

# Reconhecimento das Configurações de Mão da LIBRAS a Partir de Malhas 3D

Andres Jessé Porfirio

Orientador: Prof. Dr. Daniel Weingaertner

Universidade Federal do Paraná

# Sumário

- Introdução
- Abordagem Baseada nos Parâmetros Globais
- Utilização de Malhas 3D
- Objetivos
- Desenvolvimento
- Resultados
- Considerações Finais

# Introdução

- As línguas de sinais são as **principais ferramentas** de inclusão social dos surdos;
- Elas são utilizadas como meio de comunicação, obtenção de informações, educação, etc;
- A língua de sinais brasileira é chamada **LIBRAS;**

# Introdução

- Ferramentas têm sido desenvolvidas para facilitar a utilização dos **meios de comunicação digitais** por deficientes auditivos, ex:
  - **Reconhecimento automático de sinais;**
  - **Avatares animados em sistemas digitais;**
- Além disso, ferramentas tradutoras podem **favorecer a comunicação entre surdos e ouvintes** que não compreendem a língua de sinais;

# Introdução

- As **principais abordagens** utilizadas no reconhecimento de sinais são divididas quanto à **forma de percepção**:
  - Abordagem visual;
  - Abordagem sensorial;

# Introdução

- Abordagem Visual:
  - Baseia-se em **imagens ou sequencias de vídeo** do ator;
  - Abordagem não invasiva:
    - Não exige sensores acoplados ao ator;
    - Mais fácil de ser aplicada em ambientes reais;
  - Em alguns casos faz-se uso de **acessórios** como **luvas**;
  - Maior dificuldade de obtenção das informações de movimento,dedos, entre outros:
    - Necessidade de segmentação de imagens, rastreamento, etc;

# Introdução

- Abordagem Sensorial:
  - Faz uso de **sensores acoplados** ao corpo do ator:
    - Acelerômetros, sensores de uso médico para captura de movimentos, etc;
  - Abordagem invasiva:
    - Dificulta utilização em ambientes reais;
  - Pode causar **desconforto** ao ator;
  - Alguns sensores são disponibilizados a preços altos;

# Introdução

- Observações sobre os trabalhos encontrados na literatura:
  - Luvas e sensores limitam o uso, geram desconforto;
  - Vocabulários restritos e reconhecimento de letras do alfabeto não conferem poder de reconhecimento da língua como um todo;
- Artíficos úteis no contexto abordado, mas não para o **reconhecimento de todas os sinais/sentenças de uma língua;**



# Abordagem Baseada nos Parâmetros Globais

- Parâmetros globais são as primitivas da língua:
  - **União de parâmetros globais - Sinal**
  - **União de fonemas - Palavra**
- A estrutura fonética da LIBRAS é composta por cinco parâmetros globais:
  - **Configuração de mão;**
  - **Locação;**
  - **Movimento;**
  - **Orientação da palma da mão;**
  - **Expressão facial.**

# Abordagem Baseada nos Parâmetros Globais

- LIBRAS não são meros gestos ou mímicas e não é uma soletração do Português;
- O uso da **abordagem baseada em fonemas é mais adequada** para o reconhecimento de sinais;
- O parâmetro abordado por este trabalho é a **configuração de mão**;

# Abordagem Baseada nos Parâmetros Globais



Figura 1: Configurações de Mão da LIBRAS [1].

# Utilização de Malhas 3D

- Este trabalho difere de grande parte das publicações relacionadas:
  - Reconhecimento a partir de **malhas 3D** e não imagens/vídeos 2D;
- Vantagens:
  - Capacidade de identificar detalhes da mão e dedos independentemente da **orientação** da mão;
  - Problemas de **oclusão** podem ser evitados;
  - Invariância às condições de **iluminação**;

# Utilização de Malhas 3D

- Parte-se do princípio de que **a aquisição de malhas 3D em tempo real é um problema com solução em desenvolvimento:**
  - Novos sensores e técnicas estão sendo desenvolvidos e já apresentam resultados promissores (caso do LeapMotion – Figura 2).

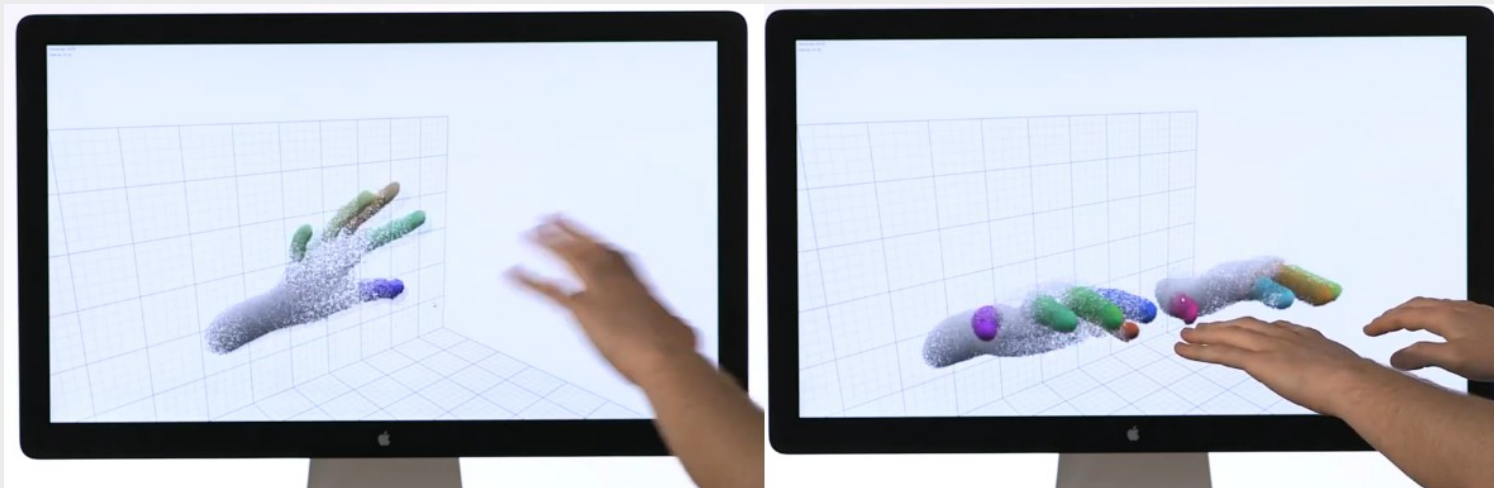


Figura 2: LeapMotion.

# Objetivos

- **Principal:**

- Reconhecer as configurações de mão da LIBRAS a partir de malhas 3D;

- **Específicos:**

- Capturar uma base de dados das configurações de mão da LIBRAS com o sensor Kinect;
- Gerar uma base de malhas 3D com o método da reconstrução por silhueta;
- Determinação e extração de características;
- Classificação das malhas 3D;
- Gerar uma base 2D das configurações de mão;

# Desenvolvimento

- O trabalho foi desenvolvido em etapas, conforme mostra a Figura 3:

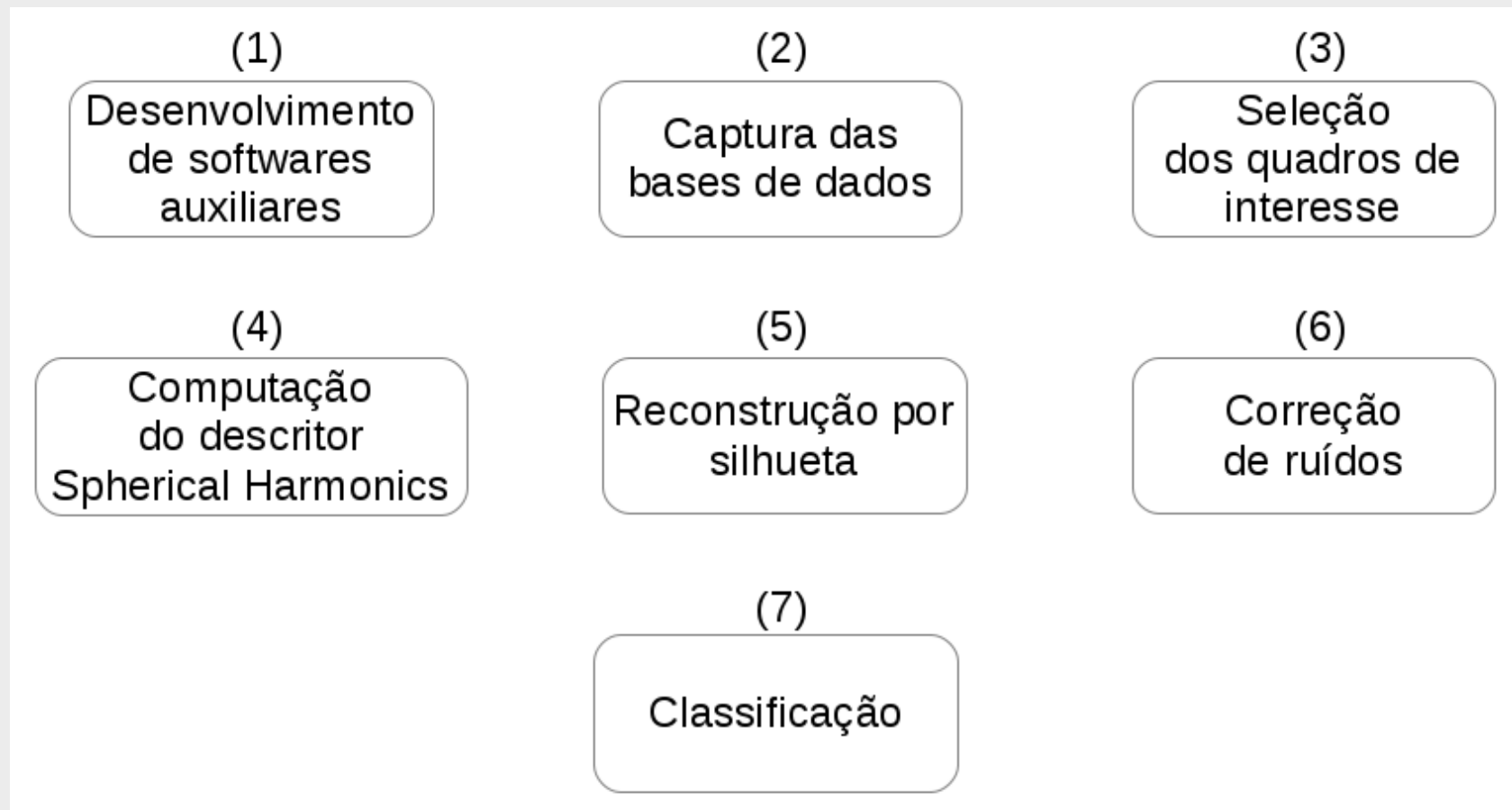


Figura 3: Etapas do desenvolvimento.

# Desenvolvimento

- Desenvolvimento de Softwares Auxiliares:
  - NiCapture 1.3:
    - Auxilia na captura de vídeos com o Kinect;
    - Organiza a base no padrão **CX\_Y\_Z**:
      - Ex: C5\_3\_2 são os dados do quinto ator, configuração de mão 3, segunda tomada.
    - Armazena o esqueleto 3D (Figura 4), dispensando a calibração e rastreamento a cada futura execução do vídeo;



Figura 4: Esqueleto 3D.



# Desenvolvimento

- Desenvolvimento de Softwares Auxiliares:
  - NiCapture 1.3:
    - Armazena os vídeos no formato **RGBDS** – Red, Green, Blue, Depth, Skeleton;
    - Exibe controles de captura no Laptop e exemplos das configurações de mão no monitor auxiliar (Figura 5);

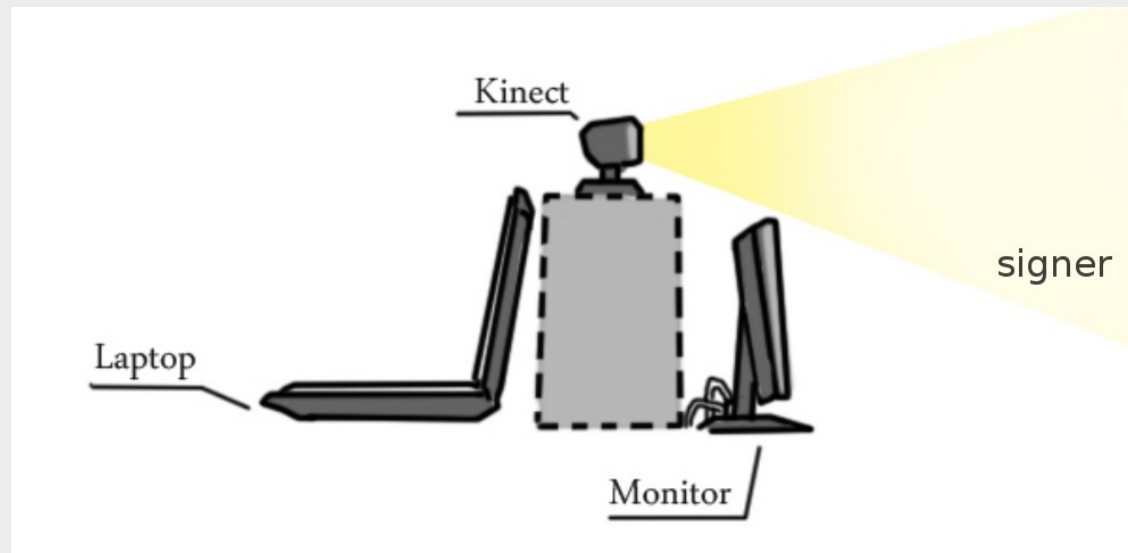


Figura 5: Ambiente de Captura.

# Desenvolvimento

- Desenvolvimento de Softwares Auxiliares:
  - NiViewSKL 1.3.1:
    - Reproduz os vídeos gerados pelo NiCapture 1.3;
    - Converte um vídeo em uma sequencia de imagens 2D com a mão segmentada;

# Desenvolvimento

- Captura da Base de Dados:
  - A configuração das capturas pode ser vista na Figura 6:

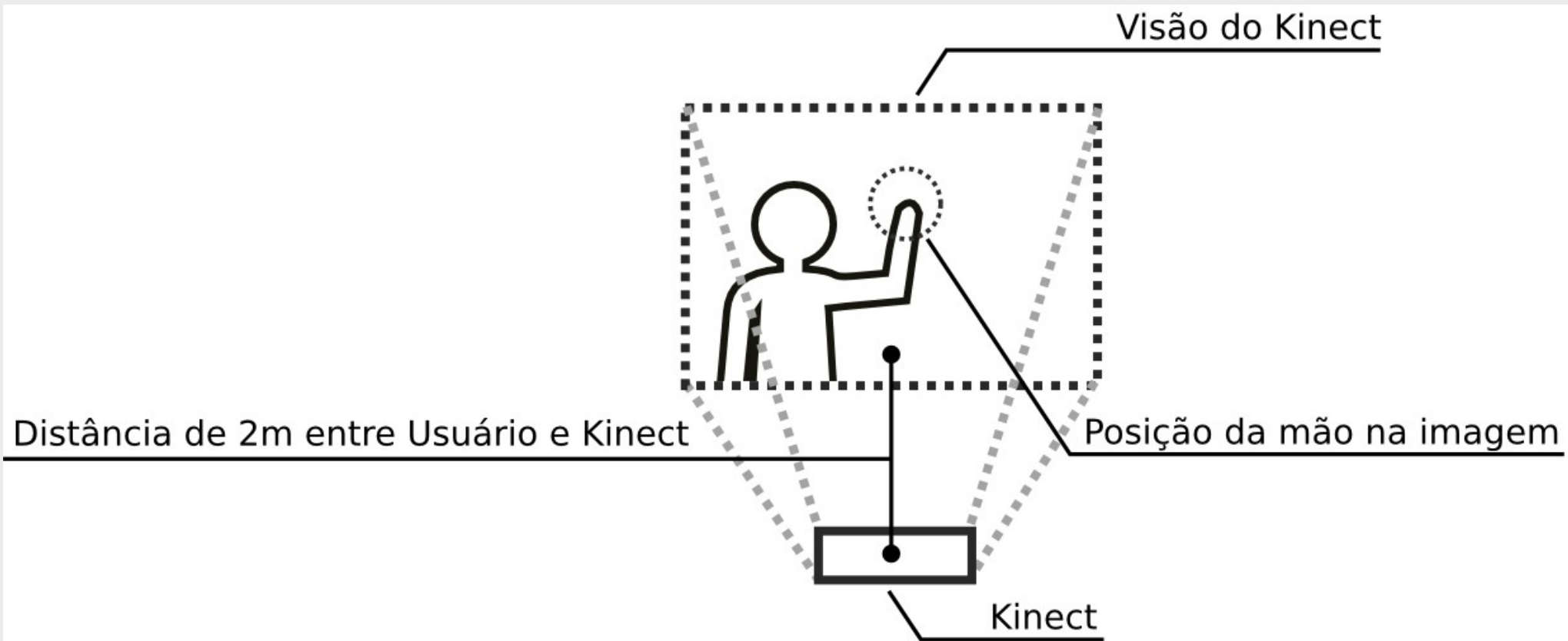


Figura 6: Configuração das Capturas.

# Desenvolvimento

- Captura da Base de Dados:
  - A sequencia de movimentos realizada pelo ator para cada configuração de mão pode ser vista na Figura 7:

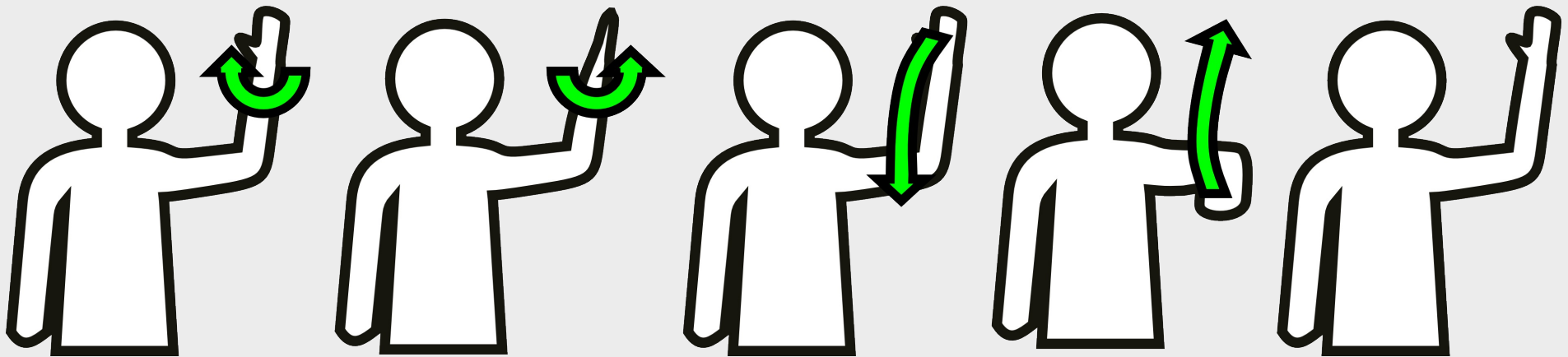


Figura 7: Sequencia de Movimentos das Capturas.

# Desenvolvimento

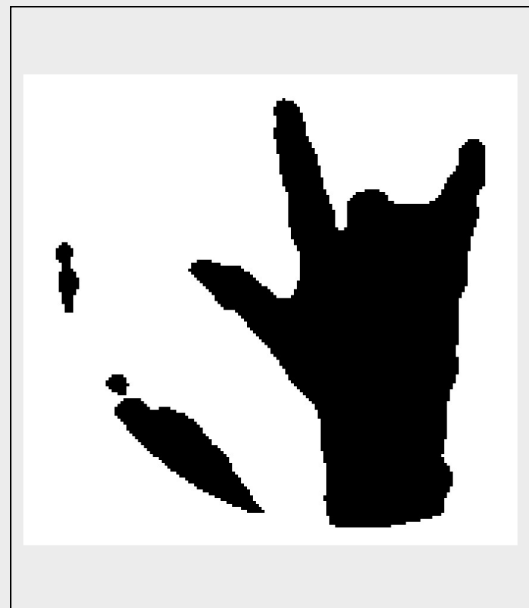
- Seleção dos Quadros de Interesse:
  - Este trabalho fez uso da reconstrução por silhueta baseada em duas imagens;
  - Esta técnica requer como entrada **duas visões** da mão: **frontal** e **lateral**;
  - A seleção dos quadros foi feita manualmente.

# Desenvolvimento

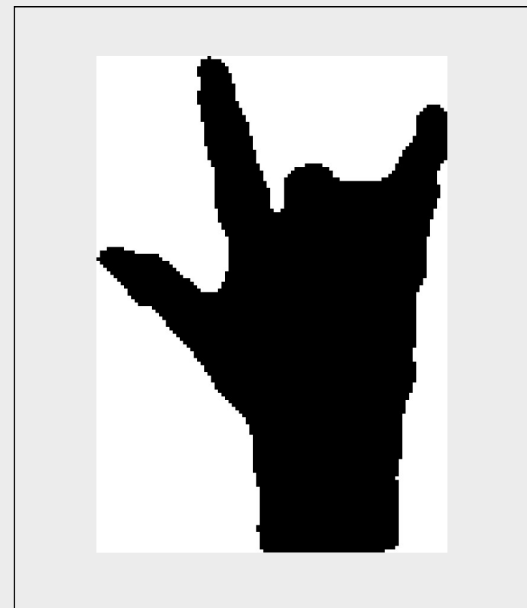
- Correção de Ruídos:
  - Através da análise das imagens foi constatado que grande parte possuía ruídos;
  - Ruídos geram **malhas incorretas**, dificultando a classificação;
  - Estas imagens foram submetidas à **correção manual**;

# Desenvolvimento

- Correção de Ruídos:
  - Principais tipos de ruídos: **Segmentação** (Figura 8).



(a)



(b)

Figura 8: Ruído de Segmentação (a), Correção (b).

# Desenvolvimento

- Correção de Ruídos:
  - Principais tipos de ruídos: **Rotação incorreta** (Figura 9).

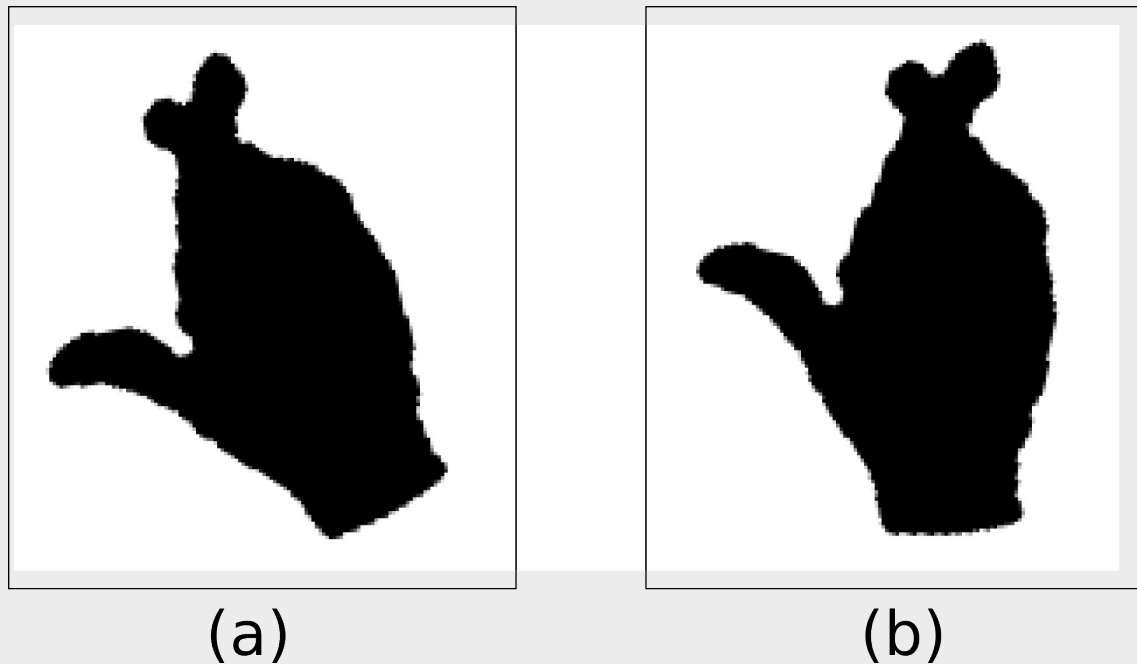
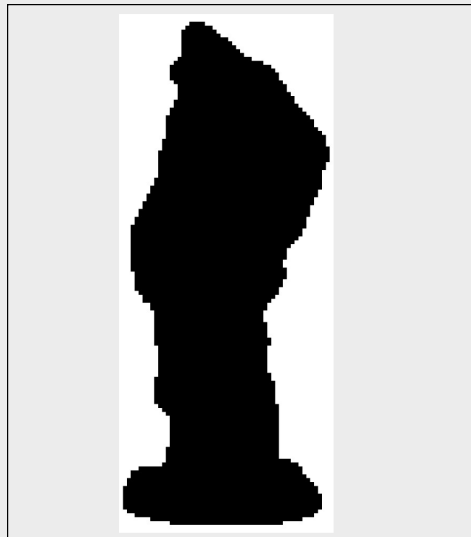


Figura 9: Rotação Incorreta (a), Correção (b).

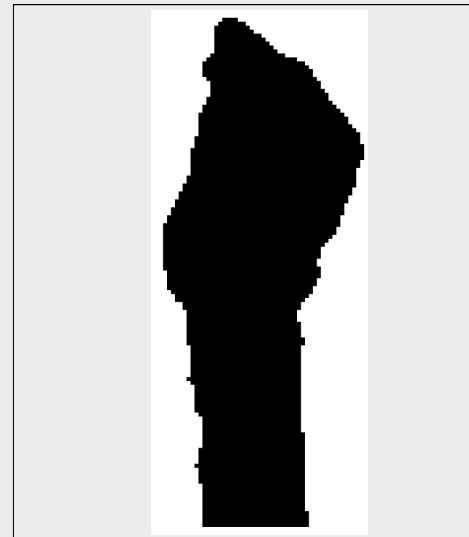


# Desenvolvimento

- Correção de Ruídos:
  - Principais tipos de ruídos: Aparecimento de artefatos como a **manga da blusa** (Figura 10).



(a)



(b)

Figura 10: Aparecimento da Manga da Blusa (a), Correção (b).

# Desenvolvimento

- Correção de Ruídos:
  - Resultados da correção:

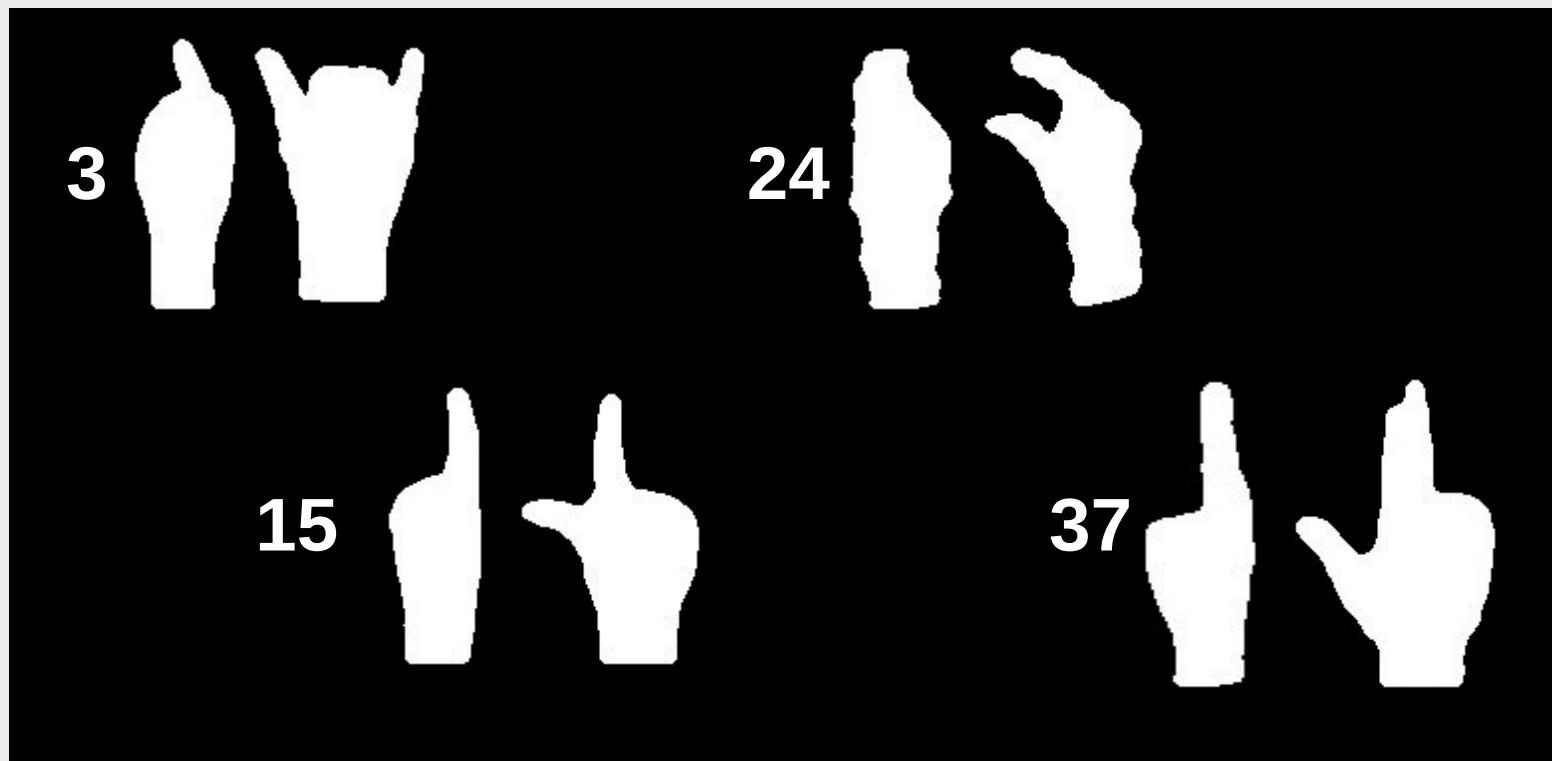


Figura 11: Exemplos de imagens corrigidas relativas às configurações de mão 3,15,24 e 37.

# Desenvolvimento

- Reconstrução por Silhueta:
  - Também conhecida como **Shape From Silhouette** ou **Visual Hull Construction**;
  - A reconstrução é feita através da estimativa da malha 3D por meio de imagens de **vários pontos de vista de um objeto** (Figura 12);

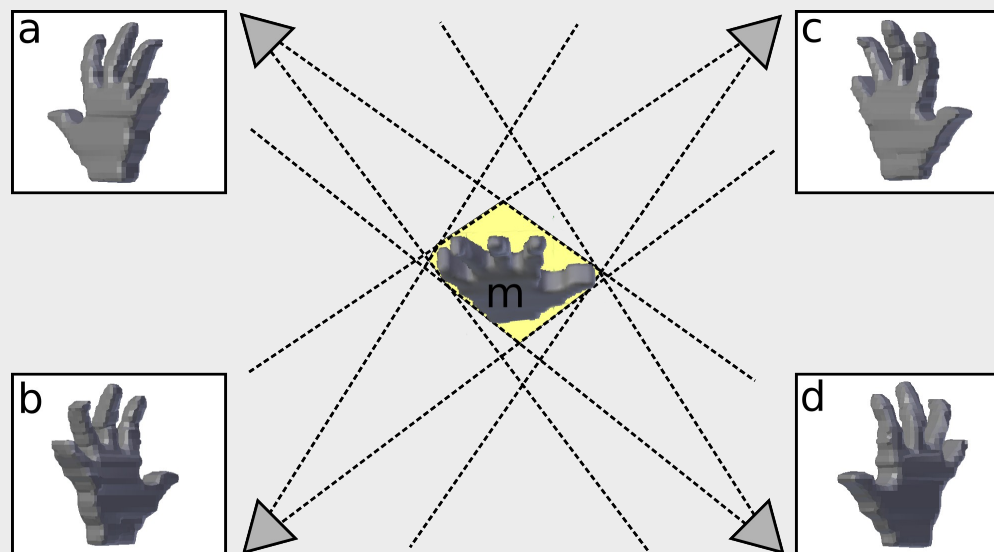


Figura 12: Reconstrução por Silhueta.

# Desenvolvimento

- Reconstrução por Silhueta:
  - Foi desenvolvido com base em dois conceitos:
    - Produção do volume da silhueta;
    - Diferença entre malhas;

# Desenvolvimento

- Reconstrução por Silhueta:
  - Produção do volume da silhueta (Figura 13):

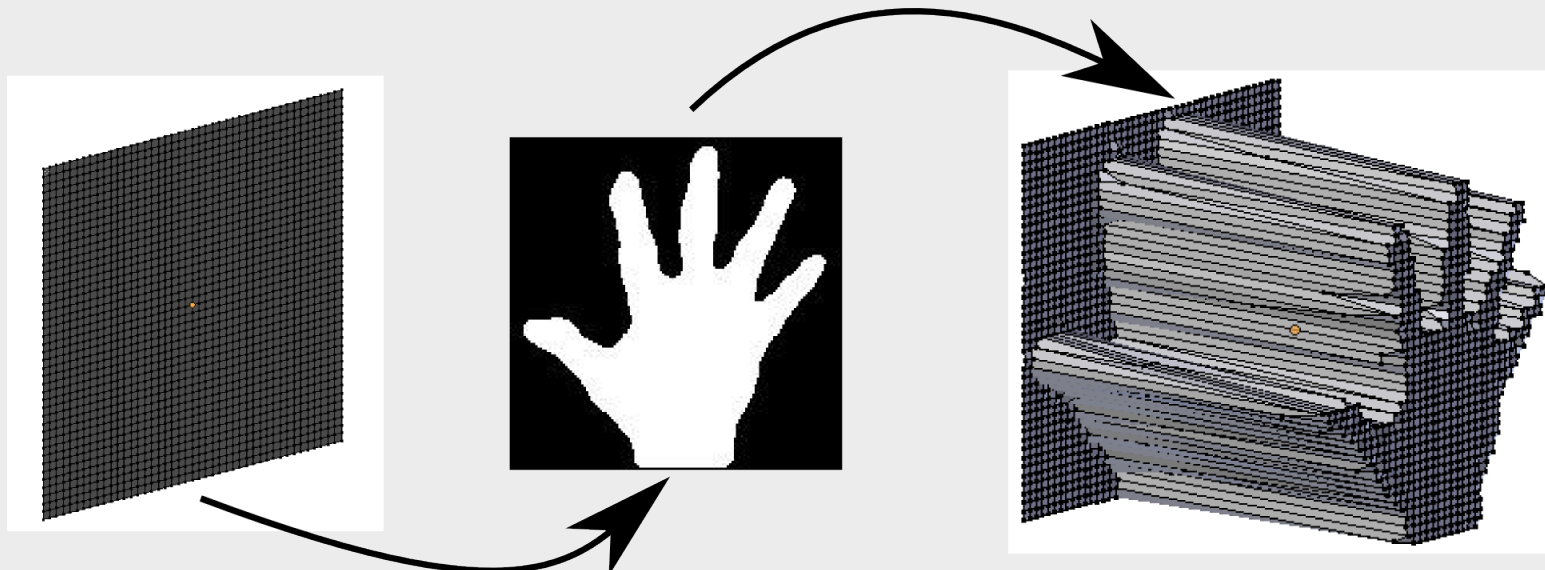


Figura 13: Produção da Silhueta.

# Desenvolvimento

- Reconstrução por Silhueta:
  - Diferença entre malhas (Figura 14):

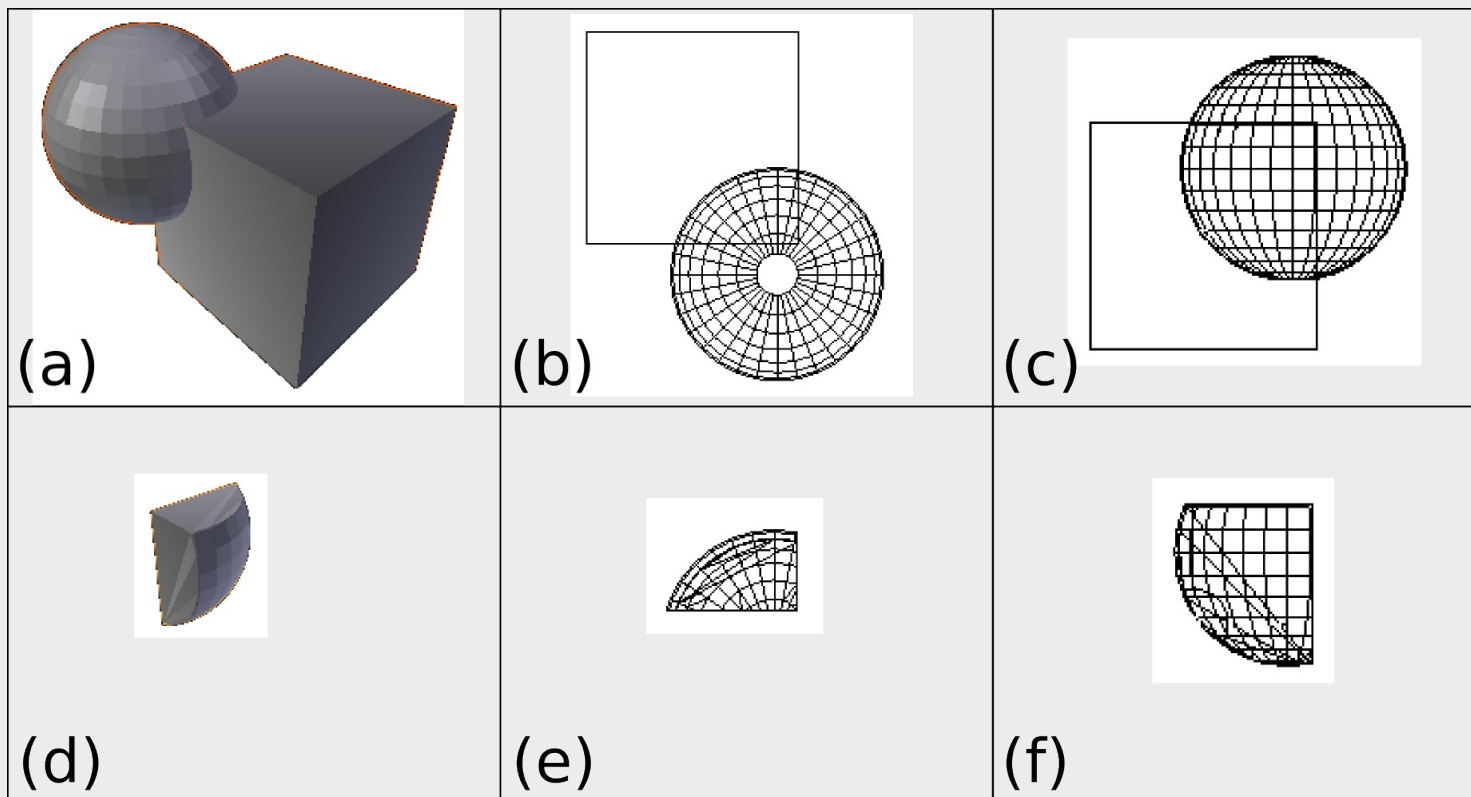


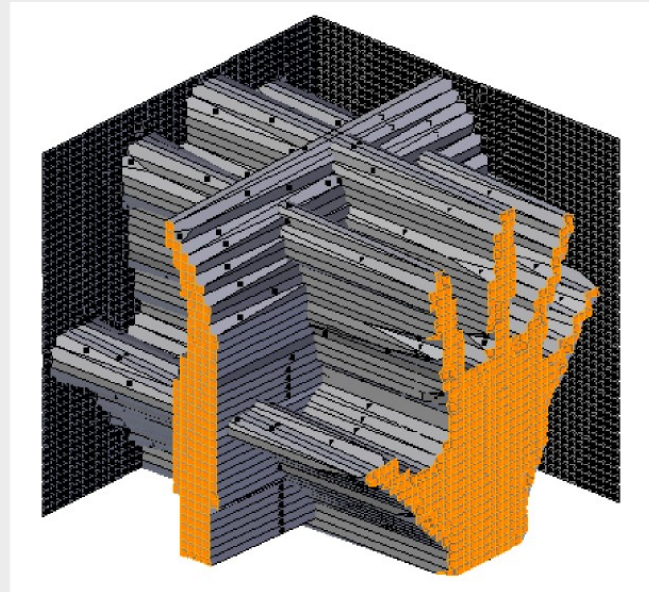
Figura 14: Diferença entre Malhas.

# Desenvolvimento

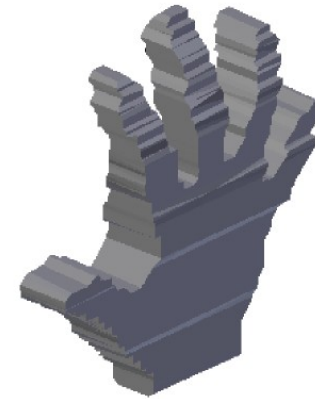
- Reconstrução por Silhueta:
  - Aplicação dos dois conceitos (Figura 15):



(a)



(b)



(c)

Figura 15: Produção dos Volumes e Diferença entre Malhas.

# Desenvolvimento

- Reconstrução por Silhueta:
  - Após a reconstrução foi aplicado um filtro de suavização e reorganização poligonal (Figura 16):

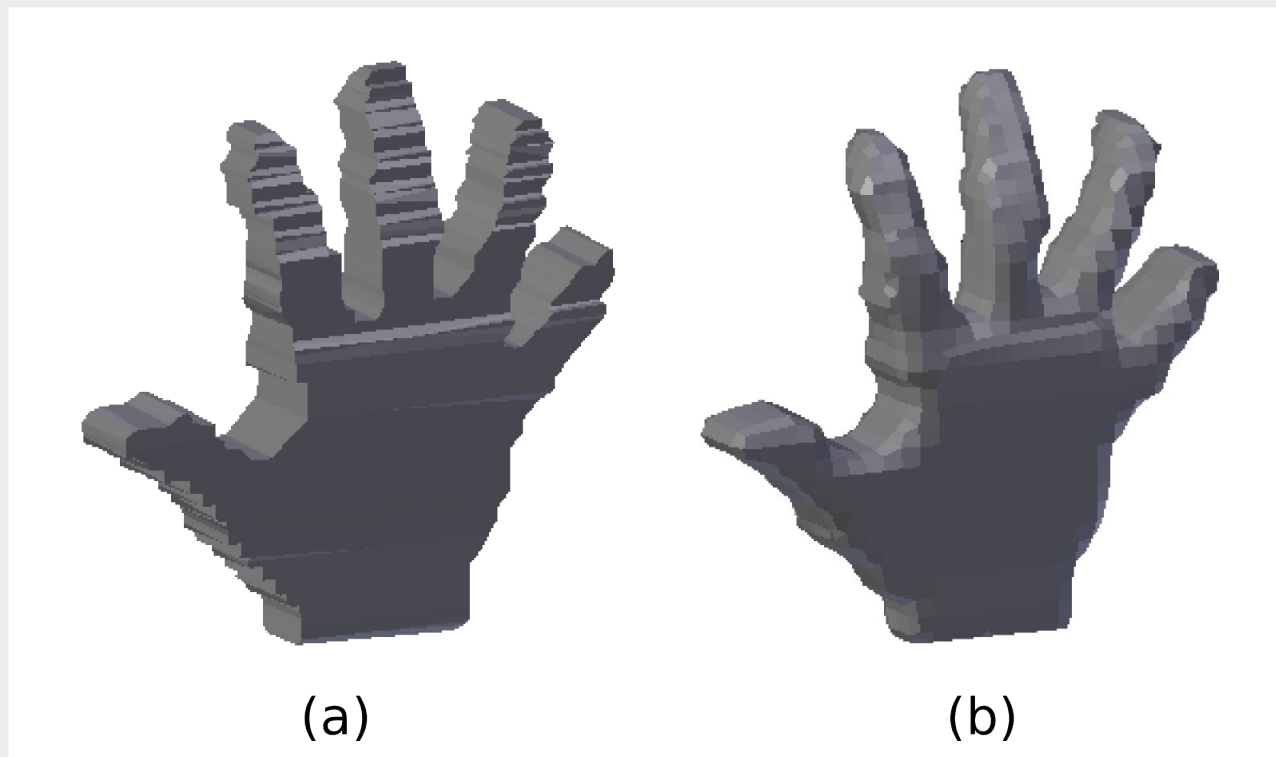


Figura 16: Pré-processamento da Malha.



# Desenvolvimento

- Reconstrução por Silhueta:
  - Exemplos de malhas 3D após pré-processamento (Figura 17):

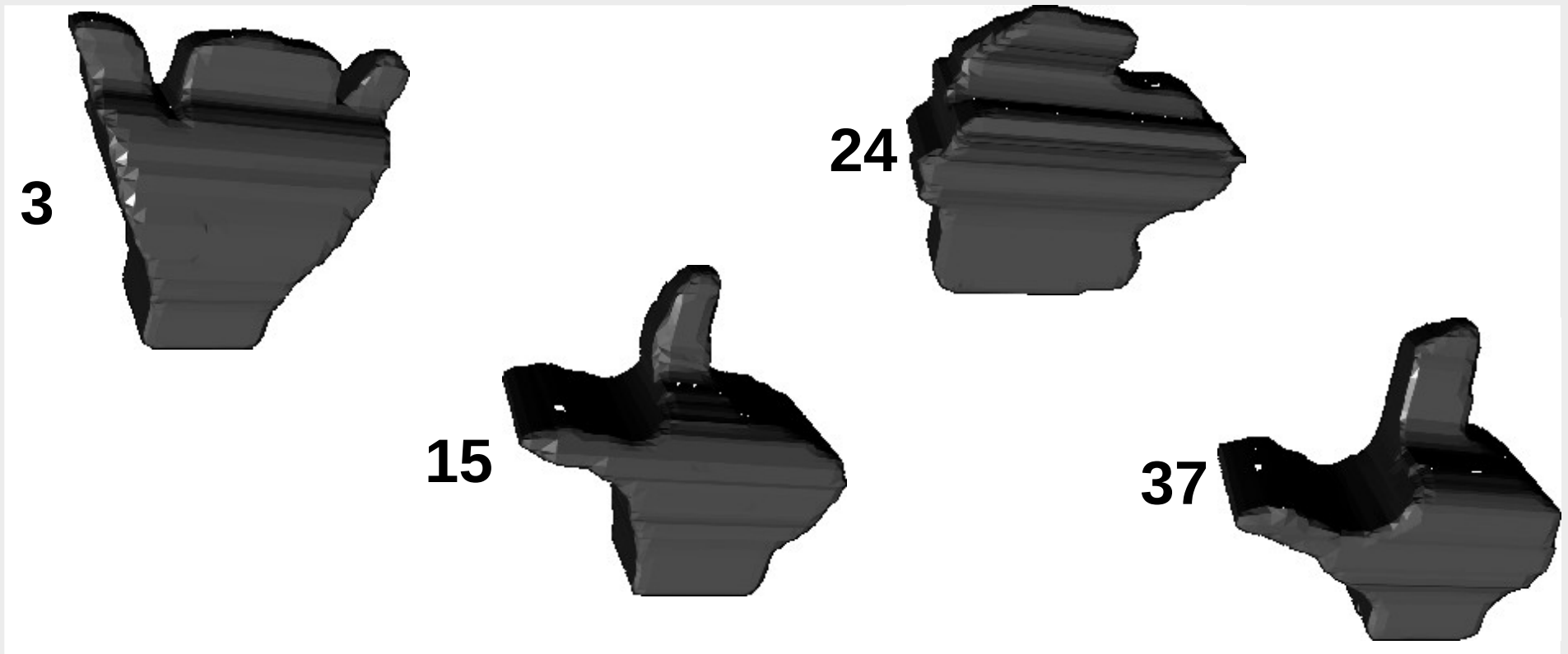


Figura 17: Exemplos de malhas 3D referentes às configurações de mão 3, 15, 24 e 37.

# Desenvolvimento

- Descritor Spherical Harmonics:
  - Descritor de malhas 3D;
  - Invariante à **iluminação, translação, rotação e escala**;
  - Calculado com o auxílio de uma ferramenta desenvolvida por Michael Kazhdan [2].

# Desenvolvimento

- Descritor Spherical Harmonics:

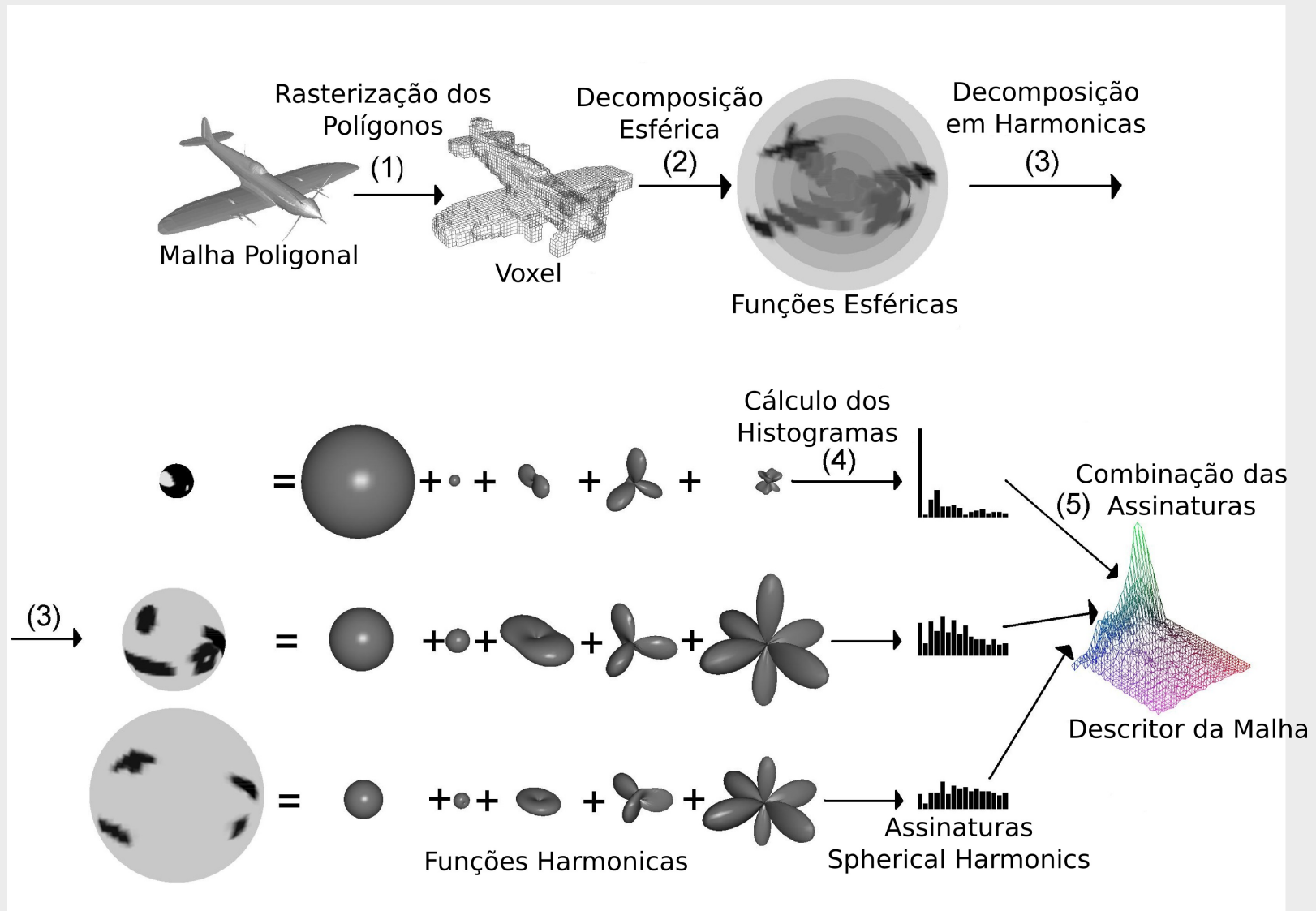


Figura 18: Computação do Descritor Spherical Harmonics.

# Desenvolvimento

- Classificação:
  - **Support Vector Machine – SVM;**
  - Classificador que faz uso de **aprendizado supervisionado;**
  - Objetivo principal a classificação de problemas com duas classes;
  - Pode ser aplicado em conjuntos de dados não separáveis linearmente;
  - Utilizável em bases com múltiplas classes através do uso da abordagem "**um-contra-um**" e **pairwise** [3].

# Desenvolvimento

- Classificação:
  - Utilização da biblioteca LibSVM;
  - Configurações do SVM:
    - Kernel Linear e Kernel RBF;
    - Ranking 1, 3 e 5;
    - Média de 10 execuções em cada configuração;
  - Configuração da base de dados:
    - 70% das instâncias para treino e 30% para testes.
    - 10 subconjuntos de dados provenientes de divisões aleatórias 70/30.

# Resultados

- Foram construídas duas bases de dados:
  - **HC-LIBRAS-RGBDS-2011:**
    - 5 atores;
    - 2 tomadas por ator;
    - 61 configurações de mão;
    - Total: 610 vídeos.
  - **HC-LIBRAS-RGB-2011:**
    - Base complementar 2D;
    - Imagens RGB extraídas da base RGBDS;
    - Visão frontal da mão;

# Resultados

- Foi construída uma base de malhas 3D:
  - Para cada dupla de imagens frontal/lateral foi gerada uma malha 3D;
  - Total: **610 malhas**, 10 para cada configuração de mão da LIBRAS;
  - Todas as malhas foram submetidas ao filtro de suavização;
  - Resultado armazenado em dois formatos:
    - **Blend** – Blender 3D.
    - **Stanford PLY** – Formato de entrada da ferramenta que calcula o descritor Spherical Harmonics;

# Resultados

- Classificação:
  - Inicialmente os arquivos de assinatura foram convertidos para o formato da LibSVM;
  - A base foi subdividida 10 vezes na proporção 70/30, gerando 10 subconjuntos de testes;
  - O descritor Spherical Harmonics foi utilizado para formar o vetor de características;



# Resultados

- Classificação:

- A Tabela 1 apresenta a média de 10 execuções do SVM com Kernel RBF:

Tabela 1: SVM com Kernel RBF.

Ranking	Taxa de acerto (%)	Desvio Padrão (%)
1	85.68	1.65
3	96.67	1.27
5	98.52	0.86

- A Tabela 2 apresenta a média de 10 execuções do SVM com Kernel Linear:

Tabela 2: SVM com Kernel Linear.

Ranking	Taxa de acerto (%)	Desvio Padrão (%)
1	86.06	2.06
3	96.83	1.31
5	98.57	0.60

# Resultados

- Classificação:
  - Análise das malhas que causaram erro de classificação:



Figura 19: Configurações de mão parecidas.

# Resultados

- Comparação com o trabalho mais similar encontrado na literatura:
  - Foi efetuado o reconhecimento das configurações de mão relativas aos 10 dígitos:







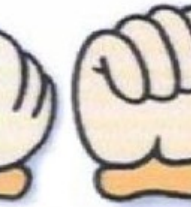

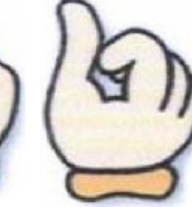











Dígito	0	1	2	3	4	5	6	7	8	9
	22	2	38	51	55	31	25	9	7	25
Conf. Mão										
Malha 3D										

Figura 20: Configurações de mão e malhas 3D relativas aos 10 dígitos.

# Resultados

- Comparação com o trabalho mais similar encontrado na literatura:
  - A configuração de mão 25 se repete para os dígitos 6 e 9;
  - A segunda ocorrência foi substituída pela configuração de mão 60;
  - Têm-se então um problema com 10 classes, similar ao trabalho de Keskin [4];

# Resultados

- Comparação com o trabalho mais similar encontrado na literatura:
  - O autor Keskin [4] obteve uma taxa de acerto de 99.9%;
  - O presente trabalho obteve uma taxa de **98.33%** de acerto em condições similares às testadas por Keskin.
  - A Tabela 3 mostra os resultados obtidos com este teste:

Tabela 3: Resultados da classificação de malhas 3D com 10 classes.

Ranking	Taxa de acerto (%)	Desvio Padrão (%)
1	98.33	1.66
3	100	0
5	100	0

# Considerações Finais

- A utilização da abordagem baseada nos parâmetros globais é promissora e viável;
- Novas técnicas e dispositivos estão sendo desenvolvidos e, futuramente, viabilizarão o reconhecimento em ambientes reais;
- O reconhecimento de configurações de mão a partir de malhas 3D é uma alternativa às técnicas clássicas que oferece benefícios como a **invariância à orientação da mão e luminosidade**;

# Referências

[1] Nelson Pimenta e Ronice Muller de Quadros. Curso LIBRAS 1 4a Edição. Editora Vozes, 2010.

[2] Spherical Harmonics, Michael Kazhdan, disponível em <http://www.cs.jhu.edu/~misha/HarmonicSignatures/>. Acesso em 10/02/13.

[3] C. Chang e C. Lin. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011.

[4] C Keskin, F. Kiraç, Y. E. Kara, e L. Akarun. Real time hand pose estimation using depth sensors. ICCV Workshops, páginas 1228–1234. IEEE, 2011.