

**UNIVERSIDADE DE SÃO PAULO**  
**ESCOLA DE ENGENHARIA DE SÃO CARLOS**  
**DEPARTAMENTO DE ENGENHARIA ELÉTRICA**

**DETECÇÃO DE FACES HUMANAS EM  
IMAGENS COLORIDAS UTILIZANDO REDES  
NEURAIS ARTIFICIAIS**

Wellington da Rocha Gouveia

Dissertação apresentada à  
Escola de Engenharia de São  
Carlos, da Universidade de São  
Paulo, como parte dos  
requisitos para a obtenção do  
título de Mestre em Engenharia  
Elétrica.

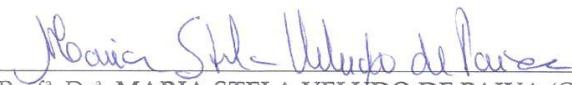
Orientadora: Profa. Dra. Maria Stela Veludo de Paiva

São Carlos  
2010

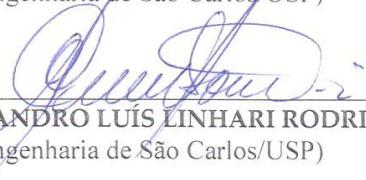
**FOLHA DE JULGAMENTO**

Candidato: Bacharel WELLINGTON DA ROCHA GOUVEIA.

Dissertação defendida e julgada em 28/01/2010 perante a Comissão Julgadora:

  
Profª. Drª. MARIA STELA VELUDO DE PAIVA (Orientadora)  
(Escola de Engenharia de São Carlos/USP)

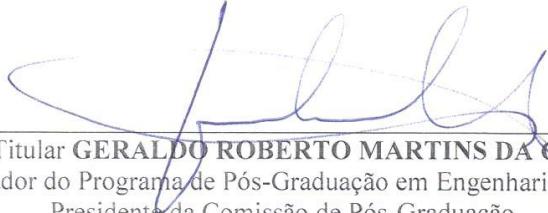
APROVADO

  
Prof. Dr. EVANDRO LUIS LINHARI RODRIGUES  
(Escola de Engenharia de São Carlos/USP)

APROVADO

  
Prof. Dr. JOÃO DO ESPÍRITO SANTO BATISTA NETO  
(Instituto de Ciências Matemáticas e de Computação/USP)

APROVADO

  
Prof. Titular GERALDO ROBERTO MARTINS DA COSTA  
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica e  
Presidente da Comissão de Pós-Graduação

## Dedicatória

---

---

Primeiramente a Deus, a minha mãe que já não esta entre nós, aos meus irmãos Wignas e Weder, à minha sobrinha Bianca, ao meu avô Justino, aos meus tios Almir e Lucia e às minhas primas Mônica, Rosana e Giovana.

## Agradecimentos

---

---

Agradeço a Deus por me presentear com a minha família, com meus colegas e amigos, que me ajudaram de diversas maneiras para a realização deste trabalho.

Agradeço a minha mãe Maria Aparecida, que mesmo não estando entre nós, sempre senti sua energia para a conquista deste trabalho.

Agradeço a toda a minha família, pelo companheirismo, paciência, carinho e incentivo. Em especial ao meu avô Justino, a minha madrinha Alzira (in memorian), as minhas primas Mônica, Rosana e Giovana, a minha sobrinha Bianca e principalmente aos meus irmãos Wignas e Weder.

Aos meus tios Almir e Lucia pelo apoio, incentivo e compreensão não só durante os estudos, mas por toda a minha a vida.

A minha orientadora, Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Stela Veludo de Paiva pelos ensinamentos, na escrita deste trabalho, mas também por acreditar no meu potencial de trabalho.

Aos meus colegas do LAVI (Laboratório de Visão Computacional), pelo apoio e companheirismo durante meus estudos, especial aos colegas André Tarallo, Ricardo Casado, Anderson Santos e Juliana Gouveia.

Aos companheiros de trabalho da Etec Paulino Botelho e da UAB/UFSCar que de alguma forma contribuíram para o desenvolvimento deste trabalho.

## Resumo

---

---

Gouveia, W. R. Detecção de Faces Humanas em Imagens Coloridas Utilizando Redes Neurais Artificiais. 2010. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2010.

A tarefa de encontrar faces em imagens é extremamente complexa, pois pode ocorrer variação de luminosidade, fundos extremamente complexos e objetos que podem se sobrepor parcialmente à face que será localizada, entre outros problemas. Com o avanço na área de visão computacional técnicas mais recentes de processamento de imagens e inteligência artificial têm sido combinadas para desenvolver algoritmos mais eficientes para a tarefa de detecção de faces. Este trabalho apresenta uma metodologia de visão computacional que utiliza redes neurais MLP (Perceptron Multicamadas) para segmentar a cor da pele e a textura da face, de outros objetos presentes em uma imagem de fundo complexo. A imagem resultante é dividida em regiões, e para cada região são extraídas características que são aplicadas em outra rede neural MLP para identificar se naquela região contém face ou não. Para avaliação do software implementado foram utilizados dois banco de imagens, um com imagens padronizadas (Banco AR) e outro banco com imagens adquiridas na Internet contendo faces com diferentes tons de pele e fundo complexo. Os resultados finais obtidos foram de 83% de faces detectadas para o banco de imagens da Internet e 88% para o Banco AR, evidenciando melhores resultados para as imagens deste banco, pelo fato de serem padronizadas, não conterem faces inclinadas e fundo complexo. A etapa de segmentação apesar de reduzir a quantidade de informação a ser processada para os demais módulos foi a que contribuiu para o maior número de falsos negativos.

**Palavras-Chave:** Detecção de Face, Imagens Coloridas, Redes Neurais, MLP, Perceptron Multicamadas, Segmentação de Imagens

## Abstract

---

---

Gouveia, W. R. Detection of Human Faces in Color Images Using Artificial Neural Networks. 2010. Dissertation (Master) – São Carlos Engineering School, University of São Paulo, São Carlos, 2010.

The task of finding faces in images is extremely complex, as there is variation in brightness, backgrounds and highly complex objects that may overlap partially in the face to be found, among other problems. With the advancement in the field of computer vision techniques latest image processing and artificial intelligence have been combined to develop more efficient algorithms for the task of face detection. This work presents a methodology for computer vision using neural networks MLP (Multilayer Perceptron) to segment the skin color and texture of the face, from other objects present in a complex background image. The resulting image is divided into regions and from each region are extracted features that are applied in other MLP neural network to identify whether this region contains the face or not. To evaluate the software two sets of images were used, images with a standard database (AR) and another database with images acquired from the Internet, containing faces with different skin tones and complex background. The final results were 83% of faces detected in the internet database of images and 88% for the database AR. These better results for the database AR is due to the fact that they are standardized, are not rotated and do not contain complex background. The segmentation step, despite reducing the amount of information being processed for the other modules contributed to the higher number of false negatives.

**Keywords:** Face Detection, Color Images, Neural Networks, MLP, Multilayer Perceptron, Segmentation of Images.

# Lista de Figuras

---



---

Figura 1 - Imagem em várias resoluções. (a) imagem original n=1. (b) n=4. (c) n=8. (d)n=16.....8	
Figura 2 - Uma face típica usada em métodos baseados em conhecimento .....	8
Figura 3 - Projeções verticais e horizontais em imagens.....	9
Figura 4 - Segmentação utilizando o método de proposto por Sirohey (1993).....	11
Figura 5 - A caixa delimita a detecção do nariz e da boca .....	12
Figura 6 - Exemplo da estrutura <i>parent vector</i> . As <i>grids</i> representam uma pirâmide de imagens .....	14
Figura 7 - Resultados obtidos para dois ou mais clusters de face .....	15
Figura 8 - (a) região da face selecionada e (b) cluster de cores no espaço de cores (RGB) ....	16
Figura 9 - Detecção da face por meio da cor da pele: (a) Imagem face típica. (b) Segmentação da cor da pele. (c) Maior região conectada de pixels de pele (BHUIYAN et al, 2003) .....	18
Figura 10 - (a) imagem original e (b) resultado da segmentação da cor da pele na imagem original (WANG e SUNG, 1999) .....	18
Figura 11 - <i>Template</i> utilizada por (YUILLE, 1992) para detectar olhos.....	21
Figura 12 - (a) Modelo utilizado para detectar a cabeça (b) Modelo utilizado para detectar os componentes faciais .....	21
Figura 13 - Projeções verticais e horizontais dos olhos obtida a partir da região interna da janela dos olhos .....	22
Figura 14 - Imagens originais e projeção delas no espaço de <i>eigenfaces</i> .....	23
Figura 15 - Algoritmo de detecção de faces baseado em RNA (ROWLEY et al. 1998a).....	25
Figura 16 - Sequência de passos do algoritmo de detecção de faces rotacionadas (ROWLEY et al, 1998a).....	27
Figura 17 - HMM para a detecção e reconhecimento de faces (NEFIAN e HAYES, 1998) .....	28
Figura 18 – Estrutura do neurônio (AMORIN, 2009) .....	38
Figura 19 - Modelo não-linear de um neurônio (HAYKIN, 2001).....	38
Figura 20 - Rede <i>feedforward</i> multicamada totalmente conectada com uma camada oculta e uma camada de saída.....	41
Figura 21 - Grafo arquitetural de um perceptron múltiplas camadas com duas camadas ocultas .....	42
Figura 22 – Direção da propagação da função sinal .....	43
Figura 23 – Gráfico da função sigmóide .....	44
Figura 24 - Diagrama em blocos da aprendizagem supervisionada .....	50
Figura 25 – Fluxograma do trabalho desenvolvido.....	55
Figura 26 – Algoritmo do trabalho desenvolvido .....	56
Figura 27 – Exemplo de imagens (a) adquirida na internet e (b) banco AR .....	57
Figura 28 – Exemplo de imagens utilizadas para teste, (a) e (b) adquiridas na internet (c) e (d) Banco AR. .....	57

Figura 29 – Exemplo de imagens utilizadas com características que dificultam a detecção .....	58
Figura 30 – Diagrama de blocos do pré-processamento da imagem.....	58
Figura 31 – Topologia da Rede Neural proposta para a segmentação de pele.....	60
Figura 32 – Diagrama de blocos da etapa de segmentação da imagem .....	60
Figura 33 – Fluxograma do método de varredura implementado .....	61
Figura 34 – Algoritmo do método de varredura implementado .....	62
Figura 35 – Exemplo de imagem após varredura.....	62
Figura 36 – Diagrama de blocos da etapa de pós-processamento da imagem.....	63
Figura 37 – Mascaras 3x3, vertical e horizontal do operador de Sobel.....	63
Figura 38 – Exemplo de imagem: (a) original, (b) segmentada, (c) binarizada e com (d) bordas .....	63
Figura 39 – Diagrama de blocos da etapa de extração de características da imagem .....	64
Figura 40 – Modelo do molde utilizado para extração de características .....	65
Figura 41 – Topologia da Rede Neural proposta para a detecção de face .....	66
Figura 40 - (a), (c) e (e) imagens originais, (b), (d) e (f) imagens processadas do banco de imagens da internet.....	68
Figura 41 - (a), (c) e (e) imagens originais, (b), (d) e (f) imagens processadas do banco de imagens AR .....	68
Figura 42 - (a) imagem original (b) imagem segmentada .....	69
Figura 43 - (a) imagem original (b) imagem segmentada.....	70
Figura 44 - (a) imagem original (b) imagem processada .....	70
Figura 45 – Gráfico de faces localizadas e faces com erosões na imagem após segmentação	71
Figura 46 - Resultado do algoritmo de detecção de face: (a) e (d) imagens originais e (b), (c), (e) e (f) imagens resultantes.....	72
Figura 47 – Gráfico com o total de faces localizadas .....	73
Figura 48 – Gráfico de com o total de faces localizadas e total de falsos positivos.....	75
Figura 49 – Artigos e periódicos utilizados na pesquisa de acordo com o ano de publicação ..	81

## Lista de Tabelas

---

---

Tabela 1 – Tabela que apresenta a quantidade de faces identificadas e faces com erosões na face.....	70
Tabela 2 – Tabela que apresenta o resultado do teste <i>T-Student</i> para o total de faces localizadas .....	71
Tabela 3 – Tabela que apresenta o total de faces e o total de faces detectadas .....	73
Tabela 4 – Tabela que apresenta o resultado do teste <i>T-Student</i> para o total de faces detectadas .....	73
Tabela 5 – Tabela que apresenta o total de pessoas e o total de falsos positivos nas imagens	74

## Lista de Abreviaturas e Siglas

---



---

MLP	Perceptron Multicamadas ( <i>Multi-Layer Perceptron</i> )
RNA	Rede Neural Artificial
PCA	Análise dos Componentes Principais ( <i>Principal Component Analysis</i> )
LDA	Análise Discriminante Linear( <i>Linear Discriminat Analysis</i> )
SOM	Mapa Auto-Organizavel ( <i>Self-Organizing Map</i> )
RGB	Vermelho, Verde e Azul ( <i>Red, Green and Blue</i> )
HSV	Tonalidade, Saturação e Valor ( <i>Hue, Saturation and Value</i> )
HSI	Tonalidade, Saturação e Intensidade ( <i>Hue, Saturation and Intensity</i> )
CMY	Ciano, Magenta e Amarelo ( <i>Cyan, Magenta and Yellow</i> )
YCrCb	Luminância, Crominância Vermelha e Crominância Azul
YIQ	Luminância, Vermelho e Azul
YES	Luminância e Componentes de Cor
CIE	<i>Comission Internationale de l'Eclairage</i>
KLT	Transformada de Karhunen-Loève ( <i>Karhunen-Loève Transform</i> )
OCR	Reconhecimento Ótico de Caracteres ( <i>Optical Character Recognition</i> )
HMM	Modelos Ocultos de Markov ( <i>Hidden Markov Models</i> )
LMS	Regra Delta ( <i>Least Mean-Square</i> )
MatLab	<i>Matrix Laboratory</i>
YPbPr	Luminância, diferença entre Azul e Luminância e diferença entre Vermelho e Luminância
YDbDr	Luminância, subtração da Luminância da cor Azul e subtração da Luminância da cor Vermelho

# Sumário

---

---

<b>DEDICATÓRIA</b>	II
<b>AGRADECIMENTOS</b>	III
<b>RESUMO</b>	IV
<b>ABSTRACT</b>	V
<b>LISTA DE FIGURAS</b>	VI
<b>LISTA DE TABELAS</b>	VIII
<b>LISTA DE ABREVIATURAS E SIGLAS</b>	IX
<b>1 INTRODUÇÃO</b>	1
1.1 Objetivo .....	2
1.2 Estrutura do Trabalho .....	2
<b>2 DETECÇÃO DE FACES</b>	5
2.1 Considerações Iniciais.....	5
2.2 Introdução .....	5
2.3 Métodos Baseados em Conhecimento.....	7
2.4 Métodos Baseados em Características Invariantes .....	10
2.5 Métodos Baseados em Templates .....	20
2.5.1 Templates Deformáveis .....	20
2.6 Métodos baseados na Aparência.....	23
2.6.1 Autofaces ( <i>Eigenfaces</i> ) .....	23
2.6.2 Redes Neurais .....	24
2.6.3 Modelos Ocultos de Markov .....	27
2.7 Considerações Finais .....	29
<b>3 FUNDAMENTAÇÃO TEÓRICA</b>	31
3.1 Considerações Iniciais.....	31
3.2 Técnicas de Detecção de Pele .....	31
3.2.1 Limiar de Classificação .....	32
3.2.2 Estatístico – Histograma de Cor .....	33
3.2.3 Estatístico - Paramétrico .....	34
3.3 Redes Neurais Artificiais.....	35
3.3.1 Modelo de um neurônio .....	37
3.3.2 Função de ativação .....	39
3.4 Arquitetura de rede.....	40
3.4.1 Redes feedforward multicamada .....	40
3.5 Perceptron Multicamadas (MLP).....	41
3.5.1 Algoritmo de Treinamento .....	44
3.6 Algoritmo de Aprendizagem.....	48
3.6.1 Aprendizagem Supervisionada .....	50
3.6.2 Considerações finais .....	51
<b>4 DETECÇÃO DE FACES HUMANAS EM IMAGENS COLORIDAS - METODOLOGIA</b>	53
4.1 Considerações Iniciais.....	53
4.2 Módulos Desenvolvidos .....	53
4.3 Metodologia.....	56
4.3.1 Pré-processamento .....	58
4.3.2 Segmentação de pele utilizando redes neurais .....	59
4.3.3 Varredura .....	61
4.3.4 Pós-processamento .....	62
4.3.5 Extração de características .....	64
4.3.6 Detecção de faces utilizando redes neurais .....	65
4.3.7 Software de desenvolvimento .....	66

<b>5</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>67</b>
5.1	Resultados obtidos com o algoritmo de segmentação.....	67
5.2	Resultados obtidos com o algoritmo de detecção de face .....	72
<b>6</b>	<b>CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS</b>	<b>77</b>
6.1	Sugestões para trabalhos futuros .....	78
6.2	Trabalho Publicado.....	79
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>81</b>
	<b>APÊNDICE A – BANCO DE IMAGENS DA INTERNET</b>	<b>89</b>
	<b>APÊNDICE B – BANCO DE IMAGENS AR</b>	<b>95</b>

# 1 Introdução

---

---

A detecção é uma das etapas iniciais do reconhecimento facial, sendo portanto, de extrema importância para o algoritmo de reconhecimento. Dada uma imagem, o objetivo da detecção de face é segmentar todas as regiões da imagem que contém uma face, independente de sua posição, orientação e condições de iluminação. Essa tarefa não é fácil, pois, as faces apresentam uma grande variação em tamanho, cor e textura. Além disso, o fundo da imagem pode possuir detalhes que sobrepõe parcial ou totalmente a face que será localizada, entre outros problemas (GOUVEIA E PAIVA, 2009; WON et all, 2001; GASPAR e PAIVA, 2005).

A combinação de técnicas das áreas de processamento de imagens e inteligência artificial, mais especificamente Redes Neurais Artificiais (RNA), Sistemas *Fuzzy* e Sistemas Inteligentes (uma combinação de RNA e Sistemas *Fuzzy*), tem sido utilizadas na implementação de algoritmos de detecção mais eficientes.

Alguns trabalhos podem ser mencionados na tarefa de detecção facial, podendo-se citar o de Jin et al (2007) e o de Zapata e Ruiz (2007) que utiliza informações de cor de pele, o de Santos (2005) que usa PCA-LDA (*Principal*

Component Analysis- Linear Discriminat Analysis) e RNA SOM (Self-Organizing Map), o de Nascimento (2005) que utiliza sistemas Fuzzy e o de Prado (2004) que utiliza proporção divina, e o trabalho de Gaspar (2006), onde foi desenvolvido um software para a extração de características faciais e o posterior reconhecimento da face.

### **1.1      Objetivo**

Esse trabalho propõe um algoritmo para detecção faces que usa RNA para segmentação de pele, separando cor de pele e objetos identificados como face, de outros objetos presentes em uma imagem. Com a segmentação de pele é reduzida a quantidade de informação para o processo de detecção.

O presente trabalho vem complementar o trabalho de Gaspar (2006), na etapa que antecede a extração de características, ou seja, identificar se existe face ou não numa dada imagem, onde essa imagem pode conter mais de uma face.

### **1.2      Estrutura do Trabalho**

Este trabalho está disposto da seguinte forma:

- Capítulo 1 – Introdução, que inclui a proposta e objetivo do trabalho;
- Capítulo 2 – Detecção de Faces, que apresenta os principais métodos utilizados para a detecção de faces humanas em imagens de interesse para esse trabalho;

- Capítulo 3 – Redes Neurais Artificiais, onde foram revisados os métodos de redes neurais artificiais utilizados para a realização deste trabalho;
- Capítulo 4 – Detecção de Faces em Imagens Coloridas - Metodologia, neste capítulo são apresentados os módulos desenvolvidos e a metodologia utilizada em cada módulo para o software de detecção de faces proposto;
- Capítulo 5 – Resultados e Discussões.
- Capítulo 6 – Conclusões e Sugestões para Trabalhos Futuros.



## **2 Detecção de Faces** ---

---

### **2.1 Considerações Iniciais**

Neste capítulo são descritos os principais trabalhos sobre detecção de faces, pesquisados na literatura.

### **2.2 Introdução**

A detecção de faces consiste na utilização de técnicas computacionais para determinar se existem faces ou não em uma imagem e, se existirem, deve retornar a localização de cada face. Embora a detecção de faces seja uma tarefa fácil para os seres humanos, a implementação de sistemas computacionais que realize esta tarefa é complexa, devido a diversas dificuldades, tais como Yang e Kriegman (2002), Lopes (2005) e Lyon e Vincent (2009):

- **Pose:** a posição de uma ou mais face presente numa imagem varia devido à posição da face em relação à câmera (frontal, 45 graus, perfil, de cabeça para baixo) e, além disso, alguns componentes faciais como olhos e nariz, podem estar parcialmente ou totalmente obstruídos.
- **Presença de componentes estruturais:** a presença de elementos na face tais como, barba, bigode e óculos, que podem modificar características como a forma, o tamanho e a cor.
- **Expressão facial:** a expressão da face afeta diretamente a aparência da face.
- **Oclusão:** as faces podem estar parcialmente obstruídas por outros objetos.
- **Orientação da imagem:** as imagens podem variar diretamente com a rotação da câmera ao redor do eixo.
- **Qualidade da imagem:** quando a imagem é formada, fatores de iluminação (espectro, fonte de distribuição e intensidade) e características da câmera (resposta de sensor e lentes) podem afetar a aparência da face.

Existem vários métodos para detecção de faces em imagens em escala de cinza e em imagens coloridas. Esses métodos estão classificados em quatro categorias, conforme apresentado por Yang (2002), as quais serão descritas nas próximas seções.

## 2.3      Métodos Baseados em Conhecimento

Os métodos de detecção de face baseados em conhecimento utilizam regras que definem o que é uma face, de acordo com uma base de conhecimento criada pelo pesquisador Yang (2002). Sabe-se que faces humanas, em geral, possuem determinados componentes como: dois olhos, um nariz e uma boca. Esses elementos encontram-se distribuídos de maneira específica sobre a face. Com este conhecimento é possível estabelecer regras que identificam uma face humana.

A construção do conjunto de regras para este método não é trivial, podendo-se encontrar algumas dificuldades. Se as regras são muito gerais, o sistema pode apresentar uma alta taxa de falsos positivos, ou seja, elementos erroneamente identificados como face. Por outro lado, se o conjunto de regras for muito específico, o sistema pode não ser eficaz ao tentar detectar faces se estas não satisfizerem todas as regras, caindo muito a precisão da detecção (YANG et al., 2002).

A técnica construída por Yang e Huang (1994) para detecção de faces, utiliza um método hierárquico baseado no conhecimento. Esse sistema consiste em três níveis de regras. O primeiro nível de regras tem como objetivo detectar os possíveis elementos da imagem candidatos a faces, retirados da mesma através de um algoritmo de janela deslizante. Basicamente o primeiro conjunto de regras descreve de maneira geral o que é uma face. Os próximos níveis são as descrições gerais da face como olhos, boca e nariz, necessários para confirmar a validade dos elementos extraídos no nível anterior.

Imagens em múltiplas resoluções são formadas através da aplicação da média em um conjunto de *pixels*, como é mostrado na figura 1. Considerando-se  $n$  um determinado número de *pixels*, geralmente contidos em uma janela quadrada de

$n \times n$  pixels, substitui-se o valor de cada pixel da janela, pelo valor médio dos pixels contidos dentro da janela. As regras são aplicadas de acordo com a resolução da imagem, por exemplo: em baixa resolução uma determinada regra identifica a região central do rosto por meio da afirmativa de que as janelas centrais, contendo olhos, nariz e boca, têm luminosidade uniforme, conforme é visto na figura 2.



Figura 1 - Imagem em várias resoluções. (a) imagem original  $n=1$ . (b)  $n=4$ . (c)  $n=8$ . (d) $n=16$

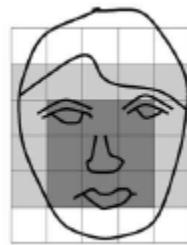


Figura 2 - Uma face típica usada em métodos baseados em conhecimento

Kotropoulos e Pitas (1997) apresentam um método para localização baseado em regras, similar aos que foram desenvolvidos por Kanade (1973) e Yang e Huang (1994). Kotropoulos e Pitas utilizam um sistema baseado em regras onde características faciais são localizadas utilizando projeções, método que Kanade utilizou com sucesso, para encontrar o limite da face em imagens. Considerando

$I(x,y)$  como um valor de intensidade da imagem de tamanho  $n \times m$  na posição  $(x,y)$ , as projeções horizontais e verticais da imagem são definidas como:

$$HI(x) = \sum_{y=1}^n I(x, y) \quad (1)$$

$$VI(y) = \sum_{x=1}^m I(x, y) \quad (2)$$

As projeções verticais são utilizadas para localizar determinados componentes faciais como, por exemplo, os olhos, por meio da detecção de mínimos locais da projeção. O mesmo tratamento é dado à projeção horizontal,  $HI$ , onde dois mínimos locais obtidos através da detecção de mudanças bruscas na projeção correspondem ao lado esquerdo e direito da face respectivamente, como pode ser visto na figura 3. A principal desvantagem deste método é que quando a imagem possui um fundo complexo a detecção da face e as características faciais ficam comprometidas (figura 3b). A detecção da face é feita pela análise dos picos presentes na projeção horizontal e os vales presentes na projeção vertical (YANG et al., 2002).

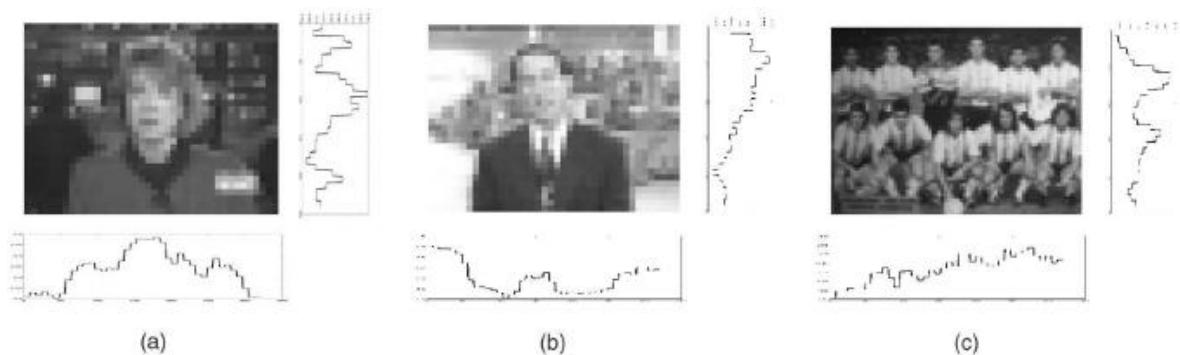


Figura 3 - Projeções verticais e horizontais em imagens

## 2.4      **Métodos Baseados em Características Invariantes**

Ao contrário dos métodos baseados em conhecimento vistos anteriormente, os métodos discutidos nesta seção incorporam técnicas que tem por objetivo encontrar características invariantes da face. Estes métodos são baseados na capacidade que os seres humanos possuem de identificar facilmente faces e objetos em diferentes posições e condições de iluminação. Pode-se descrever alguns métodos, tais como os que utilizam componentes faciais, textura e cor da pele.

Existem vários métodos propostos para detectar componentes faciais e então deduzir a presença de uma face. Componentes faciais como sobrancelhas, olhos, nariz, boca e o contorno da face são extraídos geralmente usando detectores de borda e, baseando-se nos componentes faciais extraídos, um modelo estatístico é construído para verificar a existência de face. A principal desvantagem desta abordagem é que os componentes faciais podem ser corrompidos devido às condições de iluminação, ruído e obstrução, comprometendo assim, a eficiência da abordagem (YANG et al, 2002).

Sirohey (1993) propôs um método para localização e segmentação da face em uma imagem com fundo complexo, onde é utilizado um detector de bordas e heurísticas para remover e agrupar as bordas, de modo que sejam somente preservadas as bordas do contorno da face. Uma elipse é ajustada entre a região principal e o fundo, localizando assim a face na imagem, como é mostrado na figura 4.

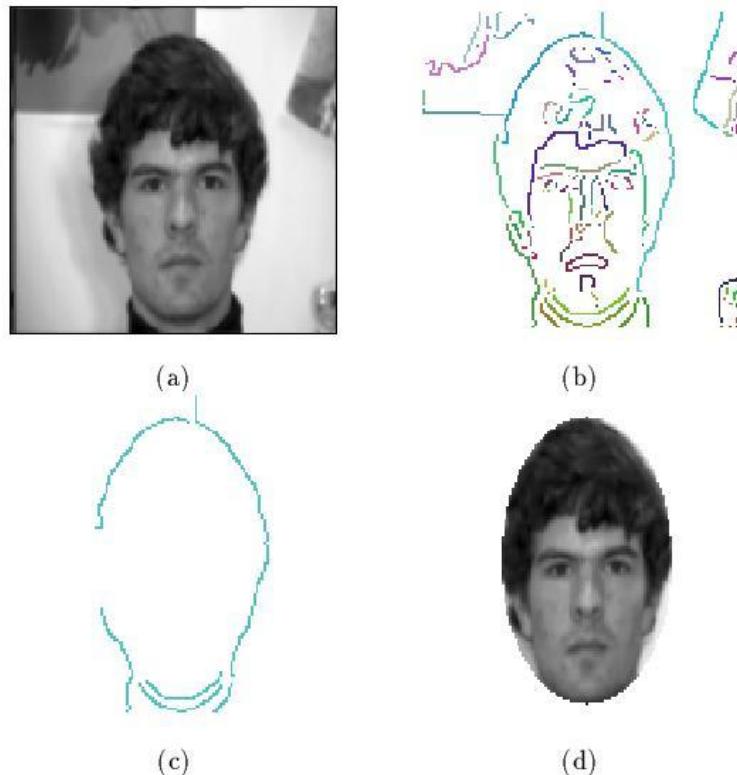


Figura 4 - Segmentação utilizando o método proposto por Sirohey (1993)

Amit et al. (1997) apresentam um método para detectar instâncias de objetos isolados em imagens com uma grande intensidade de objetos ao fundo. Este método foi aplicado para detectar e localizar apenas faces com visão frontal. A detecção é feita em dois estágios: “focalização” e classificação intensiva. A focalização é baseada em um arranjo espacial de fragmentos de borda, onde primeiramente é definida uma enorme família de arranjos espaciais e todas as constantes são definidas sobre uma larga escala de transformações fotométricas e geométricas que formam uma coleção de triângulos que são utilizados como mecanismos de atenção visual (“focalização”) por identificar possíveis grupos de bases que são utilizadas pra reduzir o aumento do numero de falsos negativos. Então, somente usando exemplos de faces, selecionam-se arranjos particulares que são mais comuns nas faces do que nos fundos. O estágio de classificação intensiva

é baseado em textura, e utiliza um conjunto de dados de faces normalizadas através de um treinamento e na coleção de regiões registradas e padronizadas na etapa de focalização. Dada uma imagem de teste, as regiões de interesse são identificadas de acordo com os arranjos espaciais de fragmentos de borda. Cada região é então classificada como face ou fundo usando a classificação intensiva como é visto no figura 5.



Figura 5 - A caixa delimita a detecção do nariz e da boca

Outra característica invariante utilizada para detectar diferentes objetos é a forma. A detecção de objetos através da forma baseia-se na capacidade que os seres humanos têm de reconhecer um objeto que ora tem uma forma determinada e

em outro momento apresenta uma forma totalmente diferente. Como exemplo tem-se a face humana que muda drasticamente dependendo do ponto de vista, onde a face vista de um ângulo frontal é totalmente diferente quando vista de perfil (LOPES, 2005).

A forma, assim como a cor, também é independente do ponto de vista. Pode ser encarada como uma distribuição probabilística, ou seja, ela é gerada a partir de um processo estocástico que fornece algum padrão a partir de permutações aleatórias (RIKERT, 1999). No método elaborado por Rikert é suposto que as imagens perceptualmente diferentes possuem distribuições similares em diferentes bandas e orientações. Tendo isso como base é utilizada uma transformada de *wavelets* guiada (*steerable wavelet transform*), ou também chamada de transformada de *wavelets* em pirâmide (*steerable wavelet pyramid*). A utilização de *wavelets* permite analisar a imagem em diferentes escalas e orientações, o que resulta em um conjunto de coeficientes independentes entre si, mas dependentes quando a mesma transformada é realizada em diferentes escalas da mesma imagem (LOPES, 2005; RIKERT, 1999).

Um conjunto de vetores denominados *parent vectors* foi utilizado para capturar esta dependência, que consiste na coleção de respostas dos filtros aplicados em diferentes escalas da imagem. Diferentes escalas da imagem são representadas utilizando um *quadtree* (figura 6), formando uma pirâmide de imagens, onde o segmento de reta representa os *pixels* nos quais os valores dos filtros formam um único *parent vector*.

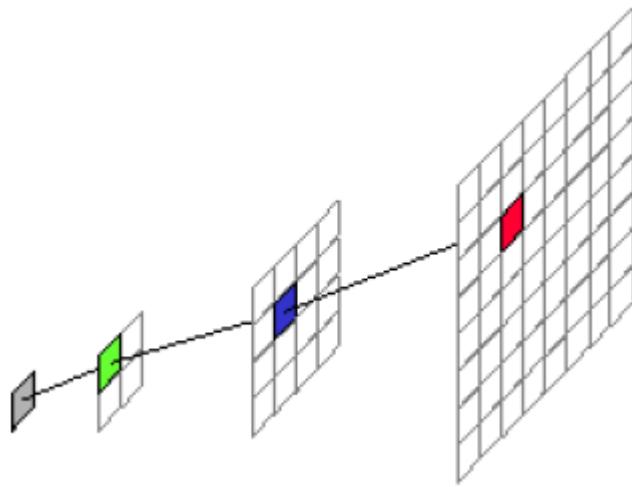


Figura 6 - Exemplo da estrutura *parent vector*. As *grids* representam uma pirâmide de imagens

Para calcular os *parent vectors*, primeiramente uma imagem de entrada é transformada em uma pirâmide de imagens, através da utilização de uma função gaussiana piramidal aplicada sobre a imagem. Como pode ser visto na figura 6, cada nível da pirâmide é reescalado em um fator de 2 em cada dimensão. Em seguida, para todos os níveis da pirâmide uma série de filtros é aplicada em cada *pixel*. Após a obtenção dos *parent vectors*, estes são agrupados em classes através da aplicação de um algoritmo de *clustering*, para que posteriormente seja possível construir um modelo gaussiano misto que generalize a distribuição.

Como pode ser visto na figura 7, o *cluster* a que pertence um determinado *parent vector* é selecionado. Vários *clusters* são tomados para verificar a presença ou não da face.



Figura 7 - Resultados obtidos para dois ou mais clusters de face

A cor da pele humana e a textura da face são características bastante utilizadas para separar a face de outros objetos presentes em uma cena com fundo complexo. As informações sobre as cores da pele constituem uma importante ferramenta para identificar áreas da face e os componentes faciais específicos. Existe uma grande variedade de cores de pele (branca, negra, amarela, etc), e existe um grande número de pesquisas que utilizam a cor da pele. Para rotular os pixels da pele são utilizados os diversos espaços de cores existentes: RGB, RGB normalizado, CMY, YUV, YPbPr, YDbDr, HSV (ou HSI), YCrCb, YIQ, YES, CIE XYZ e CIE LUV (YANG et al, 2002). Pode-se citar os seguintes trabalhos que usam alguns desses espaços de cores: Jin et al. (2007), Zapata e Ruiz (2007), Feris et al. (2000), Sobottka e Pittas (1996), Cai e Goshtasby (1999), Bhuiyan et al. (2003), Hsu et al. (2002), Wang e Sung (1999), Saber e Tekalp (1998) e Yang e Ahuja (1998).

Sobottka e Pittas (1996) propõem um método para localização e extração de características faciais utilizando formas e cores, executado a segmentação de cor no espaço HSV para localizar as regiões semelhante à pele.

A pele humana tende a formar um *cluster* no espaço de cores, podendo ser modelada por uma distribuição Gaussiana (figura 8), característica que pode ser utilizada para localizar faces em uma imagem colorida (WANG e SUNG, 1999; SABER e TEKALP, 1998).

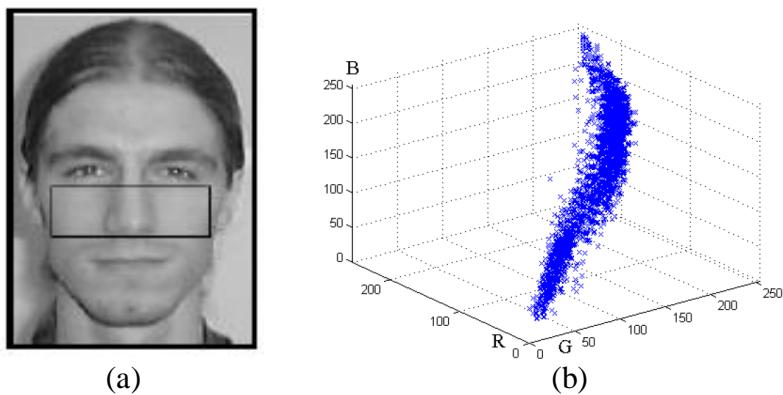


Figura 8 - (a) região da face selecionada e (b) cluster de cores no espaço de cores (RGB)

No trabalho de Bhuiyan et al. (2003), a cor da pele é utilizada para determinar a região da face. As cores relevantes e dominantes são extraídas da imagem em RGB. Em seguida a imagem é transformada para o espaço de cores YIQ descrito pelas componentes: matiz (*hue*), luminosidade e saturação. Este sistema de cores é universalmente utilizado em transmissões coloridas de televisão e produz uma transformação linear do RGB que gera Y, representando a luminosidade, e I e Q os dois canais que contém as informações para tratar cores. A matriz de transformação do espaço de cores RGB para YIQ é dada pela seguinte equação:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.320 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3)$$

Uma vez que a cor da pele tende a se aglomerar numa região do espaço de cores, um *threshold* é utilizado para detectar os *pixels* de pele (*skin pixels*). Nos experimento de Buhiyan *et al.* (2003) o seguinte *threshold*, determinado empiricamente, é utilizado:

$$(60 < Y < 200) \text{ and } (20 < I < 50) \quad (4)$$

Logo após a aplicação da equação 4 tem-se apenas os *pixels* que possivelmente fazem parte da pele humana. Para localizar efetivamente a face, toma-se a imagem com maior região conectada de *pixels* de pele. Um exemplo de resultado da aplicação desta técnica pode ser visto na figura 9. Também na figura 10 tem-se outro exemplo da aplicação da mesma técnica descrita anteriormente, notando-se que as cores da pele de diferentes pessoas são agrupadas e normalizadas no plano de cores r-g, a única diferença na aplicação da técnica é que a segmentação das cores foi feita a partir de imagens normalizadas nesse plano de cores. Segundo Wang e Sung (1999) a técnica permitiu atingir resultados com 100% de sucesso quando testada em 50 imagens obtidas a partir de câmeras CCD.

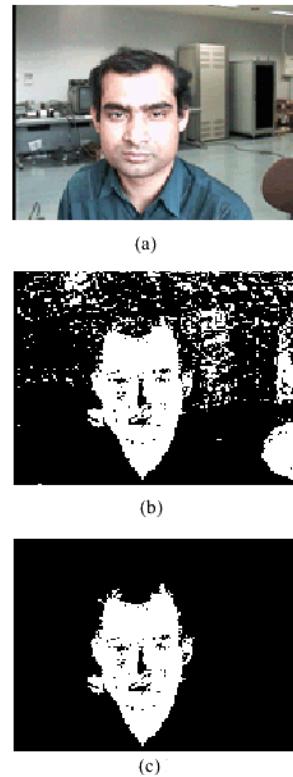


Figura 9 - Detecção da face por meio da cor da pele: (a) Imagem face típica. (b) Segmentação da cor da pele. (c) Maior região conectada de pixels de pele (BHUIYAN et al, 2003)

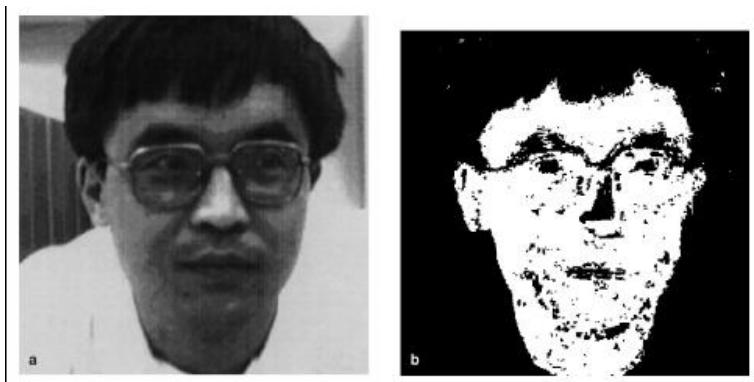


Figura 10 - (a) imagem original e (b) resultado da segmentação da cor da pele na imagem original (WANG e SUNG, 1999)

Apesar dos bons resultados apresentados, o sucesso ou fracasso das técnicas baseadas em segmentação da cor da pele é dependente do espaço de representação das cores, ou seja, o espaço de cores no qual a imagem está definida

influi na precisão da detecção da face. No trabalho de Saber e Tekalp (1998) é advertido que a utilização do espaço de cores RGB torna a técnica de segmentação de cores sensível à luminosidade. Devido a esta desvantagem outros espaços de cores foram propostos com o intuito de tornar mais robustos os algoritmos que utilizam a cor da pele. Como exemplo pode-se citar o YCbCr (*luminance-chrominance*), utilizado em várias aplicações de segmentação de pele. Entretanto Saber e Tekalp (1998) destacam a utilização do sistema de cores YES (definido pela *Society of Motion and Television Engineers*), onde Y representa a luminosidade, E e S representam as componentes de cor. A principal vantagem deste espaço de cor é que ele reduz as variações na cor (crominância) e principalmente porque permite calcular as componentes E e S a partir de uma imagem RGB apenas fazendo deslocamentos de *bits*, sem a necessidade de multiplicações, o que reflete de maneira positiva no desempenho dos algoritmos.

Entretanto, independente do espaço de cores utilizado, as técnicas de segmentação de pele permitem implementar detectores de face suficientemente rápidos para serem utilizados em tarefas de rastreamento de faces (*face tracking*) (VEZHNEVETS, 2002), úteis em áreas como segurança e interfaces inteligentes. Apesar das vantagens de se utilizar a cor para detectar a pele, um problema ocorre quando se tem objetos com cores semelhantes à pele, consequentemente provocando a ocorrência de falsos positivos. Em vista destas limitações, em tarefas como detecção e rastreamento de faces, geralmente a segmentação de pele vem acompanhada por outras técnicas que utilizam informações inerentes à forma do objeto (no caso a face), e informações do movimento para localizar e rastrear a face (YANG e AHUJA, 1998).

## 2.5      Métodos Baseados em Templates

Nos métodos baseados em *templates*, o objeto é procurado em uma imagem, utilizando um molde padrão que é definido manualmente, ou parametrizado por uma função. Uma das maneiras mais comuns de se modelar a forma de um objeto é descrevê-lo através de seus componentes geométricos básicos, como círculos, quadrados ou triângulos. A detecção do objeto consiste em localizar a melhor simetria entre o objeto presente na imagem e o seu *template* (molde) (LOPES, 2005; YANG et al., 2002).

No caso de detecção de faces o *template* mais utilizado é aquele que trata a face como uma elipse, como pode ser visto em Alattar e Rajala (1999) e Lee et al. (2001), sendo também aplicado em seqüências de imagens, como pode ser visto em Vezhnevets (2002). A técnica de *templates* é extremamente flexível, por este motivo, ela pode ser utilizada para detectar qualquer objeto que possa ser representado por formas geométricas.

### 2.5.1    Templates Deformáveis

A técnica de *templates* deformáveis proposta por Yuille et al (1992) para modelar características faciais consiste em descrever modelos de objetos através de formas geométricas simples como círculos ou parábolas. Esses modelos são elásticos e permitem o ajuste de seu tamanho, largura e outros parâmetros, como pode ser visto na figura 11. A energia é uma função de todos os parâmetros do *template* é definida através dos vales, picos e bordas presentes na imagem. O

melhor ajuste do modelo elástico é encontrado minimizando uma função de energia dos parâmetros.

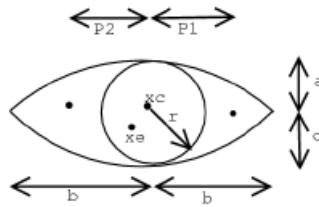


Figura 11 - *Template* utilizada por (YUILLE, 1992) para detectar olhos

Em Lee et al. (2001) é apresentada uma técnica que utiliza algoritmos genéticos para otimizar os parâmetros do *template* que modela a face. O algoritmo genético torna mais rápido o processo de busca pela face numa imagem, e permite obter a solução global. Uma vez que se tenha obtido a localização da região que contenha a face, as características faciais são detectadas dentro desta região por um processo semelhante. Um outro método que utiliza esta estratégia é visto em Alattar e Rajala (1999), que descreve um algoritmo para localizar componentes faciais. A cabeça é tratada como uma elipse podendo apresentar rotações em torno de um eixo vertical, como é apresentado na figura 12.

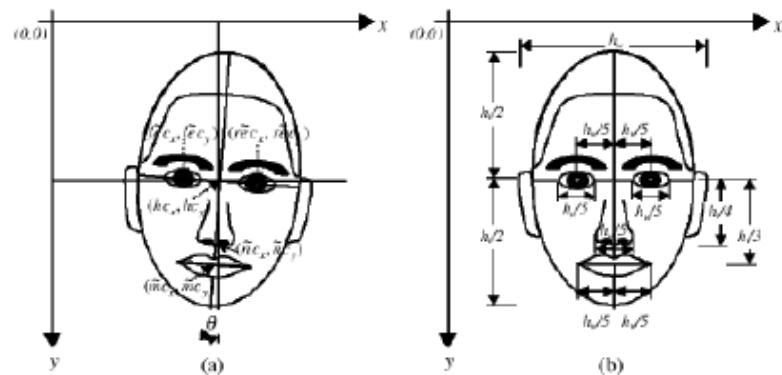


Figura 12 - (a) Modelo utilizado para detectar a cabeça (b) Modelo utilizado para detectar os componentes faciais

Após encontrar a elipse que melhor descreve a cabeça (figura 12a), os parâmetros que determinam o centro, a altura e a largura são utilizados para estimar a posição dos componentes faciais segundo suas relações geométricas com a face (figura 12b). Como as estimativas baseiam-se unicamente em grandezas geométricas, têm-se apenas localizações aproximadas dos componentes. Para localizar, por exemplo, os olhos, uma janela retangular é definida sobre a região referente a eles. Dentro da janela, a posição exata dos olhos é calculada através de projeções verticais e horizontais. Os valores presentes no gráfico apresentado na figura 13 indicam a localização das órbitas do olho esquerdo e direito respectivamente. A localização dos outros componentes faciais (nariz e boca) é semelhante à descrita para localização dos olhos.

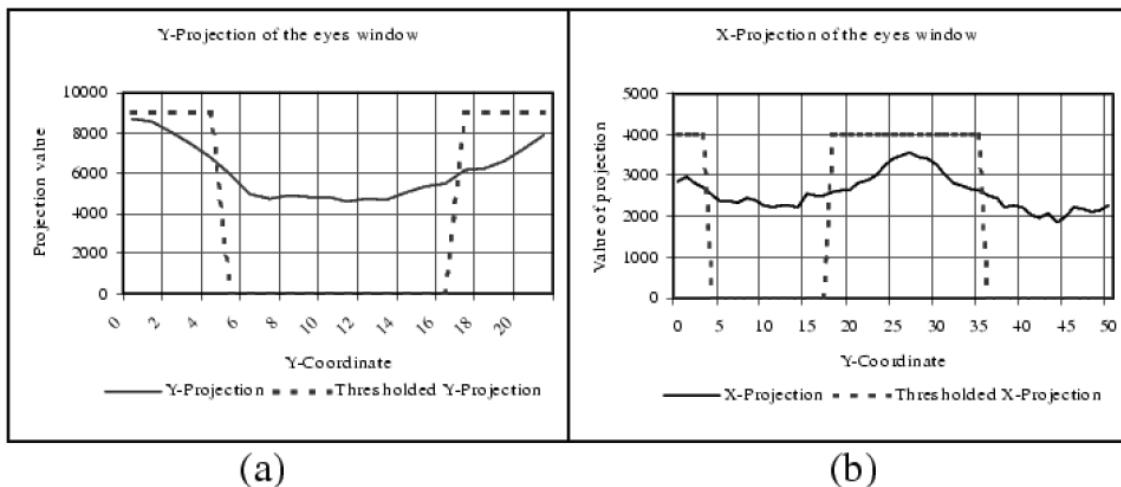


Figura 13 - Projeções verticais e horizontais dos olhos obtida a partir da região interna da janela dos olhos

## 2.6      Métodos baseados na Aparência

Nos métodos baseados na aparência, os modelos são definidos utilizando técnicas de aprendizado e treinamento, não sendo utilizados nenhum conhecimento ou características sobre o objeto previamente informado, como os vistos nos métodos descritos na sessão anterior. Desta maneira as informações necessárias para realizar a detecção são retiradas do conjunto de imagens (YANG et al, 2002; LOPES, 2005).

### 2.6.1    Autofaces (*Eigenfaces*)

O método autofaces proposto por Turk e Pentland (1991) é baseado na análise dos componentes principais (*Principal Component Analysis – PCA*), juntamente com a transformada de Karhunen-Loève (KLT) para reconhecimento e detecção de face. É similar ao trabalho de Kryby e Sirovich (1990).

O objetivo é usar a transformada KLT para encontrar os vetores que melhor representam as imagens de faces, dentro do espaço de imagens. Estes vetores são denominados de autofaces devido à semelhança que possuem com imagens de faces (figura 14).

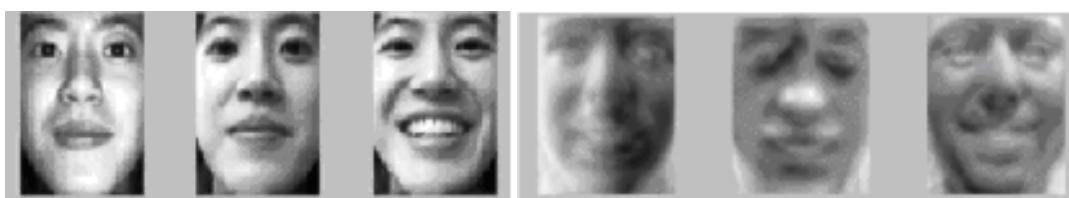


Figura 14 - Imagens originais e projeção delas no espaço de *eigenfaces*

As imagens de face são projetadas no subespaço e agrupadas. De forma similar, no treinamento de não-faces, as imagens são projetadas no mesmo subespaço e agrupadas. As imagens de face não apresentam grandes mudanças quando projetadas no espaço de autofaces, mas quando uma imagem de não-face é projetada, ela mostra-se completamente diferente (YANG et al, 2002). Pode-se utilizar esta técnica para detectar a presença de face em uma cena, apenas calculando a distância entre a região da imagem e o cluster de faces projetadas, e através de um *threshold* verificar se esta é uma face ou não (LOPES, 2005).

### 2.6.2 Redes Neurais

Diferentes arquiteturas de redes neurais artificiais foram propostas para a detecção de faces, pois a detecção de faces pode ser tratada como um problema de reconhecimento de padrões, tal como em OCR (reconhecimento ótico de caracteres), reconhecimento de objetos, etc., onde as redes neurais artificiais são aplicadas com sucesso (HSU et al, 2002; ROWLEY et al, 1998a; ROWLEY et al, 1998b).

Na detecção de faces, a utilização de redes neurais consiste em diferenciar imagens ou regiões da imagem que, possivelmente contenham uma face, daquelas imagens ou regiões que não contém face, ou seja, a função da rede é apenas classificar as imagens em face e não-face. (LOPES, 2005) .

No trabalho de Rowley et al. (1998a) é apresentado um sistema de detecção de faces baseados em redes neurais. Neste sistema uma imagem de entrada é percorrida por um filtro, que consiste em uma janela de 20x20, que extraí regiões da imagem onde podem existir faces. Logo após a extração da possível região de face,

esta é submetida à equalização de histograma e à normalização da luz, devido às várias condições de luminosidade. Se a face for maior que o tamanho da janela, a imagem é sub-amostrada repetidamente por um fator de 1.2, e o filtro é aplicado em cada sub-amostragem (figura 15). Cada bloco da imagem depois de extraído e pré-processado é enviado para uma rede neural, que decidirá se nele existe ou não uma face.

A rede é composta por várias camadas com arquiteturas diferentes, como mostrado na figura 15, denominadas campos receptivos. Cada camada avalia uma parte de imagem com a finalidade de se detectarem características que podem ser úteis para determinar a presença de face. A rede apresenta como resposta um único valor, que determina a existência ou não da face. Com os testes realizados neste trabalho (ROWLEY et al, 1998a) utilizando imagens coletadas na internet, obteve-se uma taxa de acerto (detecção correta) da ordem de 92.7%, utilizando uma rede com 78 neurônios intermediários e 4357 conexões.

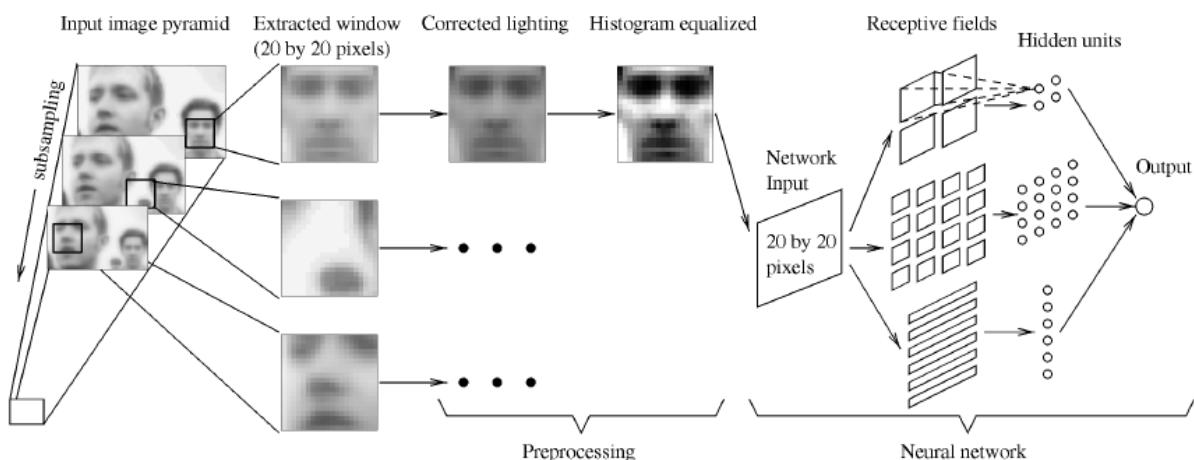


Figura 15 - Algoritmo de detecção de faces baseado em RNA (ROWLEY et al. 1998a)

A estratégia descrita anteriormente é eficaz apenas na detecção de faces frontais ou com leves variações na rotação. Diante desta limitação Rowley et al. (1998b) estendeu essa técnica de modo que a mesma seja robusta quanto a rotações da face. A modificação consistiu na utilização de outra rede neural denominada rede roteadora (*router network*), responsável por determinar o ângulo de rotação da face. A arquitetura da rede roteadora é constituída por uma camada de entrada com 400 unidades, uma camada intermediária com 15 unidades e, finalmente uma camada de saída com 36 unidades. O ângulo da face é expresso como um vetor contendo os valores das 36 saídas, onde cada unidade de saída  $i$  representa um ângulo de  $i * 10^{\circ}$ . A partir daí uma soma ponderada é obtida pela equação 5 e a direção desse vetor médio indica o ângulo da face (ROWLEY et al, 1998b).

$$\left( \sum_{i=0}^{35} output_i * \cos(\text{angle}) * 10^{\circ}, \sum_{i=0}^{35} output_i \sin(\text{angle}) * 10^{\circ} \right) \quad (5)$$

Logo após a aplicação da rede roteadora sobre uma janela, e determinada a direção da possível face, a janela é rotacionada em sentido contrário, de modo a deixar a possível face em uma vista frontal (*upright view*). Na figura 16 é apresentada a sequência de passos deste algoritmo de detecção de faces (ROWLEY et al., 1998b).

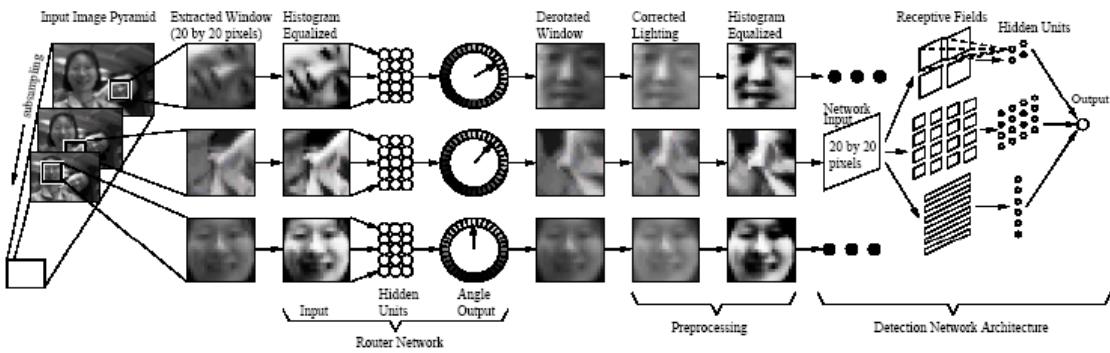


Figura 16 - Sequência de passos do algoritmo de detecção de faces rotacionadas (ROWLEY et al, 1998a)

### 2.6.3 Modelos Ocultos de Markov

Desde a década de 70, os Modelos Ocultos de Markov (*Hidden Markov Models - HMM*) tem sido largamente aplicados em várias áreas, principalmente no reconhecimento de fala (LEE et al., 1990; RABINER, 1989). Tais modelos possuem como principal característica a capacidade de caracterização de propriedades estatísticas de um sinal.

Na utilização do HMM para problemas de reconhecimento de padrão, é necessário definir um conjunto de estados ocultos do modelo, que é convertido em um conjunto de diferentes símbolos de observação, que serão utilizados para realizar o treinamento do HMM. Esse treinamento é realizado ajustando os parâmetros do modelo HMM, utilizando o algoritmo de segmentação Viterbi e o algoritmo recursivo Bam-Welch (RABINER e JUNG, 1993; YANG et al., 2002). O treinamento é um processo iterativo e pára quando a probabilidade de segmentação Viterbi é menor do que um limiar. Os parâmetros finais do HMM são obtidos através do procedimento recursivo Bam-Welch (LOPES, 2005).

O HMM também pode ser utilizado para detecção e reconhecimento de faces, como apresentado em diversos trabalhos (XU et al., 2003; LE e LI, 2004; SAMARIA, 1994; NEFIAN e HAYES, 1998).

Em Nefian e Hayes (1998) o HMM é utilizado tanto para a detecção como para o reconhecimento de faces frontais. Nesse trabalho, cada estado do HMM está associado a uma região da face (cabelo, testa, olhos, nariz e boca, sempre nesta ordem). Cada imagem do conjunto de treinamento possui largura  $W$  e altura  $H$  e é dividida em blocos que se sobrepõem, de altura  $L$  e mesma largura  $W$  (figura 17a). Todos os blocos extraídos das imagens são arranjados como um vetor-coluna, consistindo de todos os *pixels* que pertencem ao bloco. Em seguida, constrói-se uma matriz covariância e, através do uso da transformada KLT, obtém-se uma nova base de vetores. Cada vetor é então projetado na nova base, formando-se assim um conjunto de vetores de observação com um número de dimensões reduzidas.

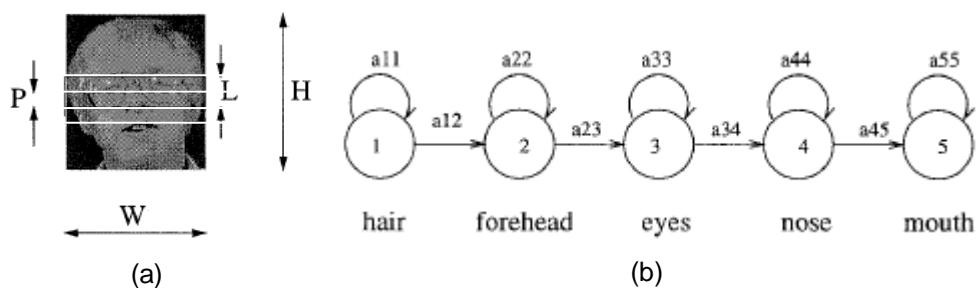


Figura 17 - HMM para a detecção e reconhecimento de faces (NEFIAN e HAYES, 1998)

A partir do conjunto de imagens de um indivíduo, extrai-se um conjunto de vetores de observação, e cada um deste é associado a um estado do HMM (figura 17b). Para que seja possível identificar uma face, independente de sua identidade, o HMM deve ser treinado com um modelo genérico de face humana obtida a partir de um grande conjunto de imagens de face. A utilização de HMM permite construir sistemas de detecção e reconhecimento robustos quanto à diferença de escala e

variações na orientação e expressões da face (NEFIAN e HAYES, 1998; SAMARIA, 1994).

## **2.7 Considerações Finais**

Neste capítulo foram apresentados os principais métodos de detecção de faces encontrados na literatura, que visam identificar a presença de faces numa imagem. Cada método preocupa-se em identificar a face ou faces presentes numa imagem com a maior taxa de acerto possível, e cada um apresenta vantagens e desvantagens. Métodos mais robustos tendem a ser de implementação mais complexa e a ter maior custo computacional, como é o caso do método de autofaces (*eigenfaces*). Métodos que utilizam a cor da pele são de implementação mais simples, mas o resultado depende do espaço de cores utilizado.



## 3 Fundamentação Teórica

---

### 3.1 Considerações Iniciais

Este capítulo descreve a base teórica utilizada nesse projeto. São apresentados os conceitos de Redes Neurais Artificiais, dando ênfase à rede MLP utilizada nesse projeto, e os modelos de cores de pele.

### 3.2 Técnicas de Detecção de Pele

A cor da pele é diferenciada pela exposição à radiação ultravioleta e inclui fatores de transmissão genética. As variações encontradas em diversos povos são relacionadas à evolução e reprodução (BECHELLI e CURBAN, 1963; BONVENTI JR e COSTA, 2003).

O problema de detecção de pele em imagens digitais tem sido pesquisado intensamente, existindo atualmente técnicas robustas para determinar os parâmetros utilizados na detecção de pele. Imagens digitais podem conter pessoas de diversas etnias e, portanto com diferentes tons de pele, que ainda podem ser alterados por

diferentes tipos de iluminação, foco e ângulos de visão variada, o que pode dificultar a detecção (RAMOS FILHO, 2006; LAURENT E BODO, 2003; CAETANO et al., 2002).

De acordo com Ramos Filho (2006) há três técnicas principais utilizadas para detecção de pele em imagens digitais, que são: limiar de classificação (regras geométricas), estatísticas paramétricas e baseadas em histogramas. Essas técnicas serão descritas nas próximas seções.

### **3.2.1 Limiar de Classificação**

As técnicas que utilizam as propriedades das cores da pele humana são técnicas que formam um agrupamento compacto em cada dimensão de espaço de cor, sendo conhecidas como técnicas baseadas em limiar de classificação. Cores baseadas na pele podem ser modeladas de várias formas utilizando um agrupamento compacto em cada dimensão do espaço de cor, agrupamento esse que pode modelado por intervalos em cada componente de cor,  $[C_{\min}^i, C_{\max}^i]$  onde  $i$  representa a dimensão do espaço de cor  $C$  (RAMOS FILHO, 2006; HSU et al, 2002).

No trabalho de Kovac et al. (2003), foi apresentada uma técnica de fácil implementação e de baixo custo computacional, que descreve os agrupamentos dos pixels, e pode ser determinada por:

$$(R > 95) \wedge (G > 40) \wedge (B > 20) \wedge (\max\{RGB\} > 15) \wedge (|R - G| > 15) \wedge (R > G) \wedge (R > B).$$

Outra técnica de limiar de classificação e que utiliza a regra de classificação foi apresentada por Hsu et al. (2002). Essa técnica baseou-se no trabalho de

Terrillon et al. (2000) que utiliza diversos espaços de cores para a detecção de pele, e que constata que espaço *rg* normalizado não é a melhor opção para detecção de pele com o objetivo de detectar face humana. O trabalho de Hsu et al. (2002) utiliza uma transformação não-linear no espaço de cor (YCbCr), visando obter um plano de crominância Cr'Cb', que seja independente do componente Y (Luminância) para o agrupamento de pele na sua base treinamento. Após obter o plano verificou-se que a projeção dos pontos de agrupamento de pele nesse novo plano forma um agrupamento bastante compacto, podendo ser descrito por uma elipse (RAMOS FILHO, 2006). A equação da elipse proposta por Hsu et al. (2002) é:

$$\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} = 1$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b - c_x \\ C_r - c_y \end{bmatrix}$$

onde  $c_x = 109,38$ ,  $c_y = 152,02$ ,  $\theta = 2,53$  radianos,  $ec_x = 1,60$ ,  $ec_y = 2,41$ ,  $a = 25,39$  e  $b = 14,03$  foram computados a partir do agrupamento de pele de  $C'_b$   $C'_r$ , onde esses valores foram obtidos a partir do experimento de Hsu et al. (2002).

### 3.2.2 Estatístico – Histograma de Cor

A técnica baseada em histograma de cor consiste em definir um modelo de cor de pele que projeta um histograma para as cores de pele e um histograma para as cores que não fazem parte do agrupamento de pele (JONES e REHG, 2002), onde os histogramas são obtidos através da marcação manual dos pixels.

As contagens dos histogramas são convertidas em probabilidades discretas,  $P(c | pele) = \frac{pele[c]}{T_p}$  e  $P(c | \overline{pele}) = \frac{n[c]}{T_n}$ , onde  $pele[c]$  indica a contagem do histograma de cores de pele associada à cor  $c$ ,  $n[c]$  indica a contagem do histograma das cores de não-pele associadas à cor  $c$ ,  $T_p$  e  $T_n$  correspondem às contagens totais obtidas dos histogramas das cores de pele e não-pele e  $P(c | pele)$  e  $P(c | \overline{pele})$  indicam as probabilidades da cor  $c$  ser pele ou não-pele (JONES e REHG, 2002; RAMOS FILHO, 2006).

Com os histogramas de cores pele e não-pele montados, o classificador de pixel de pele utiliza a equação de verossimilhança  $\frac{P(c | pele)}{P(c | \overline{pele})} \geq \Theta$ , onde  $0 \leq \Theta \leq 1$  é um limiar de classificação. O limiar  $\Theta$  pode ser ajustado de acordo com a relação de compromisso entre detecções corretas e falsos positivos. Calculando a verossimilhança de cada pixel e a comparando com o limiar  $\Theta$ , classifica-se o pixel como pele ou não pele.

### 3.2.3 Estatístico - Paramétrico

Outro modelo estatístico, além do baseado em histogramas, é o modelo paramétrico que utiliza uma distribuição de probabilidade gaussiana para modelar um conjunto de pixels manualmente marcados como pele (RAMOS FILHO, 2006). Após a modelagem gaussiana, classificar um pixel reduz-se ao cálculo da distância de Mahalanobis entre a cor do pixel  $h(s)$  e a média da distribuição. O quadrado da distância de Mahalanobis de  $x$  para  $\mu$  é definido:

$$r^2 = (x - \mu)^t \sum^{-1} (x - \mu),$$

onde  $x$  corresponde à cor do pixel e  $\mu$  o vetor de médias. Para este método foi utilizado um limiar de classificação  $\Theta$ , e montadas as curvas ROC (METZ, C. E., 1978) para comparação dos classificadores.

No trabalho de Caetano e Barone (2001), é proposta uma técnica que utiliza a mistura de gaussianas bivariadas que descrevem a projeção do agrupamento de pele nos planos HS do espaço HSV, AB do espaço CIELAB e  $Cp_1$  e  $Cp_2$  (as duas primeiras dimensões encontradas na técnica de transformação de componentes principais (TCP) (YI et al., 2003)), onde a função de distribuição de probabilidade é dada por:

$$P(c | pele) = \sum_{i=0}^k w_i P_i(c | pele),$$

onde  $w_i$  denota o peso (numero total de pixels da classe dividido pelo total de pixels da amostra) de cada uma das gaussianas que compõem a mistura (RAMOS FILHO, 2006).

### **3.3 Redes Neurais Artificiais**

O funcionamento do cérebro humano é ainda um grande mistério para o homem. O cérebro é tão poderoso que pode resolver uma grande variedade de problemas relacionados ao pensar, falar, relembrar, sentir e aprender, tendo surpreendido e inspirado muitos cientistas a tentar modelar sua operação.

O cérebro é considerado um sistema de processamento de informação altamente complexo, não-linear fazendo cálculos em paralelo. Sua estrutura e sua capacidade de organizar seus neurônios tornam possível ao ser humano realizar certas tarefas, tais como a de reconhecimento de padrões, percepção e controle

motor, muitas vezes mais rapidamente do que o mais rápido computador digital existente.

As redes neurais artificiais tentam simular o cérebro humano, modelando a maneira como o cérebro realiza uma tarefa particular ou função de interesse. A rede é normalmente implementada utilizando-se componentes eletrônicos ou é simulada por programação em um computador digital. Para alcançarem bom desempenho, as redes neurais empregam uma interligação maciça de células computacionais simples denominadas de “neurônios” ou “unidades de processamento”. Pode-se então oferecer a seguinte definição de uma rede neural vista como uma máquina adaptativa (HAYKIN, 2001):

*“Uma rede neural é um processador maciçamente paralelo e distribuído, de unidades de processamento simples, que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso. Ela se assemelha ao cérebro em dois aspectos:*

1. *O conhecimento é adquirido pela rede, a partir de seu ambiente, através de um processo de aprendizagem.*
2. *Forças de conexão entre neurônios conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.”*

O procedimento utilizado para realizar o processo de aprendizagem é chamado de algoritmo de aprendizagem, cuja função é modificar os pesos sinápticos da rede de uma forma ordenada para alcançar um objetivo de projeto desejado.

Conforme as características de cada rede neural são definidas as tarefas que podem ser executadas em sistemas computacionais, e que para os seres humanos são tarefas comuns.

Nas próximas seções serão apresentados o modelo de um neurônio, as funções de ativação, as arquiteturas de rede, o perceptron multicamadas (MLP) e os algoritmos de treinamento.

### **3.3.1      Modelo de um neurônio**

No cérebro, um neurônio é uma unidade de processamento de informação que é fundamental para a operação de uma rede neural. Cada neurônio recebe e combina sinais de muitos outros neurônios e produz sinais para o axônio conduzir certas ações, chamados de sinapses. Juntos, os neurônios formam uma grande rede, denominada de rede neural (HAYKIN, 2001).

Os axônios, que são as linhas de transmissão entre os neurônios e os dendritos, que são as zonas receptivas, constituem dois tipos de filamentos morfológicamente diferentes das células. Um axônio tem uma superfície lisa, poucos ramos, e grande comprimento, enquanto um dendrito tem uma superfície irregular e muitos ramos. A forma e o tamanho dos neurônios podem variar em diferentes partes do cérebro (figura 18).

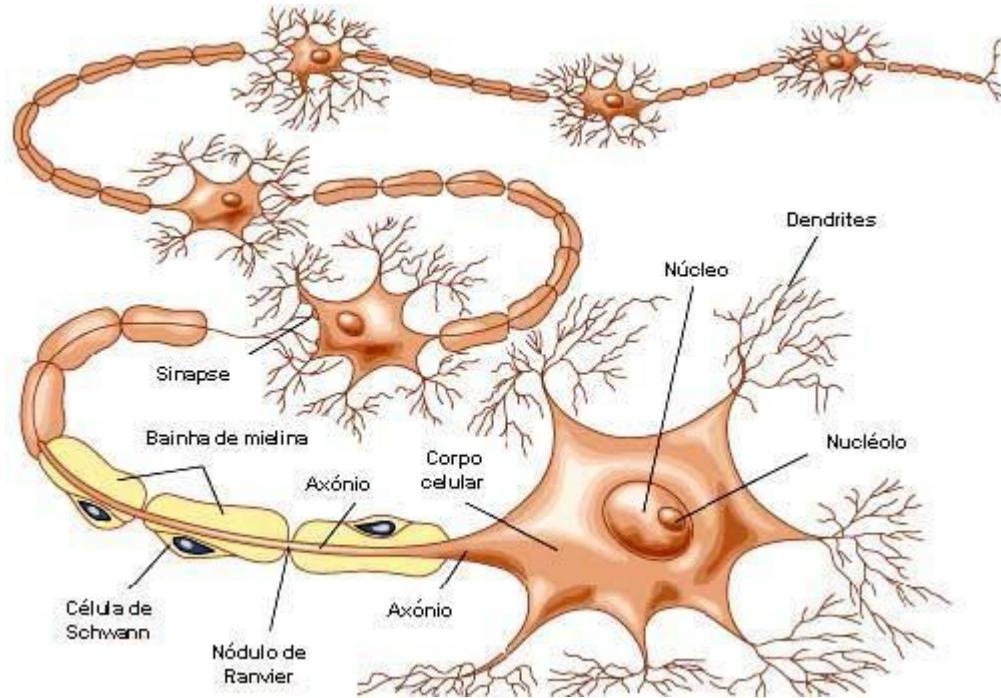


Figura 18 – Estrutura do neurônio (AMORIN, 2009)

O modelo computacional de um neurônio, inspirado nos neurônios do cérebro, é mostrado na figura 19 e constitui a base para os projetos de redes neurais artificiais. Nessa figura são identificados três elementos básicos do modelo neural (HAYKIN, 2001): conjunto de sinapses, um somador e uma função de ativação.

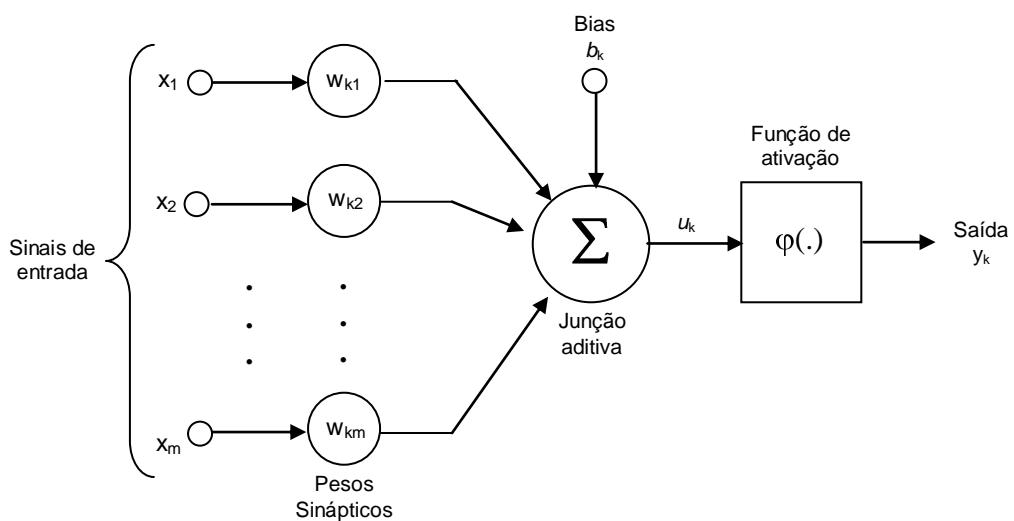


Figura 19 - Modelo não-linear de um neurônio (HAYKIN, 2001)

Nesse modelo o neurônio é formado por um conjunto de *sinapses* ou *elos da conexão*, cada uma caracterizada por um *peso* ou *força* própria. Especificamente, um sinal  $x_j$  na entrada da sinapse  $j$  conectada ao neurônio  $k$  é multiplicado pelo peso sináptico  $W_{kj}$ , onde o primeiro índice,  $k$ , refere-se ao neurônio em questão e  $j$ , o segundo, refere-se ao terminal de entrada da sinapse.

O *somador* soma os sinais de entrada, ponderados pelas respectivas sinapses do neurônio, operações que constituem um *combinador linear*.

A *função de ativação*, também conhecida como *restritiva*, limita a amplitude da saída de um neurônio.

O *bias*  $b_k$ , apresentado na figura 19, é aplicado externamente, e tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele for positivo ou negativo, respectivamente.

Pode-se descrever matematicamente um neurônio  $k$  através das equações (18) e (19):

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (6)$$

e

$$y_k = \varphi(u_k + b_k) \quad (7)$$

onde,  $u_k$  é a saída do combinador linear devido aos sinais de entrada e  $y_k$  é o sinal de saída do neurônio.

### 3.3.2 Função de ativação

A função de ativação, denotada por  $\varphi(v)$ , define o valor de saída de um neurônio em termos do nível de atividade de sua entrada (HAYKIN, 2001). Pode-se

identificar três tipos básicos de funções de ativação: função de limiar, função linear por partes e função sigmóide.

### 3.4 Arquitetura de rede

A maneira como os neurônios são organizados e interconectados define a arquitetura da rede neural. O tipo de organização está diretamente relacionado com o algoritmo de aprendizagem usado para treinar a rede. Pode-se identificar 4 tipos de arquiteturas de redes neurais:

- Redes *feedforward* unicamada (*Perceptron*)
- Redes *feedforward* multicamada
- Redes Recorrentes
- Estrutura de Lattice (Reticulada)

A arquitetura *feedforward* multicamada será descrita a seguir.

#### 3.4.1 Redes *feedforward* multicamada

Este tipo de rede é caracterizado pela presença de uma ou mais camadas ocultas, além das camadas de entrada e de saída, como ilustrado na figura 20. Os nós computacionais das camadas ocultas são chamados correspondentemente de neurônios ocultos. A função dos neurônios ocultos é intervir entre a entrada externa

e a saída de uma maneira útil. A adição de uma ou mais camadas ocultas, torna a rede capaz de extrair estatísticas de ordem elevada.

Uma rede neural é considerada totalmente conectada se cada neurônio de uma camada estiver conectado a todos os neurônios da camada subsequente, caso contrário, é considerada uma rede neural parcialmente conectada.

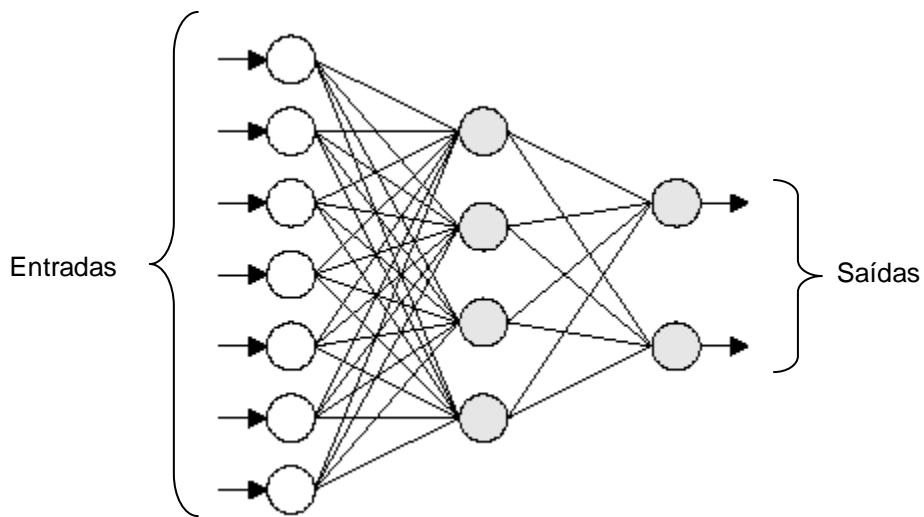


Figura 20 - Rede *feedforward* multicamada totalmente conectada com uma camada oculta e uma camada de saída.

### 3.5 Perceptron Multicamadas (MLP)

O perceptron multicamadas (*MLP, multilayer perceptron*) é uma rede do tipo *feedforward*, que contém uma ou mais camadas internas. Tem sido aplicada com sucesso, na resolução de problemas das mais diversas áreas, podendo-se citar: reconhecimento de padrões, processamento de imagens e de sinais, e previsão de séries temporais. O seu treinamento é realizado de forma supervisionada com o algoritmo de retropropagação de erro (*error back-propagation*) (HAYKIN, 2001). Este algoritmo baseia-se na regra de aprendizagem por correção de erro.

A rede perceptron multicamadas consiste de um conjunto de unidades sensoriais (nós de fonte) que constituem a camada de entrada, uma ou mais camadas ocultas de nós computacionais e uma camada de saída de nós computacionais como pode ser visto na figura 21. O vetor de entrada é representado pelas variáveis  $x_i$ , os pesos sinápticos são denotados pela letra  $w_k$ , onde  $k$  indica a camada, e a saída da rede é representada por  $y$ .

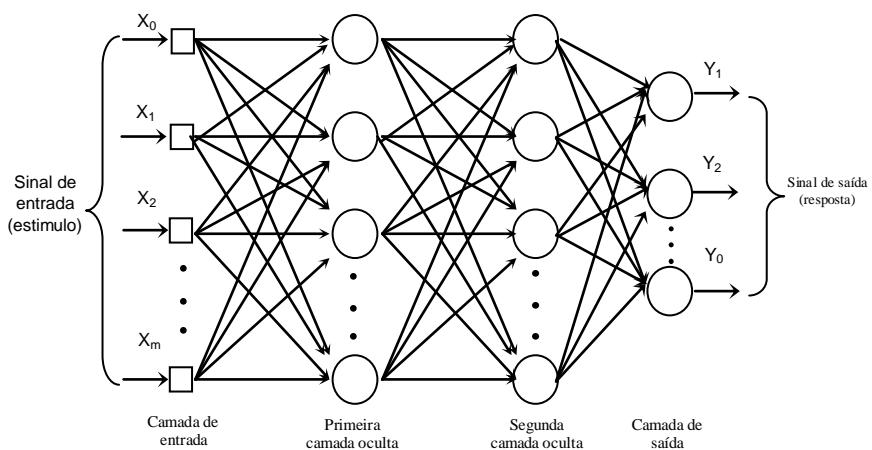


Figura 21 - Grafo arquitetural de um perceptron múltiplas camadas com duas camadas ocultas

O sinal de entrada propaga-se para frente através da rede, camada por camada, como ilustrado na figura 22, e dois tipos de sinais são encontrados nesta rede.

**1. Sinais Funcionais.** São estímulos recebidos pelos sinais de entrada (estímulo), que se propagam para frente (neurônio por neurônio) através da rede e manifestam-se no terminal de saída como um sinal de saída;

**2. Sinais de Erro.** O neurônio de saída gera um sinal de erro e se propaga para trás (camada por camada) através da rede.

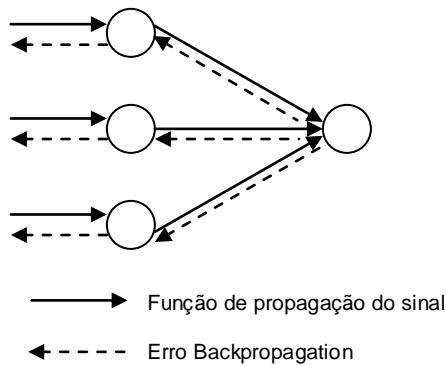


Figura 22 – Direção da propagação da função sinal

As camadas ocultas são referidas desta maneira, porque não fazem parte da saída e nem da entrada da rede, sendo a primeira camada oculta alimentada pela camada de entrada, e constituída por várias unidades sensoriais; as saídas da primeira camada são aplicadas às entradas da próxima camada oculta e assim por diante para o resto da rede (figura 21).

Essa rede apresenta três características principais:

1. O modelo de cada neurônio da rede inclui uma função de ativação não-linear. Uma forma normalmente utilizada de não-linearidade que satisfaz esta exigência é uma não-linearidade sigmóide definida pela função logística:

$$y_j = \frac{1}{1 + \exp(v_j)} \quad (8)$$

onde  $v_j$  é a soma ponderada de todas as entradas sinápticas do neurônio  $j$ ,  $y_j$  é a saída do neurônio. A figura 23 esboça o gráfico da função sigmóide, onde essa função é sempre positiva e limitada superiormente por 1 e inferiormente por 0.

2. A rede é composta de uma ou mais camadas de neurônios ocultos,

que não são parte da entrada ou da saída da rede. Estes neurônios capacitam a rede a aprender tarefas complexas extraíndo progressivamente as características mais significativas dos padrões (vetores) de entrada.

3. A rede possui um alto grau de conectividade, devido às suas sinapses. Uma modificação na conectividade da rede requer uma mudança na população das conexões sinápticas ou de seus pesos.

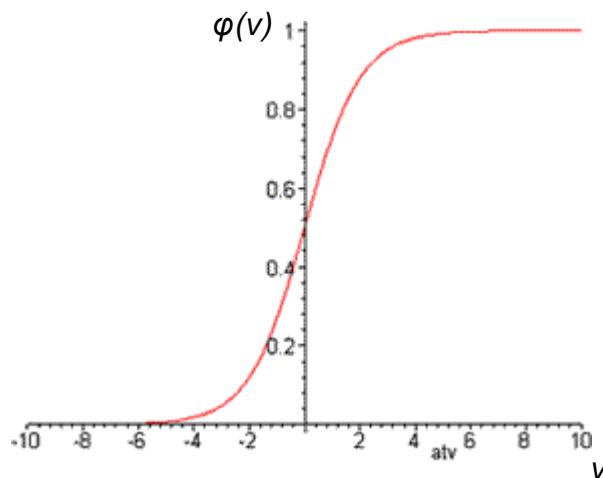


Figura 23 – Gráfico da função sigmóide

### 3.5.1 Algoritmo de Treinamento

A MLP utiliza para o seu treinamento o algoritmo de retropropagação ou *backpropagation*. Este algoritmo é constituído por dois passos principais: o passo para frente (*forward*) e o passo trás (*backward*). No passo *forward* é aplicado à rede um padrão de dados (vetor de entrada) às entradas da rede e este efeito se propaga através da rede, camada por camada.

Durante o processo de propagação, todos os pesos sinápticos da rede não se alteram. No passo *backward* os pesos sinápticos são todos ajustados de acordo com a regra de correção de erro. Contudo, a resposta gerada pela rede é subtraída da resposta desejada (alvo) para produzir um sinal de erro. Os pesos sinápticos da rede são atualizados para que a resposta gerada pela rede aproxime-se da resposta desejada (HAYKIN, 2001). Este processo é repetido diversas vezes, até que para todas as saídas e padrões de treinamento, o erro seja menor do que o especificado.

O principal objetivo do processo de treinamento é minimizar o sinal de erro entre a resposta desejada (alvo) e a atual resposta produzida pela rede. Este sinal de erro da saída do neurônio  $y_i$  na interação  $n$ , é definido por:

$$e_i(n) = d_i(n) - y_i(n) \quad (9)$$

Onde,  $d_i$  é a resposta desejada do  $i$ -th neurônio de saída.

Adicionando todos os erros quadráticos produzidos pelos neurônios de saída da rede com a respectiva  $n$ -th iteração, obtém-se a seguinte função custo ou o erro para ser minimizado:

$$E(\mathbf{w}) = \frac{1}{2} \sum_i e_i^2 = \frac{1}{2} \sum_i (d_i - y_i)^2, \quad (10)$$

A função custo é uma função continuamente diferenciada para cada peso, onde, um algoritmo do gradiente descendente pode ser usado para ajustar os pesos. Para as conexões  $w_{ij}$ , que ligam a camada oculta e a camada de saída, a regra do gradiente descendente é dada por:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta \cdot \delta_i \cdot v_j \quad (11)$$

onde,  $\eta$  é a taxa de aprendizagem do algoritmo de *backpropagation* e  $\delta_i$  é o gradiente local definido por:

$$\delta_i = \phi'(c_i) \cdot (d_i - y_i) \quad (12)$$

Onde,  $c_i$  é o sinal de entrada dos pesos aplicado para o neurônio  $y_i$ , e  $\phi(.)$  é a função de ativação do neurônio. As conexões  $w_{ij}$  são iterativamente atualizadas aplicando a seguinte equação:

$$w_{ij}^{novo} = w_{ij}^{velho} + \Delta w_{ij} \quad (13)$$

Para as conexões  $w_{jk}$ , que ligam a camada de entrada e a camada oculta, a função custo é diferenciada com o respectivo  $w_{jk}$  usando a regra da cadeia, que é dada por:

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} = -\eta \cdot \sum \frac{\partial E}{\partial v_j} \cdot \frac{\partial v_j}{\partial w_{jk}} = \eta \cdot \delta_j \cdot x_k \quad (14)$$

Na equação 14, o gradiente local  $\delta_j$  em relação ao  $w_{jk}$  é expresso pela seguinte equação:

$$\delta_j = \phi'(h_i) \sum w_{ij} \cdot \delta_i \quad (15)$$

onde,  $h_i$  é o peso do sinal de entrada aplicado para o neurônio  $v_j$ . Finalmente, todas as conexões  $w_{jk}$  são iterativamente atualizadas de acordo com a seguinte equação:

$$w_{jk}^{novo} = w_{jk}^{velho} + \Delta w_{jk} \quad (16)$$

Após o processo de treinamento, a rede é capaz de fornecer a relação

funcional entre as variáveis de entrada e saída do processo.

Existem algumas técnicas que auxiliam a melhorar o desempenho do treinamento da MLP:

- **Validação cruzada:** metodologia utilizada para testar o poder de generalização de uma MLP. Neste contexto, devem-se seguir os passos seguintes:
  - Particionar o conjunto de dados disponível em dois subconjuntos:
    - Conjunto de treinamento: utilizado para treinar a rede (80% a 90% do total);
    - Conjunto de teste: utilizado para avaliar se a rede está generalizando de forma satisfatória (10% a 20% do total).
  - Utilizar o mesmo conjunto de treinamento para ajustar todas as topologias candidatas;
  - Após o treinamento, utilizar o conjunto de teste para avaliar e escolher (validar) o melhor modelo (melhor generalização):
    - Em certos casos, uma determinada topologia pode ter desempenho satisfatório no treinamento, mas durante a fase de validação (aplicação do conjunto teste), a rede produz resultados insatisfatórios (não consegue generalizar);
    - O aumento de neurônios e de camadas não significa que a rede irá generalizar melhor;
    - Para duas topologias que estão generalizando com o mesmo grau de precisão, deve-se optar por aquela com menor número de neurônios.

- **Inserção do termo de momento:** a velocidade do algoritmo *backpropagation* pode ser aumentada (sem perigo de instabilidade) através da inclusão de um termo de momento  $\alpha$  utilizado para acelerar o algoritmo, como mostrada na equação 17:

$$\Delta w_{ji}(t+1) = w_{ji}(t) + \alpha(w_{ji}(t) - \Delta w_{ji}(t-1)) + \eta * d_j * y_i \quad (17)$$

- **Normalização dos dados:** as variáveis referentes às entradas da rede devem ser normalizadas para a faixa [0,1] se estiver sendo utilizada a função sigmóide, ou então, para [-1,1] se for utilizada a tangente hiperbólica.

### 3.6 Algoritmo de Aprendizagem

A rede neural se destaca pela sua capacidade de aprender, e isto é uma característica primordial de sua habilidade. As mudanças que ocorrem na estrutura de interconexão ou nos valores dos pesos sinápticos geram alterações no comportamento de ativação da rede e permite que seja realizado um novo aprendizado de um determinado comportamento. Desta forma, pode-se modificar o estado de ativação na saída da rede em resposta a certo “estímulo” recebido através do padrão de entrada. Contudo, a rede é capaz de estabelecer associações de entrada e saída (estímulo e resposta), a fim de se adaptar a uma situação proposta. No caso de uma rede com aprendizado supervisionado, a rede deve adaptar seus pesos de maneira a passar a responder de acordo com a resposta desejada. O método utilizado para modificar o comportamento de uma rede é conhecido como regra de aprendizado.

A definição de processo de aprendizagem em redes neurais pode ser definida da seguinte maneira:

*“Aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira através da qual a modificação dos parâmetros ocorre.” (HAYKIN, 2001).*

Um conjunto de regras bem definidas para a solução de um problema de aprendizagem é denominado um algoritmo de aprendizagem.

Esta definição de aprendizagem implica nos seguintes eventos:

1. A rede neural é **estimulada** por um ambiente;
2. A rede neural **sofre modificações** nos seus parâmetros, como resultado das estimulações;
3. A rede neural **responde de uma maneira nova** ao ambiente, devido às modificações sofridas em sua estrutura.

Com um algoritmo de aprendizagem padrão uma rede neural aprende através de um processamento iterativo do ajuste dos pesos. O tipo de aprendizagem é definido pela forma como os pesos são ajustados. Os dois principais métodos de aprendizagem são: aprendizagem supervisionada e aprendizagem não-supervisionada.

### 3.6.1 Aprendizagem Supervisionada

Essa regra de aprendizado incorpora o conceito de um supervisor ou um professor, que tem o conhecimento sobre o ambiente em que a rede está operando. Este conhecimento é representado por um conjunto de exemplos de entrada-saída ou padrões. Sendo assim, a rede deve ser capaz de medir a diferença entre seu comportamento atual e o comportamento de referência e então corrigir os pesos de maneira a reduzir este erro. O aprendizado supervisionado utiliza conhecimentos empíricos, que são representados por um conjunto de exemplos rotulados, ou seja, exemplos com pares de dados de entrada com a respectiva saída associada  $\{x_i, d_i\}$ .

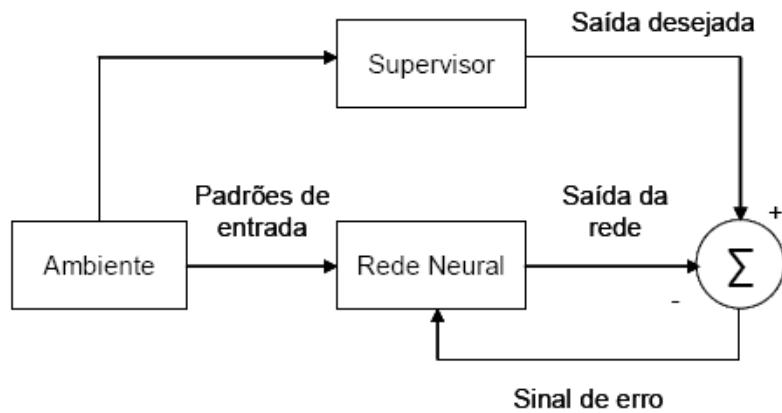


Figura 24 - Diagrama em blocos da aprendizagem supervisionada

A rede ajusta os parâmetros livres através da combinação das entradas e sinais de erro, sendo o sinal de erro a diferença entre a resposta desejada e a resposta real da rede (figura 24). Este ajuste é realizado passo a passo com o objetivo de fazer a rede “emular” um professor. Assim, se a emulação for ótima, o conhecimento do ambiente disponível ao professor é transferido para a rede neural através do treinamento. Quando esta condição é alcançada, pode-se liberar

o professor e deixar a rede neural lidar com o ambiente (HAYKIN, 2001).

### **3.6.2 Considerações finais**

As redes neurais artificiais constituem um campo da ciência da computação ligado à inteligência artificial, buscando implementar modelos matemáticos que se assemelhem às estruturas neurais biológicas. Elas apresentam grande capacidade de adaptar os seus parâmetros conforme resultados obtidos da interação com o meio externo, melhorando gradativamente o seu desempenho na solução de um determinado problema. Como mencionado neste capítulo, a sua aplicação tem se verificado nas mais diversas áreas.

A detecção de faces em imagens não é uma tarefa simples devido à complexidade do fundo, cor da pele, objetos presentes na face, expressões faciais, problemas de iluminação.



## **4 Detecção de Faces Humanas em Imagens Coloridas - Metodologia**

---

### **4.1 Considerações Iniciais**

O objetivo deste trabalho foi o desenvolvimento de um algoritmo baseado em redes neurais artificiais, para a detecção de faces humanas em imagens digitais coloridas, utilizando segmentação de pele. Neste capítulo são apresentados os módulos implementados e a metodologia de desenvolvimento de cada módulo.

### **4.2 Módulos Desenvolvidos**

O trabalho foi dividido em dois módulos principais, o de segmentação de pele e o de detecção de face. O módulo de segmentação de pele foi subdividido em quatro módulos menores, sendo eles: pré-processamento, segmentação de pele, varredura da imagem, e pós-processamento. E o módulo de detecção de face foi subdividido em dois módulos: extração de características e o módulo de detecção e

interpretação da face. A figura 25 ilustra a seqüência de processamento desses módulos e a figura 26 mostra o algoritmo do trabalho desenvolvido. Os seis módulos estão relacionados a seguir:

1. Pré-processamento, aplicado para redução de ruído e melhoria da qualidade da imagem de entrada;
2. Segmentação de Pele, aplicada para retirar da imagem todas as regiões detectadas como não pele, utilizando para isso uma rede neural MLP;
3. Varredura da Imagem, onde é realizada uma varredura em toda a imagem, que irá dividir a imagem de entrada em diversas imagens menores (regiões) sobre as quais serão executados os próximos módulos, inclusive o módulo de verificação, que determina se existe face ou não em cada região. Essa varredura é realizada utilizando um bloco com dimensão inicial de 80x80 *pixels*, e que é incrementado de 20x20 *pixels* até atingir um tamanho superior a qualquer dimensão da imagem;
4. Pós-processamento da imagem, utilizada para destacar as características relevantes para a detecção das faces, e inclui a detecção das bordas, realizada nas áreas que contem pele e que será utilizada no processo de extração de características;
5. Extração de características, onde são aplicados métodos de extração de características para cada forma localizada na etapa anterior;

6. Detecção e interpretação da face, que faz uso das características detectadas para cada forma, as quais são aplicadas em uma rede neural MLP que identificará se a forma sob análise é face ou não.

A metodologia utilizada em cada um desses módulos será detalhada nas seções seguintes.

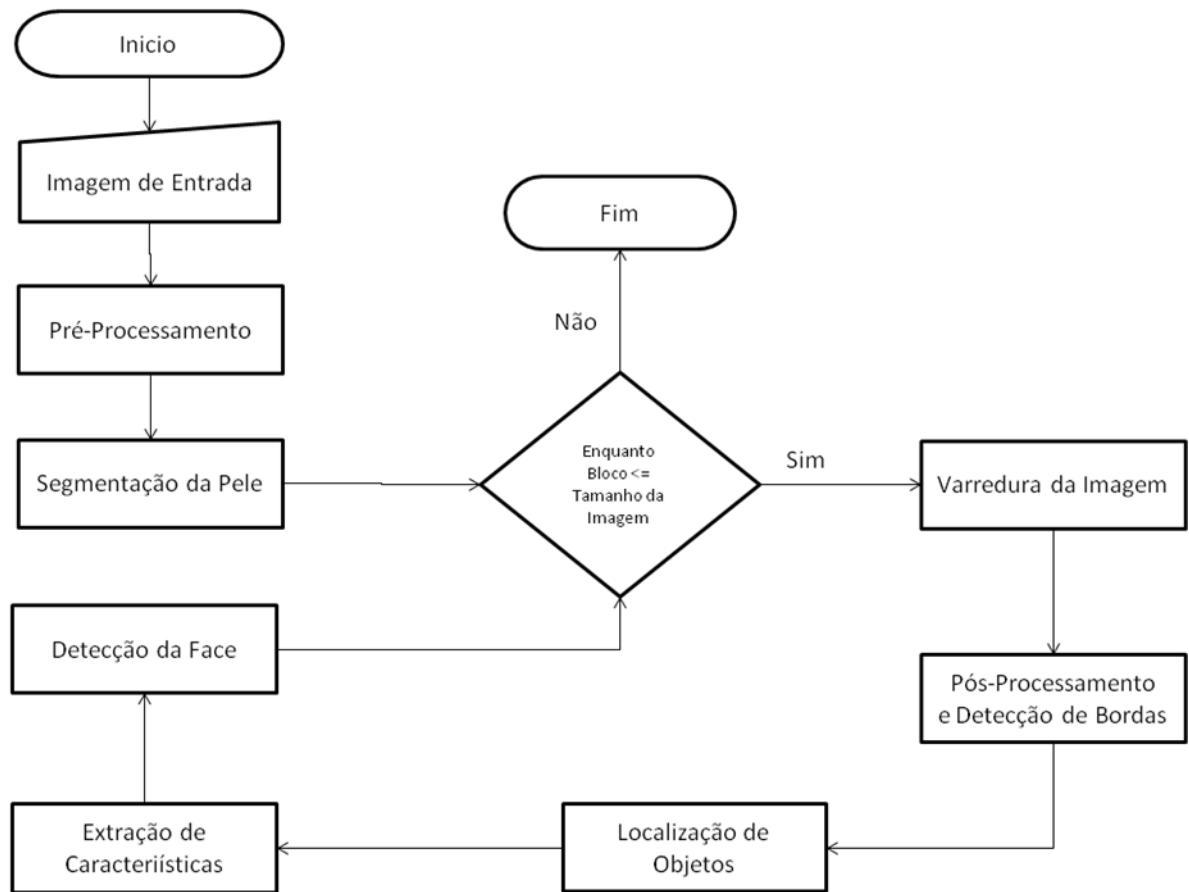


Figura 25 – Fluxograma do trabalho desenvolvido

```

    Inicio
        Im = Imagem
        Tam_Imagen = Tamanho(Im)
        Im = Pre_Processamento(Im)
        Im = Segmentacao_Pele(Im)
        Mascara = 80x80
        Enquanto (Mascara <= Tam_Imagen) faça
            Im2 = Varredura_Imagen(Im)
            Im3 = Im2
            Im2 = Pos_Processamento(Im2)
            Im2 = Deteccao_Bordas(Im2)
            Formas = Deteccao_Formas(Im2)
            Caracteristicas = Extracao_Caracteristica(Formas)
            Face = Detecao_Face(Caracteristicas)
            Se (Face) então
                Grava(Im3)
            Fim-Se
            Mascara = Mascara + 20x20
        Fim-Enquanto
    Fim_Algoritmo

```

Figura 26 – Algoritmo do trabalho desenvolvido

#### 4.3 Metodologia

Como imagens de entrada foram utilizadas dois tipos de imagens:

- Imagens com extensão JPEG, BMP e GIF, adquiridas de diferentes páginas da internet (figura 27a) com nenhuma ou várias faces, totalizando 25 imagens sendo que há um total de 80 faces nas 25 imagens;
- Imagens do banco de faces AR Face Database (MARTINEZ e BENAVENTE, 1998) (figura 27b), totalizando 25 imagens.

Todas as imagens são coloridas, contendo uma ou mais faces, com fundo de complexidade variável, com dimensões variadas e sem nenhum controle de luminosidade, como pode ser visto na figura 28.

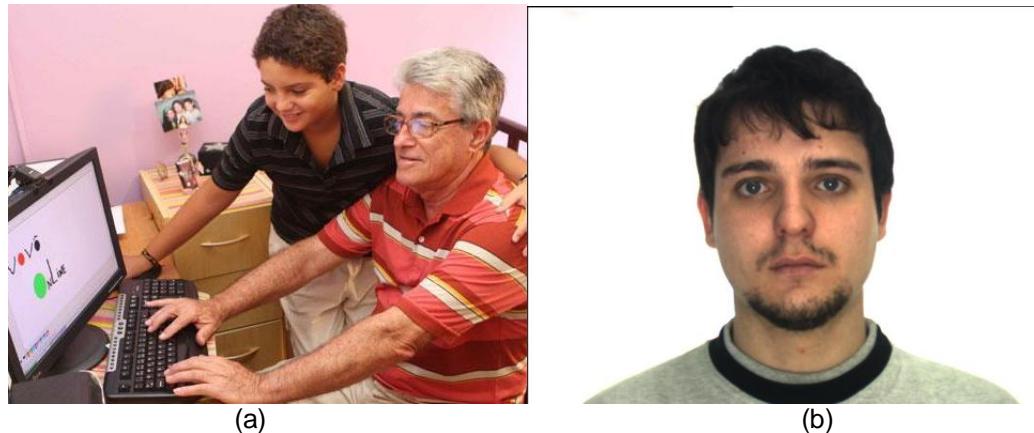


Figura 27 – Exemplo de imagens (a) adquirida na internet e (b) banco AR

É importante mencionar que nas imagens obtidas da internet houve a preocupação de se escolher imagens contendo pessoas de diferentes raças, para avaliação do módulo de segmentação de pele.

As imagens adquiridas incluem faces com diferentes ângulos de inclinação e variações de expressão, além de faces com barba, bigode e óculos, que são componentes que podem dificultar a detecção da face (figura 29).

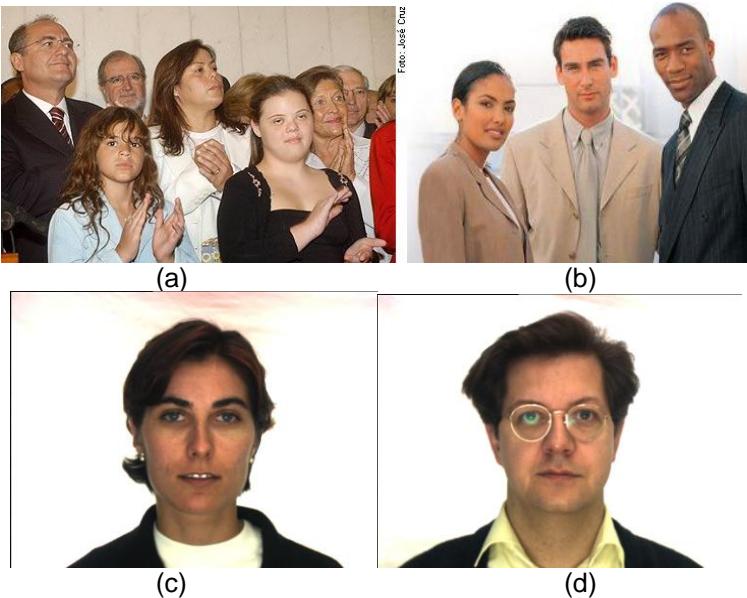


Figura 28 – Exemplo de imagens utilizadas para teste, (a) e (b) adquiridas na internet (c) e (d) Banco AR.



Figura 29 – Exemplo de imagens utilizadas com características que dificultam a detecção

#### 4.3.1     Pré-processamento

O pré-processamento é uma etapa muito importante para a segmentação da imagem e a sua função é melhorar a qualidade da imagem reduzindo a quantidade de ruídos.

Para a implementação dessa etapa foi utilizada a filtragem por mediana  $3 \times 3$ , que minimiza o ruído, mas preserva as bordas (GONZALES e WOODS, 2000). A figura 30 ilustra a aplicação desse filtro sobre uma imagem.

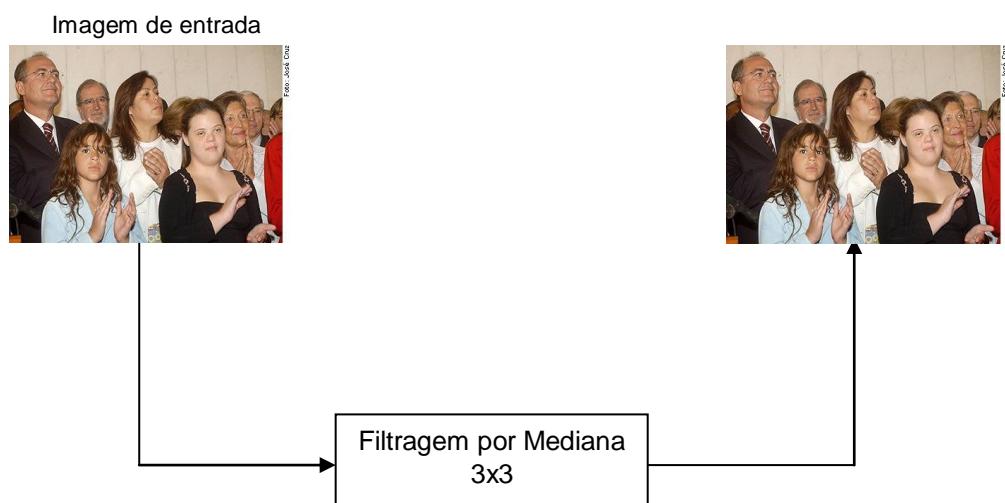


Figura 30 – Diagrama de blocos do pré-processamento da imagem

### 4.3.2 Segmentação de pele utilizando redes neurais

A etapa de segmentação de pele consiste em separar as regiões que contêm pele das demais informações da imagem.

Para a implementação dessa etapa, foi adotada a técnica de limiar de classificação de cores pele, utilizando o espaço de cores RGB, pela sua simplicidade de implementação. Inicialmente, os espaços de cores HSV e CyCbCr foram avaliados, mas o RGB foi o que apresentou melhores resultados no algoritmo implementado.

Na segmentação foi utilizada uma rede neural MLP com o algoritmo de aprendizado *backpropagation* (GOUVEIA et al., 2005; HAYKIN, 2001). A topologia dessa rede, ilustrada na figura 31, é a seguinte:

- uma camada com três entradas, onde são aplicados os valores RGB de cada *pixel* da imagem;
- duas camadas escondidas com cinco neurônios cada;
- uma camada de saída com um neurônio.

A função de ativação utilizada foi a sigmóide e o valor retornado pela função foi normalizado, de forma que valores superiores ou iguais a 0.5 são considerados pele. Nesse caso o valor do *pixel* na imagem não é alterado. No caso de valores inferiores a 0.5 é atribuída a cor preta para o pixel, como ilustrado na figura 32.

O treinamento da rede neural foi realizado utilizando a aprendizagem supervisionada, onde o conjunto de treinamento utilizado foi elaborado utilizando os padrões correspondentes aos *pixels* das regiões que contém pele e não pele. Esse conjunto foi constituído pelas 25 imagens da internet nas quais foram utilizados 25 fundos para treinamento e 25 tipos de pele. Nas imagens do Banco AR as cores do

fundo são padronizadas e a tonalidade da pele é semelhante entre as imagens, variando a iluminação.

Na figura 32 é apresentado o diagrama de blocos dessa etapa, ilustrando a imagem de entrada e a imagem resultante contendo somente as regiões de pele e com cores semelhantes à pele, o que simplifica as etapas seguintes do algoritmo.

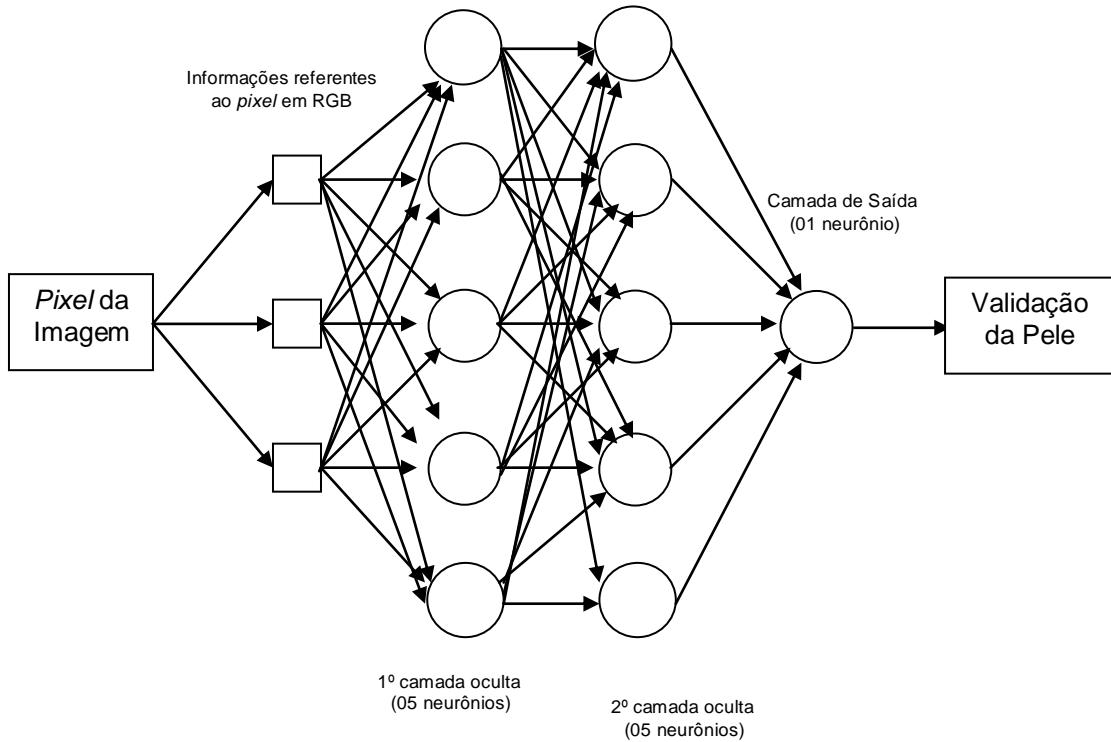


Figura 31 – Topologia da Rede Neural proposta para a segmentação de pele

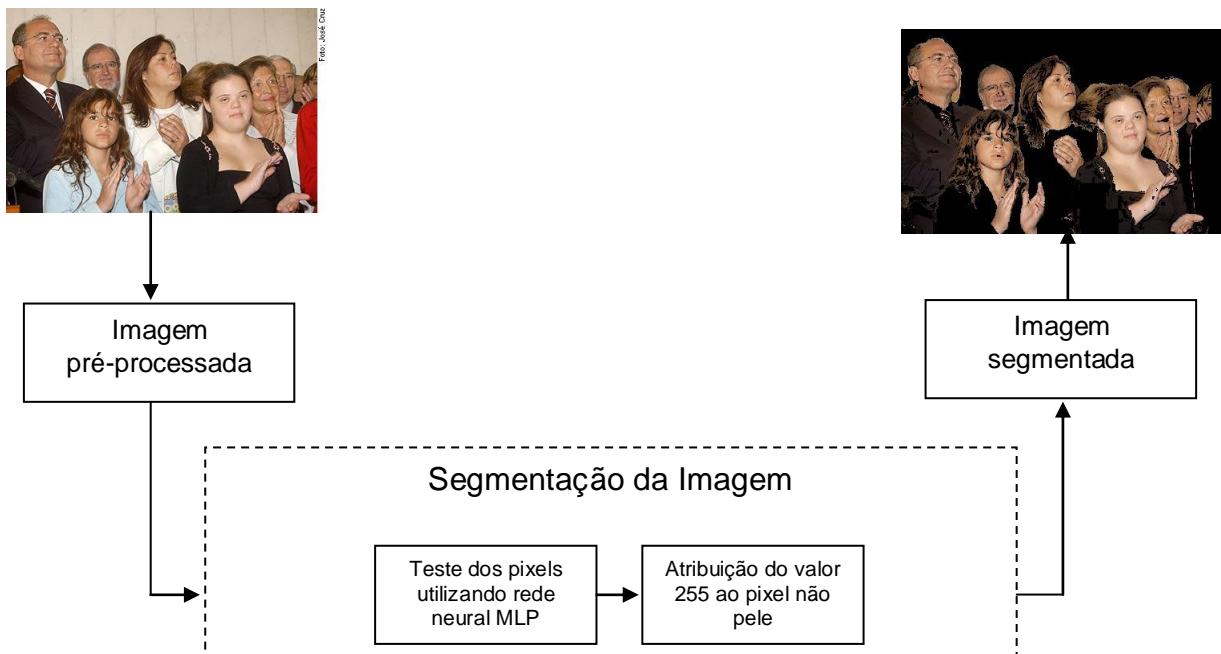


Figura 32 – Diagrama de blocos da etapa de segmentação da imagem

### 4.3.3 Varredura

Esta etapa é responsável pela divisão da imagem em pequenas partes que irão ser processadas e testadas para verificação se existe ou não face. Para isso utiliza-se uma máscara inicial de 80x80 pixels que irá varrer toda a imagem como pode ser visto na figura 33 e detalhada na figura 34. Cada região capturada por essa máscara é redimensionada em uma imagem de 100x100 pixels que será utilizada na execução dos próximos módulos (figura 35). Após essa varredura a máscara incrementa seu tamanho em 20x20 pixels e varre toda a imagem novamente, sendo esse procedimento realizado sucessivamente até o que as dimensões da máscara atinjam o tamanho da imagem.

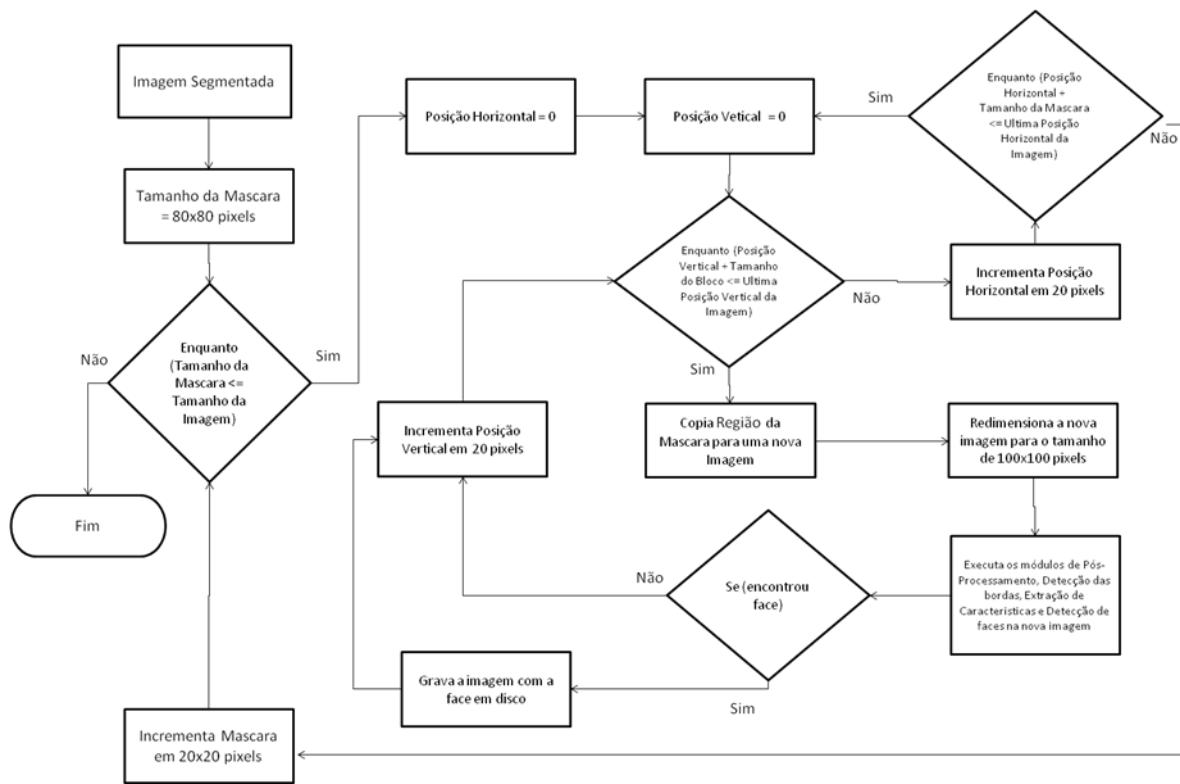


Figura 33 – Fluxograma do método de varredura implementado

```

Inicio
    Im = Imagem_segmentada
    Tam_Imagen = Tamanho(Im)
    Tam_Mascara = 80x80
    Enquanto (Tam_Mascara <= Tam_Imagen) então
        Pos_Horizontal = 0
        Enquanto (Pos_Horizontal + Tam_Mascara(Horizontal) <= Tam_Imagen(Vertical)) faça
            Pos_Vertical = 0
            Enquanto ((Pos_Vertical + Tam_Mascara(Vertical)) <= Tam_Imagen(Vertical)) faça
                Nova_Imagen = Im[Posicao Mascara]
                Nova_Imagen = Redimensiona(Nova_Imagen, 100x100)
                Pos_Processamento(Nova_Imagen)
                Binarizacao(Nova_Imagen)
                Deteccao_Bordas(Nova_Imagen)
                Extracao_Caracteristicas(Nova_Imagen)
                Detecao_Face(Nova_Imagen,Encontrou_Face)
                Se (Encontrou_Face) então
                    Grava(Nova_Imagen)
                Fim-se
                Pos_Vertical = Pos_Vertical + 20
            Fim-Enquanto
            Pos_Horizontal = Pos_Horizontal + 20
        Fim-Enquanto
        Tam_Mascara = Tam_Mascara + 20x20
    Fim-Enquanto
Fim-Algoritmo

```

Figura 34 – Algoritmo do método de varredura implementado



Figura 35 – Exemplo de imagem apóis varredura

#### 4.3.4 Pós-processamento

Na etapa de pós-processamento a imagem é binarizada utilizando o método de Otsu (GONZALES e WOODS, 2000) e na seqüência são aplicados filtros de dilatação, para corrigir pequenas falhas que tenham ocorrido na etapa de segmentação de pele (figura 36). Além disso, são aplicados filtros de aguçamento para realçar detalhes finos ou detalhes que tenham sido borrados nos passos

anteriores, e por fim para a detecção de bordas foi utilizando o operador Sobel na vertical e horizontal, cujas máscaras são apresentadas na figura 37.

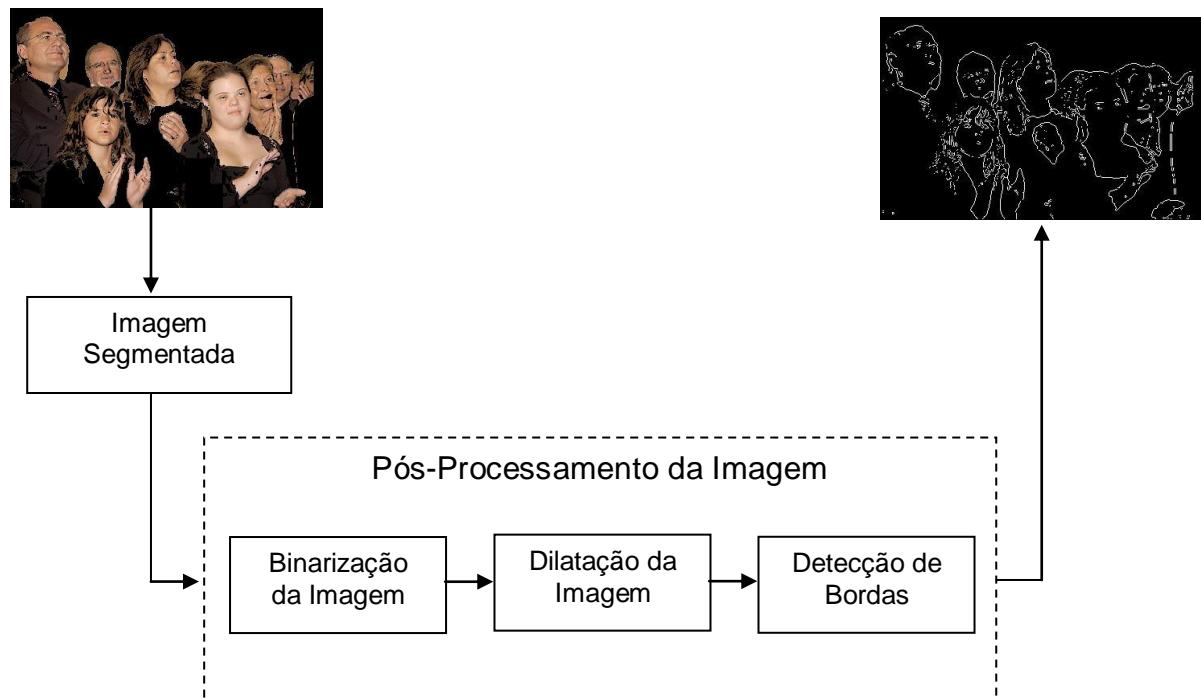


Figura 36 – Diagrama de blocos da etapa de pós-processamento da imagem

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$$

Figura 37 – Máscaras 3x3, vertical e horizontal do operador de Sobel

Um exemplo de pós-processamento é ilustrado na figura 38.



Figura 38 – Exemplo de imagem: (a) original, (b) segmentada, (c) binarizada e com (d) bordas

#### 4.3.5 Extração de características

A etapa de extração de características é muito importante no processo de detecção de face, pois reduz a quantidade de dados de entrada para o processo de detecção (BIANCHI, 2006).

Nessa etapa são extraídas características da imagem resultante dos processos anteriores, onde essas características são a presença dos olhos, do nariz e da boca, e que serão utilizadas na próxima etapa para determinar se a forma encontrada é uma face ou não. Como entrada é utilizada uma imagem, mas a saída do processo é um conjunto de dados correspondente à imagem de entrada, como é mostrado na figura 39.

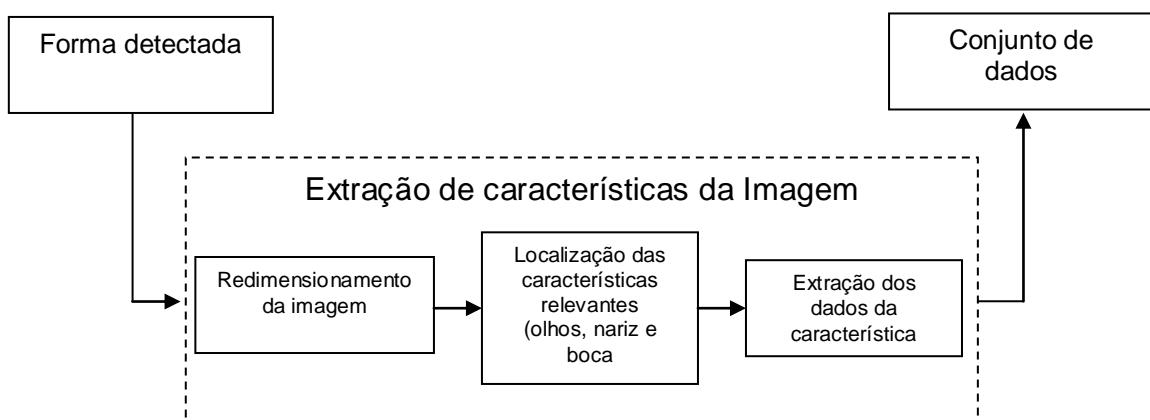


Figura 39 – Diagrama de blocos da etapa de extração de características da imagem

A extração de características é realizada nas imagens resultantes da etapa de varredura, anteriormente redimensionadas para o tamanho de 100x100 *pixels*, através de um algoritmo de varredura que irá separar e armazenar em uma matriz as coordenadas limites de todas as formas encontradas. O algoritmo faz uso de molde de dimensão 100x100 *pixels* para determinar se há formas existentes em regiões específicas da imagem que possivelmente referem-se à presença dos olhos, nariz e

boca conforme a figura 40. Essas características serão utilizadas na próxima etapa onde será analisado se a imagem é uma face ou não.

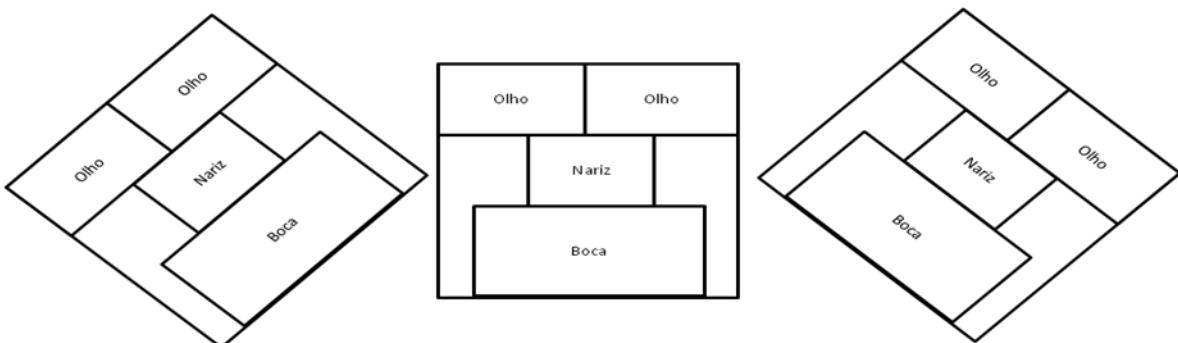


Figura 40 – Modelo do molde utilizado para extração de características

#### 4.3.6 Detecção de faces utilizando redes neurais

A utilização de redes neurais é muito importante no processo de reconhecimento de padrões e classificação. O processo de detecção de face apresentado neste trabalho consiste na utilização de uma rede neural MLP com o algoritmo de aprendizado *backpropagation*, que irá classificar, através do conjunto de dados definido pela etapa de extração de características, se a imagem selecionada é uma face ou uma não-face.

A topologia da rede MLP utilizada é a seguinte (figura 41):

- uma camada com três entradas que indicam a presença ou não dos olhos (quantidade), nariz e boca;
- duas camadas escondidas com três neurônios cada;
- uma camada de saída contendo um neurônio.

A função de ativação utilizada foi a sigmóide e o valor retornado pela função foi normalizado, de forma que valores superiores ou iguais a 0.5 são considerados face.

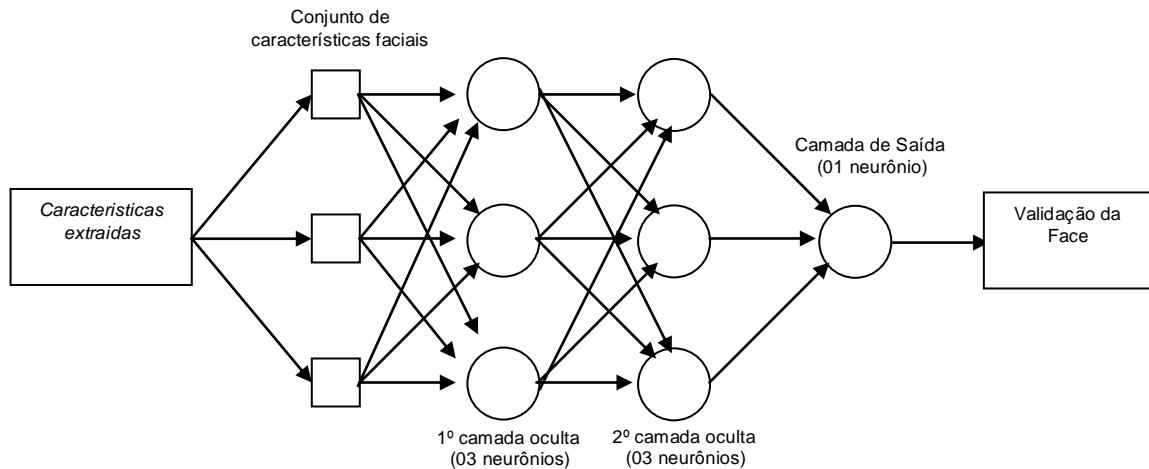


Figura 41 – Topologia da Rede Neural proposta para a detecção de face

O treinamento da rede neural para validar se é face ou não-face, foi realizado utilizando algoritmos de aprendizagem supervisionada, onde o conjunto de teste utilizado foi elaborado utilizando um total de dez vetores de características extraídas de um conjunto de imagens contendo faces e de um conjunto de não-face.

#### 4.3.7 Software de desenvolvimento

Para o desenvolvimento do algoritmo, foi utilizado o MATLAB (*Matrix Laboratory*) versão 7.1, com o *toolbox* de processamento de imagens e redes neurais. O MATLAB é um ambiente de programação que utiliza uma linguagem própria e simplifica a elaboração de trabalhos científicos. Com o auxílio dos *toolbox* que contém funções previamente implementadas, é possível obter uma maior agilidade no processo de desenvolvimento do algoritmo e na análise dos resultados obtidos.

## 5 Resultados e Discussões

---

---

Neste capítulo serão apresentados os resultados obtidos e discussões para o software de detecção de face desenvolvido.

### 5.1 Resultados obtidos com o algoritmo de segmentação

Para os testes de segmentação foram utilizados dois bancos de imagens já mencionados no capítulo 4.

As figuras 40(b) e (d) mostram as imagens resultantes da aplicação do algoritmo de segmentação nas imagens originais da internet apresentadas nas figuras 40(a) e (c), respectivamente. As figuras 41(b) e (d) mostram as imagens resultantes da aplicação do algoritmo de segmentação nas imagens originais do banco AR apresentadas nas figuras 41(a) e (c) respectivamente. Observa-se que as imagens resultantes apresentam as regiões que contêm pele, removendo-se praticamente todo o fundo da imagem.

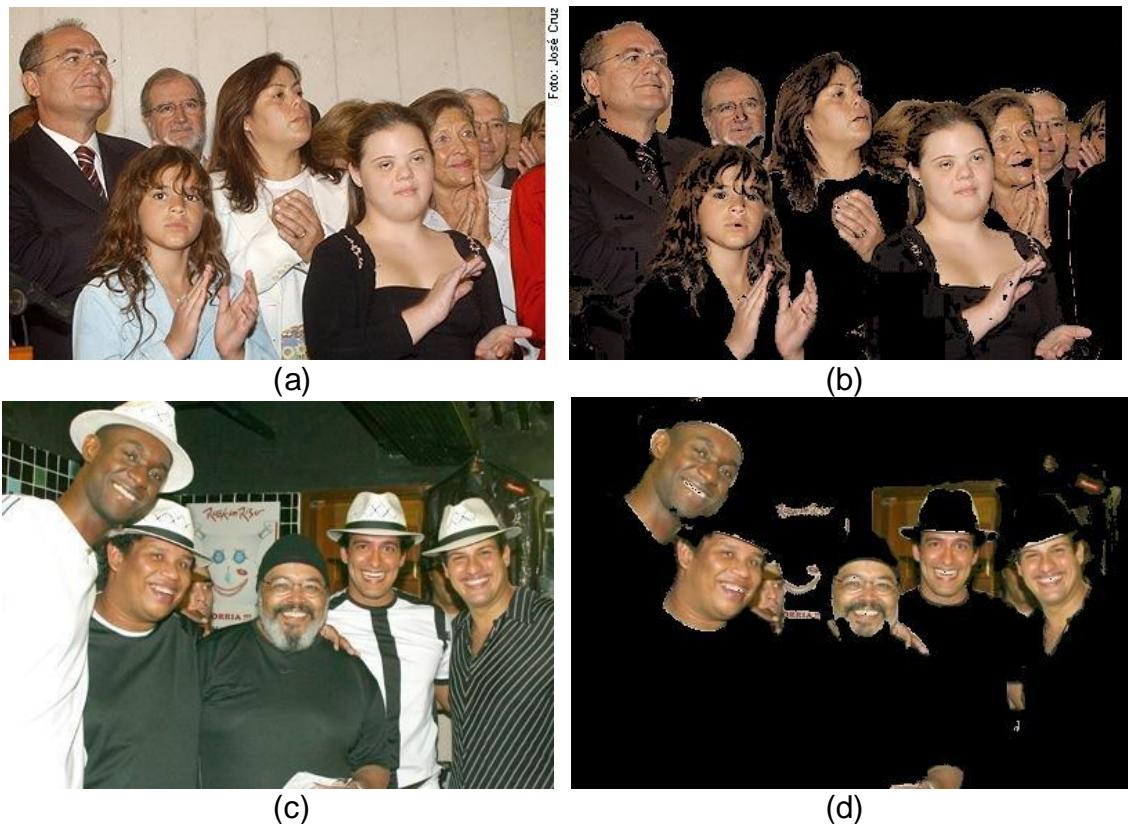


Figura 40 - (a), (c) e (e) imagens originais, (b), (d) e (f) imagens processadas do banco de imagens da internet

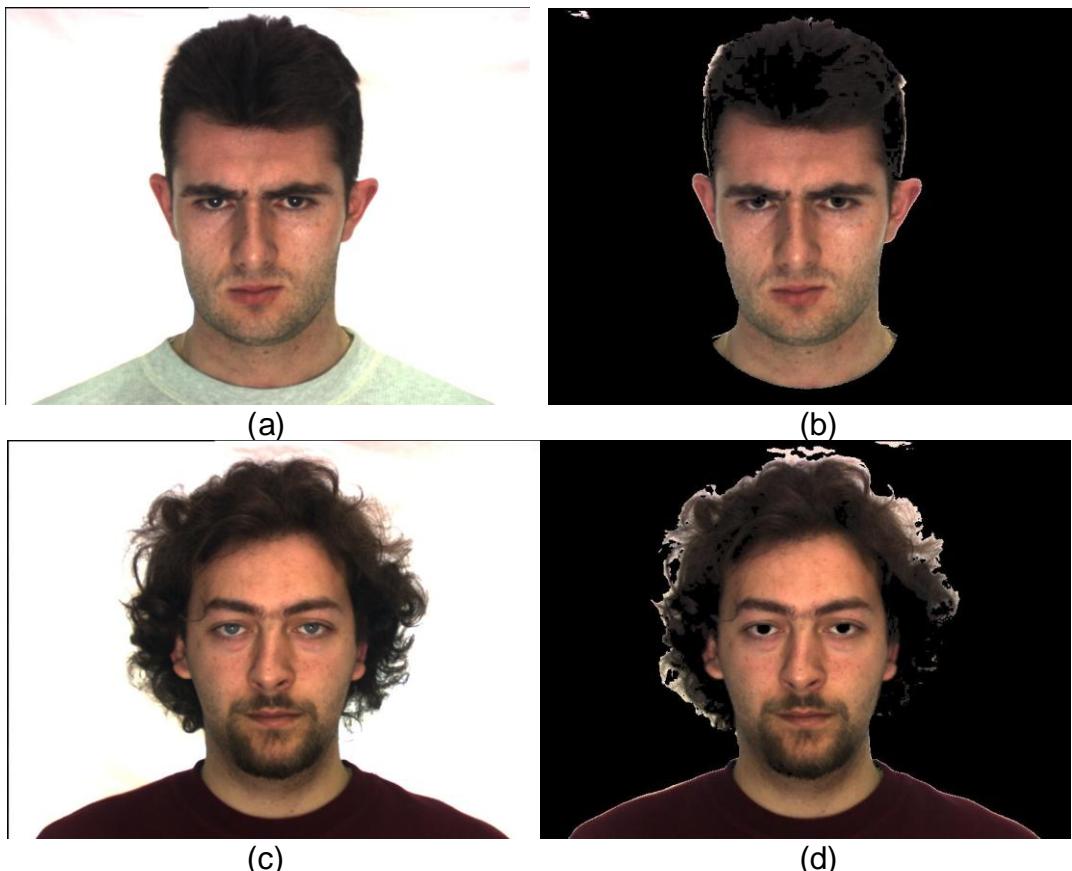


Figura 41 - (a), (c) e (e) imagens originais, (b), (d) e (f) imagens processadas do banco de imagens AR

Um dos problemas identificado no algoritmo é que ele não consegue remover o fundo das imagens que possui cores semelhantes à da pele humana como pode ser observado na imagem resultante da figura 42. Isso pode acarretar falsos positivos na identificação de face ou não face.



Figura 42 - (a) imagem original (b) imagem segmentada

Um outro problema identificado é que algoritmo pode reconhecer parcial ou completamente a pele humana como fundo, como pode ser observado na figura 43, onde as faces foram removidas completamente. Um outro exemplo é apresentado na figura 44 onde a face a face é parcialmente erodida por ser confundida com o fundo. Esse problema resulta na não identificação da face por falta de componentes estruturais.

Uma possibilidade de melhoria do resultado seria aumentar a base de treinamento para a rede neural MLP utilizada e outra possibilidade seria a aplicação de outra metodologia de detecção de pele (estatísticas paramétricas ou as baseadas em histogramas) conforme apresentado no capítulo 3.



Figura 43 - (a) imagem original (b) imagem segmentada



Figura 44 - (a) imagem original (b) imagem processada

A tabela 1 apresenta o resultado percentual e o número total de faces, faces localizadas e faces com erosões nas imagens da Internet e do Banco AR após o processo de segmentação e todas quantificadas visualmente. Essa informação é também apresentada de forma gráfica na figura 45.

Na tabela 2 é apresentado o resultado depois de aplicado o teste *T-Student* nas imagens da internet e do Banco AR.

Tabela 1 – Tabela que apresenta a quantidade de faces identificadas e faces com erosões na face

Banco de Imagens	Total de Faces		Faces Localizadas		Faces com Erosões	
Internet	80	100%	76	95%	12	15%
Banco AR	25	100%	25	100%	8	32%

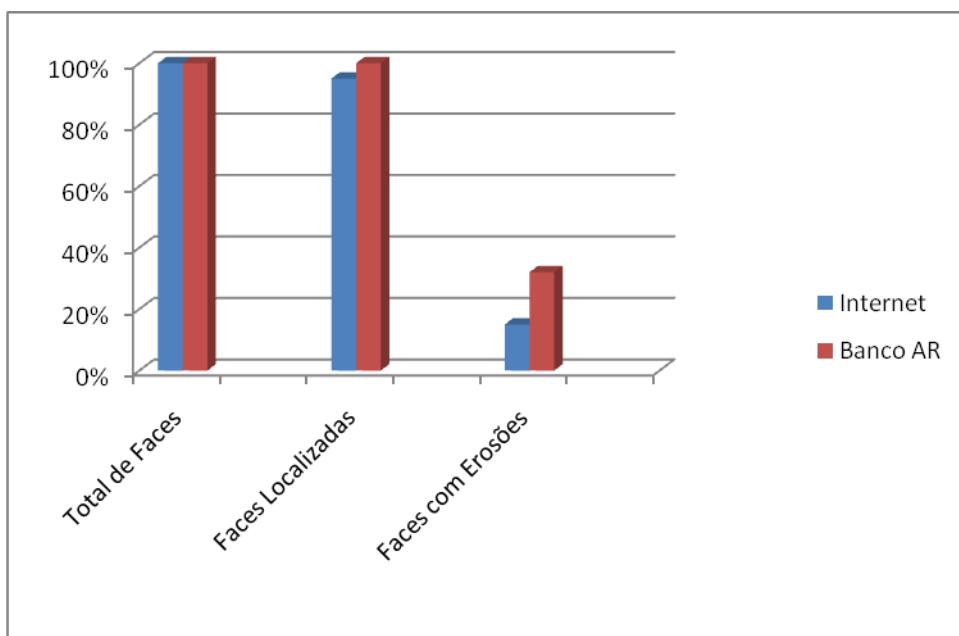


Figura 45 – Gráfico de faces localizadas e faces com erosões na imagem após segmentação

Tabela 2 – Tabela que apresenta o resultado do teste *T-Student* para o total de faces localizadas

Banco de Imagens	Taxa de significância
Internet	0,4471
Banco AR	0,0000
Média	0,2470

O que se pode concluir dos resultados obtidos ilustrados nas tabelas 1 e 2 é que as imagens do Banco AR foram a que apresentaram melhores resultados pelo fato de sua taxa de significância encontrar-se abaixo da média de significância (tabela 2), apesar de algumas faces terem sido erodidas, como pode ser observado no apêndice B. Uma das características desse banco é que a iluminação é diferente em imagens que apresentam a face de uma mesma pessoa, o que pode contribuir para o processo de erosão. Para as imagens da Internet, devido à sua diversidade 5% das faces não foi identificada (tabela 1), o que já era esperado, devido ao fato de não serem imagens padronizadas e conterem pessoas com diferentes cores de pele. No apêndice A encontra-se o resultado da segmentação para todas as imagens da Internet utilizadas.

O módulo de segmentação apresentou resultados que facilitarão a etapa de detecção de face embora, como já mencionado na discussão dos problemas encontrados, existirão falso positivos e falso negativos.

## 5.2 Resultados obtidos com o algoritmo de detecção de face

Para o teste do algoritmo de detecção de faces foram utilizadas as mesmas imagens dos testes de segmentação de pele. As figuras 46(c) e (f) apresentam os resultados após a execução do algoritmo de detecção.

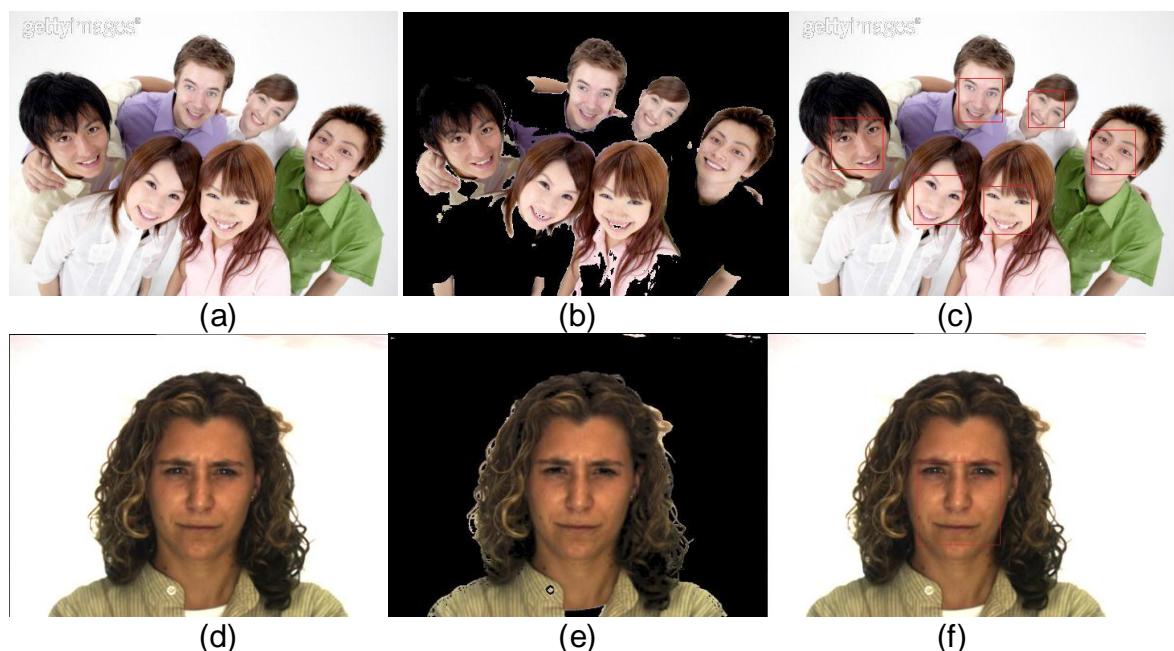


Figura 46 - Resultado do algoritmo de detecção de face: (a) e (d) imagens originais e (b), (c), (e) e (f) imagens resultantes

A tabela 3 apresenta o resultado da execução do algoritmo de detecção de faces para todas as imagens dos bancos utilizados, onde se observa que mesmo com algumas erosões nas imagens a quantidade de faces detectadas foi de 83% para o banco de imagens da Internet, e de 88% para o Banco AR. A figura 47 mostra os resultados da tabela 3 de forma gráfica.

Tabela 3 – Tabela que apresenta o total de faces e o total de faces detectadas

Banco de Imagens	Total de Faces		Faces Detectadas	
Internet	80	100%	67	83%
Banco AR	25	100%	22	88%

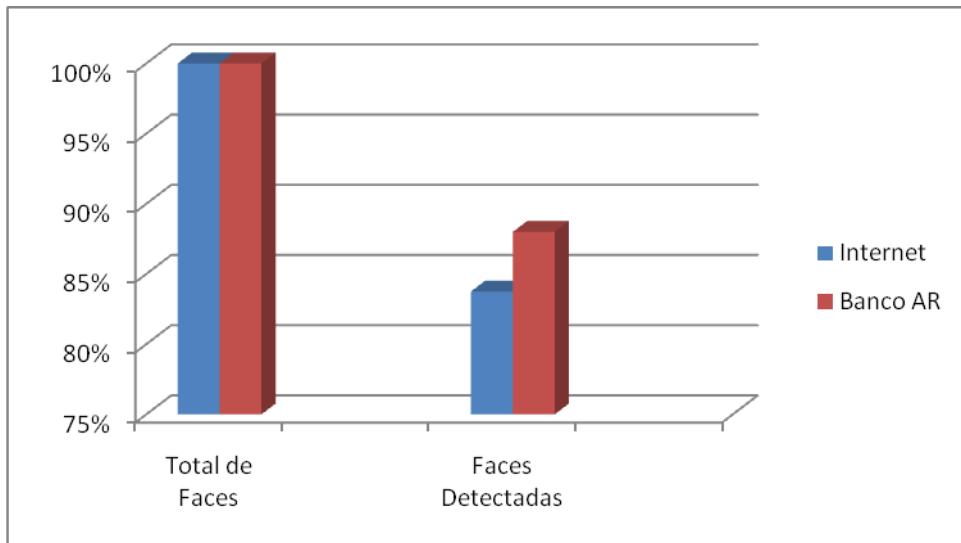


Figura 47 – Gráfico com o total de faces localizadas

Foi também aplicado o teste *T-Student* nos resultados gerados e a taxa média de significância obtida foi de 0,0248, ou seja dentro da faixa adotada (0,05). Esse teste evidenciou que os melhores resultados foram com as imagens do Banco AR que apresentaram uma taxa de significância inferior à média. Já as imagens da internet apresentaram uma taxa de significância acima da média e próxima da faixa de significância adotada, mostrando que seus resultados são inferiores àqueles obtidos com o Banco AR, conforme pode ser observado na tabela 4.

Tabela 4 – Tabela que apresenta o resultado do teste *T-Student* para o total de faces detectadas

Banco de Imagens	Faces Detectadas
Internet	0,0469

<b>Banco AR</b>	0,0028
<b>Média</b>	0,0248

Como já esperado, após a análise das tabelas 3 e 4 os melhores resultados foram obtidos para as imagens do Banco AR, pelo fato deste ser padronizado e não ter fundo tão complexo como as imagens da Internet. As faces não encontradas para o Banco AR devem-se principalmente ao problema de oclusão, já que nesse banco não há faces rotacionadas, ou com iluminação que prejudique a identificação de características.

Para as imagens da Internet as faces não encontradas devem-se aos problemas já mencionados na seção 5.1 da etapa de segmentação, problemas de oclusão e rotação da face.

Para uma melhor avaliação dos resultados é apresentada a tabela 5, que mostra os falsos positivos obtidos em cada banco. Esses resultados são também apresentados de forma gráfica na figura 48. Observa-se nessa tabela que para o Banco AR não houve falsos positivos, enfatizando que os resultados para esse Banco foram superiores aos das imagens da Internet. Para as imagens da Internet verificou-se 4% de falsos positivos, indicando que o número real de faces encontradas foi menor do que o número de faces detectadas pelo software. Esse erro deve-se principalmente aos problemas mencionados na etapa de segmentação na seção 5.1.

Tabela 5 – Tabela que apresenta o total de pessoas e o total de falsos positivos nas imagens

	<b>Imagens Diversas</b>		<b>Banco AR</b>	
<b>Total Formas</b>	81	100%	25	100%

<b>Pessoas</b>	67	83%	22	88%
<b>Falso Positivos</b>	3	4%	0	0%

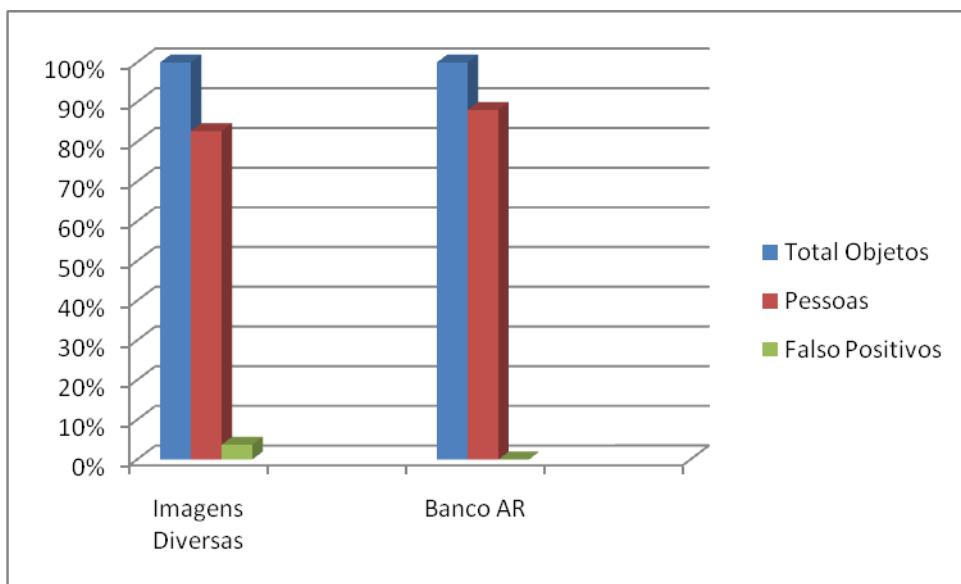


Figura 48 – Gráfico de com o total de faces localizadas e total de falsos positivos



## 6 Conclusões e Sugestões para Trabalhos Futuros

---

Este trabalho teve como objetivo o desenvolvimento de um sistema para detecção automática de faces em imagens coloridas. Para isso foram desenvolvidos dois módulos principais: módulo de segmentação de pele, que tem como objetivo remover o fundo das imagens adquiridas e o módulo de detecção de faces que tem como objetivo detectar as faces nas imagens processadas no módulo anterior.

Para avaliação do software implementado foram utilizados dois banco de imagens, um com imagens padronizadas (Banco AR) e outro banco com imagens adquiridas na Internet contendo faces com diferentes tons de pele e fundo complexo.

Após a execução do algoritmo de detecção de face, foi determinado o percentual de faces detectadas para cada banco e os resultados foram também avaliados com o teste *T-Student*. O percentual de faces detectadas foi de 83% para as imagens adquiridas na internet e de 88% para as imagens do Banco AR. O teste *T-Student* evidenciou que os resultados do Banco AR foram melhores, pelo fato

deste apresentar uma taxa de significância abaixo da média encontrada, o que já era esperado pelo fato dessas imagens serem padronizadas, não conterem faces rotacionadas e com fundo complexo.

Os principais problemas identificados para as imagens não detectadas foram: oclusão, problemas na etapa de segmentação e rotação da face. Esses problemas afetaram principalmente as imagens adquiridas na Internet pelo fato de terem um fundo complexo. As imagens do Banco AR foram afetadas principalmente pela oclusão.

O principal problema identificado no software desenvolvido encontra-se na etapa de segmentação pelo fato da rede neural considerar algumas regiões de fundo como pele, gerando assim falsos positivos, e também algumas regiões de pele como fundo, ocasionando erosão na face e assim gerando falsos negativos.

A etapa de segmentação pode ser melhorada, adotando-se uma outra técnica de segmentação de pele como aquelas já apresentadas no capítulo 3.

## **6.1 Sugestões para trabalhos futuros**

Como sugestão para trabalhos futuros propõe-se:

- Implementar um módulo de segmentação de pele por etnia;
- Na etapa de segmentação aumentar a base de treinamento;
- Na etapa de segmentação utilizar outras técnicas tais como a estatística paramétrica ou baseada em histograma;
- Na etapa de detecção implementar a rede neural com uma entrada adicional, de forma a se ter uma entrada para o olho direito e outra para o olho esquerdo.

## 6.2 Trabalho Publicado

Com este trabalho foi publicado o artigo “*Detecção de Faces Humanas em Imagens Coloridas Utilizando Redes Neurais Artificiais*”, V Workshop de Visão Computacional, (WVC’ 2009).



## Referências Bibliográficas

---



---

A fim de situar no tempo a literatura utilizada para elaboração desta pesquisa, apresenta-se a seguir o gráfico referente à literatura utilizada de acordo com o ano de publicação e quantidade, como pode ser observado na figura 49.

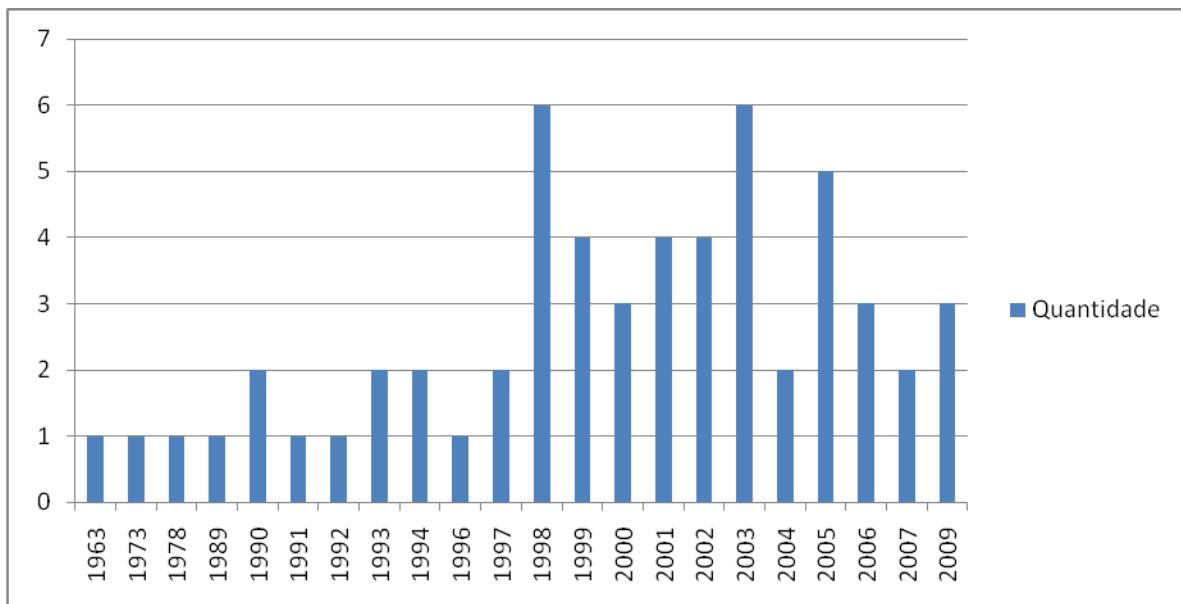


Figura 49 – Artigos e periódicos utilizados na pesquisa de acordo com o ano de publicação.

ALATTAR, A.; RAJALA S. (1999). “*Facial Features Localization in Front View Head And Shoulders Images*”, IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 6, pp.3557-3560.

AMIT, Y.; GERMAN, D.; JEDYNAK B. (1997). “*Efficient Focusing and Face Detection*”, Springer-Verlag, Face Recognition: From Theory to Applications, pp. 143-158.

AMORIM, J. (2009). “*Sistema Nervoso*”, Site, Sala de Aula, disponível em [http://www.passeiweb.com/na\\_ponta\\_lingua/sala\\_de\\_aula/biologia/biologia\\_animal/sistema\\_nervoso/sist\\_nervoso](http://www.passeiweb.com/na_ponta_lingua/sala_de_aula/biologia/biologia_animal/sistema_nervoso/sist_nervoso), ultimo acesso em: dezembro de 2009.

BECHELLI, L. M., CURBAN, G. V.; “*Compêndio de Dermatologia*”; Atheneu Editora; São Paulo; 1963.

BHUIYAN, M. A.; AMPORNARAMVETH , V.; MUTO, S.; UENO, H. (2003). “*Face Detection and Facial Feature Localization for Human-machine Interface*”, NII Journal, n. 5.

BIANCHI, M. F. (2006), “*Extração de características de imagens de faces humanas através de wavelets, PCA e IMPCA*”, Dissertação, Escola de Engenharia de São Carlos, USP – Universidade de São Paulo.

BONVENTI JR, W; COSTA, A. H. R.; “*Sistema semi-automática para detecção de pele por agrupamentos nebulosos*”; VI Simpósio Brasileiro de Automação Inteligente; Brasil; 2003.

CAETANO, T., OLABARRIAGA, S., BARONE, D.; “*Performance evaluation of single and multiple-Gaussian models for skin color modeling*”; Proceedings XV Brazilian Symposium on Computer Graphics and Image Processing SIBGRAPI; IEEE; pp. 275-282; 2002.

CAETANO, T. S., BARONE, D. A.; “*A probabilistic model for the human skin color*”; Proceedings of 11th International Conference on Image Analysis and Processing ICIAP’01; IEEE; pp. 279-284; 2001.

CAI, J.; GOSHTASBY, A.; YU, C. (1999). “*Detecting human faces in color images*”, International Workshop on Multi-Media Database Management Systems, pp. 124-131.

FERIS, R. S.; CAMPOS, T. E.; MARCONDES-JR, R. C. (2000). “*Detection and Tracking of Facial Features in Video Sequences*”, MICAI 2000: Advances in Artificial Intelligence, pp. 127-135.

GASPAR, T. L. (2006), “*Reconhecimento de faces humanas usando redes neurais MLP*”, Dissertação, Escola de Engenharia de São Carlos, USP – Universidade de São Paulo.

- GASPAR, T. L.; PAIVA, M. S. V. (2005), “*Localização de características faciais usando projeções horizontais e verticais*”, I Workshop de Visão Computacional, (WVC’ 2005), pp. 135-138.
- GONZALES, R. C.; WOODS, R. E. (2000), “*Processamento de Imagens Digitais*”, Editora Edgard Blücher.
- GOUVEIA, W. R.; PAIVA, M. S. V. (2009), “*Detecção de Faces Humanas em Imagens Coloridas Utilizando Redes Neurais Artificiais*”, V Workshop de Visão Computacional, (WVC’ 2009).
- GOUVEIA, W. R.; LORENCETTI, A. A.; OLIVETE, A. L. (2005). “*Subtração de fundo em imagens digitais utilizando redes neurais artificiais MLP*”, I Workshop de Visão Computacional, (WVC’ 2005), pp. 36-39.
- HAYKIN, S. (2001), *Redes Neurais: princípios e prática*, Editora Bookman.
- HSU, R.; Abdel-Mottaleb, M.; Jain, A. K. (2002). “*Face detection in color images*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 5, pp. 696-706.
- JIN, Z.; LOU, Z.; YANG, J.; SUN, Q. (2007), “*Face detection using template matching and skin-color information*”, Neurocomputing, vol. 70, pp. 794–800.
- JONES, M., REHG, J.; “*Statistical color models with application to skin detection; International Journal of Computer Vision*”; Kluwer; pp. 81-96; 2002.
- KANADE, T. (1973). “*Picture Processing System by Computer Complex and Recognition of Human Faces*,” doctoral dissertation, Universidade de Kyoto.
- KIRBY, M.; SIROVICH, L. (1990). “*Application of the Karhunen-Loeve procedure for the characterization of human faces*”. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pp.103-108.
- KOTROPOULOS, C.; PITAS, I. (1997). “*Rule-Based Face Detection in Frontal Views*,” IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 4, pp. 2537-2540.

KOVAC, J., PEER, P., SOLINA, F.; "Human skin colour clustering for face detection", EUROCON 2003 – International Conference on Computer as a Tool, IEEE, 2003.

LAURENT, C., LAURENT, N., BODO, Y.; "A *human skin detector combining mean shift analysis and watershed algorithm*", Proc. of International Conference on Image Processing; Vol III; IEEE.; pp. 1025-1028; 2003.

LEE, H. S; LI, H. (2004). "Recognizing frontal face images using Hidden Markov models with one training image per person", International Conference on Pattern Recognition, vol. 1, pp. 318-321.

LEE, H. W.; KIL, S.; HAN, Y.; HONG; S. H. (2001). "Automatic Face and Facial Features Detection", IEEE International Symposium on Industrial Electronics, vol. 1, pp 254-259.

LEE, K. F.; HON, H. W.; HWANG, M. Y.; HUANG, X. (1990). "Speech Recognition Using Hidden Markov Models: A CMU Perspective", Speech Communication, Elsevier Science Publishers B. V., vol. 9, pp. 497-508.

LOPES, E. C. (2005). "Detecção de Faces e Características Faciais", Relatório Técnico, Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS.

LYON, D.; VINCENT N. (2009). "Interactive Embedded Face Recognition", Journal of Object Technology, vol 8, nº 1, pp. 23-53.

MARTINEZ, A. M.; BENAVENTE, R. (1998), "The AR Face Database", CVC Technical Report 24.

METZ, C. E.; "Basic principles of ROC analysis"; Seminars in Nuclear Medicina; pp. 283-298; 1978.

NASCIMENTO, A. V.; GOZAGA, A. (2005). "Detecção de faces humanas em imagens digitais: um algoritmo baseado em lógica nebulosa", I Workshop de Visão Computacional, (WVC' 2005), pp. 96-99.

- NEIFAN, A. V.; HAYES, M. H. (1998). “*Face Detection and Recognition Using Hidden Markov Models*”, International Conference on Image Processing, vol. 1, pp. 141-145.
- PRADO, J. L. (2004), “*Investigação biométrica em imagens digitais para detecção de faces humanas através da proporção divina*”, Dissertação, Escola de Engenharia de São Carlos, USP – Universidade de São Paulo.
- RABINER L. R. (1989), “*A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*”, Proceedings of the IEEE, Vol. 77, n. 2, pp. 257-286.
- RABINER, L. R.; JUNG, B. H. (1993), “*Fundamentals of Speech Recognition*”. Prentice Hall.
- RAMOS FILHO, H. S.: “*Detecção de pele humana em Imagens veiculadas na WEB*”; Dissertação de Mestrado; Maceió; 2006.
- RIKERT, T. D. (1999). “*Texture-Based Statistical Models for Object Detection in Natural images*”, Dissertação de mestrado, Massachusetts Institute of Technology.
- ROWLEY, H. A.; BALUJA, S.; KANADE, T. (1998a), “*Neural network-based face detection*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, pp. 23-38.
- ROWLEY, H. A.; BALUJA, S.; KANADE, T. (1998b), “*Rotation Invariant Neural Network-Based Face Detection*”, IEEE Proceedings on Computer Vision and Pattern Recognition.
- SABER, E.; TEKALP, A. M. (1998), “*Frontal-view Face Detection and Facial Feature Extraction Using Color, Shape and Symmetry Based Cost Functions*”, Pattern Recognition Letters, vol. 19, pp. 669-680,
- SAMARIA, F.S. (1994). “*Face Recognition Using Hidden Markov Models*”, Tese, Univ. of Cambridge, 1994.

- SANTOS, A. R. (2005), “*Identificação de faces humanas através de PCA-LDA e redes neurais SOM*”, Dissertação, Escola de Engenharia de São Carlos, USP – Universidade de São Paulo.
- SIROHEY, S. A. (1993). “*Human face segmentation and identification*”, Relatório Técnico, Universidade de Maryland.
- SOBOTTKA, J.; PITTA, I. (1996). “*Segmentation and tracking of faces in color images*”, Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, pp. 236-241.
- TERRILLON, J., SHIRAZI, M. N., FUKAMACHI, H., AKAMATSU, S.; “*Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images*”; IEEE International Conference on Face and Gesture Recognition; pp. 54-61; 2000.
- TURK, M. A.; PENTLAND, A. P. (1991). “*Face recognition using eigenfaces*”. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586- 591.
- VEZHNEVETS, V. (2002). “*Face and Facial Feature Tracking for Natural Human-Computer Interface*”, disponível em <http://graphics.cs.msu.ru/en/publications/text/gc2002vvezh.pdf>, ultimo acesso em: dezembro de 2009.
- WANG, J.G.; SUNG, E. (1999), “*Frontal-view Face Detection and Facial Feature extraction Using Color and Morphological Operations*”, Pattern Recognition Letters, vol. 20, pp. 1053-1068.
- WON, K. W.; LAM; K. M., SIU, W.C. (2001). “*An Efficient Algorithm for Human Face Detection and Facial Feature Extraction under Different Conditions*”, Pattern Recognition Letters, vol 34, pp. 1994-2004.
- XU, T. Q.; LI, B. C. Li; WANG, B. (2003). “*Face detection and recognition using neural network and hidden Markov models*”, International Conference on Neural Networks and Signal Processing, vol. 1, pp. 228-231.

- YANG, G.; HUANG, T. S. (1994). "Human Face Detection in Complex Background", Pattern Recognition, vol. 27, no. 1, pp. 53-63.
- YANG, M. H.; KRIEGMAN, D. J.; AHUJA, N. (2002). "Detecting Faces in Images: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 24, no. 1.
- YANG, M. H.; AHUJA, N. (1998) . "Detecting human faces in color images", International Conference on Image Processing", vol 1, pp. 127-130.
- YUILLE, A. L.; HALLINAN, P. W.; COHEN, D. S. (1992). "Feature Extraction from Faces Using Deformable Templates", International Journal of Computer Vision, vol.8, n. 2, pp. 99-111.
- YI, J., PARK, J., KIM, J., CHOI, J., "Robust skin color segmentation using a 2d plane of RGB color space", Lecture Notes in Computer Science, vol. 2869, ISCIS 2003, pp. 413-420; 2003.
- ZAPATA, J.; RUIZ, R. (2007), "Biometric and Color Features Fusion for Face Detection and Tracking in Natural Video Sequences", IWINAC 2007, PART II, PP. 72–80.



## Apêndice A – Banco de Imagens da Internet

Imagen 1



Imagen 2



Imagen 3



Imagen 4



Imagen 5



Imagen 6



Imagen 7

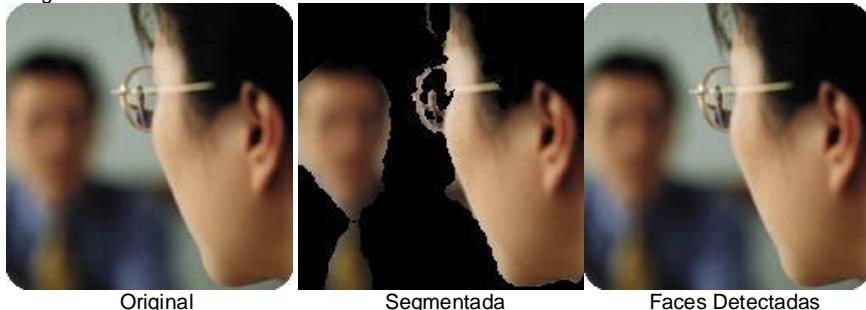


Imagen 8



Imagen 9



Imagen 10



Original

Segmentada

Faces Detectadas

Imagen 11



Original

Segmentada

Faces Detectadas

Imagen 12



Original

Segmentada

Faces Detectadas

Imagen 13



Original

Segmentada

Faces Detectadas

Imagen 14



Imagen 15



Imagen 16



Imagen 17

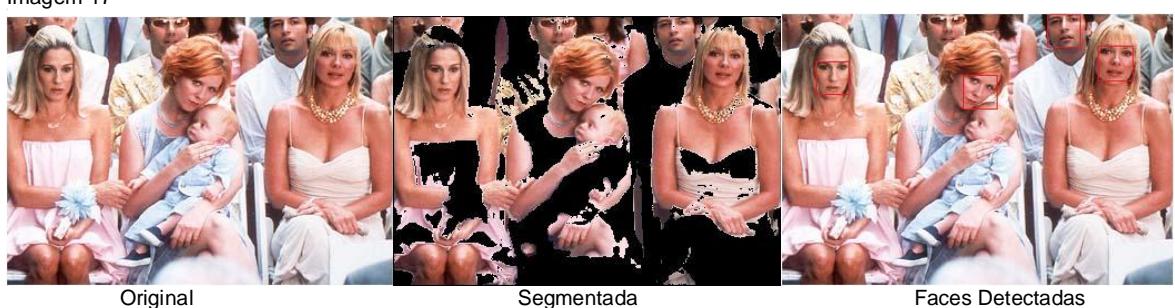


Imagen 18



Imagen 19

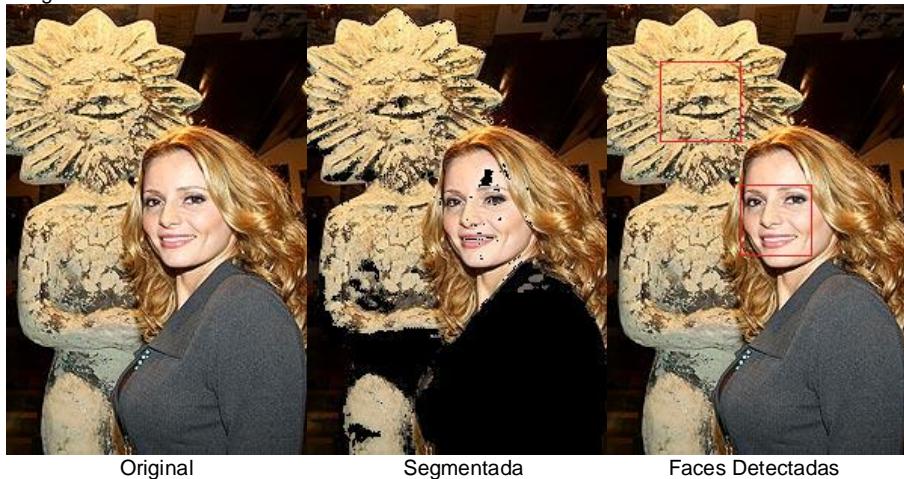


Imagen 20



Imagen 21



Imagen 22

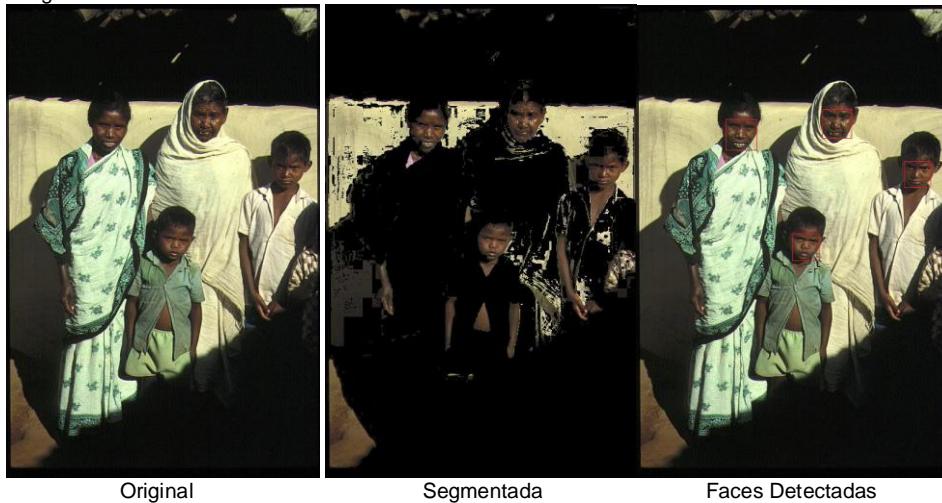


Imagen 23



Imagen 24



Imagen 25



## Apêndice B – Banco de Imagens AR

Imagen 1



Original

Segmentada

Faces Detectadas

Imagen 2



Original

Segmentada

Faces Detectadas

Imagen 3

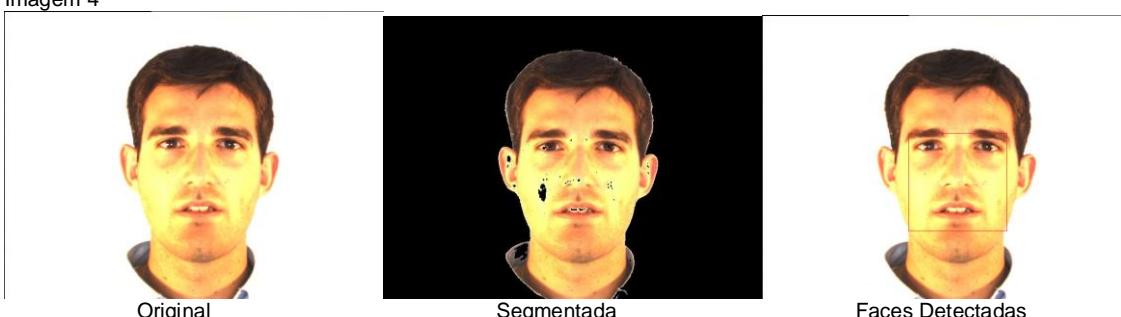


Original

Segmentada

Faces Detectadas

Imagen 4



Original

Segmentada

Faces Detectadas

Imagen 5



Original

Segmentada

Faces Detectadas

Imagen 6



Original

Segmentada

Faces Detectadas

Imagen 7



Original

Segmentada

Faces Detectadas

Imagen 8

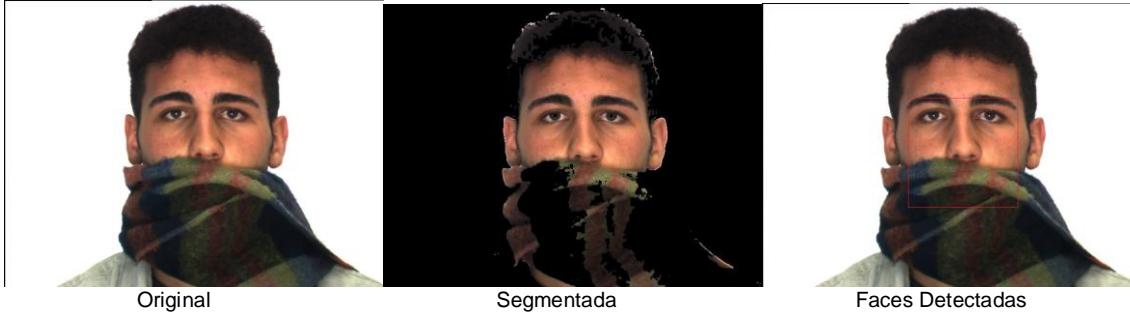


Original

Segmentada

Faces Detectadas

Imagen 9



Original

Segmentada

Faces Detectadas

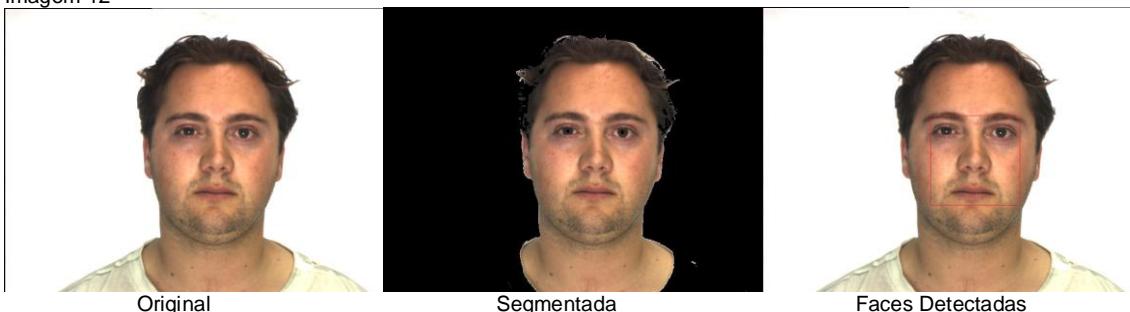
**Imagen 10****Imagen 11****Imagen 12****Imagen 13****Imagen 14**

Imagen 15

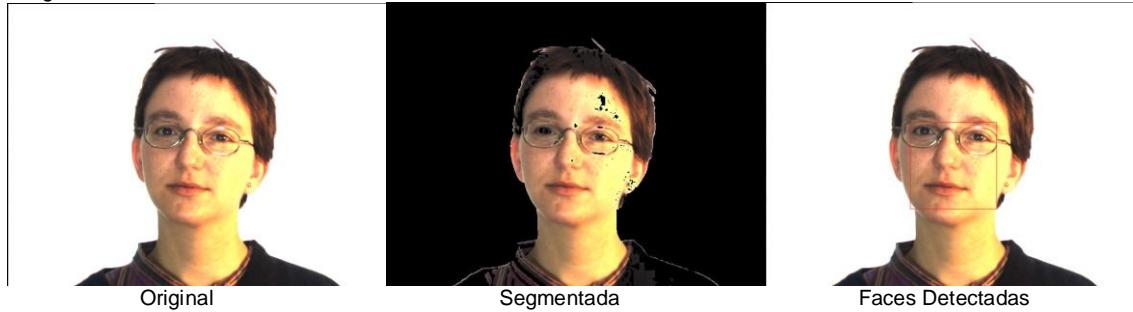


Imagen 16



Imagen 17

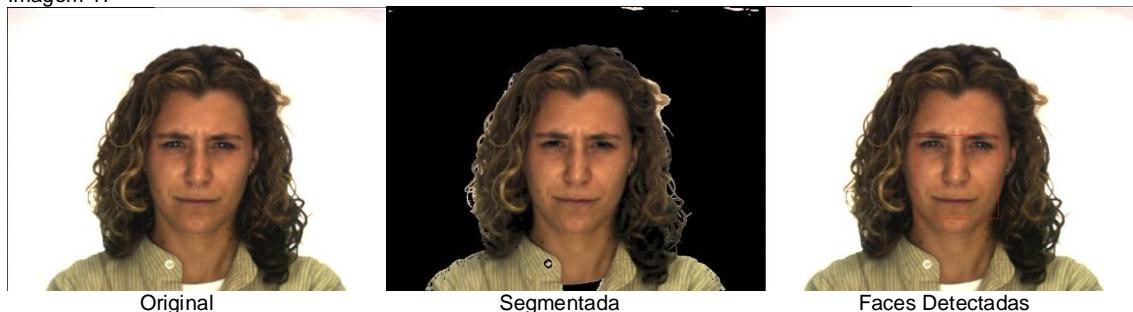


Imagen 18

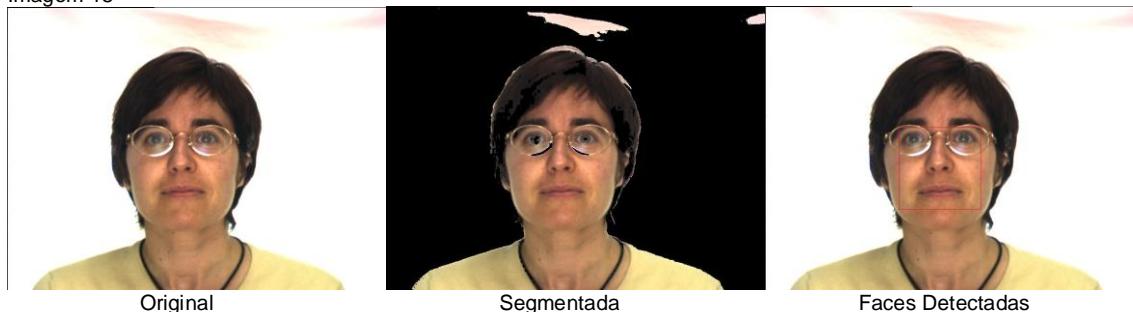


Imagen 19

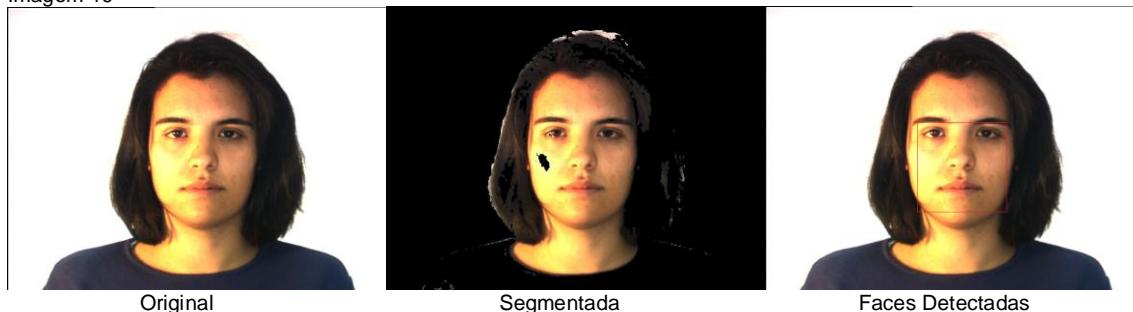


Imagen 20



Imagen 21

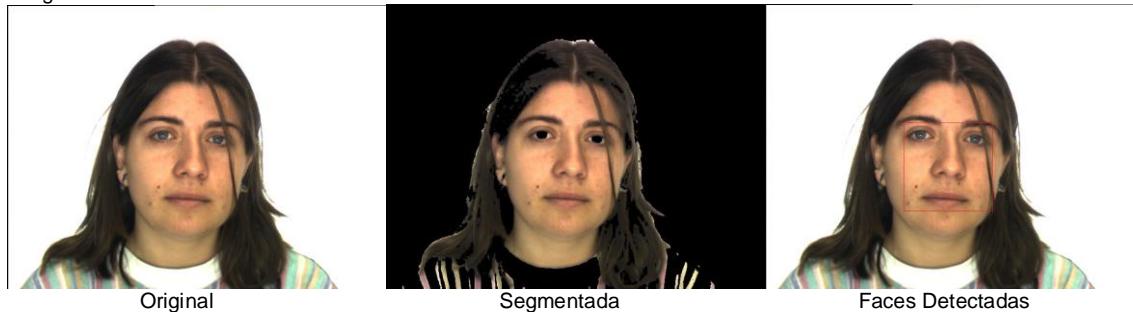


Imagen 22

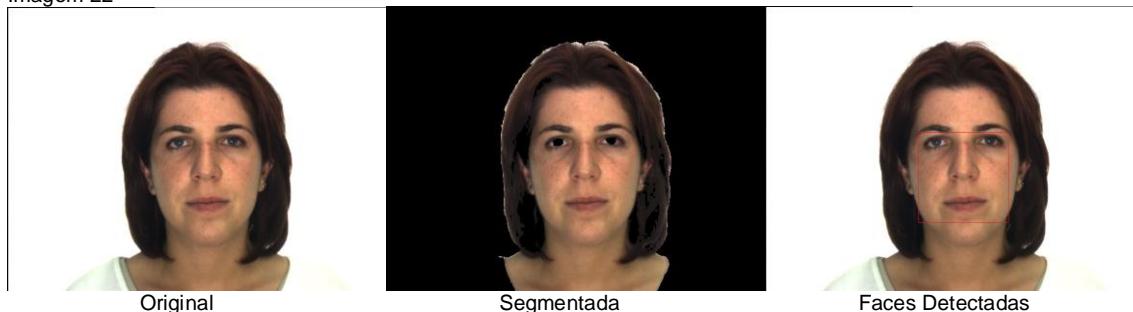


Imagen 23



Imagen 24

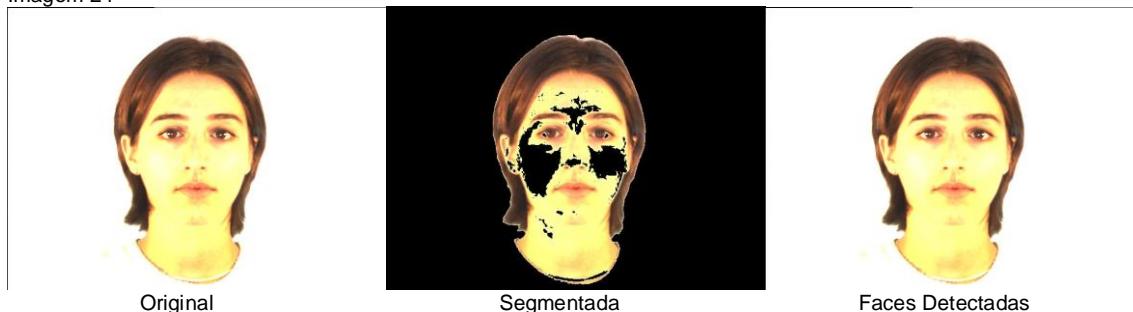


Imagen 25

