UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

ROBSON SILVA DE SOUZA

RECONHECIMENTO DAS CONFIGURAÇÕES DE MÃO DA LÍNGUA BRASILEIRA DE SINAIS –LIBRAS EM IMAGENS DE PROFUNDIDADE ATRAVÉS DA ANÁLISE DE COMPONENTES PRINCIPAIS E DO CLASSIFICADOR K-VIZINHOS MAIS PRÓXIMOS.

MANAUS
2015

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

ROBSON SILVA DE SOUZA

RECONHECIMENTO DAS CONFIGURAÇÕES DE MÃO DA LÍNGUA BRASILEIRA DE SINAIS –LIBRAS EM IMAGENS DE PROFUNDIDADE ATRAVÉS DA ANÁLISE DE COMPONENTES PRINCIPAIS E DO CLASSIFICADOR K-VIZINHOS MAIS PRÓXIMOS.

Dissertação apresentada ao Programa de Pós Graduação *Stricto Sensu* em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção de título de Mestre em Engenharia Elétrica, área de concentração Controle e Automação de Sistemas.

Orientadora: Profª. Drª. Marly Guimarães Fernandes Costa
Coorientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho

MANAUS
2015

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Souza, Robson Silva de
S729r Reconhecimento das configurações de mão da língua brasileira de sinais - LIBRAS em imagens de profundidade através da análise de componentes principais e do classificador k-vizinhos mais próximos / Robson Silva de Souza. 2015
115 f.: il. color; 31 cm.

Orientadora: Marly Guimarães Fernandes Costa
Coorientadora: Cícero Ferreira Fernandes Costa Filho
Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Amazonas.

1. deficiência auditiva. 2. surdos. 3. LIBRAS. 4. configurações de mãos. 5. Kinect. I. Costa, Marly Guimarães Fernandes II. Universidade Federal do Amazonas III. Título

ROBSON SILVA DE SOUZA

RECONHECIMENTO DAS CONFIGURAÇÕES DE MÃO DA LÍNGUA BRASILEIRA DE SINAIS-LIBRAS EM IMAGENS DE PROFUNDIDADE ATRAVÉS DA ANÁLISE DE COMPONENTES PRINCIPAIS E DO CLASSIFICADOR K-VIZINHOS MAIS PRÓXIMOS.

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

Aprovado em 11 de setembro de 2015.

BANCA EXAMINADORA

Profa. Dra. Marly Guimarães Fernandes Costa, Presidente

Universidade Federal do Amazonas- UFAM

Prof. Dr. José Raimundo Gomes Pereira, Membro

Universidade Federal do Amazonas- UFAM

Prof. Dr. Marco Antônio Gutierrez, Membro

Universidade do Estado de São Paulo- USP

AGRADECIMENTOS

Gostaria de agradecer primeiramente aos meus orientadores, Profa. Dra. Marly Guimarães Fernandes Costa e Prof. Dr. Cícero Ferreira Fernandes Costa Filho. Obtive muito conhecimento em suas disciplinas e orientação extraclasse;

Agradeço ao professor Marcos Roberto dos Santos, da Universidade do Estado do Amazonas, pelas aulas de LIBRAS;

Aos meus pais e familiares pelo suporte, que me permitiu ir até o fim nesta dissertação;

Aos meus colegas de classe que me acompanham desde a graduação: Endrews, Jonilson, Larissa e Márcio. O apoio e companheirismo vocês foram essências para concluir as disciplinas, e principalmente, para me manter focado em cada etapa do desenvolvimento deste trabalho;

Aos amigos e ex-colegas de faculdade Renata Suzan e Diego Ramon pelo incentivo ao ingresso no programa de pós-graduação;

Também gostaria de agradecer, ao corpo docente do programa de pós-graduação em engenharia elétrica-PPGEE/UFAM. Grande parte das disciplinas lecionadas contribuiu para o desenvolvimento deste trabalho;

À Universidade Federal do Amazonas e aos profissionais do Centro de Pesquisa e Desenvolvimento de Tecnologia Eletrônica e da Informação – CETELI. Graças à estrutura oferecida como o espaço físico e materiais de pesquisa, o desenvolvimento deste trabalho foi possível;

Parte dos resultados apresentados neste trabalho foram obtidos através do Projeto de pesquisa e formação de recursos humanos, em níveis de graduação e pós-graduação, nas áreas de automação industrial, *softwares* para dispositivos móveis e TV Digital, financiado pela Samsung Eletrônica da Amazônia Ltda., no âmbito da Lei no. 8.387 (art. 2º) /91.

RESUMO

De acordo com o IBGE (Censo 2010), o Brasil possui 9,7 milhões de brasileiros com algum grau de deficiência auditiva, mais de cinco por cento da população. Para a maior parte destas pessoas a língua natural principal utilizada para sua comunicação é a LIBRAS (Língua Brasileira de Sinais) e não o português. O reconhecimento de sinais visa permitir uma maior inserção sócio digital da comunidade surda através da interpretação da língua de sinais pelo computador em formato de áudio ou texto. Esta dissertação apresenta o reconhecimento de um dos parâmetros globais da LIBRAS, as configurações de mão, utilizando o classificador k-vizinhos mais próximos e a técnica de redução de dimensionalidade 2D²PCA. Um conjunto de dados, robusto e representativo das condições do cotidiano, constituído de 12.200 imagens de profundidade das 61 configurações de mão, capturadas pelo sensor Kinect® foi construído. Todas as imagens foram submetidas inicialmente a uma etapa de segmentação que buscou isolar a região da mão direita do resto do corpo. Buscando a eliminação de redundância no conjunto de dados foi implementado uma etapa de extração de características, através da técnica 2D2PCA que determina as formas mais representativas de dados a partir de combinações lineares dos pixels originais. O classificador "k-vizinhos mais próximos" foi a técnica utilizada para a etapa final de reconhecimento automático das configurações de mão. O referido classificador, implementado com k=1 e matriz de característica de 10x10, conseguiu o melhor desempenho, classificando corretamente 96,31% das amostras de testes. Foram obtidas taxas de acerto de 100% para seis configurações de mão.

Palavras-chave – deficiência auditiva, surdos, LIBRAS, configurações de mãos, reconhecimento de padrões, k-vizinhos mais próximos, Kinect®, 2D²PCA

ABSTRACT

According to Brazilian Institute of Geography and Statistic IBGE – *Instituto Brasileiro de Geografia e Estatística* (Censo 2010), Brazil has 9.7 million Brazilians with some degree of hearing impairment or deafness, more than five per cent of the population. For the majority of those persons the main natural communication method is the Brazilian Sign Language (LIBRAS – *Língua Brasileira de Sinais*), instead of the spoken Portuguese. The computer-aided recognition of signs aims to expand the social and digital inclusion of the deaf community by means of those signs translation into audio or text format.

This work presents one of the global LIBRAS parameters recognition, the hand gestures, using k-nearest neighbor's classifier and 2D²PCA dimensionality reduction technique. A robust and representative data set of daily conditions, with 12.200 depth images from 61 hand gesture, captured by Kinect® sensor, was built. Initially the images were segmented in a preprocessing step, which isolated the region of the right hand from the rest of the body. Seeking to eliminate the data set redundancy, the images were submitted to a dimensionality reduction by (2D) dimensional 2PCA technique determining the most representative forms of data from original pixels linear combination. The classifier k-nearest neighbors was the technique in the final stage in the hand gesture automatic recognition. This classifier could correctly categorize 96.31% of test samples (k = 1 and 10x10 feature matrix). Six hand gesture sets were correctly classified obtaining 100% successful rates.

Keywords— hearing impairment, deaf, LIBRAS, pattern recognition, *k-Nearest Neighbors,* Kinect®, 2D²PCA.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE QUADROS

SUMÁRIO

CAPÍTULO 1

INTRODUÇÃO

No decorrer da vida, certas pessoas passam a ter limitações na audição sejam elas causadas por uma doença natural ou causadas por acidentes. Também existem as pessoas que já nascem com deficiência auditiva. Em ambas situações, essas pessoas necessitam aprender outra forma de comunicação, que é a língua de sinais.

Segundo Viader, Pertusa e Vinardell (1999), a língua de sinais é a "língua própria das pessoas surdas, usando sua estrutura, sintaxes e gramáticas próprias, sem o uso simultâneo ou alternativo da língua falada. Respeita-se seu *status* linguístico como língua. Se expressa com elementos prosódicos e reflexões próprias".

A Língua Brasileira de Sinais (LIBRAS) é a forma de comunicação oficial utilizada pela comunidade surda brasileira. Trata-se de uma língua natural de sinais e, como tal, completamente estruturada, com suas próprias e bem definidas regras morfológicas, sintáticas e semânticas, com origem na língua de sinais francesa (CAPOVILLA e RAPHAEL, 2001).

Segundo Ramos (2006), o documento mais importante encontrado sobre a Língua Brasileira de Sinais, "*Iconographia* dos *Signaes* dos Surdos-Mudos", de autoria do aluno surdo Flausino José da Gama foi publicado em 1873. Ramos (2006) também relata que somente em 2001 foi lançado o Dicionário Enciclopédico Ilustrado da LIBRAS, em um projeto coordenado pelo Professor Doutor Fernando Capovilla (Instituto de Psicologia/USP). A LIBRAS foi oficializada em 2002 pela Lei n.º 4.857.

De acordo com o IBGE (CENSO 2010), o Brasil possui 9,7 milhões de brasileiros com algum grau de deficiência auditiva. Mais de 5% da população. A deficiência auditiva severa (pessoas com grande dificuldade ou incapazes de ouvir) foi declarada por 2,1 milhões de

pessoas. Para a maioria dessas pessoas a língua natural principal utilizada para sua comunicação é a LIBRAS e não o português.

No entanto, existe muita dificuldade de comunicação entre pessoas surdas e não surdas, pois as línguas de sinais não são apenas transcrições de línguas faladas e, também, porque as pessoas surdas nem sempre são alfabetizadas na língua oficial do seu país.

As pesquisas realizadas na área de reconhecimento automático de línguas de sinais visam possibilitar uma maior inserção sócio digital da comunidade surda. Dentro desse contexto, o reconhecimento automatizado de sinais da LIBRAS visa permitir aos computadores a interpretação da língua de sinais reproduzida por seres humanos, através da tradução de conversas entre surdos e pessoas que não compreendem a língua de sinais, a qual se dá pela conversão de dados extraídos dos sinais para forma de texto ou áudio (PORFÍRIO, 2013).

Pode-se constatar na literatura a existência de algumas propostas, iniciais e em nível de protótipos, capazes de reconhecer padrões de sinais, mas, ainda com muitas limitações.

Segundo Quan (2010) o trabalho de Shantz (1982) é o precursor no estudo do reconhecimento da língua de sinais. Desde então muitas pesquisas foram realizadas com a intenção de auxiliar os surdos, sendo a maioria dos trabalhos encontrados na literatura sobre este tema focados no alfabeto de um determinado idioma na língua de sinais respectiva. Apesar de ter surgido pela necessidade de representar as letras de forma visual, o uso do alfabeto manual é caracterizado como um Empréstimo Linguístico (BOMFIM, 2012).

Consciente de que a LIBRAS não é soletração, alguns autores concentram suas pesquisas na fonologia a qual estuda a menor unidade da língua, os fonemas. Segundo Rossi (2009) um sinal em LIBRAS pode ser formado a partir da alteração ou combinação de 5 parâmetros fonológicos: configuração de mão, ponto de articulação, orientação, movimento e expressão facial.

O presente trabalho também segue esta linha de raciocínio, propondo-se o estudo e reconhecimento de um dos parâmetros fonológicos globais, as configurações de mãos, pois diferente dos outros quatro fonemas, as configurações de mão estão presentes em todos os sinais em LIBRAS. Adicionalmente, alguns sinais em LIBRAS se diferenciam apenas pela configuração de mão.

Segundo Pimenta e Quadros (2010) na LIBRAS existem 61 configurações de mão, as quais são apresentadas na Figura 1.

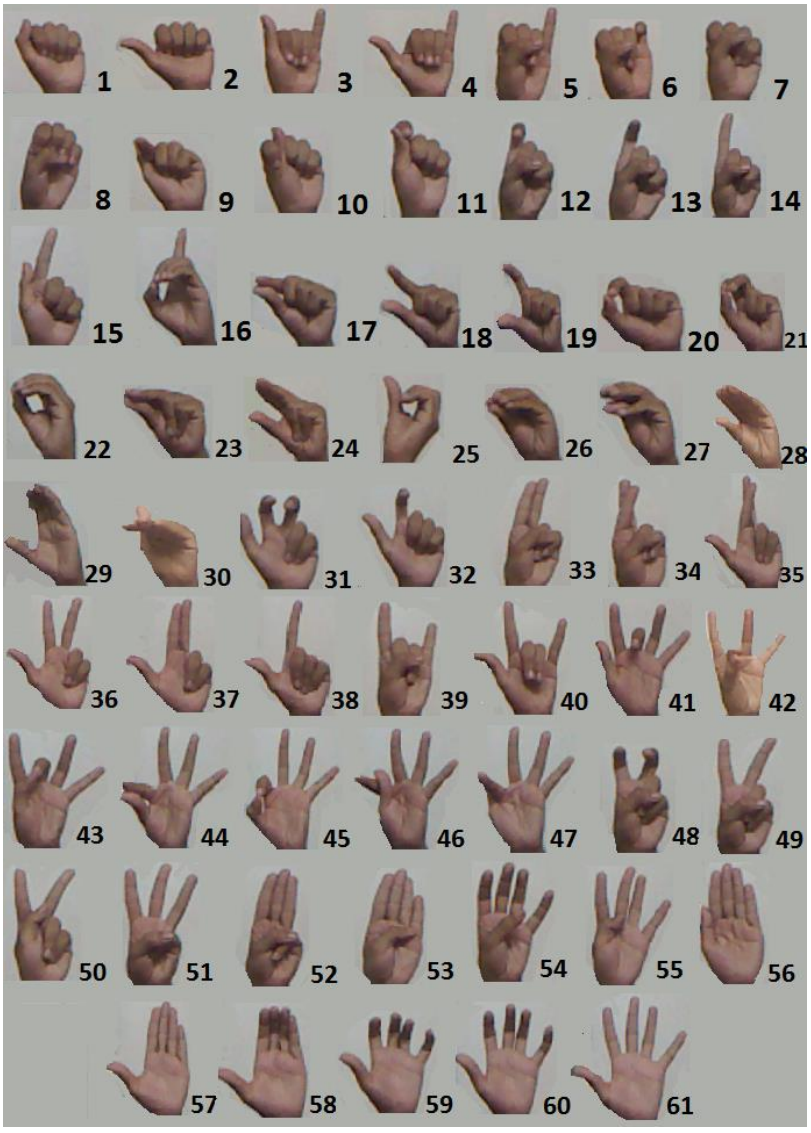


Figura 1 - Configurações de Mão da LIBRAS.

Dessa forma, o desenvolvimento de um método robusto de reconhecimento capaz de distinguir configurações de mão constitui-se em uma contribuição significativa para a construção de um sistema de reconhecimento da LIBRAS que permita uma maior inserção sócio digital da comunidade surda brasileira.

1.1 Objetivo Geral

Reconhecimento automático das 61 configurações de mãos da LIBRAS utilizando mapas de profundidade obtidos com uma câmera de profundidade, Kinect®.

1.2 Objetivos Específicos

- Construção de um banco de dados constituído de imagens RGB e de profundidade das configurações de mãos da LIBRAS;
- Desenvolvimento de um método de segmentação da mão na imagem de profundidade;
- Utilização da técnica de Redução de dimensionalidade de imagens, 2DPCA bidirecional na tarefa de reconhecimento de configurações de mão em LIBRAS;
- Desenvolvimento de um método de reconhecimento das configurações de mão.

1.3 Estrutura do trabalho

Este trabalho foi dividido em 6 capítulos. O capítulo 1 trata à caracterização do problema, o contexto histórico da área de estudo, a delimitação do trabalho e descreve o objetivo geral e os objetivos específicos da dissertação. No capítulo 2 é apresentada a revisão bibliográfica que discute as diversas técnicas que foram elaboradas por diversos autores para o reconhecimento da língua de sinais. No capítulo 3 descreve os conceitos e a fundamentação matemática empregados no desenvolvimento da dissertação. No capítulo 4 é mostrado como foi montado o aparato para se obter as imagens das configurações de mãos, bem como detalhes sobre os métodos realizados. No capítulo 5 discute-se e apresenta-se os resultados encontrados. No capítulo 6 são apresentadas as conclusões do trabalho e também projetos futuros.

CAPÍTULO 2

REVISÃO BIBLIOGRÁFICA

Neste Capítulo será apresentada uma revisão bibliográfica dos principais trabalhos na área de reconhecimento de línguas de sinais.

Na literatura encontramos duas linhas de pesquisa no que diz respeito as atividades de reconhecimento de línguas de sinais. A primeira utiliza apenas técnicas de visão computacional. A Figura 2 apresenta o diagrama de blocos de um sistema de reconhecimento de sinais que utiliza visão computacional.

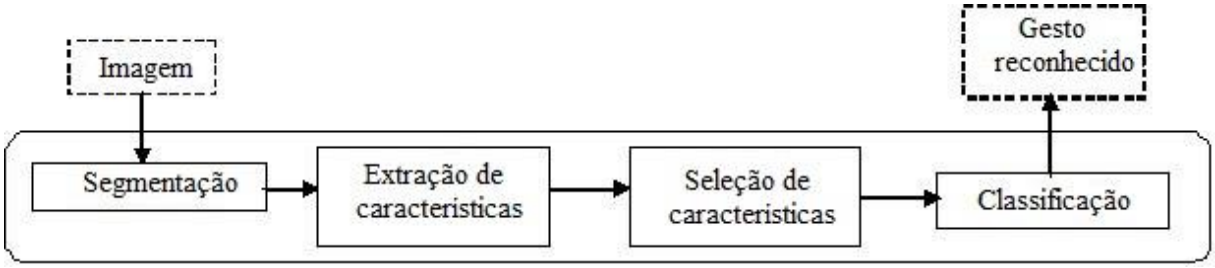


Figura 2 - Diagrama de blocos de um sistema de reconhecimento de sinais com visão computacional.

Neste caso, os dados de entrada são as imagens ou vídeos dos gestos.

A segunda linha de pesquisa associa o uso de luvas sensoriais para captura de dados dos gestos. O usuário calça a luva sensorial e, ao executar os gestos, são extraídos os dados que formarão o conjunto de características a ser utilizado pelo classificador. Nestes casos há sempre uma dependência do sistema operacional especificado pelo fabricante da luva. Adicionalmente, o uso desse aparato implica também em custos adicionais.

Na Seção 2.1 e 2.2 são apresentados os trabalhos identificados a partir de uma busca bibliográfica nas bases de dados do *Institute of Electrical and Electronic Engineers* (IEEE) e do *Engineering Village* das duas áreas de pesquisa, respectivamente.

Ao final do Capítulo é apresentada uma Tabela com um resumo da revisão bibliográfica.

2.1 Reconhecimento de gestos de língua de sinais a partir da aplicação de técnicas de visão computacional

A maioria das pesquisas cientificas encontradas na literatura que tratam do reconhecimento de gestos extraem informações significativas do mesmo a partir de imagens capturadas por câmeras de vídeo, sensores de profundidade ou outros dispositivos (ALL-JARRAH e HALLAWANI, 2001; NERIS *et al*., 2008; CARNEIRO; CORTEZ e COSTA, 2009; PIZZOLATO; ANJO e PEDROSO, 2010; PAULRAJ *et al*., 2011; ADITHYA; VINOD e GOPALAKRISHNAN, 2013).

Na maioria desses trabalhos imagens dos gestos são capturadas em um *background* homogêneo, de cor escura ou branca, a fim de facilitar a etapa de segmentação das mãos através de uma simples operação de limiarização. Estes mesmos autores utilizam redes neurais como classificador para a tarefa de reconhecimento dos gestos.

All-Jarrah e Halawani (2001) utilizaram imagens de fundo escuro (vide exemplos na Figura 3) para validar o método de reconhecimento das 30 letras do alfabeto manual árabe. A etapa de segmentação foi implementada através de um algoritmo de limiar global automático. A seguir, são calculadas a inclinação, θ_g, e o centroide da mão segmentada, bem como são identificados os pixels das bordas da mão. Após isso é realizado um realce das bordas.

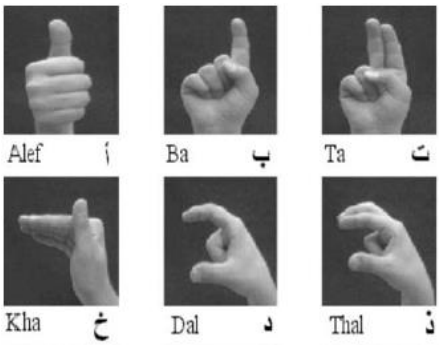


Figura 3 - Alfabeto árabe em fundo escuro.
Fonte: All-Jarrah e Halawani, 2001.

O método de extração de características é invariante a rotação, translação e escala. Esse método extrai 30 vetores, igualmente espaçados, no intervalo: (90- θ_g) $\leq$ θ $\leq$ (113 +θ_g). Os limites do referido intervalo foram obtidos empiricamente. Utilizando esses vetores de características como entrada para um classificador supervisionado *neuro-fuzzy* foi obtida uma taxa reconhecimento da representação das letras do alfabeto árabe em língua de sinais de 93,55%.

O trabalho de Carneiro; Cortez e Costa (2009) apresenta um sistema de reconhecimento da representação das 26 letras do alfabeto da língua portuguesa na Língua Brasileira de Sinais (LIBRAS). Os autores capturaram os gestos representativos em um cenário de iluminação controlada com o fundo branco. As imagens dos gestos em fundo branco são codificadas através do modelo de cor YCbCr. Dado as características de aquisição da imagem, a segmentação do gesto se resume na segmentação de regiões de pele. Então, o processo de segmentação da mão é realizado através de dois limiares globais, cujos valores foram empiricamente obtidos e aplicados aos canais Cb e Cr, respectivamente. Segundo os autores, os tons de pele se situavam no intervalo a seguir apresentado:

$$\begin{cases} 75 < Cb < 127 \\ 133 < Cr < 173 \end{cases} \quad (1)$$

Os gestos segmentados foram então caracterizados através dos 6 momentos invariantes de Hu. A classificação dos gestos se deu em duas etapas: a primeira foi implementada por Mapas Auto Organizáveis (redes SOM) que mapeiam toda a informação de treinamento separando-as por *clusters*, em que cada *cluster* detém um conjunto de classes, ou seja, cada *cluster*, após o treinamento da rede SOM, mapeia algumas das 26 classes (representações das letras do alfabeto); a segunda foi implementada através de redes neurais artificiais onde, para cada *cluster* é treinada uma rede neural *perceptron*, que identifica a classe de cada um dos gestos. A taxa de reconhecimento da representação das letras do alfabeto da língua portuguesa em LIBRAS foi de 89,66%.

Adithya; Vinod e Gopalakrishman (2013) também construíram seu banco de dados no espaço de cor YCbCr. O mesmo foi formado por 15 imagens para cada um dos 36 gestos do alfabeto da língua de sinais Indiana, capturadas em um ambiente com iluminação artificial e fundo escuro.

O mesmo método de segmentação utilizado por Carneiro; Cortez e Costa (2009) foi utilizado por Adithya; Vinod e Gopalakrishman (2013). Neste caso, a análise das imagens evidenciou outros intervalos para as componentes do espaço de cor YCbCr relativas aos tons da pele, quais sejam:

$$75 < Cb < 135 \text{ e } 130 < Cr < 180 \text{ e } Y > 80 \qquad (2)$$

A fim de obter um sistema invariante a rotação foi realizada uma transformada de distância euclidiana na imagem binarizada, apesar desta métrica possuir um alto custo computacional. Em seguida calculou-se o segundo, o terceiro e o quarto momentos centrais dos coeficientes normalizados de Fourier da projeção das linhas e colunas do mapa de distâncias resultantes da transformação de distância. E, posteriormente, uma rede neural utilizando o

algoritmo *backpropagation* foi treinada com os valores dos momentos centrais calculados na etapa anterior. A acurácia do método foi 91,11%.

Pizzolato; Anjo e Pedroso (2010) desenvolveram um método para o reconhecimento de gestos dinâmicos a partir da soletração de palavras usando gestos representativos em LIBRAS de 27 letras do alfabeto da língua portuguesa. O conjunto de dados foi criado pelos próprios autores e constituem-se de 27 sinais, 19 posturas de mão e 8 gestos (dinâmicos). Um vídeo de 3 minutos com cada um dos 45 voluntários realizando os sinais, em resolução de 640x480 e uma taxa de 10 *frames* por segundo foi adquirido em um ambiente totalmente controlado e os voluntários vestidos com camisa de manga longa preta, com apenas a mão centrada no campo de visão da câmera. Os sinais utilizados foram: A, B, C, D, E, F, G, I, J, L, M, N, O, P, R, S, T e U. O sinal J foi dividido em 3 partes (começo, meio e fim).

Inicialmente as imagens foram submetidas a seguinte sequência de pré-processamento: 1. Obtenção de uma imagem binarizada; 2. Extração de uma região de 25 x 25 pixels, contendo a região da mão. Estes 625 pixels foram então a entrada de uma rede neural *Multi Layer Perceptron* com 300 neurônios na camada escondida. Os autores também implementaram uma arquitetura de redes neurais em dois níveis, onde o primeiro foi treinado para reconhecer grupos de letras de sinais próximos: (i) A, E, M e S chamado, AEMS; (ii) F, N e T chamado (FNT); (iii) G, R e U, chamado GRU; e (iv) J3 e P, chamado J3P. No segundo nível uma outra rede foi treinada para discriminar os sinais de um mesmo grupo.

Para o teste dessa arquitetura foram criados um conjunto de teste constituído de um vídeo onde voluntários realizavam os sinais relativos a soletração de 15 palavras: arara, cisne, cobra, foca, gato, jaguatirica, leopardo, macaco, pato, perereca, peru, rato, sapo, tatu e urso. Cada estudante soletrou 1 vez cada palavra. Foram produzidos 24 filmes de cada palavra.

Feito isso, um modelo oculto de Markov (HMM -*Hidden Markov Models*) foi treinada para cada uma das 15 palavras, sendo que aquela de maior probabilidade é escolhida como correta. A taxa de reconhecimento global obtida foi de 90,7%.

Diferentemente dos autores discutidos anteriormente que aplicaram seus métodos para reconhecer sinais do alfabeto em línguas de sinais, Paulraj *et al*., (2011) projetam um sistema para reconhecer 44 gestos que representam fonemas da língua inglesa. Sete pessoas participam da construção do banco de dados. Cada pessoa realizou 10 vezes cada um dos gestos. O sistema para captura das imagens foi montado em um ambiente controlado, com fundo escuro e nível de iluminação mantida no intervalo de 15 a 18 lux. A etapa de segmentação da região de interesse é realizada pelo algoritmo descrito a seguir, o qual é aplicado a cada *pixel* da imagem para determinar se o mesmo pertence a uma região de pele ou não.

Quadro 1 - Algoritmo de detecção de pele usando o espaço de cor RGB

Passo 1: Adquirir um *frame* (imagem) RGB.

Passo 2: Separar os componentes R, G e B de *pixel*.

Passo 3: Se $R > 95\ e\ G > 40\ e\ B > 20$. Vá para o passo 4 → o *pixel* não é uma região de pele.

Passo 4: Se $Max\{R,G,B\} - Min\{R,G,B\} > 15$. Vá para o passo 5 → o *pixel* não é uma região de pele.

Passo 5: Se $|R - G| > 15\ e\ R > G\ e\ R > B$. → O *pixel* pertence a uma região de pele

Do contrário → não pertence a uma região de pele

A imagem resultante segmentada é dividida em três subimagens que correspondem a: região da cabeça, região da mão direita e região da mão esquerda.

A etapa seguinte consiste na aplicação da técnica "*Interleave*" para comprimir a imagem segmentada do gesto. Neste caso o *interleave* foi vertical. Trata-se de uma redução da dimensão da imagem (redução do número de colunas pela metade) pela comparação dos pixels de colunas alternadas com os pixels das colunas adjacentes correspondentes para a identificação do maior valor entre eles. Como a imagem segmentada é binaria, as comparações possíveis e os respectivos resultados são: *pixel* da $\text{coluna}_j = 0$ com *pixel* da $\text{coluna}_{j+1} = 0 \rightarrow \text{coluna}_j = 0$ com $\text{coluna}_{j+1} = 1 \rightarrow 1$; $\text{coluna}_j = 1$ com $\text{coluna}_{j+1} = 0 \rightarrow 1$ e $\text{coluna}_j = 1$ com $\text{coluna}_{j+1} = 1 \rightarrow 1$.

Após a aplicação do método de *Interlieve* foram extraídos os momentos invariantes 2D (num total de 120) e estes foram os dados de entrada para uma rede neural artificial (RNA) de arquitetura 120-40-4 para classificar os gestos da mão direita e da mão esquerda. A rede foi treinada com 60% das amostras (1848) e testada com as 40% restantes (1232). A saída da RNA da mão direita, bem como a da mão esquerda são combinadas e associadas ao respectivo fonema em outra RNA de arquitetura 8-3-6. O sistema proposto apresentou uma acurácia média de classificação de 92,59%.

Neris *et al*., (2008) construíram um conjunto de imagens de gestos correspondentes a representação de 26 letras latinas e 20 configurações de mãos da LIBRAS. Os referidos gestos foram adquiridos de duas pessoas. Após realizar a segmentação por um processo de limiar é aplicada a técnica de assinaturas de bits na imagem binarizada que consiste simplesmente em construir 2 vetores com a quantidade de pixels de valor 1 em cada linha e em cada coluna da imagem, respectivamente. São realizados seis tipos de treinamentos com o conjunto de dados utilizando mapas auto organizáveis. Com o conjunto de dados que envolvem os 26 sinais das letras do alfabeto foram realizados quatro treinamentos distintos. Na Tabela a seguir pode-se observar as taxas de acertos para os diversos experimentos realizados.

Tabela 1 - Taxas de acerto para o problema de reconhecimentos de sinais das letras do alfabeto.

	Test 1 R ($\mp 5^o$)	Test 2 R($\mp 10^o$)	Test 3 R($\mp 15^o$)	Test 4 S
Training 1 (V-O)	**98.0**	94.2	**78.8**	88.4
Training 2 (H-O)	**98.0**	65.3	44.2	**90.3**
Training 3 (V-O-R) ($\mp 15^o$)	92.3	**98.7**	*	82.6
Training 4 (H-O-R) ($\mp 15^o$)	86.5	90.3	*	73.0

V- Assinatura de bits na vertical, H-assinatura de bits na horizontal, O-padrões originais, R-padrões rotacionados, S-padrões escalonados.
Fonte: Neris *et al*., (2008)

O conjunto de dados completo que envolvem os 46 sinais foram utilizados em dois treinamentos distintos. Na Tabela 2 os resultados dos experimentos realizados são apresentados:

Tabela 2 - Taxas de acerto para o problema de reconhecimentos de sinais do conjunto completo de sinais.

	Test 5 R (5^o)	Test 6 R (10^o)	Test 7 R (15^o)	Test 8 S
Training 5 (V-O)	**98,9**	**84,7**	**65,2**	45,6
Training 6 (H-O)	84,7	43,4	21,7	**70,6**

V- Assinatura de bits na vertical, H-assinatura de bits na horizontal, O-padrões originais, R- padrões rotacionados, S-padrões escalonados.
Fonte: Neris *et al*., (2008).

A maior taxa de acertos foi de 98,9% e foi resultante de um treinamento utilizando assinaturas de bits na vertical (V), aplicando-as nas 46 configurações originais (O) e testando em padrões rotacionados (R) em 5°. Da Tabela 2 pode-se observar que a medida que o ângulo de rotação aumenta a taxa de acerto cai, o mesmo ocorrendo quando o método é testado com imagens re-escalonadas.

Além dos autores citados, é possível encontrar na literatura trabalhos como o de Grobel e Hienz (1996) cujo objetivo foi reconhecer 32 configurações de mão da linguagem de sinais alemã. Para facilitar o rastreamento e localização da mão os autores fazem uso de uma luva com a palma da mão e os dedos pintados com cores diferentes. As imagens foram adquiridas em um ambiente controlado com um *background* escuro. Foram adquiridas 3 imagens de cada configuração de mão realizada por 5 pessoas, totalizando 15 imagens por configuração. A segmentação foi implementada por um processo de limiar simples os quais foram definidos a partir da análise dos histogramas de cada uma das imagens componentes (R, G e B). Para cada região colorida da mão (palma e os dedos) foram obtidos os seguintes parâmetros: área, centro de gravidade, razão raio máximo/raio mínimo, retangularidade/circularidade, curvatura, circunferência e orientação. A classificação trata-se de um sistema baseado em regras, onde o conjunto total é dividido em conjuntos menores de gestos baseado nas características calculadas para cada área. A taxa de acertos foi de 94,8%.

Outros autores (RODRIGUEZ; CHÁVEZ e MENOTTI, 2012; QUAN, 2010) utilizam Momentos Invariantes de Hu para formarem seus vetores de características e utilizam *Support Vector Machine* (SVM) como classificador de sinais. No entanto Rodriguez; Chávez e Menotti (2012), também realiza experimentos com momentos de Zernike e atinge uma taxa de acerto de 96% sendo superior a Quan (2010). No trabalho de Rodriguez; Chávez e Menotti (2012) aplicou-se limiarização para segmentação das imagens enquanto Quan (2010) modelou a distribuição de cor da pele através de uma função de densidade de probabilidade conjunta gaussiana elíptica definida pela Eq. (3), em que c é um vetor de cores, p é a dimensão do vetor, μ_s é um vetor de média e $\sum s$ é a matriz de covariância. Em seus experimentos Quan (2010) atinge 95,55% de taxa de acertos.

$$p(c|skin) = \frac{1}{(2\pi)^{p/2}|\sum s|^{1/2}} \cdot e^{-\frac{1}{2}(c-\mu_s)^T \sum_s^{-1}(c-\mu_s)} \quad (3)$$

Nos últimos anos alguns pesquisadores (PORFIRIO *et al*., 2013; YI LI, 2012 e SILVA; LAMAR e BORDIM, 2013) tem partido de uma perspectiva diferente para se obter imagens da língua de sinais, através do uso do sensor Kinect® que permite além da captura de imagens 2D *truecolor*, a da profundidade da cena. Adicionalmente, o uso do sensor de profundidade possibilita o uso de *backgrounds* complexos.

Porfirio *et al*. (2013) propôs um sistema para reconhecer as 61 configurações de mão, que são um dos parâmetros globais da LIBRAS, a partir de malhas tridimensionais pelo método *shape from silhouette* que gera a malha 3D a partir das imagens frontal e lateral da mão. O autor fez a sua própria base de dados contendo 610 vídeos de cinco usuários distintos utilizando o sensor Kinect®. A segmentação das imagens é facilitada graças ao canal de profundidade deste sensor. O classificador SVM foi escolhido para distinguir as malhas 3D. A partir do método *Spherical Harmonics* foram extraídas as características utilizadas na entrada do classificador. A maior taxa de acerto média foi de 88,69%.

Yi Li (2012) também utiliza o sensor Kinect® para o reconhecimento de nove sinais que não necessariamente pertencem a uma determinada língua de sinais, sua base de dados é composta por 4 pessoas que realizam os 9 gestos, 100 vezes. Aplica-se um algoritmo de limiar, onde a região situada entre 0,5m a 0,8m é considerado como região da mão. Identificam-se as pontas dos dedos e calcula-se o número de dedos estendidos, a direção dos dedos estendidos e ângulo entre dedos consecutivos. O método de reconhecimento proposto é baseado em regras e possui três camadas de classificação:

Primeiramente os gestos são classificados pelo número de dedos estendidos.

Na segunda etapa é analisada a combinação dos dedos estendidos, se a combinação for única entre todos os gestos o processo termina.

Finalmente, os vetores de direção de todos os dedos estendidos são determinados e todos os ângulos entre pares de dedos são calculados para, então, os gestos serem classificados de

acordo com esse ângulo. A taxa de acerto alcançou o valor de 99% para o gesto denominado "*start*".

Silva; Lamar e Bordim (2013) também utilizam o sensor Kinect® para construir seu banco de dados, aplicando uma arquitetura de casamento de modelos (*Template Matching*) como classificador. Esta técnica compara imagens de testes com um subconjunto que representa cada classe. Segundo os autores do artigo uma das principais contribuições do trabalho é a descrição das métricas de avaliação e sua discussão para o reconhecimento da representação de 26 letras do alfabeto da língua inglesa na língua de sinais americana (ASL). O seu trabalho difere dos artigos abordados até o momento porque não existe propriamente um vetor característico de uma imagem por si só, pois as métricas de avaliação são adquiridas par a par. A taxa de acertos alcança o valor de 99,03%.

Binh; Shuichi e Ejima (2005) introduzem um sistema de reconhecimento de 36 gestos da ASL em ambientes sem restrições. O sistema é dividido em 3 etapas: a primeira é o rastreamento da mão em tempo real em vídeos que encontra as regiões da imagem que correspondem a pele humana (mão e partes do rosto) e as transforma em imagens monocromáticas, sendo que a localização da mão em cada *frame* é calculada através do filtro de Kalman; a segunda trata-se do treinamento dos gestos, onde são escolhidas oito imagens dentre quinze para cada um dos gestos e por último o reconhecimento de gestos utilizando o Modelo Oculto de Markov. A taxa de reconhecimento alcançou 98%.

Marcotti *et al*., (2012) demonstra as etapas do processo de implementação de um *software* classificador de imagens, denominado ClassLib, cuja finalidade é a de traduzir os sinais da LIBRAS para português. Sua base de dados é constituída de quatro conjuntos distintos de imagens, cada qual contendo 21 fotos de sinais diferentes da LIBRAS. Para cada imagem é calculada a posição relativa da mão no plano na imagem, a área relativa do objeto (área da mão fazendo o gesto em relação à "área útil" segmentada) e o momento de inércia em torno dos

eixos horizontal e vertical. A classificação foi baseada em uma arvore de decisão gerada com o algoritmo J48 do *software* Weka. Não constam dados das taxas de acerto.

2.2 Reconhecimento de gestos em língua de sinais utilizando luvas sensoriais

Mehdi e Khan (2002) implementam um projeto chamado "*Talking Hands*", eles utilizam dados oriundos de uma luva sensorial para traduzir 26 gestos da (ASL) para a língua inglesa. A luva utilizada tem sete sensores. Cada sensor retorna um valor inteiro entre 0 e 4095 que diz respeito a inclinação no ponto onde localiza-se o sensor. Esses valores formaram os vetores de características para entrada de uma rede neural artificial. A taxa de acerto do *software* foi de 88%. Os autores justificam um valor baixo na taxa de reconhecimento devido ao fato do treinamento ter sido feito com pessoas que não conheciam a língua de sinais e que fizeram leitura dos gestos num folheto.

Wang; Leu e Oz, (2006) também utiliza uma luva sensorial para extrair características de 26 sinais do alfabeto manual e 36 configurações de mão da língua de sinais americana. No entanto ele também utiliza um *Flock of Birds* que é montada no punho da pessoa que executa o sinal para a captura da posição e orientação da mão no espaço 3-D. Esses valores somados aos valores detectados pela luva sensorial (ângulos de articulação dos dedos) são usados como dados para entrada do classificador baseado nos Modelos Ocultos de Markov. O sistema dá uma taxa média de reconhecimento de 95%.

Conforme anunciado, no Quadro 2 é apresentado o sumário dos estudos analisados.

2.3 Considerações finais

A análise dos trabalhos publicados evidenciou que a maioria das técnicas elaboradas possuem em seu seio a construção de ambientes artificiais para validar suas teorias, utilizando

iluminação controlada e fundo das imagens uniforme (na cor preta ou branca) e, em alguns casos, fazendo uso de luvas coloridas para uma melhor diferenciação dos sinais, condições estas ainda distantes dos ambientes não controlados vivenciados no dia a dia dos surdos.

Em decorrência, nos últimos anos muitos pesquisadores passaram a investigar o uso de características 3D do ambiente, viabilizadas pelo uso de um sensor de profundidade que possibilita a eliminação de muitas das restrições presentes nas técnicas referidas. Mesmo com a facilitação do uso desses sensores, a maioria das pesquisas na área, no entanto, ainda se concentra no reconhecimento dos gestos representativos de letras do alfabeto nas diversas línguas de sinais. E, conforme referido no Capítulo 1 desta dissertação, a soletração de palavras é um empréstimo linguístico e, não propriamente, o cerne das línguas de sinais. Portanto os estudos que levam em consideração a fonologia da língua de sinais mostram-se mais promissores para aplicações reais de reconhecimento de língua de sinais, como a LIBRAS.

Também como pode ser observado, no Quadro 2, a maioria dos estudos utilizada conjunto de dados próprios e limitados, impossibilitando assim o *benchmark* entre as técnicas apresentadas.

Quadro 2- Sumário da Revisão Bibliográfica

Ano/ Autor (es)	Título	Objetivos	Materiais	Técnica de Segmentação	Características	Classificador	Medida de Desempenho e Resultado
2001/ Omar Al-Jarrah, Alla Halawani	*Recognition of gestures in Arabic sign Language using neuro-fuzzy systems*	Reconhecimento dos 30 gestos do alfabeto manual árabe.	**Banco de dados**: 60 imagens por gesto (indivíduos diferentes) **Imagem**: escala de cinza **Fundo**: preto	Limiarização.	30 distâncias de pixels de borda ao centro de massa uniformemente distribuídos no intervalo de $[(90+\theta g), (113+\theta g)]$ com θg sendo a direção do gesto	Sistema *Neuro-Fuzzy*	Taxa de acerto:93,55%
2005/ Nguyen Dang Binh, Enokida Shuichi, Toshiaki Ejima	*Real-Time Hand Tracking and Gesture Recognition System*	Rastreamento da mão em tempo real e Reconhecimento dos 36 gestos da ASL em ambientes sem restrições.	**Banco de dados**: 30 frames de vídeos para cada gesto de uma única pessoa. **Imagem:** escala de cinza **Fundo:** complexo (com objetos).	Filtro de Kalman onde é prevista a localização da mão em um *frame* baseado na localização detectada no *frame* anterior.	Cinco superestados constituídos de estados (conjunto de pixels correspondentes a região da mão).	Modelos Ocultos de Markov (HMM)	Taxa de acerto: 98%
2009/ Alex T.S. Carneiro, Paulo C. Cortez, Rodrigo C. S Co sta	Reconhecimento de Gestos da LIBRAS com Classificadores Neurais a partir dos Momentos Invariantes de Hu	Reconhecimento dos 26 gestos do alfabeto manual da LIBRAS.	**Banco de dados**: 50 imagens de 3 pessoas diferentes por gesto **Imagem:** RGB **Fundo:** branco e iluminação artificial.	Limiarização no espaço de cor YCbCr $77 \leq Cb \leq 127$ $133 \leq Cr \leq 173$	6 Momentos Invariantes de Hu.	Redes Neurais Artificiais.	Taxa de acerto media: 89,67%

Ano/ Autor (es)	Título	Objetivos	Materiais	Técnica de Segmentação	Características	Classificador	Medida de Desempenho e Resultado
2010/ Ednaldo B. Pizzolato, Mauro dos Santos Anjo, Guilherme C. Pedroso	*Automatic Recognition of Finger Spelling for LIBRAS based on a Two-Layer Architecture*	Reconhecimento de 27 gestos do alfabeto LIBRAS. E 10 palavras que levam em consideração as 23 letras.	**Banco de dados:** 45 pessoas realizaram apenas uma vez cada um dos 27 gestos, vestiram uma blusa de manga longa preta, e somente a mão permanecia no campo de visão da câmera. **Fundo**: preto e iluminação artificial.	Técnica de limiar	625 pixels que compõem um quadrado de tamanho de 25x25 em torno do centro da mão	Redes Neurais Artificiais E HMM	Taxa de acerto: 91,1%
2013/ Andres Jessé Porfirio, Kelly Laís Wiggers Luiz E. S. Oliveira, Daniel Weingaertner	*LIBRAS Sign Language Hand Configuration Recognition based on 3D Meshes*	Reconhecimento das 61 configurações de mão da LIBRAS utilizando o Kinect®.	**Base de dados**: 5 pessoas realizam duas vezes todos os gestos. **Imagem:** imagem RGB e de profundidade **Fundo**: escuro e uniforme (painel azul) .	Manual	Os 7 momentos invariantes de Hu , as 8 direções de Freeman e as projeções horizontal e vertical de cada imagem	SVM	Taxa de acerto: 88,69%.
2012/ Yi Li	*Hand Gesture Recognition Using Kinect®*	Reconhecimento de nove gestos não pertencentes a língua de sinais utilizando o Microsoft Kinect® for Xbox®	**Base de dados:** 4 pessoas realizam os 9 gestos 100 vezes **Imagem:** imagem de profundidade **Fundo**: complexo	Técnica de limiar: faixa de profundidade de 0,5m a 0,8 m correspondendo a região da mão.	Número de dedos estendidos, Direção dos dedos estendidos e ângulo entre dedos consecutivos.	Sistema baseado em regras	Acurácia por gesto: Gesto “Start”:99% Gesto “Star Trek”:84%

Ano/ Autor (es)	Título	Objetivos	Materiais	Técnica de Segmentação	Características	Classificador	Medida de Desempenho e Resultado
2013/ Adithya V., Vinod P. R., Usha Gopalakrishnan	*Artificial Neural Network Based Method for Indian Sign Language Recognition*	Reconhecimento automático do alfabeto e números da língua de sinais indiano	**Base de dados**: 15 imagens de cada um dos 36 gestos. **Imagem** no espaço de cor: YCbCr **Fundo**: escuro e Iluminação artificial	Técnica de limiar: 75 < Cb < 135 e 130< Cr < 180 e Y > 80.	Segundo, terceiro e quarto momentos centrais dos coeficientes normalizados de Fourier dos vectores de projeção das linhas e colunas da imagem segmentada.	Redes Neurais Artificiais	Acurácia: 91.11%
2012/ K. C. Otiniano-Rodríguez, G. Cámara-Chávez, D. Menotti	*Hu and Zernike Moments for Sign Language Recognition*	Propõe dois métodos para reconhecimento dos 24 gestos estáticos do alfabeto da Língua Internacional de Sinais	**Base de dados**: 40 imagens para letras de A a F e 100 imagens de G a Y. **Imagem**: escala de cinza (248 x 256 pixels) **Fundo:** escuro	Técnica de limiar	Momentos de Hu E Momentos de Zernike (Usados separadamente)	Máquina de vetor de suporte (SVM)	Taxa de acerto: Com Momentos de Hu: 93% Com Momentos de Zernike: 96%
2011/ Paulraj M P, Sazali Yaacob, Mohd Shuhanaz Zanar Azalan, Rajkumar Palaniappan	*A Phoneme Based Sign Language Recognition System using 2D Moment Invariant Interleaving feature and Neural Network*	Reconhecimento de 44 gestos correspondentes aos fonemas da língua inglesa.	**Base de dados:**10 imagens por gestos de 7 indivíduos **Imagem:** RGB **Fundo**: escuro com iluminação de 15-18 lux	Técnica de limiar: região das mãos R>95 e G>40 e B>20 Max {R, G, B}- Min{R, G, B}>15, |R-G|>15 e R>G e R>B	2D-Momentos Invariantes	Redes Neurais Artificiais	Acurácia: 92.59%

Ano/ Autor (es)	Título	Objetivos	Materiais	Técnica de Segmentação	Características	Classificador	Medida de Desempenho e Resultado
2008/ Marrony N. Neris, Alexandre J. Silva, Sarajane M. Peres, Franklin C. Flores	*Self Organizing Maps and Bit Signature: a study applied on Signal Language Recognition*	Reconhecimento dos 26 gestos do alfabeto e de 46 configurações de mãos (feitos separadamente) da LIBRAS	**Banco de dados**: Dois conjuntos de imagens feitos por duas pessoas. **Fundo**: preto.	Técnica de limiar	2 vetores com 236 posições cada, correspondentes à soma dos pixels na horizontal e vertical (Assinatura de bits)	Redes Auto-Organizáveis (SOM)	Taxas de acerto: Assinatura de bits horizontal:98% Assinatura de bits vertical:98% Com Configurações de mão: Assinatura de bits horizontal: 84.7% Assinatura de bits vertical: 98.9%
2007/ Paulo Marcotti, Luciana Babberg Abiuzi1, Paloma Maria Silva Rocha Rizo, Carlos Henrique Quartucci Forster	Interface para Reconhecimento da Língua Brasileira de Sinais	Demonstrar as etapas do processo de implementação de um *software* classificador de imagens, denominado *ClassLib*, cuja finalidade é a de traduzir os sinais da LIBRAS para português.	**Base de dados**: quatro conjuntos distintos de imagens, cada qual contendo 21 fotos de sinais diferentes da LIBRAS **Imagem**:RGB (320 x 240 pixels) **Fundo**: preto	Manual	Posição relativa da mão no plano na imagem, a área relativa do objeto (Área da mão fazendo o gesto em relação à "área útil "segmentada) e o momento de inércia em torno dos eixos horizontal e vertical.	A classificação foi baseada em uma arvore de Decisão.	Não constam dados das taxas de Acertos do algoritmo.

Ano/ Autor (es)	Título	Objetivos	Materiais	Técnica de Segmentação	Características	Classificador	Medida de Desempenho e Resultado
2010/ Yang quan.	*Chinese Sign Language Recognition Based On Video Sequence Appearance Modeling*	Classificação de sinais com base em informações espaciais e temporais extraídas de sequencias de vídeos dos 30 gestos do alfabeto manual chinês.	**Base de dados**: 195 imagens para cada um dos 30 sinais. **Fundo:** branco	A distribuição de cor da pele foi modelada por uma função de densidade de probabilidade conjunta Gaussiana elíptica.	7 momentos de Hu, 48 Filtros de Gabor e 128 descritores de Fourier**.**	Máquina de vetor de suporte (SVM)	Média das Taxa de acerto: 95.55%
1996/ Kirsti Grobel, Hermann Hienz	*Video-Based Handshape Recognition Using a Handshape Structure Model in Real Time*	Reconhecimento de 32 configurações de mão da linguagem de sinais alemã em tempo real utilizando uma luva colorida.	**Base de dados**: 5 pessoas realizam 3 vezes cada uma das 32 configurações de mão. **Imagem**: RGB (768 x 512 pixels) **Fundo:** escuro	Técnica de limiar	Para cada região colorida da mão é calculada o tamanho, o centro de gravidade, raio máximo / raio mínimo, retangularidade / circularidade, curvatura, circunferência e orientação.	Sistema baseado em regras.	Taxa de acerto: 94,8%

Ano/ Autor (es)	Título	Objetivos	Materiais	Técnica de Segmentação	Características	Classificador	Medida de Desempenho e Resultado
2013/ Juarez Paulino da Silva Júnior, Marcus Vinícius Lamar, Jacir Luiz Bordim	*A Study of the ICP Algorithm for Recognition of the Hand Alphabet*	Reconhecimento do alfabeto manual da ASL através de uma investigação do ICP (*Iterative Closest Point)* como um algoritmo para o casamento de formas 3D.	**Base de dados**: 20 amostras de cada uma das 26 letras a partir de um único usuário. **Imagem**: profundidade. **Fundo:** complexo	Técnica de limiar: Quadros de 128x128 pixels onde cada *pixel* representa a distância de profundidade das mãos variando de 70cm a 110cm do Kinect®	Não existe propriamente um vetor característico de uma imagem por si só, as métricas de avaliação são adquiridas par a par.	Casamento de modelos (*Template Matching*).	Acurácia: 99,0385%

CAPÍTULO 3

FUNDAMENTAÇÃO TEÓRICA

3.1 Introdução

Neste Capítulo serão apresentados os conceitos necessários para um melhor entendimento desta dissertação. Os assuntos apresentados versam sobre sensor de profundidade Kinect®, segmentação de imagens, métodos de interpolação utilizados em imagens, técnicas de redução de dimensionalidade em imagens digitais com ênfase nas técnicas baseadas na Análise das Componentes Principais (PCA). O classificador dos k-vizinhos mais próximos também é abordado neste Capítulo.

3.2 Dispositivo kinect®

3.2.1 Contexto histórico

O sensor de profundidade usado no Kinect® foi desenvolvido por Zeev Zalevsky, Alexander Shpunt, Aviad Maizels e Javier Garcia, em 2005. O Kinect® foi anunciado pela primeira vez no dia 1° de junho de 2009, sob o nome "Projeto Natal", na E3, *Electronic Entertainment Expo*. O nome "natal" faz referência a cidade brasileira de Natal no Rio Grande do Norte, isso porque um dos pesquisadores responsáveis pelo projeto, Alex Kipman, é do Brasil e escolheu o nome em homenagem a cidade. (WINGFIELD, 2010).

O sensor Kinect® foi lançado oficialmente em novembro de 2010 (MICROSOFT, 2010) para ser integrado a plataforma de jogos Xbox 360 da Microsoft. Este dispositivo foi desenvolvido pela empresa israelense *PrimeSense* em colaboração com a *Microsoft*.

Nos primeiros 60 dias depois do lançamento, mais de 8 milhões de unidades foram vendidas. Assim, o Kinect® tornou-se o consumível eletrônico vendido mais rapidamente do livro de recordes *Guinness Book*. Em janeiro de 2012, o Kinect® já tinha atingido mais de 18 milhões de unidades vendidas. (CRUZ; LUCIO; VELHO, 2012).

O Kinect® surgiu como um dispositivo inovador porque possui uma câmera capaz de detectar a posição do ambiente à sua frente em três dimensões, além disso também possui uma câmera de vídeo colorida que captura as três cores primária aditivas: vermelho, verde e azul, além de outras funcionalidades. Integrado ao vídeo game, o dispositivo permite interação dos usuários com o game a partir da realização de gestos e comando por voz.

Devido as suas funcionalidades e ao custo relativamente baixo, a comunidade cientifica logo interessou-se em utilizar o dispositivo em outras áreas como robótica e medicina. Nas primeiras aplicações, pesquisadores conectavam o sensor Xbox Kinect® aos seus computadores e utilizam a biblioteca de uso livre, chamada OpenNI, para desenvolver as mesmas. (OPENNI,2010).

Devido a pressão por parte dos desenvolvedores de aplicações, em fevereiro de 2012 a Microsoft lançou uma nova versão do sensor Kinect® chamada Kinect® for Windows (MICROSOFT, 2012), além de uma *Software Development Kit* (SDK) oficial para computadores que utilizam o sistema operacional Windows.

3.2.2 Componentes do Kinect®

O Sensor Kinect® é composto por uma câmera de profundidade, um emissor de luz infravermelha (IR), uma câmera colorida RGB, um conjunto de quatro microfones, um motor

de inclinação e um cabo USB. Na Figura 4 são mostrados os três primeiros e na Figura 5 todos os componentes do Kinect®.

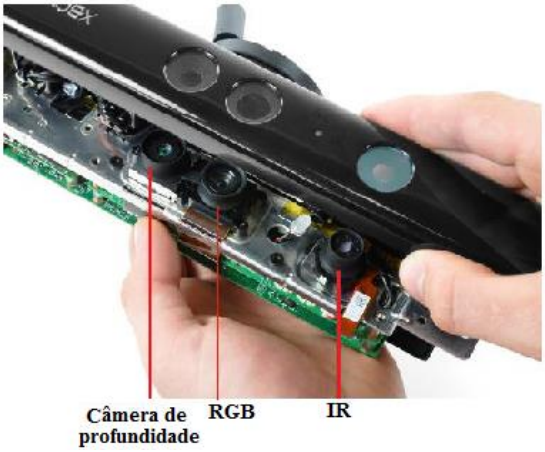


Figura 4 - Interior do Kinect® mostrando três componentes.

Fonte: Anjo, 2012.

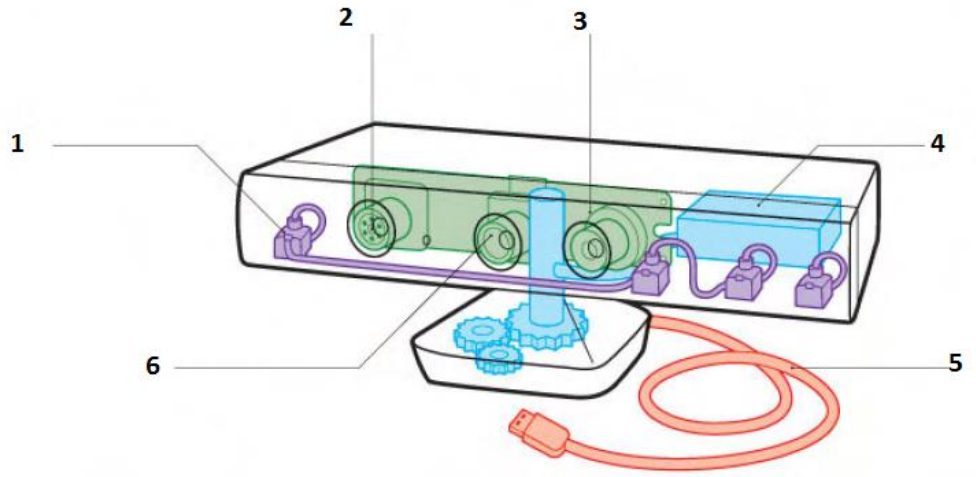


Figura 5 -Todos os componentes do Kinect® .1- conjunto de microfones, 2- Emissor de IR, 3- câmera de profundidade, 4-motor de inclinação, 5-cabo USB, 6- câmera colorida RGB.

Fonte: Ramos, 2012.

3.2.2.1 Sensor de profundidade

Sistema constituído por um emissor e por uma câmera de infravermelhos. Segundo Silva (2013) a câmera IR funciona com uma frequência de 30 Hz e transmite imagens de 640x480 pixels de 11 *bits* de profundidade resultando numa faixa dinâmica de 2048 níveis. O campo de visão é de 58º graus horizontais, de 45º graus verticais e de 70º graus na diagonal. Ressalta-se que para obtermos bons resultados o alcance deve estar entre 0,8 m e 3,5m. (SILVA, 2013).

3.2.2.2 Câmera RGB

É utilizada para obtermos valores correspondentes a cor e textura do ambiente no mapa de profundidade. Segundo Silva (2013) da mesma forma que a câmera IR, a câmera RGB também funciona a uma frequência de 30 Hz e transmite imagens de 640x480 pixels com uma resolução de intensidade de 8 bits. O Kinect® possui também a opção de aumentar a resolução da câmera, operando a 10 Quadros por segundo (fps) e 1280x1024 pixels. A própria câmera possui um conjunto de recursos incluindo balanço automático de brilho, saturação de cor, correção de defeitos e interferências.

3.2.3 Funcionamento do Kinect®.

Ramos (2012) relata que a tecnologia para aquisição de imagens pelo Kinect® é baseada no *Light coding*^{TM} desenvolvido pela *PrimeSense* que codifica o ambiente com raios

infravermelhos, a qual são considerados de classe 1 segundo o *Laser Safety* e são compatíveis com a norma 60825-1 da *International Electrotechical Commission* (IEC).

O Kinect® utiliza um sensor de imagem CMOS padrão para ler os raios que foram refletidos pela cena e transfere-os ao microcontrolador *PrimeSense* PS1080-SoC. Este microcontrolador controla a luz com uma estrutura bem definida que permite descobrir informações sobre a cena baseado na forma como essa luz é refletida através de um difusor de luz e processa os dados do sensor de imagem para fornecer uma imagem de profundidade do ambiente em tempo real (VILLAROMAN, 2011).

3.2.4 Limitações do Kinect®

Da análise dos trabalhos científicos, referidos anteriormente, que empregam o Kinect®, pode-se abstrair as seguintes limitações deste dispositivo:

- Não é aconselhável utilizar o Kinect® em lugares que recebam iluminação solar, porque parte do espectro total da radiação eletromagnética fornecida pelo Sol é composta por raios infravermelhos prejudicando a leitura dos dados de profundidade;
- Quando existem dois ou mais objetos na mesma direção, mas em diferentes profundidades, o objeto mais distante é ocluído.
- Quando se pretende usar o mapa de profundidade e a imagem RGB simultaneamente é preciso realizar uma calibração antes da aquisição das imagens.

3.3 Segmentação de imagens

3.3.1 Definição

Na segmentação são identificadas as regiões de interesse na imagem.

> Em geral a segmentação automática é uma das tarefas mais difíceis em processamento de imagens. Esse passo determina o eventual sucesso ou fracasso na análise. De fato, a segmentação efetiva quase sempre garante sucesso no reconhecimento. Por essa razão, um cuidado considerável deve ser tomado para se melhorar as chances de uma segmentação robusta (GONZALEZ, 2000 p. 295)

Existem várias formas de segmentação de imagens como por exemplo análise de histogramas, crescimento de regiões, divisão-e-união e agrupamento.

Neste trabalho a segmentação utilizada é baseada no crescimento de regiões. Portanto a próxima Seção foca-se na discussão desta técnica e sua aplicação em imagens digitais.

3.3.2 Segmentação por crescimento de regiões

Segundo Pedrini e Schwartz (2008) o processo de segmentar uma imagem utilizando o crescimento de regiões significa agrupar pixels ou sub-regiões formando regiões significativas através de um procedimento pré-definido para o crescimento. Portanto, esta técnica agrupa regiões com propriedades similares (por exemplo: nível de cinza, textura ou cor) a partir de um grupo inicial de pixels (semente) em regiões maiores até atingir um critério de parada.

Dois problemas devem ser respondidos antes da aplicação desta técnica:

- Seleção de sementes que representem de maneira fiel as regiões de interesse.
- Seleção de critérios de similaridade que dependem tanto da problemática do projeto e também do tipo de imagem que será utilizada.

O crescimento de regiões utiliza um conjunto de descritores que se baseiam em propriedades espaciais de uma única fonte de imagens. Mas a informação de conectividade ou de adjacência deve ser considerada no processo de crescimento de regiões porque empregar descritores de maneira independentes pode levar a resultados enganosos (GONZALEZ; WOODS, 2000).

A seguir apresenta-se um algoritmo de crescimento de regiões baseados em conectividade 8. Onde os pixels vizinhos de conectividade 8 de um *pixel* são os pixels posicionados ao redor desse *pixel.*

Faz-se:

- $f(x, y)$ designar um *array* imagem de entrada;
- $S(x, y)$ designar um *array* semente contendo 1's nas localizações de pontos sementes e 0's nas demais localizações;
- Q designar um predicado a ser aplicado em cada localização (x,y).

Assume-se que os *arrays* f e S são do mesmo tamanho.

Primeiramente determina-se todos os componentes conectados em $S(x, y)$ e efetua-se uma erosão em cada conjunto conectado de tal forma a reduzi-los para 1 *pixel*; rotula-se esses pixels semente com o valor 1. Designa-se 0 para todos os outros pixels em S;

Em seguida, constrói-se uma imagem f_Q na qual cada par de coordenada (x,y) obedece as seguintes condições:

O valor de $f_Q(x, y)$ será igual a 1 se o predicado Q na referida coordenada na imagem de entrada for atendido e será zero caso contrário.

Compõe-se uma imagem binária g na qual o valor 1 é atribuído a todos os pixels em f_Q que sejam 8-conectados a um *pixel* semente contido em S.

E finalmente rotula-se cada componente conectado em g com uma região diferente. Esta será a imagem segmentada obtida através da aplicação do algoritmo de crescimento de regiões.

3.4 Métodos de interpolação

O termo “interpolar” é quase sempre utilizado com significado de “reconstruir”. Interpolar é predizer (ou estimar) o valor da variável em estudo num ponto não amostrado (LANDIM, 2002);

A interpolação de imagens é utilizada em um grande número de aplicações na área de processamento de imagens digitais, tais como redimensionamento, deformação e restauração de imagens. Segundo Rézio; Schwartz e Pedrini (2011) há vários métodos de interpolação e existem três que são mais comumente utilizados para a reamostragem dos pixels (formação de uma nova imagem a partir da imagem inicial): interpolação do vizinho mais próximo (interpolação de ordem zero), interpolação bilinear e interpolação bicúbica.

Neste trabalho foram realizados testes com esses três métodos de interpolação para o redimensionamento das imagens (Seção 4.10) e também na aplicação de ampliação de imagens do banco de dados (Seção 4.8). Nas seções seguintes estes métodos são analisados e no Capítulo 5 desta dissertação será discutido o impacto na escolha de cada um desses métodos nas taxas de acertos finais para classificação das configurações de mão.

3.4.1 Vizinho mais próximo (Interpolação de ordem zero)

Segundo Akyama (2010) o método do vizinho mais próximo replica o valor do *pixel* geometricamente mais próximo ao novo *pixel* da imagem. É um processo rápido e fácil de implementar. A Figura 6 mostra um exemplo da aplicação do método do vizinho mais próximo.

$$J(x,y)=\begin{bmatrix}8 & 5 & 7\\ 4 & 2 & 3\\ 6 & 1 & 9\end{bmatrix} \qquad V(x,y)=\begin{bmatrix}8 & 8 & 5 & 5 & 7\\ 8 & 8 & 5 & 5 & 7\\ 4 & 4 & 2 & 2 & 3\\ 4 & 4 & 2 & 2 & 3\\ 6 & 6 & 1 & 1 & 9\end{bmatrix}$$

(a) (b)

Figura 6- Aplicação do método do vizinho mais próximo. a) vetor bidimensional original b) aplicação do método.

Fonte: Akyama, 2009.

3.4.2 Interpolação bilinear

Silva (2009) explica que a interpolação bilinear trata-se de uma extensão da interpolação linear para aplicação em funções de duas variáveis. O objetivo é aplicar a interpolação linear em uma direção e depois em outra direção perpendicular à primeira.

Por exemplo para encontrar o valor de uma função desconhecida f no ponto $P = (x, y)$. utilizam-se os valores de quatro pontos conhecidos de f, $Q_{11} = (x_1, y_1)$, $Q_{12} = (x_2, y_2), Q_{21} = (x_3, y_3)$ *e* $Q_{22} = (x_4, y_4)$ de acordo com a Figura 7.

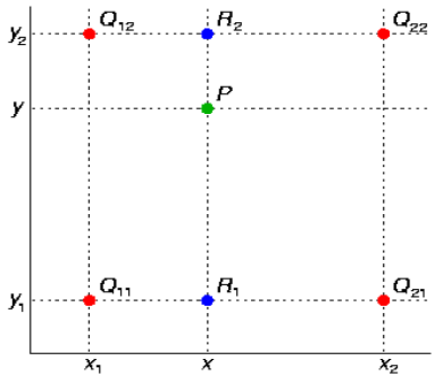


Figura 7 - Interpolação bilinear

Fonte: Silva, 2009.

Primeiro executa-se a interpolação linear na direção x de acordo com a Eq. (4) e a Eq. (5).

$$f(R_1) = \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x_2 - x}{x_2 - x_1} f(Q_{21}) \quad (4)$$

$$f(R_2) = \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x_2 - x}{x_2 - x_1} f(Q_{22}) \quad (5)$$

Em que $R_1 = (x, y_1)$ *e* $R_2 = (x, y_2)$. Em seguida aplica-se a interpolação na direção y, de acordo com a Eq. (6):

$$f(P) = \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y_2 - y}{y_2 - y_1} f(R_2) \quad (6)$$

3.4.3 Interpolação bicúbica

Silva (2009) afirma que a interpolação bicúbica preserva detalhes presentes na imagem ao custo de tempo adicional para execução da interpolação. Neste tipo de interpolação, o valor f (x, y) de uma função f no ponto (x, y) é calculado como uma média ponderada dos dezesseis pontos mais próximos a ele, montando uma matriz 4x4. Dois polinômios de ordem três de interpolação são utilizados, um para cada direção. A interpolação bicúbica é calculada pela Eq. (7).

$$f(x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j \quad (7)$$

Por exemplo suponhamos que se deseja encontrar o ponto $P = (j + x, k + y)$, ilustrado na Figura 8. As equações para interpolação segundo o eixo horizontal e vertical são dadas respectivamente pela Eq. (8) e pela Eq. (9):

$$a_{j+x,k} = \frac{1}{6}(a_{j-1,k}R_1 + a_{j,k}R_2 + a_{j+1,k}R_3 + a_{j+x+2,k}R_4) \quad (8)$$

$$a_{j+x,k} = \frac{1}{6}(a_{j+x,k-1}R_1 + a_{j+x,k}R_2 + a_{j+x,k+1}R_3 + a_{j+x,k+2}R_4) \quad (9)$$

Onde os coeficientes R_1a R_4são dados pelas Eq. (10) a Eq. (13):

$$R_1 = (3 + x)^3 - 4(2 + x)^2 + 6(1 + x)^3 - 4x_3 \quad (10)$$

$$R_2 = (2 + x)^3 - 4(2 + x)^3 + 6x^3 \quad (11)$$

$$R_3 = (1 + x)^3 - 4(2 + x)^3 \quad (12)$$

$$R_4 = x^3 \quad (13)$$

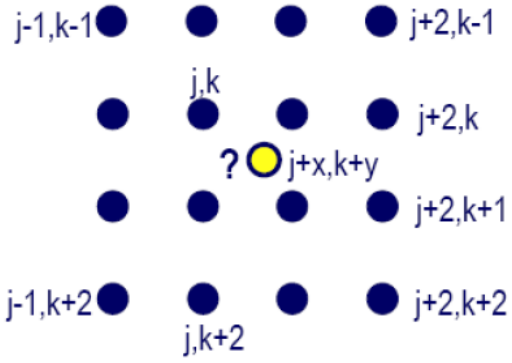


Figura 8 - Interpolação bicúbica

Fonte: Silva, 2009.

3.5 Técnicas de redução de dimensionalidade

Segundo Johnson e Wichern (2002) a redução dos dados ou simplificação dos dados é um dos principais objetivos de se utilizar métodos de análise estatística multivariada, onde o fenômeno estudado é apresentado de forma reduzida, mas sem perder valores significativos de informações.

Existem várias técnicas para redução da dimensão de dados, como regressões múltiplas, correlações canônicas e análises de discriminantes. Um dos métodos estatísticos para redução de dados mais utilizados é o método da Análise de Componentes Principais que será abordado na Seção seguinte.

3.5.1 Análise de Componentes Principais (PCA)

Tino (2005) afirma que a análise das componentes principais é uma técnica de transformação de variáveis. Se cada variável medida pode ser considerada como um eixo de variabilidade, estando correlacionada com outras variáveis, esta análise transforma os dados a fim de descrever a mesma variabilidade total existente, com o mesmo número de eixos originais, porém não mais correlacionados entre si.

A análise de componentes principais avalia a matriz de covariância dos padrões a fim de constatar covariância nula entre as características.

Considere um padrão de treinamento representado pelo vetor a, de dimensão mn. A matriz de covariância, S, é calculada a partir dos N vetores de treinamento $A = [a_1, a_{2,\dots} a_N]$ dada pela Eq.(14):

$$S = \frac{1}{N}(A - M)(A - M)^T \tag{14}$$

Em que M é a matriz cujas colunas contêm as médias dos valores das colunas de A, ou seja, sua i-ésima coluna é dada pela Eq.(15):

$$M_i = \frac{1}{N}\sum_{j=1}^{N} a_j \tag{15}$$

Para $i = 1, 2, 3, \ldots N$. Da definição, a matriz de covariâncias tem a forma dada pela Eq. (16):

$$S = \begin{bmatrix} c_{11} & \cdots & c_{1mn} \\ \vdots & \ddots & \vdots \\ c_{m1} & \cdots & c_{mm} \end{bmatrix} \tag{16}$$

Esta matriz possui as variâncias na diagonal principal e nas demais posições a covariância entre as direções. Essa matriz é simétrica e real, de modo que sempre é possível encontrar um conjunto de autovetores ortonormais (ANTON e RORRES, 2004).

O próximo passo é encontrar a matriz de transformação X formada pelos q autovetores $[u_1, u_{2,\ldots} u_q]$ associados aos q maiores autovalores da matriz de covariância S. Assim, para a redução de dimensão, a matriz de transformação é da forma dada pela Eq. (17):

$$X = [u_1, u_{2,\ldots} u_q] \tag{17}$$

Dado um conjunto de treinamento A, o algoritmo PCA encontra um transformador linear X de forma a representar A em outro espaço, Y, dado pela Eq.(18) a qual possui matriz de covariância diagonal.

$$Y = X^T A \tag{18}$$

3.5.2 Análise de componentes principais bidimensional (2DPCA)

Como vimos na Seção anterior para calcular as componentes principais de uma imagem (2D), esta é transformada em vetor de imagem (1D). Segundo Yang *et al* (2004) esse procedimento resulta em um vetor com espaço de dimensão elevada. A análise de componentes principais bidimensional (2DPCA) foi proposta em Yang *et al* (2004). Esta técnica não transforma a imagem 2D em um vetor 1D, a matriz de covariância é construída diretamente das matrizes das imagens de treinamento.

Seja A_i a i-ésima matriz das imagens de treinamento, de dimensão mxn, $\overline{A}$ a média de todas as imagens de treinamento e N é o número de imagens de treinamento. No método 2DPCA, a matriz de covariâncias G_H (matriz de dispersão total) do conjunto de treinamento é obtido pela Eq. (19).

$$G_H = \frac{1}{N}\sum_{i=1}^{N}(A_i - \overline{A})^T (A_i - \overline{A}) \tag{19}$$

Da mesma forma que ocorre no PCA convencional, a matriz X que realiza a transformação linear é formada pelos autovetores que correspondem aos q maiores autovalores da matriz de covariância G_H.

$$Y = AX \tag{20}$$

Na Eq. (20) a dimensão da matriz de projeção X é $n\ x\ q$ e, como consequência, a dimensão da matriz de característica Y é $m\ x\ q$, com $q < n$.

3.5.3 2DPCA bidirecional ou (2D)²PCA

O método 2DPCA realiza a redução de dimensão somente na direção horizontal de uma imagem. Zhang e Zhou (2005) propôs o método 2DPCA bidirecional com o intuito de fazer reduções na direção horizontal e vertical. Esta técnica também conhecida como (2D)²PCA consiste em encontrar duas matrizes de dispersão: $X = [q_1, q_{2,\dots} q_n]$ e $Z = \left[z_1, z_{2,\dots} z_n\right]$.

A matriz X é obtida pelo método 2DPCA de acordo com a Seção anterior e a matriz Z é obtida pelo método "*alternative* 2DPCA", que toma por base o cálculo da matriz de dispersão total.

$$G_v = \frac{1}{N}\sum_{i=1}^{N}(A_i - \overline{A})(A_i - \overline{A})^T \tag{21}$$

Feito isso, calcula-se os autovetores da matriz G_v que servirão para realizar a projeção da imagem original, reduzindo a dimensão na direção vertical. A matriz G_v possui dimensão $m\ x\ m$. Sendo X de dimensão $n\ x\ d$ e Z de dimensão $m\ x\ d$. A matriz de características C_k , de dimensão $d\ x\ q$, é obtida projetando-se a imagem A_k nas matrizes X e Z simultaneamente e expressa pela Eq.(22):

$$C_k = Z^T A_k X \tag{22}$$

3.6 Classificador k-vizinhos mais próximos

Segundo Oliveira (2008) esta técnica realiza a classificação de um novo objeto, com base no conhecimento das classes dos objetos que se encontram mais próximos deste, no espaço

de atributos. Estes métodos precisam de uma métrica para comparar as distâncias entre os diferentes objetos. A distância Euclidiana é um tipo de métrica bastante utilizada para classificação por vizinhança. Mas pode-se usar toda a família de distâncias Minkowski como critério de avaliação dos vizinhos mais próximos dada pela Eq.(23):

$$d(x,y) = (\sum_{i=1}^{n} |x_i - y_i|^r)^{\frac{1}{r}} \; com \; \text{r} \; \epsilon \; \text{IR}, \text{r} \geq 1 \qquad (23)$$

Para r =1: distância Manhattan ou Hamming;

Para r =2: distância Euclidiana;

Para r =∞ : distância Tshebyshev;

Para o caso em que o número de vizinhos k é igual a 1, realiza-se o cálculo de todas as distâncias entre um objeto de teste e todos os objetos do conjunto de treino. Para cada objeto de teste apresentado é atribuído a classificação do seu vizinho mais próximo (Figura 9).

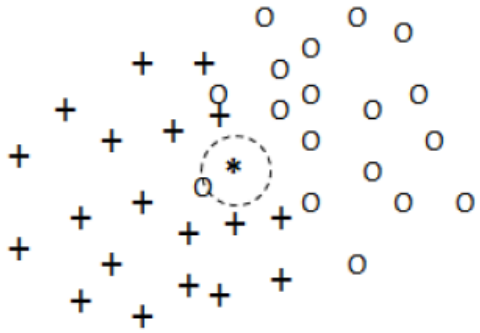


Figura 9 - Regra do vizinho mais próximo com k=1

Fonte: Silva, 2008.

A regra dos k-vizinhos mais próximos (KNN) é uma extensão da regra do vizinho mais próximo. Leva-se em consideração não apenas a classe de um único vizinho mais próximo, mas sim das classes dos k (número a ser definido) vizinhos mais próximos. Dessa forma, a

classe atribuída ao objeto de teste será a classe mais votada entre os k-vizinhos mais próximos do conjunto de treinamento (Figura10).

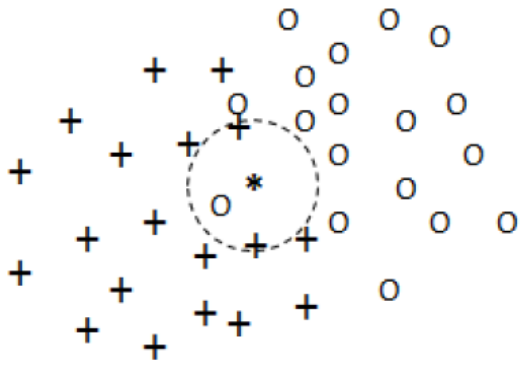


Figura 10 - Regra do vizinho mais próximo com k=3.

Fonte: Silva, 2008.

CAPÍTULO 4

MATERIAIS E MÉTODOS

O desenvolvimento deste trabalho teve como foco atuar no reconhecimento de configurações de mão da LIBRAS em cenas capturadas pelo dispositivo Kinect® em ambiente fechado.

A Figura 11 traz um diagrama de blocos da proposta deste projeto.

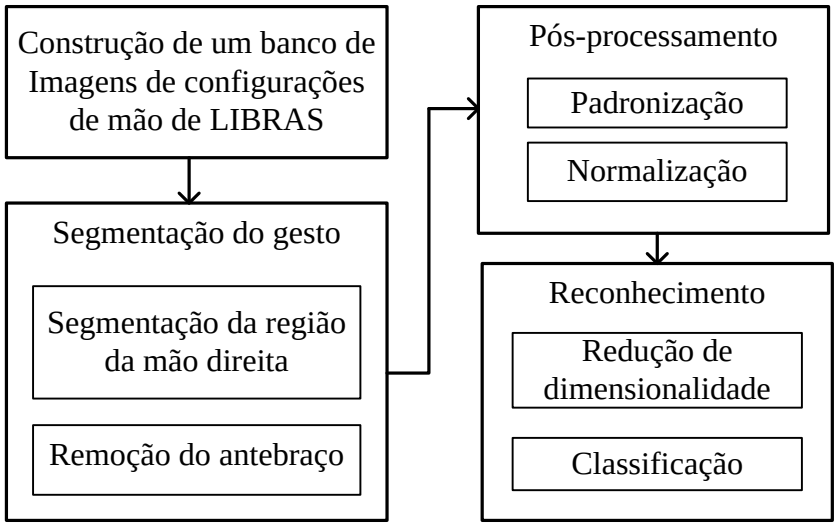


Figura 11- Diagrama em blocos da proposta

4.1 Materiais

4.1.1 Ambiente de desenvolvimento

O desenvolvimento dos algoritmos desse projeto foi realizado inteiramente no MATLAB ® versão 2014ª. Esses algoritmos implementam as etapas de segmentação, pós-processamento e reconhecimento do gesto, referidas na Figura 11.

A infraestrutura de *hardware* utilizada para o desenvolvimento da dissertação foi um computador pessoal (PC) e um sensor de profundidade, Kinect®. Este último foi conectado ao computador via cabo USB. Os dados foram transmitidos através de um *feed* sem criptografia.

O computador utilizado neste projeto possui as características apresentadas no Quadro 3.

Quadro 3- Características do computador utilizado

Sistema Operacional	*Windows seven Ultimate*
Processador	Intel(R) Core (TM) i3 2,0GHz
Memória RAM	3,00GB
Tipo de Sistema	Sistema Operacional de 64 *bits*

As características do Kinect® utilizado são as relatadas na Seção 3.2.2. Na Figura 12 é mostrado tal dispositivo.

Figura 12 - Kinect® utilizado para captura das imagens.

4.1.2 Construção do Banco de dados de imagens

Tendo em vista a indisponibilidade de imagens do fonema “configuração de mãos” de LIBRAS e, adicionalmente, com a finalidade de contribuir para o avanço da pesquisa na área,

foi construído um banco de imagens robusto e representativo de configurações de mãos de LIBRAS.

Ressalta-se que a construção do referido banco foi realizada em conjunto com o mestrando Jonilson Roque dos Santos, que também desenvolveu dissertação dentro da mesma temática (SANTOS, 2015).

O esquema de configuração física utilizado pode ser visto na Figura 13. Onde se pode observar que a distância do Kinect® ao *background* foi fixada em torno de 1,40m.

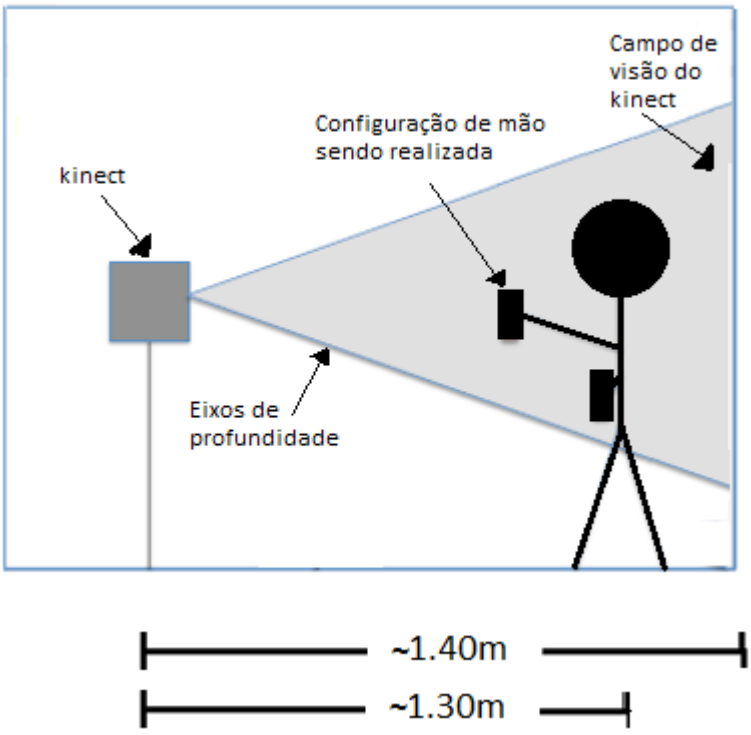


Figura 13 - Esquema de configuração física utilizada para adquirir as imagens das configurações de mão.

Para aquisição das imagens foi utilizado o próprio pacote de desenvolvimento de *software* (SDK) do dispositivo Kinect®.

As imagens foram adquiridas em dois ambientes físicos: 1) nas dependências da Escola Estadual Augusto Carneiro dos Santos, uma Escola de Educação Especial para surdos, situada

na Avenida Joaquim Nabuco 2274, Praça 14 de Janeiro, Manaus e 2) no Laboratório de Processamento Digital de Imagens do Centro de Tecnologia Eletrônica e da Informação-CETELI da Universidade Federal do Amazonas – UFAM. Exatos 70% do banco de dados de imagens foram criados no ambiente 1 e os outros 30% no ambiente 2.

Dez pessoas colaboraram para obtenção das imagens, das quais 7 são alunos surdos com idade entre 18 e 20 anos e que, portanto, conheciam as configurações de mão. Os outros 3, situavam-se na faixa de idade de 24 e 25 anos, não apresentavam tal deficiência, nem possuíam familiaridade com a LIBRAS. Uma outra característica desse banco é possuir diversidade de gênero: 8 homens e 2 mulheres. A participação da Escola e dos alunos voluntários foi devidamente autorizada pela direção da Escola, assim como os voluntários não surdos assinaram termo de consentimento. Na Figura 14 podemos ver cada umas das pessoas que, de forma voluntária, auxiliaram na construção do banco.

Figura 14 - Voluntários que contribuíram para construção do banco de dados.

Cada voluntário realizou cada uma das 61 configurações de mão de LIBRAS.

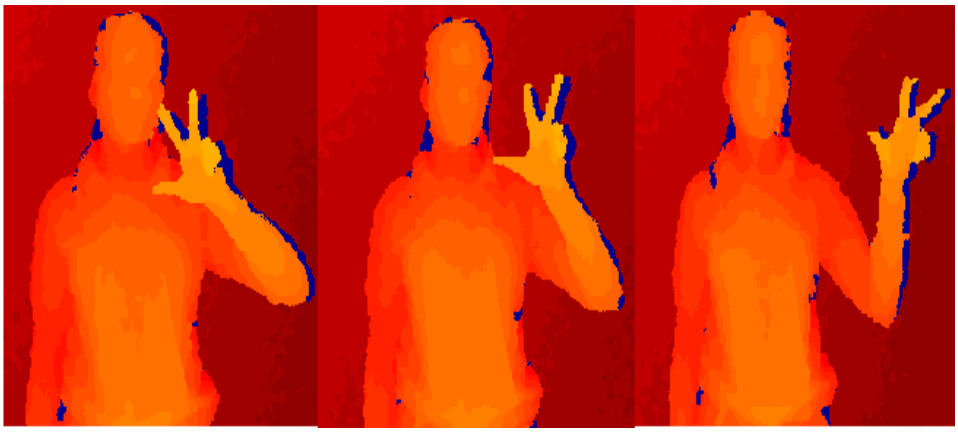
Para se obter um conjunto de dados robusto e representativo, foi solicitado aos voluntários que movessem a mão no momento da execução das configurações, pois dessa forma teríamos configurações de mão feitas em posições e inclinações diversas e inserindo assim um desafio maior ao sistema, na tarefa de reconhecimento correto das configurações de mão.

De cada uma das 61 configurações de mão foram capturadas 200 cenas (20 por indivíduo) correspondendo a posições e inclinações diversas. Dessa forma, o banco construído é composto de 12.200 cenas. Para cada cena foram gerados dois arquivos: um corresponde a imagem *truecolor* no formato *bitmap* de 640x480 pixels e resolução de intensidade de 8 *bits*; já o segundo corresponde a um mapa de profundidade em formato *txt* com resolução de intensidade de 11 *bits*. Os dados correspondentes aos mapas de profundidade são os que foram usados no presente desenvolvimento.

A Figura 15 mostra exemplos de imagens do banco de imagens construído. Na Figura 15(a) apresenta-se a imagem *truecolor* da configuração de mão nº 36, realizada em diferentes posições e inclinações. Na Figura15(b) apresenta-se o mapa de profundidade das respectivas imagens *truecolor*.

(a)

Figura 15 - Exemplos de imagens de uma mesma configuração de mão capturadas em diferentes posições e inclinações. (a) imagem *truecolor*.



(b)

Figura 16 (continuação) - Exemplos de imagens de uma mesma configuração de mão capturadas em diferentes posições e inclinações. (a) imagem *truecolor* e (b) imagem de profundidade

4.2 Segmentação do gesto

4.2.1 Segmentação da mão direita

Como foi comentado na Seção 3.3.1 para separar a região da mão e antebraço do restante do corpo foi escolhida a técnica de segmentação por crescimento de regiões.

O Toolbox de Processamento de Imagens do MATLAB® não possui uma função para Crescimento de Região. Foi então implementada a função *regiongrow* de acordo com Gonzalez, Woods e Eddins (2009), cuja sintaxe é apresentada a seguir.

Sintaxe: *regiongrow* (F, S, T), onde:

F: é a imagem;

S: valor de intensidade para a Semente na imagem F.

T: Escalar que define um Limiar Global para a imagem F.

Como foi salientado na Seção 3.3.2 que dois problemas imediatos devem ser resolvidos para execução deste método:

- A seleção de semente que represente a região de interesse.
- Seleção de critérios de similaridade.

4.2.1.1 Seleção do ponto semente

Como mencionado na Seção 3.2.3 o emissor de infravermelho do Kinect® emite um feixe simples que é dividido em feixes múltiplos por uma rede de difração para criar um padrão de manchas constantes (pontos de nuvens) projetados na cena. O mapa de profundidade, transformado em uma imagem de profundidade, tem 640 x 480 pixels. O valor de cada *pixel* corresponde a distância do ponto que está sendo capturado até o sensor Kinect®.

Partindo do pressuposto que ao realizar gestos, os surdos sempre posicionam suas mãos à frente dos seus corpos, a definição do ponto semente (S) de cada imagem é determinado automaticamente como aquele que se situa mais próximo do dispositivo Kinect®, ou seja é aquele que apresenta a menor distância, d_{min}, da cena ao dispositivo, conforme ilustra a Figura16.

4.2.1.2 Definição do critério de similaridade

Fazendo Δd corresponder a:

$$\Delta d = d_P - d_{min} \tag{24}$$

Em que a distância entre um dado *pixel*, *p*, e o eixo de referência do Kinect® é d_p (vide Figura 16).

Então, se *p* é 8-conectado ao *pixel* semente, *S*, e se *Δd* for menor ou igual a um valor de limiar *T*, obtido experimentalmente, o *pixel*, *p*, passa a ser considerado como pertencente a região da mão direita.

.

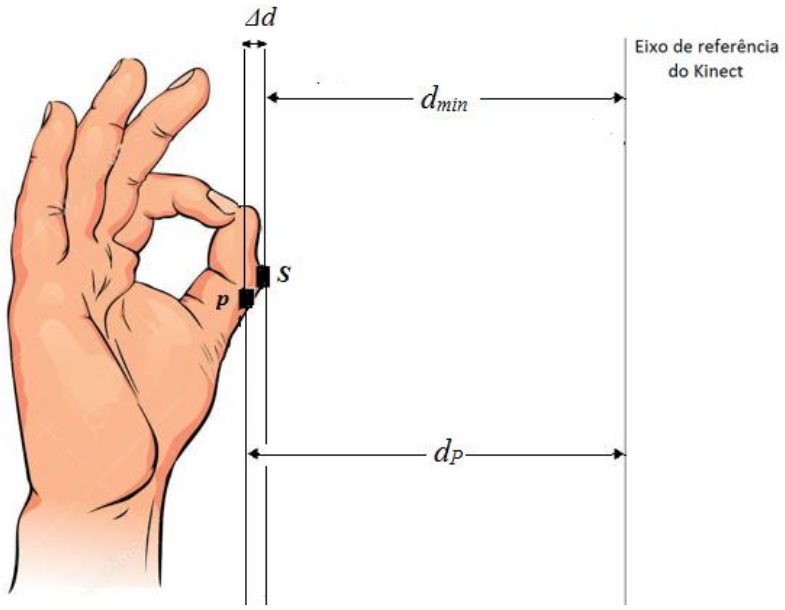


Figura 17 - Ilustração da representação do *pixel* semente S e de um *pixel* 8-conectado à semente.

Para determinação do valor de T foram feitos vários experimentos nos quais o valor de T variou de 50 mm até 100 mm, com passo de 10 mm (os valores do mapa de profundidades são fornecidos pelo Kinect® em milímetros).

Os experimentos consistiram em, para cada valor de T, aplicar o procedimento de crescimento de regiões em todas as imagens do banco e avaliar o resultado da segmentação, notadamente se somente a região da mão permanecia na imagem.

Dessa forma nos experimentos objetivou-se atingir uma grande quantidade de pixels que caracterizasse a configuração de mão. Dava-se início aos experimentos sempre na primeira imagem do banco de dados, aplicava-se determinado valor de T e avaliavam-se os resultados. Se ao aplicar o limiar, a imagem resultante não contivesse valores de pixels de profundidade

das pontas dos dedos, por exemplo, o experimento encerrava-se e iniciava-se um novo experimento com outro valor de T.

Nos experimentos iniciais que consistiram em aplicar o valor de T igual a 50mm, obtiveram-se imagens com valores de pixels pertencentes a mão, porém poucos pixels foram agrupados com esse limiar.

Nos experimentos utilizando o limiar T igual a 90mm, todas as 12.200 imagens de profundidades foram segmentadas devidamente. Ao aumentar o valor de T para 100mm, partes do corpo dos voluntários aparecem nas imagens resultantes da segmentação.

A Figura 17 mostra um exemplo de uma imagem correspondente a configuração de mão 55, onde podemos notar uma diferença significante quando se aplica um valor de T=90mm e T=100mm.

a) b)

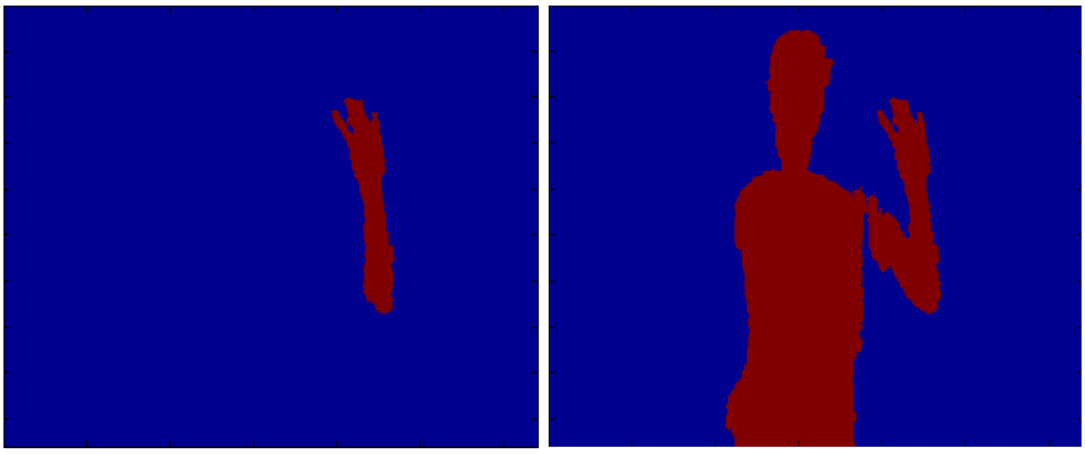


Figura 18 - Configuração de mão 55. a) limiar T=90mm, b) limiar de T=100mm

Dessa forma a análise global dos experimentos inferiu como melhor limiar o valor de 90mm.

4.2.2 Remoção do antebraço

Como observado na Figura 17(a) a região segmentada inclui parte do antebraço, pois o mesmo está, em quase todas as configurações de mão de LIBRAS, na mesma distância da mão. Dessa forma, faz-se necessário a remoção deste.

4.2.2.1 Determinação da orientação da mão

Considerando a característica do banco construído, de que cada um dos gestos se apresenta em diferentes posições e inclinações. O conhecimento da inclinação do gesto é fundamental para as etapas subsequentes.

Assim, essa etapa consistiu em determinar o ângulo θ, que define a orientação dos eixos principais de inércia em relação ao centroide da região da mão segmentada. O ângulo θ é definido pela Eq. (25):

$$\theta = \frac{1}{2} arctg\left(\frac{2\mu_{1,1}}{\mu_{2,0}-\mu_{0,2}}\right) \quad (25)$$

em que $\mu_{1,1}$, $\mu_{2,0}$ e $\mu_{0,2}$ são os momentos de segunda ordem que são calculados usando a Eq. (26):

$$\mu_{p,q} = \sum_{x=1}^{M}\sum_{y=1}^{N}(x-x_c)^p\,(y-y_c)^q\,f(x,y) \quad (26)$$

em que x_c, y_c correspondem ao centro de massa da imagem *f(x,y)* de dimensões MxN. Os valores de $p\ e\ q$ variam de 0 a 2.

Para um melhor entendimento consideremos a Figura 18. Nela, podemos observar os quadrantes do plano cartesiano onde a mão se posiciona durante a captura das imagens.

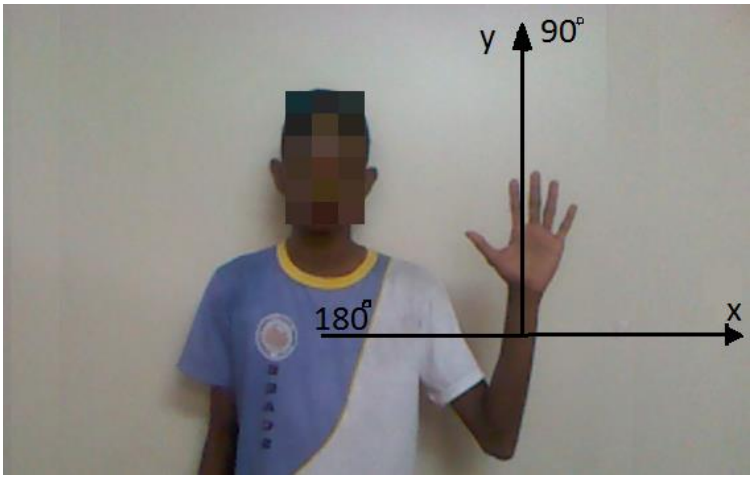


Figura 19 - Quadrantes no plano cartesiano onde as configurações de mão se posicionam.

A orientação dos gestos das imagens do banco encontra-se entre 45° e 135°

4.2.2.2 Reorientação das configurações de mão

Para o prosseguimento do projeto foi preciso reorientar as inclinações das configurações de mão para um ângulo de 90° em relação ao eixo cartesiano x positivo.

A imagem rotacionada é obtida pela aplicação da matriz de rotação (R) apresentada na Eq. (27) a cada um dos pixels da imagem segmentada.

$$R = \begin{bmatrix} cos\beta & -sen\beta & 0 \\ sen\beta & cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (27)$$

em que β é a diferença entre o ângulo calculado (θ) e o ângulo de 90°.

Para o redimensionamento das imagens foram utilizadas todas as técnicas de interpolação explicadas na Seção 3.4: vizinho mais próximo, interpolação bilinear e interpolação bicúbica.

4.2.2.3 Remoção do antebraço

Finalmente, a remoção do antebraço, propriamente dita, foi operacionalizada a partir dos seguintes procedimentos:

1. Obtenção da projeção vertical da imagem padronizada binarizada (IPB). A projeção vertical, Pv, é obtida pela contagem do número de pixels de valor 1 em cada linha da imagem. Pv é um vetor coluna de dimensões $Mx1$ em que cada elemento, Pv_i, (com *i* variando de 1 a M) é obtido de acordo com a Eq. (28):

$$Pv_i = \sum_i I_{PB}(i,j) \tag{28}$$

2. Identificação do valor da linha, $i_{Pv_{max}}$, correspondente ao maior valor projetado, $Pv_{max} = \max(Pv_i)\ com\ i = 0,1,2,3,\dots,M$, conforme ilustrada a Figura 19.

3. Obtenção da linha de corte do antebraço, i_c, ilustrada na Figura 19, a partir da Eq. (29):

$$i_c = i_{Pv_{max}} + \overline{\Delta_i} \tag{29}$$

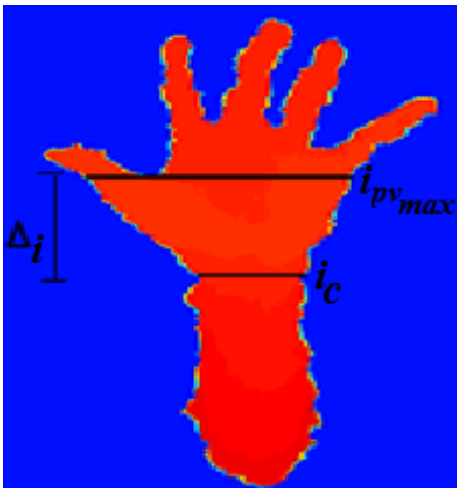


Figura 20 - Ilustração do processo de remoção do antebraço: $i_{Pv_{max}}$ linha correspondente ao maior valor da projeção vertical e i_c. linha de corte do antebraço.

O valor de $\overline{\Delta_i}$ foi definido experimentalmente, através de análise estatística de 610 imagens do banco (1 imagem/configuração/indivíduo). É calculada a média aritmética $\overline{\Delta_i}$ dos valores de Δ_i, (vide Figura 19), a qual aponta para onde mais se concentram os dados da distribuição. A expressão para calcular a média aritmética é dada pela Eq. (30):

$$\overline{\Delta_i} = \frac{1}{N} \sum_{i=1}^{N} \Delta_i \tag{30}$$

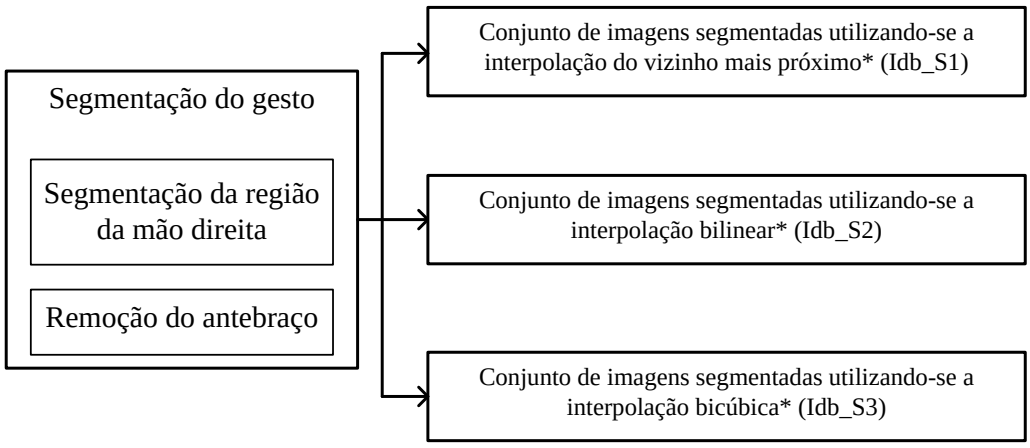
em que N corresponde ao número de imagens do experimento, no caso 610 (conjunto de treinamento).

Para determinar o valor da linha, $i_{Pv_{max}}$, de cada uma das 610 imagens é construído um algoritmo para calcular esse valor de forma automática. E para determinar o valor da linha de corte do antebraço, i_c, é realizada uma inspeção visual nas 610 imagens de profundidade e subjetivamente é determinado o valor da linha de corte para cada uma dessas imagens.

De acordo com a figura 19 é calculado o valor de Δ_i através da subtração da linha de corte, i_c, pela linha correspondente ao maior valor projetado $i_{Pv_{max}}$. Em seguida é utilizada a Eq. (30) para calcular o valor de $\overline{\Delta_i}$ a qual foi obtido o valor de 36.

Uma vez estimada a linha de corte, i_c, para todas as imagens de profundidade a região do antebraço pode, então, ser removida. Vide Figura 20.

Ressalta-se que ao final dessa etapa de segmentação ter-se-á três conjuntos de imagens segmentadas, conforme ilustra a Figura 20.



*Os conjuntos se diferenciam na etapa de reorientação dos gestos

Figura 21 - Detalhamento dos conjuntos de imagens resultantes de diferentes métodos de interpolação utilizados na etapa de reorientação dos gestos

4.3 Pós-processamento

O objetivo desta etapa é obter uma imagem padronizada e normalizada, necessária à etapa seguinte de reconhecimento das configurações de mão. A padronização compreende o recorte das imagens segmentadas no limite da região da mão para, em seguida, as mesmas serem remapeadas em uma imagem de área padrão, segundo os métodos de ampliação mencionados no Capítulo 3. Por último, a imagem padronizada é submetida a uma normalização dos pixels.

4.3.1 Padronização das imagens segmentadas

A padronização é realizada em duas fases: o recorte das imagens segmentadas no limite da região da mão e, posteriormente, o mapeamento da região da mão em uma área padrão.

O recorte das imagens se deu pela eliminação de linhas e colunas. Para isso construiu-se um algoritmo que consiste na identificação dos 4 pontos que definem os limites do retângulo que circunscreve a mão, L_1, L_2, C_1 e C_2, mostrados na imagem exemplo da Figura 21. Em seguida, calcula-se a nova dimensão MxN da imagem: M, número de linhas, é a diferença entre L_2 e L_1 e N, número de colunas é a diferença entre C_2 e C_1 como mostra a Figura 21.

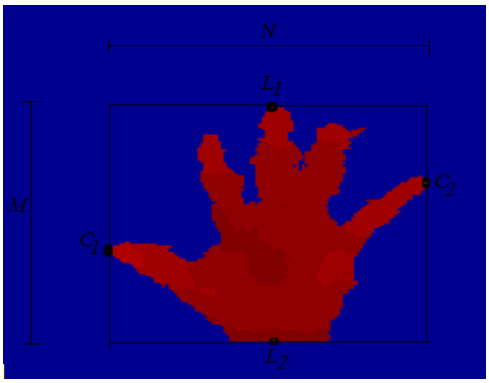


Figura 22- Imagem segmentada exemplo onde se apresentam as linhas e colunas que definem o retângulo que circunscreve a mão segmentada.

O algoritmo de recorte das imagens apresentado resulta em imagens de dimensões diversas. No entanto, o cálculo da matriz de covariância na etapa de extração de características, requer que as imagens possuam a mesma dimensão. Dessa forma, as imagens recortadas são submetidas a um processo de padronização de área através do remapeamento dos pixels para uma dimensão padrão utilizando os processos de interpolação já mencionados no Capítulo 3. A dimensão padrão definida é aquela correspondente ao maior valor de M e de N apresentados pelas imagens do banco. Dessa forma as imagens deverão ter 135x139 pixels.

Assim, os fatores de ampliação vertical (n_{v_i}) e horizontal (n_{h_i}) da *i-ésima* imagem do banco são dados pelas Equações (31) e (32), respectivamente:

$$n_{v_i} = \frac{135}{M_i} \tag{31}$$

$$n_{h_i} = \frac{139}{N_i} \tag{32}$$

As figuras 22 e 23 mostram exemplos do resultado da padronização das imagens.

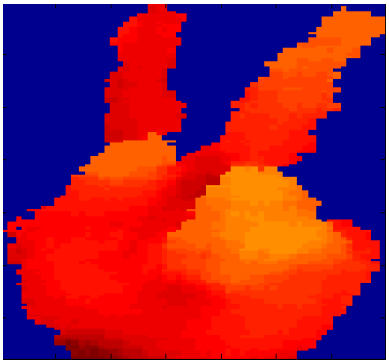


Figura 23 - Exemplo de padronização de uma imagem da configuração de mão 50, utilizando-se o método de interpolação bicúbica.

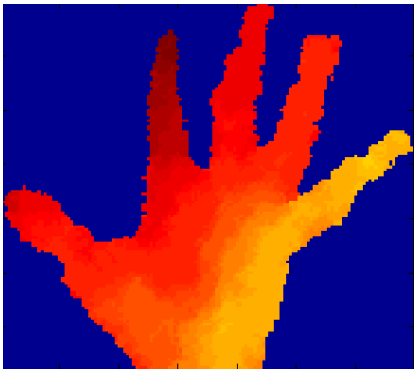


Figura 24 - Exemplo de padronização de uma imagem da configuração de mão 61, utilizando-se o método de interpolação bicúbica.

4.3.2 Normalização dos valores dos pixels

Conforme já mencionado, a distância do *background* ao Kinect® foi fixado em 1,40m, enquanto os voluntários se posicionavam a uma distância média de 1,30m. Nenhuma restrição foi feita à posição da mão, de tal forma que os mesmos ficassem livres para realização das configurações das mãos naturalmente.

Dessa forma houve a necessidade de normalizar a profundidade dos gestos. O método proposto é realizado em todos os pixels diferentes de zero e consiste em subtrair as distâncias de profundidade de cada *pixel* pela menor distância presente no mapa.

Recordando, na etapa de segmentação do gesto o conjunto de imagens inicial gerou 3 conjuntos distintos de imagens segmentadas (Idb_S1, Idb_S2, Idb_S3) em face da utilização de três métodos de interpolação na etapa de remapeamento, para o estabelecimento de todos os gestos em uma mesma orientação. Da mesma forma, esses três conjuntos de imagens, multiplicam-se em face da utilização da interpolação vizinho mais próximo, bilinear e bicúbica, na etapa de padronização da mão segmentada. A Figura 24 ilustra 9 conjuntos distintos de imagens. Nela, P1, P2 e P3 referem-se aos conjuntos de dados gerados com os métodos de interpolação: vizinho mais próximo, bilinear e bicúbica, respectivamente.

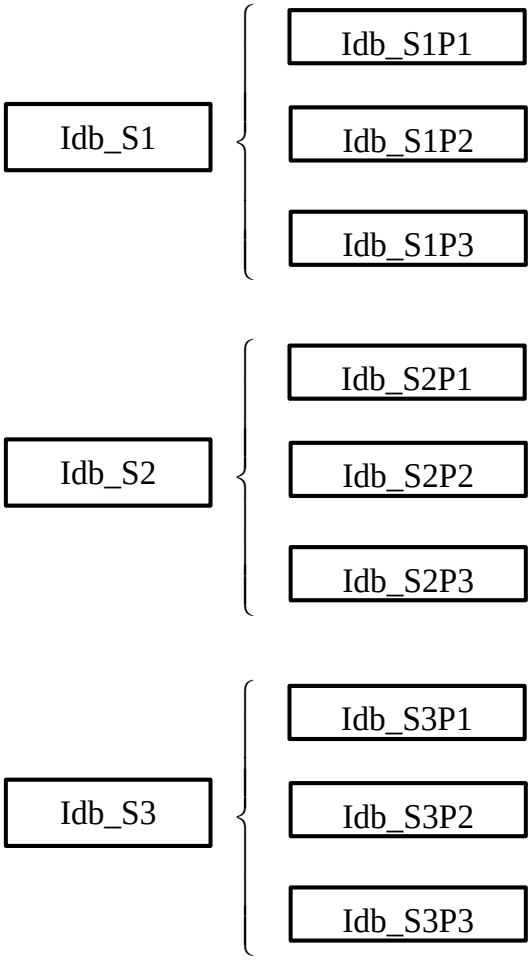


Figura 25 - Detalhamento dos conjuntos de imagens resultantes de diferentes métodos de interpolação utilizados na etapa de padronização.

4.4 Reconhecimento das configurações de mão

A etapa final, reconhecimento das configurações de mão, foi implementada através da aplicação de uma técnica de redução de dimensionalidade no conjunto de dados (mapas de profundidade das mãos segmentadas), seguida da classificação da configuração da mão.

4.4.1 Redução de dimensionalidade dos dados

O objetivo desta etapa é representar cada mapa de profundidade do banco de dados por sua matriz de característica correspondente. Para isso, determinam-se os vetores de projeção através do método 2D^{2}PCA, cujos detalhes estão descritos na Seção 3.5.3.

Na implementação desta técnica, o conjunto de dados foi dividido em duas partes: 6.100 imagens são utilizadas para etapa de treinamento e 6.100 imagens para teste da eficiência do método. Na seleção das imagens de treinamento e teste foi utilizada o método *interleave*. Por exemplo, para as 20 imagens de um gesto/individuo, a primeira vai para o conjunto de treinamento e a segunda para o conjunto de teste e, assim sucessivamente. A técnica 2D^{2}PCA é então aplicada aos dados de treinamento que são usados para construção dos vetores de projeção. Após, projetam-se todas as imagens de cada conjunto nos referidos vetores.

Nessa etapa foram realizados experimentos variando-se a dimensionalidade de 5x5, 6x6, 7x7, 8x8, ..., 20x20 utilizando-se cada um 9 dos conjuntos de dados, totalizando, 144 experimentos como mostra a Figura 25.

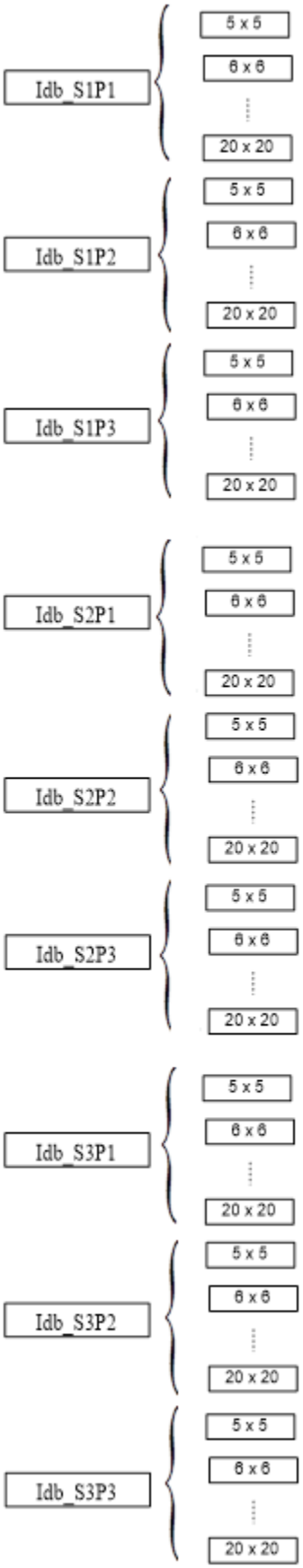


Figura 26 - Detalhamento dos experimentos realizados na etapa de redução de dimensionalidade nos 9 conjuntos de dados.

4.4.2 Etapa de Classificação

O classificador implementado foi o k-vizinhos mais próximos.

Para tal, de posse de todas as matrizes de característica resultantes da etapa de redução de dimensionalidade, calcula-se a distância entre uma matriz de característica da imagem de teste e cada matriz de característica das imagens de treinamento. De acordo com a Seção 3.6 podemos usar toda a família de distâncias Minkowski, como critério de avaliação dos vizinhos mais próximos para classificar as imagens. Entretanto, nesta dissertação foi utilizada somente a distância Euclidiana.

Nessa etapa foram realizados experimentos variando-se o valor de k de 1 até 20 (vide Figura 26), em incrementos de 1, em cada um dos 144 conjuntos de dados resultantes da etapa de redução de dimensionalidade, totalizando 2880 experimentos.

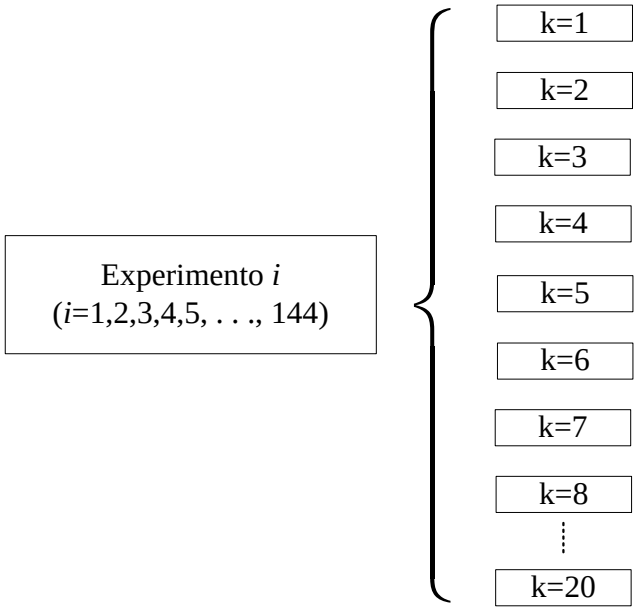


Figura 27 - Detalhamento dos experimentos realizados na etapa de classificação.

CAPÍTULO 5

RESULTADOS E DISCUSSÕES

Inicialmente, são apresentados os resultados das etapas de segmentação e pós-processamento, considerando os diversos métodos de interpolação utilizados. Em seguida são apresentados os resultados da aplicação de redução de dimensionalidade dos dados utilizando $2D^2PCA$, e o resultado da etapa final de reconhecimento implementada através do classificador k-vizinhos mais próximos. Ressalta-se que os resultados de todos os experimentos realizados são apresentados no apêndice desta dissertação. Na última seção é apresentada uma discussão dos resultados obtidos.

5.1 Etapa de segmentação do gesto

As Figuras 27 a 32 apresenta um exemplo das imagens resultantes dos processos de segmentação e pós-processamento utilizando interpolação de ordem zero em ambas etapas.

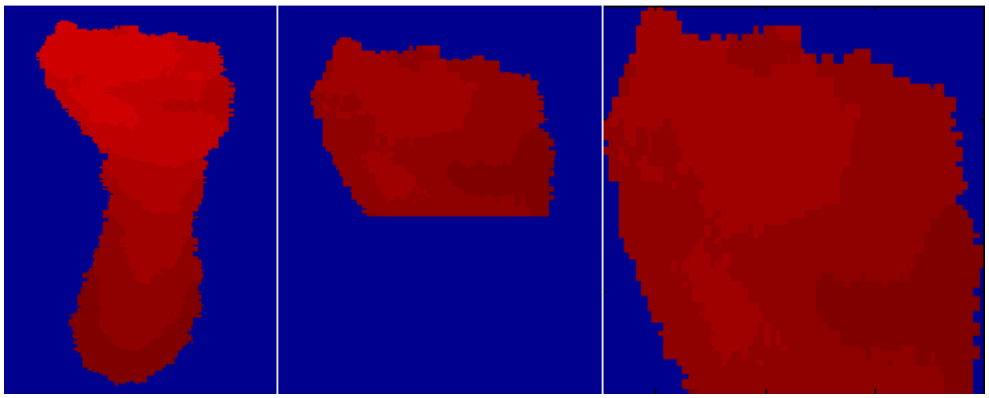


Figura 28 - Exemplo de imagens resultantes dos seguintes processamentos (a) segmentação da mão (b) remoção do antebraço e (c) pós-processamento (padronização) referentes a configuração de mão 1.

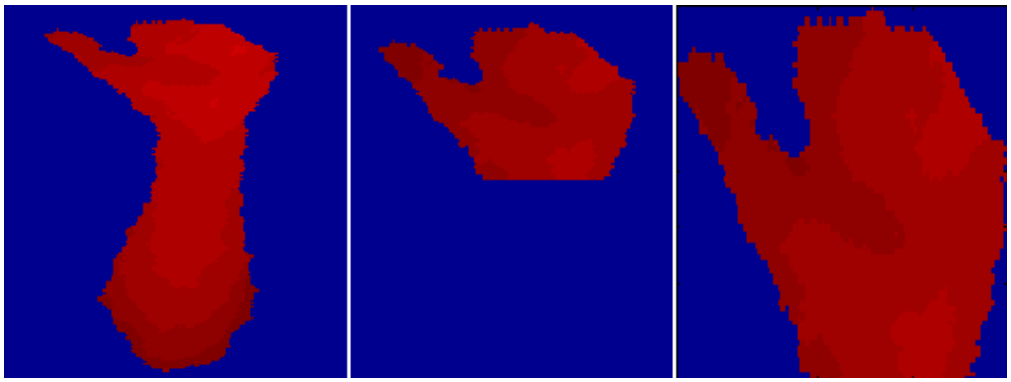


Figura 29 - Exemplo de imagens resultantes dos seguintes processamentos (a) segmentação da mão (b) remoção do antebraço e (c) pós-processamento (padronização) referentes a configuração de mão 2.

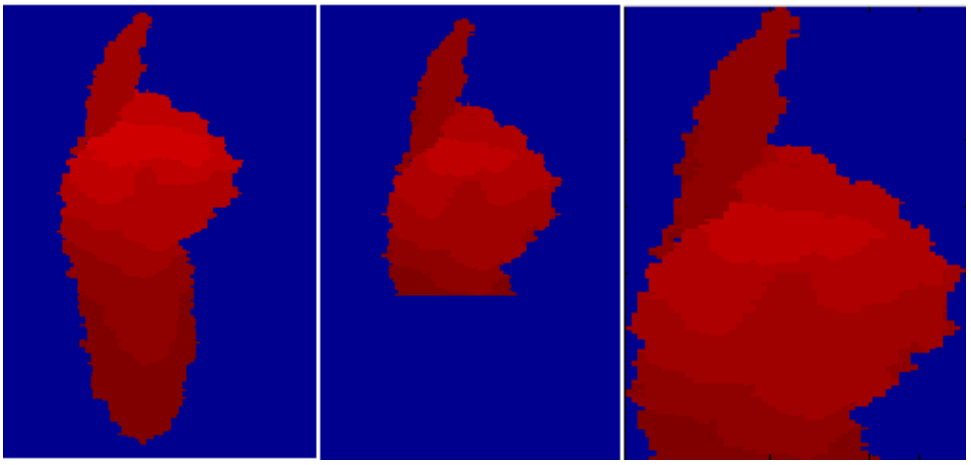


Figura 30 -Exemplos de imagens resultantes dos seguintes processamentos (a) segmentação da mão (b) remoção do antebraço e (c) pós-processamento (padronização) referentes a configuração de mão 14.

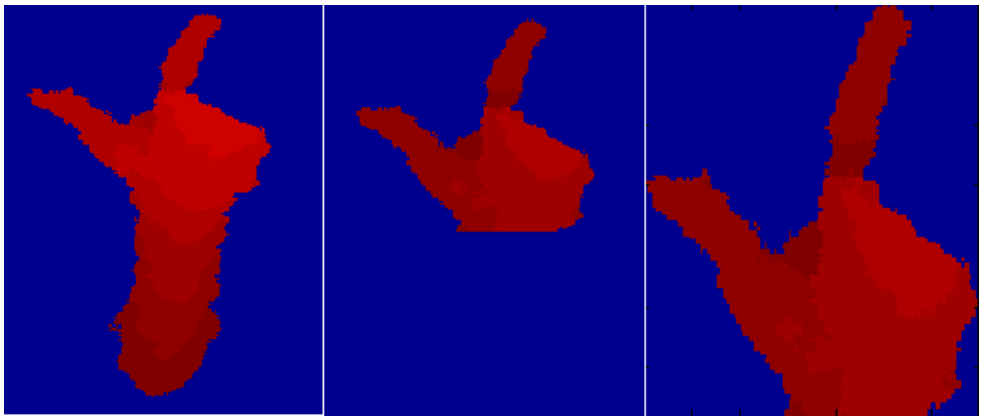


Figura 31 - Exemplo de imagens resultantes dos seguintes processamentos (a) segmentação da mão (b) remoção do antebraço e (c) pós-processamento (padronização) referentes a configuração de mão 38.

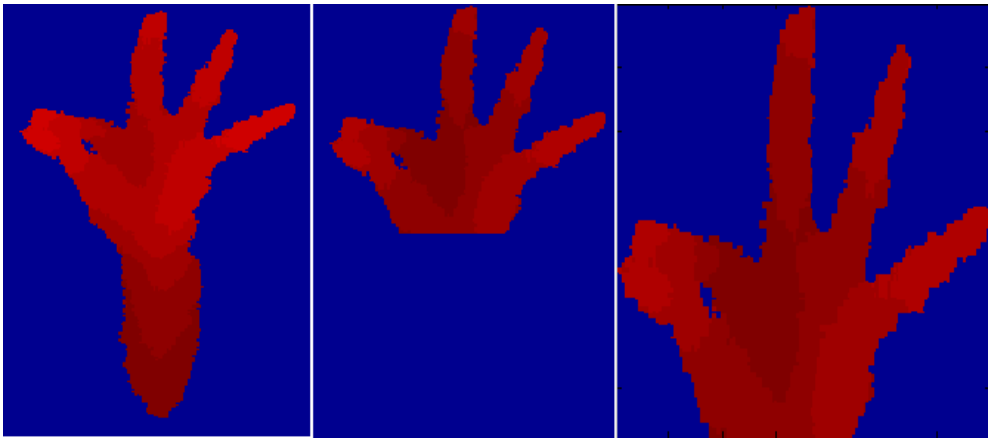


Figura 32 - Exemplo de imagens resultantes dos seguintes processamentos (a) segmentação da mão (b) remoção do antebraço e (c) pós-processamento (padronização) referentes a configuração de mão 44.

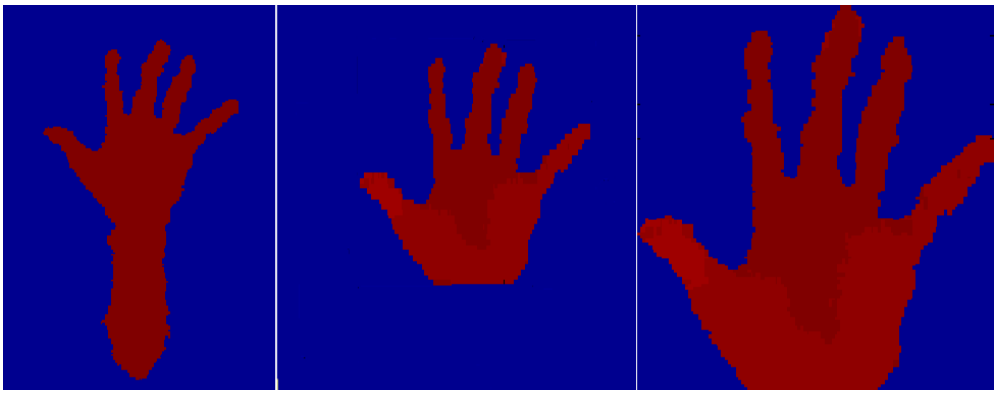
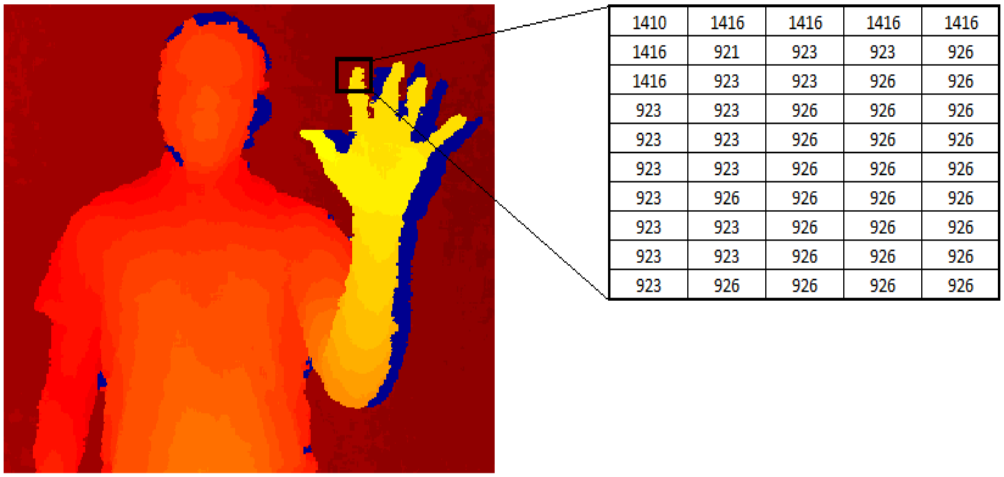


Figura 33 - Exemplo de imagens resultantes dos seguintes processamentos (a) segmentação da mão (b) remoção do antebraço e (c) pós-processamento (padronização) referentes a configuração de mão 61.

A efetividade evidente da técnica de remoção do antebraço implementada, a partir da estimativa da linha de corte, nos exemplos apresentados nas figuras 27 a 32, se estende por todo o conjunto de 12.200 imagens.

5.2 Etapa de pós-processamento da segmentação do gesto

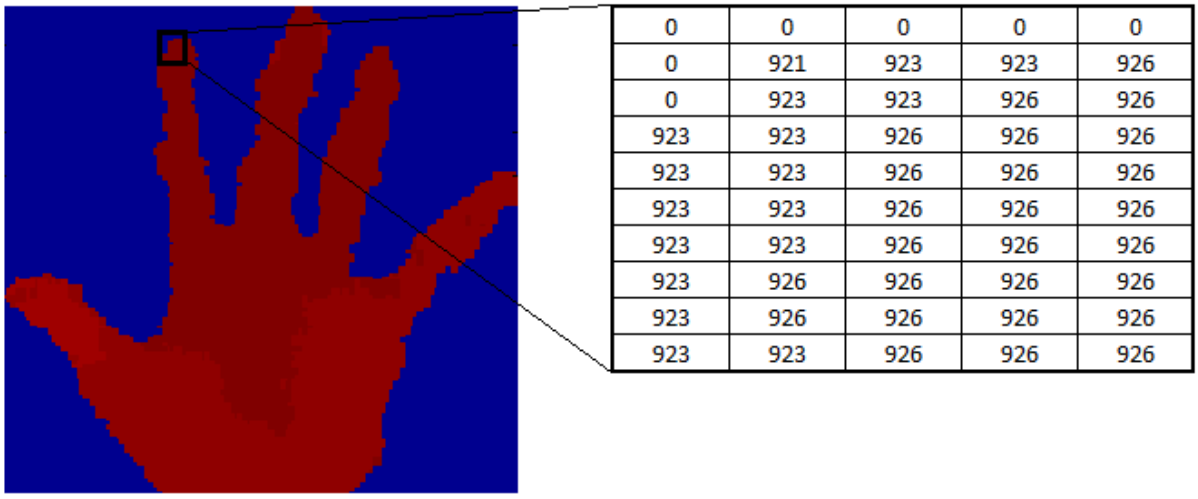
Na Figura 33 é apresentado um exemplo de imagem da mão direita realizando a configuração de mão 61 e uma matriz de pixels (10x5), correspondente a uma região de borda do dedo indicador.



1410	1416	1416	1416	1416
1416	921	923	923	926
1416	923	923	926	926
923	923	926	926	926
923	923	926	926	926
923	923	926	926	926
923	926	926	926	926
923	923	926	926	926
923	923	926	926	926
923	926	926	926	926

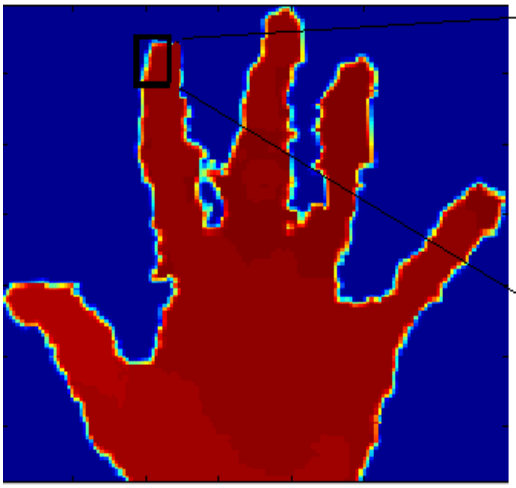
Figura 34 - Exemplo de imagem realizando a configuração de mão 61 e uma matriz de pixels (10x5), correspondente a uma região de borda do dedo indicador.

Nas figuras seguintes, 34 a 36, são apresentadas as imagens resultantes da padronização da referida imagem exemplo, quando se variou o método de interpolação: ordem zero, bilinear e bicúbica, respectivamente.



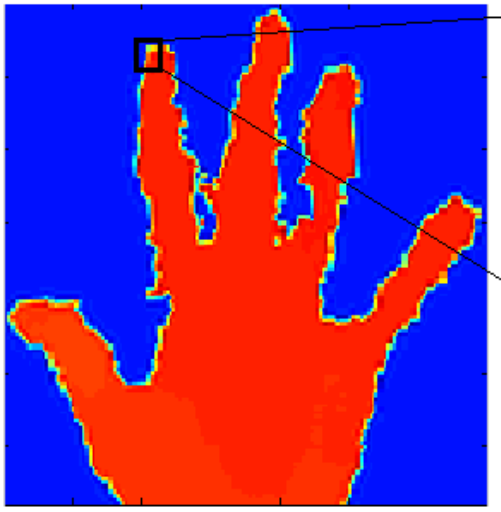
0	0	0	0	0
0	921	923	923	926
0	923	923	926	926
923	923	926	926	926
923	923	926	926	926
923	923	926	926	926
923	923	926	926	926
923	926	926	926	926
923	926	926	926	926
923	923	926	926	926

Figura 35 – Resultado do pós-processamento da imagem mostrada na Figura 33, utilizando interpolação por ordem zero.



0	0	0	0	0
80,66045	229,2022	253,9664	278,3896	303,6149
234,1615	669,9706	735,95	797,7393	861,3835
286,8541	852,7753	892,2809	904,2839	915,2494
486,8199	891,2898	924,3383	925,7093	925,9615
676,1049	903,518	925,1917	925,9512	925,9948
788,9898	923,5176	925,394	925,9951	926
750,5894	923,4755	925,2258	925,9938	926
791,6074	924,1132	925,5154	925,9961	926
858,2328	924,3192	925,5086	925,996	926

Figura 36 - Resultado do pós-processamento da imagem mostrada na Figura 33, utilizando interpolação bilinear.



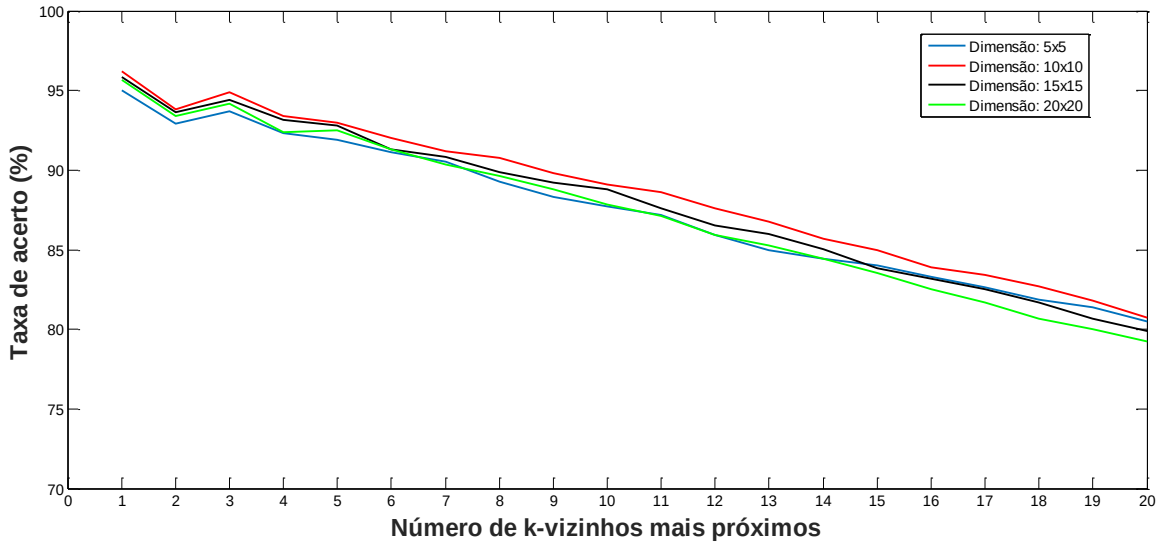
0	0	0	0	0
0,0006	-0,005	-0,0273	0,22178	4,5722
-0,0033	0,032	0,15506	-0,712	-46,653
-0,0073	0,0602	0,33599	-3,1364	-23,603
0,00837	-0,145	-0,3745	-3,5124	503,61
-0,0413	0,7745	0,89443	-116,95	739,03
0,0865	-0,91	-3,7674	45,5213	823,62
0,4897	-5,818	-19,324	501,546	863,78
1,06001	-10,4	-47,6	746,975	933,16
3,21523	-11,72	-115,75	682,193	937,41

Figura 37 - Resultado do pós-processamento da imagem mostrada na Figura 33, utilizando interpolação bicúbica.

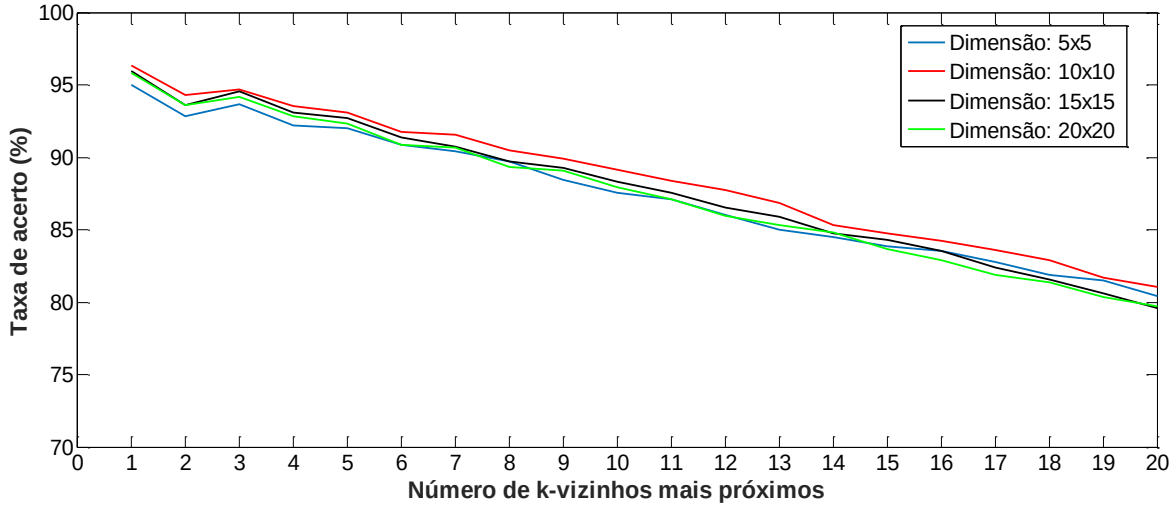
5.3 Etapa de Reconhecimento

Conforme a Figura11, o reconhecimento das configurações de mão compreende as etapas de redução de dimensionalidade e classificação. A primeira, foi implementada através da técnica 2D²PCA. Nessa implementação foram realizados experimentos variando-se a dimensionalidade dos dados de 5x5, 6x6, 7x7, 8x8, ..., 20x20. Já a etapa de classificação, implementada através do classificador k-vizinhos mais próximos, foi avaliada com k variando de 1 a 20, em associação com os experimentos da redução de dimensionalidade referidos.

Nos gráficos seguintes são mostradas as taxas de acertos de reconhecimento das configurações de mão, quando se escolhe determinada técnica de interpolação para cada uma das duas etapas citadas e varia-se as dimensões dos dados através da técnica 2D²PCA. Também é mostrado o impacto na taxa de acertos quando o número de k-vizinhos mais próximos é alterado.

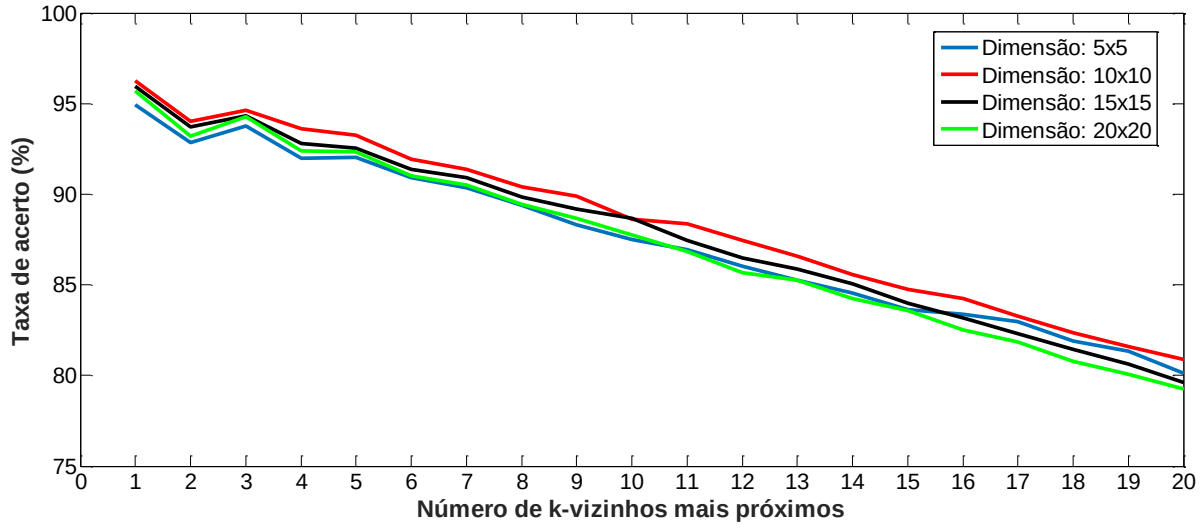


(a)



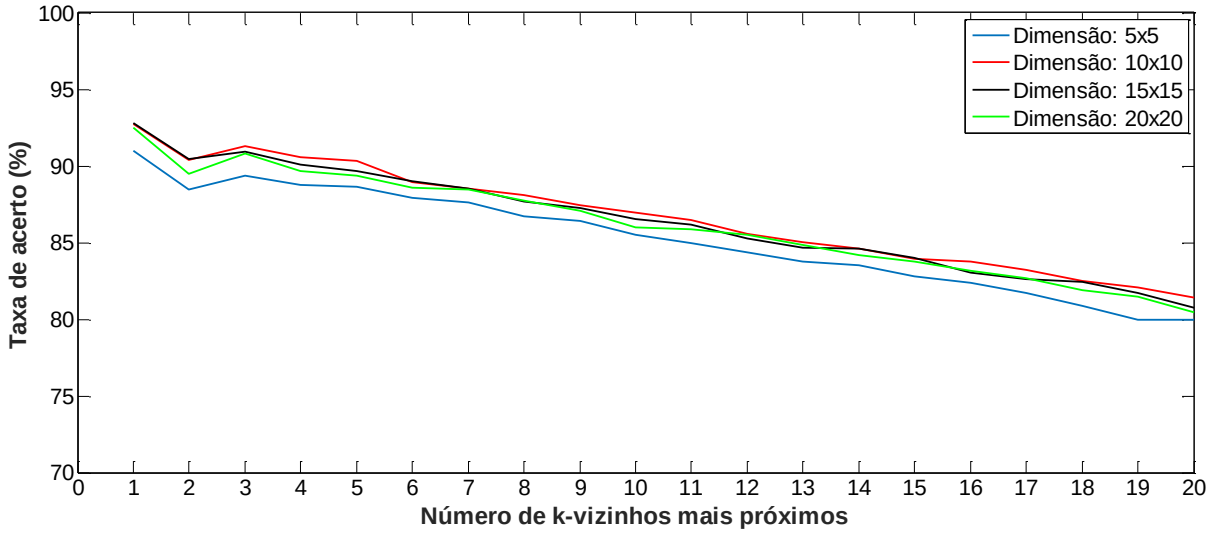
(b)

Figura 37 - Taxa global média de acerto para o conjunto de teste, quando a dimensão dos dados foi reduzida para 5x5, 10x10, 15x15 e 20x20 e o número de vizinhos do classificador variou de 1 a 20 em incrementos de 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (a) interpolação de ordem zero (vizinho mais próximo) para etapa de segmentação e para etapa de padronização das imagens; (b) Interpolação por vizinho mais próximo para etapa de segmentação e interpolação bilinear para etapa de padronização das imagens.



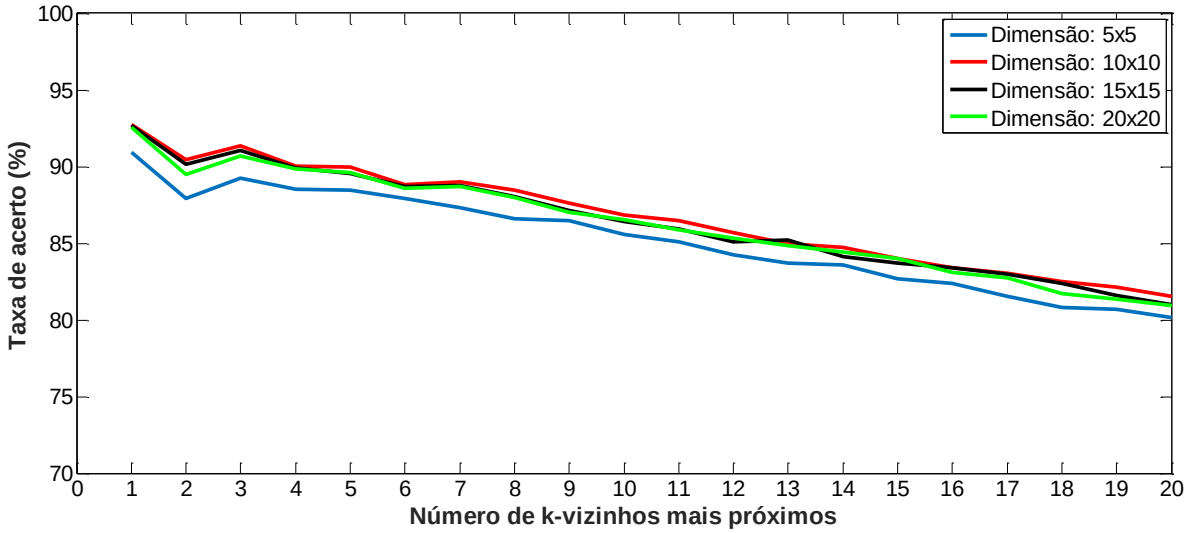
(c)

Figura 38(continuação) - Taxa global média de acerto para o conjunto de teste, quando a dimensão dos dados foi reduzida para 5x5, 10x10, 15x15 e 20x20 e o número de vizinhos do classificador variou de 1 a 20 em incrementos de 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (c) Interpolação por vizinho mais próximo para etapa de segmentação e interpolação bicúbica para etapa de padronização das imagens.

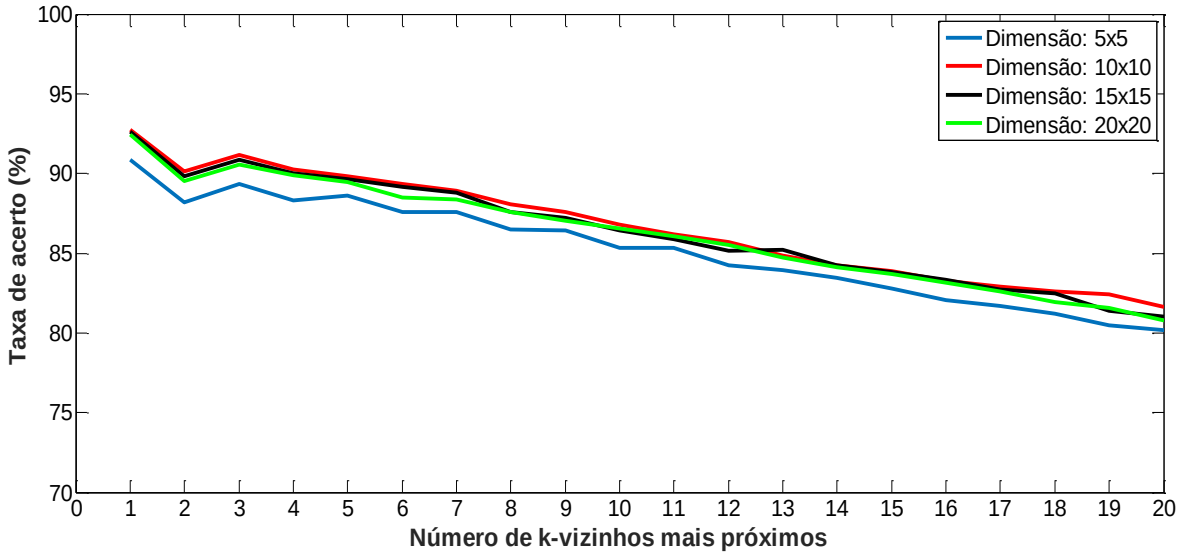


(a)

Figura 38 - Taxa global média de acerto para o conjunto de teste, quando a dimensão dos dados foi reduzida para 5x5, 10x10, 15x15 e 20x20 e o número de vizinhos do classificador variou de 1 a 20 em incrementos de 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (a) Interpolação bilinear para etapa de segmentação e interpolação por vizinho mais próximo para etapa de padronização das imagens.

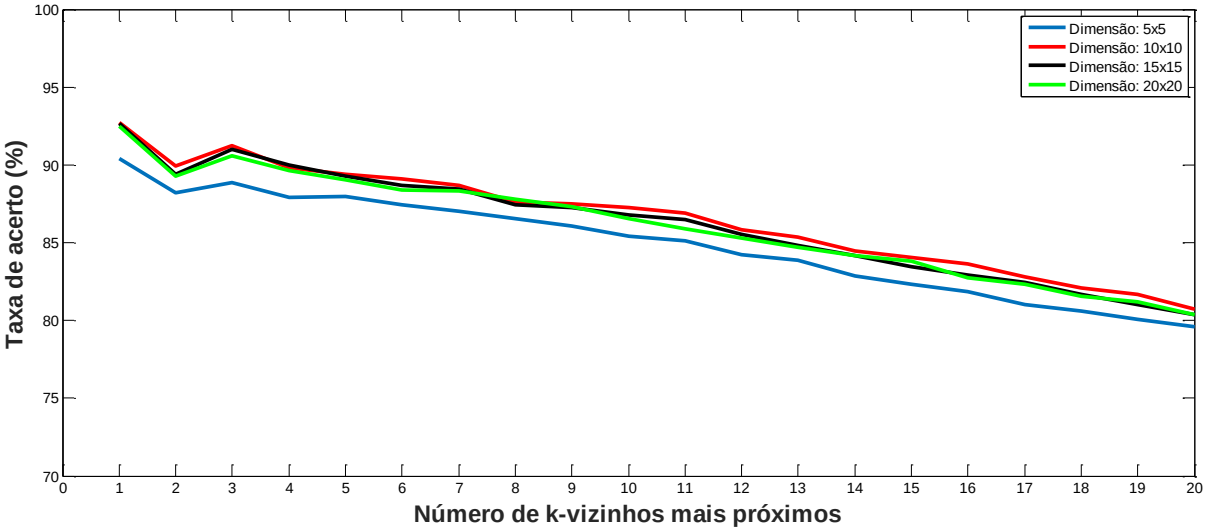


(b)

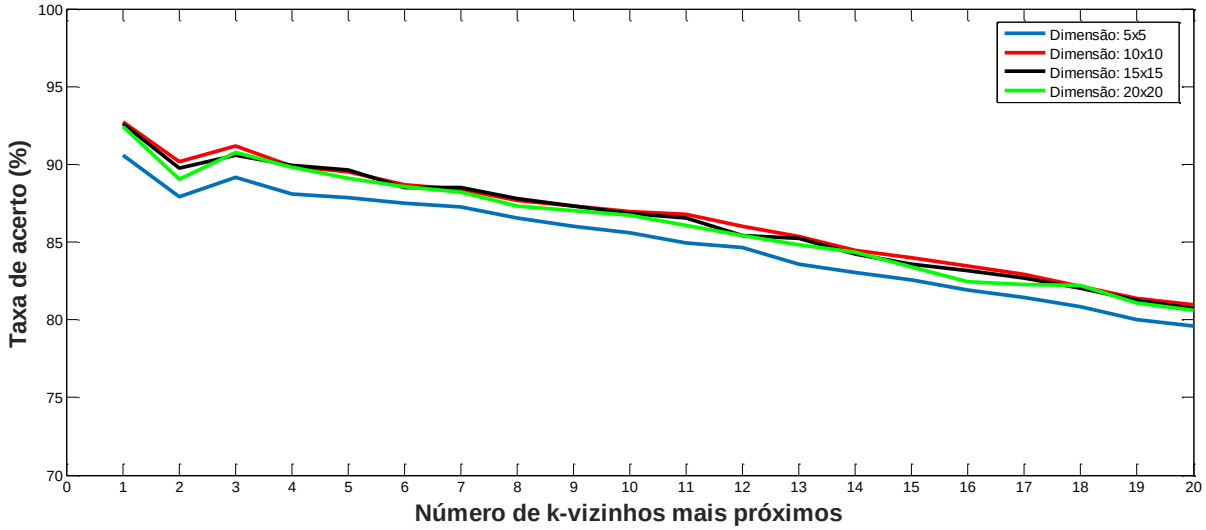


(c)

Figura 39 (continuação) - Taxa global média de acerto para o conjunto de teste, quando a dimensão dos dados foi reduzida para 5x5, 10x10, 15x15 e 20x20 e o número de vizinhos do classificador variou de 1 a 20 em incrementos de 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (b) interpolação bilinear para etapa de segmentação e para etapa de padronização das imagens; (c) Interpolação bilinear para etapa de segmentação e interpolação bicúbica para etapa de padronização das imagens.

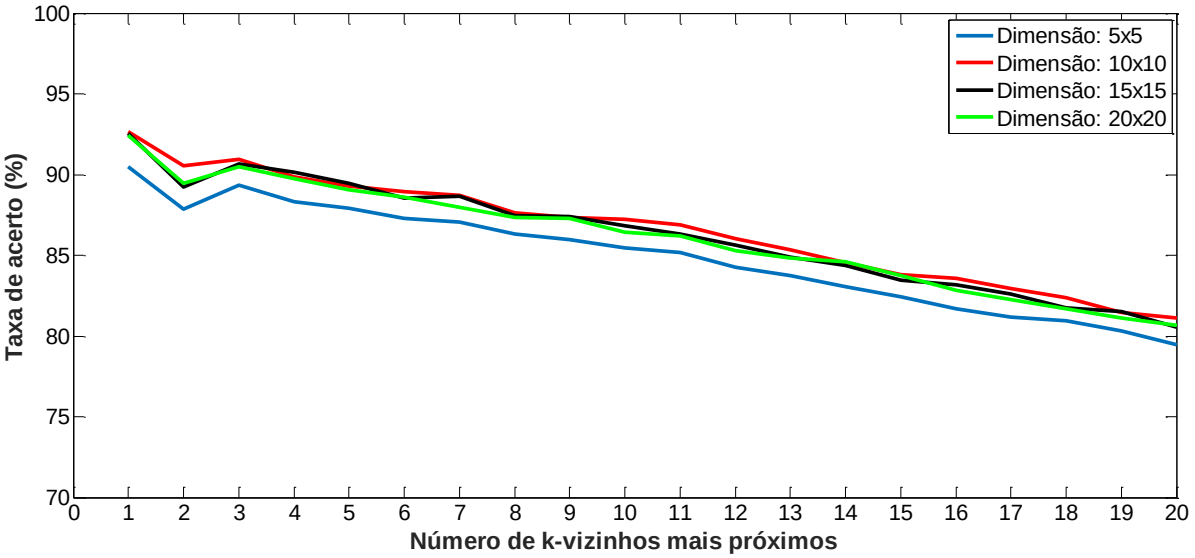


(a)



(b)

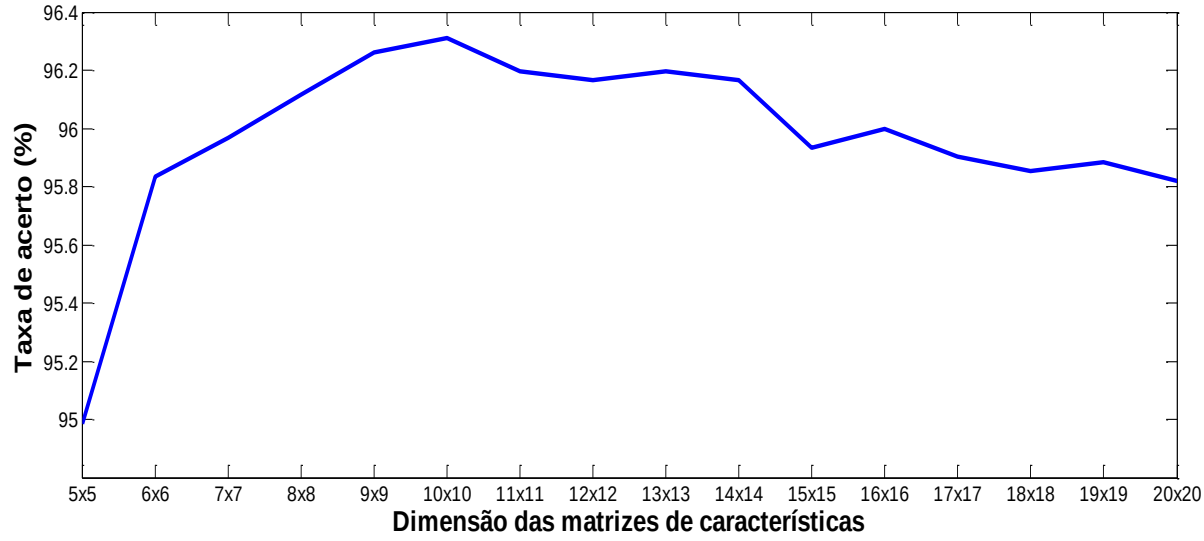
Figura 39 - Taxa global média de acerto para o conjunto de teste, quando a dimensão dos dados foi reduzida para 5x5, 10x10, 15x15 e 20x20 e o número de vizinhos do classificador variou de 1 a 20 em incrementos de 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (a) interpolação bicúbica para etapa de segmentação e interpolação por vizinho mais próximo para etapa de padronização das imagens; (b) Interpolação bicúbica para etapa de segmentação e interpolação bilinear para etapa de padronização das imagens.



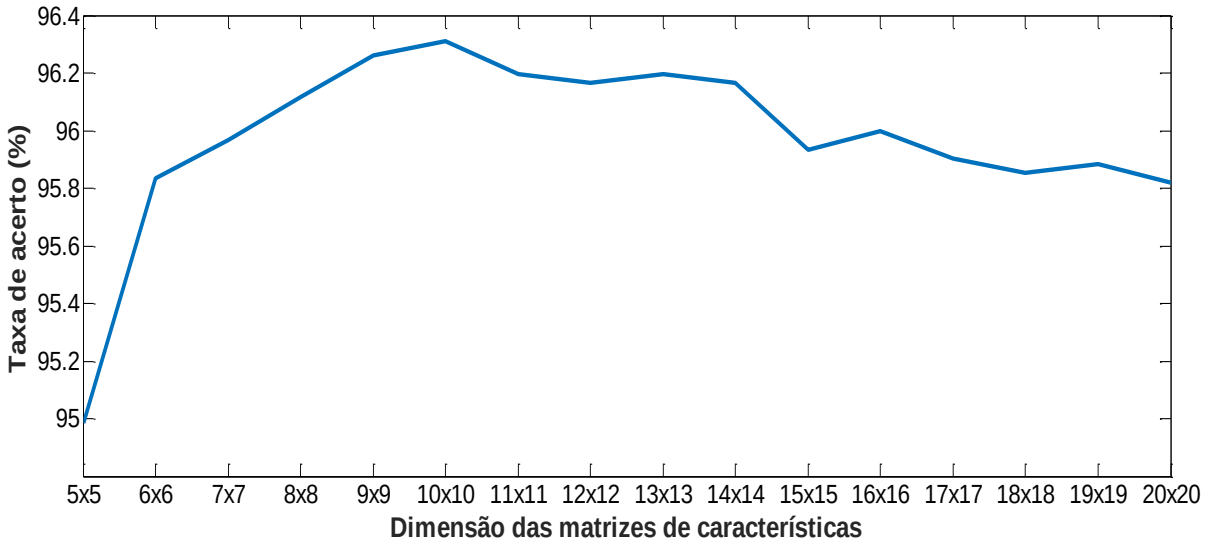
(c)

Figura 40 (continuação) - Taxa global média de acerto para o conjunto de teste, quando a dimensão dos dados foi reduzida para 5x5, 10x10, 15x15 e 20x20 e o número de vizinhos do classificador variou de 1 a 20 em incrementos de 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (c) interpolação bicúbica para etapa de segmentação e interpolação por vizinho mais próximo para etapa de padronização das imagens.

Nos gráficos seguintes é apresentado o comportamento da taxa global média de acerto em função da redução de dimensionalidade dos dados quando o número de vizinhos do classificador é fixado em k=1. Ressalta-se que a redução de dimensionalidade resultou em matrizes de características com número de linhas e colunas iguais.

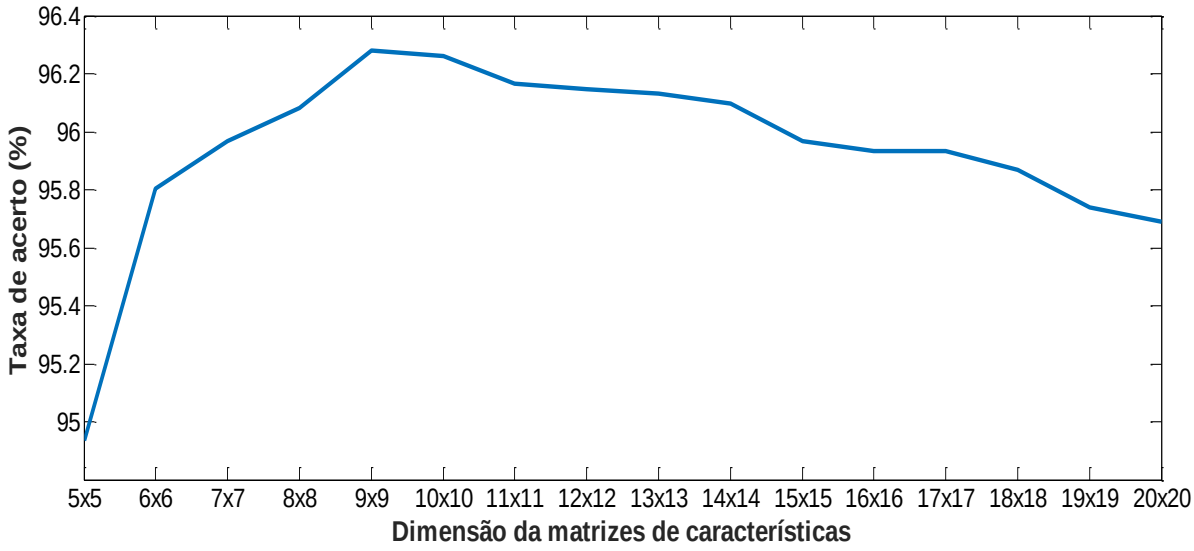


(a)



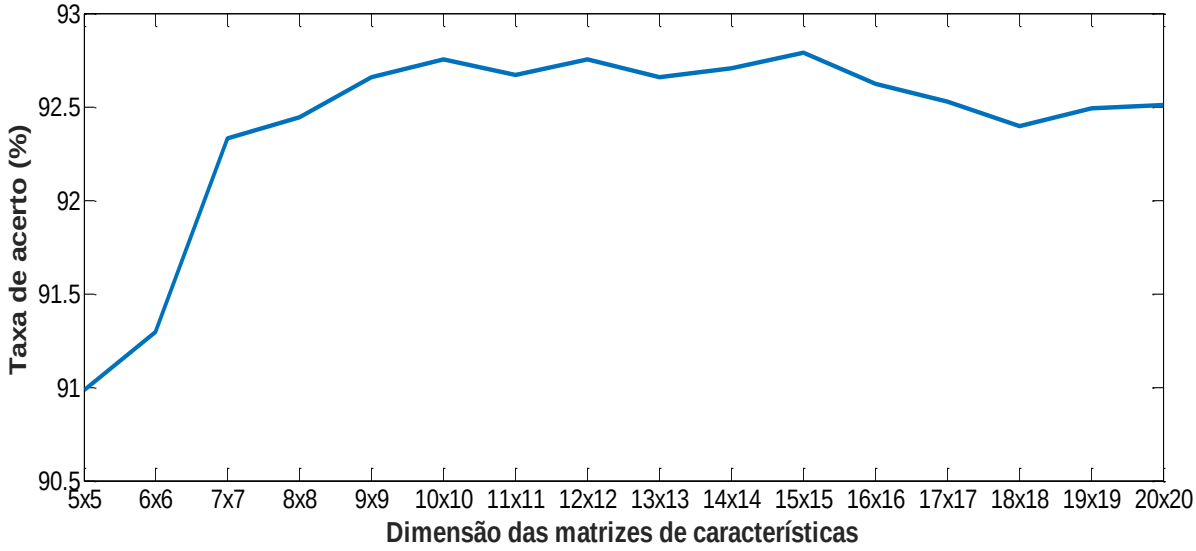
(b)

Figura40 - Taxa global média de acerto para o conjunto de teste em função da redução da dimensão dos dados quando o número de vizinhos do classificador foi fixado em 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (a) Interpolação por vizinho mais próximo para as etapas de segmentação e para padronização das imagens. (b) Interpolação por vizinho mais próximo para etapa de segmentação e interpolação bilinear para padronização das imagens.



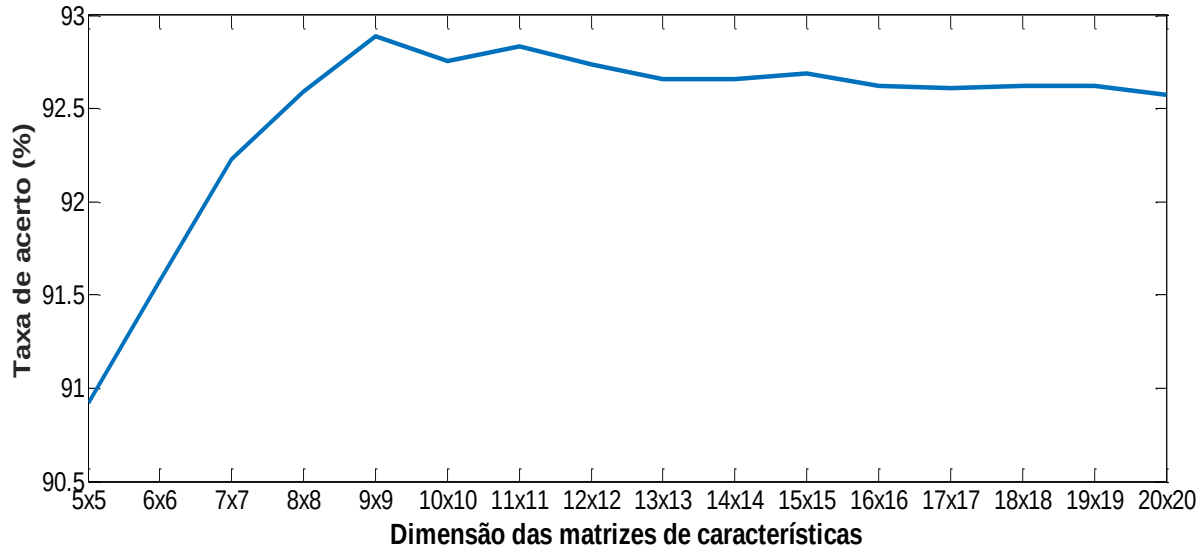
(c)

Figura 41 (continuação) - Taxa global média de acerto para o conjunto de teste em função da redução da dimensão dos dados quando o número de vizinhos do classificador foi fixado em 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (c) Interpolação por vizinho mais próximo para as etapas de segmentação e interpolação bicúbica para padronização das imagens.

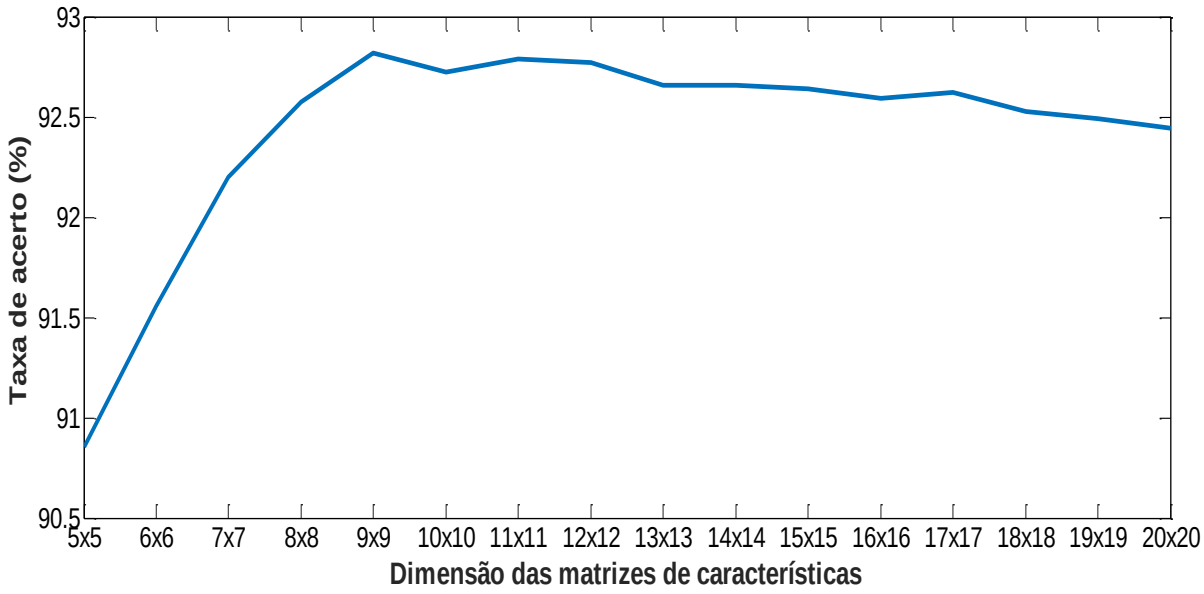


(a)

Figura 41 - Taxa global média de acerto para o conjunto de teste em função da redução da dimensão dos dados quando o número de vizinhos do classificador foi fixado em 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (a) Interpolação bilinear para etapa de segmentação e interpolação por vizinho mais próximo para padronização das imagens.

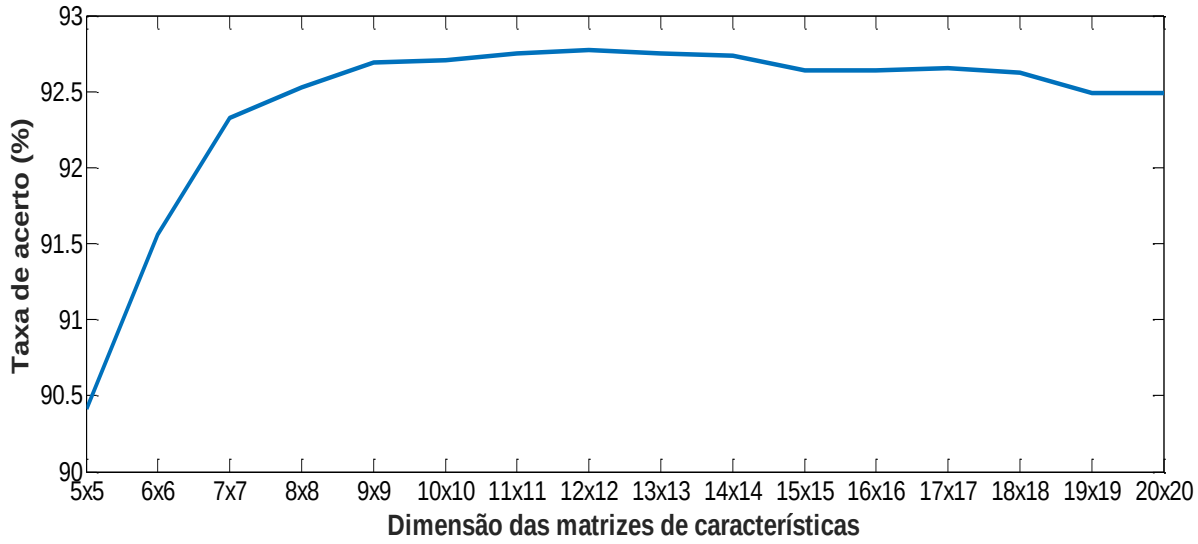


(b)

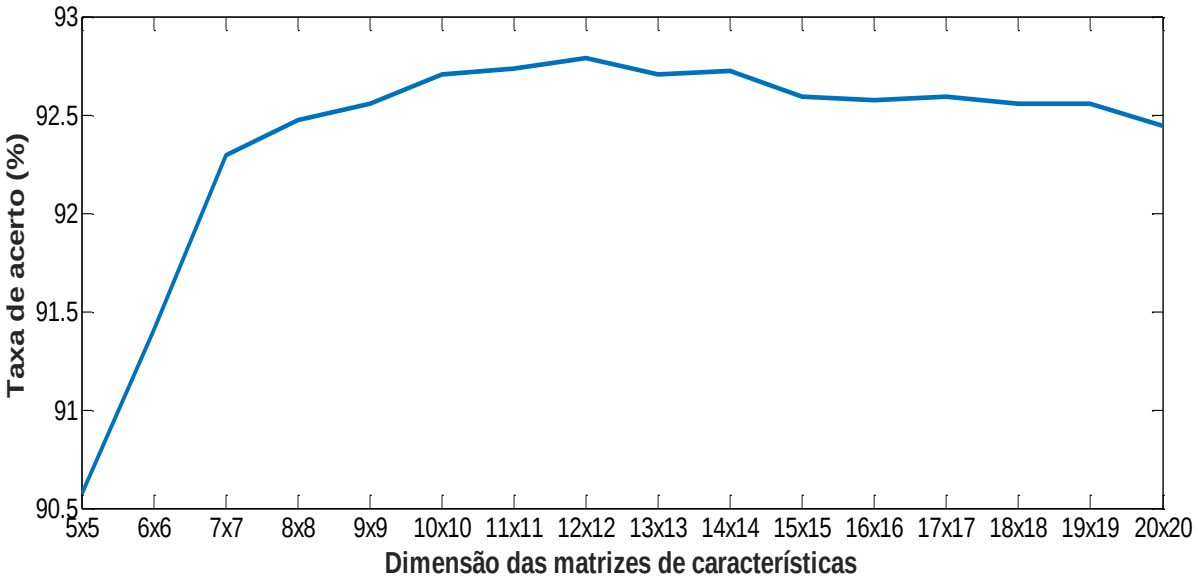


(c)

Figura 42(continuação) - Taxa global média de acerto para o conjunto de teste em função da redução da dimensão dos dados quando o número de vizinhos do classificador foi fixado em 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (b) Interpolação bilinear para as etapas de segmentação e para padronização das imagens. (c) Interpolação bilinear para etapa de segmentação e interpolação bicúbica para padronização das imagens.

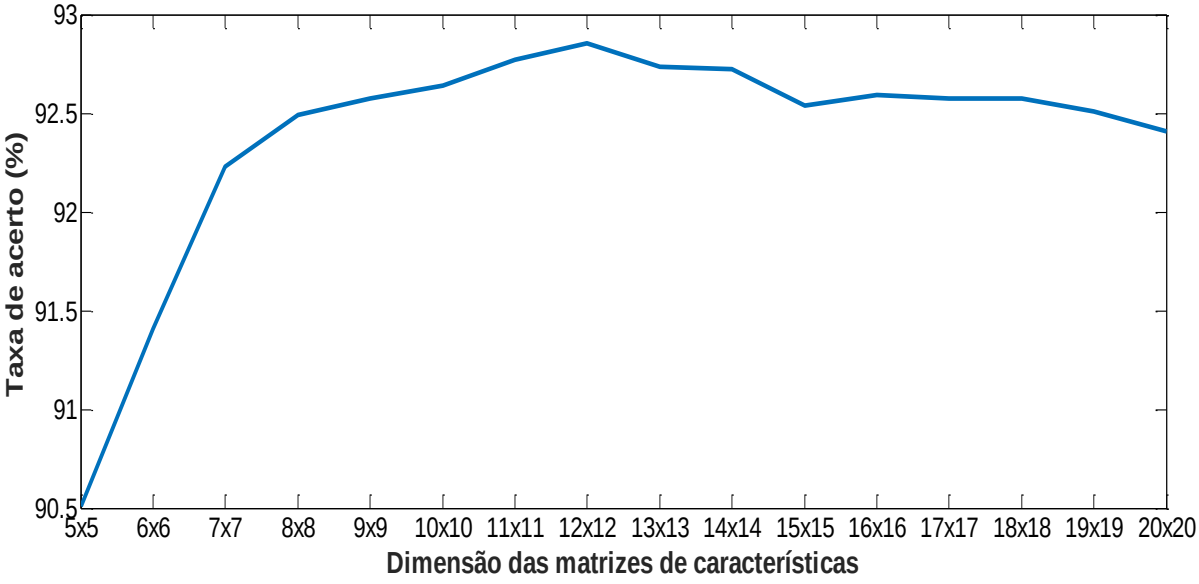


(a)



(b)

Figura 42 - Taxa global média de acerto para o conjunto de teste em função da redução da dimensão dos dados quando o número de vizinhos do classificador foi fixado em 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (a) Interpolação bicúbica para etapa de segmentação e interpolação por vizinho mais próximo para padronização das imagens. (b) Interpolação bicúbica para etapa de segmentação e interpolação bilinear para padronização das imagens.



(c)

Figura 43 (continuação) - Taxa global média de acerto para o conjunto de teste em função da redução da dimensão dos dados quando o número de vizinhos do classificador foi fixado em 1, em função do tipo de interpolação utilizado nas etapas de segmentação e padronização das imagens. (c) Interpolação bicúbica para as etapas de segmentação e para padronização das imagens.

Na Figura 43 apresenta-se a matriz de confusão que mostra as classes reais e as classes atribuídas pelo classificador k-vizinhos mais próximos de cada uma das 61 classes (configurações de mão) para o conjunto de 6100 amostras de teste, utilizando dados de dimensão 10x10 e o classificador com k=1. O número de acertos, para cada classe, se localiza na diagonal principal, M(Ci,Cj) para i=j, da referida matriz e os demais elementos M(Ci,Cj), para i≠j, representam erros na classificação.

Resultados detalhados dos demais experimentos realizados são apresentados no Apêndice.

Classe Real	Classe atribuída pelo classificador																																																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
1	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	91	8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
5	0	0	0	0	98	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	97	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	3	0	0	0	0	1	93	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	1	1	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	1	0	0	0	0	0	0	94	2	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	4	0	0	0	0	0	1	0	0	91	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	3	0	0	0	0	0	1	0	0	2	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	91	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	7	90	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	1	5	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	95	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	96	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	6	1	0	0	2	0	87	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	98	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	94	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	9	89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2	0	95	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	97	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	97	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	94	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	2
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	96	0	0	0	0	0	0	0	0	1	0	0	0	0	0
48	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	1	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	1	1	0	0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	97	0	0	0	0	0	0	0	0	0	0	0
51	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	91	6	1	0	1	0	0	0	0	0	0
52	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95	2	0	0	2	0	0	0	0	0
53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	97	0	0	1	0	0	0	0	0
54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	2	0	0	0	0	0	0
55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	1	0	0	0	0	0
56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	93	0	0	0	0	0
57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	0	0	0	3
58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	2	0
59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	3	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	89	1
61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	98

Figura 44 - Matriz de confusão para o conjunto de teste quando os dados foram reduzidos para a dimensão de 10x10 através da técnica 2D²PCA em associação com o classificador k-vizinhos mais próximos com k=1.

5.4 Tempo de processamento

O Quadro 4 mostra o tempo de processamento em cada uma das etapas necessárias ao reconhecimento da CM, quando a infraestrutura de *hardware* descrita no quadro 3 é usada e o método de reconhecimento foi configurado da seguinte forma: interpolação por vizinho mais próximo na etapa de segmentação e interpolação bilinear para etapa de padronização das imagens e uma redução de dimensionalidade de 10x10, associada ao uso do classificador k-vizinhos mais próximos com k igual a 1. Os tempos apresentados, são valores médios, e estão relacionados ao processamento de uma única imagem.

Quadro 4 – Tempo de processamento equivalente a um único experimento.

Etapa	Tempo de processamento médio (segundos)
Segmentação + pós-processamento	0,324
Redução de dimensionalidade: 2D²PCA	0,086
Classificação: KNN	0,022

5.5 Análise e discussão dos resultados

Nos gráficos representados pelas figuras 37 a 39 é feita uma comparação dos resultados quando os dados têm suas dimensões reduzidas para 5x5, 10x10, 15x15 e 20x20, respectivamente, utilizando a técnica 2D²PCA. Podemos perceber que independentemente do tipo de interpolação utilizada nas etapas de segmentação e padronização das imagens, ao aumentarmos o número de k-vizinhos a taxa de acerto global do sistema de reconhecimento cai drasticamente.

A partir desta análise é fixado o valor de k=1. Então foram construídos os gráficos apresentados nas figuras 40 a 42 com o intuito de se evidenciar a melhor técnica de interpolação para as etapas de segmentação e padronização das imagens e também concluir sobre a dimensão da matriz de característica que melhor represente os dados originais.

Em cada um destes 9 gráficos o eixo das abscissas corresponde a dimensão da matriz de característica onde adotou-se o valor inicial de 5x5 e no eixo y encontram-se as taxas de acerto correspondentes. Percebe-se em todos os gráficos um crescimento na taxa de acerto conforme a dimensão das matrizes de características é aumentada, alcançando um ponto máximo e posteriormente decrescendo.

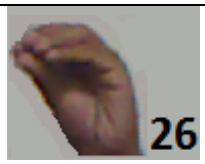
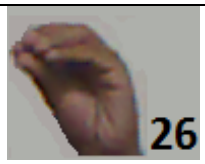
A interpolação por vizinho mais próximo mesmo não exigindo qualquer cálculo e por simplesmente repetir o valor do *pixel* mais próximo mostrou-se a melhor técnica de interpolação na etapa de segmentação das configurações de mão e a interpolação bilinear mostrou-se a melhor técnica na etapa de padronização das imagens. O desempenho destas técnicas pode ser observado na Figura 40(b) onde a taxa de acerto atinge um máximo de 96,31% quando a dimensão da matriz de características é 10x10. Quando utiliza-se a interpolação bicúbica nas etapas de segmentação e pós-processamento o desempenho do sistema de reconhecimento é de 92,57% como pode ser observado na Figura 42(c).

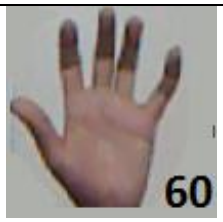
Com base nessas informações construiu-se a matriz de confusão mostrada na Figura 43, onde constata-se elevadas taxas de acertos, a menor taxa de acerto obtida foi de 87% para a configuração de mão 23. Obteve-se ótimas taxas de acerto (acima de 99%) para as configurações de mão 2, 4, 19, 31, 38, 39 e 55. Uma taxa de acerto de 100% foi alcançada nas configurações de mão 15, 20, 29, 40, 44 e 45.

Na Tabela 3 são mostrados os principais erros de classificação do método proposto e o percentual de erro, por configuração de mão (CM), no conjunto de 100 imagens de teste/CM.

Tabela 3 - Comparação entre classe esperada e classe atribuída pelo classificador

Classe esperada	Classe atribuída	Erro de classificação
		8%
		3%
		4%
		3%
		9%
		7%
		5%
		3%

Classe esperada	Classe atribuída	Erro de classificação
		6%
		4%
		9%
		4%
		3%
		6%
		7%
		3%

Classe esperada	Classe atribuída	Erro de classificação
		9%

Das 6100 amostras de teste apenas 3,68% foram classificadas erroneamente. Uma justificativa para ter ocorrido isso é o fato de o Kinect® ainda possuir limitações na detecção precisa dos dedos das mãos.

Outro fator que contribuiu para causar equívocos na classificação é ilustrado na Figura 44. A referida figura apresenta exemplos da configuração de mão 51 do banco de dados construído. Como foi mencionado na Seção 4.1.2 foi solicitado aos voluntários que movessem a mão no momento da execução das configurações, ficando os mesmos livres para realização de movimentos quando da realização das configurações. Dessa forma, em certas posições, o gesto é feito lateralmente em relação ao sensor de profundidade. No caso da CM51, os dedos que nesta configuração são separados, aparecem parcialmente justapostos causando confusão com a CM 52.

A técnica de segmentação por crescimento de regiões utilizada foi eficaz, pois toda a região da mão direita foi agrupada, não havendo outras regiões do corpo conectadas a mão. Nesta etapa utilizou-se interpolação de ordem zero que conservou os valores de pixels de profundidade capturados pelo kinect® e isso contribuiu de forma decisiva na etapa de reconhecimento.

(a) (b)

(c) (d) (e)

Figura 45 – Ilustração de erro de classificação entre as CM 51 e CM52 (a) configuração de mão 51 (b) configuração de mão 52 (c),(d) e (e) exemplos de execução da CM51

CAPÍTULO 6

CONCLUSÃO

Na revisão da literatura constatou-se que muitas pesquisas foram realizadas com a intenção de auxiliar a comunicação de pessoas surdas, sendo a maioria dos trabalhos restrita ao endereçamento de reconhecimento do alfabeto manual, o qual constitui-se, conforme já mencionado, um empréstimo linguístico e não em língua de sinais propriamente dita.

A análise dos trabalhos publicados evidenciou que a maioria das técnicas elaboradas possui em seu seio a construção de ambientes artificiais para validar suas teorias, utilizando iluminação controlada e fundo das imagens uniforme (na cor preta ou branca) e, em alguns casos, fazendo uso de luvas coloridas para uma melhor diferenciação dos sinais, condições estas ainda distantes dos ambientes não controlados vivenciados no dia a dia dos surdos. A única exceção é o trabalho de Porfírio (2013) que faz uso de imagens de profundidade.

Esta dissertação propôs o desenvolvimento de um método robusto de reconhecimento capaz de distinguir configurações de mão (CM), um dos principais fonemas da LIBRA, o qual está presente em todos os sinais da referida língua.

Nesta dissertação foram utilizadas imagens de profundidade e, portanto, a construção de ambientes artificiais facilitadores do processamento da imagem, sobretudo da etapa de segmentação da configuração da mão na cena, não foi necessária. Prova é que a segmentação da mão foi implementada com técnicas já estabelecidas, como o algoritmo de crescimento de regiões.

Com a finalidade de contribuir para o avanço da pesquisa na área, foi construído, em conjunto com (SANTOS, 2015) um banco de imagens robusto e representativo das 61 configurações de mãos de LIBRAS, composto de 12.200 imagens.

As imagens passaram por uma etapa de segmentação e pós-processamento. Foi desenvolvido um algoritmo de redução de dimensionalidade de imagens, através da técnica 2D²PCA, resultando nas formas mais representativas de dados, a partir das combinações lineares das variáveis originais e, finalmente, esses dados foram submetidos ao classificador k-vizinhos mais próximos para o reconhecimento das CM. O referido classificador demonstrou ser eficaz no reconhecimento das configurações de mão, pois seu uso redundou numa taxa global de acerto média de 96,31%. Essa alta taxa de acerto é respaldada pelo fato da mesma ter sido obtida em um conjunto de imagens robusto e representativo das condições do cotidiano, uma vez que as imagens foram capturadas, sem restrição de iluminação, em posições (diferentes distâncias em relação ao corpo) e ângulos de rotação diferentes (na faixa de 45º e 135º). Soma-se a isso a quantidade de imagens/configuração muito superior as utilizadas por outros autores (200 imagens/CM, totalizando 12200 imagens).

Diferentemente, do que se viu na literatura, as altas taxas de acerto do reconhecimento das CM da LIBRAS foram obtidas com técnicas de fácil implementação e baixo custo computacional, enquanto que os resultados apresentados na literatura foram, em maioria, obtidos com técnicas mais complexas, como redes neurais. Tendo em vista que um sistema de reconhecimento de LIBRAS deverá endereçar a identificação dos outros fonemas, ponto de articulação, orientação, movimento e expressão facial, é relevante que a implementação de cada uma dessas tarefas de reconhecimento apresente boas características com respeito a complexidade e ao custo computacional, para que o sistema como um todo seja exequível.

Apesar dessa dissertação ter cumprido os objetivos propostos no início de seu desenvolvimento, melhorias podem ser incorporadas para obtenção de uma taxa maior de acertos. Dentro desse contexto, sugere-se a avaliação de outras propostas de classificação, como as redes convolucionais.

A continuidade da pesquisa deve também ser dirigida para o reconhecimento dos outros parâmetros fonológicos: ponto de articulação, orientação, movimento e expressão facial.

REFERÊNCIAS

ADITHYA, V.; VINOD, P.R.; GOPALAKRISHNAN, U., **Artificial neural network based method for Indian sign language recognition**, Information & Communication Technologies (ICT), 2013 IEEE Conference on , pp.1080,1085, 11-12 April 2013

AKYAMA, Márcio Teruo; **Interpolação de imagens baseada em clustering;** Paraná, 2010. Dissertação de mestrado-Programa de Pós-Graduação em engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná, Paraná, 2010.

ALL-JARRAH, O.; HALAWANI, A. **Recognition of gestures in Arabic Sign Language using neuro-fuzzy systems**, Artificial Intelligence 133, pp. 117-138, 2001.

ANJO, Mauro dos Santos; **Avaliação das técnicas de segmentação, modelagem e classificação para o reconhecimento automático de gestos e proposta de uma solução para classificar gestos da LIBRAS em tempo real.** São Carlos, 2012. Dissertação de mestrado-Centro de ciências exatas e de tecnologia da Universidade Federal de São Carlos, São Carlos, 2012.

ANTON, H., RORRES, C. **Álgebra Linear com aplicações**, Bookman, Porto Alegra, 2004.

BINH, N., SHUICHI, E., EJIMA, T., **Real-Time Hand Tracking and Gesture Recognition System**, Intelligence Media Laboratory, Kyushu Institute of Technology, 2005.

BOMFIM, D., A; **Apostila para o curso de Formação em LIBRAS, Educação e Surdez.** Minas Gerais, 2012.

CAPOVILLA, F. C.; RAPHAEL, W. D. **Dicionário Enciclopédico Ilustrado Trilíngue da Língua de Sinais Brasileira**. São Paulo, Brasil: Editora da Universidade de São Paulo, 2001.

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**. In: Interaction 09 - South America, 2009, São Paulo. Interaction 09 - South America, 2009. p. 190 - 195.

CRUZ, L; LUCIO, D; VELHO, L. **Kinect and rgbd images: challenges and applications.** páginas 36 – 49, Los Alamitos, CA, USA, 2012.

GONZALEZ, R. C.; WOODS, R. E. 3rd. Edition. **Digital Image Processing**, Prentice Hall, 2000. 295p.

GONZALEZ, R.C. WOODS R.E; EDDINS S.L**. Digital Image Processing Using MATLAB**. Upper Saddle River, Prentice Hall, 2009. 578p.

GROBEL, K.; HIENZ, H., **Video-based handshape recognition using a handshape structure model in real time,** Pattern Recognition, 1996., Proceedings of the 13th International Conference on , vol.3, pp.446,450 vol.3, 25-29 Aug 1996.

JOHNSON, R. A; WICHERN, D. W. **Applied Multivariate Statistical Analysis.** Texas, USA, 2002.

LANDIM, P. M. B. Introdução à Confecção de Mapas pelo Software Surfer. **Texto didático 08. Laboratório de geomatemática**, Rio Claro: UNESP, p. 22, 2002.

MARCOTTI, P, ABIUZI, L. B., RIZOL P. M. S. R., e FORSTER, C. H. Q. **Interface para reconhecimento da língua brasileira de sinais**. XVIII Simpósio Brasileiro de Informática na Educação - SBIE - Mackenzie, 2007.

MEHDI, S. A.; KHAN, Y. N. **Sign Language Recognition Using Sensor Gloves**, In Proc. of the 9th International Conference on Neural Information Processing, Vol. 5, pp. 2204–2206, Singapore, 2002 .

MICROSOFT. **Xbox 360 + Kinect**, nov 2010. Disponível em: http://www.xbox.com/pt-BR/Kinect/Home-new. Acesso em: jan. 2015.

MICROSOFT. **Kinect for Windows**, fev 2012. Disponível em: http://www.microsoft.com/en-us/Kinectforwindows/. Acesso em: jan 2015.

NERIS, M. N., SILVA, A. J., PERES, S. M. e FLORES, F. C. **Self organizing maps and bit signature: a study applied on signal language recognition,** in International Joint Conference on Neural Networks, 2008, p. NN0814

NINH, N.D.; SHUICHI, E. e EJIMA, T. **Real-Time Hand Tracking and Gesture Recognition System**, 2005.

OLIVEIRA, Sandra Maria; **Técnicas Geométricas de Condensação para o classificador K-NN;** Aveiro, 2008. Dissertação de mestrado-Universidade de Aveiro Departamento de matemática, Portugal, 2008.

OPENNI. OpenNI, nov 2010. **Biblioteca livre para o desenvolvimento da interoperabilidade de dispositivos de interação natural.** Disponível em: http://www.openni.org/. Acesso em: jan.2015.

PAULRAJ, M.P.; YAACOB, S.; AZALAN, M.S.Z.; PALANIAPPAN, R., **A phoneme based sign language recognition system using 2D moment invariant interleaving feature and Neural Network**, Research and Development (SCOReD), 2011 IEEE Student Conference on , vol., no., pp.111,116, 19- 20 Dec. 2011.

PEDRINI, H.; SCHWARTZ, W. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações**, Thomson Learning, São Paulo, 2008.

PIMENTA, N.; QUADROS, R. M de. **Curso LIBRAS 1.** 4ª Edição. Editora Vozes, 2010.

PIZZOLATO, E. B.; ANJO, M. S.; PEDROSO, G. C. - **Automatic Recognition of Finger Spelling for LIBRAS based on a Two Layer Architecture.** In: ACM Symposium On Applied Computing, 2010, Sierre. Proceedings of the 25th Symposium on Applied Computing, 2010. p. 970-974.

PORFIRIO, A., WIGGERS, K., OLIVEIRA, S., e WEINGAERTNER, D. **Libras sign language hand configuration recognition based on 3D meshes**. IEEE Int. Conf. on Systems, Man, and Cybernetics, 2013.

QUAN, Y. **Chinese sign language recognition based on video sequence appearance modeling**, Industrial Electronics and Applications (ICIEA), 2010 the 5th IEEE Conference on, vol., no., pp.1537, 1542, 15-17 June 2010

RAMOS, Marcos Roberto de; **Uso do sensor Kinect® para medir a regularidade na distribuição de fertilizantes sólidos.** Ponta Grossa, 2012. Dissertação de mestrado-Universidade Estadual de Ponta Grossa, Ponta Grossa, 2012.

RAMOS, C. R. LIBRAS: **A Língua de Sinais dos Surdos Brasileiros**, Petrópolis, 2006. Disponível em: <http://www.editora-arara-azul.com.br/pdf/artigo2.pdf>. Acesso em: 05 Dezembro 2014.

RÉZIO, A.C.C; SCHWARTZ, W.R, PEDRINI, H. P. **Super-Resolução de Imagens Baseada em Aprendizado Utilizando Descritores de Características**. X Congresso Brasileiro de Inteligência Computacional (CBIC), Fortaleza-CE, Brazil, November 08-11, 2011.

RODRÍGUEZ, K. C. O; CHÁVEZ, G. C; MENOTTI, D. **Hu e Zernike Moments for Sign Language Recognition.** In: The 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition, 2012, Las Vegas. IPCV'12, 2012.

ROSSI, A. R.; ROSSI, J. M. A. **Curso de LIBRAS básico**. Centro de Ensino, Pesquisa e Extensão e atendimento em Educação Especial, Universidade Federal de Uberlândia, 2009.

SANTOS, J.R. **Reconhecimento das configurações da libras baseado na analise de discriminante de fisher bidimensional utilizando imagens de profundidade**. Manaus, 2015. Dissertação de mestrado-Programa de Pós Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, Manaus, 2015.

SHANTZ, M. e POIZNER, H. **A computer program to synthesize american sign language**. Behavior Research Methods & Instrumentation, Vol. 14(5),467-474, 1982.

SILVA, Gabriel de França Pereira e; **Photodoc- um ambiente para processamento de imagens de documentos adquiridas por câmeras digitais portáteis;** Recife, 2009. Dissertação (Mestrado). Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Pernambuco, Recife, 2009.

SILVA, Joaquim Guilherme Vasconcelos Gonçalves da; **Sistema de aumento de segurança para cadeira de rodas baseada num sensor Kinect;** Bragança, 2013. Dissertação de mestrado-Escola Superior de Tecnologia e de Gestão do Instituto Politécnico de Bragança, Bragança, 2013.

SILVA, J.P. da; LAMAR, M.V.; BORDIM, J.L., **A study of the ICP algorithm for recognition of the hand alphabet**, Computing Conference (CLEI), 2013 XXXIX Latin American , pp.1,9, 7-11 Oct. 2013.

SILVA, Marcelo Mourão; **Uma abordagem Evolucionária Para o Aprendizado Semi-Supervisionado em Máquinas de Vetores de Suporte;** Belo Horizonte, 2008. Dissertação de mestrado-Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais, Belo Horizonte, 2008.

SILVA, J.P da.; LAMAR, M.V.; BORDIM, J.L., **A study of the ICP algorithm for recognition of the hand alphabet**, Computing Conference (CLEI), 2013 XXXIX Latin American , vol., no., pp.1,9, 7-11 Oct. 2013

The Kinect patent - method and system for object reconstruction, 2005, http://www.wipo.int/patentscope/search/en/WO2007043036. Acesso em: jan. 2015.

TINO, Vicente Fernandes; **Utilização de análise de componentes principais na regulagem de máquinas de injeção plástica;** Rio de Janeiro, 2005. Dissertação de mestrado-Programa de Pós-Graduação de engenharia da Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2005.

VIADER, Maria P. F.; PERTUSA, Esther; VINARDELL, Marta**. Importância das estratégias e recursos do educador surdo no processo de ensino/aprendizagem da língua escrita**. In: SKLIAR, C. (Org.). Atualidade da educação bilíngue para surdos. Porto Alegre Mediação, 1999. p. 47-57.

VILLAROMAN, N., ROWE, D., e SWN, B. (2011). **Teaching Natural User Interaction using OpenNI and the Microsoft Kinect Sensor**. In Proceedings of the 2011 conference on Information technology education, SIGITE '11, pages 227{232, New York, NY, USA.

WANG, H.; LEU, M.C, Oz C (2006) **American sign language recognition using multi-dimensional hidden Markov models**. J Inf Sci Eng, vol 22, no. 5, pp 1109–1123.

WINGFIELD, N. **Kinect põe em jogo bilhões em P & D da Microsoft**, nov de 2010. Disponível em: http://blog.tribunadonorte.com.br/tecnologiacomentada/kinect-poe-em-jogo-bilhoes-em-pd-da-microsoft/45469/. Acesso em: jan. 2015.

YANG, J.; ZHANG, D.; YONG, X.; YANG, J. **Two-dimensional discriminant transform for face recognition** , Pattern Recognition Society, Vol. 38, P. 1125-1129, 2005.

Yi Li, **Hand gesture recognition using Kinect**. Software Engineering and Service Science (ICSESS), 2012 IEEE 3rd International Conference on, pp.196,199, 22-24 June 2012

ZHANG, D.; ZHOU, Z. (2D)²PCA: **Two-directional two-dimensional PCA for efficient face representation and recognition**, Neurocomputing, Vol. 69, P. 224–231, 2005

APÊNDICE

EXPERIMENTOS REALIZADOS

A Tabela a seguir apresenta os resultados de alguns dos 2.880 experimentos realizados.

Tabela A1 – Taxa de acerto global obtida em função das configurações dos experimentos realizados

Tipo de Interpolação		Dimensão da matriz de característica	Classificação	Taxa de acerto(%)
Segmentação	**Padronização**	**Redução de dimensionalidade**	**k-vizinhos mais próximos**	
Vizinho mais próximo	Vizinho mais próximo	5X5	1	95,01
			5	91,75
			9	88,45
		6X6	1	95,81
			5	92,88
			9	89,83
		7X7	1	95,91
			5	93,39
			9	90,16
		8X8	1	96,09
			5	93,19
			9	90,19
		9X9	1	96,24
			5	93,21
			9	89,91
		10X10	1	96,19
			5	93,11
			9	89,60
		11X11	1	96,26
			5	92,86
			9	89,70
		12X12	1	96,13
			5	92,68
			9	89,55
Vizinho mais próximo	Bilinear	5X5	1	94,98
			5	92,09
			9	88,67
		6X6	1	95,83
			5	93

Tipo de Interpolação		Dimensão da matriz de característica	Classificação	Taxa de acerto(%)
Segmentação	**Padronização**	**Redução de dimensionalidade**	**k-vizinhos mais próximos**	
			9	89,88
Vizinho mais próximo	Bilinear	7X7	1	95,96
			5	93,34
			9	90,29
		8X8	1	96,11
			5	93,44
			9	89,96
		9X9	1	96,26
			5	93,032
			9	90,18
		10X10	1	96,31
			5	93,24
			9	89,72
		11X11	1	96,19
			5	92,81
			9	89,45
		12X12	1	96,16
			5	92,75
			9	89,78
Vizinho mais próximo	Bicúbica	5X5	1	94,93
			5	92,04
			9	88,32
		6X6	1	95,80
			5	93,03
			9	89,80
		7X7	1	95,96
			5	93,21
			9	90,26
		8X8	1	96,08
			5	93,22
			9	90,18
		9X9	1	96,27
			5	92,96
			9	90,18
		10X10	1	96,26
			5	93,13
			9	90,03

Tipo de Interpolação		Dimensão da matriz de característica	Classificação	Taxa de acerto(%)
Segmentação	**Padronização**	**Redução de dimensionalidade**	**k-vizinhos mais próximos**	
Vizinho mais próximo	Bicúbica	11X11	1	96,16
			5	92,85
			9	89,29
		12X12	1	96,14
			5	92,90
			9	89,34
Bilinear	Vizinho mais próximo	5X5	1	90,98
			5	88,67
			9	86,39
		6X6	1	91,29
			5	88,83
			9	87
		7X7	1	92,32
			5	89,49
			9	87,54
		8X8	1	92,44
			5	89,55
			9	87,55
		9X9	1	92,65
			5	89,96
			9	87,90
		10X10	1	92,75
			5	90,13
			9	87,42
		11X11	1	92,67
			5	90,06
			9	87,55
		12X12	1	92,75
			5	89,72
			9	87,81
Bilinear	Bilinear	5X5	1	90,91
			5	88,70
			9	86,59
		6X6	1	91,57
			5	89,049
			9	87,065

Tipo de Interpolação		Dimensão da matriz de característica	Classificação	Taxa de acerto(%)
Segmentação	**Padronização**	**Redução de dimensionalidade**	**k-vizinhos mais próximos**	
Bilinear	Bilinear	7X7	1	92,22
			5	89,39
			9	87,37
		8X8	1	92,59
			5	89,68
			9	87,57
		9X9	1	92,88
			5	90,098
			9	87,72
		10X10	1	92,75
			5	89,90
			9	87,44
		11X11	1	92,83
			5	90,04
			9	87,70
		12X12	1	92,73
			5	90,04
			9	87,88
Bilinear	Bicúbica	5X5	1	90,85
			5	88,52
			9	86,59
		6X6	1	91,55
			5	88,93
			9	86,91
		7X7	1	92,19
			5	89,44
			9	87,39
		8X8	1	92,57
			5	89,57
			9	87,26
		9X9	1	92,81
			5	90,016
			9	87,59
		10X10	1	92,72
			5	89,96
			9	87,50
		11X11	1	92,78
			5	90

Tipo de Interpolação		Dimensão da matriz de característica	Classificação	Taxa de acerto(%)
Segmentação	**Padronização**	**Redução de dimensionalidade**	**k-vizinhos mais próximos**	
		11x11	9	87,80
		12X12	1	92,77
			5	89,72
			9	87,78
Bicúbica	Vizinho mais próximo	5X5	1	90,40
			5	87,95
			9	86,16
		6X6	1	91,55
			5	88,96
			9	86,73
		7X7	1	92,32
			5	89,24
			9	87,44
		8X8	1	92,52
Bicúbica	Vizinho mais próximo		5	89,67
			9	87,42
		9X9	1	92,68
			5	89,40
			9	87,45
		10X10	1	92,70
			5	89,32
			9	87,52
		11X11	1	92,75
			5	89,49
			9	87,75
		12X12	1	92,77
			5	89,57
			9	87,95
		5X5	1	90,57
			5	87,93
			9	85,88
		6X6	1	91,40
			5	88,90
			9	87,09
		7X7	1	92,29
			5	89,13
			9	87,31

Tipo de Interpolação		Dimensão da matriz de característica	Classificação	Taxa de acerto(%)
Segmentação	Padronização	Redução de dimensionalidade	k-vizinhos mais próximos	
Bicúbica	Bilinear	8X8	1	92,47
			5	89,32
			9	87,72
		9X9	1	92,55
			5	89,65
			9	87,42
		10X10	1	92,70
			5	89,47
			9	87,49
		11X11	1	92,73
			5	89,63
			9	87,77
		12X12	1	92,78
			5	89,29
			9	87,80
Bicúbica	Bicúbica	5X5	1	90,50
			5	88,01
			9	85,73
		6X6	1	91,40
			5	88,81
			9	86,98
		7X7	1	92,22
			5	89,22
			9	87,22
		8X8	1	92,49
			5	89,47
			9	87,37
		9X9	1	92,57
			5	89,45
			9	87,70
		10X10	1	92,63
			5	89,54
			9	87,42
		11X11	1	92,77
			5	89,78
			9	87,81
			1	92,85

Tipo de Interpolação		Dimensão da matriz de característica	Classificação	Taxa de acerto(%)
Segmentação	**Padronização**	**Redução de dimensionalidade**	**k-vizinhos mais próximos**	
Bicúbica	Bicúbica	12x12	5	89,31
			9	87,81