



UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

Alex Torquato Souza Carneiro

Sistema de Reconhecimento do Alfabeto da LIBRAS por Visão Computacional e Redes Neurais

FORTALEZA – CEARÁ
AGOSTO 2010

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

ALEX TORQUATO SOUZA CARNEIRO

**Sistema de Reconhecimento do Alfabeto da LIBRAS por
Visão Computacional e Redes Neurais**

*Dissertação de Mestrado submetida
ao corpo docente da Coordenação
do Curso de Pós-Graduação em
Engenharia de Teleinformática da
Universidade Federal do Ceará como
parte dos requisitos para obtenção do
grau de **Mestre em Engenharia de
Teleinformática.***

Área de Concentração: Sinais e
Sistemas

Orientador : Prof. Dr. Paulo César
Cortez

FORTALEZA – CEARÁ

AGOSTO 2010

Ficha elaborada pela bibliotecária Umbelina Caldas Neta - CRB558-CE

C288s Carneiro, Alex Torquato Souza
Sistema de reconhecimento do alfabeto da LIBRAS por visão computacional e redes neurais / Alex Torquato Souza Carneiro, 2010.
98f. ; il.; enc.

Orientador: Prof. Dr. Paulo César Cortez
Área de concentração: Sinais e sistemas
Dissertação (Mestrado) - Universidade Federal do Ceará, Departamento de Engenharia de Teleinformática, Fortaleza, 2010.

1. Teleinformática. 2. Sinais e Sistemas. 3. Visão por computador.
I. Cortez, Paulo César (orient.). II. Universidade Federal do Ceará – Programa de Pós- Graduação em Engenharia de Teleinformática. III. Título.

CDD 621.38

ALEX TORQUATO SOUZA CARNEIRO

SISTEMA DE RECONHECIMENTO DO ALFABETO LIBRAS POR VISÃO COMPUTACIONAL E REDES NEURAIAS

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Engenharia de Teleinformática, da Universidade Federal do Ceará, como requisito parcial para a obtenção do grau de Mestre em Engenharia de Teleinformática.

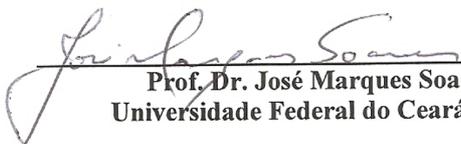
Área de concentração sistemas e sistemas.

Aprovada em 23/04/2010.

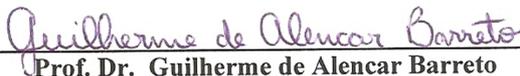
BANCA EXAMINADORA



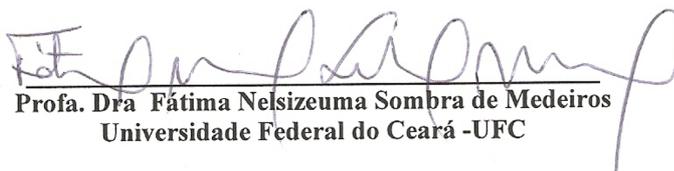
Prof. Dr. Paulo César Cortez (Orientador)
Universidade Federal do Ceará -UFC



Prof. Dr. José Marques Soares
Universidade Federal do Ceará -UFC



Prof. Dr. Guilherme de Alencar Barreto
Universidade Federal do Ceará -UFC



Profa. Dra Fátima Nelsizeuma Sombra de Medeiros
Universidade Federal do Ceará -UFC



Prof. Dr. Waslon Terlizzi Araujo Lopes
Universidade Federal de Campina Grande - UFCG

Resumo

Atualmente, os surdos encontram dificuldades de se comunicar com as pessoas ouvintes, principalmente pela escassez de formas alternativas à comunicação verbal. Este trabalho aborda o ambiente social em que a comunidade surda se encontra inserida atualmente, destacando-se a falta de assistência social e tecnológica necessária à inclusão do surdo na sociedade contemporânea. A atividade proposta neste trabalho consiste no desenvolvimento de um sistema de visão computacional, intitulado Sistema de Reconhecimento da LIBRAS por Visão Computacional e Redes Neurais, capaz de fazer a conversão do alfabeto da Língua Brasileira de Sinais (LIBRAS) para o alfabeto latino, o qual é adotado por diversas línguas, inclusive o português falado no Brasil. Este sistema é composto por quatro etapas principais: aquisição da imagem, segmentação da pele, extração de características da imagem e classificação do gesto. O sistema proposto obteve taxas de acerto de até 82,67% nos testes realizados, deixando-o em equivalência aos sistemas semelhantes encontrados na literatura.

Palavras-chave: Língua de Sinais Brasileira, Redes Neurais, Visão Computacional.

Abstract

Nowadays , deaf people have difficulties to communicate with hearing ones, mainly because the verbal communication has no substitutes. This work approaches the social environment that the deaf community is inserted, specially the lack of social and technologic support for inclusion in contemporary society. The propused activity of this work is to develop a computer vision system, called System of LIBRAS Alphabet Recognition Based on Computer Vision and Neural Networks, able to convert the alphabet of Brazilian Sign Language (LIBRAS) to latin alphabet, which is used for many languages, including Brazilian Portuguese. This system consists of four main parts: image acquisition, segmentation of the skin, extration of image's features and gesture classification. The proposes system has obtained hit rate of up to 82.67% in the tests, leaving it in equivalence to others systems found in literature.

Keywords: Brazilian Sign Language, Neural Networks, Computational Vision.

Aos meus pais, Gilberto e Mônica,
à minha irmã Viviane,
à minha namorada Cleíze
e aos irmãos da Santa Mãe Igreja Católica
pela ajuda e atenção nos momentos de necessidade.

Agradecimentos

Antes de todos, agradeço à Deus, meu Senhor e Redentor, que esteve sempre comigo na pessoa de Jesus Cristo, principalmente quando mais precisei.

À Santa Maria, Mãe de Deus, Santo Antônio e os anjos do Senhor pela proteção e intercessão.

Ao meu Orientador Prof. Dr. Paulo César Cortez, pela confiança depositada, paciência e orientação.

Aos professores, funcionários e agregados ao Departamento de Engenharia de Teleinformática, ao Programa de Pós-Graduação em Engenharia de Teleinformática e ao Laboratório de Teleinformática (LATIN/LESC), em especial o Rodrigo Costa, por todo o auxílio e dedicação.

Aos meus pais Gilberto e Mônica por todas as condições que me deram para concluir este trabalho e minha irmã Viviane pelos bons momentos de conversa.

À minha namorada Maria Cleíze pela atenção, apoio, carinho e incentivo.

Aos meus amigos, colegas e demais familiares.

À Associação dos Surdos do Ceará (ASCE), juntamente à Federação Nacional de Educação e Integração de Surdos (FENEIS-CE) e à Associação de Pais e Amigos de Deficientes Auditivos (APADA-CE) por toda a ajuda neste trabalho.

À Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) pelo suporte financeiro.

A todos que colaboraram de forma direta ou indireta na elaboração deste trabalho.

Uma hora de sofrimento por amor e submissão à Vontade de Deus
vale mais do que muitos dias de trabalho,
feitos com menos amor.

S. Francisco de Sales

Sumário

Lista de Figuras	xi
Lista de Tabelas	xiii
Lista de Siglas	xiii
1 Introdução	1
1.1 Motivação	2
1.2 Objetivos deste trabalho	3
1.3 Organização da dissertação	3
1.4 Produção científica	4
2 Fundamentação teórica	5
2.1 Pré-processamento	6
2.2 Segmentação	7
2.2.1 Segmentação por limiarização	9
2.2.2 Segmentação por modelo de probabilidade	9
2.3 Pós-processamento	9
2.3.1 Crescimento de região	10
2.3.2 Detecção de bordas	10
2.3.3 Morfologia matemática	10
2.4 Extração de atributos	11
2.4.1 Momentos Invariantes de Hu	11

2.4.2	Descritores de Fourier	12
2.4.3	Componentes de velocidade	14
2.5	Classificação	14
2.5.1	Mapas Auto-Organizáveis	15
2.5.2	<i>Perceptron</i> Simples	16
2.5.3	<i>Perceptron</i> Multi Camada	18
2.5.4	Funções de Base Radial	19
2.6	Resumo do capítulo	21
3	Estado da arte em reconhecimento de gestos manuais	22
3.1	Sistemas de reconhecimento de gestos baseados em sensores mecânicos	22
3.2	Sistemas de reconhecimento de gestos baseados em visão computacional	23
3.2.1	Algoritmos presentes na literatura especializada	24
4	Metodologia	28
4.1	Visão geral do sistema	29
4.2	Aquisição das imagens	30
4.3	Etapa de pré-processamento	31
4.4	Etapa de segmentação da pele	34
4.4.1	Algoritmo de segmentação por conjunto de limiares no espaço de cores RGB	35
4.4.2	Algoritmos de segmentação por limiarização simples	35
4.4.3	Algoritmo de segmentação por modelo de probabilidade	36
4.5	Pós-processamento	37
4.6	Etapa de extração de atributos da mão	39
4.7	Pré-classificação dos gestos	41
4.8	Classificação dos gestos	44
5	Resultados e discussão	46
5.1	Resultados do pré-processamento	46
5.2	Resultados da segmentação	48
5.3	Resultados do pós-processamento	51

5.4	Resultados da classificação	51
5.4.1	Avaliação individual dos resultados	52
5.4.2	Avaliação geral dos atributos	55
5.5	Discussão dos resultados da classificação	56
5.6	Concepção do sistema baseada nos resultados	56
6	Conclusões, Contribuições e Trabalhos Futuros	60
6.1	Trabalhos Futuros	61
	Apêndice A LIBRAS	63
A.1	Alfabeto da LIBRAS	64
	Apêndice B Análise dos classificadores supervisionados	69
	Referências Bibliográficas	98

Lista de Figuras

2.1	diagrama de blocos de um sistema de visão computacional genérico conforme Gonzalez e Woods (2008).	5
2.2	(a) histograma original, (b) histograma desejado e (c) histograma obtido.	7
2.3	modelo de neurônio artificial.	15
2.4	arquitetura de rede neural do tipo PS.	17
2.5	modelo da arquitetura multi camadas da rede MLP, com uma camada oculta.	19
2.6	modelo de arquitetura da rede RBF.	20
3.1	Cyberglove [®] em uso.	23
3.2	exemplo de um sistema de reconhecimento de gestos baseado em visão computacional.	24
3.3	sistema proposto por Zahedi <i>et al.</i> (2006).	26
3.4	sistema proposto por Pistori <i>et al.</i> (2006).	27
4.1	processos de um sistema de visão computacional simplificado.	28
4.2	alfabeto da LIBRAS.	29
4.3	diagrama do sistema proposto.	30
4.4	imagens relativas à letra 'A' do alfabeto da LIBRAS de cada uma das pessoas que serviram de modelo.	30
4.5	(a) imagem original e os resultados dos algoritmos de segmentação (b) em RGB, (c) em YCbCr, (d) em HSV e (e) pelo modelo de probabilidade.	32

4.6	resultado de referência dos algoritmos de segmentação (a) imagem original colorida e (b) imagem segmentada.	32
4.7	histogramas relativos aos canais (a) Cb e (b) Cr da imagem de referência.	33
4.8	resultado da separação dos histogramas em pele e fundo.	34
4.9	regiões com falhas no processo de segmentação.	37
4.10	imagens obtidas pelo processo de pós-processamento (a) a primeira, (b) a segunda aplicação do algoritmo e (c) resultado final.	38
4.11	processo de detecção de bordas (a) imagem colorida original, (b) imagem binária segmentada, e (c) contorno da imagem.	39
4.12	(a) imagem colorida original, (b) imagem binária segmentada e (c) imagem em níveis de cinza segmentada.	40
4.13	sub-amostragem do contorno original: (a) contorno original, contorno com (b) 10, (c) 30 e (d) 50 amostras.	40
4.14	exemplo de aplicação da rede SOM para <i>clustering</i> (a) aplicação da rede SOM sobre todos os dados e (b) resultado após o processo de <i>clustering</i>	42
4.15	diagrama do sistema de classificação em dois passos.	44
5.1	(a)-(c) imagens originais, (d)-(f) imagens obtidas após a aplicação do algoritmo de especificação de histograma.	47
5.2	(a)-(c) resultados obtidos pela segmentação manual e (d)-(f) resultados obtidos pela segmentação das imagens das Figuras 5.1(a) a 5.1(c) e (g)-(i) resultados obtidos pela segmentação das imagens das Figuras 5.1(d) a 5.1(f).	47
5.3	imagens originais de seis pessoas: (a) primeira, (b) segunda, (c) terceira, (d) quarta, (e) quinta e (f) sexta.	48
5.4	resultados da segmentação ideal.	49
5.5	resultados da segmentação para os algoritmos (a) - (f) limiarização em YCbCr, (g) - (l) limiarização em HSV, (m) - (r) limiares em RGB e (s) - (x) modelo de probabilidade.	50
5.6	resultados da segmentação após a etapa de pós-processamento.	51
5.7	sistema SIRENE.	58
A.1	gestos correspondentes as palavras (a) "eu", (b) "te" e (c) "amo" em LIBRAS.	63

A.2	imagens relativas as letras do alfabeto da LIBRAS representadas por gestos estáticos.	64
A.3	imagens relativas as letras do alfabeto da LIBRAS representadas por gestos dinâmicos. (a)-(d) H , (e)-(h) J , (i)-(l) K , (m)-(p) X , (q)-(t) Y e (u)-(x) Z	65
B.1	resultados das redes neurais (a) MLP e (b) RBF para o sexto <i>cluster</i> da Tabela 5.10.	71
B.2	resultados da rede neural MLP para o sétimo <i>cluster</i> da Tabela 5.10.	72
B.3	resultados das redes neurais (a) MLP e (b) RBF para o oitavo <i>cluster</i> da Tabela 5.10.	74
B.4	resultados das redes neurais (a) MLP e (b) RBF para o nono <i>cluster</i> da Tabela 5.10.	75
B.5	resultados da rede neural MLP para o décimo <i>cluster</i> da Tabela 5.10.	76
B.6	resultados das redes neurais (a) MLP e (b) RBF para o primeiro <i>cluster</i> da Tabela 5.11.	79
B.7	resultados da rede neural RBF para o segundo <i>cluster</i> da Tabela 5.11.	81
B.8	resultados das redes neurais (a) MLP e (b) RBF para o terceiro <i>cluster</i> da Tabela 5.11.	82
B.9	resultados da rede neural RBF para o quarto <i>cluster</i> da Tabela 5.11.	84
B.10	resultados das redes neurais (a) MLP e (b) RBF para o oitavo <i>cluster</i> da Tabela 5.11.	85
B.11	resultados das redes neurais (a) MLP e (b) RBF para o nono <i>cluster</i> da Tabela 5.11.	87
B.12	resultados das redes neurais (a) MLP e (b) RBF para o décimo <i>cluster</i> da Tabela 5.11.	89

Lista de Tabelas

4.1	algoritmos de segmentação da pele.	35
4.2	incidência de classes por <i>cluster</i>	43
4.3	correspondências entre classes e <i>cluster</i>	43
4.4	saídas desejadas para uma rede neural com 5 neurônios de saída.	45
5.1	resultados obtidos com os algoritmos de segmentação.	49
5.2	resultados obtidos com o pós-processamento.	52
5.3	resultados da classificação obtidos para a primeira pessoa.	53
5.4	resultados da classificação obtidos para a segunda pessoa.	53
5.5	resultados da classificação obtidos para a terceira pessoa.	53
5.6	resultados da classificação obtidos para a quarta pessoa.	54
5.7	resultados da classificação obtidos para a quinta pessoa.	54
5.8	resultados da classificação obtidos para a sexta pessoa.	54
5.9	resultados da classificação para cada um dos conjuntos de atributos.	55
5.10	<i>clusters</i> obtidos com a representação pelos Momentos Invariantes de Hu para a segunda pessoa.	56
5.11	<i>clusters</i> obtidos com a representação por cinquenta descritores de Fourier para a quinta pessoa.	57
5.12	<i>clusters</i> obtidos com a representação por cinquenta descritores de Fourier para a sexta pessoa.	57
B.1	matriz de confusão para o quinto <i>cluster</i> da Tabela 5.10 com a rede neural PS.	70

B.2	matriz de confusão para o sexto <i>cluster</i> da Tabela 5.10 com a rede neural PS.	72
B.3	matriz de confusão para o sétimo <i>cluster</i> da Tabela 5.10 com a rede neural PS.	73
B.4	matriz de confusão para o oitavo <i>cluster</i> da Tabela 5.10 com a rede neural PS.	73
B.5	matriz de confusão para o nono <i>cluster</i> da Tabela 5.10 com a rede neural PS.	76
B.6	matriz de confusão para o décimo <i>cluster</i> da Tabela 5.10 com a rede neural PS.	77
B.7	matriz de confusão para o décimo segundo <i>cluster</i> da Tabela 5.10 com a rede neural PS.	77
B.8	tabela de confusão final para os resultados das Tabelas 5.10 e B.1 à B.7.	78
B.9	matriz de confusão para o primeiro <i>cluster</i> da Tabela 5.11 com a rede neural MLP com 48 neurônios ocultos.	80
B.10	matriz de confusão para o segundo <i>cluster</i> da Tabela 5.11 com a rede neural PS.	81
B.11	matriz de confusão para o terceiro <i>cluster</i> da Tabela 5.11 com a rede neural MLP com 35 neurônios ocultos.	83
B.12	matriz de confusão para o quarto <i>cluster</i> da Tabela 5.11 com a rede neural PS.	84
B.13	matriz de confusão para o oitavo <i>cluster</i> da Tabela 5.11 com a rede neural MLP com 48 neurônios ocultos.	86
B.14	matriz de confusão para o nono <i>cluster</i> da Tabela 5.11 com a rede neural MLP com 32 neurônios ocultos.	88
B.15	matriz de confusão para o décimo <i>cluster</i> da Tabela 5.11 com a rede neural MLP com 43 neurônios ocultos.	90
B.16	tabela de confusão final para os resultados das Tabelas 5.11 e B.9 à B.15.	91

Lista de Siglas

ASCE	Associação dos Surdos do Ceará
APADA	Associação de Pais e Amigos de Deficientes Auditivos
FENEIS	Federação Nacional de Educação e Integração de Surdos
HMM	Modelos Ocultos de Markov (<i>Hidden Markov Models</i>)
HSV	Matiz - Saturação - Valor (<i>Hue - Saturation - Value</i>)
HSI	Matiz - Saturação - Intensidade (<i>Hue - Saturation - Intensity</i>)
LIBRAS	Língua de Sinais Brasileira
PCA	Análise das Componentes Principais (<i>Principal Components Analysis</i>)
PDI	Processamento Digital de Imagens
PDS	Processamento Digital de Sinais
PS	<i>Perceptron</i> Simples
RNA	Rede Neural Artificial
RGB	Vermelho - Verde - Azul (<i>Red - Green - Blue</i>)
MLP	<i>Perceptron</i> Multi-Camadas (<i>Multi-Layer Perceptron</i>)
SOM	Mapas Auto-Organizáveis (<i>Self-Organizing Maps</i>)
SL	<i>Software</i> Livre
SIVp	<i>Scilab Image and Video Processing toolbox</i>
YCbCr	Luminância - Crominância azul - Crominância vermelha (<i>Luma - Blue Chrominance - Red Chrominance</i>)

Capítulo 1

Introdução

As línguas de sinais constituem uma forma de linguagem visual e são usadas por portadores de deficiência auditiva e demais pessoas que se relacionam com deficientes auditivos em todo o mundo, como educadores e familiares, sendo adotada em todo o país. Cada nação tem sua própria língua de sinais, como exemplo, a Língua de Sinais Americana (ASL) usada nos Estados Unidos, a Língua de Sinais Quebequiana (LSQ) usada no Canadá, Língua de Sinais Francesa (LSF) usada na França, a Língua Gestual Portuguesa usada em Portugal (LGP) e a Língua de Sinais Brasileira (LIBRAS) usada no Brasil. Algumas destas línguas de sinais apresentam dialetos e regionalismos, semelhantemente às linguagens verbais (JOKINEN, 2006).

No Brasil, órgãos como o Instituto Brasileiro de Geografia e Estatística (IBGE) realizam levantamentos de informações sobre a população com respeito, inclusive, à surdez. Segundo o último censo realizado por este órgão, o Brasil tinha, em 2000 (ano do último censo), 1.066.400 pessoas surdas ou com grande dificuldade permanente de ouvir (IBGE, 2003).

Conforme registrado por Amaral e Coutinho (2002), Capovilla (2008) e Maganhotte Júnior, Faria e Bortolozzi (2002), pessoas portadoras de deficiência auditiva têm grande dificuldade de comunicação com a Língua Portuguesa (escrita). Esta dificuldade impõe barreiras sociais aos deficientes e, muitas vezes, os privam dos direitos básicos dos cidadãos como saúde, lazer e educação. Maganhotte Júnior, Faria e Bortolozzi (2002) afirmam ainda que muitos dos *softwares* de auxílio a deficientes auditivos no Brasil são importados, não atendendo às necessidades

específicas dos usuários brasileiros.

Capovilla (2008) e Capovilla *et al.* (2004) mostram a importância do uso de instrumentos e recursos multimídia na educação e avaliação de deficientes auditivos, e ainda ressaltam a escassez de recursos que atendam às necessidades próprias da comunidade surda brasileira. Guazzelli (2007) e Reichert (2006) apresentam o papel da informática e de outros sistemas multimídia no contexto da educação de deficientes auditivos e mostram como a interação com recursos multimídia pode ajudar na aprendizagem do deficiente auditivo, em todos os campos do conhecimento.

Para suprir as necessidades de tecnologia assistiva à comunidade surda brasileira, estão sendo desenvolvidos sistemas como o SIGUS (PISTORI *et al.*, 2006), o Sign WebMessage (SOUZA; VIEIRA, 2006), o F-LIBRAS (BAPTISTA, 2007) e o Rybená (RYBENÁ, 2009). Todos estes estão sendo desenvolvidos no Brasil e encontram-se, ainda, em fase inicial. O principal objetivo destes sistemas é fornecer subsídios tecnológicos para favorecer a aprendizagem de portadores de deficiência auditiva.

1.1 Motivação

A necessidade de recursos tecnológicos da comunidade surda brasileira associada à inexistência de sistemas funcionais e acessíveis para atender suas necessidades específicas foram os propulsores do desenvolvimento deste trabalho.

Os problemas enfrentados pela comunidade surda, tanto no Brasil como internacionalmente, consistem em dificuldade de aprendizagem de formas de comunicação. Em decorrência das privações sócio-culturais dos surdos, estes têm grandes dificuldades de obtenção e permanência no emprego e aprendizagem de novos ofícios (SILVEIRA, 2006; FRIEDHEIN *et al.*, 2005), o que é agravado pela informatização da sociedade e dos meios de produção. Conseqüentemente, os surdos passam a ser excluídos da sociedade por não poderem se inserir e absorver o processo de desenvolvimento da sociedade (AMARAL; COUTINHO, 2002).

Em decorrência das dificuldades de relacionamento pessoal e inserção social enfrentados diariamente pelos surdos, estes muitas vezes não conseguem ter acesso aos seus direitos básicos. A adoção de uma interface que possibilite a comunicação direta entre o surdo e uma pessoa ouvinte que não tenha conhecimentos sobre a LIBRAS pode viabilizar o acesso dos surdos aos órgãos e instituições que atendam

suas necessidades fundamentais, tais como saúde, lazer e educação.

Desta forma, o sistema proposto neste trabalho, pretende facilitar e motivar a aprendizagem da LIBRAS e facilitar a comunicação entre os surdos e com a sociedade moderna e informatizada, em que a tecnologia encontra-se presente em todos os segmentos, servindo de base para novas tecnologias assistivas.

1.2 Objetivos deste trabalho

O objetivo geral deste trabalho é propor um sistema em *software* livre capaz de reconhecer os gestos relativos ao alfabeto da LIBRAS.

Os objetivos específicos deste trabalho são:

- ▶ revisão dos principais métodos de visão computacional utilizados para interação entre pessoas e computadores;
- ▶ estudo de métodos de segmentação de pele utilizados nos sistemas de visão computacional;
- ▶ estudo de métodos de realce para aumentar a robustez da segmentação em ambientes com iluminação não controlada;
- ▶ estudo de métodos de extração de características da mão segmentada;
- ▶ estudo de métodos de classificação de gestos; e
- ▶ avaliação da capacidade de discriminação de gestos dos diversos métodos estudados.

Através deste trabalho, são avaliados diferentes algoritmos para cada etapa do sistema a fim de determinar os mais indicados para compor o sistema de reconhecimento de gestos proposto.

1.3 Organização da dissertação

Este trabalho está organizado em 6 capítulos, os quais são descritos a seguir.

O Capítulo 2 apresenta e descreve conceitualmente os algoritmos necessários ao entendimento deste trabalho, fundamentando a teoria necessária para a compreensão dos demais capítulos. O Capítulo 3 apresenta o estado da arte sobre reconhecimento

de gestos das mãos, abordando as técnicas e recursos necessários para possibilitar o reconhecimento de gestos. No Capítulo 4, o sistema proposto e o ambiente de desenvolvimento são descritos. E o Capítulo 5 apresenta e discute os principais resultados obtidos neste trabalho. Por fim, no Capítulo 6, são avaliadas as conclusões, contribuições deste trabalho e propostas de perspectivas futuras.

1.4 Produção científica

No período de desenvolvimento deste trabalho foram publicados os seguintes artigos:

- ▶ **Carneiro, A. T. S.**; Cortez, P. C; Costa, R. C. S. Reconhecimento de Gestos da LIBRAS com Classificadores Neurais a partir dos Momentos Invariantes de Hu. Anais da 1a Conferência Regional do Interaction Design Association (IxDA): Interaction '09 | South-America (evento apoiado pela SBC). Páginas 190-195. ISBN: 978-85-63115-00-3.

- ▶ **Carneiro, A. T. S.**; Cortez, P. C; Costa, R. C. S. Correction of high lighting using histogram matching for skin segmentation on white background images. Anais da 1a Conferência Regional do Interaction Design Association (IxDA): Interaction '09 | South-America (evento apoiado pela SBC). Páginas 196-201. ISBN: 978-85-63115-00-3.

Fundamentação teórica

Os sistemas de visão computacional são divididos conceitualmente conforme o diagrama apresentado na Figura 2.1 (GONZALEZ; WOODS, 2008). Cada elemento do diagrama corresponde a uma técnica ou conjunto de técnicas que compõe o sistema completo.

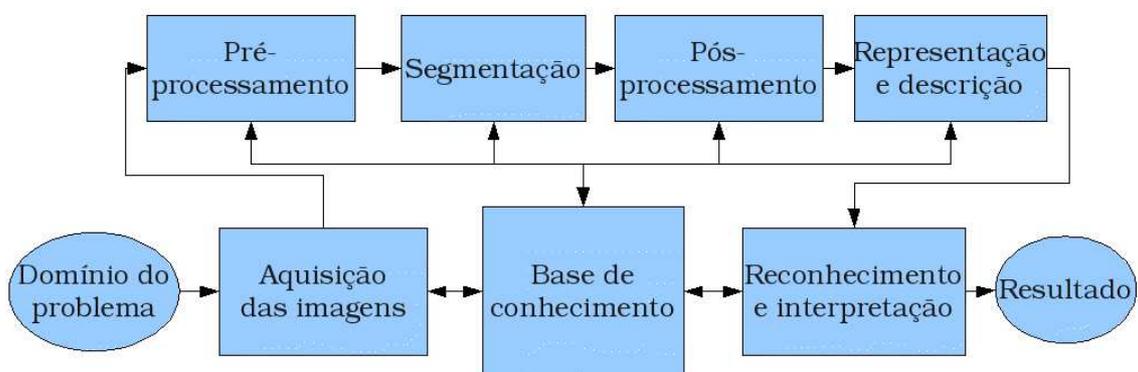


Figura 2.1: diagrama de blocos de um sistema de visão computacional genérico conforme Gonzalez e Woods (2008).

Seguindo o fluxo da informação na Figura 2.1, o elemento **Domínio do problema** representa o contexto no qual o sistema trabalha e a fonte de informação para o sistema. O bloco **Aquisição das imagens** corresponde ao dispositivo físico e a interface de aquisição das imagens que correspondem à entrada do sistema. Os blocos **Pré-processamento**, **Segmentação**, **Pós-processamento**, **Representação e descrição** e **Reconhecimento e interpretação** implementam, respectivamente, os algoritmos que adequam as imagens capturadas às restrições das etapas seguintes, por exemplo algoritmos de realce, o isolamento do objeto ou região

de interesse na imagem restante da imagem, os algoritmos que seguem a segmentação adequando a imagem à etapa seguinte, os algoritmos de extração de características do objeto ou região de interesse na imagem e o algoritmo de reconhecimento de padrões que realiza o processo de classificação com base nas características do bloco anterior. O elemento **Resultado** é a saída final do sistema. O bloco **Base de conhecimento** representa o conhecimento prévio acerca do problema, tais como parâmetros de adequação dos algoritmos e dados provenientes do ambiente em questão previamente processados.

2.1 Pré-processamento

A etapa de pré-processamento consiste em um método de realce baseado na aplicação do algoritmo de especificação de histograma, o qual especifica e aplica uma transformação em cada *pixel* da imagem original para modificar seu histograma $p_r(r_k)$, de modo que este passe a ter uma forma desejada, previamente determinada, $p_z(z_k)$. Considerando uma imagem em níveis de cinza, em que o valor de cada pixel está no intervalo $[0, L - 1]$, em que L corresponde ao total de níveis de cinza, o histograma $p_r(r_k)$ de uma imagem é definido por

$$p_r(r_k) = \frac{n(r_k)}{N}, \quad (2.1)$$

em que N é o total de *pixels* da imagem, r_k é o k -ésimo nível de cinza e $n(r_k)$ corresponde ao número de *pixels* que possuem este nível de cinza.

O primeiro passo do algoritmo de especificação de histograma é a aplicação da transformação $T(\cdot)$ dada por (GONZALEZ; WOODS, 2008)

$$T(r) = \int_0^r p_r(w)dw, \quad (2.2)$$

em que w é uma variável muda. O segundo passo consiste em tomar o histograma desejado $p_z(z)$, e aplicar a transformação $G(\cdot)$:

$$G(z) = \int_0^z p_z(t)dt, \quad (2.3)$$

em que t é uma variável muda. Fazendo $G(z) = T(r)$, obtém-se que $z = G^{-1}[T(r)]$,

ou seja, gera-se uma nova imagem, cujo histograma é $p_z(z_k)$.

Para ilustrar o funcionamento do algoritmo de especificação de histograma, a Figura 2.2 apresenta um histograma original $p_r(r)$, um histograma desejado $p_z(z)$ e o histograma obtido que, para o exemplo apresentado, é uma aproximação de $p_z(z)$.

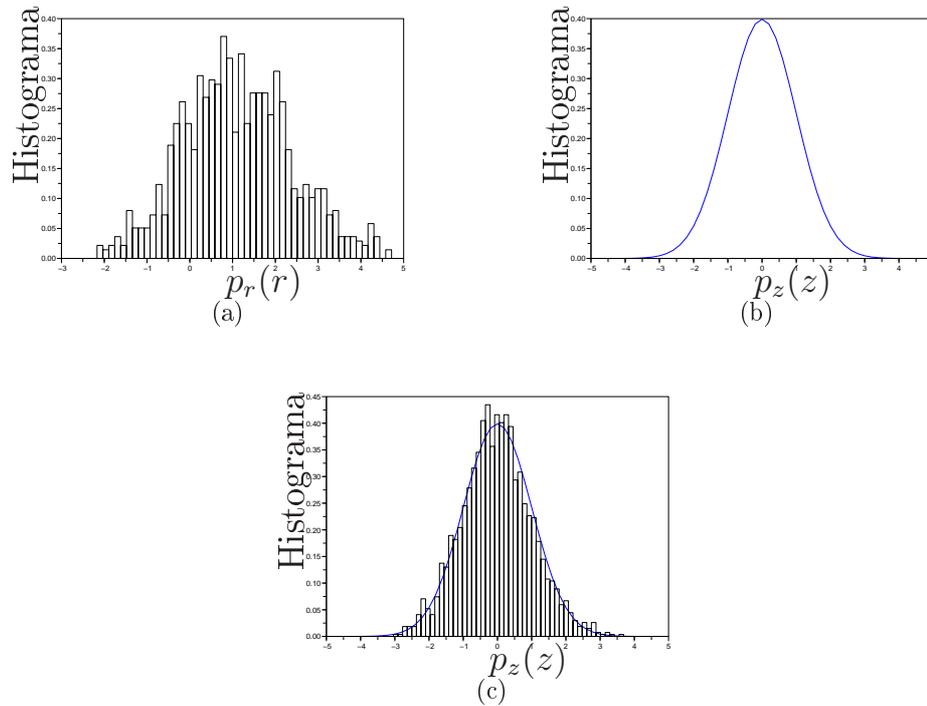


Figura 2.2: (a) histograma original, (b) histograma desejado e (c) histograma obtido.

Como o método é formulado no espaço contínuo, a implementação em computação discreta fornece resultados aproximados.

2.2 Segmentação

A segmentação está dentre as primeiras etapas dos sistemas de reconhecimento de gestos baseados em visão computacional, podendo ou não ser precedida de algum pré-processamento. Os algoritmos de segmentação de pele utilizados neste trabalho são selecionados com base na eficiência e na complexidade computacional.

Oliveira *et al.* (2006) descrevem os benefícios da representação colorida para a imagem a fim de que a segmentação forneça os melhores resultados, pois, as imagens coloridas contém mais informações do que imagens em níveis de cinza ou monocromáticas. Dentre os trabalhos disponíveis na literatura, a grande maioria

destes realiza a segmentação da pele apenas em imagens coloridas. Uma exceção é o trabalho de Zahedi (2007), que apresenta um algoritmo para segmentar a pele em imagens em níveis de cinza e outro para imagens coloridas.

Dentre os diversos algoritmos de segmentação da pele para imagens coloridas, destacam-se por uma segmentação eficiente associada a um baixo custo computacional, os métodos de limiarização e classificação por modelo de probabilidade. Cada um dos métodos de segmentação depende da codificação de cor da imagem, podendo ser Vermelho-Verde-Azul (RGB), Luminância-Crominância azul-Crominância vermelha (YCbCr) e Matiz-Saturação-Valor (HSV), ou Matiz-Saturação-Intensidade (HSI).

O dispositivo de aquisição de imagens usado captura as imagens codificadas em RGB em que cada *pixel* da imagem é representado por um vetor de três componentes $I(x, y) = [R, G, B]^T$, a partir deste padrão é possível recodificar a imagem para os padrões YCbCr $I(x, y) = [Y, Cb, Cr]^T$ e HSV $I(x, y) = [H, S, V]^T$ usando, respectivamente, (GONZALEZ; WOODS, 2008)

$$\begin{aligned} Y &= 0,299R + 0,587G + 0,114B, \\ Cb &= 0,564(B - Y), \\ Cr &= 0,713(R - Y), \end{aligned} \tag{2.4}$$

e

$$\begin{aligned} V &= \max \{R, G, B\}, \\ S &= \begin{cases} \frac{V - \min\{R, G, B\}}{V}, & \text{se } V \neq 0, \\ 0, & \text{se } V = 0, \end{cases} \\ H &= \begin{cases} \frac{G - B}{6S}, & \text{se } V = R, \\ \frac{1}{3} + \frac{B - R}{6S}, & \text{se } V = G, \\ \frac{2}{3} + \frac{R - G}{6S}, & \text{se } V = B. \end{cases} \end{aligned} \tag{2.5}$$

O processo de segmentação consiste em subdividir uma imagem em objetos e/ou regiões que a constituem, a segmentação é finalizada quando o objeto ou região de interesse está devidamente isolada do restante da imagem (GONZALEZ; WOODS, 2008).

2.2.1 Segmentação por limiarização

Conforme Gonzalez e Woods (2008), a limiarização é a técnica de segmentação mais difundida em aplicações de processamento digital de imagens. O processo de segmentação por limiarização isola os objetos ou regiões de interesse rotulando cada *pixel* da imagem $I(x, y)$ de acordo com o seu valor, conforme apresentado a seguir

$$L(x, y) = \begin{cases} 1, & \text{se } I(x, y) \in R_d, \\ 0, & \text{se } I(x, y) \notin R_d, \end{cases} \quad (2.6)$$

em que $L(x, y)$ é o rótulo correspondente ao *pixel* $I(x, y)$ e R_d denota o conjunto de valores de *pixels* correspondentes ao objeto ou região de interesse. Usualmente, o conjunto R_d consiste de uma região determinada por limiares, por isto este algoritmo é denominado segmentação por limiarização.

2.2.2 Segmentação por modelo de probabilidade

Segundo Fukunaga (1990), o processo de segmentação por modelo de probabilidade consiste de duas etapas: estimação da probabilidade de cada *pixel*, com relação a um modelo estatístico específico $I_p(x, y) = \text{prob}(I(x, y))$, seguida da seguinte limiarização

$$L(x, y) = \begin{cases} 1, & \text{se } I_p(x, y) \geq P, \\ 0, & \text{se } I_p(x, y) < P, \end{cases} \quad (2.7)$$

sendo $I_p(x, y)$ a probabilidade do *pixel* $I(x, y)$ pertencer ao objeto ou região de interesse e P é a probabilidade mínima para que um *pixel* seja rotulado como pertencente ao objeto ou região de interesse.

2.3 Pós-processamento

Após a segmentação, a imagem é submetida a etapa de pós-processamento, o qual consiste em algoritmos e técnicas para remoção de erros provenientes do processo de segmentação e adequação da imagem para a etapa seguinte, que é a etapa de extração de atributos.

2.3.1 Crescimento de região

O algoritmo de crescimento de região é usado para corrigir as falhas decorrentes do processo de segmentação. Este consiste em agrupar *pixels* ou pequenas regiões em regiões maiores baseando-se em um critério pré-definido (GONZALEZ; WOODS, 2008).

O algoritmo consiste em, dado um ponto inicial, chamado semente, pertencente à região de interesse, os pontos vizinhos a este são analisados para a confirmação do critério. Este processo é repetido até que a região de interesse esteja completamente definida.

2.3.2 Detecção de bordas

A detecção de bordas em imagens digitais consiste em aplicar uma técnica de realce de detalhes, por exemplo o gradiente Sobel. O gradiente de Sobel gera duas imagens, em que cada uma apresenta as componentes horizontais e verticais da imagem original, então utiliza-se a imagem de magnitude do gradiente. A técnica de gradiente consiste em convolver a imagem original com as matrizes

$$\mathbf{G}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ e } \mathbf{G}_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \quad (2.8)$$

em que a convolução da imagem original com \mathbf{G}_x e \mathbf{G}_y resulta, respectivamente, em \mathbf{I}_x e \mathbf{I}_y . A imagem de magnitude do gradiente é dada pela soma matricial $I_G(x, y) = \|I_x(x, y)\| + \|I_y(x, y)\|$ (GONZALEZ; WOODS, 2008).

2.3.3 Morfologia matemática

A morfologia matemática é uma área do processamento digital de imagens que aborda a extração de informações de imagens a partir de transformações de forma (FELIX, 2007). A morfologia matemática é baseada em dois operadores básicos: erosão e dilatação. Além dos operadores básicos, outro elemento que define a transformação morfológica é o elemento estruturante, o qual caracteriza-se pela sua forma geométrica (GONZALEZ; WOODS, 2008; FELIX, 2007).

Segundo Gonzalez e Woods (2008), as operações morfológicas de erosão

(operador \ominus) e dilatação (operador \oplus) entre uma imagem A e um elemento estruturante B são definidas, respectivamente, por

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad \text{e} \quad A \oplus B = \left\{z | \left[(\hat{B})_z \cap A \right] \subseteq A \right\}, \quad (2.9)$$

como A e B são conjuntos de \mathbb{Z}^2 , z é um elemento genérico de \mathbb{Z}^2 , então $(B)_z$ denota o elemento B transladado de z e $(\hat{B})_z$ denota o elemento B refletido sobre sua origem transladado de z .

2.4 Extração de atributos

Após o pós-processamento, a próxima etapa do sistema de visão computacional é a extração dos descritores de forma. Neste trabalho são utilizados dois conjuntos distintos de atributos, os Momentos Invariantes de Hu e os Descritores de Fourier os quais são descritores geométricos de formas em imagens (GONZALEZ; WOODS, 2008; CONSEIL; BOURENNANE; MARTIN, 2007).

2.4.1 Momentos Invariantes de Hu

Hu (1962) apresenta o método dos Momentos Invariantes, que consistem em um método de análise de formas para reconhecimento visual de padrões. Estes momentos são formados por um conjunto de parâmetros geométricos de objetos em imagens, os quais são independentes a distância, translação e rotação¹ do objeto.

O momento centrado de ordem pq é definido por (HU, 1962)

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy, \quad (2.10)$$

em que $[\bar{x} \ \bar{y}]^T$ corresponde ao centro de massa da imagem $f(x, y)$ e são definidos por

$$\bar{x} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy} \quad \text{e} \quad \bar{y} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy}. \quad (2.11)$$

Os momentos invariantes de Hu são dados por

¹Os Momentos Invariantes de Hu são independentes quanto à rotação em torno de qualquer eixo perpendicular ao plano da imagem.

$$\begin{aligned}
\phi_1 &= m_{20} + m_{02}, \\
\phi_2 &= (m_{02} - m_{20})^2 + 4m_{11}^2, \\
\phi_3 &= (m_{30} - 3m_{12})^2 + (m_{03} - 3m_{21})^2, \\
\phi_4 &= (m_{30} + m_{12})^2 + (m_{03} + m_{21})^2, \\
\phi_5 &= (m_{30} - 3m_{12})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{03} + m_{21})^2) \\
&\quad + (m_{03} - 3m_{21})(m_{03} + m_{21})((m_{03} + m_{21})^2 - 3(m_{30} + m_{12})^2), \\
\phi_6 &= (m_{20} - m_{02})((m_{30} + m_{12})^2 - (m_{03} + m_{21})^2) \\
&\quad + 4m_{11}(m_{30} + m_{12})(m_{03} + m_{21}),
\end{aligned} \tag{2.12}$$

em que $m_{pq} = \frac{\mu_{pq}}{\mu_{00}}$, corresponde ao momento centrado de ordem pq normalizado pelo momento centrado de ordem nula. Esta normalização concede a invariância dos momentos de Hu à distância.

Com base nestes momentos, pode-se definir um vetor de características $\underline{x} = [\phi_1 \ \phi_2 \ \phi_3 \ \phi_4 \ \phi_5 \ \phi_6]^T$ utilizado como conjunto de informações sobre a geometria da mão.

2.4.2 Descritores de Fourier

Outro conjunto de atributos para descrição de formas são os descritores de Fourier. Conseil, Bourennane e Martin (2007) e Zhang e Lu (2001) definem os descritores de Fourier a partir das assinaturas de forma. Estas, por sua vez, são obtidas a partir dos pontos de contorno do objeto representados em um plano de coordenadas complexas

$$z_m = x_m + jy_m, \tag{2.13}$$

em que (x_m, y_m) correspondem às coordenadas do m -ésimo ponto sobre o contorno da imagem e $j = \sqrt{-1}$. Segundo os autores, a Equação (2.13) apresenta melhores resultados quando modificada para retirar o *offset* do objeto coincidindo assim o centro deste com o ponto de origem do plano, fazendo

$$z_m = (x_m - x_C) + j(y_m - y_C), \tag{2.14}$$

em que

$$x_C = \frac{1}{M} \sum_{m=0}^{M-1} x_m \quad \text{e} \quad y_C = \frac{1}{M} \sum_{m=0}^{M-1} y_m, \quad (2.15)$$

sendo M o número total de pontos do contorno.

Zhang e Lu (2001) apresentam em seu trabalho três assinaturas de forma: distância ao centróide, assinatura da curvatura e função angular cumulativa. A assinatura de forma com base na distância ao centróide r_m é definida por

$$r_m = ((x_m - x_C)^2 + (y_m - y_C)^2)^{\frac{1}{2}}. \quad (2.16)$$

A assinatura de forma baseada na curvatura K_m é definida por

$$K_m = \theta_m - \theta_{m-1}, \quad (2.17)$$

em que

$$\theta_m = \arctg \left(\frac{y_m - y_{m-w}}{x_m - x_{m-w}} \right), \quad (2.18)$$

para um determinado intervalo w , ou ainda por

$$K_m = \phi_m - \phi_{m-1}, \quad (2.19)$$

em que ϕ_m é a assinatura de forma função angular cumulativa, definida por

$$\phi_m = [\theta_m - \theta_0] \bmod \{2\pi\}. \quad (2.20)$$

Os Descritores de Fourier são definidos como a Transformada Discreta de Fourier de uma das assinaturas de forma, representada genericamente por s_m (GONZALEZ; WOODS, 2008)

$$u_n = \frac{1}{M_s} \left| \sum_{m=0}^{M_s-1} s_m e^{\frac{-j2\pi nm}{M_s}} \right|, \quad (2.21)$$

observa-se que o limite superior do somatório é M_s , este valor representa o número total de pontos do contorno após uma sub-amostragem de s_m .

2.4.3 Componentes de velocidade

Além dos atributos geométricos, Momentos de Hu e Descritores de Fourier, também são consideradas informações sobre a velocidade da mão. As componentes de velocidade são calculadas a partir das coordenadas do centro da imagem, determinadas durante o cálculo dos atributos geométricos nas equações (2.11) e (2.15). Por se tratarem de elementos em um tempo discreto, as componentes de velocidade são aproximadas por uma diferenciação discreta (OPPENHEIM; SCHAFER; BUCK, 1999).

Para os Momentos Invariantes de Hu, as componentes de velocidade da mão para o i -ésimo *frame* (v_x, v_y) são dadas por

$$v_x = \frac{d\bar{x}}{dt} \approx \bar{x}_i - \bar{x}_{i-1} \quad \text{e} \quad v_y = \frac{d\bar{y}}{dt} \approx \bar{y}_i - \bar{y}_{i-1}, \quad (2.22)$$

em que (\bar{x}_i, \bar{y}_i) correspondem às coordenadas do centro de massa no i -ésimo *frame*, obtidos pela Equação (2.11).

Para os Descritores de Fourier, as componentes de velocidade da mão para o i -ésimo *frame* são dadas por

$$v_x = \frac{dx_C}{dt} \approx x_{C(i)} - x_{C(i-1)} \quad \text{e} \quad v_y = \frac{dy_C}{dt} \approx y_{C(i)} - y_{C(i-1)}, \quad (2.23)$$

em que $(x_{C(i)}, y_{C(i)})$ correspondem as coordenadas do centro do contorno no i -ésimo *frame*, obtidos pela Equação (2.15).

2.5 Classificação

De posse do vetor de atributos para cada imagem, composto pelos Momentos Invariantes de Hu ou pelos Descritores de Fourier, a última categoria de algoritmos a serem apresentados são os classificadores. Neste trabalho, são utilizados classificadores baseados em algoritmos de Redes Neurais Artificiais (RNA).

As RNAs são algoritmos implementados com base em elementos de processamento simples (neurônios artificiais) interligados, com capacidade de aprendizagem a partir dos dados. A literatura apresenta duas categorias de redes neurais: as redes de aprendizagem supervisionada e as redes de aprendizagem não-supervisionada (HAYKIN, 1999; MARQUES, 2005).

O modelo de neurônio artificial consiste em uma combinação linear das entradas \underline{x} , seguida de uma função de ativação $\phi(\cdot)$, conforme ilustrado na Figura 2.3.

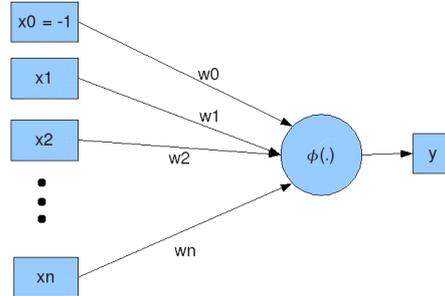


Figura 2.3: modelo de neurônio artificial.

As RNAs de aprendizagem supervisionada caracterizam-se pela aprendizagem baseada em tentativa e erro, em que existe um supervisor que informa a rede neural se esta forneceu um resultado correto ou errado, enquanto que as RNAs de aprendizagem não-supervisionada não apresentam a figura do supervisor durante a aprendizagem (MARQUES, 2005).

Este trabalho implementa as duas abordagens de classificação. Na primeira categoria, destacam-se os algoritmos *Perceptron Simples* (PS), *Perceptron Multi Camada* (MLP) e Funções de Base Radial (RBF), enquanto que a segunda utiliza o algoritmo de Mapas Auto-Organizáveis (SOM).

2.5.1 Mapas Auto-Organizáveis

Os Mapas Auto-Organizáveis são redes neurais de aprendizagem não-supervisionada usados em situações que envolvem quantização vetorial ou análise de agrupamentos (*clustering*). As redes SOM são compostas por neurônios, identificados por um vetor de pesos, sendo o i -ésimo neurônio \underline{m}_i definido por $\underline{m}_i = [\mu_{i1} \mu_{i2} \dots \mu_{in}]^T \in \mathbb{R}^n$. Nestes mapas, os neurônios são usualmente dispostos em uma grade de uma ou duas dimensões (KOHONEN, 2001).

A aprendizagem, ou treinamento, da rede SOM acontece em dois passos principais para cada padrão de entrada, dado genericamente por $\underline{x} = [\phi_1 \phi_2 \dots \phi_n]^T \in \mathbb{R}^n$.

O primeiro passo é determinar o neurônio mais semelhante ao padrão de entrada \underline{x} , chamado de protótipo vencedor representado por \underline{m}_c (KOHONEN, 2001; HERTZ;

PALMER; KROGH, 1998)

$$c = \underset{\forall i}{\operatorname{arg\,min}} \{ \|\underline{x} - \underline{m}_i\| \}, \quad (2.24)$$

sendo $\|\cdot\|$ a norma euclidiana de um vetor, de modo que

$$\|\underline{x} - \underline{m}_c\| = \min_{\forall i} \{ \|\underline{x} - \underline{m}_i\| \}. \quad (2.25)$$

Uma vez determinado o neurônio \underline{m}_c , o segundo passo do treinamento da rede SOM consiste em atualizar os pesos \underline{m}_i de todos os neurônios da rede com base na seguinte equação

$$\underline{m}_i(t+1) = \underline{m}_i(t) + h_{ci}(t) (\underline{x}(t) - \underline{m}_i(t)), \quad (2.26)$$

em que $\underline{m}_i(t)$ é o vetor de pesos no instante t , $\underline{m}_i(t+1)$ é o vetor de pesos no instante $t+1$ e $h_{ci}(t)$ é a função vizinhança dada por

$$h_{ci}(t) = \alpha(t) e^{\left(-\frac{\|\underline{r}_c - \underline{r}_i\|^2}{2\sigma^2(t)} \right)}, \quad (2.27)$$

em que os parâmetros $\alpha(t)$ e $\sigma(t)$ são funções monotonicamente decrescentes que representam, respectivamente, a taxa de aprendizagem do algoritmo e o tamanho da vizinhança do protótipo vencedor no instante t e r_i representa a posição do i -ésimo neurônio na rede neural. Este assume valores escalares para redes unidimensionais ou vetoriais para redes com mais de uma dimensão (KOHONEN, 2001).

O algoritmo de aprendizagem é repetido até que uma condição de parada seja estabelecida, por exemplo um número fixo de épocas ou o erro médio quadrático (KOHONEN, 2001; HERTZ; PALMER; KROGH, 1998).

2.5.2 *Perceptron* Simples

As redes neurais do tipo *Perceptron* Simples (PS) são RNAs de aprendizagem supervisionada, formadas por uma única camada de neurônios artificiais² conforme apresentado na Figura 2.4 (HAYKIN, 1999; MARQUES, 2005; HERTZ; PALMER; KROGH, 1998). Nesta Figura, $\{x_0, x_1, \dots, x_n\}$ correspondem as entradas, $\phi(\cdot)$ é a

²Usualmente adota-se o modelo de neurônios proposto por McCulloch e Pitts em 1942 (MARQUES, 2005).

função de ativação e $\{y_1, y_2, \dots, y_m\}$ são a saída da rede neural.

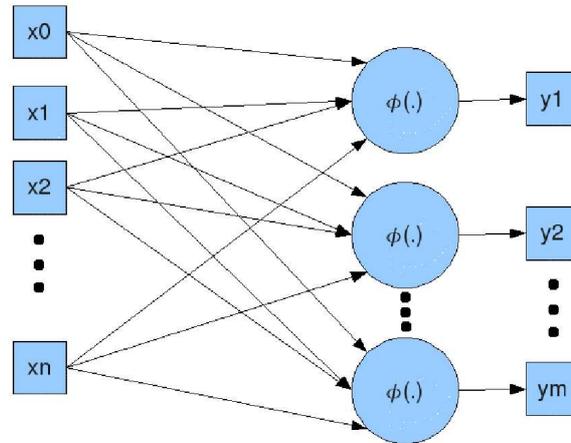


Figura 2.4: arquitetura de rede neural do tipo PS.

A saída do j -ésimo neurônio y_j é dada, em função do vetor de entrada $\underline{x} = [x_0 = -1, x_1, x_2, \dots, x_n]^T$ e do vetor de pesos $\underline{w}_j = [w_{j0}, w_{j1}, \dots, w_{jn}]^T$, pela seguinte equação (HAYKIN, 1999; MARQUES, 2005)

$$y_j = \phi(\underline{w}_j^T \underline{x}). \quad (2.28)$$

Uma das funções de ativação $\phi(\cdot)$ mais comuns literatura, a qual é adotada neste trabalho, é a tangente hiperbólica, dada por

$$\phi(a) = \frac{1 - e^{-a}}{1 + e^{-a}}, \quad (2.29)$$

para uma variável qualquer a .

Após definir a quantidade de neurônios e a função de ativação, a rede PS encontra-se apta a ser treinada. O treinamento da rede PS segue os seguintes passos, os quais são aplicados a cada neurônio de forma independente (MARQUES, 2005)

- i. no instante $t = 0$, antes da primeira iteração, o vetor \underline{w}_j é inicializado com valores aleatórios;
- ii. apresenta-se uma entrada \underline{x} , cuja saída desejada para a rede PS seja conhecida (d_j);
- iii. computa-se a saída y_j fornecida pela rede neural, conforme a Equação (2.28);

iv. atualizar o vetor \underline{w}_j segundo a equação

$$\underline{w}_j(t) = \underline{w}_j(t-1) + \Delta \underline{w}_j, \quad (2.30)$$

em que $\Delta \underline{w}_j(t) = \alpha \underline{x}(t)e(t)$; e α corresponde ao ganho de adaptação ou passo de aprendizagem e $e_j = d_j - x_j$ corresponde ao erro de saída;

v. retornar ao segundo passo até que uma condição de parada seja obtida.

As redes PS são de simples implementação, porém sua capacidade de resolver problemas limita-se a situações cujos dados são linearmente separáveis, ou seja, a classificação pode ser feita mediante superfícies de decisão lineares (hiperplanos) (HAYKIN, 1999; MARQUES, 2005; HERTZ; PALMER; KROGH, 1998). As redes MLP são uma extensão das redes PS, as quais são capazes de resolver problemas não-linearmente separáveis.

2.5.3 *Perceptron* Multi Camada

As redes neurais *Perceptron* Multi Camada (MLP) representam uma generalização das redes PS, sendo que uma rede MLP é constituída de duas ou mais camadas de redes PS, conforme apresentado na Figura 2.5, em que a saída de uma camada é a entrada da camada seguinte³ (HAYKIN, 1999).

O método de treinamento de uma rede neural MLP mais conhecido é o algoritmo de retropropagação, o qual consiste em propagar o erro da rede neural no sentido inverso ao fluxo de dados para alterar os pesos de cada neurônio. Assim, o erro percorre o sentido da saída para a entrada da rede durante o treinamento. O processo de atualização dos pesos dos neurônios para uma rede MLP é semelhante à rede PS, em que as diferenças aparecem na propagação do sinal de entrada. Neste caso, a saída de uma camada é a entrada da camada seguinte, como apresentado anteriormente, e o termo $\Delta \underline{w}_j$, para o caso da rede MLP, é calculado para cada elemento Δw_{ji} através de (HAYKIN, 1999; MARQUES, 2005)

$$\Delta w_{ji} = \alpha e_j \dot{\phi}(\nu_j) y_i, \quad (2.31)$$

em que $\dot{\phi}(\nu_j)$ é a derivada da função de ativação em relação a ν_j e

³As camadas anteriores à camada de saída são chamadas camadas ocultas ou camadas escondidas.

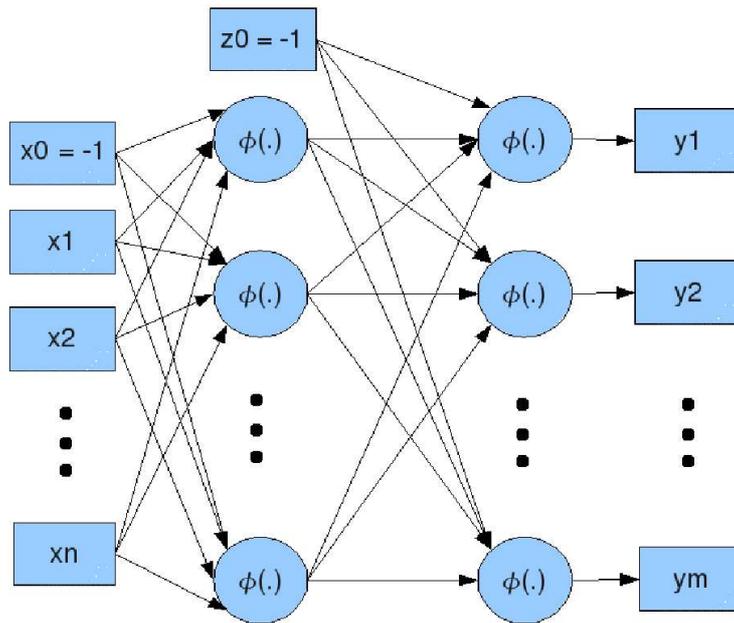


Figura 2.5: modelo da arquitetura multi camadas da rede MLP, com uma camada oculta.

$$\nu_j = \sum_{i=0}^p w_{ji} y_i, \quad (2.32)$$

sendo p o total de neurônios na camada oculta. Haykin (1999) descreve detalhadamente como obter o resultado $\Delta w_{ji} = \alpha e_j \dot{\phi}(\nu_j) y_i$ para todos os neurônios da rede MLP.

2.5.4 Funções de Base Radial

As redes de Funções de Base Radial têm uma arquitetura semelhante às redes MLP de uma camada oculta, conforme apresentado na Figura 2.6 (HAYKIN, 1999; MARQUES, 2005; HERTZ; PALMER; KROGH, 1998).

Uma rede RBF é constituída por uma camada escondida, com função de ativação local em \mathbb{R}^n , ligada a uma camada de saída. A saída da i -ésima unidade da camada escondida, representada por z_i , é definida por

$$z_i = B_i(\underline{x}), \quad (2.33)$$

em que

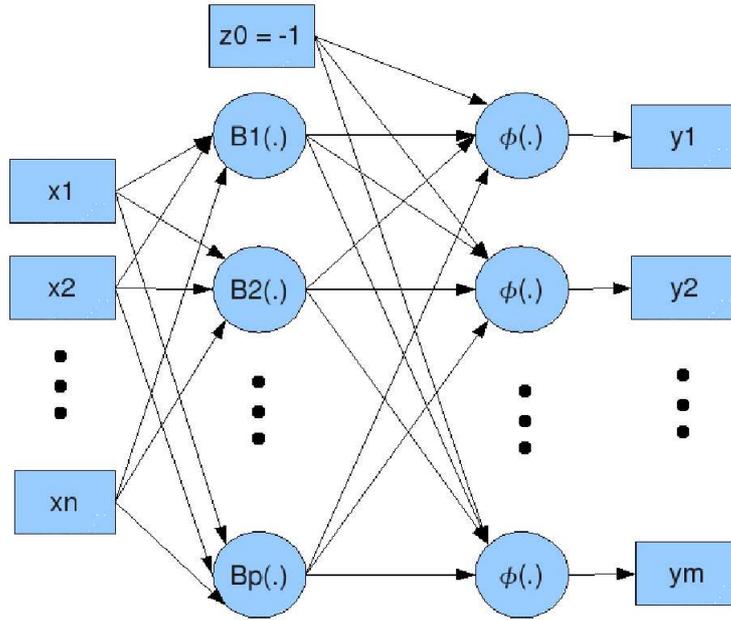


Figura 2.6: modelo de arquitetura da rede RBF.

$$B_i(\underline{x}) = \frac{1}{(2\pi)^{n/2} \sigma_i^n} e^{-\frac{\|\underline{x} - \underline{\mu}_i\|^2}{2\sigma_i^2}}, \quad (2.34)$$

nos quais $\underline{\mu}_i$ é o vetor de pesos da i -ésima unidade da camada escondida, σ_i a variância da função de base radial e n é o número de dimensões de \underline{x} . A estimação de σ_i^2 é feita pelo método de máxima verossimilhança, o qual é dado por (MARQUES, 2005)

$$\sigma_i^2 = \frac{1}{N_i} \sum_{x \in X_i} \|x - \mu_i\|^2, \quad (2.35)$$

em que X_i é o conjunto de todos os elementos x tais que

$$\|\underline{x} - \underline{\mu}_i\| = \min_{\forall k \leq p} \left\{ \|\underline{x} - \underline{\mu}_k\| \right\}, \quad (2.36)$$

sendo N_i o número de elementos de X_i e $k \in \mathbb{N}$.

Os valores de $\underline{\mu}_i$ são determinados como os centróides de um algoritmo de *clustering* (KOHONEN, 2001; MARQUES, 2005).

2.6 Resumo do capítulo

Neste capítulo são apresentados as técnicas e os algoritmos utilizados em cada uma das etapas do sistema proposto neste trabalho.

O conhecimento teórico acerca dos algoritmos é necessário ao entendimento deste trabalho e a partir do conhecimento teórico apresentado, é possível compreender além da proposta deste trabalho, também os demais trabalhos correlatos a este, os quais são apresentados no capítulo seguinte.

Capítulo 3

Estado da arte em reconhecimento de gestos manuais

Os sistemas de reconhecimento de gestos das mãos dividem-se em dois grupos principais: baseados em sensores mecânicos (BEDREGAL; DIMURO; COSTA, 2007; GABARDO, 2005) e baseados em visão computacional (STARNER; WEAVER; PENTLAND, 1996; ZAHEDI *et al.*, 2006; ZAHEDI, 2007; COOPER; BOWDEN, 2007; RIBEIRO, 2006). Existem ainda sistemas de visão que fazem uso de recursos, tais como luvas com indicadores, para facilitar o reconhecimento dos gestos (LAMAR; BHUIYAN; IWATA, 2003).

Na próxima seção são apresentados trabalhos que utilizam sensores mecânicos e na Seção 3.2 são apresentados trabalhos que utilizam visão computacional.

3.1 Sistemas de reconhecimento de gestos baseados em sensores mecânicos

Os sistemas de reconhecimento de gestos baseados em sensores mecânicos usam luvas com sensores inteligentes como a CyberGlove^{®1} (IMMERSION, 2007), apresentada na Figura 3.1. Este dispositivo consiste em uma luva com 18 ou 22 sensores de posição e movimento angulares com precisão de 0,5 grau, e comunica-se com computadores pessoais via Bluetooth[®]. A CyberGlove[®] é acompanhada de um *software* que implementa as interfaces para o usuário.

¹As informações sobre a CyberGlove[®] estão disponíveis para acesso no endereço Web: <http://www.cyberglovesystems.com/products/cyberglove-ii/overview>



Figura 3.1: Cyberglove[®] em uso.

Por se tratar de um sistema de *hardware* e *software* proprietários, a CyberGlove[®] é dependente da plataforma do sistema, ou seja, o usuário não tem a liberdade de usar o equipamento em um sistema operacional diferente do qual é especificado pelo fabricante. Além disto, o preço elevado impossibilita o acesso à maior parte da população. O valor estimado² do sistema completo, de um conjunto de *hardware* e *software* semelhante e equivalente, para reconhecimento de gestos é US\$ 18.300,00, sendo US\$ 14.800,00 para o *hardware* e US\$ 3.500,00 para o *software* próprio de reconhecimento de gestos.

3.2 Sistemas de reconhecimento de gestos baseados em visão computacional

Comparados aos sistemas baseados em sensores mecânicos, os sistemas de visão computacional têm um custo reduzido, o *hardware* necessário, que consiste de uma câmera³ ligada a um computador, é facilmente encontrado por valores de até R\$ 90,00 e *software* pode ser obtido sem custo para o usuário final, como em sistemas *freeware*, tornando-os mais acessíveis que a categoria de sistemas apresentados na seção anterior. Uma grande vantagem de aplicações de visão computacional é a possibilidade de ser desenvolvido independente da plataforma, por se tratar de um *hardware* mais difundido e acessível, enquanto que os sistemas baseados em sensores mecânicos são desenvolvidos para uma plataforma específica, assim, os sistemas de visão computacional fornecem ao usuário a liberdade de escolher qual ambiente de *hardware* e *software*, em particular o sistema operacional, deseja usar.

²Valores retirados do *web site* (HITLab, Universidade de Washington): <http://www.hitl.washington.edu/research/knowledgebase/virtual-worlds/haptics-faq.html> acessado em 29 de abril de 2010.

³Como o computador é um equipamento comum aos dois sistemas, o seu custo não é considerado para fins de comparação.

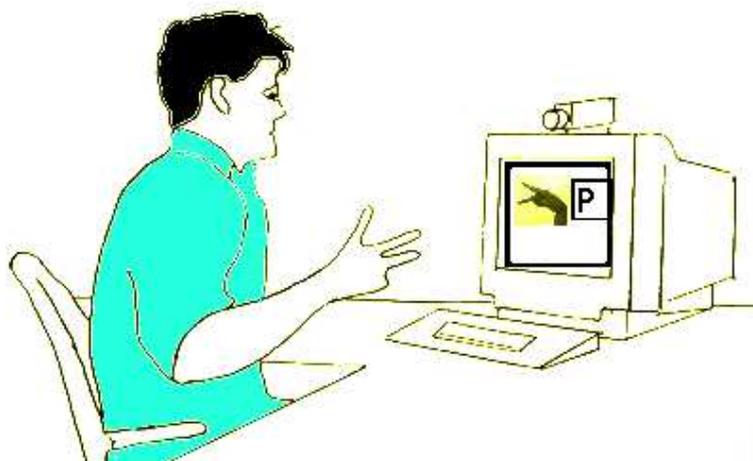


Figura 3.2: exemplo de um sistema de reconhecimento de gestos baseado em visão computacional.

A liberdade de escolha do sistema operacional utilizado é muito importante, pois, para certas pessoas, o custo impossibilita sua utilização. A maioria da população surda possui dificuldades em obter fontes de renda em decorrência de suas limitações.

Dentre os trabalhos disponíveis na literatura e que abordam a temática de reconhecimento de gestos, o que considera de forma mais transparente a problemática de custo e acesso dos usuários ao sistema é o SIGUS (PISTORI *et al.*, 2006). O SIGUS consiste em um sistema integrado de *software* que dispõe de recursos para desenvolver aplicações de visão computacional. Souza e Pistori (2005), Dias, Souza e Pistori (2006) e Souza, Dias e Pistori (2007) apresentam aplicações de análise e reconhecimento da LIBRAS desenvolvida no SIGUS.

3.2.1 Algoritmos presentes na literatura especializada

Os trabalhos científicos de reconhecimento de gestos produzidos durante o desenvolvimento do sistema SIGUS (DIAS; SOUZA; PISTORI, 2006; SOUZA; DIAS; PISTORI, 2007), e a maioria dos trabalhos disponíveis na literatura (VOGLER; METAXAS, 1999; CAMPR *et al.*, 2008; TRUYENQUE, 2005; ZAHEDI, 2007; STARNER; WEAVER; PENTLAND, 1996; WU, 2001), utilizam os Modelos Ocultos de Markov (HMM) para realizar o reconhecimento dos gestos (PAPOULIS; PILLAI, 2002). Por outro lado, outros trabalhos utilizam outros algoritmos de reconhecimento de padrões, como redes neurais artificiais (RNA) (HANDOUYAYA; ZIOU; WANG, 1999; LAMAR; BHUIYAN; IWATA,

2003), algoritmos estatísticos baseados em distribuições de múltiplas gaussianas (RIBEIRO, 2006; RIBEIRO; GONZAGA, 2006) e distância bayesiana (CONSEIL; BOURENNANE; MARTIN, 2007).

Os principais descritores utilizados como entrada dos classificadores são os Momentos Invariantes de Hu (ZAHEDI, 2007; RIBEIRO, 2006; CONSEIL; BOURENNANE; MARTIN, 2007; CAMPR *et al.*, 2008; SOUZA; PISTORI, 2005; ZAHEDI *et al.*, 2006), e alguns autores usam os Descritores de Fourier (CONSEIL; BOURENNANE; MARTIN, 2007; CAMPR *et al.*, 2008; HANDOUYAYA; ZIOU; WANG, 1999) para comparar o desempenho dos sistemas desenvolvidos. Existem ainda outros descritores selecionados a partir da análise das componentes principais (PCA) (BOWDEN; SARHADI, 2000; WU, 2001) e baseados em funções de momento (HANDOUYAYA; ZIOU; WANG, 1999).

Os Momentos Invariantes de Hu são calculados após a segmentação da pele na imagem e os descritores de Fourier são calculados após a obtenção do contorno da mão, o qual é obtido pela detecção das bordas da imagem segmentada. A literatura apresenta diversos algoritmos de segmentação de pele, sendo que os mais comuns são os de limiarização (PHUNG; BOUZERDOUM; CHAI, 2005) ou baseados em modelos de probabilidade (REAES, 2006). A seguir são apresentados métodos de segmentação utilizados em diversos sistemas de reconhecimento de gestos existentes.

- ▶ Zahedi (2007): para imagens em níveis de cinza é utilizada uma limiarização enquanto que para imagens coloridas é utilizado um classificador bayesiano (MARQUES, 2005);
- ▶ Ribeiro (2006): limiarização e mistura de gaussianas;
- ▶ Conseil, Bourennane e Martin (2007): limiarização;
- ▶ Campr *et al.* (2008): mistura de gaussianas.

Nos trabalhos que abordam reconhecimento de gestos, geralmente não existe a preocupação com o efeito da variação da iluminação do ambiente na segmentação de pele, pois procura-se controlar as condições do ambiente. Contudo, outros trabalhos presentes na literatura estudam a influência da iluminação para a segmentação da pele (ZHENG *et al.*, 2005; SIGAL; SCLAROFF; ATHITSOS, 2000).

Dentre os trabalhos estudados, Zahedi *et al.* (2006) e Pistori *et al.* (2006) são os mais próximos a esta proposta. A Figura 3.3 apresenta o diagrama de blocos do sistema proposto por Zahedi *et al.* (2006).

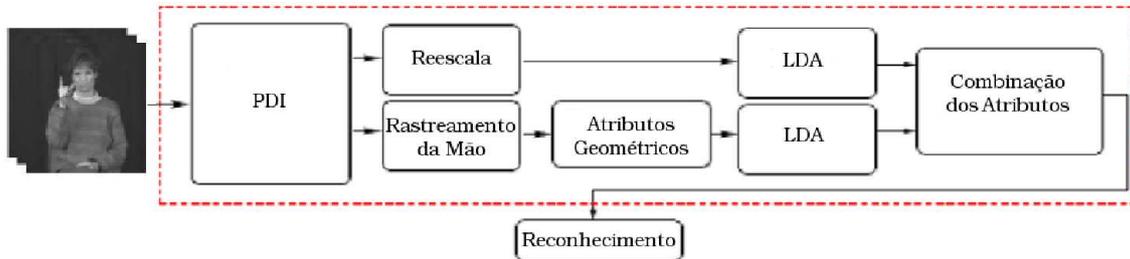


Figura 3.3: sistema proposto por Zahedi *et al.* (2006).

Zahedi *et al.* (2006) propõe um sistema de reconhecimento de gestos da ASL, o qual usa uma combinação ponderada de diferentes atributos em um *framework* estatístico de um sistema de reconhecimento da língua falada com um vasto vocabulário. Este sistema faz uma abordagem de segmentação não-explicita baseada em movimento, os atributos usados na descrição das imagens consistem em um conjunto de nove atributos geométricos bem conhecidos, a saber: área da mão, comprimento do contorno da mão, coordenadas x e y do centro geométrico, do ponto superior direito e do ponto inferior esquerdo da mão e compactação, os momentos de inércia, orientação e excentricidade do objeto, baseados nos momentos centrados, e os momentos de Hu. Como o sistema utiliza um *framework* estatístico de um sistema de reconhecimento da língua falada, e estes, segundo a autora, apresentam classificadores bayesianos, a classificação dos gestos é feita por um classificador bayesiano modificado.

A Figura 3.4 apresenta o diagrama de blocos do sistema proposto por Pistori *et al.* (2006). O sistema SIGUS é baseado em *software* livre e faz uso de uma série de aplicativos de grande difusão, tais como Java Media Framework, ImageJ e WEKA, para fornecer um ambiente de desenvolvimento de sistemas para inclusão digital de pessoas com necessidades especiais. O objetivo do sistema SIGUS é aumentar o número de programas computacionais destinados a pessoas com necessidades especiais, por exemplo limitações locomotoras ou surdez, fornecendo em um único ambiente gráfico um conjunto de ferramentas de processamento digital de imagens, como o ImageJ, e aprendizado de máquina, como o WEKA e o Adaptools, já existentes.

Comparando o sistema proposto nesta dissertação com os demais trabalhos

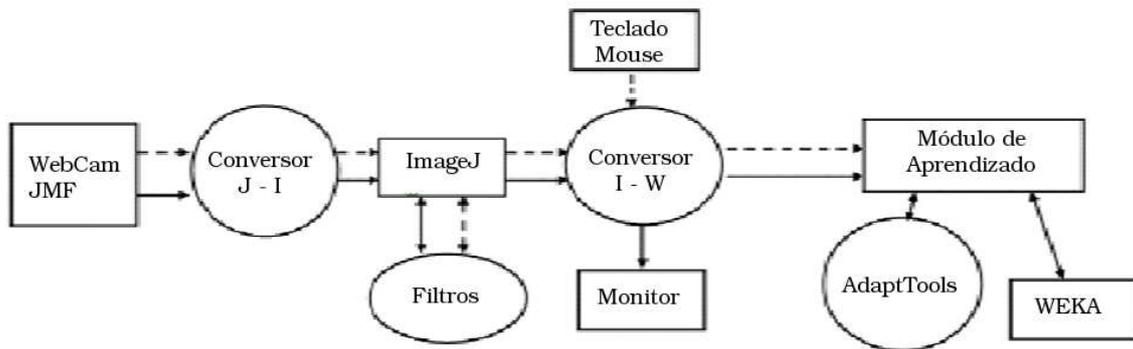


Figura 3.4: sistema proposto por Pistori *et al.* (2006).

apresentados neste capítulo, observa-se neste trabalho o emprego de algoritmos de menor complexidade e o foco específico em reconhecimento de gestos da LIBRAS, enquanto que os demais utilizam técnicas mais avançadas que são capazes de reconhecer gestos com maior complexidade, principalmente no que diz respeito ao movimento, e, por serem em sua maioria sistemas desenvolvidos em outros países, não são adequados às particularidades da LIBRAS.

Tendo conhecimento dos trabalhos apresentados neste capítulo, pode-se descrever o sistema proposto, bem como toda a metodologia empregada nesta dissertação.

Capítulo 4

Metodologia

Neste Capítulo é feita a descrição do Sistema de Reconhecimento do Alfabeto da LIBRAS por Visão Computacional e Redes Neurais. O protótipo do sistema é desenvolvido e avaliado em um computador do tipo *laptop* da marca Semp Toshiba®, modelo IS-1522 E-5596, com a seguinte configuração de *hardware*: processador Intel® Core 2 Duo T5300 1,73GHz, 2GB de memória RAM e 160GB de HD.

O sistema operacional utilizado é o GNU/Linux Ubuntu 8.04. O ambiente de simulação utilizado é o Scilab, em que são implementados os processos típicos de um sistema de visão computacional, conforme apresentado sinteticamente na Figura 4.1.

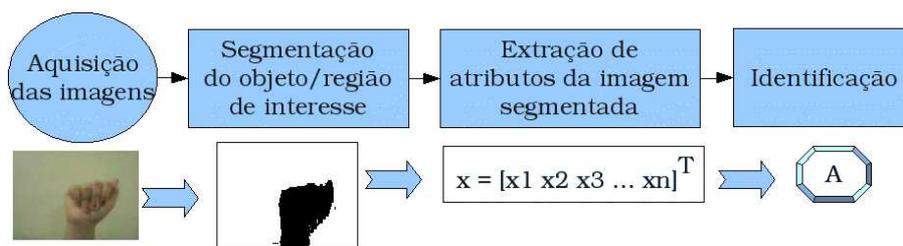


Figura 4.1: processos de um sistema de visão computacional simplificado.

O Scilab¹ contém originalmente funções e operações para manipulação de matrizes, porém não dispõe de suporte nativo para manipulação de imagens. O pacote de extensão *Scilab Image and Video Processing toolbox* (SIVp) acrescenta ao Scilab as funções necessárias para a manipulação de imagens, incluindo funções de

¹O ambiente de simulação Scilab é disponível no endereço *Web*: <http://www.scilab.org/> juntamente com toda a documentação e pacotes de extensão.

leitura e escrita de arquivos de imagens em diversos padrões e conversões entre os formatos de imagens, tais como RGB, YCbCr e HSV, necessários para os algoritmos de segmentação.

Os algoritmos apresentados no capítulo anterior são avaliados em conjunto neste capítulo, em que cada retângulo da Figura 4.1 é substituído por um conjunto de algoritmos correspondentes.

4.1 Visão geral do sistema

Esta dissertação propõe um sistema de visão computacional baseado em redes neurais, organizados de forma a viabilizar o reconhecimento dos gestos que compõem o alfabeto da LIBRAS, apresentado na Figura 4.2.

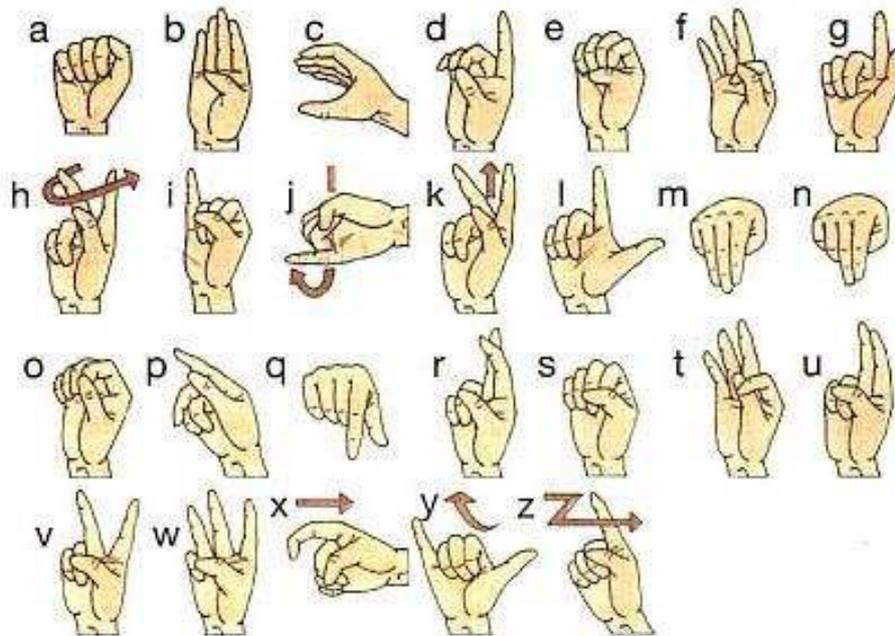


Figura 4.2: alfabeto da LIBRAS.

Os gestos do alfabeto da LIBRAS são detalhadamente analisados no Apêndice A. Todavia, é importante observar que seis dos vinte e seis gestos são dinâmicos, ou seja, caracterizam-se por movimento². O diagrama do sistema é apresentado na Figura 4.3.

A seguir são apresentadas as peculiaridades de cada elemento do diagrama da Figura 4.3, de acordo com os algoritmos avaliados.

²A presença de movimentos nos gestos faz com que seja importante inserir informações sobre este movimento juntamente aos descritores geométricos.

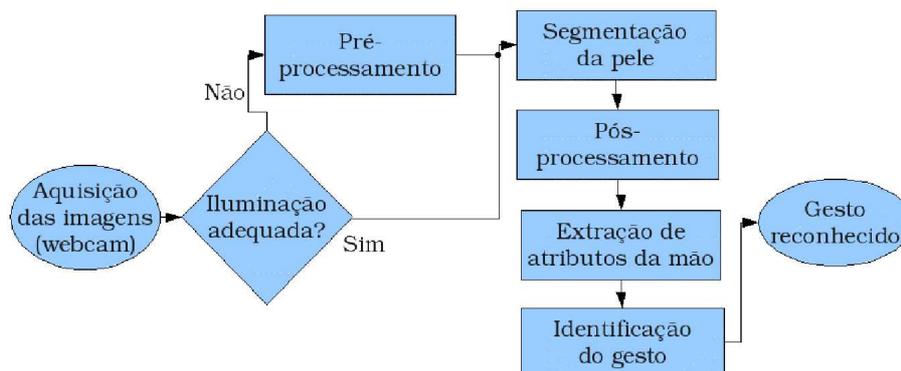


Figura 4.3: diagrama do sistema proposto.

4.2 Aquisição das imagens

A primeira etapa do sistema é a aquisição das imagens que são capturadas por uma *webcam* modelo WebPro 2 da marca LG com padrão de captura VGA (resolução das imagens de 640 por 480 *pixels*), utilizando-se o *software* Scilab com a *toolbox* SIVp.

As imagens de teste foram geradas nas dependências da Associação dos Surdos do Ceará (ASCE) em local com iluminação artificial. As pessoas que posaram para a geração das imagens são profissionais, estudantes e usuários da LIBRAS, totalizando seis pessoas, incluindo crianças, jovens e adultos. A Figura 4.4 apresenta uma imagem da mão de cada uma das pessoas que colaboraram para a geração das imagens.

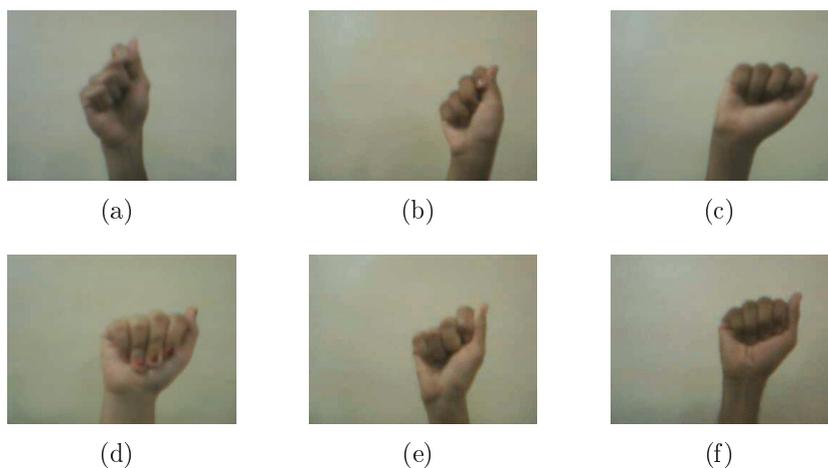


Figura 4.4: imagens relativas à letra 'A' do alfabeto da LIBRAS de cada uma das pessoas que serviram de modelo.

Segue a descrição de cada uma das pessoas que serviram de modelo para as filmagens:

- ▶ a primeira pessoa que serviu de modelo foi um jovem do sexo masculino;
- ▶ a segunda pessoa que serviu de modelo foi uma criança do sexo feminino;
- ▶ a terceira pessoa que serviu de modelo foi uma criança do sexo masculino;
- ▶ a quarta pessoa que serviu de modelo foi uma criança do sexo feminino;
- ▶ a quinta pessoa que serviu de modelo foi um adulto do sexo feminino;
- ▶ a sexta pessoa que serviu de modelo foi um adulto do sexo masculino.

Ao todo foram geradas 7800 imagens, concentradas em 156 vídeos, adquiridos a uma taxa de 11 *frames* por segundo, de 50 *frames* cada (26 vídeos, um por letra, por pessoa). A taxa de 11 *frames* por segundo foi utilizada como um limite mínimo para as câmeras disponíveis atualmente, de modo que é improvável que algum usuário utilize uma câmera com taxa de captura de *frames* inferior a 11 por segundo. As imagens foram geradas no padrão RGB com 3 *bytes* por *pixel*.

4.3 Etapa de pré-processamento

Para a primeira análise, as imagens são submetidas a cada um dos algoritmos de segmentação apresentados na Seção 4.4 para uma avaliação individual por algoritmo. Com base nos resultados de cada algoritmo de segmentação, aplicados a uma mesma imagem sem nenhuma forma de pré-processamento, estes são analisados a fim de se determinar o melhor algoritmo de segmentação para as condições onde as imagens são capturadas. Um exemplo está apresentado na Figura 4.5, em que a Figura 4.5(a) é uma imagem original e as Figuras 4.5(b), 4.5(c), 4.5(d) e 4.5(e) correspondem, respectivamente, aos resultados obtidos com os algoritmos de segmentação em RGB, YCbCr, HSV e por modelo de probabilidade apresentados na seção a seguir.

O método de pré-processamento necessita dos resultados da segmentação para ser desenvolvido, pois este utiliza um histograma de referência para mapear o histograma das demais imagens e esse é obtido como sendo o histograma da imagem que obtiver o melhor resultado de segmentação.

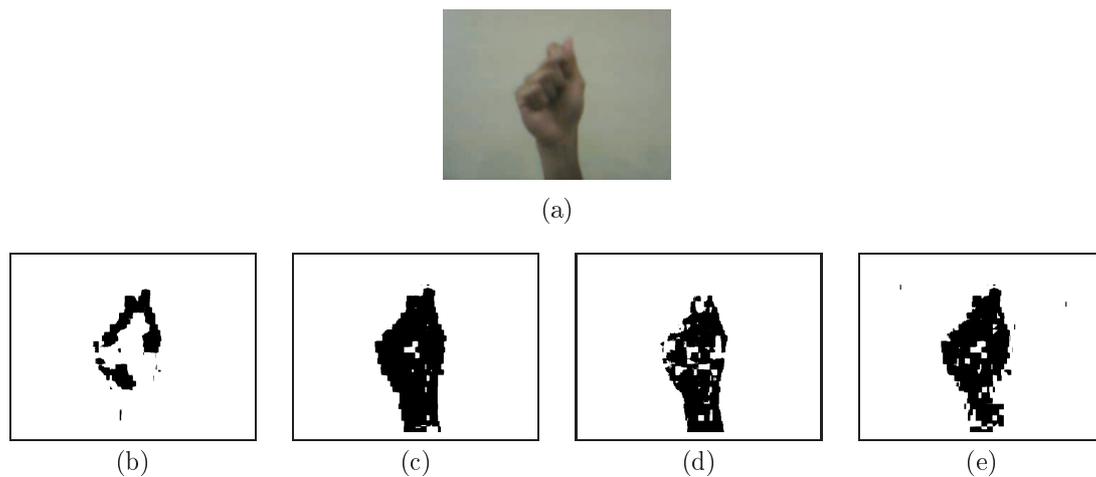


Figura 4.5: (a) imagem original e os resultados dos algoritmos de segmentação (b) em RGB, (c) em YCbCr, (d) em HSV e (e) pelo modelo de probabilidade.

Caso algum dos algoritmos apresente uma segmentação satisfatória em todas as imagens, as imagens obtidas são direcionadas para a etapa de segmentação, utilizando o algoritmo que apresenta os melhores resultados. Porém, se nenhum dos algoritmos se mostrar eficiente para todas as imagens disponíveis, então estas são submetidas à etapa de correção de histograma. Esta etapa consiste em reproduzir o histograma da imagem que fornece o melhor resultado de segmentação nas demais imagens, fazendo com que todas as imagens apresentem resultados de segmentação semelhantes ao melhor resultado possível. A Figura 4.6 apresenta um exemplo dos melhores resultados de segmentação dentre os testes realizados.

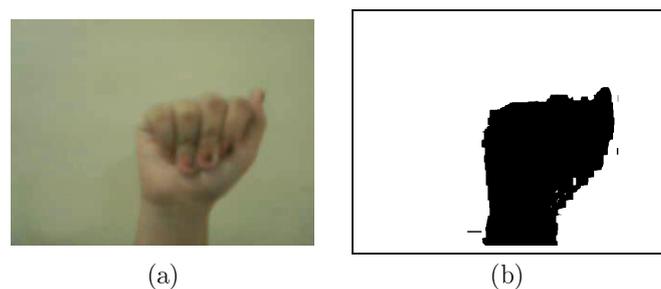


Figura 4.6: resultado de referência dos algoritmos de segmentação (a) imagem original colorida e (b) imagem segmentada.

O resultado apresentado na Figura 4.6 é obtido pelo algoritmo de segmentação em YCbCr, o qual consiste de uma limiarização simples. A Figura 4.7 apresenta os histogramas dos canais Cb e Cr da imagem original, apresentada na Figura 4.6(a).

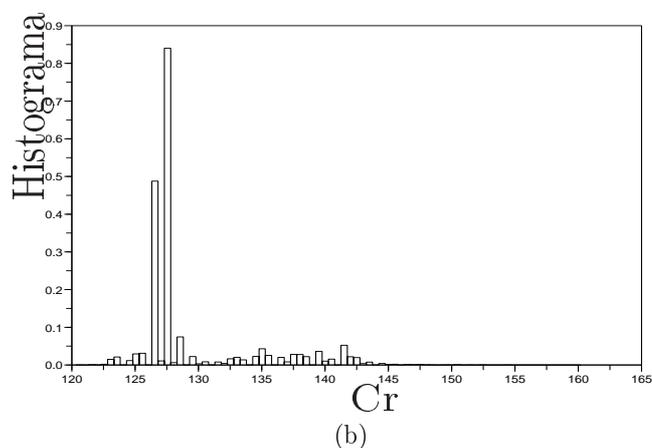
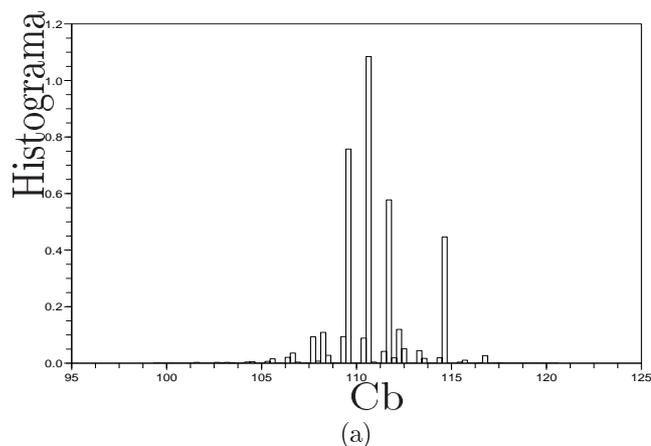


Figura 4.7: histogramas relativos aos canais (a) Cb e (b) Cr da imagem de referência.

Observa-se na Figura 4.7(a) que as componentes do canal Cb estão todas concentradas em um intervalo sem que seja possível identificar as componentes de pele e do fundo, de modo que, para o ambiente em que as imagens são geradas, as componentes Cb não influenciam significativamente o algoritmo. Assim, o discriminante da pele encontra-se apenas no canal Cr , no qual observa-se duas concentrações principais, em que a primeira, maior, é centrada próximo de 125 e a outra, menor, é centrada próximo de 137. Segundo Mahmoud (2008), Conseil, Bourennane e Martin (2007), Phung, Bouzerdoum e Chai (2005), o intervalo de valores no canal Cr que compreende a pele é $133 \leq Cr \leq 173$, conclui-se então que a segunda concentração, centrada em 137, corresponde à região de pele.

Neste caso, aplica-se a segmentação por limiar obtendo-se os resultados sobre o histograma apresentado na Figura 4.8, em que a região à esquerda da reta tracejada

corresponde ao fundo da imagem e a região a direita corresponde a pele.

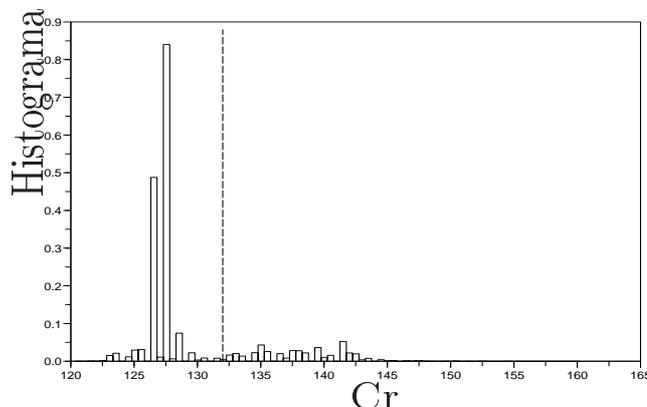


Figura 4.8: resultado da separação dos histogramas em pele e fundo.

A determinação do histograma desejado, o qual é usado na Equação (2.3), é especificado como duas distribuições gaussianas: uma para representar a pele e a outra o fundo. Neste caso, as médias e variâncias são obtidas com base nas médias e variâncias das distribuições apresentadas nos gráficos da Figura 4.8.

O algoritmo de especificação de histograma, descrito na Seção 2.1, é aplicado aos canais que são usados pelo algoritmo de segmentação em questão, neste caso os canais Cb e Cr . Após a transformação, os canais são reinseridos na imagem resultante, a qual é submetida aos algoritmos de segmentação para avaliação da influência do algoritmo de especificação de histograma no processo de segmentação das imagens.

4.4 Etapa de segmentação da pele

Os métodos de segmentação de pele usados no desenvolvimento do sistema SIRENE são apresentados na Tabela 4.1, com os respectivos espaços de cor e sua referência.

As seções a seguir descevem detalhadamente cada um dos algoritmos apresentados na Tabela 4.1.

Tabela 4.1: algoritmos de segmentação da pele.

Espaço de cor	Regras	Referências
RGB	$R > 95, G > 40, B > 20,$ $\max \{R, G, B\} - \min \{R, G, B\} > 15,$ $ R - G > 15, R > G e R > B$	(RIBEIRO, 2006)
YCbCr	$77 \leq Cb \leq 127 e 133 \leq Cr \leq 173$	(MAHMOUD, 2008)
HSV	$0^\circ \leq H \leq 50^\circ e 0,23 \leq S \leq 0,68$	(RIBEIRO, 2006)
YCbCr	$p_X(\underline{x}) = \frac{1}{2\pi \mathbf{C} ^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \mathbf{C}^{-1}(\underline{x}-\underline{\mu})} > \mu_X + \sigma_X$	(REAES, 2006)

4.4.1 Algoritmo de segmentação por conjunto de limiares no espaço de cores RGB

O primeiro algoritmo de segmentação de pele consiste na verificação dos seguintes limiares sobre as componentes de cor de cada *pixel* da imagem codificada no padrão de cor RGB (RIBEIRO; GONZAGA, 2006; RIBEIRO, 2006; RASOOL, 2003):

$$\begin{aligned}
 &R > 95, G > 40, B > 20, \\
 &\max \{R, G, B\} - \min \{R, G, B\} > 15, \\
 &|R - G| > 15, R > G e R > B.
 \end{aligned} \tag{4.1}$$

Estes limiares delimitam uma região específica no espaço RGB. No caso deste algoritmo, o custo computacional torna-se relativamente elevado, pois todas as condições devem ser avaliadas para cada *pixel* a cada nova imagem adquirida.

4.4.2 Algoritmos de segmentação por limiarização simples

Os algoritmos de segmentação por limiarização, em geral, possuem o menor custo computacional dentre os algoritmos de segmentação.

Os dois algoritmos de segmentação da pele por limiarização simples utilizados neste trabalho têm a mesma fundamentação, porém distinguem-se pela codificação

de cor ao qual são aplicados. O primeiro, que usa a codificação YCbCr, assume que os *pixels* pertencentes à pele humana, para os canais Cb e Cr, estão entre os valores (MAHMOUD, 2008; CONSEIL; BOURENNANE; MARTIN, 2007; PHUNG; BOUZERDOUM; CHAI, 2005):

$$77 \leq Cb \leq 127 \text{ e } 133 \leq Cr \leq 173. \quad (4.2)$$

Seguindo a ordem da Tabela 4.1, o algoritmo seguinte, que usa o espaço de cor HSV, consiste na aplicação de limiares para imagens coloridas codificadas neste padrão. Este método assume que os *pixels* pertencentes a pele humana, para os canais H e S, estão entre os valores (PHUNG; BOUZERDOUM; CHAI, 2005; RIBEIRO, 2006):

$$0^\circ \leq H \leq 50^\circ \text{ e } 0,23 \leq S \leq 0,68. \quad (4.3)$$

Após a conversão para a codificação utilizada, o custo computacional de ambos os algoritmos é o menor, quando comparado aos demais algoritmos de segmentação.

4.4.3 Algoritmo de segmentação por modelo de probabilidade

Diferente dos demais algoritmos, o algoritmo de segmentação por modelo de probabilidade apresenta maior robustez do ponto de vista estatístico, pois, considera aspectos da distribuição normalmente encontrada dos *pixels* de pele em imagens (FUKUNAGA, 1990).

Este algoritmo de segmentação de pele consiste em gerar uma imagem de probabilidades a partir das componentes Cb e Cr da imagem original. Cada *pixel* da imagem de probabilidades é um valor escalar dado por (REAES, 2006):

$$p(\underline{x}) = \frac{1}{2\pi|\mathbf{C}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \mathbf{C}^{-1}(\underline{x}-\underline{\mu})}, \quad (4.4)$$

em que \underline{x} é um vetor com as componentes Cb e Cr de um *pixel* da imagem, $\underline{\mu}$ é um vetor, de valor constante, com as componentes Cb e Cr médias para a pele humana, \mathbf{C} é a matriz de covariância obtida com *pixels* correspondentes às regiões de pele e $|\cdot|$ é o operador determinante. O valor $p(\underline{x})$ corresponde a probabilidade com que um *pixel* na imagem original pertença a uma região de pele.

Reaes (2006) apresenta as variáveis presentes na Equação (4.4) como segue

$$\underline{x} = [Cr \quad Cb]^T, \quad (4.5)$$

$$\underline{\mu} = [148,875 \quad 114,872]^T \quad (4.6)$$

e

$$\mathbf{C}^{-1} = \begin{bmatrix} 0,044 & 0,0318 \\ 0,0318 & 0,042 \end{bmatrix}. \quad (4.7)$$

Sabe-se que, para qualquer matriz \mathbf{A} inversível, $\frac{1}{|\mathbf{A}|} = |\mathbf{A}^{-1}|$ (CALLIOLI; DOMINGUES; COSTA, 1990), então como os demais termos da Equação (4.4) são conhecidos tem-se todos os parâmetros da função densidade de probabilidade.

4.5 Pós-processamento

A segmentação da pele neste trabalho é empregada conforme discutido na seção anterior, porém, para a etapa de extração de atributos, é necessário pós-processar a imagem segmentada para excluir regiões de fundo que tenham sido erroneamente segmentadas como pele, apresentadas como regiões vermelhas e regiões de pele não identificadas, apresentadas como regiões azuis na Figura 4.9.

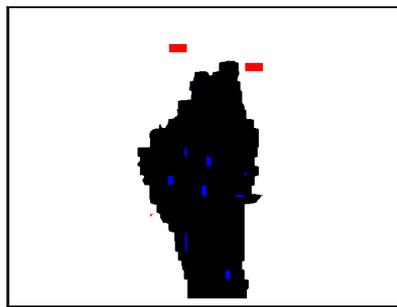


Figura 4.9: regiões com falhas no processo de segmentação.

O pós-processamento da imagem é realizado por um algoritmo de crescimento de região, o qual consiste em, dados um ponto qualquer em uma imagem, $\underline{p} = [x \ y]^T$, e uma propriedade de interesse, agrupar os pontos adjacentes a \underline{p} que apresentem a mesma propriedade de interesse.

O procedimento consiste em duas aplicações do algoritmo: a primeira para preencher as regiões de pele não identificadas e a segunda para retirar as regiões de fundo identificadas erroneamente. A Figura 4.10 apresenta um exemplo de imagens obtidas após cada uma das aplicações do algoritmo. Nesta figura, as regiões identificadas pelo algoritmo são apresentadas em cinza.

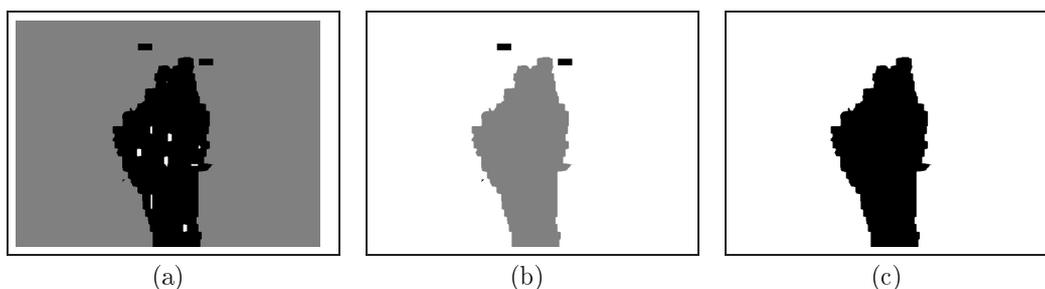


Figura 4.10: imagens obtidas pelo processo de pós-processamento (a) a primeira, (b) a segunda aplicação do algoritmo e (c) resultado final.

A semente escolhida para a primeira aplicação do algoritmo é o canto superior esquerdo, pois este ponto não é identificado como pele em nenhuma imagem. Para a segunda aplicação do algoritmo, a semente escolhida é o centro de massa da imagem definido na Equação (2.11). Em ambos os casos, a propriedade de interesse é seu próprio valor que corresponde ao fundo e região de pele, respectivamente, visto que a imagem segmentada é binária.

As regiões preta e branca são unidas, tornando-se uma única região preta, e a região cinza passa a ser branca, na Figura 4.10(a). Enquanto que, na Figura 4.10(b), as regiões branca e preta são unidas, resultando em uma única região branca, e a região cinza passa a ser preta. O resultado final obtido é apresentado na Figura 4.10(c).

Para o uso dos Descritores de Fourier, como estes são calculados a partir de uma assinatura de forma do contorno da região segmentada, a imagem precisa ser submetida a um algoritmo de detecção de bordas. O método adotado neste trabalho para detecção de bordas é o gradiente de Sobel seguido de uma operação morfológica de erosão com um elemento estruturante quadrado de dimensões 3×3 (GONZALEZ; WOODS, 2008), a aplicação da operação morfológica tem por finalidade o afinamento do contorno obtido pelo gradiente de Sobel. A Figura 4.11 apresenta um exemplo deste processo.

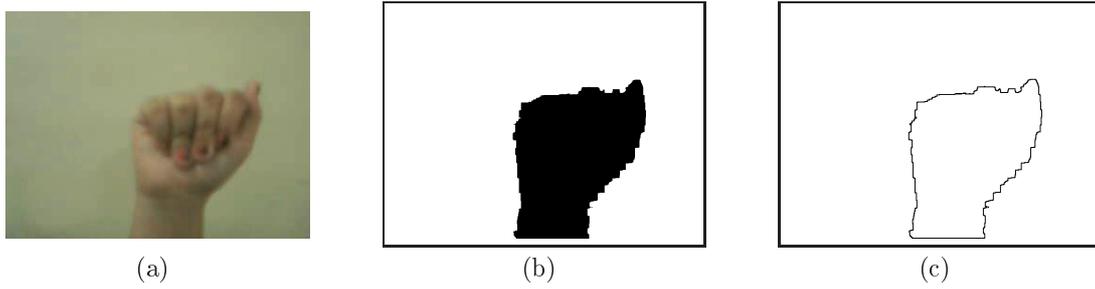


Figura 4.11: processo de detecção de bordas (a) imagem colorida original, (b) imagem binária segmentada, e (c) contorno da imagem.

Assim, as etapas de segmentação e pós-processamento são concluídas e as imagens são direcionadas à etapa seguinte, extração de atributos.

4.6 Etapa de extração de atributos da mão

A imagem segmentada segue o sentido do fluxo da informação do sistema para a etapa de extração de atributos conforme mostrado na Figura 4.3. Os atributos escolhidos neste trabalho são os Momentos Invariantes de Hu e os Descritores de Fourier, apresentados na Seção 2.4.

São empregadas duas formas de descrição das imagens. A primeira cujo vetor de atributos é constituído pelos seis Momentos Invariantes de Hu e as componentes de velocidade apresentadas na Equação (2.22), e a segunda cujo vetor de atributos é constituído por dez, trinta ou cinquenta Descritores de Fourier e as componentes de velocidade apresentadas na equação 2.23.

Os Momentos Invariantes de Hu $\{\phi_1 \phi_2 \phi_3 \phi_4 \phi_5 \phi_6\}$ são calculados conforme apresentado na Seção 2.4.1 usando as imagens segmentadas convertidas para níveis de cinza aplicando o produto elemento a elemento da imagem segmentada com a imagem original convertida para níveis de cinza, como apresentado na Figura 4.12.

Para cada imagem, as componentes de velocidade v_x e v_y , calculadas conforme apresentado no Capítulo 2, são usadas na composição do vetor de atributos \underline{x} .

O vetor de atributos extraídos da imagem é, então

$$\underline{x} = [v_x \ v_y \ \phi_1 \ \phi_2 \ \phi_3 \ \phi_4 \ \phi_5 \ \phi_6]^T. \quad (4.8)$$

Para a obtenção do vetor de atributos com os Descritores de Fourier, conforme

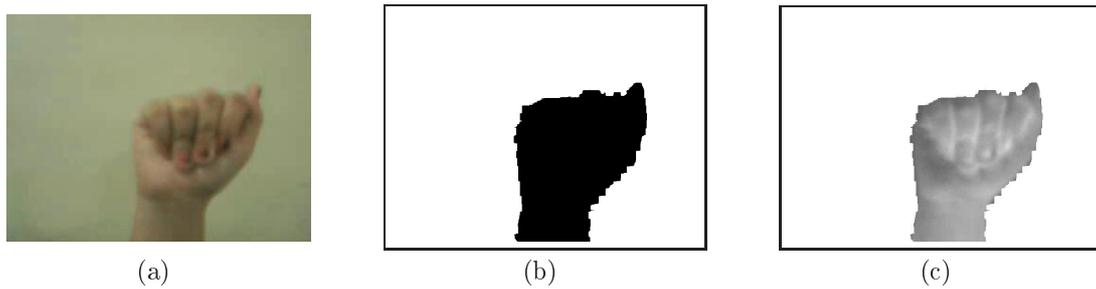


Figura 4.12: (a) imagem colorida original, (b) imagem binária segmentada e (c) imagem em níveis de cinza segmentada.

apresentado na Seção 2.4.2, após a extração do contorno é calculada a assinatura de forma de distância ao centróide. Em seguida, o vetor de atributos com a assinatura de forma é sub-amostrado de modo a possuir uma quantidade de componentes pré-definida. Neste trabalho, os descritores de Fourier apresentam dez, trinta e cinquenta componentes. A Figura 4.13 apresenta os pontos sobre a imagem que correspondem ao contorno original e aos contornos após a sub-amostragem.

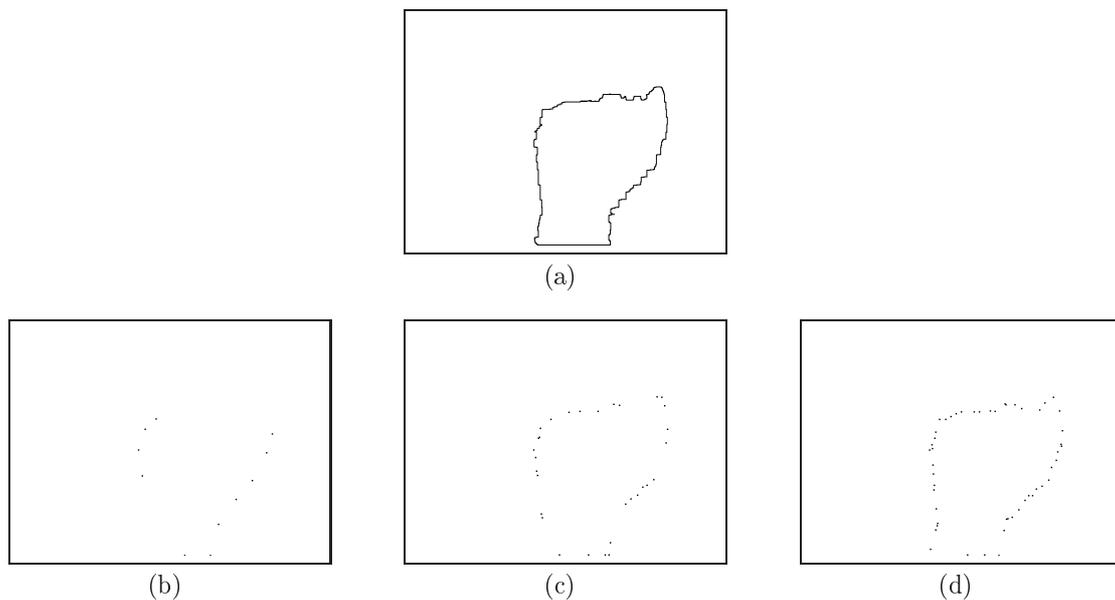


Figura 4.13: sub-amostragem do contorno original: (a) contorno original, contorno com (b) 10, (c) 30 e (d) 50 amostras.

Os Descritores de Fourier são calculados sobre os pontos do contorno resultantes da sub-amostragem da imagem binária, de modo que o vetor de atributos extraídos da imagem é, então

$$\underline{x} = [v_x \ v_y \ u_1 \ u_2 \ \dots \ u_N]^T, \quad (4.9)$$

em que $N \in \{10, 30, 50\}$.

Tanto para o vetor relativo aos Momentos de Hu como para o vetor relativo aos Descritores de Fourier, antes do vetor de atributos ser utilizado para a etapa final do sistema, que é a classificação, as componentes são normalizadas de modo que cada atributo esteja compreendido no intervalo $[0, 1]$.

4.7 Pré-classificação dos gestos

A classificação dos gestos ocorre em duas etapas: a pré-classificação e a classificação final. De posse dos vetores de atributos \underline{x} para todas as imagens disponíveis, a pré-classificação dos gestos consiste em agrupar os vetores em *clusters*, de modo que cada *cluster* contenha um número reduzido de classes.

O algoritmo de *clustering* utilizado neste trabalho é a rede SOM, descrito na Seção 2.5.1. A arquitetura de rede SOM adotada neste trabalho é a unidimensional. A Figura 4.14 apresenta um exemplo de aplicação da rede SOM unidimensional em um espaço de entrada bidimensional.

O resultado obtido no processo de *clustering* é apresentado na Figura 4.14(b), e ilustra o que acontece com os dados relativos aos vetores de atributos \underline{x} , porém em duas dimensões.

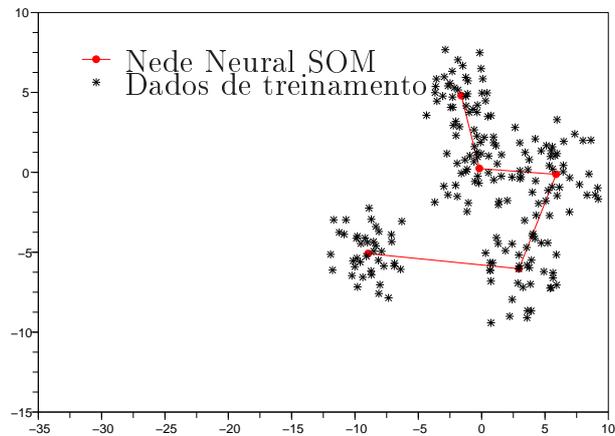
O processo de *clustering* consiste em rotular cada um dos vetores de atributos com o índice, c_w , do protótipo da rede SOM mais próximo

$$c_w = \arg \min_{\forall_i} \{ \|\underline{x} - \underline{m}_i\| \}, \quad (4.10)$$

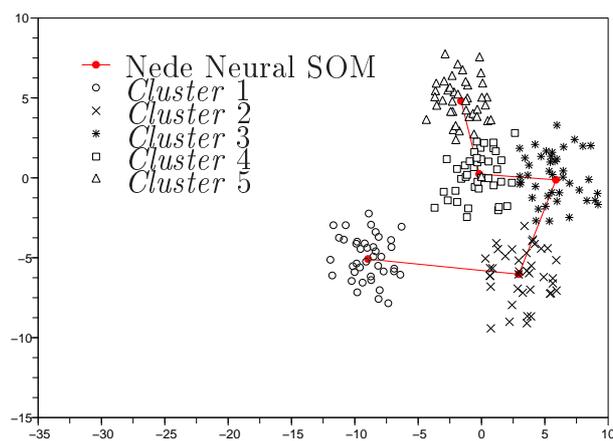
em que $\|\cdot\|$ denota a norma euclidiana.

De posse dos rótulos de cada vetor de atributos, é necessário identificar quais classes pertencem a quais *clusters*. Para isto, computa-se a incidência de cada classe em cada um dos *clusters*, obtendo-se uma matriz de incidência, como no exemplo apresentado na Tabela 4.2 em que são agrupados 300 dados, divididos em 3 classes de 100 dados cada, em 5 *clusters*.

Para determinar as classes e os *clusters* correspondentes, calcula-se a incidência cumulativa em cada classe, por *cluster*, da seguinte forma, tomando o exemplo da



(a)



(b)

Figura 4.14: exemplo de aplicação da rede SOM para *clustering* (a) aplicação da rede SOM sobre todos os dados e (b) resultado após o processo de *clustering*.

Classe 1 da Tabela 4.2:

- i. ordena-se os *clusters* em ordem crescente de incidência por classe, segundo o exemplo adotado, a sequência dos *clusters* para a Classe 1 é [4, 1, 5, 2, 3];
- ii. determina-se a incidência mínima acumulada Inc_{min} , neste caso supõe-se $Inc_{min} = 80\%$;
- iii. calcula-se a incidência acumulada de cada *cluster* obedecendo a ordem do

Tabela 4.2: incidência de classes por *cluster*.

<i>Cluster</i>	Classe 1	Classe 2	Classe 3
1	12	28	0
2	4	0	84
3	2	4	11
4	75	2	2
5	7	66	3

passo 1, para o exemplo a incidência acumulada para a Classe 1 é dada por [75%, 87%, 94%, 98%, 100%];

- iv. a classe em questão pertence então aos *clusters*, cuja incidência acumulada seja menor do que Inc_{min} e o que apresenta incidência acumulada imediatamente acima do valor Inc_{min} , de modo que, para a classe usada como exemplo pertence aos *clusters* 4 e 1.

Seguindo os passos descritos anteriormente, obtém-se a matriz de correspondências entre classes e *clusters* apresentada na Tabela 4.3.

Tabela 4.3: correspondências entre classes e *cluster*.

<i>Cluster</i>	Classe 1	Classe 2	Classe 3
1	X	X	-
2	-	-	X
3	-	-	-
4	X	-	-
5	-	X	-

A matriz de correspondências para 3 classes e 5 *clusters* do exemplo é apresentada na Tabela 4.3.

Para este trabalho foram usadas 26 classes, relativas as 26 letras do alfabeto da LIBRAS, e uma quantidade variável de *clusters*. Em cada caso, a quantidade de protótipos é determinada de tal modo que cada *cluster* apresente no máximo 6 classes, para os vetores de atributos dos Momentos Invariantes de Hu, e 8 classes, para os vetores de atributos dos Descritores de Fourier ³.

³Os valores 6 e 8 são obtidos após a execução de exaustivos testes com os dados obtidos.

4.8 Classificação dos gestos

Os dados pertencentes a cada cluster é direcionado a uma rede neural supervisionada, apresentadas nas Seções 2.5.2, 2.5.3 e 2.5.4, conforme mostrado no diagrama da Figura 4.15.

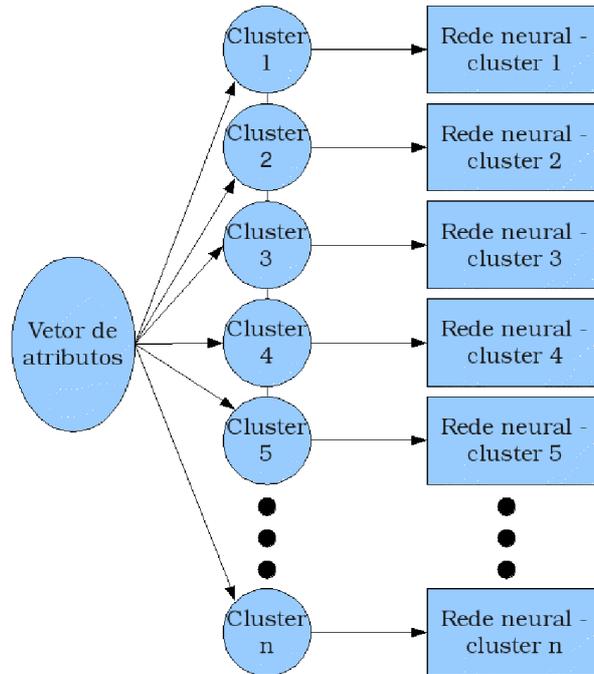


Figura 4.15: diagrama do sistema de classificação em dois passos.

Como apresentado na Seção 2.5.4, a rede neural RBF é constituída de duas etapas: um algoritmo de *clustering*, que calcula os centróides das funções $B_i(\underline{x})$, e a função de ativação $\phi(\underline{z})$ é a tangente hiperbólica, Equação (2.29), definida como a saída de uma rede neural PS.

O treinamento das redes neurais é feito com os dados de cada *cluster* correspondente, ou seja, a rede neural do *cluster* i (genérico) é treinada exclusivamente com dados deste *cluster*.

Para as redes neurais MLP e RBF, que apresentam camada oculta, o número de neurônios da camada oculta é determinado mediante a análise dos resultados obtidos, todavia, os valores para este número, N_h , usados neste trabalho são

$$N_h \in \left[\left\lfloor \frac{N_{in} + N_s}{2} \right\rfloor, \max \{N_{in}, N_s\} \right], \quad (4.11)$$

para a rede MLP e

para a rede MLP e

$$N_h \in [\min \{N_{in}, N_s\}, \max \{N_{in}, N_s\}], \quad (4.12)$$

para a rede RBF, em que N_{in} corresponde à quantidade de atributos de entrada e N_s à quantidade de neurônios na camada de saída.

Cada rede neural apresenta tantos neurônios na camada de saída quantas sejam a quantidade de classes do *cluster* correspondente, de modo que, se um *cluster* tem Nc classes, a sua rede neural correspondente tem $N_s = Nc$ neurônios de saída, dos quais cada neurônio de saída representa uma das classes pertencentes ao *cluster*. Sendo cada classe representada por um neurônio de saída, os padrões de saída desejados são como os apresentados na Tabela 4.4.

Tabela 4.4: saídas desejadas para uma rede neural com 5 neurônios de saída.

Classe correspondente	Saída dos neurônios				
	1	2	3	4	5
1	A	-A	-A	-A	-A
2	-A	A	-A	-A	-A
3	-A	-A	A	-A	-A
4	-A	-A	-A	A	-A
5	-A	-A	-A	-A	A

Conforme recomendado pela literatura, quando os neurônios de saída usam a função tangente hiperbólica, o valor de A usado neste trabalho é $A = 0,95$, não sendo assim assintótico para a função de ativação (valor 1), todavia é próximo de 1 (HAYKIN, 1999).

Definida a metodologia empregada nesta dissertação, obtém-se os resultados para os diversos algoritmos de cada uma das etapas do sistema, os quais são apresentados no próximo capítulo.

Capítulo 5

Resultados e discussão

O Sistema de Reconhecimento do Alfabeto da Libras por Visão Computacional e Redes Neurais implementa uma interface de interação humano-computador por visão computacional baseada no reconhecimento de gestos da mão.

Neste capítulo são apresentados os resultados obtidos com os processos de pré-processamento, segmentação, pós-processamento e com as técnicas usadas nas etapas de pré-classificação e classificação dos gestos, além de comparar as diferentes técnicas e algoritmos utilizados.

5.1 Resultados do pré-processamento

Para comprovar a eficiência do método de realce proposto (Seção 2.1), a Figura 5.1 apresenta um resultado típico obtido com a aplicação do algoritmo de especificação de histograma cujos resultados da segmentação são mostrados na Figura 5.2.

As imagens de teste apresentadas nas Figuras 5.1(a), 5.1(b) e 5.1(c) foram obtidas em um ambiente sem controle de iluminação, influenciando fortemente no resultado da segmentação, como é visto nas Figuras 5.2(d), 5.2(e) e 5.2(f).

A taxa de acerto percentual para os algoritmos de segmentação utilizada neste trabalho, r_{segm} , é baseada em similaridade de conjuntos e dada por

$$r_{segm} = 100 \times \frac{N_{pele} [\min \{I_{ref}, I_{segm}\}]}{N_{pele} [\max \{I_{ref}, I_{segm}\}]}, \quad (5.1)$$

em que I_{ref} corresponde à imagem de referência, obtida pela segmentação manual, I_{segm} corresponde à imagem segmentada resultante dos algoritmos de segmentação

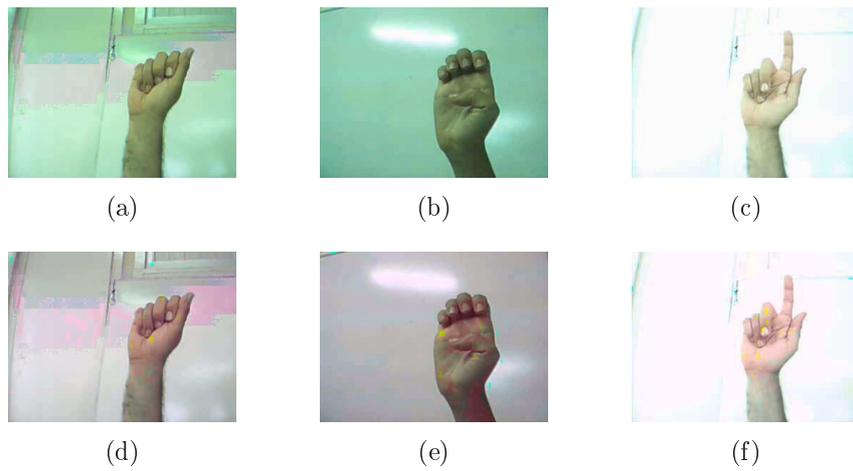


Figura 5.1: (a)-(c) imagens originais, (d)-(f) imagens obtidas após a aplicação do algoritmo de especificação de histograma.

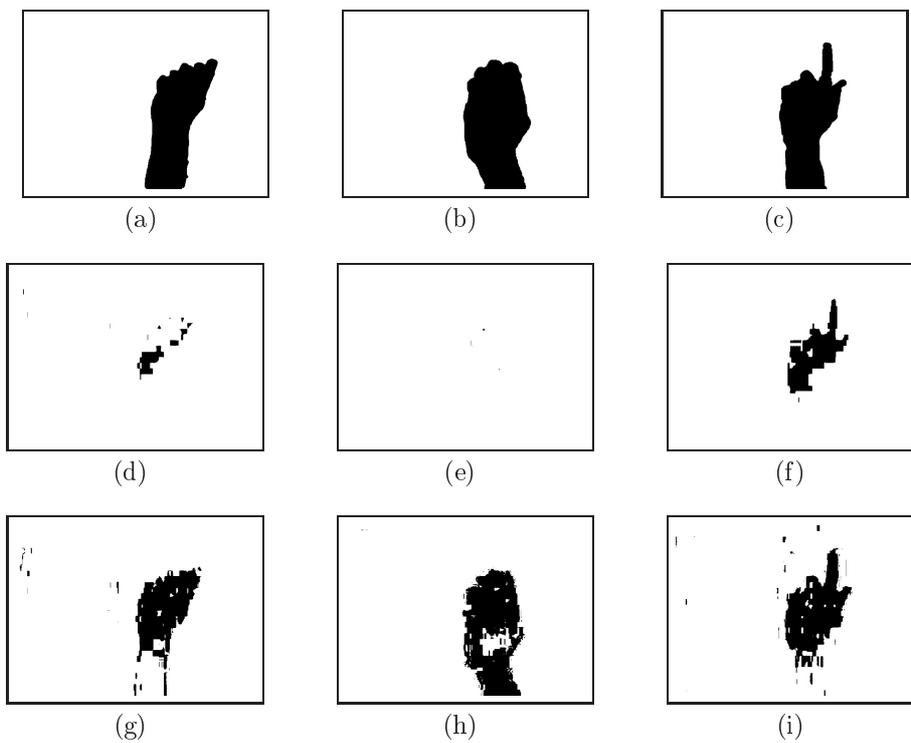


Figura 5.2: (a)-(c) resultados obtidos pela segmentação manual e (d)-(f) resultados obtidos pela segmentação das imagens das Figuras 5.1(a) a 5.1(c) e (g)-(i) resultados obtidos pela segmentação das imagens das Figuras 5.1(d) a 5.1(f).

e $N_{pele} [\cdot]$ representa o número de *pixels* de pele identificados. Assim, o numerador da fração representa a região de pele corretamente reconhecida e o denominador representa a região total de pele das duas imagens, incluindo os erros de segmentação.

Esta métrica de análise do resultado da segmentação é uma contribuição deste trabalho.

As taxas de acerto percentuais para as imagens das Figuras 5.2(d), 5.2(e) e 5.2(f) são, respectivamente, $r_{segm\{1\}} = 10,7\%$, $r_{segm\{2\}} = 0,288\%$ e $r_{segm\{3\}} = 41,8\%$, enquanto que para as imagens das Figuras 5.2(g), 5.2(h) e 5.2(i) são $r_{segm\{1\}} = 63,4\%$, $r_{segm\{2\}} = 79,4\%$ e $r_{segm\{3\}} = 67,1\%$.

Os resultados apresentados são obtidos com o uso da segmentação por limiarização em YCbCr.

5.2 Resultados da segmentação

Nesta seção são apresentados os resultados relativos à aplicação dos algoritmos correspondentes às duas primeiras etapas do sistema: pré-processamento e segmentação. Como apresentado na Seção 4.3, o algoritmo de especificação de histograma, utilizado como pré-processamento, depende dos resultados da segmentação e somente é aplicado quando os algoritmos de segmentação não apresentam resultados satisfatórios.

Os resultados de segmentação para cada um dos algoritmos apresentados na Seção 4.4 são obtidos e analisados a partir das imagens apresentadas na Figura 5.3.

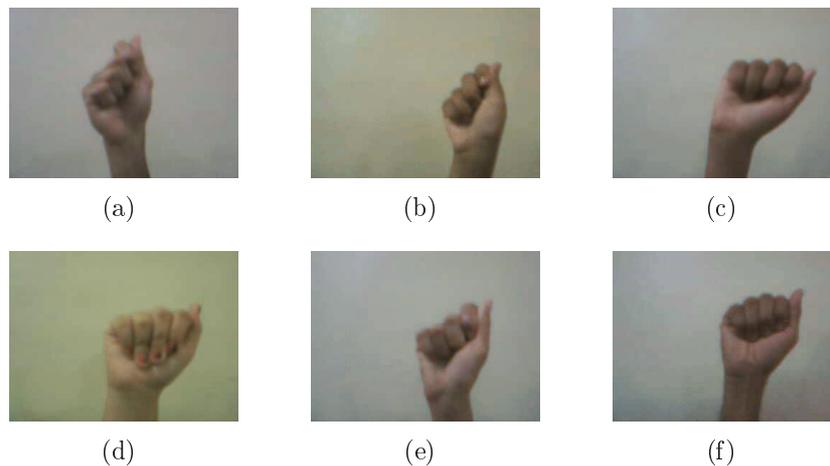


Figura 5.3: imagens originais de seis pessoas: (a) primeira, (b) segunda, (c) terceira, (d) quarta, (e) quinta e (f) sexta.

Os resultados da segmentação para cada um dos algoritmos é comparado com uma segmentação ideal, feita manualmente, apresentada na Figura 5.4.

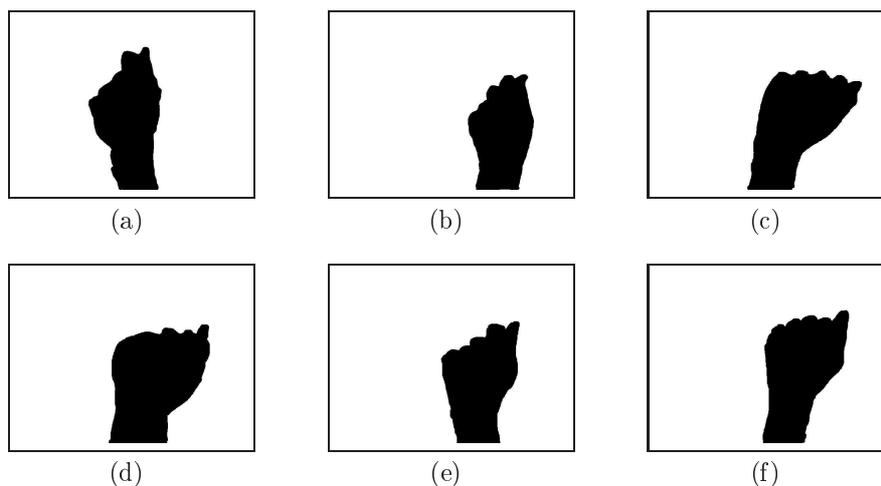


Figura 5.4: resultados da segmentação ideal.

Os resultados da segmentação com o algoritmo de limiarização em YCbCr (Seção 4.4.2), em HSV (Seção 4.4.2), em RGB (Seção 4.4.1) e por modelo de probabilidade (Seção 4.4.3) são apresentados nas Figuras 5.5(a) a 5.5(f), 5.5(g) a 5.5(l), 5.5(m) a 5.5(r) e 5.5(s) a 5.5(x), respectivamente.

As taxas de acerto percentual obtidas no processo de segmentação são apresentadas na Tabela 5.1.

Tabela 5.1: resultados obtidos com os algoritmos de segmentação.

Pessoas	Taxas de acerto por algoritmo de segmentação			
	YCbCr	HSV	RGB	Probabilidade
primeira	87,1%	69,4%	30,3%	78,7%
segunda	92,2%	68,4%	50,7%	90,3%
terceira	90,0%	79,3%	38,3%	85,2%
quarta	92,2%	66,4%	59,6%	84,8%
quinta	90,9%	68,3%	47,7%	83,2%
sexta	82,0%	82,0%	28,2%	74,2%
média	89,1%	72,3%	42,5%	82,7%

Como o algoritmo de segmentação por limiarização em YCbCr apresenta taxas de acerto percentual mais elevadas, com média próxima a 90%, não justifica-se o uso do algoritmo de pré-processamento para não aumentar o custo computacional do sistema.

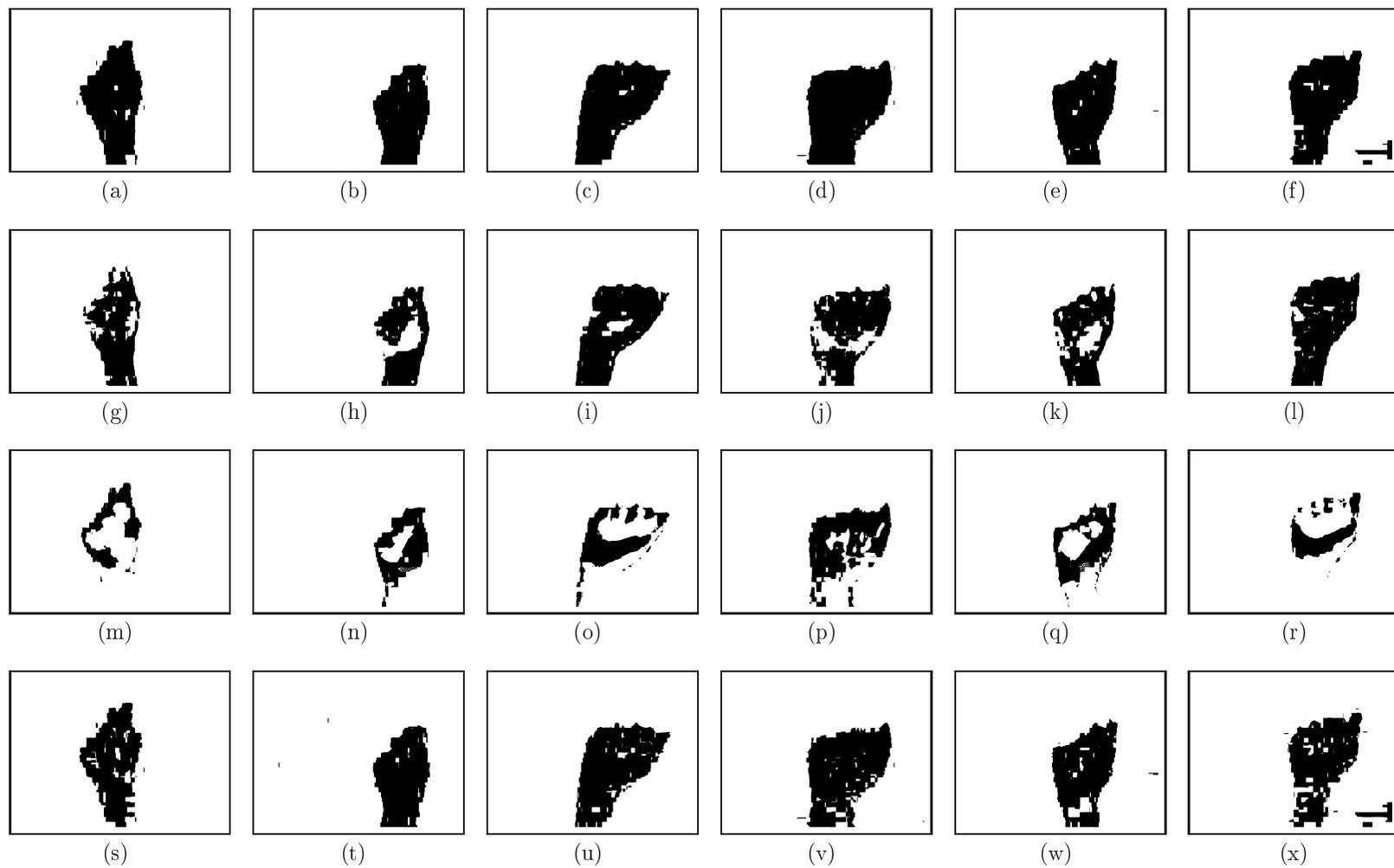


Figura 5.5: resultados da segmentação para os algoritmos (a) - (f) limiarização em YCbCr, (g) - (l) limiarização em HSV, (m) - (r) limiares em RGB e (s) - (x) modelo de probabilidade.

5.3 Resultados do pós-processamento

O pós-processamento é aplicado para corrigir imperfeições decorrentes de falhas na segmentação e adequar a imagem para a etapa seguinte. Como visto na seção anterior, o algoritmo de segmentação por limiarização em YCbCr apresenta os melhores resultados, com taxa de acerto percentual de 89,12%. Nesta seção, apenas os resultados da correção da segmentação obtidas por este algoritmo são analisados.

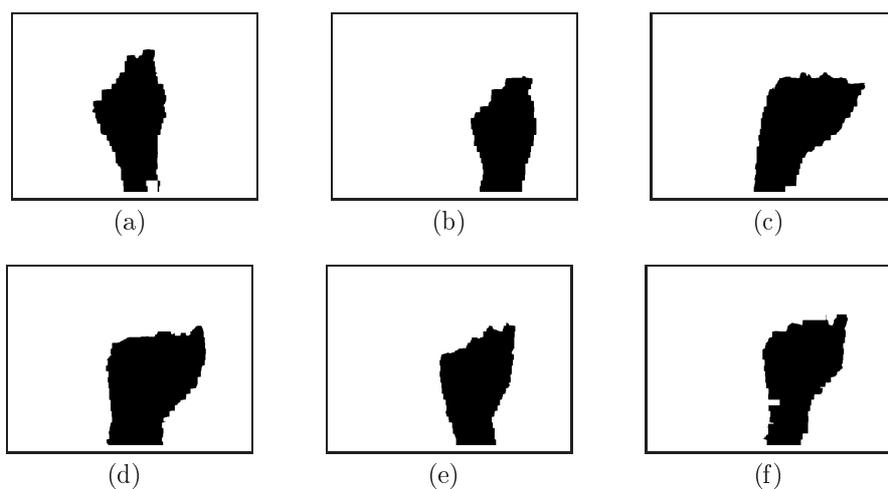


Figura 5.6: resultados da segmentação após a etapa de pós-processamento.

A correção dos erros da segmentação faz com que os vetores de atributos extraídos apresentem maior semelhança para um mesmo gesto, evitando dispersões entre vetores de atributos de uma mesma classe que ocorreriam devido a presença destes erros. A Tabela 5.2 apresenta os resultados obtidos após a aplicação do método de pós-processamento.

O resultado mostrado na Figura 5.6(f) justifica o uso do pós-processamento, visto que, conforme a Tabela 5.2, a melhora da segmentação é superior a 5%, observa-se a remoção de uma área fora da mão localizada no canto inferior direito da imagem que havia sido erroneamente segmentada como pele, vista na Figura 5.5(f). Falhas de segmentação desta natureza ocasionam discrepâncias nos atributos das imagens, de modo que a correção da segmentação é essencial.

5.4 Resultados da classificação

No processo de classificação, os dados correspondentes aos vetores de atributos de cada uma das imagens são submetidos à etapa de pré-classificação. Nesta etapa,

Tabela 5.2: resultados obtidos com o pós-processamento.

Pessoas	Segmentação	Com o	Ganho de
	original	pós-processamento	segmentação
primeira	87,1%	88,6%	1,5%
segunda	92,2%	92,6%	0,4%
terceira	90,0%	91,3%	1,3%
quarta	92,2%	92,6%	0,4%
quinta	90,9%	92,3%	1,4%
sexta	82,0%	87,6%	5,6%
média	89,1%	90,8%	1,7%

os dados são agrupados de acordo com a semelhança em *clusters*. Os *clusters* gerados compreendem os subconjuntos de dados que são submetidos à classificação final. Os critérios de avaliação dos atributos são, principalmente, a taxa de acerto percentual média na classificação dos gestos, consistindo na taxa de acerto percentual média final, o desvio padrão nas taxas de acerto e, com menor relevância, o número de *clusters* gerados, em que deseja-se ter o menor número de *clusters* possível.

A taxa de acerto percentual média na classificação dos gestos r_{gestos} é definida por¹

$$r_{gestos} = 100 \times \frac{\sum_{i=1}^{N_{clusters}} r_i N[C_i]}{\sum_{i=1}^{N_{clusters}} N[C_i]}, \quad (5.2)$$

em que $N_{clusters}$ corresponde ao número de *clusters*, r_i é a maior taxa de acerto obtida para cada algoritmo de classificação aplicado ao i -ésimo *cluster* e $N[C_i]$ é o número de elementos deste *cluster*.

5.4.1 Avaliação individual dos resultados

Os resultados obtidos para a primeira, segunda, terceira, quarta, quinta e sexta pessoas são apresentados nas Tabelas 5.3 à 5.8, respectivamente.

O melhor resultado da Tabela 5.3 com taxa de acerto média de 69,68% e 22

¹O cálculo do desvio padrão é feito de forma análoga ao da taxa de acerto média.

Tabela 5.3: resultados da classificação obtidos para a primeira pessoa.

Conjunto de atributos	Número de <i>clusters</i>	Taxas de acerto			Desvio padrão
		média	máxima	mínima	
Momentos Invariantes de Hu	22	69,68%	81,46%	54,61%	5,819%
10 Descritores de Fourier	28	46,70%	54,07%	40,15%	3,669%
30 Descritores de Fourier	16	56,08%	61,16%	49,29%	2,902%
50 Descritores de Fourier	22	61,97%	68,31%	54,77%	3,193%

clusters é obtido com a descrição das imagens pelos Momentos Invariantes de Hu.

Tabela 5.4: resultados da classificação obtidos para a segunda pessoa.

Conjunto de atributos	Número de <i>clusters</i>	Taxas de acerto			Desvio padrão
		média	máxima	mínima	
Momentos Invariantes de Hu	12	68,72%	77,21%	55,42%	4,628%
10 Descritores de Fourier	20	49,13%	56,09%	42,44%	3,263%
30 Descritores de Fourier	15	63,82%	69,57%	56,94%	2,870%
50 Descritores de Fourier	10	70,65%	75,73%	64,14%	2,575%

O melhor resultado da Tabela 5.4 com taxa de acerto média de 70,65% e 10 *clusters* é obtido com a descrição das imagens por 50 Descritores de Fourier.

Tabela 5.5: resultados da classificação obtidos para a terceira pessoa.

Conjunto de atributos	Número de <i>clusters</i>	Taxas de acerto			Desvio padrão
		média	máxima	mínima	
Momentos Invariantes de Hu	27	49,34%	65,85%	43,14%	7,178%
10 Descritores de Fourier	35	47,03%	50,84%	44,16%	1,531%
30 Descritores de Fourier	16	63,44%	69,56%	56,22%	3,527%
50 Descritores de Fourier	10	66,76%	72,12%	60,37%	2,954%

O melhor resultado da Tabela 5.5 com taxa de acerto média de 66,76% e 10 *clusters* é obtido com a descrição das imagens por 50 Descritores de Fourier.

Tabela 5.6: resultados da classificação obtidos para a quarta pessoa.

Conjunto de atributos	Número de <i>clusters</i>	Taxas de acerto			Desvio padrão
		média	máxima	mínima	
Momentos Invariantes de Hu	12	63,93%	84,78%	51,38%	8,967%
10 Descritores de Fourier	13	46,41%	54,52%	38,97%	3,458%
30 Descritores de Fourier	14	69,80%	75,31%	62,87%	2,801%
50 Descritores de Fourier	10	73,05%	77,73%	67,50%	2,152%

O melhor resultado da Tabela 5.6 com taxa de acerto média de 73,05% e 10 *clusters* é obtido com a descrição das imagens por 50 Descritores de Fourier.

Tabela 5.7: resultados da classificação obtidos para a quinta pessoa.

Conjunto de atributos	Número de <i>clusters</i>	Taxas de acerto			Desvio padrão
		média	máxima	mínima	
Momentos Invariantes de Hu	11	78,47%	92,20%	65,58%	7,874%
10 Descritores de Fourier	11	55,75%	62,64%	48,23%	3,420%
30 Descritores de Fourier	10	77,29%	81,72%	71,11%	2,283%
50 Descritores de Fourier	10	82,67%	88,27%	75,52%	2,743%

O melhor resultado da Tabela 5.7 com taxa de acerto média de 82,67% e 10 *clusters* é obtido com a descrição das imagens por 50 Descritores de Fourier.

Tabela 5.8: resultados da classificação obtidos para a sexta pessoa.

Conjunto de atributos	Número de <i>clusters</i>	Taxas de acerto			Desvio padrão
		média	máxima	mínima	
Momentos Invariantes de Hu	14	77,10%	87,07%	69,81%	4,732%
10 Descritores de Fourier	13	48,94%	55,58%	41,89%	3,311%
30 Descritores de Fourier	15	71,03%	77,08%	63,27%	2,802%
50 Descritores de Fourier	11	72,67%	80,54%	61,00%	4,253%

O melhor resultado da Tabela 5.8 com taxa de acerto média de 77,10% e 14 *clusters* é obtido com a descrição das imagens pelos Momentos Invariantes de Hu.

Como apresentado na Seção 2.2, a segmentação é uma etapa crítica do sistema, porém, observando-se os resultados das Tabelas 5.3 à 5.8, que uma melhor segmentação não implica em melhores resultados de classificação.

5.4.2 Avaliação geral dos atributos

Dentre os quatro conjuntos de atributos, os que apresentam melhores resultados são os Momentos Invariantes de Hu e 50 Descritores de Fourier, conforme apresentado na Tabela 5.9.

Tabela 5.9: resultados da classificação para cada um dos conjuntos de atributos.

Conjunto de atributos	Número médio de <i>clusters</i>	Taxas de acerto			Desvio padrão
		média	máxima	mínima	
Momentos Invariantes de Hu	16,3	67,87%	81,43%	56,66%	6,533%
10 Descritores de Fourier	20,0	48,99%	55,62%	42,64%	3,109%
30 Descritores de Fourier	14,3	66,92%	72,40%	59,95%	2,284%
50 Descritores de Fourier	12,1	71,29%	77,12%	63,88%	2,798%

Além de apresentar as maiores taxas de acerto médias segundo a Tabela 5.9, os Momentos Invariantes de Hu e os 50 Descritores de Fourier apresentam os melhores resultados para todas as pessoas que serviram de modelo, sendo que os 50 Descritores de Fourier contribuem com dois terços dos melhores resultados e os Momentos Invariantes de Hu com um terço dos melhores resultados.

Entre os 50 Descritores de Fourier e os Momentos Invariantes de Hu, os primeiros são os atributos mais adequados para representar as imagens, tendo em vista a estabilidade das taxas de acerto, pois em todos os casos os desvios padrões obtidos pelos 50 Descritores de Fourier são menores que os obtidos pelos Momentos Invariantes de Hu.

Outra vantagem dos Descritores de Fourier é que estes são obtidos pelo contorno da mão, o que os torna menos dependentes da segmentação, desde que esta apresente resultados semelhantes para um mesmo conjunto de imagens.

5.5 Discussão dos resultados da classificação

Os conjuntos atributos que melhor atenderam as condições especificadas foram os Momentos Invariantes de Hu e a representação por cinquenta descritores de Fourier. Nas Tabelas 5.10, 5.11 e 5.12 são apresentados os *clusters* obtidos para as descrições pelos Momentos Invariantes de Hu para a segunda pessoa, por 50 Descritores de Fourier para a quinta e a sexta pessoas, destacando-se as classes compreendidas por cada *cluster* e o percentual de elementos de cada classe por *cluster* (valor entre parênteses).

Tabela 5.10: *clusters* obtidos com a representação pelos Momentos Invariantes de Hu para a segunda pessoa.

Índices dos <i>clusters</i>	Classes compreendidas					
1	-	-	-	-	-	-
2	J (52)	-	-	-	-	-
3	-	-	-	-	-	-
4	F (88)	-	-	-	-	-
5	K (16)	R (92)	U (98)	-	-	-
6	G (90)	H (56)	K (68)	P (74)	T (98)	Z (46)
7	H (34)	J (12)	L (92)	V (100)	W (98)	Z (34)
8	A (98)	B (100)	C (32)	I (92)	M (94)	S (96)
9	E (100)	X (96)	Y (24)	-	-	-
10	C (68)	N (96)	Q (96)	Y (60)	-	-
11	D (100)	-	-	-	-	-
12	J (18)	O (96)	-	-	-	-

As taxas de acerto médias relativas aos dados das Tabelas 5.10 e 5.12 são medianas, próximas a 70%, enquanto que a taxa de acerto média relativa aos dados da Tabela 5.11 corresponde ao melhor resultado obtido pelo sistema, com quase 83% de acerto.

5.6 Concepção do sistema baseada nos resultados

Os resultados dos diversos algoritmos apresentados neste capítulo permitem concluir quais os algoritmos mais adequados a cada uma das etapas do sistema

Tabela 5.11: *clusters* obtidos com a representação por cinquenta descritores de Fourier para a quinta pessoa.

Índices dos <i>clusters</i>	Classes compreendidas							
	1	A (96)	I (94)	J (54)	M (100)	N (100)	Q (98)	S (100)
2	B (96)	E (96)	U (100)	-	-	-	-	-
3	C (88)	D (82)	H (52)	K (22)	O (92)	V (82)	W (100)	-
4	H (12)	Y (54)	-	-	-	-	-	-
5	P (34)	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-
7	P (42)	-	-	-	-	-	-	-
8	H (16)	L (80)	-	-	-	-	-	-
9	F (96)	G (96)	J (18)	K (70)	R (84)	T (90)	-	-
10	J (12)	X (84)	Y (12)	Z (74)	-	-	-	-

Tabela 5.12: *clusters* obtidos com a representação por cinquenta descritores de Fourier para a sexta pessoa.

Índices dos <i>clusters</i>	Classes compreendidas							
	1	P (10)	-	-	-	-	-	-
2	P (32)	-	-	-	-	-	-	-
3	J (8)	X (90)	-	-	-	-	-	-
4	B (100)	F (84)	H (96)	J (14)	K (78)	S (76)	T (62)	W (28)
5	C (26)	G (64)	U (50)	V (88)	W (68)	-	-	-
6	L (74)	Y (12)	-	-	-	-	-	-
7	C (28)	O (24)	P (34)	-	-	-	-	-
8	C (24)	L (10)	O (64)	Y (22)	-	-	-	-
9	D (18)	G (26)	R (76)	U (44)	Y (48)	-	-	-
10	A (78)	D (72)	I (98)	J (58)	Q (76)	T (28)	-	-
11	E (94)	M (98)	N (98)	Z (78)	-	-	-	-

SIRENE. A Figura 5.7 apresenta o diagrama com os algoritmos que forneceram os melhores resultados.

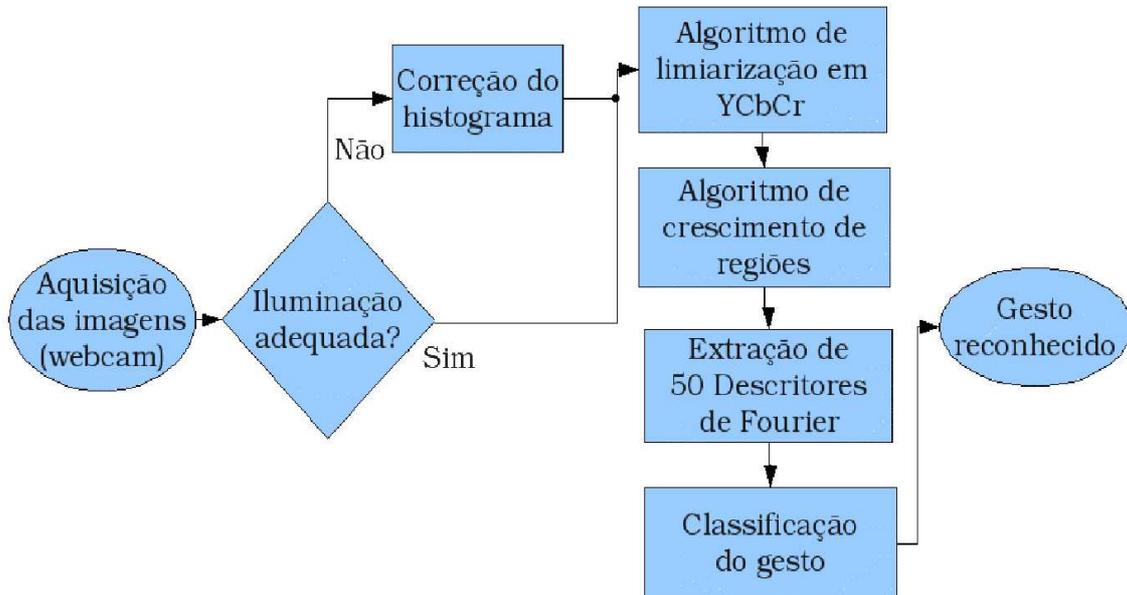


Figura 5.7: sistema SIRENE.

No diagrama da Figura 5.7, o bloco de **Classificação do gesto** não apresenta os algoritmos específicos, pois, consiste na aplicação da rede neural SOM para pré-classificação do gesto seguida das redes neurais supervisionadas. As RNAs supervisionadas apresentam resultados diferentes para cada *cluster* da rede SOM, todavia percebe-se que a rede neural RBF não apresenta resultados superiores aos das redes neurais PS e MLP que justifiquem seu uso neste sistema, sendo então descartada. Neste caso, ambas as redes neurais são aplicadas em cada um dos *clusters*, para o caso da MLP, variando-se o número de neurônios ocultos, a fim de se determinar a melhor configuração de rede neural para cada *cluster*.

Uma vez determinados os algoritmos empregados no sistema, define-se a forma de utilizá-los de maneira coerente, como a seguir:

- i. com a *webcam* devidamente conectada, o sistema é iniciado;
- ii. são capturadas as imagens do usuário para cada gesto do alfabeto da LIBRAS;
- iii. as imagens são submetidas ao algoritmo de segmentação da pele;

- iv. se o algoritmo de segmentação da pele fornece resultados satisfatórios, as imagens são submetidas à etapa de pós-processamento, se não as imagens são submetidas à etapa de pré-processamento (algoritmo de correção de histograma), depois segmentadas para, em seguida, serem submetidas à etapa de pós-processamento (algoritmo de crescimento de regiões);
- v. após o pós-processamento, são extraídos os 50 Descritores de Fourier das imagens;
- vi. tendo os atributos de todas as imagens, estes são normalizados para o intervalo $[0, 1]$ e submetidos à pré-classificação;
- vii. de posse dos *clusters*, estes são submetidos às redes neurais PS e MLP para determinar-se a arquitetura que forneça os melhores resultados (maior média de acertos com o menor desvio padrão) para cada um dos *clusters*;
- viii. com o sucesso da classificação final, o sistema está pronto para ser utilizado.

Na condição atual, o sistema como proposto consiste de um conjunto de técnicas e algoritmos capazes de fazer o reconhecimento dos gestos da LIBRAS de modo *off-line*.

Com isto, o sistema proposto nesta dissertação está completamente descrito e avaliado, o capítulo a seguir apresenta as conclusões e contribuições acerca desta dissertação e propõe sugestões de trabalhos futuros.

Capítulo 6

Conclusões, Contribuições e Trabalhos Futuros

Esta dissertação propõe e descreve um sistema de visão computacional para reconhecimento de gestos do alfabeto da Língua de Sinais Brasileira (LIBRAS). Desenvolvido e baseado em plataformas de *software* livre e *hardware* de baixo custo, para torná-lo acessível ao maior contingente possível da população.

O sistema é implementado no ambiente de simulação Scilab, fazendo uso da *toolbox* SIVp. O sistema proposto é concebido a partir de um conjunto de imagens geradas nas dependências da Federação Nacional de Educação e Integração de Surdos (FENEIS-CE) e da Associação de Pais e Amigos de Deficientes Auditivos (APADA-CE) com profissionais do uso e/ou ensino da LIBRAS, estas imagens são submetidas a diversos algoritmos de realce, segmentação da pele humana, pós-processamento e extração de atributos, os atributos extraídos das imagens são utilizados em algoritmos de classificação baseados em redes neurais, afim de determinar-se quais os algoritmos que fornecem os melhores resultados.

O algoritmo de realce proposto neste trabalho é baseado na técnica de especificação de histograma e é empregado quando o ambiente em que o sistema está sendo utilizado não tem condições de iluminação adequadas. Para a segmentação da pele, são avaliados quatro algoritmos sendo três de segmentação por limiarização e o outro baseado em um modelo de probabilidade. Os algoritmos de pós-processamento condicionam as imagens segmentadas à etapa de extração de atributos, para esta são avaliados três conjuntos de Descritores de Fourier e os Momentos Invariantes de Hu.

Os atributos extraídos das imagens são classificados em *clusters* por uma rede neural SOM e cada *cluster* é submetido a três algoritmos de classificação supervisionada, em que cada um é baseado em uma das redes neurais: PS, MLP e RBF.

Após a análise de cada um dos algoritmos, é feita a seleção dos algoritmos mais eficientes, de modo que o sistema proposto apresenta taxas de acerto na ordem de 70% nos testes realizados para todas as pessoas que serviram de modelo para geração das imagens, comprovando a adequação do sistema para reconhecimento dos gestos do alfabeto da LIBRAS. Além disso, a forma como o sistema SIRENE é concebido possibilita que seja usado ou adaptado para fazer o reconhecimento de outros conjuntos de gestos.

Assim, o sistema tem condições de ser usado com sucesso, inclusive, em outras interfaces de interação homem-máquina baseada em comunicação gestual e visual, auxiliando pessoas incapazes de interagir com computadores de outra forma.

As principais contribuições deste trabalho são: a análise dos algoritmos de segmentação de pele, com a proposta de um método de realce que aumenta a eficácia dos algoritmos de segmentação, em procurou-se abordar técnicas conhecidas da literatura que fossem de baixo custo computacional. Outra contribuição são a proposta e a análise do método de pós-processamento baseado no algoritmo de crescimento de regiões aplicado a imagens binárias para correção de falhas na segmentação, o que melhora significativamente o resultado da segmentação, o uso das imagens em níveis de cinza para obtenção dos Momentos Invariantes de Hu, possibilitando o reconhecimento de gestos estáticos com formas semelhantes, e o método de pré-classificação dos gestos que possibilita o reconhecimento de quantidades elevadas de gestos distintos.

6.1 Trabalhos Futuros

O sistema proposto nesta dissertação abre possibilidades de desenvolvimento de várias tecnologias de auxílio a pessoas com limitações físicas, por se tratar de uma interface de interação homem-máquina baseada em visão computacional, o que é base para diversos sistemas de acessibilidade. Outros trabalhos futuros podem ser desenvolvidos a partir deste, ou ainda partes do sistema proposto podem ser usadas em outros sistemas, como a segmentação da pele com os algoritmos de realce e/ou pós-processamento e os descritores geométricos.

A evolução do sistema proposto, a partir do estado atual, dar-se-á com as seguintes atividades:

- ▶ implementação dos algoritmos em linguagens de programação C/C++;
- ▶ avaliar o sistema em condições de uso em tempo-real;
- ▶ estudar e avaliar outros modelos e abordagens para o pré-processamento, a fim de conceber uma técnica de correção adaptativa da iluminação;
- ▶ avaliar outras formas de atributos para descrição das imagens e dos movimentos de cada gesto;
- ▶ estudar outras abordagens de classificadores a fim de aumentar as taxas de acerto obtidas;
- ▶ disponibilizar o sistema funcional para uso da sociedade sob licença *freeware*;
- ▶ estudar os impactos sociais do sistema para a comunidade surda.

Assim é concluída esta dissertação, com a graça de Deus e as bênçãos de Jesus e Maria.

Apêndice A

LIBRAS

No Brasil, a LIBRAS é a linguagem natural para a comunidade surda. As instituições que divulgam e ensinam a LIBRAS são:

- ▶ Instituto Nacional de Educação de Surdos (INES);
- ▶ Federação Nacional de Educação e Integração de Surdos (Feneis);
- ▶ Associação de Pais e Amigos de Deficientes Auditivos (Apada).

A LIBRAS é uma linguagem gestual, que compreende gestos e expressões faciais, manuais e corporais para compor as sentenças (SOUZA; VIEIRA, 2006). Com relação aos gestos das mãos, a LIBRAS possui um conjunto com 46 diferentes gestos, ressaltando que as sentenças formadas apenas com gestos manuais podem diferir apenas no movimento feito pelas mãos (DIAS; SOUZA; PISTORI, 2006). A Figura A.1 apresenta a frase "eu te amo" em LIBRAS gerado pelo *software* Rybená (RYBENÁ, 2009).

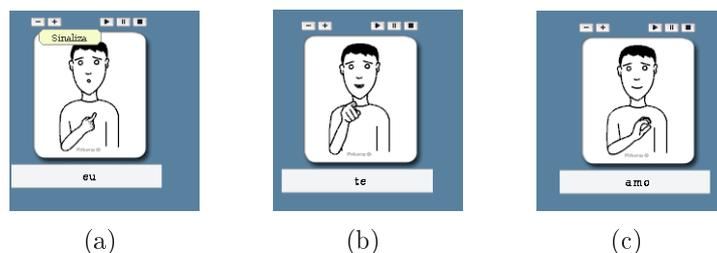


Figura A.1: gestos correspondentes as palavras (a) "eu", (b) "te" e (c) "amo" em LIBRAS.

A LIBRAS é uma língua muito rica que inclui gestos para representar desde as letras do alfabeto, foco deste trabalho, até expressões idiomáticas completas. A seguir são apresentadas características específicas do alfabeto da LIBRAS.

A.1 Alfabeto da LIBRAS

Com relação ao alfabeto da LIBRAS, este é constituído por vinte gestos estáticos, sem movimento, e seis gestos dinâmicos, que apresentam movimento. Os gestos estáticos correspondem às letras {A, B, C, D, E, F, G, I, L, M, N, O, P, Q, R, S, T, U, V, W}, apresentados, respectivamente, nas Figuras A.2(a) à A.2(t), e os gestos dinâmicos correspondem às letras {H, J, K, X, Y, Z}, apresentados na Figura A.3.

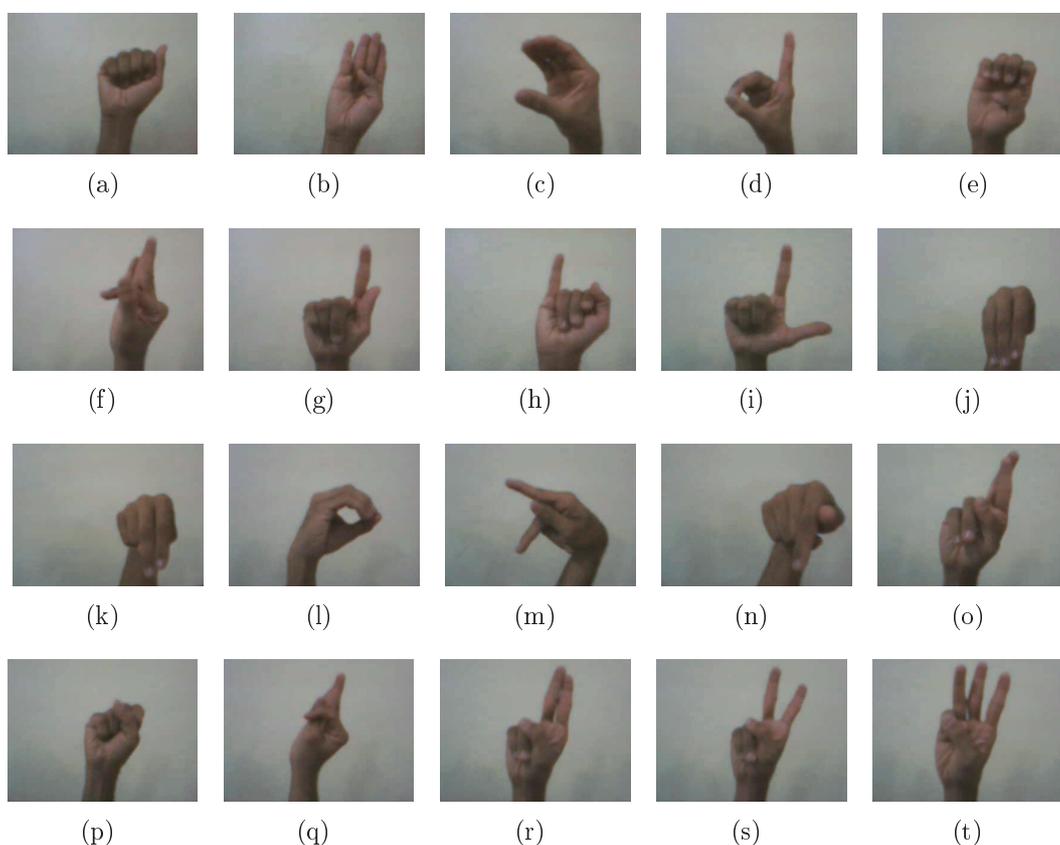


Figura A.2: imagens relativas as letras do alfabeto da LIBRAS representadas por gestos estáticos.

Capovilla e Raphael (2005) apresenta as características de cada gesto para uma coletânea de aproximadamente três mil verbetes, dentre os quais estão as letras do alfabeto da LIBRAS, as quais são descritas conforme apresentado a seguir:

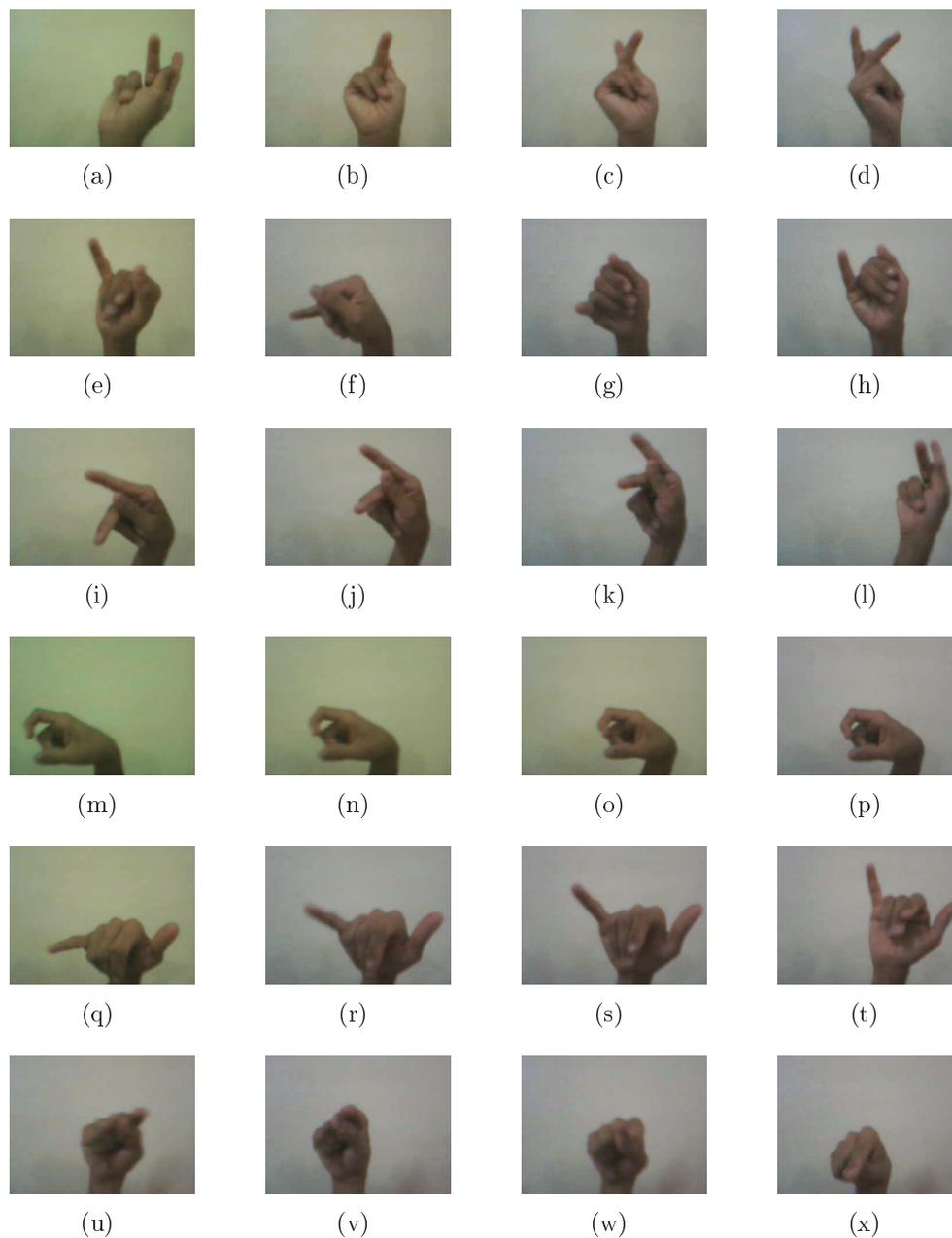


Figura A.3: imagens relativas as letras do alfabeto da LIBRAS representadas por gestos dinâmicos. (a)-(d) **H**, (e)-(h) **J**, (i)-(l) **K**, (m)-(p) **X**, (q)-(t) **Y** e (u)-(x) **Z**.

- ▶ a letra 'A' é representada com a mão direita vertical fechada, palma para frente e polegar tocando a lateral do indicador;
- ▶ a letra 'B' é representada com a mão direita vertical aberta, dedos unidos e polegar dobrado contra a palma;
- ▶ a letra 'C' é representada com a mão direita vertical, palma para a esquerda,

dedos unidos e curvados e polegar paralelo aos demais dedos e também curvado;

- ▶ a letra 'D' é representada com a mão direita vertical, palma para a esquerda, indicador distendido, demais dedos curvados contra a palma e o polegar fechado sobre eles;
- ▶ a letra 'E' é representada com a mão direita vertical, palma para frente dedos curvados contra a palma e polegar dobrado contra a palma;
- ▶ a letra 'F' é representada com a mão direita vertical aberta, palma para frente, dedo indicador flexionado e o polegar tocando a lateral do indicador;
- ▶ a letra 'G' é representada com a mão direita vertical fechada, palma para frente, dedo indicador distendido e polegar tocando a lateral do indicador;
- ▶ a letra 'H' é representada com a mão direita vertical fechada, palma para frente, dedos indicador e médio distendidos com o médio inclinado para frente e o polegar tocando a lateral deste. Girar a mão pelo pulso para dentro;
- ▶ a letra 'I' é representada com a mão direita vertical, palma para frente, dedo mínimo distendido demais dedos fechados contra a palma e polegar fechado sobre estes;
- ▶ a letra 'J' é representada com a mão direita em "I", palma para frente e inclinada para baixo. Mover a mão para baixo, girá-la pelo pulso para a esquerda e então virar a palma para cima;
- ▶ a letra 'K' é representada com a mão direita vertical fechada, palma para frente, dedos indicador e médio distendidos com o médio inclinado para frente e o polegar tocando entre o indicador e o médio. Mover a mão para cima;
- ▶ a letra 'L' é representada com a mão direita vertical fechada, palma para frente, dedos indicador e polegar distendidos com o polegar apontando para a esquerda;
- ▶ a letra 'M' é representada com a mão direita vertical, palma para dentro, dedos indicador, médio e anular distendidos para baixo e demais dedos fechados contra a palma;

- ▶ a letra 'N' é representada com a mão direita vertical fechada, palma para dentro, dedos indicador e médio distendidos para baixo;
- ▶ a letra 'O' é representada com a mão direita vertical, palma para a esquerda, dedos unidos e curvados tocando a ponta do polegar;
- ▶ a letra 'P' é representada com a mão direita horizontal, palma para baixo, dedos anular e mínimo fechados, médio apontando para baixo, indicador apontando para frente e palma do polegar tocando a lateral do dedo médio;
- ▶ a letra 'Q' é representada com a mão direita vertical fechada, palma para dentro, dedo indicador distendido e apontando para baixo e polegar tocando a lateral do indicador;
- ▶ a letra 'R' é representada com a mão direita vertical fechada, palma para frente, dedos indicador e médio distendidos e entrelaçados com o dorso do indicador tocando a palma do médio;
- ▶ a letra 'S' é representada com a mão direita horizontal fechada, palma para a esquerda e palma do polegar tocando os demais dedos;
- ▶ a letra 'T' é representada com a mão direita vertical aberta, palma para frente e indicador flexionado sobre o dorso do polegar;
- ▶ a letra 'U' é representada com a mão direita vertical fechada, palma para frente, dedos indicador e médio distendidos e unidos;
- ▶ a letra 'V' é representada com a mão direita vertical fechada, palma para frente, dedos indicador e médio distendidos e afastados um do outro;
- ▶ a letra 'W' é representada com a mão direita vertical fechada, palma para frente, dedos indicador, médio e anular distendidos e afastados um do outro;
- ▶ a letra 'X' é representada com a mão direita horizontal fechada, palma para a esquerda e indicador distendido e curvado. Mover a mão, ligeiramente, para trás;
- ▶ a letra 'Y' é representada com a mão direita vertical fechada, palma para frente, dedos mínimo e polegar distendidos. Mover a mão em um pequeno arco para frente e para cima;

- a letra 'Z' é representada com a mão direita em "1"horizontal¹, palma para baixo, indicador apontando para frente. Mover a mão ligeiramente para a direita, diagonalmente, para a esquerda e para baixo, e então para a direita novamente.

Observando as descrições e figuras apresentadas, percebe-se uma semelhança forte entre as letras que compreendem os conjuntos {A, E, S}, {D, G, Z}, {F, T}, {M, N, Q}, {I, J} e {H, K, V}; para os dois últimos conjuntos, a diferença entre cada gesto é principalmente evidenciada pelo movimento.

¹A mão em "1" é representada com a mão direita vertical fechada, palma para dentro e indicador distendido para cima

Análise dos classificadores supervisionados

Neste apêndice são analisados os resultados dos classificadores supervisionados, apresentados nas seções 2.5.2, 2.5.3 e 2.5.4, para os *clusters* contidos nas Tabelas 5.10 e 5.11. Em todos os testes realizados neste trabalho, os classificadores supervisionados são avaliados cem vezes, cada uma com cem épocas de treinamento e utilizando taxas de aprendizagem de $\alpha = 0,01$, estes valores são adotados baseando-se nos testes e nas recomendações da literatura.

Os classificadores baseados nas redes neurais MLP e RBF apresentam resultados diferentes quando o número de neurônios ocultos é alterado, então, para cada *cluster*, a quantidade de neurônios ocultos é variada conforme as equações 4.11 e 4.12.

Os resultados obtidos, para cada configuração das redes neurais MLP e RBF, são comparados entre si e com o resultado da rede neural PS, caso estes apresentem semelhança, a ordem de preferência é PS, MLP e RBF, obedecendo o custo computacional de cada algoritmo.

As figuras apresentadas a seguir mostram os gráficos que correspondem às taxas de acerto percentual média, máxima e mínima e ao desvio padrão para cada número de neurônios ocultos e as tabelas apresentam as matrizes de confusão média para cada *cluster* que compreenda mais de duas classes.

As informações que seguem apresentam as taxas de acerto percentual dos classificadores correspondentes às redes neurais MLP e RBF, variando o número de neurônios na camada oculta, para cada *cluster*, e as Tabelas B.1 à B.7 apresentam

as matrizes de confusão referentes aos melhores resultados obtidos para cada *cluster* da Tabela 5.10, Capítulo 5.

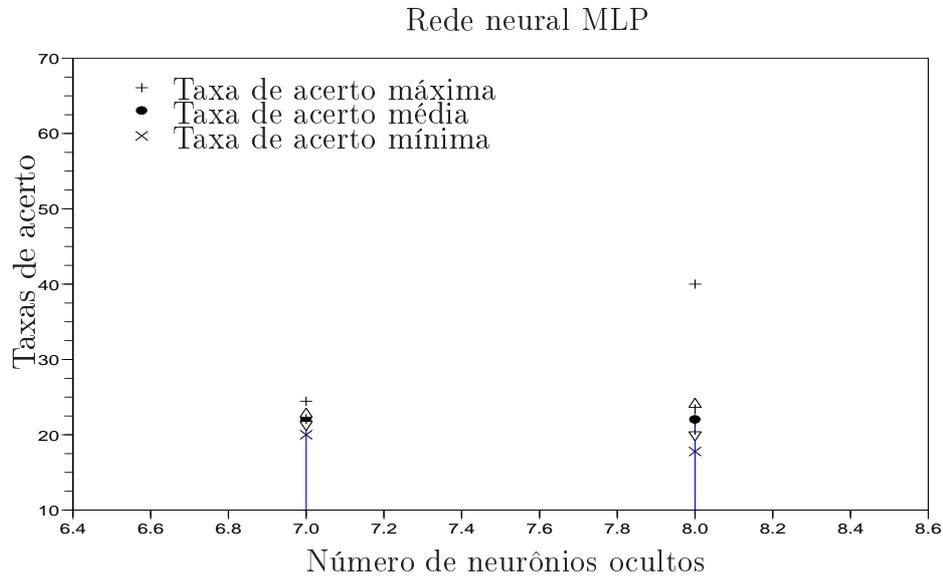
Os resultados das redes neurais MLP e RBF para o quinto *cluster* da Tabela 5.10 são ambos constantes e iguais a 45,45% para todas as quantidades de neurônios ocultos.

O classificador baseado na rede neural PS apresenta 59,36% de acerto para o quinto *cluster* da Tabela 5.10. De modo que, o melhor resultado é o obtido com a rede neural PS, cuja matriz de confusão está apresentada na Tabela B.1.

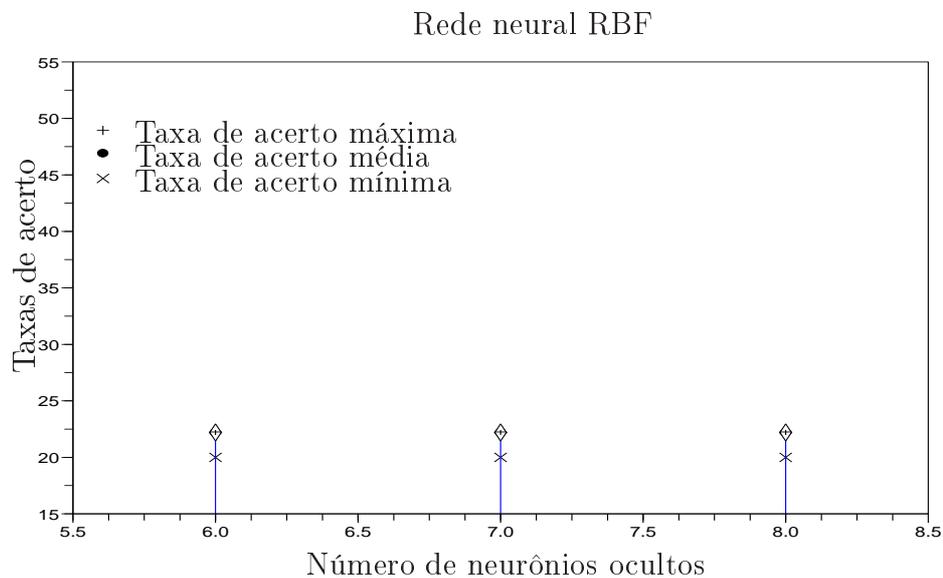
Tabela B.1: matriz de confusão para o quinto *cluster* da Tabela 5.10 com a rede neural PS.

Saídas	Entradas		
	K	R	U
K	0,00	0,00	0,00
R	0,47	4,14	1,08
U	1,53	5,86	8,92

Os resultados das redes neurais MLP e RBF para o sexto *cluster* da Tabela 5.10 são apresentados na Figura B.1.



(a)



(b)

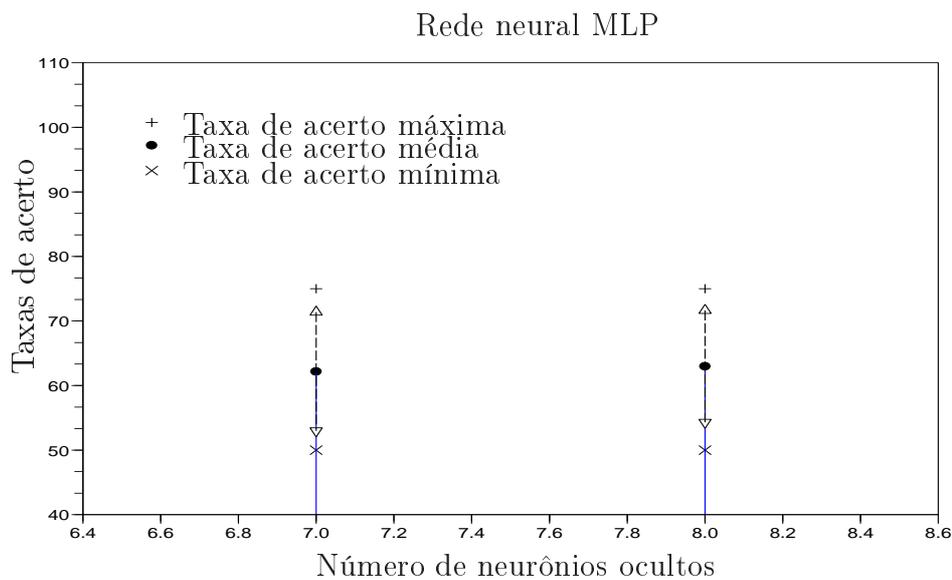
Figura B.1: resultados das redes neurais (a) MLP e (b) RBF para o sexto *cluster* da Tabela 5.10.

O classificador baseado na rede neural PS apresenta 53,02% de acerto para o sexto *cluster* da Tabela 5.10. Assim, o melhor resultado é obtido com a rede neural PS, cuja matriz de confusão está apresentada na Tabela B.2.

Tabela B.2: matriz de confusão para o sexto *cluster* da Tabela 5.10 com a rede neural PS.

Saídas	Entradas					
	G	H	K	P	T	Z
G	7,26	0,00	6,20	0,23	0,16	3,49
H	0,00	0,00	0,00	0,00	0,00	0,00
K	0,00	0,00	0,00	0,00	0,00	0,00
P	0,00	0,00	0,04	6,93	0,17	0,00
T	1,74	6,00	0,76	0,84	9,67	1,51
Z	0,00	0,00	0,00	0,00	0,00	0,00

Os resultados da rede neural MLP para o sétimo *cluster* da Tabela 5.10 são apresentados na Figura B.2, os resultados da rede neural RBF para este *cluster* são constantes e iguais a 25,00%.

Figura B.2: resultados da rede neural MLP para o sétimo *cluster* da Tabela 5.10.

O classificador baseado na rede neural PS apresenta 75,00% de acerto para o sétimo *cluster* da Tabela 5.10. De modo que, o melhor resultado é o obtido com a rede neural PS, cuja matriz de confusão está apresentada na Tabela B.3.

Tabela B.3: matriz de confusão para o sétimo *cluster* da Tabela 5.10 com a rede neural PS.

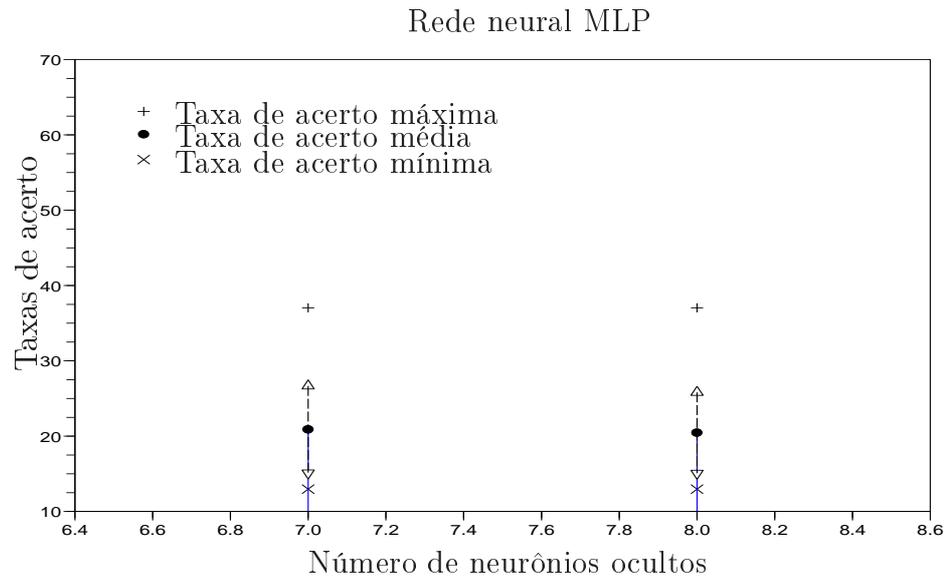
Saídas	Entradas					
	H	J	L	V	W	Z
H	0,00	0,00	0,00	0,00	0,00	0,00
J	0,00	0,00	0,00	0,00	0,00	0,00
L	0,04	0,00	10,0	0,00	0,00	0,00
V	2,40	0,00	0,00	10,0	0,00	0,00
W	1,56	2,00	0,00	0,00	10,0	4,00
Z	0,00	0,00	0,00	0,00	0,00	0,00

Os resultados das redes neurais MLP e RBF para o oitavo *cluster* da Tabela 5.10 são apresentados na Figura B.3.

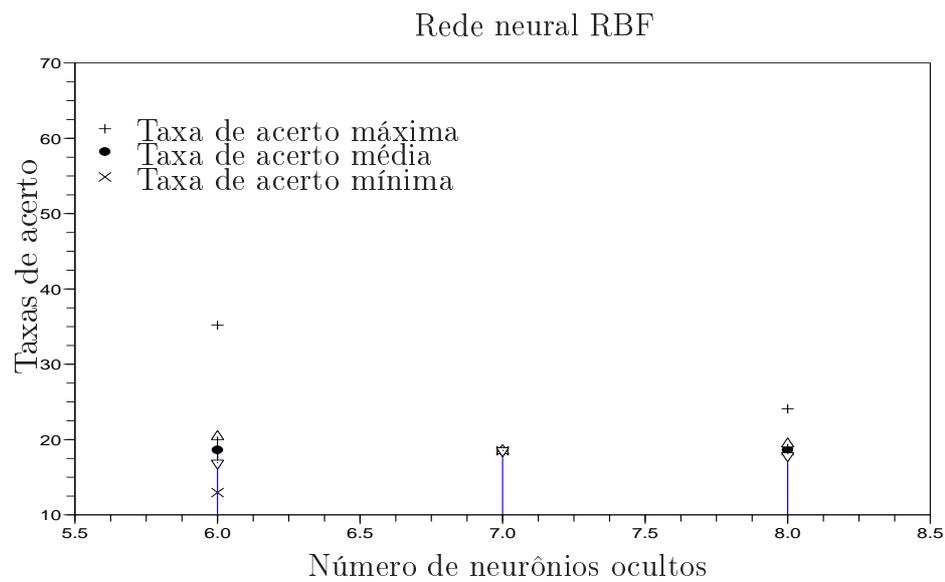
O classificador baseado na rede neural PS apresenta 48,38% de acerto para o oitavo *cluster* da Tabela 5.10. Desta forma, o melhor resultado é o obtido com a rede neural PS, cuja matriz de confusão está apresentada na Tabela B.4.

Tabela B.4: matriz de confusão para o oitavo *cluster* da Tabela 5.10 com a rede neural PS.

Saídas	Entradas					
	A	B	C	I	M	S
A	2,97	0,79	6,20	1,61	0,00	5,65
B	4,24	9,21	0,00	0,78	0,00	2,38
C	0,00	0,00	0,00	0,00	0,00	0,00
I	0,00	0,00	0,33	2,01	0,00	0,00
M	2,79	0,00	3,41	0,27	10,0	0,03
S	0,00	0,00	0,26	5,33	0,00	1,94



(a)

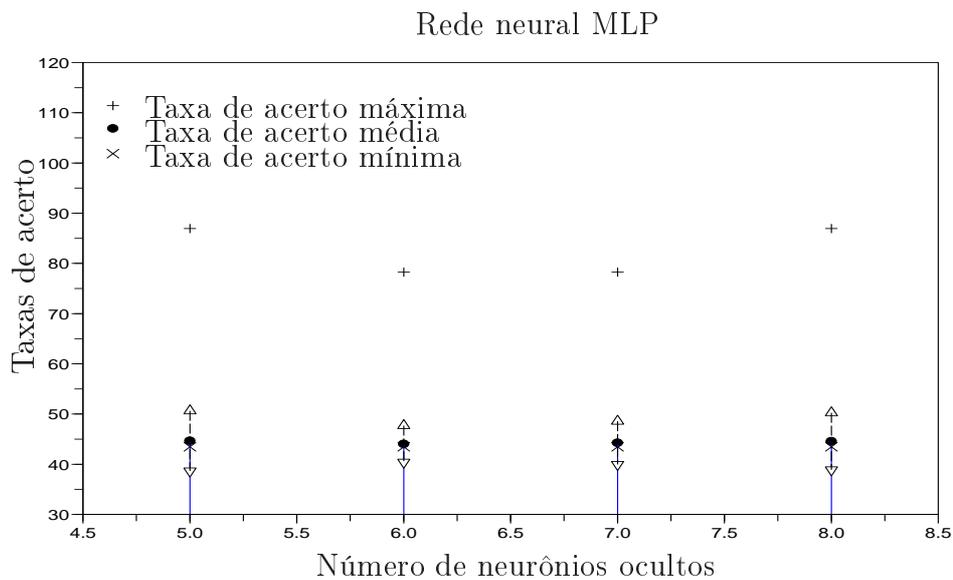


(b)

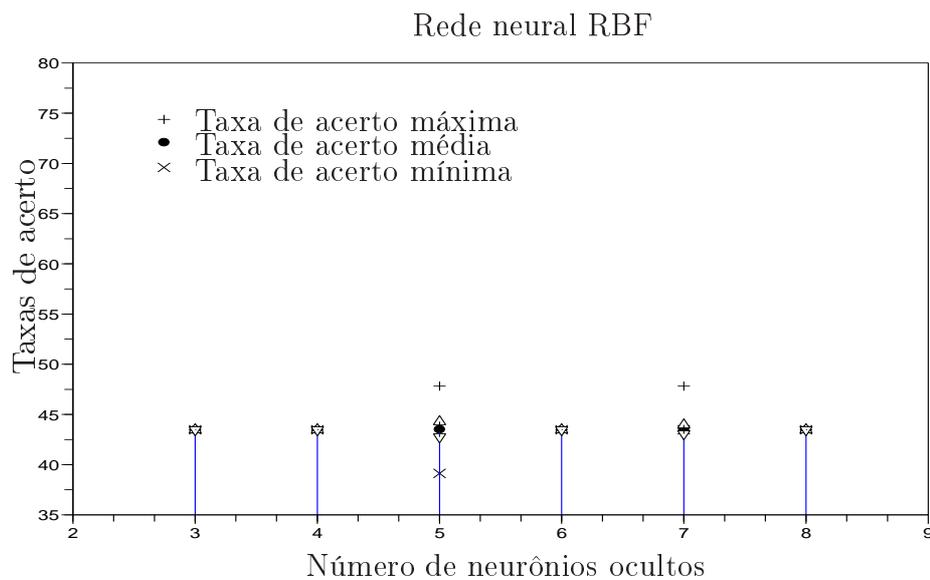
Figura B.3: resultados das redes neurais (a) MLP e (b) RBF para o oitavo *cluster* da Tabela 5.10.

Os resultados das redes neurais MLP e RBF para o nono *cluster* da Tabela 5.10 são apresentados na Figura B.4.

O classificador baseado na rede neural PS apresenta 86,52% de acerto para o nono *cluster* da Tabela 5.10. Assim, o melhor resultado é o obtido com a rede neural



(a)



(b)

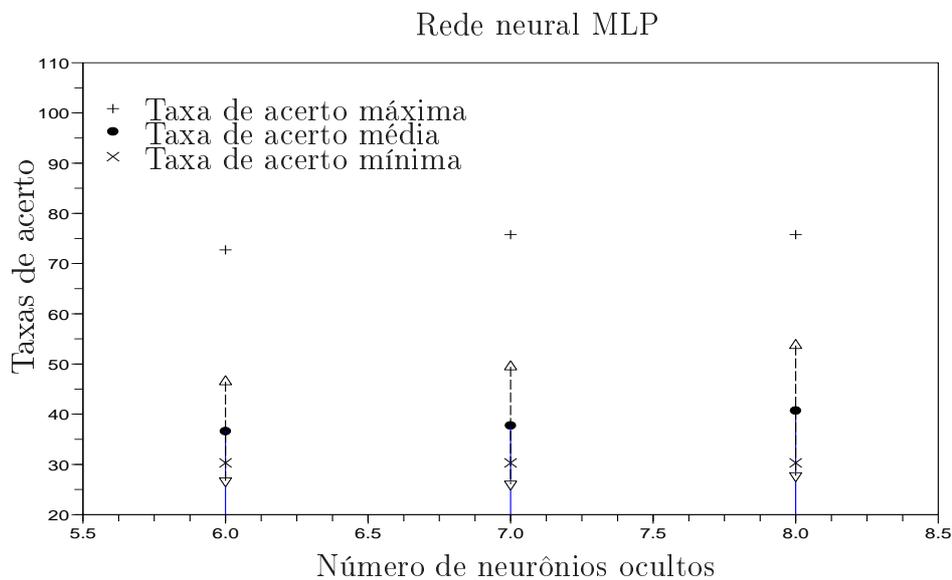
Figura B.4: resultados das redes neurais (a) MLP e (b) RBF para o nono *cluster* da Tabela 5.10.

PS, cuja matriz de confusão está apresentada na Tabela B.5.

Tabela B.5: matriz de confusão para o nono *cluster* da Tabela 5.10 com a rede neural PS.

Saídas	Entradas		
	E	X	Y
E	10,0	0,10	0,00
X	0,00	9,90	3,00
Y	0,00	0,00	0,00

Os resultados da rede neural MLP para o décimo *cluster* da Tabela 5.10 são apresentados na Figura B.5, os resultados da rede neural RBF para este *cluster* são constantes e iguais a 30,30%.

Figura B.5: resultados da rede neural MLP para o décimo *cluster* da Tabela 5.10.

O classificador baseado na rede neural PS apresenta 80,66% de acerto para o décimo *cluster* da Tabela 5.10. De modo que, o melhor resultado é o obtido com a rede neural PS, cuja matriz de confusão está apresentada na Tabela B.6.

Tabela B.6: matriz de confusão para o décimo *cluster* da Tabela 5.10 com a rede neural PS.

Saídas	Entradas			
	C	N	Q	Y
C	0,95	0,00	0,00	0,00
N	2,86	10,0	0,33	0,00
Q	2,94	0,00	9,67	0,00
Y	0,25	0,00	0,00	6,00

Os resultados das redes neurais MLP e RBF para o décimo segundo *cluster* da Tabela 5.10 são ambos constantes e iguais a 83,33% para todas as quantidades de neurônios ocultos.

O classificador baseado na rede neural PS também apresenta 83,33% de acerto para o décimo segundo *cluster* da Tabela 5.10. De modo que o melhor resultado é o obtido com a rede neural PS, cuja matriz de confusão está apresentada na Tabela B.7.

Tabela B.7: matriz de confusão para o décimo segundo *cluster* da Tabela 5.10 com a rede neural PS.

Saídas	Entradas	
	J	O
J	0,00	0,00
O	2,00	10,0

Tomando por base as Tabelas 5.10 e B.1 à B.7 obtém-se a Tabela B.8, cujos resultados para cada letra, em que a primeira coluna da Tabela corresponde a cada uma das letras juntamente aos *clusters* que compreendem a respectiva letra, em seguida a taxa de acerto percentual para cada uma das letras e as últimas colunas apresentam as letras em que observa-se erro e o erro percentual para cada uma das letras (valor entre parênteses).

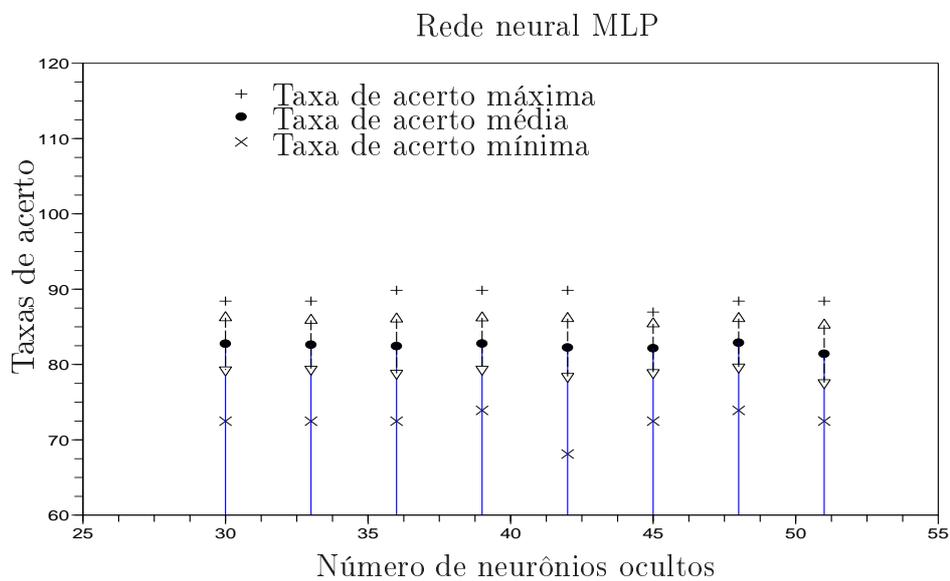
Tabela B.8: tabela de confusão final para os resultados das Tabelas 5.10 e B.1 à B.7.

Letras e <i>clusters</i>	Taxas de acerto	Taxas de erro
A {8}	29,7	B(42,4), M(27,9)
B {8}	92,1	A(7,9)
C {8, 10}	8,64	I(3,0), M(31,0), S(2,36), N(26,0), Q(26,7), Y(2,27)
D {11}	100	—
E {9}	100	—
F {4}	100	—
G {6}	80,6	T(19,3)
H {6, 7}	0,00	T(60,0), L(0,4), V(24,0), W(15,6)
I {8}	20,1	A(16,1), B(7,8), M(2,7), S(53,3)
J {2, 7, 12}	56,5	W(21,7), O(21,7)
K {5, 6}	0,00	R(5,22), U(17,0), G(68,9), P(0,4), T(8,4)
L {7}	100	—
M {8}	100	—
N {10}	100	—
O {12}	100	—
P {6}	86,6	G(2,9), T(10,5)
Q {10}	96,7	N(3,3)
R {5}	41,4	U(58,6)
S {8}	19,4	A(56,5), B(23,8), M(0,3)
T {6}	96,7	G(1,6), P(1,7)
U {5}	89,2	R(10,8)
V {7}	100	—
W {7}	100	—
X {9}	99,0	E(1,0)
Y {9, 10}	66,7	X(33,3)
Z {6, 7}	0,00	G(38,8), T(16,8), W(44,4)

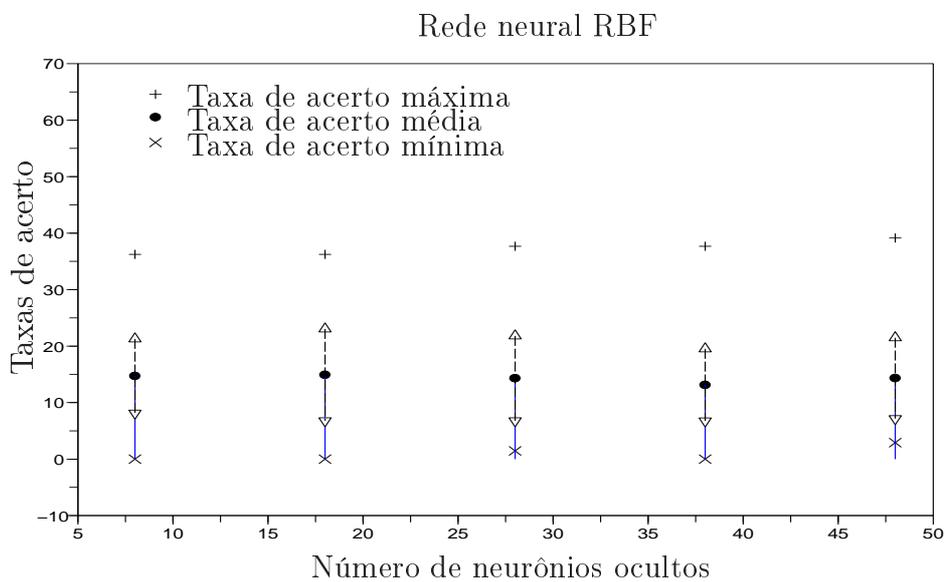
As informações que seguem apresentam as taxas de acerto percentual dos classificadores correspondentes às redes neurais MLP e RBF, variando o número de neurônios na camada oculta, para cada *cluster*, e as Tabelas B.9 à B.15 apresentam a matriz de confusão referente ao melhor resultado obtido para cada *cluster* da Tabela

5.11, Capítulo 5.

Os resultados das redes neurais MLP e RBF para o primeiro *cluster* da Tabela 5.11 são apresentados na Figura B.6.



(a)



(b)

Figura B.6: resultados das redes neurais (a) MLP e (b) RBF para o primeiro *cluster* da Tabela 5.11.

O classificador baseado na rede neural PS apresenta 76,56% de acerto para o

Os resultados da rede neural RBF para o segundo *cluster* da Tabela 5.11 são apresentados na Figura B.7, os resultados da rede neural MLP para este *cluster* são constantes e iguais a 100,0%.

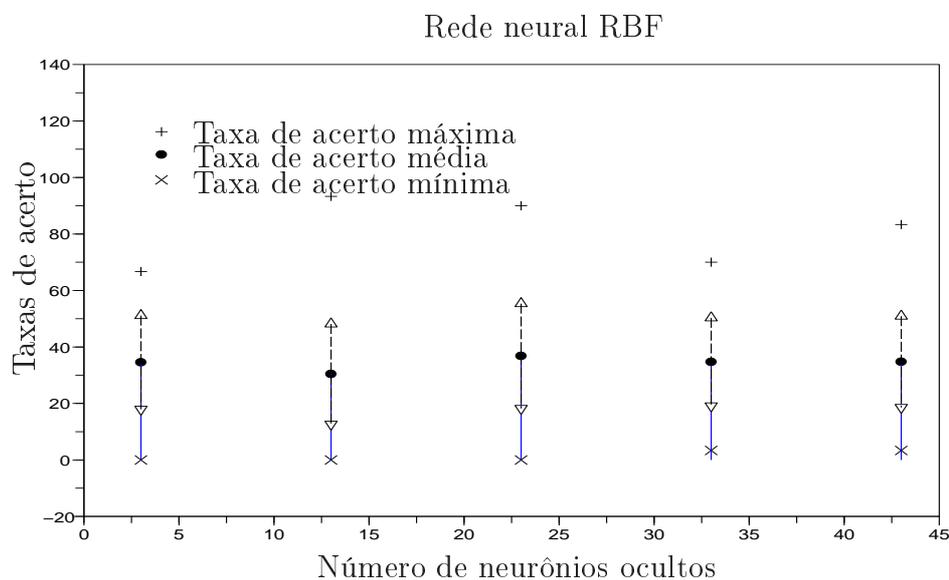


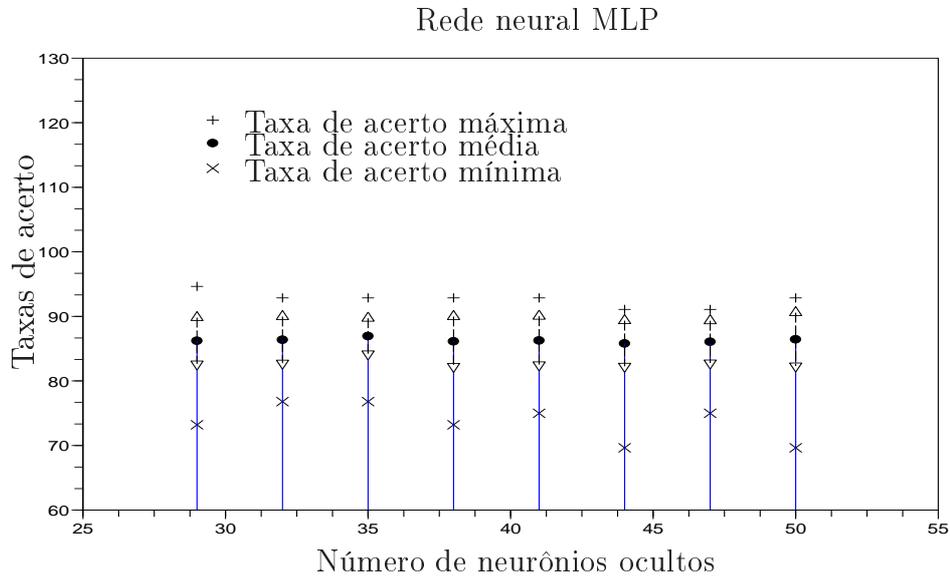
Figura B.7: resultados da rede neural RBF para o segundo *cluster* da Tabela 5.11.

O classificador baseado na rede neural PS apresenta 100,0% de acerto para o segundo *cluster* da Tabela 5.11. Assim, o melhor resultado é o obtido com a rede neural PS, cuja matriz de confusão está apresentada na Tabela B.10.

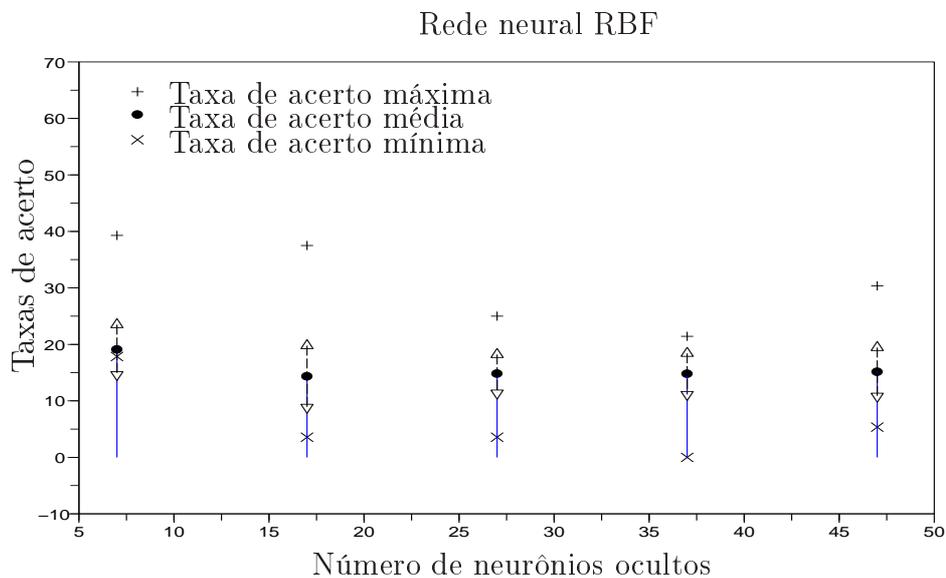
Tabela B.10: matriz de confusão para o segundo *cluster* da Tabela 5.11 com a rede neural PS.

Saídas	Entradas		
	B	E	U
B	10,0	0,00	0,00
E	0,00	10,0	0,00
U	0,00	0,00	10,0

Os resultados das redes neurais MLP e RBF para o terceiro *cluster* da Tabela 5.11 são apresentados na Figura B.8.



(a)



(b)

Figura B.8: resultados das redes neurais (a) MLP e (b) RBF para o terceiro *cluster* da Tabela 5.11.

O classificador baseado na rede neural PS apresenta 82,07% de acerto para o terceiro *cluster* da Tabela 5.11. Neste caso, o melhor resultado é o obtido com a rede neural MLP com 35 neurônios ocultos com taxa de acerto de 86,94%, cuja matriz

de confusão está apresentada na Tabela B.11.

Tabela B.11: matriz de confusão para o terceiro *cluster* da Tabela 5.11 com a rede neural MLP com 35 neurônios ocultos.

Saídas	Entradas						
	C	D	H	K	O	V	W
C	7,46	0,34	0,34	0,00	0,04	0,05	0,00
D	0,27	8,56	0,00	0,02	0,00	0,26	0,00
H	0,02	0,00	5,66	0,24	0,00	0,06	0,27
K	0,00	0,00	0,00	0,00	0,00	0,00	0,00
O	1,02	0,00	0,00	0,00	9,96	0,00	0,00
V	0,01	0,10	0,00	2,16	0,00	7,32	0,00
W	0,22	0,00	0,00	0,58	0,00	1,31	9,73

Os resultados da rede neural RBF para o quarto *cluster* da Tabela 5.11 são apresentados na Figura B.9, os resultados da rede neural MLP para este *cluster* são constantes e iguais a 100,0%.

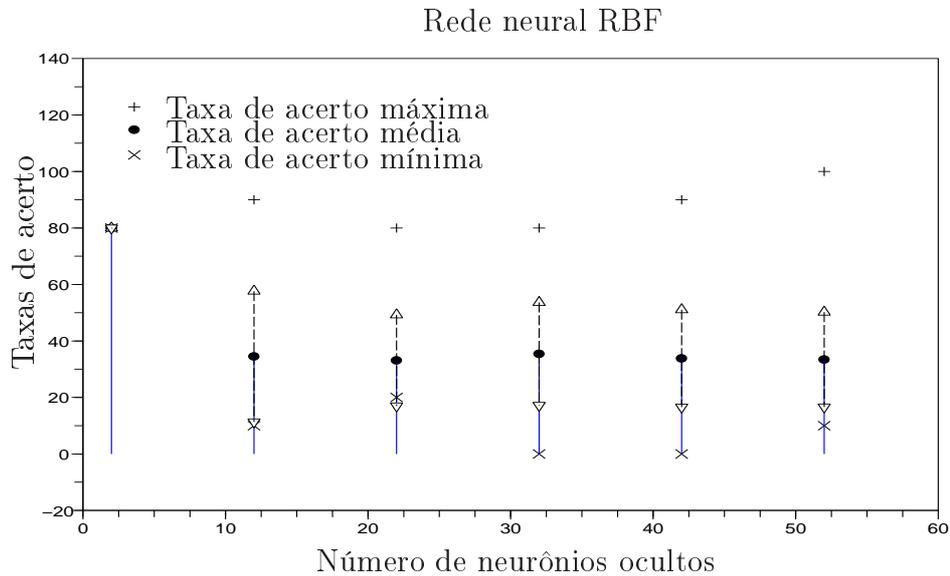


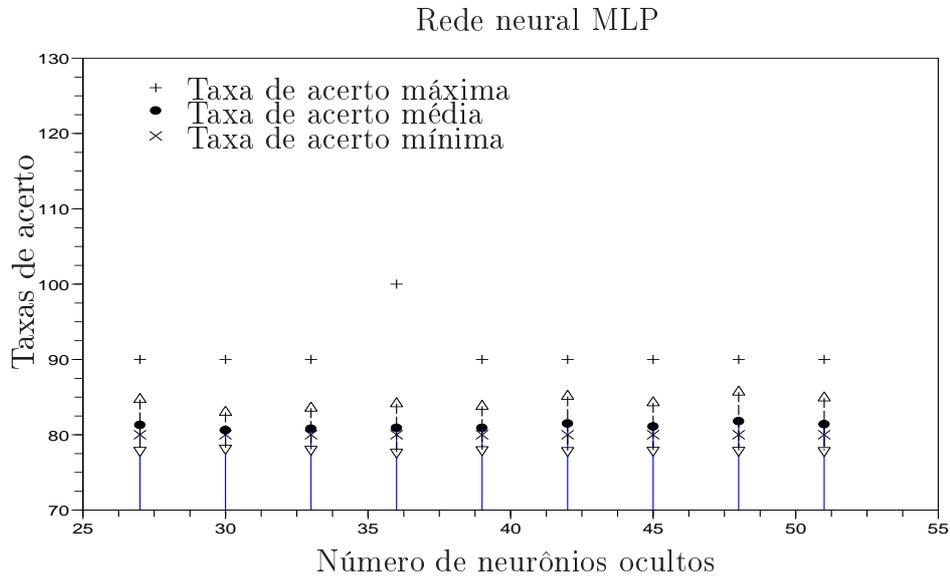
Figura B.9: resultados da rede neural RBF para o quarto *cluster* da Tabela 5.11.

O classificador baseado na rede neural PS apresenta 79,99% de acerto para o quarto *cluster* da Tabela 5.11. De modo que, o melhor resultado é o obtido com a rede neural PS, cuja matriz de confusão está apresentada na Tabela B.12.

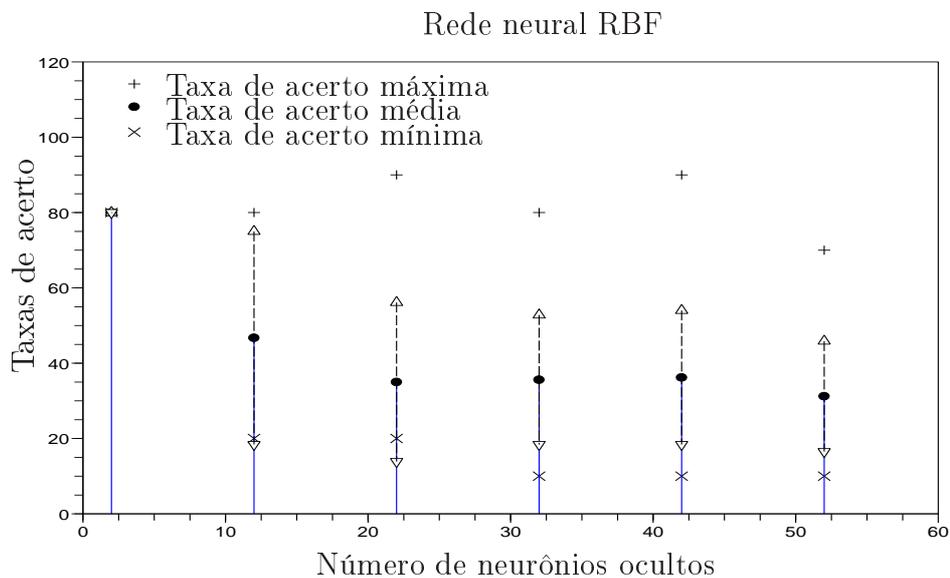
Tabela B.12: matriz de confusão para o quarto *cluster* da Tabela 5.11 com a rede neural PS.

Saídas	Entradas	
	H	Y
H	0,00	0,00
Y	2,00	8,00

Os resultados das redes neurais MLP e RBF para o oitavo *cluster* da Tabela 5.11 são apresentados na Figura B.10.



(a)



(b)

Figura B.10: resultados das redes neurais (a) MLP e (b) RBF para o oitavo *cluster* da Tabela 5.11.

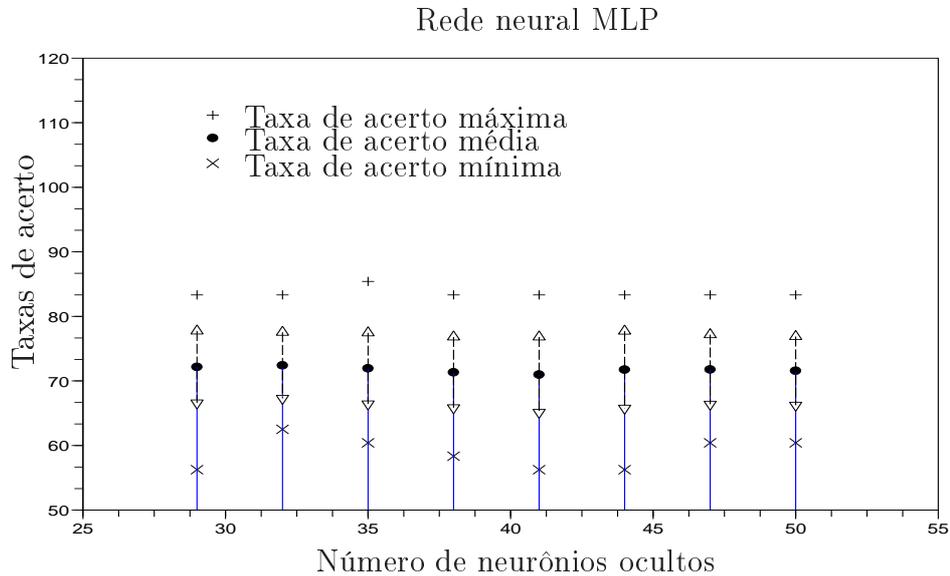
O classificador baseado na rede neural PS apresenta 79,99% de acerto para o oitavo *cluster* da Tabela 5.11. Assim, o melhor resultado é o obtido com a rede neural MLP com 48 neurônios ocultos com taxa de acerto de 81,79%, cuja matriz

de confusão está apresentada na Tabela B.13.

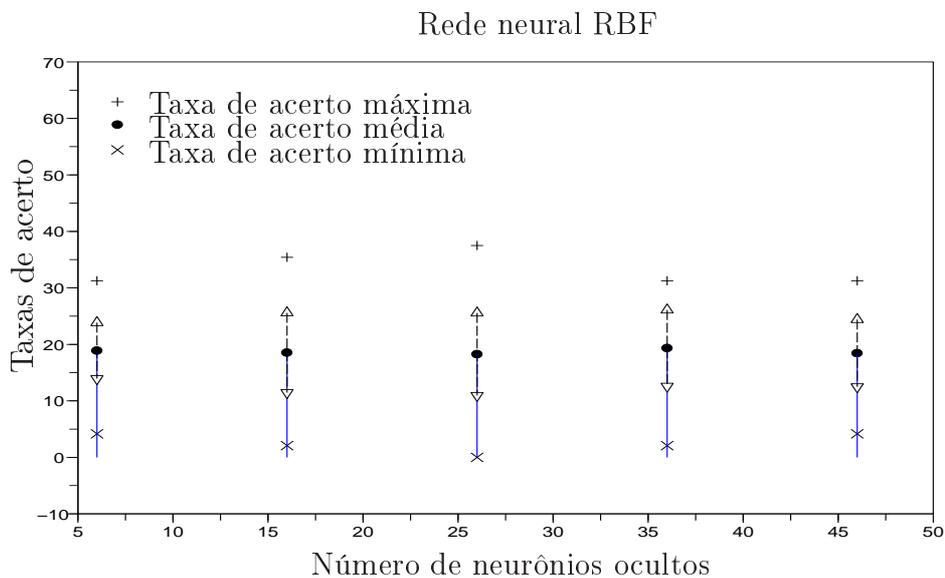
Tabela B.13: matriz de confusão para o oitavo *cluster* da Tabela 5.11 com a rede neural MLP com 48 neurônios ocultos.

Saídas	Entradas	
	H	L
H	0,18	0,00
L	1,82	8,00

Os resultados das redes neurais MLP e RBF para o nono *cluster* da Tabela 5.11 são apresentados na Figura B.11.



(a)



(b)

Figura B.11: resultados das redes neurais (a) MLP e (b) RBF para o nono *cluster* da Tabela 5.11.

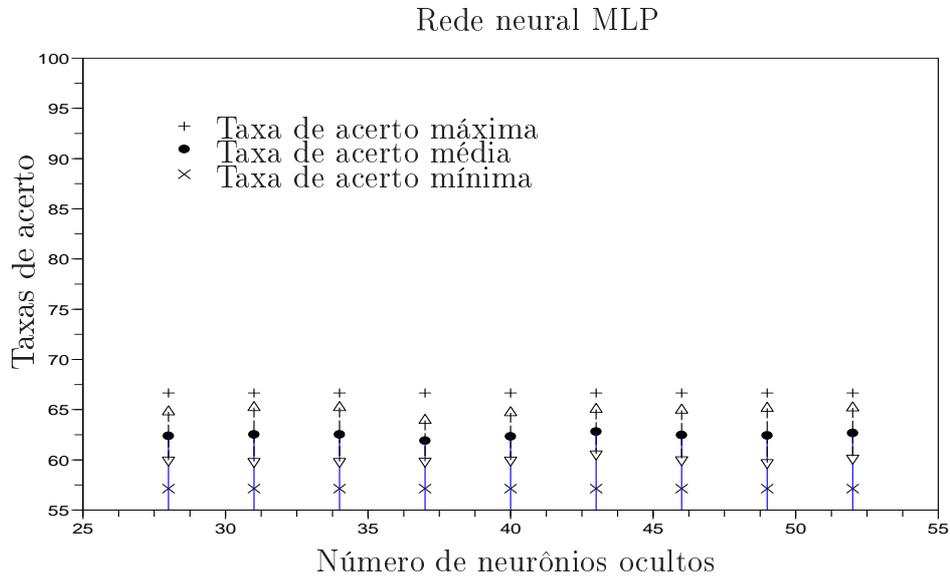
O classificador baseado na rede neural PS apresenta 67,95% de acerto para o nono *cluster* da Tabela 5.11. De modo que, o melhor resultado é o obtido com a rede neural MLP com 32 neurônios ocultos com taxa de acerto de 72,41%, cuja matriz

de confusão está apresentada na Tabela B.14.

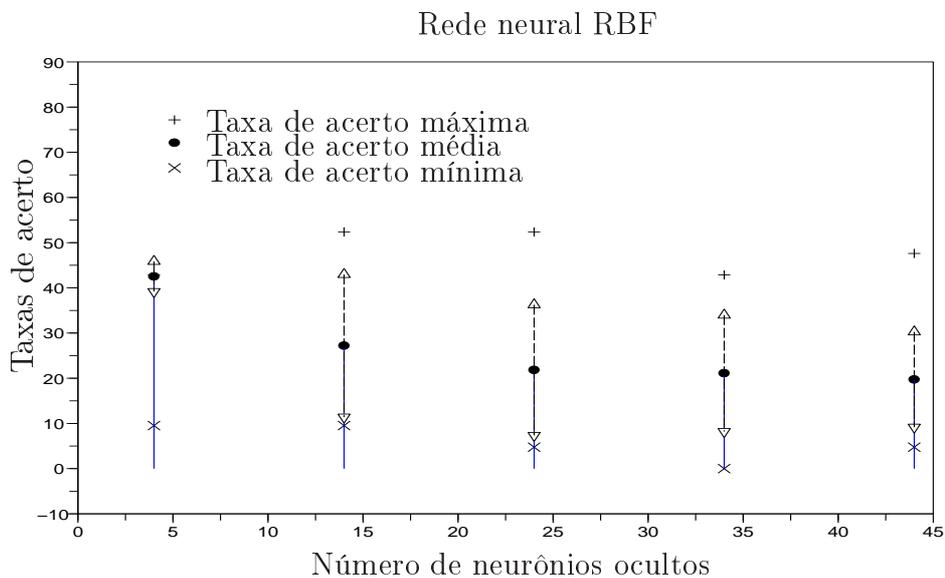
Tabela B.14: matriz de confusão para o nono *cluster* da Tabela 5.11 com a rede neural MLP com 32 neurônios ocultos.

Saídas	Entradas					
	F	G	J	K	R	T
F	8,85	1,71	0,01	0,28	0,00	0,21
G	0,13	7,24	0,00	0,00	0,20	0,00
J	0,00	0,00	0,00	0,00	0,00	0,00
K	0,37	0,00	0,01	4,59	0,00	3,46
R	0,00	1,05	0,00	0,85	9,75	1,00
T	0,65	0,00	1,98	1,28	0,05	4,33

Os resultados das redes neurais MLP e RBF para o décimo *cluster* da Tabela 5.11 são apresentados na Figura B.12.



(a)



(b)

Figura B.12: resultados das redes neurais (a) MLP e (b) RBF para o décimo *cluster* da Tabela 5.11.

O classificador baseado na rede neural PS apresenta 52,04% de acerto para o décimo *cluster* da Tabela 5.11. Neste caso, o melhor resultado é o obtido com a rede neural MLP com 43 neurônios ocultos com taxa de acerto de 62,80%, cuja matriz

de confusão está apresentada na Tabela B.15.

Tabela B.15: matriz de confusão para o décimo *cluster* da Tabela 5.11 com a rede neural MLP com 43 neurônios ocultos.

Saídas	Entradas			
	J	X	Y	Z
J	0,00	0,00	0,00	0,00
X	1,68	7,32	1,00	2,13
Y	0,00	0,00	0,00	0,00
Z	0,32	1,68	1,00	5,87

Tomando por base as Tabelas 5.11 e B.9 à B.15 obtém-se a Tabela B.16. Nesta tabela estão os resultados para cada letra, em que a primeira coluna da tabela corresponde a cada uma das letras juntamente aos *clusters* que compreendem a respectiva letra, em seguida a taxa de acerto percentual para cada uma das letras e as últimas colunas apresentam as letras em que observa-se erro e o erro percentual para cada uma das letras (valor entre parênteses).

Tabela B.16: tabela de confusão final para os resultados das Tabelas 5.11 e B.9 à B.15.

Letras e <i>clusters</i>	Taxas de acerto	Taxas de erro
A {1}	86,3	I(0,1), N(8,4), Q(2,2), S(3,0)
B {2}	100	—
C {3}	82,9	D(3,0), H(0,2), O(11,3), W(2,4), V(0,1)
D {3}	95,1	C(3,8), V(1,1)
E {2}	100	—
F {9}	88,5	G(1,3), K(3,7), T(6,5)
G {9}	72,4	F(17,1), R(10,5)
H {3, 4, 8}	58,4	C(3,4), Y(20,0), L(18,2)
I {1}	89,9	Q(8,9), S(1,2)
J {1, 9, 10}	58,3	A(1,1), I(0,3), S(0,3), F(0,1), K(0,1), T(19,8), X(16,8), Z(3,2)
K {3, 9}	45,9	D(0,2), H(2,4), V(21,6), W(5,8), F(2,8), R(8,5), T(12,8)
L {8}	100	—
M {1}	100	—
N {1}	49,7	A(15,3), J(0,9), M(33,2), S(0,9)
O {3}	99,6	C(0,4)
P {5, 7}	100	—
Q {3}	100	—
R {9}	97,5	G(2,0), T(0,5)
S {1}	82,7	A(7,6), N(9,7)
T {9}	48,1	F(2,3), K(38,5), R(11,1)
U {2}	100	—
V {3}	81,3	C(0,5), D(2,9), H(0,6), W(14,6)
W {3}	97,3	H(2,7)
X {10}	81,3	Z(18,7)
Y {4, 10}	80,0	X(10,0), Z(10,0)
Z {10}	57,9	J(4,0), M(18,2), N(0,5), Y(19,3)

As Tabelas B.8 e B.16 apresentam os resultados da classificação por letra, em que observam-se letras que têm altas taxas de acerto em ambos os casos, por exemplo $\{B, D, E, L, M, O, Q, W\}$ e outras que têm baixas taxas em ambos os casos,

$\{H, K, Z\}$. Considerando que as letras que apresentam as menores taxas de acerto são caracterizadas por movimento e outras letras, como $\{X, Y\}$, também contém movimentos, porém são reconhecidas eficientemente pelo sistema, conclui-se que existe a necessidade de aprimorar a descrição dos gestos para melhor representar a geometria das posturas da mão e os movimentos para cada gesto.

Referências Bibliográficas

AMARAL, M. A.; COUTINHO, A. A criança surda: Educação e inserção social. *Análise Psicológica*, v. 20, n. 3, p. 373–378, jul 2002.

BAPTISTA, F. *F-Libras: Ambiente integrado de ensino-aprendizagem para a Língua Brasileira de Sinais*. Dissertação (Mestrado) — Centro de Ensino "Eurípedes Soares da Rocha", Marília, SP, 2007.

BEDREGAL, B. R. C. *et al.* Hand gesture recognition in an interval fuzzy approach. *Tendências em Matemática Aplicada e Computacional*, v. 8, n. 1, p. 21–31, 2007.

BOWDEN, R.; SARHADI, M. Building temporal models for gesture recognition. *Proceedings of British Machine Vision Conference*, 2000.

CALLIOLI, C. A. *et al.* *Álgebra Linear e Aplicações*. 6a. ed. São Paulo: Atual, 1990. ISBN 8570562977.

CAMPR, P. *et al.* Sign-language-enabled information kiosk. *Proceedings of Summer Workshop on Multi-Modal Interfaces*, p. 24 – 33, 2008.

CAPOVILLA, F. C. Recursos para educação de crianças com necessidades especiais e articulação entre educação especial e inclusiva. *O Mundo da Saúde*, v. 32, n. 2, p. 208 – 214, 2008.

CAPOVILLA, F. C. *et al.* Avaliando compreensão de sinais da LIBRAS em escolares surdos do ensino fundamental. *Interação em Psicologia*, v. 8, n. 2, p. 159 – 169, 2004.

CAPOVILLA, F. C.; RAPHAEL, W. D. *Enciclopédia da língua de sinais brasileira: o mundo do surdo em LIBRAS*. São Paulo: Editora Universidade de São Paulo - EDUSP, 2005.

CONSEIL, S. *et al.* Comparison of Fourier descriptors and Hu moments for hand posture recognition. *Proceedings of Eusipco*, p. 1960 – 1964, 2007.

COOPER, H.; BOWDEN, R. Large lexicon detection of sign language. *Human-Computer Interaction (HCI)*, p. 88 – 97, 2007.

DIAS, J. B. *et al.* Conjunto de treinamento para algoritmos de reconhecimento de LIBRAS. *II Workshop de Visão Computacional*, p. 354 – 359, 2006.

FELIX, J. H. da S. *Sistema de visão computacional para detecção e quantização de efisema pulmonar*. Dissertação (Mestrado) — Universidade Federal do Ceará, Fortaleza, CE, 2007.

FRIEDHEIN, J. *et al.* Dificuldades e capacitações profissionais das pessoas surdas de Santa Maria. *Anais do IX Simpósio de Ensino, Pesquisa e Extensão – SEPE*, 2005.

FUKUNAGA, K. *Introduction to Statistical Pattern Recognition*. 2a. ed. San Diego: Academic Press, 1990. ISBN 0-12-269851-7.

GABARDO, T. M. *Reconhecimento Computacional da Língua Brasileira de Sinais*. 2005. Monografia, UnicenP, Curitiba - PR.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 3a. ed. New Jersey: Prentice Hall, 2008. ISBN 978-0-13-168728-8.

GUAZZELLI, V. B. Possibilidades de interação de alunos surdos com ferramentas de comunicação computadorizadas. *II Escola Regional de Licenciatura em Computação - Erlicomp*, 2007.

HANDOUYAYA, M. *et al.* Sign language recognition using moment-based size functions. *Vision Interface '99*, p. 210–216, 1999.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2a. ed. New Jersey: Prentice Hall, 1999. ISBN 0-13-273350-1.

- HERTZ, J. A. *et al. Introduction to the Theory of Neural Computation*. 1a. ed. Redwood City: Westview, 1998. ISBN 0-201-51560-1.
- HU, M.-K. Visual pattern recognition by moment invariants. *Ire Transactions on Information Theory*, p. 179–187, 1962.
- IBGE. *Censo Demográfico 2000*. Jun 2003. Web – <http://www.ibge.gov.br/home/presidencia/noticias/27062003censo.shtm>. Acessado em 1 de agosto de 2009.
- IMMERSION. *CyberGlove® II Wireless Glove*. Feb 2007. Web – http://www.metamotion.com/images/CyberGloveII_feb07_v2.pdf. Acessado em 22 de abril de 2009.
- JOKINEN, M. *Fact Sheet: Sign Language*. Oct 2006. World Federation of the Deaf. Helsinki, Finland.
- KOHONEN, T. *Self-Organizing Maps*. 3a. ed. New York: Springer, 2001. ISBN 3-540-67921-9.
- LAMAR, M. V. *et al. Reconhecimento de posturas manuais usando redes neurais. Proceedings of the VI Brazilian Conference on Neural Networks - VI Congresso Brasileiro de Redes Neurais*, p. 307 – 312, 2003.
- Maganhotte Júnior, A. *et al. Crianças surdas, interface muda! multimeios no auxílio da aprendizagem da escrita da língua portuguesa por crianças surdas. Revista Eletrônica de Iniciação Científica - REIC*, 2002.
- MAHMOUD, T. M. A new fast skin color detection technique. *World Academy of Science, Engineering and Technology*, v. 43, p. 501 – 505, 2008.
- MARQUES, J. S. *Reconhecimento de padrões: métodos estatísticos e neuronais*. 2a. ed. Lisboa: Ensino da Ciência e da Tecnologia, 2005. ISBN 972-8469-08-X.
- OLIVEIRA, T. M. de *et al. Segmentação de pele humana em imagens de vídeo utilizando wavelet e redes neurais. II Workshop de Visão Computacional*, p. 129–134, 2006.
- OPPENHEIM, A. V. *et al. Discrete-Time Signal Processing*. 2a. ed. New Jersey: Prentice Hall, 1999.

- PAPOULIS, A.; PILLAI, S. U. *Probability, Random Variables and Stochastic Processes*. 4a. ed. Columbus: McGraw Hill Companies, 2002. ISBN 0073660116.
- PHUNG, S. L. *et al.* Skin segmentation using color pixel classification: analysis and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 27, n. 1, p. 148 – 154, 2005.
- PISTORI, H. *et al.* Sigus - plataforma de apoio ao desenvolvimento de sistemas para inclusão digital de pessoas com necessidades especiais. *IV Congresso Iberoamericano sobre Tecnologias de Apoio a Portadores de Deficiência - IBERDISCAP*, p. 20–22, 2006.
- RASOOL, S. B. *Skin Segmentation: Combining Color and Edge Information*. Dhahran, Saudi Arabia, 2003.
- REAES, F. M. *Reconhecimento de Faces em Imagens: Projeto Beholder*. 2006. Monografia, USP, São Paulo - SP.
- REICHERT, A. R. *Mídia televisiva sem som*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2006.
- RIBEIRO, H. L. *Reconhecimento de gestos usando segmentação de imagens dinâmicas de mãos baseada no modelo de mistura de gaussianas e cor de pele*. Dissertação (Mestrado) — Escola de Engenharia de São Carlos (EESC), São Carlos, SP, 2006.
- RIBEIRO, H. L.; GONZAGA, A. Reconhecimento de gestos de mão usando o algoritmo GMM e vetor de características de momentos de imagem. *II Workshop de Visão Computacional*, p. 13–18, 2006.
- RYBENÁ. *Solução Rybená: Comunicação Priorizando a Acessibilidade*. 2009. Web – <http://www.rybena.com.br/rybena/com/default/index.jsp>. Acessado em 30 de abril de 2009.
- SIGAL, L. *et al.* Estimation and prediction of evolving color distributions for skin segmentation under varying illumination. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 2152 – 2159, 2000.
- SILVEIRA, F. F. *O trabalho: representações sociais e a identidade dos surdos*. Brasília, DF, 2006.

- SOUZA, K. P.; PISTORI, H. Implementação de um extrator de características baseado em momentos da imagem. *XVIII Brazilian Symposium on Computer Graphics and Image Processing - SIBGRAPI, III Workshop de Trabalhos de Iniciação Científica em Computação Gráfica e Processamento de Imagens, WICCGPI*, 2005.
- SOUZA, K. P. de *et al.* Reconhecimento automático de gestos da língua brasileira de sinais utilizando visão computacional. *III Workshop de Visão Computacional*, 2007.
- SOUZA, V. C. de; VIEIRA, R. Uma proposta para tradução automática entre LIBRAS e Português no Sign WebMessage. *Proceeding of the Brazilian Symposium on Artificial Intelligence 2006*, 2006.
- STARNER, T. *et al.* *Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video*. Massachusetts Institute of Technology, USA, 1996.
- TRUYENQUE, M. A. Q. *Uma Aplicação de Visão Computacional que Utiliza Gestos da Mão para Interagir com o Computador*. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, Rio de Janeiro, RJ, 2005.
- VOGLER, C.; METAXAS, D. Parallel hidden markov models for american sign language recognition. *Proceedings of the International Conference on Computer Vision*, p. 22 – 25, 1999.
- WU, H. Dynamic gesture recognition using pca with multi-scale theory and hmm. *Proceedings of SPIE*, v. 4550, p. 132 – 139, 2001.
- ZAHEDI, M. *Robust Appearance-based Sign Language Recognition*. Tese (Doutorado) — Rheinisch-Westfälischen Technischen Hochschule Aachen, Aachen, Germany, 2007.
- ZAHEDI, M. *et al.* Geometric features for improving continuous appearance-based sign language recognition. *Proceedings of British Machine Vision Association*, v. 3, p. 1019 – 1029, 2006.

ZHANG, D.; LU, G. A comparative study on shape retrieval using Fourier descriptors with different shape signatures. *Int. Conf. on Intelligent Multimedia and Distance Education*, 2001.

ZHENG, J.-H. *et al.* Adaptive skin detection under unconstrained lighting conditions using a bigaussian model and illumination estimation. *Image Analysis and Stereology*, v. 24, p. 21 – 33, 2005.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)