bem alinhados serão $M1$, $M2$, $M3$, $M4$ e $M5$, onde o Mi é o i -ésimo mais bem alinhado.

Para a configuração de mão $S1$, $P1$ é a probabilidade do usuário ter realizado $S1$, dado que $M1$ foi o melhor alinhamento. $P1$ é calculado pela divisão:

$$\frac{\text{nº de ocorrências de } M1 \text{ em primeiro lugar para } S1}{\text{nº de ocorrências de } M1 \text{ em primeiro lugar para todos os SematosEmas}}$$

$P2$ é a probabilidade do usuário ter realizado $S1$, dado que $M2$ foi o segundo melhor alinhamento. $P2$ é calculado pela divisão:

$$\frac{\text{n}^\circ \text{ de ocorrências de } M2 \text{ em segundo lugar para } S1}{\text{n}^\circ \text{ de ocorrências de } M2 \text{ em segundo lugar para todos os SematosEmas}}$$

$P3$ é a probabilidade do usuário ter realizado $S1$, dado que $M1, M2, M3, M4$ e $M5$, foram os cinco melhores alinhamentos.

$$\frac{1}{5} \cdot \sum_{i=1}^5 \frac{\text{n}^\circ \text{ de ocorrências de } Mi \text{ entre os 5 primeiros lugares para } S1}{\text{n}^\circ \text{ de ocorrências de } Mi \text{ entre os 5 primeiros lugares para todos os SematosEmas}}$$

$P4$ é a distribuição homogênea de probabilidades, a qual corresponde a $1/(\text{total de SematosEmas})$, no caso dos testes, $1/24$.

Calcula-se então para cada configuração de mão a seguinte probabilidade:

$$P(S) = w1 \cdot P1 + w2 \cdot P2 + w3 \cdot P3 + w4 \cdot P4$$

Os pesos $w1, w2, w3$ e $w4$ foram escolhidos respeitando o critério de que a primeira resposta tem maior influência no resultado, seguida da segunda e assim sucessivamente. Algumas foram escolhidas arbitrariamente e testadas, sendo que a que apresentou melhores resultados foi a de valores 0,6, 0,25, 0,1 e 0,05 para $w1, w2, w3$ e $w4$, respectivamente.

Tabela 3: Resultados do ICP para dois testes com o SematosEma *C*.

Teste 1		Teste 2	
0.000008	Mão aberta - dedos distendidos e paralelos	0.000002	C
0.000015	4	0.000032	D
0.00003	C	0.000056	1
0.000038	Mão aberta - dedos flexionados	0.000061	Y
0.000043	D	0.000064	E
0.00005	E	0.000076	S
0.000084	G	0.000085	Mão aberta - dedos distendidos e paralelos
0.000117	T	0.000093	A
0.000118	Y	0.000094	K
0.000134	K	0.000117	Mão aberta - dedos curvados e separados
0.000161	Mão aberta - dedos curvados e separados	0.000118	G
0.000218	A	0.000135	4
0.000222	Mão aberta - dedos unidos e curvados	0.000183	Mão aberta - dedos unidos e curvados
0.000226	N	0.000205	I
0.000247	Mão aberta - dedos separados	0.00022	L
0.000276	L	0.000229	Mão aberta - dedos flexionados
0.000299	Z	0.000235	N
0.000308	Mão aberta - dedos unidos	0.000251	Mão aberta - dedos separados
0.000334	Mão aberta - ponta dos dedos unidos	0.000291	Mão aberta - dedos unidos
0.000354	1	0.000308	Z
0.000553	S	0.000316	T
0.00058	X	0.000554	X
0.00059	I	0.000554	Mão aberta - ponta dos dedos unidos
0.000804	F	0.000808	F

7 RECONHECIMENTO DO MORFEMA

O reconhecimento do MorfEma é realizado após a captação do sinal. Primeiramente, analisam-se as probabilidades dos movimentos e segmenta-se o sinal através do movimento mais provável entre cada frame, isto é, frames com probabilidades de movimentos iguais são considerados como pertencentes a um único segmento do MorfEma.

Paralelamente a isto, o sinal é filtrado, retirando-se frames em que há transição da direção do movimento, pois, por exemplo, quando o usuário movimenta a mão para a esquerda e depois para a direita, é frequente que haja um pequeno trecho de posição constante entre os segmentos de movimentação para a esquerda e para a direita. Esse ruído é causado pela dificuldade em sincronizar a troca de movimento com a taxa de captura do sistema. Por consequência, torna-se desnecessário utilizar o processo do ICP nos frames filtrados, diminuindo, portanto, o tempo total do reconhecimento.

Segmentado o sinal pelo movimento, a probabilidade de cada MorfEma ter sido realizado é a multiplicação da probabilidade para cada um dos seus segmentos. Esta, por sua vez, é feita com outras quatro probabilidades: movimento do segmento, posição da mão no início do movimento, posição da mão no final do movimento e média das probabilidades do SematosEma de configuração de mão ao longo dos frames do segmento. A probabilidade de MorfEmas com número de segmentos diferente do capturado é zero, pois a probabilidade dos segmentos excedentes ou faltantes terem sido feitos é zero.

7.1 Banco de dados de MorfEmas

Os sinais foram escolhidos do livro Deit-LIBRAS(3) dentre os sinais cujos todos os SematosEmas eram compreendidos pelo sistema e de forma que os sinais escolhidos permitissem avaliar cada componente do reconhecimento e também

o sistema integrado. Escolheu-se, por exemplo, sinais de mesmo movimento e posições com configurações de mão diferentes, sinais de movimentos diferentes e mesma configuração de mão e sinais com número de movimentos diferentes.

O banco de dado de MorfEmas utilizado nos testes foi composto por 12 sinais:

7.1.1 Coordenar

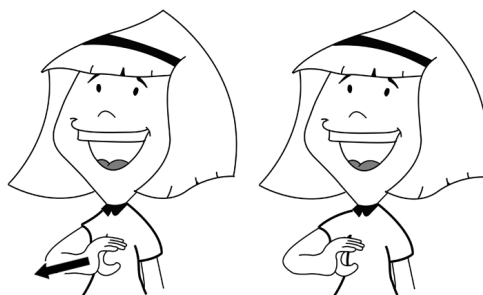


Figura 13: Coordenar: Mão em C, palma para a esquerda, tocando o peito. Movê-la para a frente.

7.1.2 Depressa

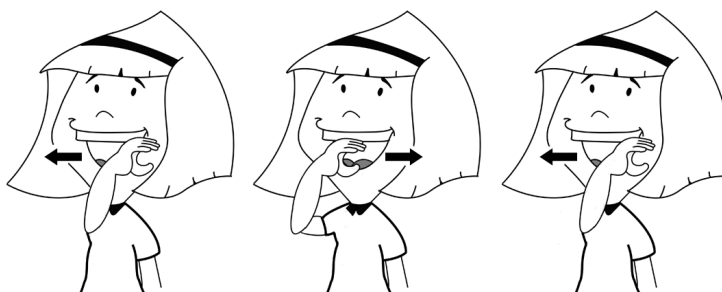


Figura 14: Depressa: Mão em C, palma para a esquerda. Movê-la rapidamente para a esquerda e para a direita, diante da boca.

7.1.3 Deus

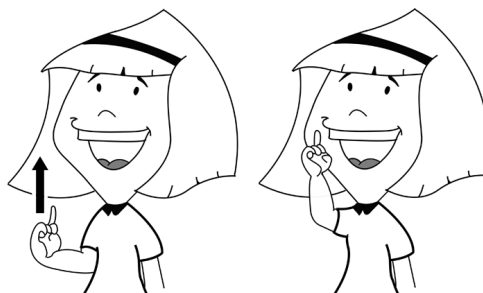


Figura 15: Deus: Mão em D, palma para a esquerda. Movê-la para cima, acima da cabeça.

7.1.4 Elevar

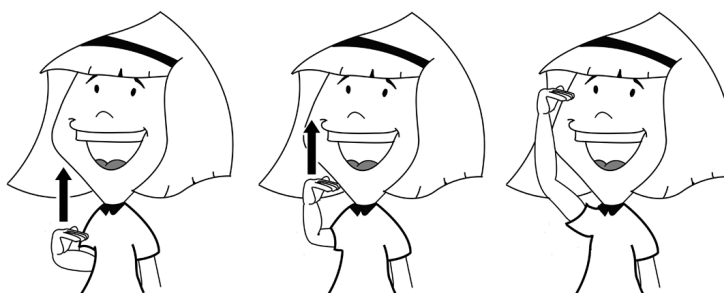


Figura 16: Elevar: Mão aberta com os dedos flexionados. Movê-la do peito até a altura da cabeça.

7.1.5 Em cima

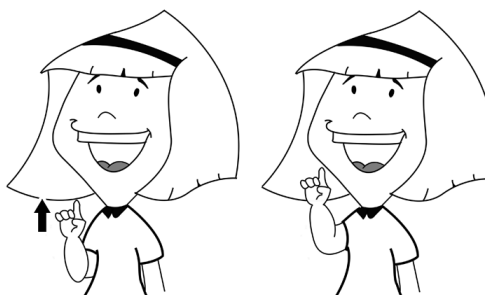


Figura 17: Em cima: Mão em 1, palma para frente, na altura do ombro. Mover a mão para cima.

7.1.6 Gêmeo

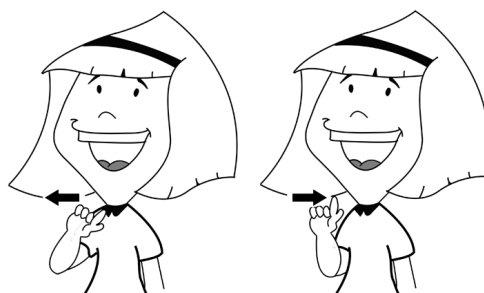


Figura 18: Gêmeo: Mão em G, palma para a frente. Mover a mão para a esquerda e para a direita.

7.1.7 Genro

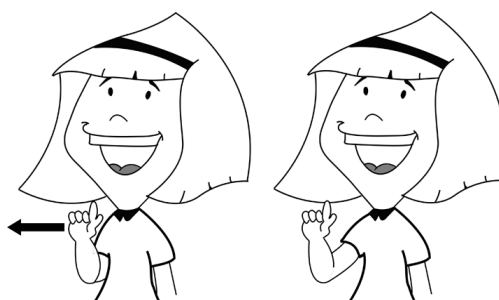


Figura 19: Genro: Mão em G, palma para frente. Movê-la para a direita.

7.1.8 Lado direito

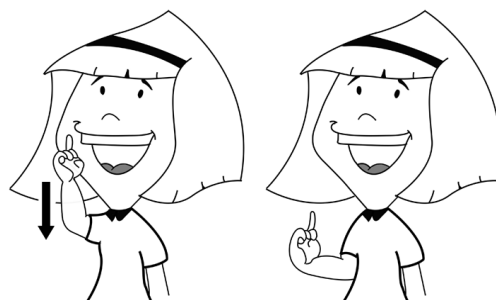


Figura 20: Lado direito: Mão em D, palma para frente, ao lado direito do corpo. Mover a mão para baixo.

7.1.9 Nora

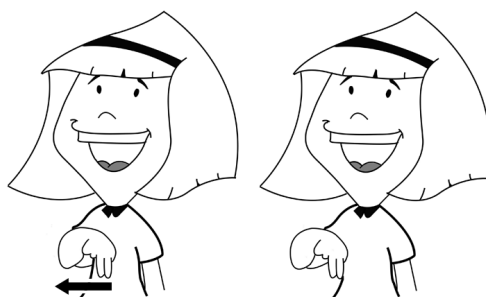


Figura 21: Nora: Mão em N. Movê-la para a direita.

7.1.10 Norte

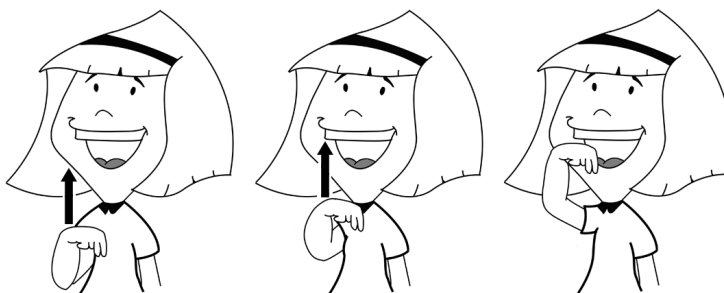


Figura 22: Norte: Mão em N. Movê-la para cima.

7.1.11 Para cima

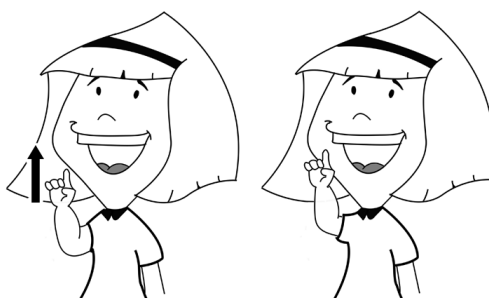


Figura 23: Para cima: Mão em 1, palma para a frente, na altura da cabeça. Movê-la ligeiramente para cima.

7.1.12 Teto

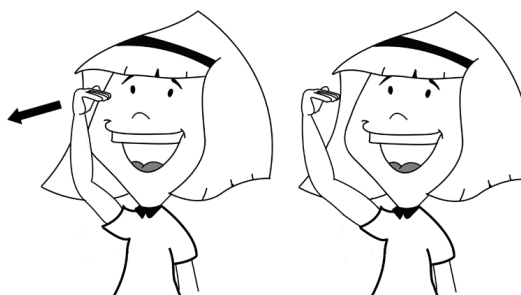


Figura 24: Teto: Mão aberta e dedos flexionados, palma para a esquerda, em cima da cabeça. Mover a mão para frente.

7.2 Paralelização

O cálculo do ICP é o processo de maior custo computacional, correspondendo a até 98,6% do tempo total do reconhecimento do sinal. O alinhamento de um frame com um modelo leva em média 2,04s, totalizando 48,96s para os 24 modelos do banco de dados. Todavia, cada alinhamento depende apenas do frame atual e de um modelo, o que permite a paralelização do processo, de forma que cada alinhamento seja feito em uma *thread* independente da *thread* que contém o programa principal de análise dos sinais.

O processamento das probabilidades dependem diretamente de todos os resultados do ICP para um frame, por isso, é preciso assegurar a sincronização das *threads*, isto é, todas as *threads* devem ter terminado antes que as probabilidades sejam calculadas.

Essa paralelização foi feita com o *Boost*, uma biblioteca de suporte para linguagem C++ que oferece auxílio para programação concorrente, estruturas de dados, entre outras funcionalidades. A implementação desse processo é bastante simples, pois todas as *threads* compartilham a memória e suas posições com mesmo custo.

Com a paralelização, conseguiu-se que tempo médio de alinhamento de um frame com todos os modelos fosse reduzido para 7,54s, ou seja, 6,5 vezes mais rápido.

8 IMPLEMENTAÇÃO

O programa de reconhecimento de sinais foi desenvolvido em linguagem C++ no ambiente de desenvolvimento VisualStudio e os testes foram realizados em um computador com as especificações descritas na tabela 4.

Tabela 4: Especificações do computador utilizado nos testes

Processador	Intel TM Core®i7-3970X CPU @3.50GHz
Sistema Operacional	Windows 7 Professional 64-bits
Memória	24.0 GB
Espaço em disco	1TB

Através de testes foi confirmado que não são necessários 30 frames/*s* para o reconhecimento. Devido ao custo computacional aumentar proporcionalmente à quantidade de frames, a taxa de utilizada foi de 8 frames/*s*. Foi estipulado o tempo de realização do sinal em 25 frames pois os sinais de teste podem ser feitos nesse período.

8.1 Interface gráfica

A interface gráfica do programa desenvolvido é bastante simples e intuitiva: a maior parte da janela transmite o vídeo que está sendo captado pela câmera do sensor.

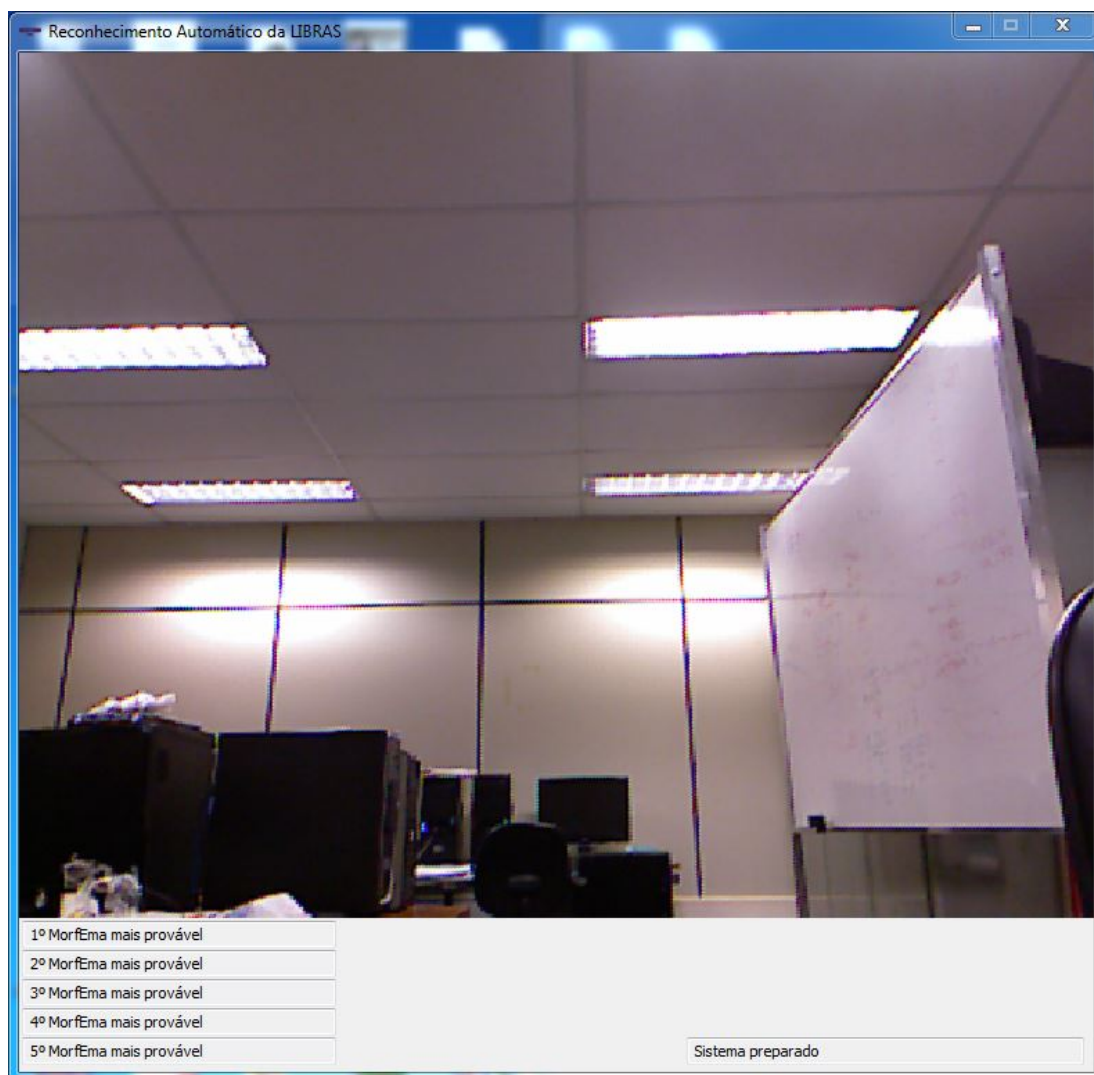


Figura 25: Interface Homem Máquina (IHM) ao iniciar o programa.

Quando um usuário é reconhecido pelo sistema, a sua mão direita é envolta por um quadrado amarelo que acompanha o movimento da mão. Após três segundos, o contorno se torna verde, simbolizando que o sinal começou a ser captado. Mais três segundos depois, o contorno se torna vermelho, simbolizando o término da captura. Quando o usuário realiza um sinal de menor duração, deve permanecer com a mão na posição final do sinal, até que termine a captura. A figura 26 ilustra essa sequência.

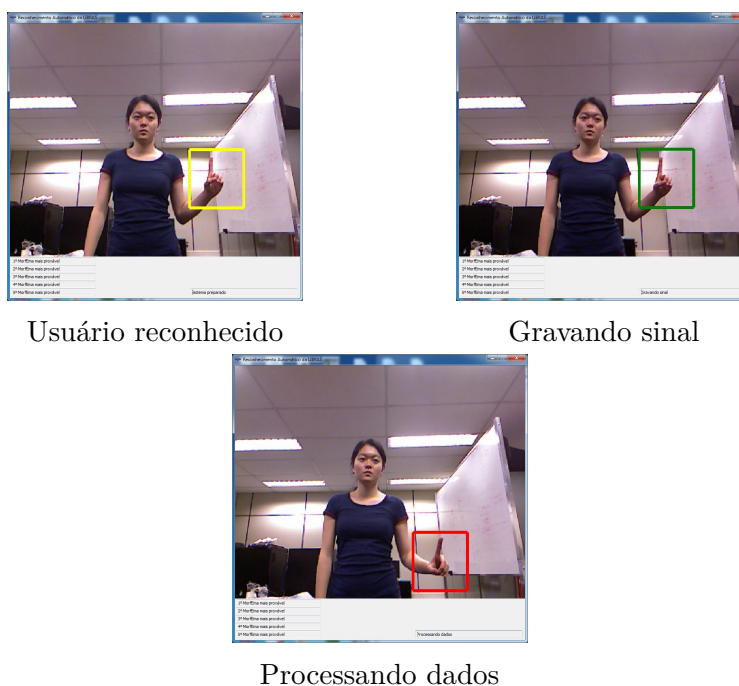


Figura 26: Sequencia da IHM ao analisar um sinal.

Quando o sistema terminar de processar o reconhecimento do sinal, mostra os cinco sinais mais prováveis de terem sido realizados nas caixas inferiores esquerdas da janela. Quando o processamento termina o usuário pode realizar outro sinal para ser reconhecido.

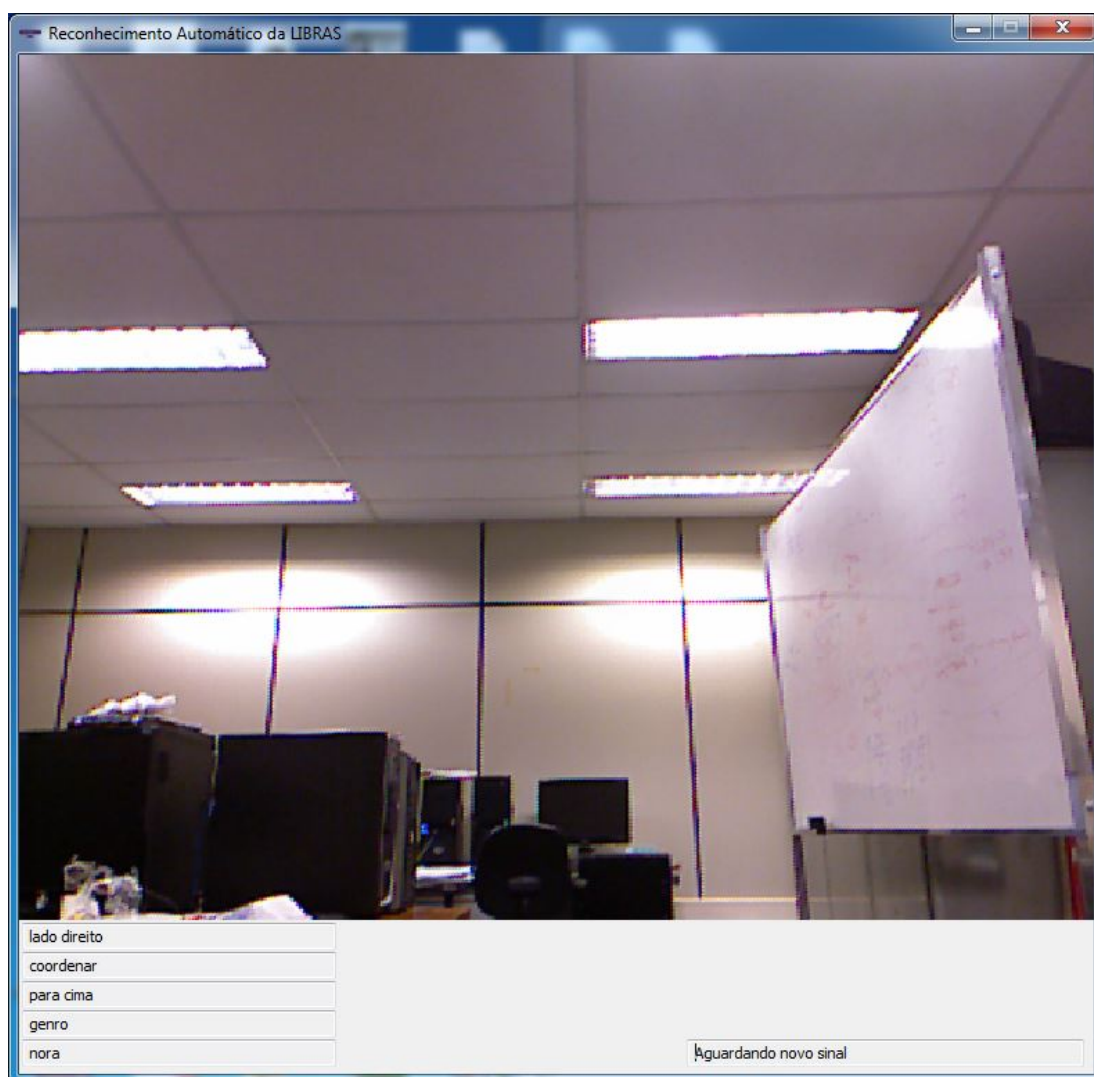


Figura 27: Interface Homem Máquina (IHM) após reconhecimento do sinal.

9 RESULTADOS

O sistema foi testado 48 vezes (quatro vezes para cada um dos MorfEmas no banco de dados), sendo que o usuário em todos os testes é a mesma pessoa cuja mão compõe o banco de dados dos modelos de SematosEmas. O tempo médio de reconhecimento de cada sinal foi de 127s por teste. Dentro dessa amostra, o sistema reconheceu corretamente 31 sinais, correspondendo a uma taxa de acerto de 65%. Por outro lado, entre os 17 sinais não reconhecidos, o sinal correto foi classificado 13 vezes como o segundo mais provável, 3 vezes como terceiro e somente uma vez como quarto. A distribuição dos resultados pode ser vista na figura 28.

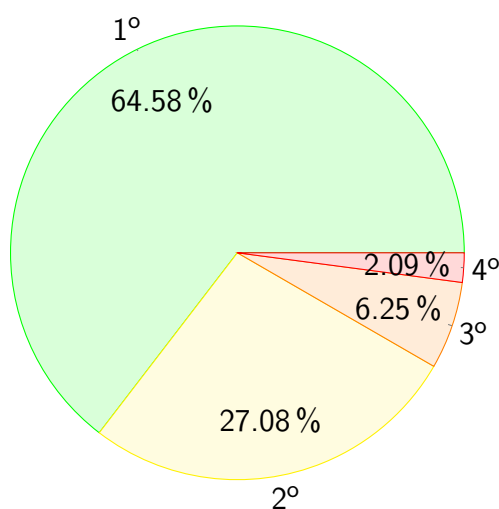


Figura 28: Resultados dos testes: posições da resposta correta na resposta do sistema

Verificando a taxa de acerto em que o sinal correto é classificado em primeiro e segundo lugar (91,67%) têm-se um indício de que, apesar de não trabalhar com o processamento de linguagem natural, o sistema pode proporcionar reconhecimentos próximos o suficiente para possibilitar o entendimento de uma sequência de sinais. Entretanto, nos casos em que o sinal real aparece como segundo mais provável, caberia ao usuário atestar o significado através do contexto.

Analisando-se, também, a porcentagem de acerto individual de cada MorfEma nos quatro testes realizados para cada, obteve-se os resultados apresentados na figura 29.

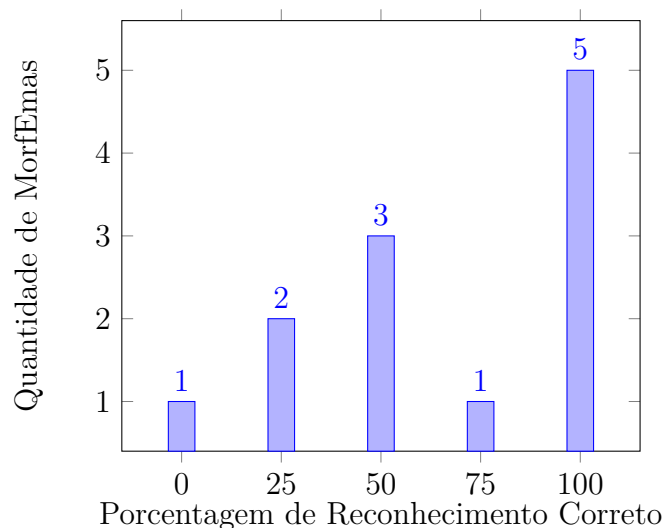


Figura 29: Gráfico da porcentagem de acerto de MorfEmas em primeiro lugar pela quantidade.

9.1 Limitações do sistema

O protótipo entregue não reconhece todos os SematosEmas listados nos requisitos do projeto. Os SematosEmas de Articulação de mão foram limitados pois o custo computacional aumenta proporcionalmente com esse banco de dados. Os de articulação de braço não foram analisados pois já é avaliado o local da articulação, que já indica as articulações de braço escolhidas anteriormente. A orientação das palmas da mão e o movimento em curvas não foram analisados porque durante o projeto foram encontrados problemas que consumiram mais tempo do que o planejado, forçando o corte desse tópico.

10 CONCLUSÃO

O protótipo criado para avaliar a viabilidade do reconhecimento da LIBRAS por visão computacional serviu a seu propósito, pois os testes realizados com ele foram satisfatórios tendo em vista o tempo e os recursos disponíveis para desenvolvê-lo. Entretanto, o sistema possui limitações, tais como reconhecimento de um número reduzido de SematosEmas em relação a todos os existentes; tempo de resposta longo e variável, não permitindo o processamento on-line e tempo de captura fixo em, aproximadamente, 3 segundos para cada sinal.

Tais limitações poderiam ser superadas com a implementação de todos os SematosEmas, o que acarreta em maior custo computacional para o alinhamento. Esse problema, por sua vez, pode ser solucionado pelo emprego de mais recursos computacionais, já que o alinhamento de frames e modelos é facilmente paralelizável. Com maior disponibilidade de tempo e pessoas, seria possível aprimorar o processo de *machine learning*, tornando assim, as respostas mais próximas do correto.

O projeto mostrou que a abordagem de reconhecimento do MorfEmas através da avaliação de seus SematosEmas é bastante promissora, pois, mesmo que um SematosEma seja avaliado erroneamente, tanto por presença de ruídos ou por limitação do sistema, a utilização de todos os SematosEmas faz com que a resposta final esteja próxima do esperado.

A existência da classificação dos sinais da LIBRAS por SematosEmas abre caminho para que o reconhecimento dos sinais seja feito de forma automática por meio de visão computacional.

11 SUGESTÕES PARA TRABALHOS FUTUROS

Como sugestões de trabalhos futuros há a implementação da identificação de mais SematosEmas, como de articulação de mão não utilizados neste trabalho, de orientação da palma da mão, de configuração do braço e de expressão facial. Tais implementações aumentarão o custo computacional e o tempo do reconhecimento ainda mais. Para melhorar isso seria aconselhável otimizar a implementação do algoritmo ICP, tanto em tempo quanto em precisão. Também seria ideal que o banco de dados utilizasse mais pessoas e mais testes na sua composição, para que o aprendizado de máquina seja mais efetivo.

REFERÊNCIAS

- 1 CENTRE, W. M. *WHO — Deafness and hearing loss*. World Health Organization, 2014. Acessado em: 2014-04-06. Disponível em: <http://www.who.int/mediacentre/factsheets/fs300/en/>.
- 2 GUIMARAES, C. et al. Technological Artifacts for Social Inclusion: Structure of the Brazilian Sign Language (LIBRAS), Gestures for Citizenship. In: *Proceedings of the IADIS International Conference on WWW/Interne*. [S.l.: s.n.], 2010. p. 267.
- 3 CAPOVILLA, F. C.; RAPHAEL, W. D.; MAURICIO, A. C. L. *NOVO DEIT-LIBRAS: Dicionário Enciclopédico Ilustrado Trilíngue da Língua de Sinais Brasileira (libras) Baseado em Linguística e Neurociências Cognitivas*. 3^ª edição. ed. [S.l.]: EDUSP, 2013. ISBN 978-85-314-1433-6.
- 4 DIAS, D. B.; PERES, S. M. Algoritmos bio-inspirados aplicados ao reconhecimento de padrões da LIBRAS: enfoque no parâmetro movimento. In: *16^o Simpósio Internacional de Iniciação Científica da Universidade de São Paulo*. [s.n.], 2008. Disponível em: <http://uspdigital.usp.br/siicusp/cdOnlineTrabalhoVisualizarResumo?numeroInscricaoTrabalho=2592&numeroEdicao=16>.
- 5 DIGIAMPIETRI, L. A. et al. Um Sistema de Informação Extensível Para o Reconhecimento Automático de LIBRAS. In: *VIII Simpósio Brasileiro de Sistemas de Informação (SBSI 2012)*. [s.n.], 2012. Disponível em: <http://www.lbd.dcc.ufmg.br/colecoes/sbsi/2012/0045.pdf>.
- 6 HALL, M. et al. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, ACM, New York, NY, USA, v. 11, n. 1, p. 10–18, nov. 2009. ISSN 1931-0145. Disponível em: <http://doi.acm.org/10.1145/1656274.1656278>.
- 7 LI, Y. et al. A sign-component-based framework for Chinese sign language recognition using accelerometer and sEMG data. *IEEE transactions on bio-medical engineering*, IEEE, v. 59, n. 10, p. 2695–704, out. 2012. ISSN 1558-2531. Disponível em: <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6170877>.
- 8 YANG, Q. Chinese sign language recognition based on video sequence appearance modeling. In: *2010 5th IEEE Conference on Industrial Electronics and Applications*. IEEE, 2010. p. 1537–1542. ISBN 978-1-4244-5045-9. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5514688>.
- 9 VOGLER, C.; METAXAS, D. ASL recognition based on a coupling between HMMs and 3D motion analysis. In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. Narosa Publishing

- House, 1998. p. 363–369. ISBN 81-7319-221-9. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=710744>.
- 10 VOGLER, C.; SUN, H.; METAXAS, D. A framework for motion recognition with applications to American sign language and gait recognition. In: *Proceedings Workshop on Human Motion*. IEEE Comput. Soc, 2000. p. 33–38. ISBN 0-7695-0939-8. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=897368>.
- 11 PORFIRIO, A. J. Reconhecimento das Configurações de mão da LIBRAS a Partir de Malhas 3D. 2013, Disponível em: <http://www.inf.ufpr.br/vri/alumni/2013-AndresPorfirio/pos-defesa-2013-03-26.pdf>.
- 12 CHEN, X. et al. *Kinect Sing Language Translator expands communication possibilities*. Disponível em: http://research.microsoft.com/en-us/collaboration/stories/kinectforsignlanguage_cs.pdf.
- 13 ZAFRULLA, Z. et al. A novel approach to American Sign Language (ASL) phrase verification using reversed signing. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. IEEE, 2010. p. 48–55. ISBN 978-1-4244-7029-7. ISSN 2160-7508. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5543268>.
- 14 SANTOS, A. P. T. et al. KINLIB - Protótipo de Ensino de LIBRAS Utilizando o KINECT. 2011, Disponível em: <http://engenharia.anhembib.br/tcc-11/cc0-02.pdf>.
- 15 OIKONOMIDIS, I.; KYRIAZIS, N.; ARGYROS, A. A. Tracking the articulated motion of two strongly interacting hands. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012. p. 1862–1869. ISSN 1063-6919. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6247885>.
- 16 PORTOGENTE. *Libras - Cursos - Portogente*. Acessado em: 2014-03-16. Disponível em: <http://portogente.com.br/cursos/comunicacao/libras-80423>.
- 17 GUIMARAES, C. et al. Evaluation of a computational description model of Libras (Brazilian Sign Language): Bridging the gap towards information access. In: *2011 FIFTH INTERNATIONAL CONFERENCE ON RESEARCH CHALLENGES IN INFORMATION SCIENCE*. IEEE, 2011. p. 1–10. ISBN 978-1-4244-8670-0. ISSN 2151-1349. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6006855>.
- 18 ANTUNES, D. R. et al. A framework to support development of Sign Language human-computer interaction: Building tools for effective information access and inclusion of the deaf. In: *2011 FIFTH INTERNATIONAL CONFERENCE ON RESEARCH CHALLENGES IN INFORMATION SCIENCE*. IEEE, 2011. p. 1–12. ISBN 978-1-4244-8670-0. ISSN 2151-1349. Disponível em: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6006832>.

- 19 DUDUCHI, M.; CAPOVILLA, F. C. BuscaSigno. In: *Proceedings of VII Brazilian symposium on Human factors in computing systems - IHC '06*. New York, New York, USA: ACM Press, 2006. p. 21. ISBN 1595934324. Disponível em: <http://dl.acm.org/citation.cfm?id=1298023.1298027>.
- 20 MICROSOFT. *Kinect*. Acessado em: 2014-03-04. Disponível em: <http://www.xbox.com/en-US/kinect>.
- 21 BLOG, M. P. C. *Part 1 - Introduction to Microsoft Kinect - Microsoft Pakistan Community Blog - Site Home - MSDN Blogs*. Acessado em: 2014-04-01. Disponível em: <http://blogs.msdn.com/b/pakistan/archive/2013/01/26/part-1-introduction-to-microsoft-kinect.aspx>.
- 22 BLOG, M. P. C. *Part 2 - Getting started with Microsoft Kinect - Microsoft Pakistan Community Blog - Site Home - MSDN Blogs*. Acessado em: 2014-04-01. Disponível em: <http://blogs.msdn.com/b/pakistan/archive/2013/02/03/part-2-getting-started-with-microsoft-kinect.aspx>.
- 23 RUSU, R. B.; COUSINS, S. 3D is here: Point Cloud Library (PCL). In: *IEEE International Conference on Robotics and Automation (ICRA)*. Shanghai, China: [s.n.], 2011.
- 24 IZADI, S. et al. Kinect Fusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. In: *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*. [s.n.], 2011. p. 559. ISBN 9781450307161. Disponível em: [http://dl.acm.org/citation.cfm?id=2047270\\$delimiter"026E30F\\$nhhttp://dl.acm.org/citation.cfm?doid=2047196.2047270](http://dl.acm.org/citation.cfm?id=2047270$delimiter).
- 25 NEWCOMBE, R. A. et al. KinectFusion: Real-time dense surface mapping and tracking. In: *2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011*. [S.l.: s.n.], 2011. p. 127–136.
- 26 SHOTTON, J. et al. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, ACM, v. 56, n. 1, p. 116, jan. 2013. ISSN 00010782. Disponível em: http://dl.acm.org/ft_gateway.cfm?id=2398381&type=html.

APÊNDICE A – MODELOS



A



C



D



E



F



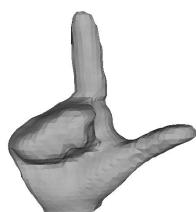
G



I



K



L



N



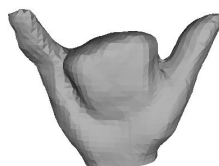
S



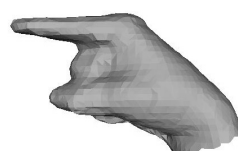
T



X



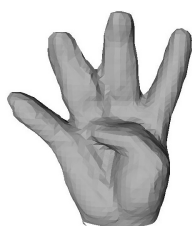
Y



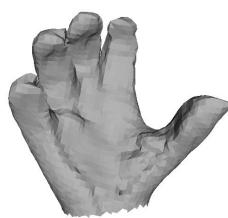
Z



1



4



Mão aberta - dedos curvados e separados



Mão aberta - dedos distendidos e paralelos



Mão aberta - dedos flexionados



Mão aberta - dedos separados



Mão aberta - dedos unidos



Mão aberta - dedos unidos e curvados



Mão aberta - ponta dos dedos unidos