

Universidade Federal de Alagoas  
Instituto de Computação



Dissertação de Mestrado

# **Deteccção de Pele Humana em Imagens Veiculadas na Web**

Heitor Soares Ramos Filho  
heitor.ramos@gmail.com

Orientador:  
Prof. Dr. Alejandro C. Frery

Maceió, Fevereiro de 2006

Heitor Soares Ramos Filho

# **Detecção de Pele Humana em Imagens Veiculadas na Web**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Curso de Mestrado em Modelagem Computacional de Conhecimento do Departamento de Tecnologia da Informação da Universidade Federal de Alagoas.

Orientador:

Prof. Dr. Alejandro C. Frery

Maceió, Fevereiro de 2006

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**  
**Divisão de Tratamento Técnico**  
**Bibliotecária Responsável: Helena Cristina Pimentel do Vale**

R175d Ramos Filho, Heitor Soares.  
Detecção de pele humana em imagens veiculadas na web / Heitor Soares  
Ramos Filho. – Maceió, 2006.  
ix, 127f. : il.

Orientador: Alejandro C. Frery.  
Dissertação (mestrado em Modelagem Computacional de Conhecimento) –  
Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2006.

Bibliografia: f. 121-127.

1. Processamento de imagens. 2. Modelagem computacional. 3. Detecção de  
pele – Análise estatística. I. Título.

CDU: 004.932



Ata da defesa de dissertação do aluno  
Heitor Soares Ramos Filho

Realizou-se no dia 13 de fevereiro de dois mil e seis, às 09:00 horas, no auditório do Programa de Pós-Graduação do Instituto de Matemática, da Universidade Federal de Alagoas, a defesa de dissertação de Mestrado em Modelagem Computacional de Conhecimento, intitulada "Detecção de Pele Humana em Imagens Veiculadas na Web", apresentada por Heitor Soares Ramos Filho, diplomado em Engenharia Elétrica, como requisito parcial para a obtenção do grau de Mestre em Modelagem Computacional de Conhecimento, à seguinte comissão examinadora:

Professor Roberto Marcondes César Junior  
Instituto de Matemática e Estatística – IME/USP

Professor Glauber José Ferreira Tomaz da Silva  
Instituto de Computação – IC/UFAL

Professor Alejandro César Frery Orgambide  
Instituto de Computação – IC/UFAL

A dissertação foi considerada aprovada pela Comissão. Finalizados os trabalhos, lavrou-se a presente ata, que vai assinada por mim e pelos membros da Comissão.

Maceió, 13 de fevereiro de 2006.

José Vitor de Menezes Torres  
Secretário

Professor Roberto Marcondes César Junior  
Doutor em Física – Universidade de São Paulo/Brasil

Glauber José Ferreira Tomaz da Silva  
Doutor em Ciência da Computação – Universidade Federal de Pernambuco/Brasil

Alejandro César Frery Orgambide  
Doutor em Computação Aplicada – Instituto Nacional de Pesquisas Espaciais/Brasil

# Agradecimentos

Primeiramente gostaria de agradecer aos meus pais Heitor e Santina Ramos, não apenas pela concepção, mas principalmente pela criação e por todos os princípios morais e éticos que me foram passados. Em todos os momentos de minha vida todos os esforços que estavam aos seus alcances foram realizados no sentido de me propiciar uma educação consistente.

À minha esposa Karina por todo o amor, compreensão e paciência, em especial durante a confecção deste trabalho. Certamente sem o apoio dela este trabalho teria sido muito mais difícil e a longa jornada para sua confecção seria árdua.

Aos meus sobrinhos Caio, Taísa e Lívia pelos choros, gritos, barulhos e infinitas alegrias que de alguma forma contribuíram consistentemente para o sucesso desta etapa da minha vida.

Ao amigo e Prof. Dr. Evandro Costa por sempre me manter próximo à academia e por todas as oportunidades que me foram dadas até hoje.

Ao Mario Cardona pela idéia, pelas interações, pelo apoio e pela paciência para ler, responder e contribuir com os diversos e-mails trocados durante a execução deste trabalho.

Ao “Guru” Prof. Dr. Alejandro Frery, meu amigo e orientador, por toda a paciência, disponibilidade, sabedoria e pela confiança depositada em mim para a elaboração deste trabalho.

A todas pessoas do meu convívio, minhas irmãs Chris e Grace, meu irmão Ruy e meus cunhados que direta ou indiretamente me apoiaram e contribuíram para esta vitória.

Não poderia esquecer de agradecer a todos aqueles que contribuem com a comunidade de Software Livre, desde os idealizadores a todos os colaboradores anônimos, pois sem estas ferramentas este trabalho seria muito mais difícil.

# Resumo

A detecção de pele humana em imagens digitais é utilizada para diversas aplicações como detecção de faces, reconhecimento de gestos e detecção de pornografia. A forma mais comum de detecção de pele encontrada na literatura é através da cor. A variação de iluminação pode redundar em efeitos nocivos à detecção de pele, pois a aparência da cor de um objeto é diretamente relacionada com a forma em que ele é iluminado. Para a detecção de pele pela cor exclusivamente, estratégias robustas às variações de iluminação e modelos descrevam corretamente o agrupamento das cores da pele devem ser utilizados. Ao enfrentarmos o problema de detecção de pele em ambientes onde não há controle sobre as características da imagem, não encontramos resultados satisfatórios na literatura, principalmente quando se refere à tentativa de minimizar os efeitos da variação de iluminação. As estratégias de correção de cor presentes na literatura melhoram consideravelmente a detecção de pele em algumas situações específicas, mas degradam esta classificação em outras situações. Neste trabalho, avaliamos o desempenho de sete diferentes modelos de detecção de pele, com quatro diferentes tipos de dados de entrada e propusemos uma estratégia para escolha das imagens que serão submetidas à correção de cor e o tipo de técnica de correção de cor mais adequado para esta imagem. A técnica que utiliza um modelo gaussiano bivariado, utilizando as duas primeiras componentes após aplicarmos transformação de componentes principais ao dados RGB da amostra de pele utilizada para treinamento resultou na melhor técnica abordada nesse trabalho ao utilizarmos a correção de cor proposta. Os resultados obtidos são comparados por meio de diversas métricas derivadas da matriz de confusão, e se mostram pelo menos tão bons quanto os alcançados por técnicas disponíveis na literatura.

# Abstract

Face detection, gesture recognition and pornography content assessment are some of the applications that require the detection of human skin in digital imagery. Most methods employ color as the main feature for this task. Whenever the acquisition conditions are controlled, there is available information about illumination, resolution and geometry, making the skin detection problem a relatively easy task for which there are plenty of results in the literature. The problem becomes more challenging in less structured conditions, mainly because of the influence illumination conditions have on the apparent color of objects. There are proposals for color correction that lead to both good and bad classification results, depending on the input data. When dealing with Web imagery, little can be assumed about their content or about the conditions in which they were acquired, and robust techniques are needed for skin detection. This MSc thesis makes a qualitative assessment of seven skin detection models and of four different types of input data. A heuristic is proposed for deciding if an image requires color correction and, if needed, which is the best suited technique. Results are compared by means of measures derived from confusion matrices, and our approach produces competitive classification products.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	3
1.2	Objetivos . . . . .	5
1.3	Metodologia . . . . .	5
1.4	Estrutura do Texto . . . . .	7
<b>2</b>	<b>Preliminares</b>	<b>9</b>
2.1	Imagem Digital . . . . .	9
2.2	Interpretação dos Dados de uma Imagem Digital . . . . .	18
2.3	Classificação em Imagens Digitais . . . . .	20
2.3.1	Classificação Não-supervisionada . . . . .	22
2.3.2	Classificação Supervisionada . . . . .	22
2.4	Avaliação de Classificadores . . . . .	25
2.5	Espaços de Cor . . . . .	29
2.5.1	RGB . . . . .	29
2.5.2	Cor RGB Normalizada . . . . .	30
2.5.3	XYZ . . . . .	30
2.5.4	HSV . . . . .	32
2.5.5	YCbCr . . . . .	33
2.5.6	Sistemas de Cor de Percepção Uniforme . . . . .	34
2.5.7	CIELAB . . . . .	35
2.6	Transformação de Componentes Principais . . . . .	36
2.7	Análise de Dados . . . . .	40
<b>3</b>	<b>Sistema de Detecção de Imagens Pornográficas</b>	<b>44</b>
3.1	Motivação . . . . .	44
3.2	Arquitetura . . . . .	49
3.2.1	Pré-Processador de Imagem . . . . .	50
3.2.2	Processador de Imagem . . . . .	52
3.2.3	Extrator de Atributos . . . . .	52
3.2.4	Qualificador . . . . .	54

---

<b>4</b>	<b>Detecção de Pele</b>	<b>56</b>
4.1	Arquitetura do Sistema de Detecção de Pele . . . . .	57
4.2	Análise dos Dados . . . . .	59
4.3	Redução de dimensionalidade . . . . .	62
4.4	Correção de Cor . . . . .	70
4.5	Técnicas de Detecção de Pele . . . . .	77
4.5.1	Limiar de Classificação . . . . .	79
4.5.2	Estatístico - Histograma de Cor . . . . .	80
4.5.3	Estatístico - Paramétrico . . . . .	84
<b>5</b>	<b>Resultados</b>	<b>90</b>
<b>6</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>102</b>
6.1	Conclusões . . . . .	102
6.2	Trabalhos Futuros . . . . .	106
<b>A</b>	<b>Matrizes de confusão para as Técnicas de Detecção de Pele Avaliadas</b>	<b>108</b>
<b>B</b>	<b>Exemplos de Detecção de Pele em Imagens Tipicamente Encontradas na Web</b>	<b>117</b>

# Lista de Figuras

2.1	Esquema de estudo e interpretação de um sistema físico . . .	10
2.2	Geração de uma imagem digital multiespectral a partir de uma cena . . . . .	12
2.3	Resposta linear de um sensor remoto a uma certa radiação incidente . . . . .	13
2.4	Curvas de resposta espectral do olho humano . . . . .	15
2.5	Paradigmas de abstração no estudo de cor . . . . .	17
2.6	Classes em um espaço multiespectral bidimensional . . .	21
2.7	Regra de classificação em uma imagem multiespectral . .	22
2.8	Curvas de nível de distribuições gaussianas bidimensionais . . . . .	24
2.9	Curva ROC para quatro classificadores . . . . .	28
2.10	Espaço de cromaticidade XYZ . . . . .	31
2.11	Espaço de cor HSV . . . . .	32
2.12	Espaço de cor CIELAB . . . . .	36
2.13	Transformação de Componentes Principais (2D) . . . . .	39
2.14	Bandas das imagens original e transformada . . . . .	41
2.15	Elementos de um boxplot . . . . .	43
3.1	Arquitetura do sistema T <sub>0</sub> RADO (Módulo analisador de imagem) . . . . .	51
3.2	Abordagem das elipses . . . . .	54
3.3	Qualificador . . . . .	55
4.1	Arquitetura do sistema de detecção de pele . . . . .	58
4.2	Marcação dos <i>pixels</i> de pele e não-pele para construir a base de treinamento . . . . .	60
4.3	Boxplot dos canais RGB da amostra de pele . . . . .	62
4.4	Agrupamento de cores de pele no espaço RGB . . . . .	63
4.5	Agrupamento de cores de pele no espaço HSV . . . . .	63
4.6	Agrupamento de cores de pele no espaço CIELAB . . . . .	64

---

4.7	Opções para redução de dimensionalidade dos dados de pele	65
4.8	Scatter Plot dos dados de treinamento projetados no plano encontrado por TCP . . . . .	67
4.9	Scatter Plot dos dados de treinamento projetados no plano HS . . . . .	68
4.10	Scatter Plot dos dados de treinamento projetados no plano LA . . . . .	69
4.11	Frequências de ocorrência dos pixels (plano HS) . . . . .	70
4.12	Imagem com um desvio no seu espectro de cor causado por iluminação fluorescente . . . . .	72
4.13	Exemplos da correção de cor proposta por Rein-Lien et al. (2002) . . . . .	75
4.14	Limite da taxa de acerto pelo método do histograma . . . . .	83
4.15	Suavização do histograma de pele . . . . .	84
4.16	Suavização do histograma de não-pele . . . . .	85
4.17	Separação dos <i>pixels</i> no plano HS em 9 classes distintas . . . . .	88
4.18	Separação dos <i>pixels</i> no plano HS em 2 classes distintas . . . . .	88
4.19	Separação dos <i>pixels</i> no plano CP1,CP2 em 9 classes distintas . . . . .	89
4.20	Separação dos <i>pixels</i> no plano CP1,CP2 em 2 classes distintas . . . . .	89
5.1	Curva ROC para os métodos analisados sem a correção de cor . . . . .	95
5.2	Curva ROC para os métodos analisados após a correção de cor . . . . .	97
5.3	Curva ROC para os métodos analisados com e sem correção de cor . . . . .	99
5.4	Detecção de pele em uma imagem azulada (iluminação fluorescente) . . . . .	100
5.5	Detecção de pele em uma imagem amarelada (iluminação incandescente) . . . . .	100

# Lista de Tabelas

2.1	Denominação das cores de acordo com a comprimento de onda da radiação incidente . . . . .	16
2.2	Especificação de uma Matriz de Confusão . . . . .	25
4.1	Dados estatísticos da amostra de pele . . . . .	61
5.1	Técnicas Avaliadas e seus dados de entrada . . . . .	91
5.2	Notação utilizada no experimento de avaliação das técnicas de detecção de pele . . . . .	93
A.1	Matriz de confusão para o classificador – Regra geométrica de Kovac et al. (2003) sem correção de cor (geoPeer) . . . .	108
A.2	Matriz de confusão para o classificador – Regra geométrica de Kovac et al. (2003) com correção de cor (geoPeercc) . . .	109
A.3	Matriz de confusão para o classificador – Regra geométrica de Rein-Lien et al. (2002) sem correção de cor (geoHsu) . .	109
A.4	Matriz de confusão para o classificador – Regra geométrica de Rein-Lien et al. (2002) com correção de cor (geoHsucc) .	110
A.5	Matriz de confusão para o classificador – Histograma sem correção de cor (bin64) . . . . .	110
A.6	Matriz de confusão para o classificador – Histograma com correção de cor (bin64cc) . . . . .	111
A.7	Matriz de confusão para o classificador – Histograma suavizado sem correção de cor (spline) . . . . .	111
A.8	Matriz de confusão para o classificador – Histograma suavizado com correção de cor (splinecc) . . . . .	112
A.9	Matriz de confusão para o classificador – Mistura Gaussiana Bivariada (PCA) sem correção de cor (pca2x) . . . . .	112
A.10	Matriz de confusão para o classificador – Mistura Gaussiana Bivariada (PCA) com correção de cor (pca2xcc) . . . . .	113

---

A.11	Matriz de confusão para o classificador – Mistura Gaussiana 9x Variada (PCA) sem correção de cor (pca9x) . . . . .	113
A.12	Matriz de confusão para o classificador – Mistura Gaussiana 9x Variada (PCA) com correção de cor (pca9xcc) . . . . .	114
A.13	Matriz de confusão para o classificador – Mistura Gaussiana Bivariada (HS) sem correção de cor (hs2x) . . . . .	114
A.14	Matriz de confusão para o classificador – Mistura Gaussiana Bivariada (HS) com correção de cor (hs2xcc) . . . . .	115
A.15	Matriz de confusão para o classificador – Mistura Gaussiana 9x Variada (HS) sem correção de cor (hs9x) . . . . .	115
A.16	Matriz de confusão para o classificador – Mistura Gaussiana 9x Variada (HS) com correção de cor (hs9xcc) . . . . .	116

# Capítulo 1

## Introdução

O ser humano se encontra rodeado de informações que são percebidas através do sentido da visão. A radiação eletromagnética proveniente dos objetos atinge a retina, que por sua vez transforma essa radiação em pulsos elétricos formando a psicovisão da imagem correspondente da cena de interesse. O termo “psicovisão” é utilizado pois o cérebro é responsável por receber e interpretar os sinais elétricos provenientes da retina e nos fornece o sentido da visão.

O subsistema de formação de imagens do ser humano é bastante limitado quanto à porção da radiação que consegue perceber (Lira Chávez 2002), considerando que apenas uma pequena faixa do espectro eletromagnético pode ser transformada em estímulos visuais. Em contrapartida, o ser humano é capaz de analisar e interpretar de forma extremamente rápida as imagens que recebe.

Apesar de que, em algumas tarefas, a automatização da análise de imagens pode superar as habilidades da psicovisão (Lira Chávez 2002), como na detecção automática de bordas de cenas complexas, tarefas mais complexas são difíceis de serem reproduzidas por sistemas computacionais.

O subsistema de visão dos seres humanos é capaz de extrair informações de imagens com poucos dados, com informações faltantes, na presença de ruído (até muito intenso), com diversos tipos de ilumina-

ção (configuração geométrica e composição da luz) e de detectar objetos que estão aparentemente sombreados e/ou incompletos. Esta extração é realizada de forma rápida mesmo sem conhecimento prévio da cena e utiliza um amplo banco de dados (baseado na experiência) sobre os objetos que compõem a cena. Algumas técnicas de visão computacional podem realizar tarefas semelhantes, só que com custos computacionais elevados.

A maioria dos sistemas automáticos de detecção de objetos em imagens digitais se baseia no uso de modelos gerais de diversos tipos como, por exemplo, geométricos e estatísticos. A categorização de um elemento de interesse na imagem é realizada contrastando certos atributos do elemento com aqueles considerados típicos de cada um dos modelos previamente estabelecidos. O reconhecimento de um objeto implica em sua classificação (Lira Chávez 2002).

Nesse contexto, a detecção de pele humana em imagens digitais é trabalhada através de diversas abordagens distintas onde modelos matemáticos são montados para que seja extraída a informação de pele da imagem digital. A grande maioria dos sistemas de detecção de pele usa a cor como principal informação útil (atributo) para que sejam construídos modelos matemáticos tratáveis e precisos sobre a cor da pele.

Nesta dissertação trabalharemos com algumas técnicas de detecção de pele em imagens pictóricas que estão descritas na Seção 4.5, com as restrições impostas por sistemas baseados na Web. Todas as técnicas tratadas nesse trabalho procuram identificar a pele humana em imagens digitais através da cor; outros possíveis atributos, como por exemplo textura, não serão considerados. Os conceitos fundamentais sobre imagens digitais e sobre cor estão descritos nas Seções 2.1 e 2.5, bem como o ferramental matemático utilizado nessa dissertação, está descrito no Capítulo 2.

Diversos sistemas como os de detecção de imagens pornográficas (ver Capítulo 3), de detecção de faces (Rein-Lien et al. 2002) e de recuperação de imagens baseada no conteúdo (Smeulders et al. 2000) entre outros, podem utilizar as posições de cores identificadas como sendo

de pele como um tratamento inicial; o trabalho prossegue com a extração de outros atributos derivados. Dessa forma, para esses sistemas, a detecção dos *pixels* de pele se torna essencial e os erros de detecção se propagarão e influenciarão na qualidade desses sistemas. Portanto um sistema de detecção de pele que apresente uma alta taxa de acerto para uma baixa taxa de erros para uma grande diversidade de imagens é desejável em diversas aplicações.

Um dos maiores problemas enfrentados pelos sistemas de detecção de pele é falta de robustez com relação à variação de iluminação no ambiente em que as imagens são adquiridas. Este problema é tratado aqui pelo uso de técnicas de correção de cor apresentadas na Seção 4.4.

A avaliação do desempenho das diversas técnicas de detecção de pele é realizada pelo uso de técnicas supervisionadas que empregam bases de dados de treinamento e de testes. Uma discussão sobre a construção da base de treinamento é feita na Seção 4.2, enquanto a avaliação comparativa das técnicas avaliadas é apresentada no Capítulo 5.

## 1.1 Motivação

O projeto de pesquisa descrito no Capítulo 3 foi o grande motivador para a realização deste trabalho. Este projeto de pesquisa pretende desenvolver um sistema de detecção de pornografia em imagens na Web. Diversas abordagens para detecção de pele têm sido propostas na literatura mas não foi encontrada uma comparação ou medição de desempenho que atendessem aos requisitos de detecção de pornografia.

Jones & Rehg (2002) propõem uma técnica de detecção de pele que, para o treinamento do sistema classificador, utiliza imagens de diversas características; esta talvez tenha sido a maior base de imagens já utilizada para este fim, já que é da ordem de 13640 imagens rotuladas com aproximadamente  $10^9$  pixels.

Além desse trabalho, há na literatura diversos trabalhos de detecção de face que utilizam detecção de pele como uma etapa anterior à

detecção de face. Diversas técnicas, principalmente as estatísticas, foram encontradas mas avaliações e comparações de desempenho não são comumente encontradas e amplamente discutidas. Essa é uma das contribuições deste trabalho, tal como será detalhado mais adiante.

Portanto, para a especificação de um módulo de detecção de pele robusto às variações de iluminação e de paisagens de fundo, precisou-se de uma avaliação rigorosa e de uma comparação entre as diversas técnicas apresentadas na literatura. Este trabalho foi realizado com a finalidade de realizar essa avaliação e de propor uma técnica de detecção de pele que atendesse aos seguintes requisitos:

- ✓ **Desempenho:** O sistema proposto será utilizado em um ambiente onde as imagens serão avaliadas em tempo real, não sendo permitido um grande atraso nesta avaliação.
- ✓ **Robustez:** O sistema proposto será utilizado em ambiente Web onde não se tem controle sobre a qualidade das imagens. Dessa forma, o sistema deverá detectar pele humana em imagens com iluminação inadequada, com diversos tipos de paisagens de fundo e com variadas resoluções, sendo robusto às variações desses parâmetros.

Diversas outras aplicações poderão fazer uso dos resultados de detecção de pele alcançados neste trabalho, uma vez que as restrições impostas ao sistema obedecem os requisitos impostos pela detecção de nudez a que, por sua vez, é uma aplicação genérica. Como exemplo de um trabalho com menos escalabilidade, podemos citar o de Rein-Lien et al. (2002) que, por estar fortemente atrelado à detecção de face e por supor que há de fato uma face em cada imagem de entrada, as hipóteses de trabalho são mais restritivas do que as por nós consideradas. Outro exemplo de aplicação a um domínio mais controlado do que o nosso é o trabalho de Bretzner et al. (2002) que, por visar o reconhecimento de gestos de mãos, supõe a existência de uma mão na cena. No nosso caso, pode haver ou não pele na cena e, com isso, as nossas hipóteses são muito mais gerais.

Devido aos requisitos apresentados, a detecção de pele a ser utilizada pelo sistema de detecção de nudez proposto no Capítulo 3, desse sistema

## 1.2 Objetivos

Os seguintes objetivos serão atingidos após o término do projeto:

1. avaliar as técnicas de detecção de pele humana em imagens veiculadas na Web;
2. propor uma técnica de detecção de pele humana em imagens veiculadas na Web que apresente um desempenho adequado para a detecção de nudez;
3. especificar e desenvolver algoritmos que auxiliem na classificação dos *pixels* de imagens referentes à pele humana;
4. oferecer um módulo de detecção de pele humana para um sistema de detecção de nudez.

## 1.3 Metodologia

Para a realização deste trabalho, e tendo em vista os objetivos descritos na Seção 1.2, iremos trabalhar para especificar e desenvolver um sistema de classificação onde uma imagem digital (para uma discussão detalhada sobre imagens digitais, ver Seção 2.1), terá seus *pixels* classificados como pele ou não-pele.

Para desenvolver este classificador seguiremos os seguintes passos:

- ✓ **Criação da Base de Treinamento:** A base de treinamento é composta de um conjunto de imagens, onde os *pixels* de pele e de não-pele são manualmente marcados com cores pouco freqüentes em imagens digitais. Essa base de treinamento é utilizada para treinar

os algoritmos de correção de cor e de detecção de pele presentes neste trabalho (Seção 4.2).

- ✓ **Criação da Base de Testes:** A base de testes é composta de imagens com os *pixels* marcados da mesma forma que a base de treinamento, mas é formada por imagens diferentes que servem para avaliar a desempenho dos classificadores utilizados neste trabalho (Seção 4.2).
- ✓ **Especificação da Correção de Cor:** Para evitar erros grosseiros de classificação, o sistema de detecção de pele proposto irá, antes de avaliar se um determinado *pixel* pertence ou não à classe de pele humana, corrigir as possíveis distorções de cor provocadas por variações de iluminação (Seção 4.4).
- ✓ **Especificação do Modelo de Detecção de Pele:** Esta etapa consiste em especificar e desenvolver o classificador que irá determinar se um *pixel* pertence ou não ao agrupamento de pele (Seção 4.5).
- ✓ **Avaliação dos Classificadores:** Nesta etapa será feita uma avaliação detalhada de todas as técnicas avaliadas nesse trabalho. Para isto, foi montado um experimento onde foram colhidos indicadores de desempenho das diversas técnicas de detecção de pele, com e sem a utilização da correção de cor (Capítulo 5).

Diversos sistemas foram criados para o desenvolvimento dos classificadores, para isto foi utilizada a linguagem de programação C em ambiente Gnu/Linux, Versão-2.6.11, compilador Gcc, Versão-3.3.6 e a plataforma de desenvolvimento de software Eclipse, Versão-3.1.0 com o módulo CDT, Versão-3.0.0. Foram utilizadas duas bibliotecas de processamento de imagens, a plataforma Gandalf, Versão-1.3.12 e a biblioteca OpenCV, Versão-0.9.5. A biblioteca opencv se demonstrou muito eficiente para a utilização de funções de alto nível para processamento de imagens, como por exemplo algumas funções de morfologia matemática enquanto a biblioteca gandalf se demonstrou muito

eficiente para as funções de baixo nível como o acesso aos atributos dos *pixels*.

Foi utilizada também a biblioteca `Gnu Scientific Library(gsl)`, Versão-1.7-2 para facilitar a tarefa de ordenação de vetores, implementação da função densidade de probabilidade da distribuição gaussiana, entre outras funções matemáticas.

Para o tratamento estatístico dos dados e, principalmente com a finalidade de prototização, foi utilizada a plataforma `R`, Versão-2.2.0 (R Development Core Team 2005). É válido ressaltar a importância das ferramentas “Software Livre” para o desenvolvimento dessa dissertação.

## 1.4 Estrutura do Texto

Esta dissertação está dividida em cinco capítulos e dois apêndices. Neste primeiro capítulo foi descrito o contexto geral para a utilização do sistema de detecção de pele. Também foi descrita a metodologia aplicada para o desenvolvimento do classificador.

No segundo capítulo será apresentada uma discussão geral sobre diversos conceitos utilizados neste trabalho. Neste capítulo serão detalhadas, dentre outras, a noção de imagem digital, e as principais operações de extração, manipulação e interpretação dos dados dessas imagens.

No terceiro capítulo, será descrito e especificado um sistema de detecção de pornografia que foi o grande motivador para o desenvolvimento deste trabalho.

No quarto capítulo serão apresentadas as técnicas aplicadas neste trabalho para o desenvolvimento dos classificadores utilizados para detecção de pele humana em imagens digitais. Neste capítulo, será detalhado o tratamento estatístico empregado nos dados extraídos das imagens e serão avaliados diversos modelos para detecção de pele. Será realizada uma avaliação dos classificadores construídos e identificados os classificadores que apresentam a melhor desempenho segundo diversos critérios de avaliação.

Os resultados numéricos das avaliações dos classificadores estão apresentados no Apêndice A e diversas imagens que ilustram a utilização da técnica de detecção de pele indicada neste trabalho estão apresentadas no Apêndice B.

No quinto e último capítulo serão apresentadas as conclusões sobre este trabalho e diversas indicações para trabalhos futuros que visam dar continuidade a este trabalho.

# Capítulo 2

## Preliminares

Este capítulo apresenta os conceitos e definições preliminares que serão empregados no decorrer desta dissertação. Os conceitos apresentados nessa Seção foram adaptados dos livros de Gomes & Velho (1994) de Lira Chávez (2002) de Richards & Xiuping (1999) e de Fukunaga (1972) e formam a base necessária para a aplicação das técnicas utilizadas nesse trabalho.

### 2.1 Imagem Digital

Para a detecção de pele em imagens digitais veiculadas na Web, iremos trabalhar seguindo a metodologia encontrada em Lira Chávez (2002) para interpretação de um sistema físico através da análise de dados multiespectrais organizados de forma bidimensional. Este esquema está ilustrado na Figura 2.1.

Nesse esquema, temos que informações a respeito do sistema físico representado pela cena (f) são capturadas através de um dispositivo chamado sensor (h). O sensor introduz diversos fenômenos de degradação (v) nas informações coletadas. Os valores observados na imagem multiespectral (g) estão relacionados tanto às propriedades físicas da cena observada quanto à física de funcionamento do sensor, bem como ao meio por onde a informação trafegou entre o primeiro e o segundo.

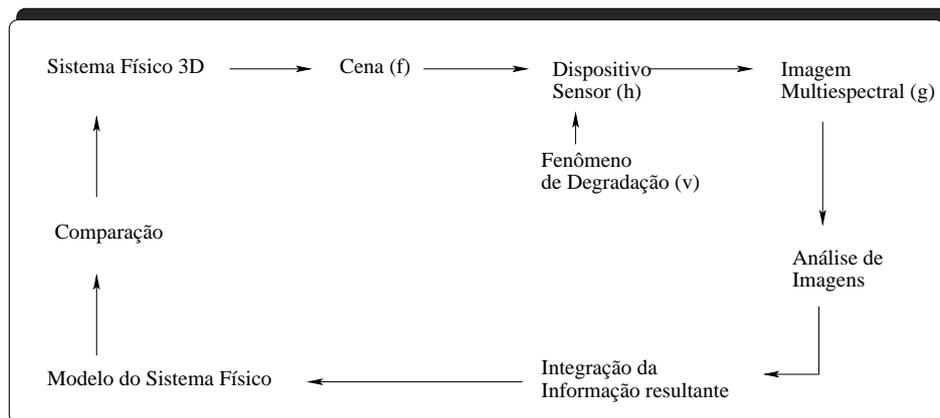


Figura 2.1: Esquema de estudo e interpretação de um sistema físico  
 Fonte: Lira Chávez (2002, p. 46)

A imagem multiespectral é a resposta finita do dispositivo sensor e dos mecanismos de interferência envolvidos na medição das variáveis da cena. A imagem é, portanto, uma simplificação da cena, limitada pelo comportamento do sensor e pelos fenômenos de interferência que degradam a informação que desejamos obter da mesma.

Esta interferência pode, em certas ocasiões, ser desprezada e em outras não. Um exemplo dessa interferência é a iluminação proveniente de fontes de luz que não sejam totalmente brancas, como o é a luz solar sob certas circunstâncias. Outra possível interferência é decorrente da geometria da iluminação da cena e do meio de propagação da radiação refletida, como no caso da fotografia subaquática. Segundo Lira Chávez (2002) a presença de interferências não é uma limitação das situações experimentais existentes no tratamento de imagens mas uma característica inerente das ciências experimentais. Cabe ao pesquisador procurar minimizar os efeitos das interferências.

A etapa de geração da imagem multiespectral é seguida pela fase de análise de imagens. De forma esquemática, a fase de análise de imagens pode ser representada pela adição simbólica:

Processamento de Imagens  
+ Reconhecimento de Padrões  
Análise de imagens

Por sua vez, processamento de imagens é um conjunto de transformações no espaço onde se encontram originalmente os dados multiespectrais com o objetivo de evidenciar ou realçar um certo conjunto de padrões espaciais ou espectrais. No presente trabalho, esta etapa está representada pela detecção de pele que irá realçar padrões para diversos sistemas como os detecção de face, detecção de gestos e detecção de nudez, por exemplo. Para finalizar o ciclo da metodologia indicada, iremos juntar todas as informações extraídas da imagem e comparar o modelo gerado a partir do sistema físico com as características extraídas da imagem. Quando tivermos um modelo que, após uma análise qualitativa e quantitativa, esteja de forma aceitavelmente próximo ao sistema físico modelado, teremos o sistema de detecção de pele pronto.

Usaremos a Figura 2.2 para identificar os elementos que compõem uma imagem digital capturada de uma cena. Neste trabalho, trataremos apenas dos sistemas reflexivos de cor pois são os utilizados pela captura da cena para a formação da imagem digital de nosso interesse. Outros sistemas de cor, como os emissivos, podem ser encontrados na literatura mas não fazem parte do objeto de estudo desse trabalho (ver, por exemplo, Gomes & Velho 1994).

A cena é composta de elementos físicos que interagem com a fonte de radiação fazendo com que uma parte dessa radiação seja refletida e atinja o sistema óptico do dispositivo que irá capturar a cena e gerar a imagem digital. Estes elementos físicos são chamados de Campo Instantâneo de Visão (CIV).

O CIV é uma subárea ou uma subregião da cena. Quando parte da radiação refletida no CIV atinge o sensor óptico, ela é direcionada através do subsistema de condução de radiação e focalizada nos filtros de radiação. Cada um destes filtros é responsável por selecionar um intervalo de energia  $\Delta E_i$  ou de comprimento de onda  $\Delta \lambda_i$ ,  $1 \leq i \leq \gamma$ , do total da radiação incidente para cada um dos  $\gamma$  sensores que estão,

teoricamente, perfeitamente calibrados de forma que o  $i$ -ésimo sensor foi projetado para detectar a radiação no intervalo  $\Delta E_i$ .

Podemos usar tanto  $\Delta E_i$  quanto  $\Delta \lambda_i$  para designar uma faixa do espectro visível, pois a equação de Planck (Equação (2.1)), descreve a relação entre as duas grandezas:

$$E = \frac{hc}{\lambda}, \quad (2.1)$$

onde  $E$  é a energia de cada fóton de luz (partículas de um raio luminoso),  $h \approx 6.626 \times 10^{-34} \text{ J.s}$  (medida em Joules vezes segundos) é a constante de Planck e  $\lambda$  é o comprimento de onda da radiação eletromagnética.

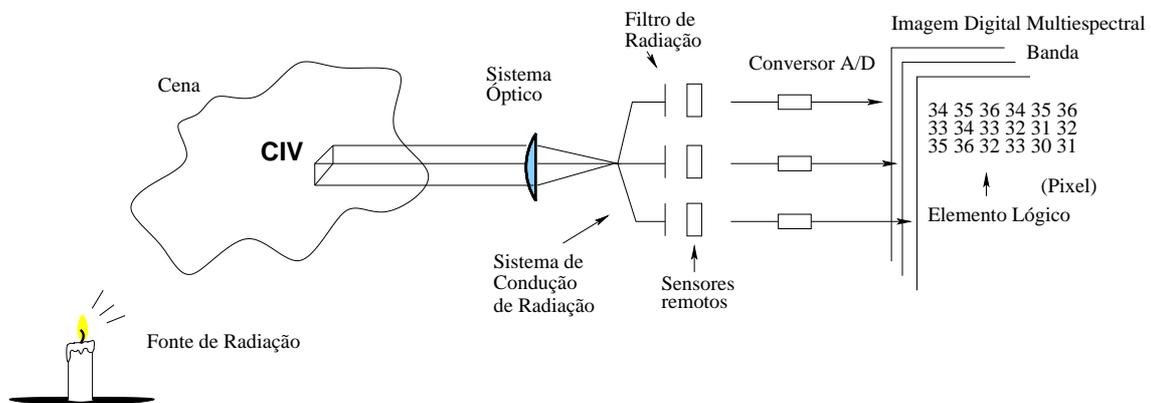


Figura 2.2: Geração de uma imagem digital multiespectral a partir de uma cena

Fonte: Lira Chávez (2002, p. 16)

Para cada faixa de comprimento de onda  $\Delta \lambda_i$  da radiação incidente, os sensores medem a quantidade de radiação por unidade de tempo e geram uma tensão elétrica cuja intensidade é proporcional à energia medida. A tensão gerada pelos sensores, à medida que estes observam sucessivos CIVs, é enviada para um conversor analógico/digital (A/D) que faz a quantização dessa tensão elétrica em uma escala, usualmente os inteiros do intervalo  $[0, 255]$ ; este fato faz com que os valores, uma vez arredondados, possam ser armazenados em palavras de 8 bits em um computador digital.

Os sensores são calibrados para gerar o valor 255 para o CIV mais brilhante da cena e zero para o CIV mais escuro. A resposta do sensor pode ser considerada linear em relação à radiação incidente na região de trabalho. Esta resposta possui um valor de sinal de saída mesmo quando a radiação incidente é nula; este valor é chamado de “polarização do sensor”.

O ganho de um sensor é proporcional à inclinação da reta formada pela radiação incidente e a resposta do sensor. Esses parâmetros estão ilustrados na Figura 2.3 e podem ser ajustados para que os sensores respondam de forma diferente a diferentes condições de iluminação procurando, assim, uma resposta linear.

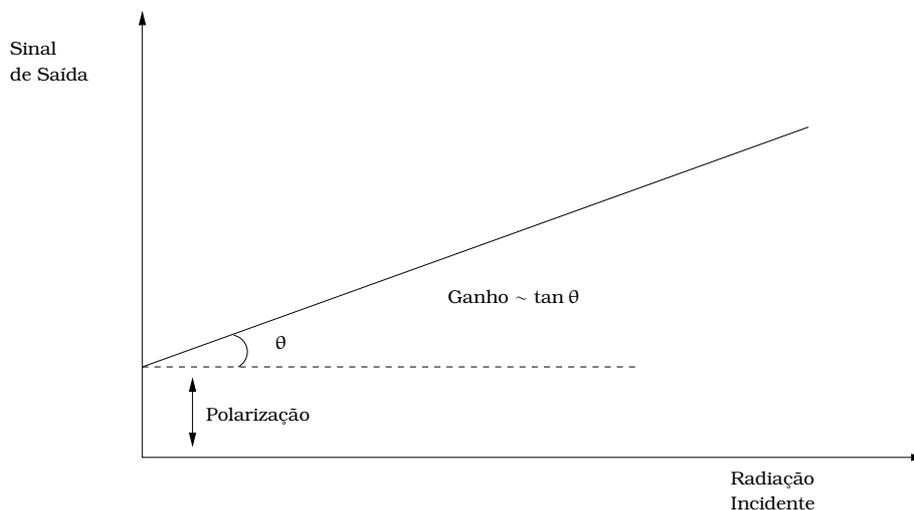


Figura 2.3: Resposta linear de um sensor remoto a uma certa radiação incidente

Fonte: Lira Chávez (2002, p. 18)

Para cada faixa de comprimento de onda  $\Delta\lambda_i$  e para cada CIV, gera-se um número, como resultado da medida da radiação refletida pelo CIV, cujo valor é armazenado (em disco magnético ou óptico, por exemplo) em uma posição lógica que varia com a posição relativa do CIV correspondente na cena (Figura 2.2). Esta posição lógica está associada a um conjunto de coordenadas  $S = \{0, \dots, m - 1\} \times \{0, \dots, n - 1\}$ , denominado “suporte”.

Cada CIV é idealmente representado em uma imagem digital por um *pixel*, de modo que haja um mapeamento único entre um CIV e um *pixel*. Por sua vez, um *pixel*  $h(s)$  é um vetor associado à posição  $s \in S$ , e é composto pelos valores quantizados da radiação incidente, obtidos pelos sensores. Um *pixel* é denotado por:

$$h(s) = (b_1(s), b_2(s), \dots, b_{i-1}(s), b_i(s), b_{i+1}(s), \dots, b_\gamma(s)), \quad (2.2)$$

onde, para cada  $1 \leq i \leq \gamma$ ,  $b_i(s) \in \mathbb{R}$  é o valor numérico obtido para o intervalo de energia  $\Delta E_i$  e  $\gamma$  é o número de sensores utilizados no sistema de visão. Tal como já foi mencionado, é frequente quantizar estes valores, e dentre as possíveis quantizações a mais comum é a que utiliza uma base de oito bits. Com isso,  $b_i \in \{0, 1, \dots, 2^k - 1\}$ , com  $k = 8$ .

Tipicamente, cada sensor é sintonizado em uma faixa do espectro, isto é, ele privilegia a energia incidente no intervalo  $\Delta E_i$  ou nos comprimentos de onda  $\Delta \lambda_i$ .

Um *pixel* é um o conjunto de todos os  $b_i$  correspondentes a um CIV. O índice  $s$  corresponde à posição  $(m, n)$  do *pixel* na imagem digital. Temos, então, que o CIV é um elemento físico da cena enquanto um *pixel* é um elemento lógico da imagem digital.

Na Figura 2.2 temos a representação de um sistema de captura de imagens para a formação de uma imagem digital. Neste sistema temos a presença de três sensores para capturar três canais diferentes da radiação incidente. Na Equação (2.2) temos uma generalização, pois nela são utilizados  $\gamma$  sensores para formar uma imagem multiespectral.

A teoria de Young postula que a retina humana possui três tipos de fotorreceptores responsáveis pela captação da luz incidente (ver em Gomes & Velho 1994). Cada um desses receptores ou canais apresenta uma curva diferente de resposta à luz incidente, como apresentado na Figura 2.4. Nota-se que as curvas de resposta espectral do olho têm um ponto de máximo exatamente nas faixas vermelho, verde e azul do espectro visível (ver Tabela 2.1, página 16).

Esses três canais definem um sistema tricromático para a formação

de cor, que também pode ser encontrado na literatura sob a denominação LMH, por baixa (L – *low*), média (M – *medium*) e alta (H – *high*) frequência. Um quarto fotoreceptor também é encontrado na retina, mas é efetivamente utilizado apenas em baixas iluminações.

Devido ao fato de termos exatamente três fotoreceptores responsáveis pela formação da cor, três componentes numéricas são necessárias para descrever as cores visíveis pelos seres humanos. Assim sendo, a forma específica da Equação (2.2) de representação dos *pixels* na captura de imagens visíveis pelos seres humanos é mostrada na Equação (2.3).

$$h(s) = (R(s), G(s), B(s)) \quad (2.3)$$

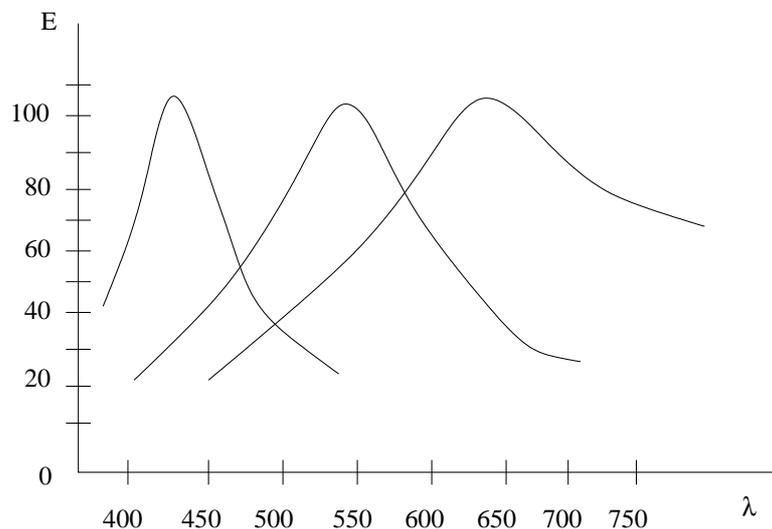


Figura 2.4: Curvas de resposta espectral do olho humano  
Fonte: Gomes & Velho (1994, p. 71)

As componentes R, G e B de uma imagem, quando associadas aos canais vermelho, azul e verde, respectivamente, irão produzir uma imagem realista; esta é conhecida como “operação de reconstrução”. A representação RGB é, talvez, a mais popular em computação gráfica e no armazenamento de imagens pictóricas. Uma discussão mais detalhada da representação RGB e outras existentes na literatura está apresentada na Seção 2.5.

Se  $A$  denota um subconjunto não vazio de  $S$ , então  $h_A$  denota todos os *pixels* associados a ele, isto é,  $h_A = \{h(s) : s \in A \subset S\}$ .

Agora estamos aptos a formalizar o conceito de imagem digital RGB com quantização em oito bits para cada canal como apresentado na Equação (2.4):

$$h: S \rightarrow ([0, 255] \cap \mathbb{N})^3; \quad (2.4)$$

por brevidade, costuma ser utilizada a notação  $K = [0, 255] \cap \mathbb{N}$ . Outras quantizações podem ser utilizadas na prática.

Do ponto de vista perceptual, cada comprimento de onda está associado a uma cor diferente. Nem todo comprimento de onda é capaz de sensibilizar o olho humano, ficando apenas os comprimentos de onda entre 380 nm a 780 nm dentro da faixa visível do espectro luminoso. A Tabela 2.1 fornece a denominação usual de cada faixa do espectro visível.

Tabela 2.1: Denominação das cores de acordo com a comprimento de onda da radiação incidente

<b>Cor</b>	<b>Faixa do Espectro</b>
Violeta	380 - 440 nm
Azul	440 - 490 nm
Verde	490 - 565 nm
Amarelo	565 - 590 nm
Laranja	590 - 630 nm
Vermelho	630 - 780 nm

A cor é o resultado da percepção da luz incidente na região visível do espectro, sendo a presença da luz que cria a sensação de cor. A luz é um fenômeno físico, mas a cor depende da interação da luz com o sistema visual, sendo assim um fenômeno psicofísico (Gomes & Velho 1994).

Ao descrevermos um objeto, usualmente utilizamos a sua cor. Ao dizermos que um determinado objeto é vermelho, qual o real significado disso? A definição de vermelho é idêntica para duas pessoas diferentes?

Devido ao fato de que iremos trabalhar com modelos de cor de pele para a detecção de pele em imagens digitais, essas perguntas nos re-

metem à necessidade de uma especificação precisa do significado das cores, ou seja, uma especificação do modelo matemático para tornar a definição de cor precisa. Existem vários modelos e o mais adequado está diretamente relacionado ao contexto de utilização.

A Figura 2.5 ilustra os níveis de abstração do estudo das cores. Primeiramente devemos entender o processo de formação de cores no universo físico como mostrado na Figura 2.2. Em seguida, devemos estudar os modelos matemáticos de cor como os elementos matemáticos descritos pelas Equações (2.2), (2.3) e (2.4). A partir daí, estudamos os diversos modelos de representação de cor. Nessa etapa temos os espaços de cor descritos na Seção 2.5. A última etapa do estudo de cores em imagens digitais é a codificação das cores no computador; esta última não será tratada detalhadamente neste trabalho pois a partir de qualquer formato de armazenamento utilizado, como o JPEG, GIF, TIFF, PNG entre outros, podemos extrair a informação no formato da Equação (2.3). Este formato é suficiente para a extração das informações necessárias para a detecção de pele como objetiva este trabalho.

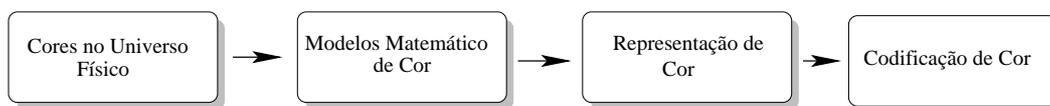


Figura 2.5: Paradigmas de abstração no estudo de cor Fonte: Gomes & Velho (1994, p. 64)

O “cubo” definido pelo produto cartesiano  $K \times K \times K$ , denotado  $K^3$ , é chamado de codificação de cor do espaço de cor “RGB”. Outras representações de cor também são utilizadas como HBS, HVS, CIELAB e YCbCr. Sem perda de generalidade, qualquer mapeamento um-a-um  $[0, 1]^3 \leftrightarrow C \subset \mathbb{R}^3$  é chamado de “espaço de cor”.

Na Seção 2.5 iremos tratar de alguns espaços de cor que foram utilizados no desenvolvimento das técnicas de detecção de pele em imagens digitais; na próxima seção iremos introduzir alguns conceitos referentes à interpretação dos dados de uma imagem digital.

## 2.2 Interpretação dos Dados de uma Imagem Digital

Ao obtermos os dados de uma imagem no formato digital, quantizados espacialmente em *pixels* e quantizados radiometricamente através das intensidades referentes a cada canal de cor (ver Seção 2.1), dispomos de diversas possibilidades para extração de informações (Richards & Xiuping 1999).

Uma possibilidade envolve o uso de um computador para examinar cada *pixel* individualmente com o objetivo de fazer julgamentos sobre os *pixels* examinados baseado nos seus atributos. Para a nossa aplicação, o atributo que será analisado é a cor do *pixel* e o julgamento que será feito é se ele pertence à classe pele ou à classe não-pele. Esta abordagem é conhecida por “análise quantitativa” uma vez que os *pixels* com as mesmas características são contados e agrupados para análise posterior, geralmente através de modelos matemáticos.

Outra abordagem envolve um analista/interpretador humano extraíndo informações através da sua percepção visual dos dados da imagem. Esta abordagem é conhecida como “foto-interpretação” (Richards & Xiuping 1999) ou “interpretação de imagens” e seu sucesso depende do aproveitamento, pelo especialista, das informações espaciais, espectrais e temporais presentes na imagem. Para um melhor entendimento da segunda abordagem podemos citar como exemplo das informações espaciais: formatos, tamanhos, orientação e texturas; informações temporais como a mudança de um objeto de uma imagem para outra (em sequencias de imagens) e as características espectrais que são extraídas pelo analista através de sua experiência sobre as características de reflectâncias espectrais (ver Seção 2.1) dos objetos e como essas características são adquiridas pelas câmeras que capturam a imagem.

As duas abordagens apresentadas para interpretação de imagens digitais são amplamente utilizadas e tem suas próprias características e aplicações. Por exemplo, em imagens de sensoriamento remoto (ver

Richards & Xiuping 1999) as duas abordagens são utilizadas e se complementam. Diversas aplicações se beneficiam das duas abordagens, onde a foto-interpretação é facilitada pela análise quantitativa, e a análise quantitativa é avalizada por um analista de imagens.

A aplicação que norteia este trabalho faz uso das duas abordagens, sendo que em etapas diferentes. Na etapa de treinamento do sistema (ver a descrição de cada etapa na Seção 1.3) as duas abordagens são amplamente utilizadas: um analista humano estará intervindo e analisando a imagem para o refinamento e a validação do modelo matemático de detecção de pele humana. Na etapa de produção do sistema, a abordagem de análise quantitativa é utilizada para a classificação dos *pixels* e a etapa de foto-interpretação será utilizada apenas quando houver uma necessidade de refinar o modelo.

Em termos gerais, a abordagem análise quantitativa é mais precisa do que a foto-interpretação em algumas situações e menos precisa em outras. Por exemplo, se apresentarmos a um analista humano uma imagem com apenas um *pixel* e perguntarmos se esse *pixel* é referente à pele humana, ele não terá nenhuma informação adicional para inferir sobre a questão, em contrapartida, na análise quantitativa o computador pode indicar qual a probabilidade daquele *pixel* pertencer à classe de pele. Outras características como cálculo de áreas também são realizadas com mais sucesso pela primeira abordagem. Por outro lado, a abordagem foto-interpretação produz resultados bastante superiores quando as características a serem extraídas da imagem são de alto nível como na detecção e reconhecimento de faces e gestos, julgamento de pornografia e outras aplicações que envolvem informações espaciais.

Quando, ao invés de um único *pixel*, a análise é baseada em cenas completas, a foto-interpretação apresenta também resultados superiores aos da análise quantitativa devido ao aproveitamento de diversos atributos onde o ser humano pode se utilizar para inferir sobre o problema, como por exemplo a presença de pessoas.

Para nossa aplicação, a foto-interpretação é menos utilizada já que o sistema visa automatizar o processo de detecção de pele humana em

imagens veiculadas na Web. Dessa forma, grandes quantidades de imagens serão utilizadas, não sendo viável para um analista humano intervir diretamente em todas elas. O analista humano só deverá intervir para um retreinamento do modelo, caso se faça necessário. Na próxima seção, apresentaremos os principais conceitos relativos a abordagem da análise quantitativa: a classificação.

## 2.3 Classificação em Imagens Digitais

Para diversas aplicações como detecção de faces, reconhecimento de gestos e detecção de pornografia, a detecção de pele é utilizada como base para a extração dos atributos que são determinantes para a classificação. Nesses casos, os sistemas automatizados não podem fazer uso da foto-interpretação para detectar a pele e a análise quantitativa se faz necessária devido a impraticabilidade da utilização da foto-interpretação, seja por características inerentes à aplicação como no cálculo de áreas em sensoriamento remoto (Richards & Xiuping 1999), como também pelo grande volume de dados a serem analisados como na detecção de pornografia na Web.

A Figura 2.6 mostra uma das metodologias mais utilizadas que serve de suporte para a especificação de algoritmos para a classificação em imagens digitais. Nesta figura, temos a disposição dos *pixels* de uma imagem hipotética de forma que em um eixo temos um dos atributos extraídos da imagem e no outro eixo, temos outro atributo. Nessa abordagem, estamos considerando um espaço de atributos bidimensional. Para o caso de apenas duas classes, podemos apresentar apenas uma classe, onde o complemento é dado pela outra classe. A detecção pele pode ser enquadrada nessa segunda categoria, onde os atributos utilizados são as informações extraídas sobre a cor em cada uma das bandas da imagem.

A classificação envolve rotular os *pixels* de forma que cada rótulo represente uma determinada classe. Esta etapa é baseada na informa-

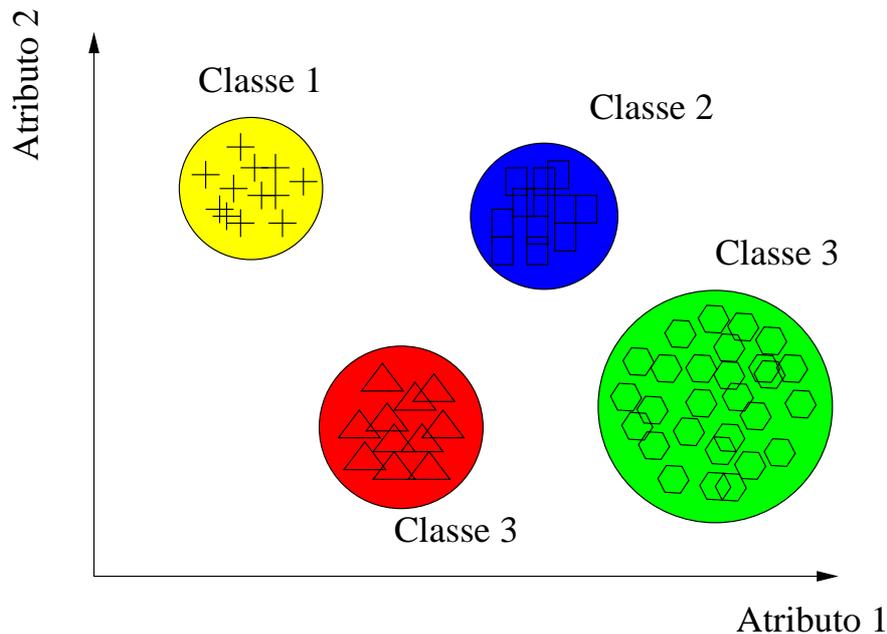


Figura 2.6: Classes em um espaço multiespectral bidimensional

ção espectral dos dados disponíveis. A Figura 2.7 ilustra esse processo onde temos uma sequência de atributos de uma determinada imagem (por exemplo os canais de um espaço de cor), e um *pixel* tem um determinado valor para cada um desses atributos, constituindo o vetor de atributos. O processo de classificação se dá com o mapeamento de um *pixel* em um mapa de rótulos. Esse processo também é conhecido como alocação, categorização e rotulação.

Para classificação em imagens pictóricas, podemos considerar a cor como atributo e utilizá-la para classificar objetos na cena. A maioria dos espaços de representação de cor é tridimensional sugerindo a utilização de um classificador tridimensional. Para a detecção de pele proposta nesse trabalho, utilizaremos a cor como atributo para classificação e utilizaremos algumas operações em espaços de cor para trabalharmos com um classificador bidimensional. As discussões sobre a redução de dimensionalidade  $3D \rightarrow 2D$  e as técnicas utilizadas para o classificador estão apresentadas no Capítulo 4.

Existem dois grandes grupos procedimentos de classificação: (i) Clas-

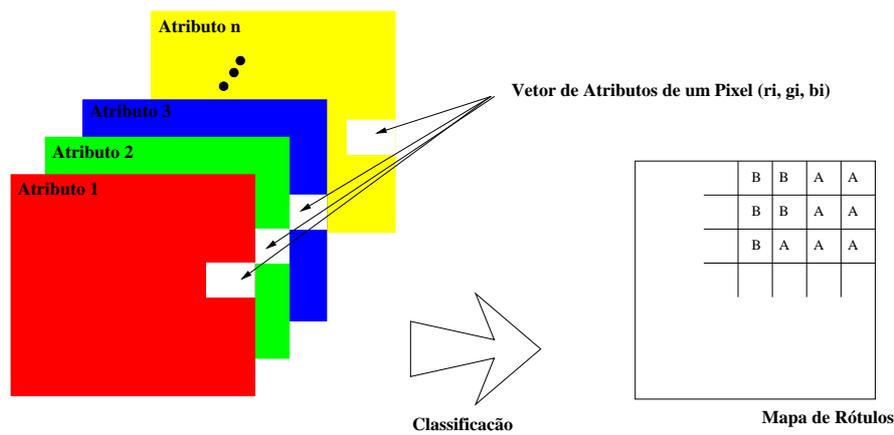


Figura 2.7: Regra de classificação em uma imagem multiespectral

sificação Não-supervisionada e (ii) Classificação Supervisionada, que serão descritos a seguir.

### 2.3.1 Classificação Não-supervisionada

As técnicas não-supervisionadas classificam cada *pixel* de uma imagem em uma determinada classe sem o conhecimento prévio do número de classes presentes. Essas técnicas comumente empregam métodos de agrupamento de dados. Estes classificadores podem ser utilizados para a determinação do número de classes e da localização dessas classes no espaço de atributos, bem como também para determinar a classe que um dado *pixel* pertence. Um analista humano poderá, a posteriori associar o conjunto de *pixels* de cada classe a um determinado padrão que se deseje reconhecer em uma imagem. Estes procedimentos costumam ser computacionalmente caros. Para a nossa aplicação, os classificadores supervisionados são mais utilizados já que sabe-se, *a priori* a quantidade de classes que se deseja classificar os *pixels*.

### 2.3.2 Classificação Supervisionada

As técnicas de classificação supervisionadas são as ferramentas analíticas utilizadas para a extração de informação quantitativa dos dados

das imagens. Diversas técnicas de classificação supervisionadas estão presentes na literatura, sendo que as técnicas estatísticas são as mais freqüentemente encontradas e fazem parte das técnicas analisadas neste trabalho.

As técnicas estatísticas supõem que as observações oriundas de cada classe são ocorrências de variáveis aleatórias com distribuições de probabilidade diferentes no espaço multiespectral. Dessa forma, teremos uma distribuição multivariada (Miller 1964), onde o número de componentes corresponde ao número de dimensões do espaço de atributos. A distribuição descreve a probabilidade de um dado *pixel* pertencer a uma dada classe. A distribuição de probabilidade mais utilizada para classificação é a distribuição gaussiana; ela nos fornece um modelo matemático tratável para auxiliar no processo de classificação.

A Figura 2.8 mostra um espaço multiespectral onde cada classe é modelada por uma gaussiana bidimensional (ver Seção 4.5.3 para maior detalhamento sobre estas distribuições). Essas curvas são chamadas “curvas de nível”, e cada curva é associada a um valor equiprovável do *pixel* pertencer à classe.

A região geométrica formada pelos pontos onde duas densidades de probabilidade possuem o mesmo valor é conhecida como “fronteira de decisão”. Essas regiões possuem, em geral, formas arbitrárias mas no caso gaussiano elas são sempre hiperquádricas (ver Duda et al. 2001, p. 36)

Uma distribuição gaussiana bivariada é totalmente especificada pelo vetor de médias  $\mu$  e pela matriz de covariância  $\Sigma$ ; se esses parâmetros são conhecidos para todas as classes, a classificação se dá calculando as probabilidades do dado *pixel* pertencer a cada uma das classes, sendo que a classe que apresentar a maior probabilidade será a mais apropriada para a classificação do *pixel*. Pelo fato de considerar que as probabilidades a priori de cada *pixel* pertencer a cada classe serem iguais, a regra é conhecida como “classificador de máxima verossimilhança”. Caso fossem consideradas probabilidades a priori diferenciadas, seria um classificador bayesiano. Este tipo de classificador

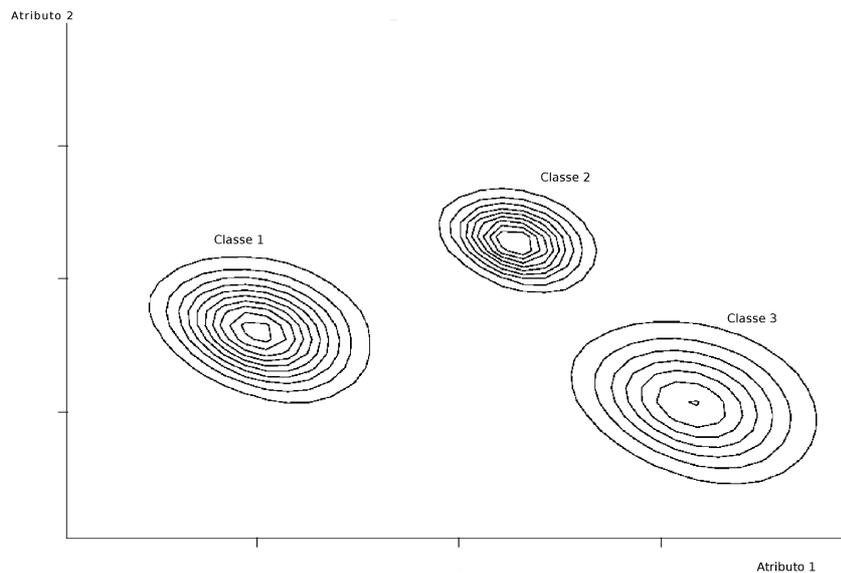


Figura 2.8: Curvas de nível de distribuições gaussianas bidimensionais

supervisionado é paramétrico pois devemos conhecer ou estimar os parâmetros da distribuição. Esses parâmetros são em geral estimados a partir de uma base de *pixels* de treinamento, determinando  $\hat{\mu}$  e  $\hat{\Sigma}$ . A metodologia para o treinamento utilizada neste trabalho está descrita na Seção 4.2. Após o treinamento, o classificador está apto a fazer a classificação para o qual ele foi treinado.

Alguns classificadores supervisionados não utilizam modelos funcionais de distribuição de probabilidade e, com isso, não necessitam passar pela fase de estimação dos parâmetros. Estes classificadores são chamados “não-paramétricos”. Neste trabalho avaliamos um estimador não paramétrico baseado no trabalho de Jones & Rehg (2002), que utiliza histogramas da cor da pele e da cor de não-pele. Este classificador está descrito na Seção 4.5.

Um outro tipo de classificador supervisionado utilizado neste trabalho é o baseado em regras que descrevem geometricamente a amostra de pele. Essas regras geométricas descrevem o agrupamento de pele e são

extraídas, geralmente, por técnicas de *data mining*. Esse classificador também é considerado não-paramétrico e avaliamos os classificadores descritos nos trabalhos de Kovac et al. (2003) e o de Rein-Lien et al. (2002).

Após a especificação dos classificadores, é desejável uma metodologia para a avaliação de desempenho da qualidade da classificação. Para isto, a próxima seção descreve a metodologia utilizada neste trabalho.

## 2.4 Avaliação de Classificadores

Neste trabalho iremos comparar a classificação dos *pixels* de imagens digitais tipicamente encontradas na Web. Esta classificação se dará entre os *pixels* que pertencem à classe de pele e os *pixels* que pertencem à classe de não-pele. Para realizarmos essa comparação utilizamos uma metodologia confiável e sistemática que iremos descrever nesta seção. Os conceitos fundamentais para a avaliação e posterior comparação dos classificadores de duas classes (pele e não-pele) serão formalizados. Esses conceitos podem ser encontrados mais detalhadamente em Hamilton et al. (2002).

Uma das ferramentas mais utilizadas para a avaliação de classificadores é a matriz de confusão (Fukunaga 1972). A matriz de confusão contém informações sobre os valores reais e os valores encontrados pelos classificadores, ou seja, contém estatísticas sobre os acertos e os erros dos classificadores. Na Tabela 2.2, temos a especificação de uma matriz confusão para classificadores de duas classes.

Tabela 2.2: Especificação de uma Matriz de Confusão

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>a</b>	<b>b</b>
	Positivo	<b>c</b>	<b>d</b>

Na Tabela 2.2 as entradas estão associadas aos seguintes valores:

- a:** é o número de classificações corretas da classe considerada como negativo;
- b:** é o número de classificações incorretas da classe considerada como positivo (erro de omissão);
- c:** é o número de classificações incorretas da classe considerada como negativo (erro de comissão);
- d:** é o número de classificações corretas da classe considerada como positivo.

Para a aplicação que esse trabalho se propõe, um classificador de duas classes será utilizado, onde teremos as classes pele e não-pele. Dessa forma, uma matriz confusão exatamente com a mesma estrutura que a apresentada na Tabela 2.2 será utilizada. Para o nosso caso, utilizaremos a classe pele como a classe considerada como positivo e a classe não-pele como a classe considerada como negativo. Assim, estaremos interessados em controlar a qualidade do classificador de pele, tendo não-pele como a classe complementar.

Diversos termos são definidos a partir dos valores da matriz de confusão para um classificador de duas classes. Esses termos estão definidos a seguir:

- ✓ **Acurácia:**  $AC = \frac{a+d}{a+b+c+d}$ ;
- ✓ **Verdadeiros Positivos:**  $TP = \frac{d}{c+d}$ ;
- ✓ **Falsos Positivos:**  $FP = \frac{b}{a+b}$ ;
- ✓ **Verdadeiros Negativos:**  $TN = \frac{a}{a+b}$ ;
- ✓ **Falsos Negativos:**  $FN = \frac{c}{c+d}$ ;
- ✓ **Precisão:**  $P = \frac{d}{b+d}$ .

O parâmetro acurácia pode não ser interessante para avaliação de desempenho de um classificador quando o número de casos negativos é muito maior do que o número de casos positivos. Por exemplo, vamos supor que temos 1000 casos, 995 são casos negativos e 5 são positivos. Se o sistema classificar todos como negativo, a acurácia será de 99,5%, mesmo o classificador não acertando nenhum dos casos positivos. Outros possíveis valores para avaliação de desempenho de classificadores estão listados abaixo:

✓ **Média Geométrica:**  $g_1 = \sqrt{TP \times P}$

✓ **Média Geométrica:**  $g_2 = \sqrt{TP \times TN}$

✓ **Medida F:**  $F = \frac{(\beta^2 + 1)P \times TP}{\beta^2 P + TP}$ , com  $\beta \in \mathbb{R}_+$ .

Estas medidas estão apresentadas em Hamilton et al. (2002) e valem zero quando nenhum dado da categoria positivo é classificado corretamente.

Uma outra forma de avaliação de desempenho de classificadores amplamente utilizada nos trabalhos de detecção de pele é através das curvas ROC (*Receiver Operating Characteristic*). As curvas ROC são gráficos que apresentam a taxa de falsos positivos no eixo das abscissas e a taxa de verdadeiros positivos no eixo das ordenadas. O ponto (0, 1) representa o classificador perfeito. O ponto (0, 0) representa o classificador que classifica todos os casos como negativos e o ponto (1, 1) representa o classificador que classifica todos os casos como positivos.

Alguns classificadores possuem um parâmetro, ou limiar de controle, que pode ser ajustado para aumentar TP ao custo de aumentar FP, ou para diminuir FP ao custo de diminuir TP. Cada valor do limiar de controle produz um par (FP, TP), e um conjunto desses pares pode ser utilizado para construir a curva ROC. Um classificador de regra fixa ou sem limiar de controle é representado como um ponto correspondente ao par (FP, TP) do classificador; por simplicidade, diremos que esse ponto é também uma curva ROC. As curvas ROC para os classificadores avaliados nesse trabalho estão apresentadas no Capítulo 5.

Um exemplo de quatro curvas ROC é mostrado na Figura 2.9. Nesta Figura, temos os classificadores de regra fixa  $\diamond$  e  $\triangle$ , e os outros classificadores representados pelas linhas vermelha e azul. Mais adiante será discutido como comparar curvas ROC.

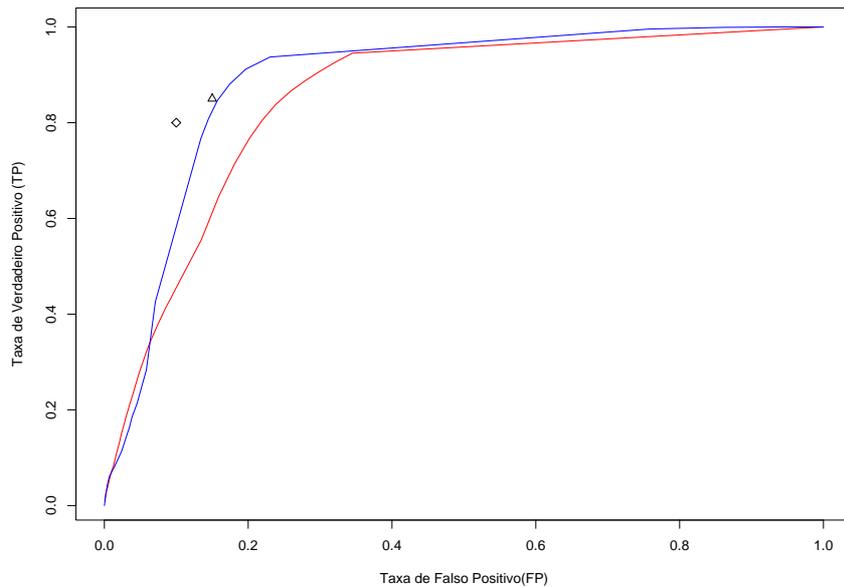


Figura 2.9: Curva ROC para quatro classificadores

Uma curva ROC possui algumas características interessantes para a avaliação de classificadores e são amplamente utilizadas nos classificadores utilizados para detecção de pele; por exemplo uma curva ROC encapsula todas as informações contidas na matriz de confusão, uma vez que FN e TN são os complementos de TP e FN respectivamente. Além disso, as curvas ROC provêm uma forma visual para examinar a habilidade do classificador para classificar corretamente os casos positivos e para visualizar o compromisso entre o variação de FN e a respectiva variação de TP.

Hamilton et al. (2002) apresentam duas abordagens para comparação de classificadores através das curvas ROC: (i) área sob a curva, e (ii) distância euclidiana. Incidentalmente, convém notar que a defini-

ção de distância euclidiana constante na referência está errada, sendo a correta a dada pela Equação (2.5).

A área sob a curva é uma medida simples de comparação entre os classificadores mas não contempla os classificadores que não fazem uso do limiar de controle, enquanto a distância euclidiana permite a comparação entre quaisquer tipos de classificador. A definição corrigida da distância euclidiana ( $AC_d$ ) apresentada por Hamilton et al. (2002) está mostrada na Equação (2.5).

$$AC_d = \sqrt{W(1 - TP)^2 + (1 - W)FP^2}, \quad (2.5)$$

onde o valor  $0 \leq W \leq 1$  é utilizado para atribuir relevância para os falsos positivos ou para os falsos negativos.

Pode-se verificar que  $AC_d$  vale zero para um classificador perfeito e  $\sqrt{2}$  para um classificador que classifica todos os casos incorretamente. Para um classificador que utiliza limiar de classificação  $AC_d$  será o menor valor dentre os vários pares (FP, TP).

Na próxima seção, iremos apresentar os diversos espaços de cor utilizados nesse trabalho e suas principais características.

## 2.5 Espaços de Cor

Diversos espaços de cor vêm sendo utilizados em diferentes aplicações nas áreas de Colorimetria, Computação Gráfica, Transmissão de Vídeo e Visão Computacional. Cada espaço de cor apresenta diferentes propriedades e características úteis ou nocivas para cada aplicação. Nessa seção, descreveremos os espaços de cor que foram utilizados em nosso trabalho com a finalidade de detecção de pele.

### 2.5.1 RGB

O espaço RGB é o espaço de cor utilizado pelo Monitor de Tubos de Raios Catódicos (monitor comumente encontrado em computadores).

Neste espaço de cor, as cores são geradas a partir da combinação de três cores básicas, o vermelho, o verde e o azul (*red* (R), *green* (G) e *blue* (B) em língua inglesa). É um dos espaços de cor mais utilizados em processamento e armazenamento de imagens digitais. Este espaço é o cubo de lados  $[0, \max_R]$ ,  $[0, \max_G]$  e  $[0, \max_B]$  e cada cor neste espaço é um ponto nesse cubo, isto é, uma cor é um ponto da forma (R,G,B) onde  $0 \leq R \leq \max_R$ ,  $0 \leq G \leq \max_G$  e  $0 \leq B \leq \max_B$ . Usualmente  $\max_R = \max_G = \max_B = K$ , com  $K = 1$  ou  $K = 255$ .

### 2.5.2 Cor RGB Normalizada

Esta cor é obtida a partir do espaço RGB através da transformação  $(R, G, B) \mapsto (r, g, b)$  dada por

$$r = \frac{R}{R + G + B} \quad g = \frac{G}{R + G + B} \quad b = \frac{B}{R + G + B},$$

com isso a soma das três componentes da cor normalizada é  $r + g + b = 1$ .

Apenas duas componentes possuem informação sobre todo o espaço de cor RGB) sendo que uma delas pode ser omitida sem perda de informação reduzindo dimensionalidade. As componentes  $r$  e  $g$  são conhecidas como “cores puras” pois a normalização diminui a dependência de  $r$  e  $g$  do brilho. Uma importante propriedade deste espaço de cor é que, sob certas condições, ignorando a iluminação ambiente, as componentes do espaço de cor RGB Normalizado permanecem constantes perante variações da orientação de uma superfície lisa em relação à fonte de luz irradiante. Esta propriedade, aliada ao baixo custo computacional envolvido, tornou este espaço de cor popular entre pesquisadores.

### 2.5.3 XYZ

O espaço de cor XYZ é o espaço de cor correspondente aos três estímulos percebidos pelos fotoreceptores do nosso sistema visual adotado pela CIE e adota as seguintes premissas:

1. As componentes de cor devem ser positivas;
2. Devemos obter o maior número possível de cores espectrais com algumas coordenadas de cromaticidade nulas;
3. Duas primárias devem ter luminância zero.

A Figura 2.10 mostra o espaço de cromaticidade no espaço de cor XYZ.

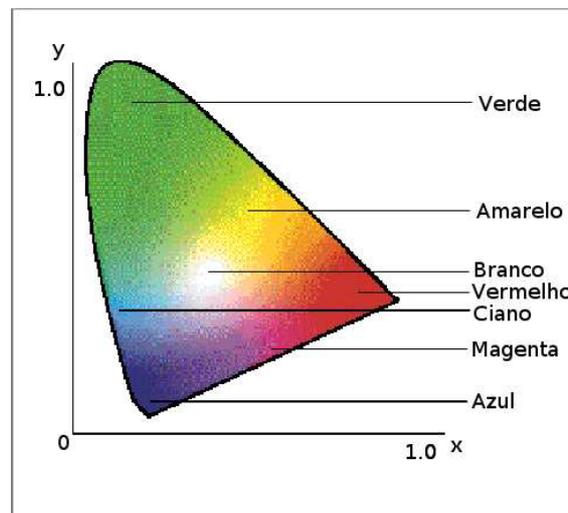


Figura 2.10: Espaço de cromaticidade XYZ

A transformação entre RGB e XYZ utilizada neste trabalho está apresentada na Equação (2.6).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (2.6)$$

onde supomos  $\max_R = \max_G = \max_B = 1$ . Diversos valores das constantes da matriz de transformação podem ser utilizados dependendo do branco de referência adotado. Neste trabalho foi utilizado o padrão D65 (Poynton 2002).

### 2.5.4 HSV

Os espaços de cor baseados em Matiz-Saturação (*Hue-Saturation*) foram introduzidos pela necessidade de especificar as propriedades perceptivas da cor de forma numérica. Estes espaços descrevem a cor através de atributos fortemente relacionados com a percepção da cor, baseados nos conceitos artísticos matiz, saturação e tonalidade. A Matiz (*Hue*) define a cor dominante (como vermelho, verde, azul, ou uma mistura de duas delas) de uma área. A saturação mede a coloração de uma área proporcional ao seu brilho, e varia de um cinza neutro, passando pelos tons pastéis até as cores totalmente saturadas. A componente *Value* é relativa a iluminação. A Figura 2.11 ilustra a o espectro de cor representado pelo espaço HSV.

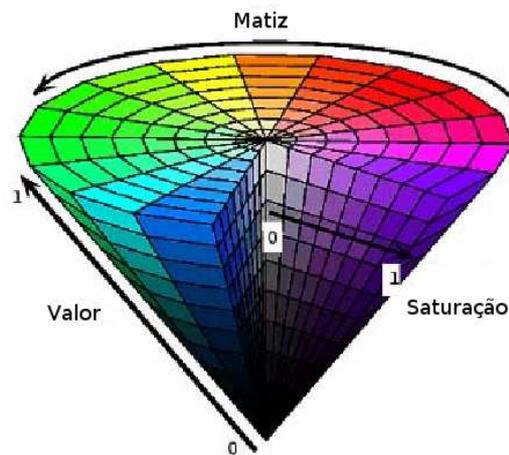


Figura 2.11: Espaço de cor HSV  
Fonte: Skarbek & Koschan (1994, p. 43)

A intuitividade das componentes e a discriminação entre os componentes de luminância e crominância fazem desse espaço de cor populares nos trabalhos de detecção de pele. Diversas propriedades interessantes da componente matiz como a invariância da matiz quando submetida a iluminação através de luz branca e também a invariância a orientação de iluminação em superfícies lisas foram destacadas por Skarbek & Koschan (1994).

Porém, Poynton (2002) destaca diversas propriedades indesejáveis como a descontinuidade da componente matiz quando definida sobre todos os reais, e o fato de  $V$  (equação (2.7)) não corresponder exatamente com a percepção do brilho pelo sistema humano de visão. As equações (2.7), (2.7) e (2.7) descrevem a transformação do espaço de cor RGB (no cubo unitário) para o HSV.

$$\begin{aligned} H &= \arccos \frac{\frac{1}{2}((R-G) + (R-B))}{\sqrt{((R-B)^2 + (R-B)(G-B))}}, \\ S &= 1 - 3 \frac{\min(R, G, B)}{R + G + B}, \\ V &= \frac{R + G + B}{3}. \end{aligned}$$

O sistema de coordenadas polares do plano Matiz-Saturação (HS) resulta numa natureza cíclica do espaço de cor. Esta característica pode ser inconveniente para modelos de pele paramétricos onde é necessária fazer agrupamentos de cores para uma melhor classificação. Esta dificuldade pode ser contornada utilizando uma transformação para coordenadas cartesianas através da Equação (2.7):

$$X = S \cos H, \quad Y = S \sin H. \quad (2.7)$$

### 2.5.5 YCbCr

O espaço de cor YCbCr é caracterizado por uma transformação afim do espaço RGB, bastante utilizada pela padronização do sinal de TV da Europa. Neste espaço, as cores são representadas pelas componentes Luma ( $Y$ ) e por duas componentes de crominância como apresentado na Equação 2.8.

$$\begin{aligned} Y &= 0.257R + 0.504G + 0.098B + 16, \\ Cr &= R - Y, \\ Cb &= B - Y. \end{aligned}$$

Este espaço supõe que as componentes Cb e Cr possuem a informação sobre a crominância e a componente Y possui a informação sobre luminância, sendo, por este motivo, atrativo para alguns pesquisadores como Rein-Lien et al. (2002).

### **2.5.6 Sistemas de Cor de Percepção Uniforme**

A denominação “cor da pele” não é uma propriedade física do objeto de interesse e sim um fenômeno de percepção visual dos seres humanos, sendo assim um conceito subjetivo. Entretanto, o uso de representações de cor semelhantes à sensibilidade do sistema de visão dos seres humanos pode ajudar a obter uma alta desempenho nos algoritmos de detecção de pele.

O espaço de Cor CIELAB descreve um sistema de percepção uniforme tal como os propostos por Wyszecki (1986) e padronizados por CIE (*Commission Internationale de L'Eclairage*); este espaço de cor será detalhado na Seção 2.5.7. Os termos “Sistema de Cor de Percepção Uniforme” fazem referência a que uma pequena perturbação num valor de uma componente do espaço de cor é igualmente perceptível ao longo do intervalo deste valor. Para esclarecer este conceito, considere o seguinte exemplo (Poynton 2002): O controle de volume de um rádio é projetado para ser perceptualmente uniforme, ou seja, um aumento no volume será igualmente perceptível em termos de aumento de intensidade sonora independente do volume inicial ser baixo ou alto. Se o controle de volume fosse fisicamente linear, a natureza logarítmica da percepção sonora do ouvido humano tornaria a percepção sonora mais intensa no início da escala do controle.

O espaço de cor RGB não é caracterizado como de percepção uniforme. As transformações não-lineares apresentadas no espaço CIELAB procuram corrigir essa situação. Essas transformações são complexas e computacionalmente mais caras do que as transformações para os outros espaços de cor apresentados neste trabalho.

### 2.5.7 CIELAB

Dois problemas referentes aos espaços de cor “tri-estímulos” e aos espaços de cromaticidade são citados na literatura. O primeiro problema é que as suas especificações não são facilmente relacionadas com dimensões psicofísicas, e o segundo problema é que esses espaços de cor não são perceptualmente uniformes.

Em busca de solucionar eses problemas, a CIE especificou dois espaços de cor sendo o primeiro mais útil para especificar cores de objetos que irradiam iluminação e o segundo para as cores de superfícies (que refletem a iluminação). O CIELAB é o espaço de cor que representa a segunda categoria.

O CIELAB especifica a percepção de cores em termos de um espaço tridimensional. O eixo L, conhecido como o de iluminação, apresenta valores no intervalo  $[0, 100]$ , sendo o valor 100 para a cor branca e 0 para a cor preta. Os outros dois eixos, A e B, são relativos à crominância; o primeiro representa as variações do vermelho e do verde (vermelho nos valores mais baixos e verde nos valores mais altos) e o segundo do amarelo para o azul (amarelo para os valores mais baixos e azul para os valores mais baixos).

Sendo assim, valores de  $A = 0$  e  $B = 0$  são acromáticos e representam a escala de cinza desde o preto até o branco.

A Figura 2.12 ilustra as informações anteriores onde vemos um corte do espaço tridimensional CIELAB e podemos perceber as variações de cor dos eixos A e B e a variação de luminância do eixo L.

A transformação do XYZ para o CIELAB é dada pela Equação 2.8

$$l = 116fY - 16 \quad a = 500(fX - fY) \quad b = 200(fY - fZ) \quad (2.8)$$

onde,

$$f\xi = \begin{cases} \left(\frac{\xi}{\xi_n}\right)^{1/3} & \text{se } \frac{\xi}{\xi_n} > 0.008856 \\ \frac{7.787\xi}{\xi_n} + \frac{16}{116} & \text{caso contrário.} \end{cases}$$

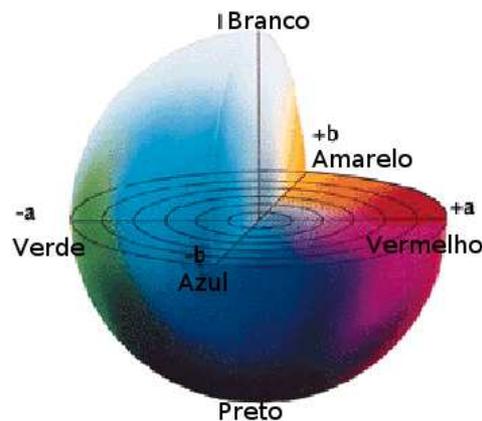


Figura 2.12: Espaço de cor CIELAB

onde  $\xi \in \{X, Y, Z\}$  e  $\xi_n \in \{X_n, Y_n, Z_n\}$ . As constantes  $[X_n, Y_n, Z_n]$  são os valores do branco de referência.

## 2.6 Transformação de Componentes Principais

De acordo com o apresentado na Seção 2.5, a maioria dos espaços de cor utilizados para detecção de pele é definida em um espaço tridimensional. Com o objetivo de diminuição do custo computacional e melhorar a tratabilidade do modelo matemático para detecção de pele, se faz necessária a redução de dimensionalidade. Esta etapa irá diminuir a complexidade dos modelos estatísticos utilizados nesse trabalho pois, com o aumento da quantidade de dados, esses modelos se tornam mais complexos e computacionalmente mais caros. Algumas propostas de redução de dimensionalidade envolvem apenas a transformação entre espaços de cor e o descarte da componente de luminância. Essa abordagem nem sempre produz resultados bons (uma discussão mais ampla é encontrada em Rein-Lien et al. 2002); por este motivo, avaliamos também a transformação de componentes principais.

A transformação de componentes principais, também conhecida como transformação de Karhunen-Loève, é uma das ferramentas mais poderosas para o tratamento de objetos multidimensionais, por exemplo

sinais. Ela é uma transformação entre espaços  $\gamma$ -dimensionais, mas para os propósitos deste trabalho será suficiente descrevê-la para espaços tridimensionais dado que as imagens aqui tratadas são pictóricas, isto é, elas originalmente estão armazenadas nas suas componentes R, G e B.

Os canais de uma imagem multiespectral são freqüentemente altamente correlacionados, de modo que as bandas sejam similares tanto visual quanto numericamente, caracterizando a redundância espectral. A análise de imagens por suas bandas espectrais individuais pode ser ineficiente devido a esta redundância.

A transformação de componentes principais é uma técnica de realce que procura reduzir a redundância espectral. Isto é feito pela geração de uma nova imagem cujas bandas individuais apresentam informações não disponíveis em outras bandas (ver Richards & Xiuping 1999, INPE 2006).

No que segue será apresentada esta transformação dando ênfase à aplicação visada neste trabalho, isto é, imagens pictóricas compostas por três bandas. Convem notar, no entanto, que esta transformação pode ser aplicada a imagens com qualquer número de bandas acima de dois.

Esta transformação é derivada da matriz de covariância entre as bandas e gera um novo conjunto de imagens, de modo que cada valor de *pixel* é uma combinação linear dos valores originais. O número de componentes principais é igual ao número de bandas espectrais da imagem original e podem ordenar-se pela da variância de cada componente de cor; assim, a primeira componente principal pode ter a maior variância e a última a menor variância.

A análise de componentes principais é uma técnica estatística poderosa que pode ser utilizada para redução do número de variáveis e para fornecer uma visão estatisticamente privilegiada do conjunto de dados. Dessa forma, a técnica que iremos apresentar será útil tanto para redução de dimensionalidade, onde o espaço 3D (geralmente RGB) de uma imagem pode ser reduzido para 2D, desprezando a componente

principal com menor variância (componente que carrega menos informação), como também podemos utilizá-la para escolha do conjunto de características mais relevantes extraídas das imagens com a finalidade de obter uma boa classificação.

A Figura 2.13 ilustra uma transformação de componentes principais em duas dimensões que corresponde à rotação do eixo original para coincidir com os eixos de máxima e mínima variância dos dados. Nesta figura temos que  $C_{p_1}$  corresponde à primeira componente principal (a que tem a maior variância) e que  $C_{p_2}$  corresponde à segunda componente principal (a que tem a menor variância) e que “Canal 1” e “Canal 2” correspondem aos dados originais. Os eixos pontilhados correspondem as direções para onde apontam os autovetores da matriz de covariância dos dados originais.

A transformação de componentes principais pode ser descrita nas seguintes etapas:

1. Calcular  $\Sigma$ , a matriz de covariância dos dados (suponhamos que ela seja positiva definida pois estamos tratando de variâncias).
2. Decompor  $\Sigma$  nos autovetores  $U$  e autovalores  $\lambda$ . Essa matriz será diagonalizável uma vez que a matriz de covariância é positiva definida (Krzanowski 1995).
3. Calcular a nova imagem *pixel a pixel*; multiplicando o valor de cada *pixel* pela matriz dos autovetores.

Os autovalores representam o comprimento dos eixos das componentes principais de uma imagem e são medidos na unidade da variância. Associado a cada autovalor, existe um vetor de módulo unitário chamado autovetor. Os elementos de cada autovetor são fatores de ponderação que definem a contribuição de cada banda original para uma componente principal, numa combinação linear (INPE 2006). Os autovetores representam as direções dos eixos das componentes principais.

A formalização matemática do processo supracitado pode ser vista da seguinte maneira: a transformação de componentes principais da

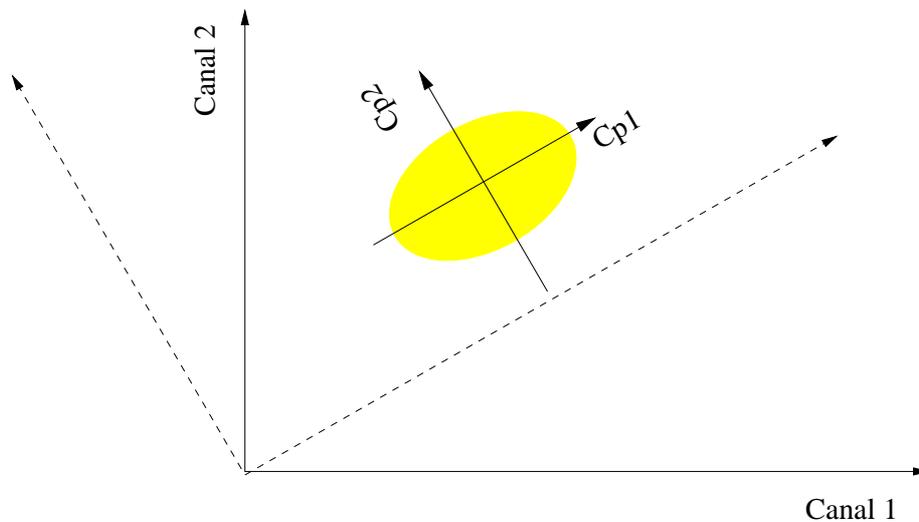


Figura 2.13: Transformação de Componentes Principais (2D)

imagem  $f: S \rightarrow \mathbb{R}^p$  consiste em calcular a imagem  $g: S \rightarrow \mathbb{R}^p$  onde  $g(s) = (g_1(s), \dots, g_p(s))$  e para todo  $s \in S$  temos que

$$g_i(s) = \sum_{1 \leq j \leq p} \alpha_{i,j} f_j(s),$$

onde para cada  $1 \leq i \leq p$ ,  $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,p})$  é o autovetor  $i$  da matriz de covariância da imagem  $f$ .

Desta maneira, a transformação de componentes principais de uma imagem de  $p$  bandas gera uma nova imagem de  $p$  bandas definida sobre o mesmo suporte. A transformação se realiza sobre o espaço de atributos.

Veremos a seguir, na Figura 2.14, um exemplo de aplicação da transformação de componentes principais, aplicada a uma imagem pictórica. As Figuras 2.14(a), 2.14(b) e 2.14(c) mostram as componentes R, G e B da imagem original, enquanto as Figuras 2.14(d), 2.14(e) e 2.14(f) mostram as três componentes principais derivadas, em ordem decrescente de variância. Nesta figura, foi utilizada a matriz de covariância da imagem original para realizar a transformação de componentes principais com o objetivo de ilustrar a redistribuição de informações entre os ca-

nais.

É interessante notar como a informação nos dados originais está espalhada entre as três bandas. Já nos dados transformados ela está mais concentrada na primeira banda e muito mais ausente na terceira. Há, de fato, uma redistribuição da informação; é notável como na primeira componente há informação do céu (nuvens pouco visíveis na banda R, mas aparentes na banda G) e mais variabilidade nos tons de pele (que aparece mais escura e menos contrastada na banda G).

## 2.7 Análise de Dados

Neste trabalho iremos utilizar e avaliar diversas técnicas estatísticas para análise de imagens digitais, bem como trataremos de diversos modelos estatísticos para cor de pele. Alguns conceitos relacionados à análise de dados são necessários para o entendimento da utilização dessas técnicas. Os principais conceitos relativos a esta área estão apresentados nessa seção.

Seja  $\mathbf{y} = (y_1, \dots, y_n)$  uma amostra e  $\mathbf{y}_\bullet = (y_{1:n}, \dots, y_{n:n})$  a amostra  $\mathbf{y}$  ordenada do menor para o maior, isto é,  $y_{1:n} \leq y_{2:n} \leq \dots \leq y_{n:n}$ . Segundo Bustos & Frery (1992), uma análise descritiva de dados deve conter, pelo menos, as seguintes informações a respeito da amostra  $\mathbf{y}$ :

- o tamanho e os valores mínimo e máximo da amostra  $\mathbf{y}$ , que são, respectivamente,  $n$ ,  $y_{\min} = y_{1:n}$  e  $y_{\max} = y_{n:n}$ ;
- a média amostral e o desvio padrão amostral, que são, respectivamente,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \text{ e } \hat{\sigma} = \sqrt{\widehat{\text{Var}}(\mathbf{y})},$$

onde  $\widehat{\text{Var}}(\mathbf{y}) = n^{-1} \sum_{i=1}^n (y_i - \bar{y})^2$  é denominada a variância amostral. A variância é uma medida de dispersão dos valores de uma variável aleatória em torno da média;



(a) Banda R, original

(b) Banda G, original

(c) Banda B, original



(d) Primeira componente

(e) Segunda componente

(f) Terceira componente

Figura 2.14: Bandas das imagens original e transformada

- a mediana amostral, que é dada por

$$q_{\frac{1}{2}}(\mathbf{y}) = \begin{cases} y_{(n+1)/2:n} & \text{se } n \text{ for ímpar,} \\ (y_{n/2:n} + y_{n/2+1:n})/2 & \text{caso contrário;} \end{cases}$$

- os quartis inferior e superior dados, respectivamente, por

$$q_{\frac{1}{4}} = y_{[n/4]:n} \quad \text{e} \quad q_{\frac{3}{4}} = y_{[3n/4]:n};$$

utilizamos a notação usual de colchetes para o inteiro mais próximo, isto é,  $[x]$  é a parte inteira de  $x + \frac{1}{2}$ ;

- a assimetria amostral e o excesso de curtose amostral, que são, respectivamente,

$$\hat{\gamma}_1 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^3 / \hat{\sigma}^3 \quad \text{e} \quad \hat{\gamma}_2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^4 / \hat{\sigma}^4 - 3.$$

A assimetria amostral é o grau de desvio, ou afastamento da simetria, de uma amostra, e a curtose amostral é o grau de achatamento da amostra em relação à lei gaussiana;

- e o histograma, que é um dos elementos gráficos mais importantes para descrever uma amostra de valores reais. Esta função define-se contando o número de observações que estão contidas em intervalos, ou seja, o histograma da amostra  $\mathbf{y}$  associado a uma partição  $\mathcal{J}$  é a função  $(x_\iota, \mathbf{H}(\mathbf{y}, \iota))$ , onde  $\mathbf{H}(\mathbf{y}, \iota) = \#\{\ell : y_\ell \in I_\iota\}$  com  $0 \leq \iota \leq k$ . A partição  $\mathcal{J} = \{I_0, \dots, I_k\}$  do intervalo  $I$  é definida de tal forma que  $[y_{\min}, y_{\max}] \subseteq I$ , e  $x_\iota$  é o ponto central de cada intervalo  $I_\iota$  para todo  $0 \leq \iota \leq k$ .

Uma outra descrição gráfica útil de um conjunto de dados é o *boxplot*, cujos elementos são mostrados na Figura 2.15 e onde IQR denota a distância interquartil dada por  $q_{\frac{1}{4}} - q_{\frac{3}{4}}$  e os pontos denotados com ‘•’ são considerados suspeitos ou *outliers* no corpo da amostra.

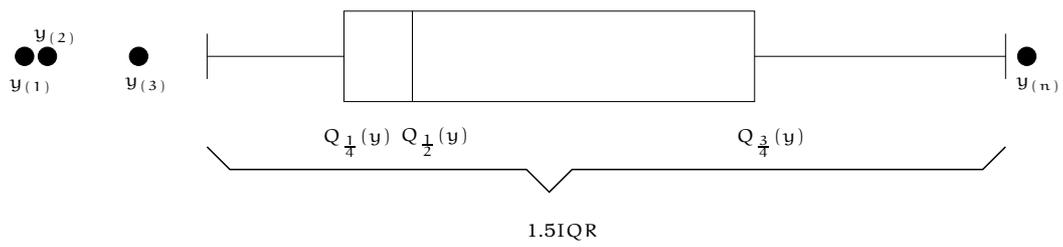


Figura 2.15: Elementos de um boxplot

No próximo capítulo, iremos apresentar o sistema de detecção de imagens pornográficas que o sistema de detecção de pele deste trabalho será acoplado.

## Capítulo 3

# Sistema de Detecção de Imagens Pornográficas

Este capítulo discute a necessidade de detecção de imagens pornográficas na Web e descreve uma proposta para um sistema de detecção dessas imagens. A Seção 3.1 cumpre o primeiro objetivo deste capítulo e a Seção 3.2 cumpre o segundo objetivo.

### 3.1 Motivação para detecção de imagens pornográficas

O crescimento da Internet tem proporcionado uma grande mudança nos hábitos diários da população mundial. Serviços como *Internet Home Banking*, Comércio Eletrônico e Governo Eletrônico, entre outros tantos, têm feito com que as pessoas executem tarefas do cotidiano de forma ágil e eficiente.

As Comunidades Virtuais da Internet têm servido para agrupar pessoas em torno de interesses específicos para debates, encontros, conversas informais e diversão. Elas vêm se tornando um verdadeiro universo virtual onde muitas das atividades do dia-a-dia são realizadas de forma eficiente e prazerosa. Esses universos virtuais vêm propiciando, também, atividades impossíveis de serem realizadas no universo real.

Toda nova tecnologia traz benefícios e possíveis problemas, frequentemente ambos sempre existiram, e são apenas potencializados pelas tecnologias.

O crescente uso da Internet tem evidenciado um problema inerente a todas as mídias veiculadas (ver Resnick & Miller 1996): a possível exposição indiscriminada a materiais que não são apropriados para qualquer audiência. Este problema tem promovido debates sobre a utilização ou não de censura do material tido como ofensivo, sem que se chegue a um consenso.

O fato é que determinados conteúdos encontrados nas diversas mídias distribuídas na Internet são considerados não aceitáveis para determinadas empresas ou corporações. Para determinadas corporações ou entidades, aplicar censura com o objetivo de diminuir a procrastinação e evitar a exposição imprópria a materiais inadequados parece ser uma atitude aceitável.

Imagens pornográficas ou com conteúdo ofensivo podem ser enquadradas neste tipo de mídia.

A veiculação de material impróprio não é uma prerrogativa nem uma novidade decorrente da Internet. A exibição de conteúdo erótico ou pornográfico é recorrente em praticamente todas as culturas, mas como esse material é tratado e avaliado varia entre a plena aceitação e a condenação absoluta.

Segundo Chan et al. (1999), o tráfego indevido é danoso pois ele ocasiona negação de serviço e fere a responsabilidade cibernética. A negação de serviço ocorre quando o tráfego indevido ocasiona o bloqueio, a perda ou o atraso do tráfego legítimo pela exaustão das capacidades da rede. O conceito de responsabilidade cibernética se refere a legalidade de trafegar conteúdo indecente e/ou pornográfico principalmente em corporações. A questão da legalidade é bastante polêmica sendo que países como a Inglaterra (ver detalhes em Smith 2001) já discutem o problema em sua legislação e os Estados Unidos da América através da *The Communication Decency Act (CDA), 1996*, já apresentou uma iniciativa de censura mas este ato foi revogado pela suprema corte dos

Estados Unidos em 26 de junho de 1997.

A principal dificuldade para censurar mídia imprópria é o volume de dados disponíveis na Internet, bem como a sua natureza dinâmica. Uma estratégia manual de censura seria muito dispendiosa em termos de recursos humanos. Diante dessa situação, estratégias automáticas ou semi-automáticas de censura poderiam prover uma solução desejável para este problema.

Técnicas de processamento digital de imagens, visão computacional e Inteligência Artificial podem ser aliadas para obtermos uma boa qualidade no julgamento de imagem para que seja decidido se esta irá ou não ser veiculada.

Uma solução que vem sendo empregada com relativo sucesso é a de banir o acesso a sítios Web conhecidamente provedores de conteúdo inadequado. Com isso, não é o conteúdo que é analisado senão onde ele está. Tecnologias como os servidores *proxy* (ver um exemplo em Squid 2005, SquidGuard 2005), agem filtrando acesso às páginas Web cadastradas em listas negras.

Os sistemas *proxy* têm sido muito utilizados mas geram a tarefa administrativa dispendiosa do cadastramento de todos os sítios Web aos que será proibido o acesso. Para minimizar o trabalho manual dos administradores dos sistemas, se faz uso de listas negras geradas automaticamente por robôs Web<sup>1</sup>. O grande problema desses sistemas é a grande quantidade de páginas cadastradas indevidamente e a quantidade, ainda maior, de páginas não cadastradas nas listas negras. Este fato faz com que os administradores dos *proxy* despendam muito tempo para adaptar as listas negras geradas automaticamente à realidade de suas corporações.

Um problema diferente é o que os provedores de hospedagem enfrentam ao tentar implementar políticas automáticas ou semi-automáticas de controle de conteúdo. Eles devem evitar que mídias impróprias sejam armazenadas ou publicadas sob o seu domínio. Esse problema

---

<sup>1</sup>Sistemas que varrem exaustivamente a Internet em busca de informações; são também utilizados por engenhos de busca para indexação de páginas.

é central nos álbuns virtuais e nos sistemas gerenciadores de comunidades virtuais. Um exemplo da primeira categoria é o Fotolog.net (<http://www.fotolog.net>) enquanto que para a segunda temos o Orkut (<http://www.orkut.com>).

Os filtros anti-spam (ver Vaughan-Nichols 2003) compreendem um outro tipo de tecnologia utilizada para censura de conteúdo na Internet. Eles analisam os conteúdos das caixas postais dos usuários de um determinado servidor de e-mail, e censuram o que for classificado como conteúdo não solicitado ou impróprio.

Nenhuma das tecnologias supracitadas costuma tratar imagens digitais como conteúdo a ser analisado. Os filtros anti-spam mais empregados, por exemplo, analisam apenas textos.

Vale ressaltar que o conceito de pornografia não possui fronteiras bem definidas, dificultando a operacionalização da censura de imagens, e que questões morais, políticas e legais sobre o problema estão fora do escopo desse sistema. O termo “imagem pornográfica” será usado no restante desse texto sem uma formalização do conceito de pornografia.

Na próxima seção iremos propor um sistema para automatizar a censura de imagens.

O projeto de detecção de pele, cerne desta dissertação, começou com a idéia do desenvolvimento de um sistema de detecção de pornografia denominado *TaRADO* (*Tracking and Ranking Antropomorphic Deformable Objects*). O *TaRADO* propõe uma solução de software que automatize a classificação de imagens, nos mais diversos formatos veiculados na Internet, como ofensivas ou não ofensivas. As imagens de conteúdo ofensivo em questão são aquelas que contenham nudez humana. Essa dissertação tem por objetivo, especificar, implementar e avaliar o desempenho de um sistema de detecção de pele que irá fazer parte do sistema de nudez discutido nesse capítulo.

A tarefa de detecção de nudez em imagens não é trivial devido a fatores como a diversidade de cores de pele, diferentes pontos de vista de visualização e diferentes disposições geométricas da(s) pessoa(s) que variam de acordo com a postura. Segundo Forsyth & Fleck (1996) a des-

coberta de cor e textura de pele humana por si só não redundam em bons resultados para identificação de objetos 3D. Deve-se utilizar informação a respeito das superfícies e formas e suas posições espaciais relativas para evitar as variações de objetos do mesmo tipo, as mudanças de postura e dos pontos de vista.

Tendo em vista essas dificuldades, foi projetada uma arquitetura, detalhada na Seção 3.2, que, para detecção de imagens pornográficas, utilizará primeiramente detecção de pele humana. Logo após, o TORADO utilizará métodos de captura de informações sobre a imagem em questão, utilizando técnicas de visão computacional com o intuito de encontrar padrões comuns às imagens pornográficas. Este conjunto de características servirá para alimentar um sistema qualificador de informações para que a imagem possa ser julgada. O sistema de detecção de pele proposto, implementado e avaliado nessa dissertação, comporá este sistema de nudez, sendo um módulo essencial pois todas as outras etapas usarão as informações extraídas pelo sistema de detecção de nudez proposto.

Vários sistemas de classificação automática de conteúdo presentes nas várias áreas de aplicação como os IDS (*Intrusion Detection Systems*, ver Snort Team 2004) ou em sistemas anti-spam (ver Vaughan-Nichols 2003) e também os sistemas anti-nudez, apresentam uma grande quantidade de falso negativos, para a nossa aplicação isso redundam em que os sistemas deixam passar conteúdo não desejado, e os falsos positivos, onde os sistemas utilizam uma classificação muito rígida e bloqueiam conteúdos legítimos. Para minimizar a ocorrência de falsos positivos e falsos negativos, o TORADO utilizará mecanismos de inferência que aprendem quando informados que agiram de forma errada e que possibilitam que o administrador do sistema configure o classificador como mais ou menos rígido.

## 3.2 Arquitetura

O sistema de detecção de pornografia proposto nessa dissertação, funcionará como um módulo que poderá ser acoplado a diversos outros sistemas como:

**Servidor Web:** Um servidor web poderá fazer uso deste sistema nos casos em que pretende-se filtrar o conteúdo armazenado nele. Esta funcionalidade será útil em casos nos que os servidores hospedam conteúdo; são exemplos desses servidores: os albuns digitais na Web, as comunidades virtuais na Web e os provedores de hospedagem de conteúdo Web.

**Servidor MMC:** Os servidores de mensagem multimídia para as redes de telefonia celular (MMC) funcionam como intermediadores das mensagens enviadas através dos aparelhos celulares dos usuários da telefonia celular. Um módulo de filtragem de pornografia serviria para impedir que fotos pornográficas fossem trocadas, quando isto se fizer necessário.

**Servidor de E-mail:** Um dos maiores problemas enfrentados pelo serviço de e-mail oferecido na rede pública de dados (Internet) é a prática indiscriminada de *spam*, essas mensagens não solicitadas, muito frequentemente, apresentam conteúdo pornográfico que poderá ser detectado por este módulo e avisado ao servidor de e-mail que não fará a entrega da mensagem na caixa postal do usuário.

**Servidor Proxy:** Em diversas situações, como no ambiente corporativo, o controle de acesso à Web é aplicado e o sistema de detecção poderá informar aos servidores *proxy* a presença de pornografia em um sítio Web de modo que este servidor *proxy* bloqueie o acesso àquele sítio.

**Sistemas de Auditoria:** Diversos crimes digitais estão associados à pornografia; um exemplo dessa prática criminosa no mundo ocidental

é a pedofilia. Sistemas automáticos de detecção de pornografia poderiam ser úteis às entidades que são responsáveis por investigar as suspeições dessas condutas como também auxiliar na detecção deste tipo de tráfego na Internet.

A Figura 3.1 ilustra a proposta para a arquitetura do sistema ToRADO composto pelos seguintes módulos:

1. Pré-Processador de Imagem
2. Processador de Imagem
3. Extrator de Atributos
4. Qualificador

Cada módulo é descrito nas seções seguintes.

### 3.2.1 Pré-Processador de Imagem

O módulo de pré-processamento receberá a imagem proveniente do sistema que está fazendo uso do módulo ToRADO com a finalidade de:

1. abstrair o formato gráfico, e
2. detectar e prevenir ataques ao sistema

Este módulo será uma camada que tornará o sistema totalmente independente do formato gráfico a ser tratado entregando para o módulo seguinte uma seqüência de *bits* num formato RGB (*Red, Blue, Green*) que represente todo o espaço de cor das imagens em questão. Os formatos dos arquivos de entrada a serem considerados são os padrões de páginas Web, isto é, GIF, JPG e PNG. Imagens que são exibidas através de programas (Java ou Flash, tipicamente), bem como outros formatos gráficos que não são padrões da Web (TIF, BMP, EPS por exemplo), não serão considerados neste trabalho.

O outro papel importante deste módulo é a prevenção de ataques. Ataque ao sistema pode ser visto como qualquer técnica que tenha como

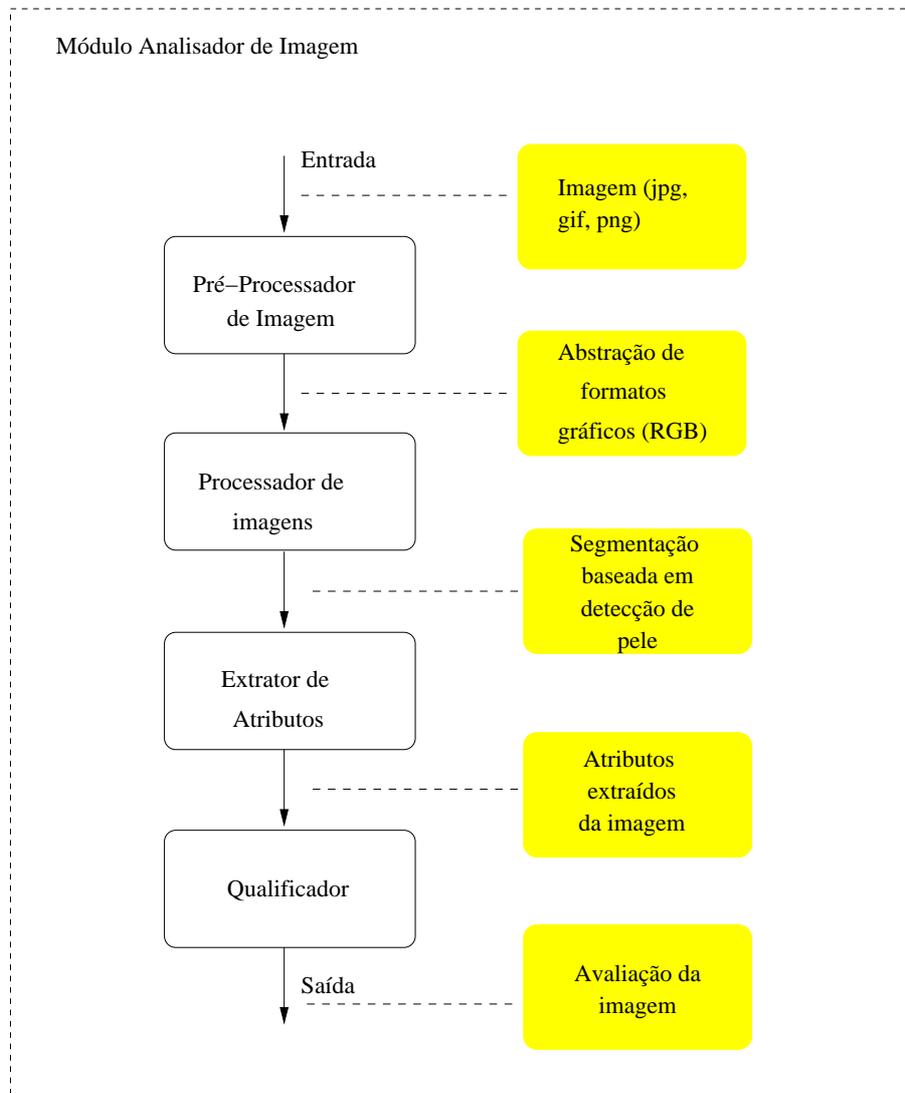


Figura 3.1: Arquitetura do sistema TarADO (Módulo analisador de imagem)

finalidade enganar a detecção de imagens pornográficas e deixar passar imagens que seriam censuradas caso esta técnica não fosse utilizada. Um bom exemplo é o ataque do mosaico (Cabral & Frery 2003) que corta a imagem em pedaços menores dificultando a ação do módulo de extração de atributos.

### **3.2.2 Processador de Imagem**

O módulo de processamento gráfico é responsável pela detecção de pele humana sendo o objeto de estudo dessa dissertação. A primeira etapa do TORADO será a detecção de pele humana.

Levando em consideração que as imagens armazenadas em hospedeiros de páginas Web apresentam características como cores de pele de várias etnias, vários tipos de iluminação, foco e ângulos de visão variados, e que não se tem controle sobre as imagens ali depositadas, técnicas que sejam robustas às variações desses parâmetros devem ser escolhidas para compor o módulo detecção de pele humana.

A técnica de detecção de pele humana a ser utilizada está descrita no Capítulo 4 e várias referências (ver Forsyth & Fleck 1996, Fleck et al. 1996, Caetano & Barone 2001, Caetano et al. 2002) tratam o problema de forma bastante objetiva. Vários modelos foram avaliados para a escolha do que apresentou o melhor desempenho quanto à qualidade da detecção.

A detecção de pele humana em imagens é uma etapa muito importante para a automatização da censura de imagens pornográficas pois os módulos subseqüentes irão agir em na imagem resultante da aplicação dessas técnicas. Falhas nessa etapa irão ser propagadas para as etapas posteriores podendo levar a erros de classificação grosseiros.

### **3.2.3 Extrator de Atributos**

Detectar pele humana nas imagens é, sem dúvida alguma, uma tarefa imprescindível para a censura de imagens que contenham nudez hu-

mana, mas não é suficiente para indicar que a imagem é imprópria. O percentual de pele humana encontrado numa imagem pode ser um bom indicador para desconfiarmos que esta imagem pode ter pessoas nuas mas outras técnicas da visão computacional devem ser utilizadas para tratar o problema.

Forsyth & Fleck (1996) utilizam agrupamento de partes do corpo humano detectadas através de aproximação do esqueleto humano por cilindros. Esta técnica produz um resultado interessante mas gera bastante falsos negativos quando alguma parte do corpo está coberta por peças de roupa ou quando a imagem não é de corpo inteiro e sim de partes do corpo humano, não permitindo o casamento de padrões com o esqueleto humano, além de apresentar um alto custo computacional.

O ToRADO pretende utilizar um conjunto maior de dados extraídos da imagem utilizando várias características extraídas da imagem, de modo que qualquer evidência de algum padrão conhecido como possível atributo de uma imagem imprópria seja utilizada como entrada de um qualificador.

Nesta etapa, já teremos a imagem com as coordenadas de pele identificadas podendo, com isso, extrair os atributos necessários à classificação. Exemplos dos atributos que serão tratados no ToRADO estão listados abaixo:

- Percentual de *pixels* detectados como pele.
- Probabilidade média dos *pixels* de pele.
- Número de componentes de pele conectados.
- Percentual de *pixels* detectados como não-pele.
- Ajuste de elipses global e local (Zheng et al. 2004).

A Figura 3.2, reproduzida de Zheng et al. (2004), apresenta a abordagem das elipses, onde a Figura 3.2(a), mostra a imagem original, a Figura 3.2(b) mostra um ajuste de uma elipse global, onde todos os

pontos associados à pele são utilizados para o ajuste da elipse e a Figura 3.2(c) mostra os ajustes de duas elipses, sendo uma para cada agrupamento; no ajuste de elipse global, a elipse representa um único objeto que se deseja reconhecer e no ajuste de elipse local, mais de um objeto é tratado, onde cada objeto é representado por uma elipse.

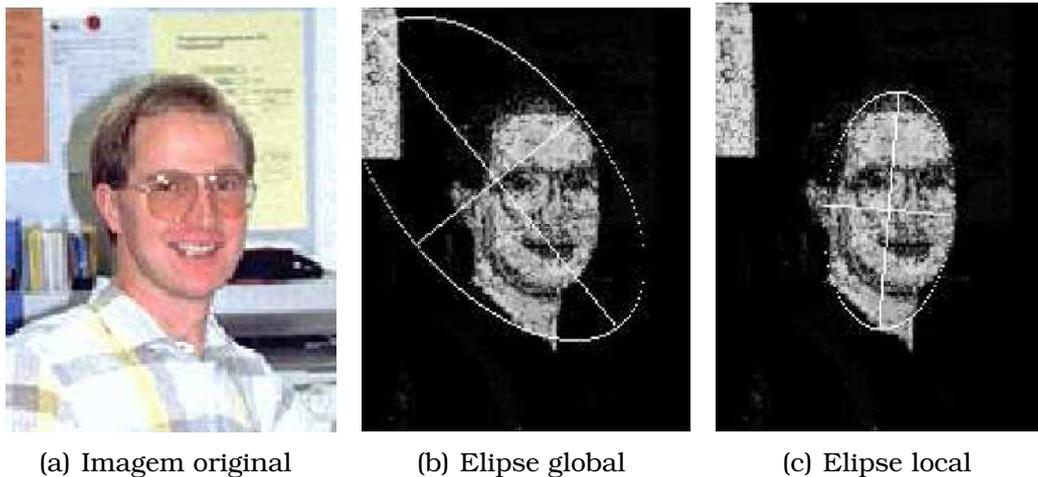


Figura 3.2: Abordagem das elipses

Após ajustar as elipses, outros atributos podem ser tratados como os listados abaixo:

- Probabilidade de pele média nos *pixels* contidos nas elipses.
- Número regiões de pele nas imagens.
- Distância do centróide da região maior ao centro da imagem.
- Ângulo do maior eixo da elipse local com o eixo horizontal.
- Razão entre o eixo maior e o eixo menor da elipse local.
- Probabilidade média de pele nos *pixels* não contidos nas elipses.

### 3.2.4 Qualificador

O módulo qualificador de imagens é utilizado para desempenhar o papel de indicar uma nota após o julgamento da imagem. Esta nota será a

saída do sistema. A partir dessa nota, todos os sistemas que utilizarem o TDRADO como apoio para a censura de imagens pornográficas tratarão a imagem da forma que lhes for mais conveniente. A escolha de valores maiores ou menores da nota como um limite de tolerância aceitável será configurado pelo usuário e tornará o sistema mais ou menos rígido na avaliação.

Diversos qualificadores podem ser utilizados para desempenhar o papel representado na Figura 3.3 onde vários atributos servem de entrada para o qualificador que gera uma avaliação final indicando se a imagem é pornográfica ou não.

Nesta etapa, vários qualificadores como as redes bayesianas e as redes neurais (ver ambos em Russell & Norvig 1995) são candidatos naturais para desempenhar esta função.

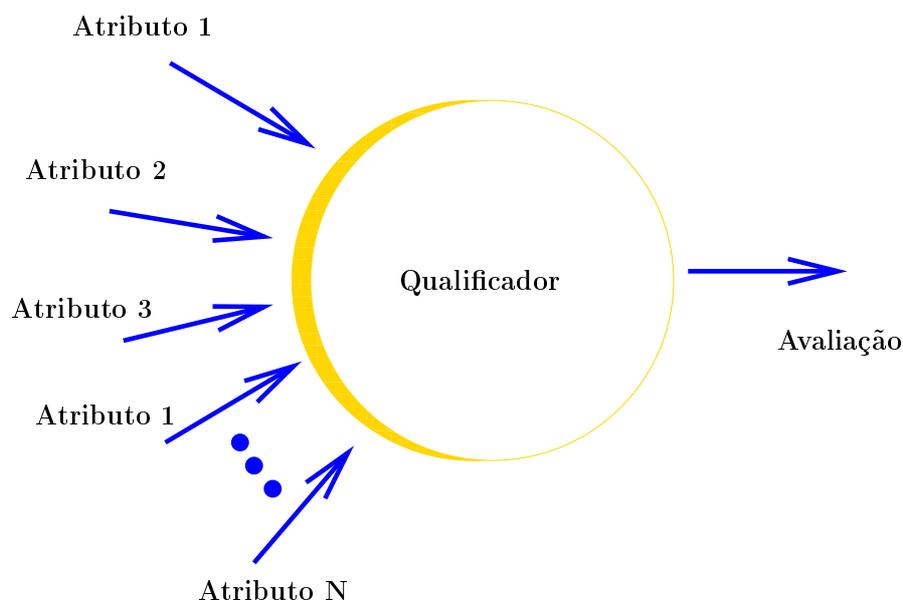


Figura 3.3: Qualificador

O sistema aqui apresentado, está sendo desenvolvido através de um projeto de pesquisa e o módulo de detecção de pele aqui apresentado é o primeiro módulo desenvolvido desse sistema. O sistema de detecção de pele com todos os seus módulos e técnicas utilizadas, é o objeto de estudo dessa dissertação e está apresentado no próximo capítulo.

# Capítulo 4

## Detecção de Pele

A detecção de pele é tarefa central deste trabalho e será abordada com detalhes neste capítulo. Em primeiro lugar apresentaremos a arquitetura do sistema proposto, de forma que o sistema seja subdividido em módulos que apresentam funcionalidades distintas (Seção 4.1).

Para o desenvolvimento do sistema aqui proposto, uma amostra de treinamento foi montada com diversas imagens colhidas na Web. Essa amostra está descrita na Seção 4.2 e as suas principais características são evidenciadas. Na Seção 4.3, apresentamos uma técnica utilizada com a finalidade de eliminar as informações redundantes da amostra, para diminuir o custo computacional na elaboração do classificador de pele. Diversos problemas de desvios do modelo (*outliers*) foram detectados ao especificarmos um modelo baseado na amostra de treinamento e uma técnica para minimizar o problema ocasionado pelos *outliers* está apresentada na Seção 4.4. Uma vez que os dados foram transformados para um formato mais tratável e os efeitos nocivos causados pelos *outliers* foram minimizados, diversos modelos matemáticos foram avaliados para a detecção de pele e estão apresentados na Seção 4.5. Uma comparação criteriosa entre os modelos é apresentada no Capítulo 5 e a técnica de correção de cor que visa minimizar os problemas dos *outliers* é avaliada. Diversos exemplos de detecção de pele em imagens colhidas na Web utilizando o classificador que apresentou o melhor desempenho

e a técnica de correção de cor estão apresentados no Apêndice B.

## **4.1 Arquitetura do Sistema de Detecção de Pele**

Com a finalidade de desenvolver o sistema de detecção de pele, foi projetada a arquitetura apresentada na Figura 4.1 que é composta pelos seguintes módulos:

1. Correção de cor
2. Redução de dimensionalidade
3. Modelos Matemáticos

O módulo de correção de cor tem por finalidade corrigir as variações de iluminação no ambiente em que as imagens são adquiridas e também as variações de respostas que cada câmera fotográfica poderá apresentar. Uma discussão completa sobre esse módulo e uma proposta para utilização de correção de cor para detecção de pele está apresentada na Seção 4.4.

O módulo de redução de dimensionalidade é responsável por reduzir a dimensionalidade dos dados de entrada —tipicamente 3D (RGB)— com a finalidade de reduzir o custo computacional e simplificar a tarefa de modelagem dos dados. Após a redução de dimensionalidade o sistema deverá poder descrever a amostra com menos informações do que a amostra original mas de forma que a perda de informações seja a menor possível ou de forma que se consiga separar a informação mais útil da informação menos útil para o processo de detecção de pele. Algumas técnicas de redução de dimensionalidade estão tratadas na Seção 4.3.

De acordo com a metodologia de análise de dados descrita na Seção 4.2, temos dois conjuntos de imagens: (i) a base de treinamento e (ii) a base de testes. Para a especificação do modelo matemático que irá compor o último módulo apresentado na Figura 4.1, foi utilizada a

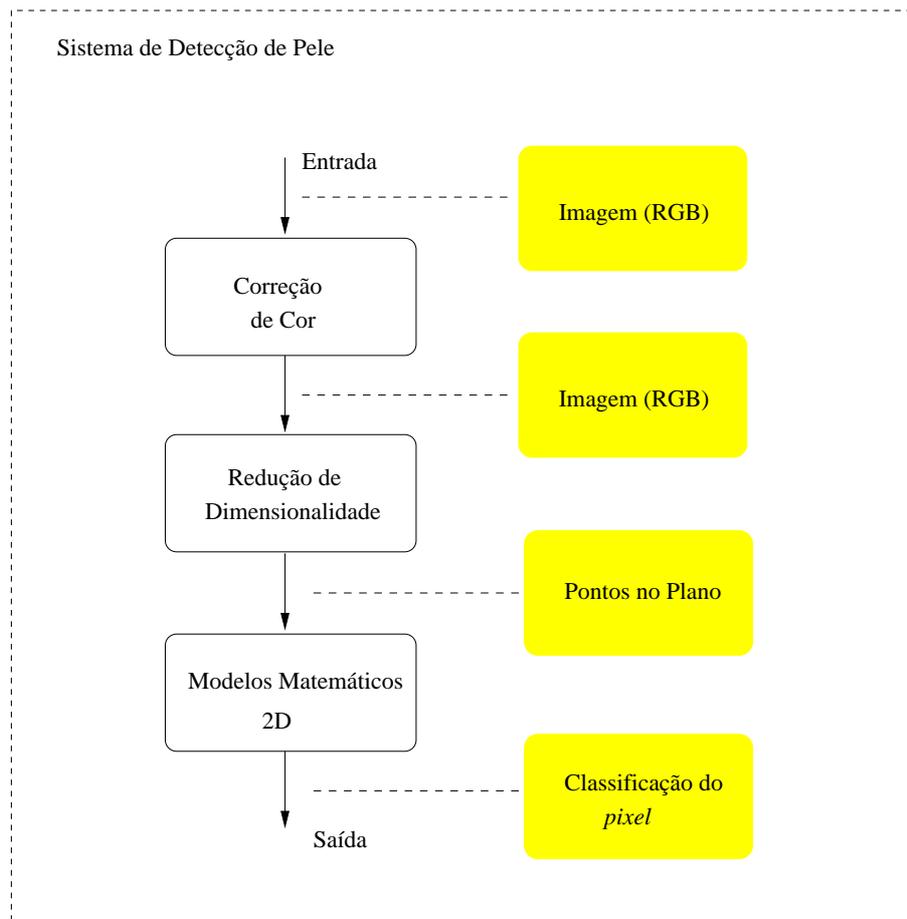


Figura 4.1: Arquitetura do sistema de detecção de pele

saída do módulo de redução de dimensionalidade nas imagens de treinamento após a correção de cor. Vários modelos foram especificados como apresentados na Seção 4.5. Para a avaliação do modelo especificado, a base de testes é utilizada. No Capítulo 5 temos uma discussão comparativa entre os modelos discutidos, as técnicas de correção de cor e as alternativas de redução de dimensionalidade.

Nas próximas Seções apresentaremos uma discussão sobre as bases de treinamento e de testes e em seguida iremos detalhar os módulos que compõem a arquitetura do sistema de detecção de pele apresentado.

## 4.2 Análise dos Dados

Visando o objetivo principal de classificação dos *pixels* de uma imagem como pele e não-pele foi montada uma base de dados de *pixels* para treinamento de alguns algoritmos descritos na Seção 4.5. Esta base de treinamento foi montada a partir de cem imagens obtidas em sítios *Web* como [www.orkut.com](http://www.orkut.com), [www.feias.com](http://www.feias.com), [www.fotolog.net](http://www.fotolog.net) onde diversas fotos de pessoas estão ali postadas. Como o sistema de detecção de pele proposto nessa dissertação irá compor um sistema de detecção de pornografia, diversas fotos dessa base de imagens são fotos pornográficas, portanto vários exemplos dessas imagens estão presentes nesta dissertação.

Vários *pixels* foram manualmente rotulados como pele e não-pele. A Figura 4.2(a) mostra a imagem original, enquanto a Figura 4.2(b) mostra na cor verde as posições claramente identificadas como pele e na cor vermelha aquelas onde não havia pele. Nenhum *pixel* diretamente na fronteira entre o corpo de uma pessoa e um objeto foi rotulado e nem aqueles que, apesar de fazer parte da pele humana, apareciam sombreados.

Jones & Rehg (2002) utilizaram em seus trabalhos, uma base de imagens rotuladas de maneira bastante semelhante a apresentada nessa dissertação. Apesar de que eles anunciam a disponibilização dessa base

de imagens para fins acadêmicos, nenhuma das inúmeras tentativas de obtenção deste acervo, teve sucesso.

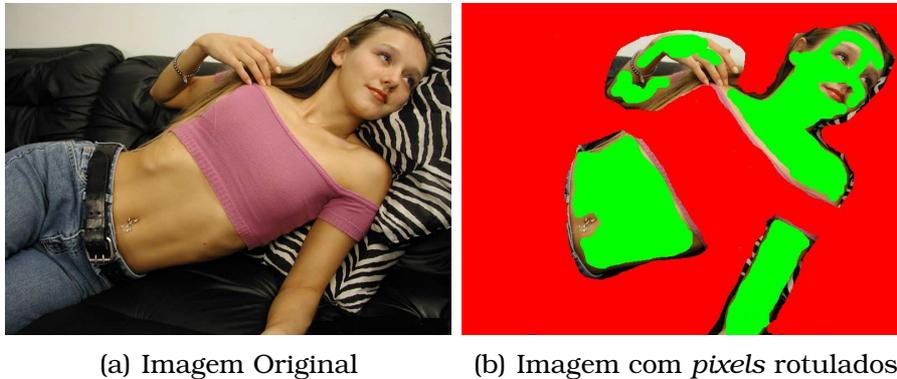


Figura 4.2: Marcação dos *pixels* de pele e não-pele para construir a base de treinamento

De posse das imagens originais e das rotuladas, foi montado um banco de dados utilizando o formato BerkeleyDb (Sleepycat 2005) onde foram utilizada três colunas para armazenar as informações necessárias; na primeira, o valor RGB do *pixel* (campo com três valores), na segunda a frequência de ocorrências como pele e, na última coluna, a frequência de ocorrência como não-pele. Com essas informações, podemos especificar modelos matemáticos e extrair informações sobre as duas classes, pele e não-pele.

Os dados foram armazenados no formato RGB pois este é o mais comumente encontrado nas bibliotecas de processamento de imagem e, como foi discutido na Seção 2.5, dispomos das transformações desse formato para os outros formatos de interesse.

O banco de dados foi povoado com 821.778 registros. Esta quantidade de registros difere de 16.777.216 (número de cores do espaço RGB de 8 bits), pois grande parte das cores que podem ser representadas neste espaço, não são frequentemente encontradas no tipo de imagem que se encontra em nossa base de dados. Os *pixels* que povoam esse banco de dados foram extraídos de um conjunto de cem imagens que contém pessoas de várias etnias como caucasianos, africanos, asiáticos

e hispânicos, entre outros, e de várias partes do corpo humano.

O fato de não termos informações sobre todos as possíveis cores do espaço RGB, apesar de aparentemente inócuo, terá consequências importantes ao utilizarmos o classificador baseado no histograma de cores. Isso se deve à falta de informações ocasionada pela ausência dessas cores. Este problema está relatado na Seção 4.5.2 na página 80, e duas possíveis soluções estão apresentadas.

Nos 821.778 registros armazenados, estamos levando em conta a redundância onde temos 9.060.832 marcados como não-pele e 438.409 como pele.

Os dados estatísticos referentes à amostra de cores de pele, em cada canal do espaço de cor RGB, estão apresentados na Tabela 4.1:

Tabela 4.1: Dados estatísticos da amostra de pele

<b>Estatística</b>	<b>R</b>	<b>G</b>	<b>B</b>
Mínimo	7	0	0
Máximo	255	255	255
Média	185,90	131,60	95,48
Mediana	186	129	92
Desvio Padrão	39,99	39,96	43,35
Coefficiente de Variação ( $\hat{v} = \hat{\sigma}/\bar{x}$ )	0,21	0,30	0,45
Assimetria Amostral	-0,31	0,029	0,24
Curtose Amostral	-0,27	-0,41	-0,34

A Figura 4.3 mostra o boxplot da amostra de pele nos canais RGB. Vemos claramente que os *pixels* da amostra de pele apresentam uma característica onde o canal vermelho, em geral, contribui mais para a formação da cor, em seguida o canal verde e por último, o canal azul. Percebe-se também que os dados são visualmente simétrico e que os *outliers* aparecem no canal R, abaixo do primeiro quartil, no canal G, tanto abaixo do primeiro quartil quanto acima do terceiro quartil e no canal B sempre acima do terceiro quartil.

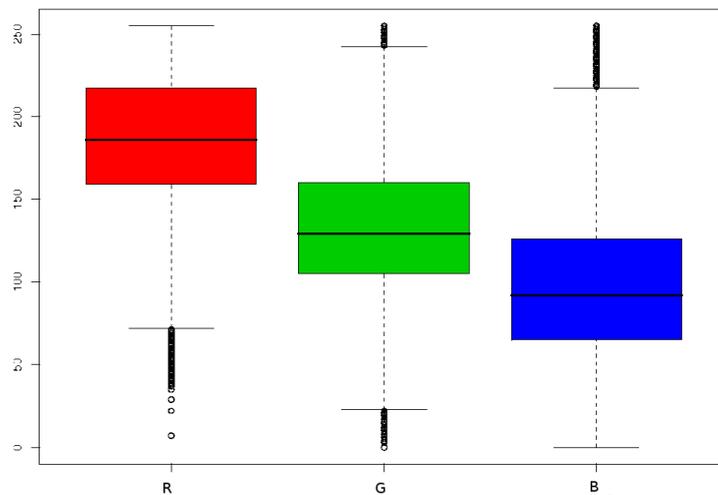


Figura 4.3: Boxplot dos canais RGB da amostra de pele

### 4.3 Redução de dimensionalidade

Uma das metas fundamentais deste trabalho é a construção de modelos expressivos e tratáveis para a descrição de cores correspondentes a regiões de pele em imagens na Web. Tal como já vimos, toda cor é um ponto em algum espaço tridimensional; podemos, portanto, visualizar em perspectiva um conjunto representativo destes pontos. Há vários espaços tridimensionais aptos a representar cores. Nas Figuras 4.4, 4.5 e 4.6 são exibidos todos os pontos da amostra de pele da base de treinamento em três espaços de cor diferentes: RGB (Figura 4.4), HSV (Figura 4.5) e CIELAB (Figura 4.6).

Dentre os vários modelos possíveis para descrever a amostra de pele, neste trabalho adotamos o uso de ferramentas estatísticas dada a variabilidade inerente dos dados tratados. Nesse contexto, quanto maior a dimensão dos dados maior a complexidade dos modelos e o custo computacional para usá-los. Portanto, podemos considerar uma redução de dimensionalidade com o intuito de reduzir o custo computacional e simplificar a tarefa de modelagem dos dados apresentados nas Figuras 4.4,

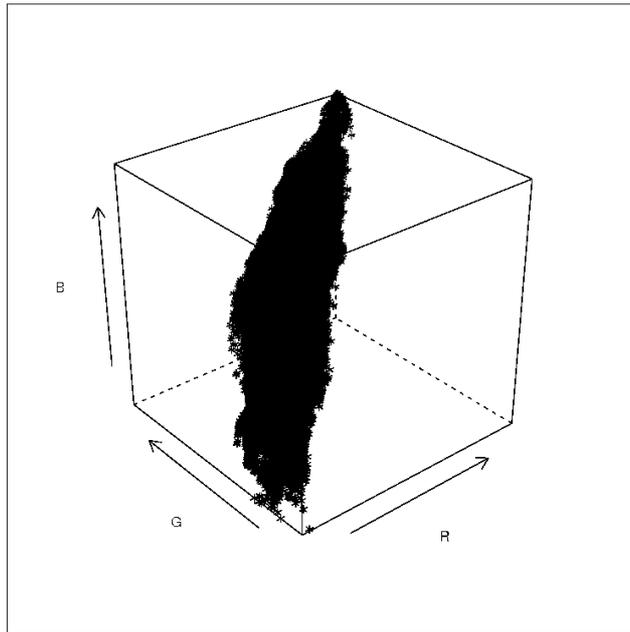


Figura 4.4: Agrupamento de cores de pele no espaço RGB

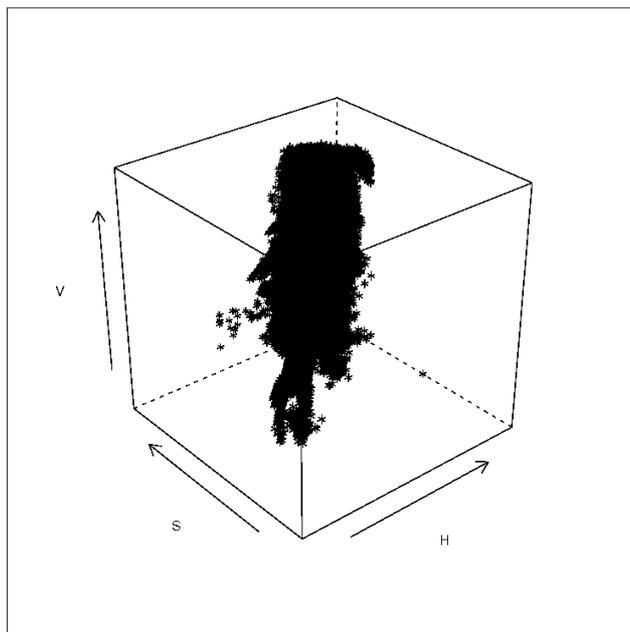


Figura 4.5: Agrupamento de cores de pele no espaço HSV

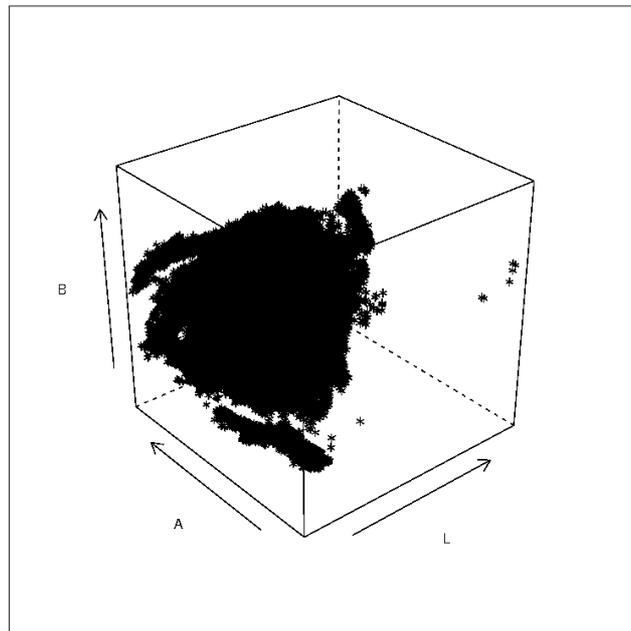


Figura 4.6: Agrupamento de cores de pele no espaço CIELAB

4.5 e 4.6.

A redução de dimensionalidade pode ser feita de várias formas, mas neste trabalho verificamos duas maneiras apresentadas na Figura 4.7.

Na primeira opção, usamos a transformação de componentes principais, técnica utilizada no trabalho de Yi et al. (2003) e descrita com mais detalhes na Seção 2.6 página 36, com o objetivo de encontrar um plano no espaço RGB em que se possa expressar o maior número de informações possível do agrupamento de cores de pele, reduzindo a dimensão de três para duas. Essa técnica gera, de forma única, três novos eixos a partir dos dados RGB do agrupamento de cor de pele, de forma que cada eixo está alinhado com as direções que apresentam mais informações sobre o agrupamento de dados.

A segunda opção consiste em utilizar uma mudança do espaço de cores de RGB para um espaço de cor onde a cromaticidade (2D) e a luminância (1D) sejam separadas em suas componentes e, neste espaço, descartar a informação de luminância. Diversos autores (Rein-Lien et

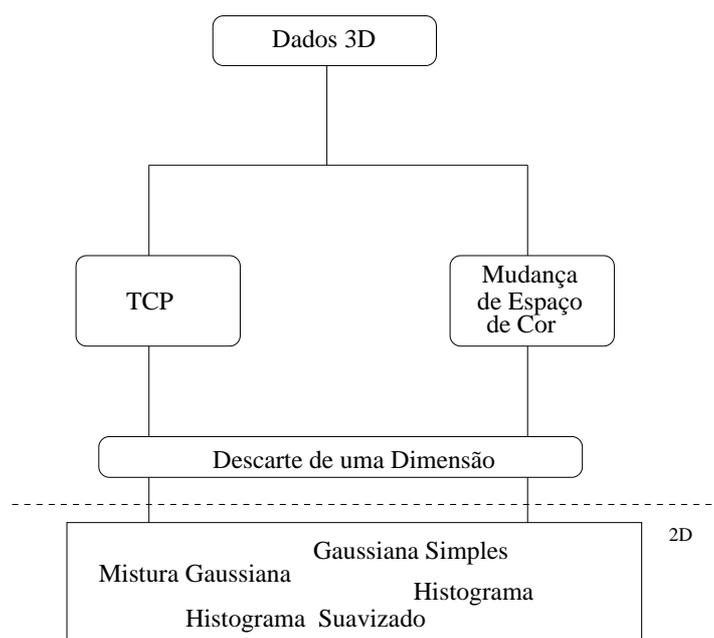


Figura 4.7: Opções para redução de dimensionalidade dos dados de pele

al. 2002, Zarit et al. 1999, Caetano & Barone 2001) utilizam essa abordagem para reduzir a dimensionalidade.

O descarte da dimensão que representa a luminância é, em princípio, possível dado que queremos detectar pele independentemente da intensidade com que ela é iluminada. A cor sem o atributo associado ao brilho é bem representada pela crominância. Ao descartar o brilho, espera-se tornar a análise mais independente das condições de iluminação, que são tipicamente muito variáveis, reduzir a complexidade dos modelos e também diminuir o custo computacional do processamento.

Um exemplo para a segunda técnica é a utilização do espaço HSV, através da transformação não-linear apresentada na Seção 2.5.4 ou para o espaço CIELAB apresentado na Seção 2.5.7, dispensando a dimensão de luminância desses espaços (V para o espaço HSV e L para o espaço CIELAB).

Utilizando a primeira técnica, obtivemos os dados apresentados na Figura 4.8. Nesta figura, temos claramente que os *pixels* referentes a cor de pele formam um agrupamento visualmente compacto neste

plano. De acordo com a técnica utilizada, este plano corta o cubo RGB e é o que contém a maior quantidade de informações do agrupamento de pele. Paralelamente aos eixos são mostrados os *boxplots* referentes a cada componente.

Utilizando a segunda técnica, que utiliza a transformação para o espaço de cor HSV, as dimensões H e S foram projetadas no sistema cartesiano e resultou nos dados apresentados na Figura 4.9; nesta figura, também estão presente os *boxplots* com as mesmas características dos da Figura 4.8.

Na Figura 4.9 podemos visualizar claramente dois agrupamentos distintos. Após uma análise mais detalhada sobre os dois agrupamentos, foi detectado que o agrupamento menor, localizado no canto inferior esquerdo, surge devido a uma contaminação nos dados de pele provocados por uma imagem com problemas de iluminação. Esta imagem está apresentada na Figura 4.12.

O problema de modificação da cor da pele ocasionado pela variação de iluminação está melhor discutido na Seção 4.4.

Na Figura 4.11 vemos a frequência de ocorrência de cada ponto do agrupamento de pele, evidenciando os dois agrupamentos citados, e a maior relevância do agrupamento mais à direita (agrupamento maior).

As componentes do plano HS do espaço de cor HSV, exibidas na Figura 4.9 (convertido em coordenadas cartesianas), apresentam uma correlação menor do que a observada entre as componentes do plano LA do espaço de cor CIELAB (Figura 4.10), mas apresenta maior correlação do que as duas primeiras componentes geradas pela técnica de TCP, (Figura 4.8). Este resultado nos leva a trabalhar com o plano encontrado pela técnica TCP dado que ele contém mais informações sobre o agrupamento de pele. A projeção dos dados do espaço RGB para o plano encontrado pela técnica TCP apresenta baixo custo computacional.

Segundo Shin et al. (2002), a mudança do espaço de cor objetivando melhorar a separabilidade entre as classes de pele e não-pele e aumentar a similaridade entre as diferentes tonalidades de cor da pele aumentando a robustez no sentido da variação de iluminação não apresenta

os resultados esperados. Nesse sentido, comparamos um dos espaços de cor mais utilizados (HSV) com a técnica de TCP.

Uma comparação quantitativa sobre a qualidade de detecção de pele entre o plano HS e o plano determinado por TCP está apresentada no Capítulo 5.

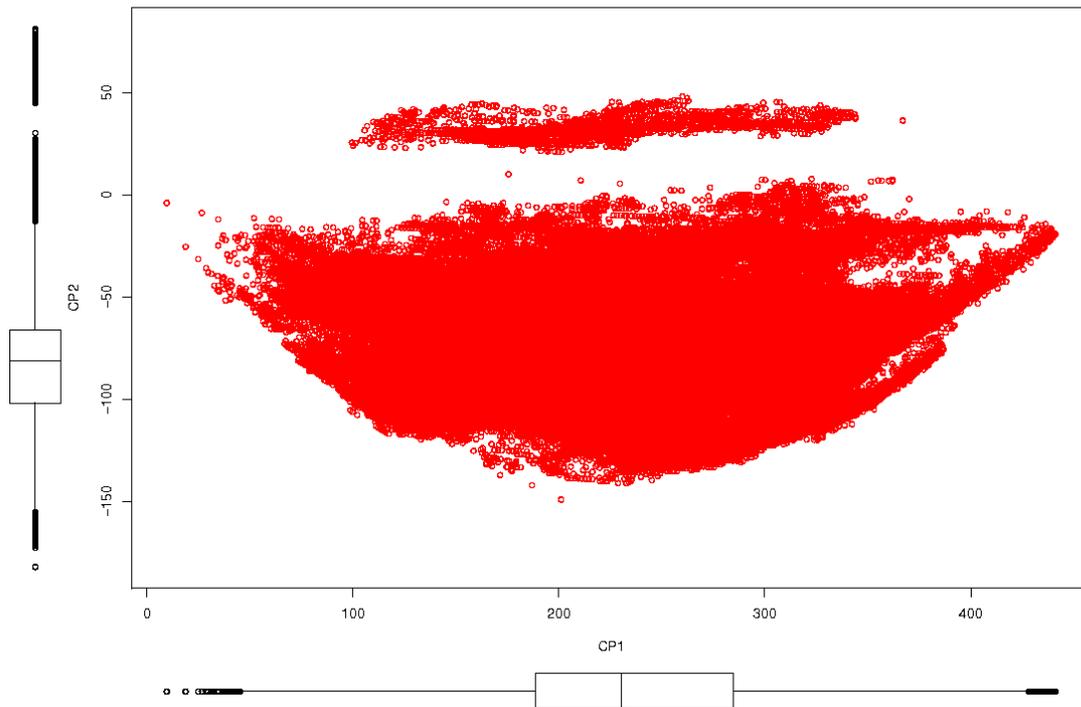


Figura 4.8: Scatter Plot dos dados de treinamento projetados no plano encontrado por TCP

De posse de dados em duas dimensões, podemos montar modelos matemáticos que os descrevam com boa precisão e utilizá-los para classificar os *pixels* de imagens que não pertençam ao conjunto de treinamento. Os modelos de pele utilizados nesse trabalho estão descritos na Seção 4.5, página 77.

Antes de discutirmos os modelos matemáticos que descrevem o agrupamento de pele, iremos, na próxima seção, discutir o problema da variação da cor da pele com a variação da iluminação do ambiente onde

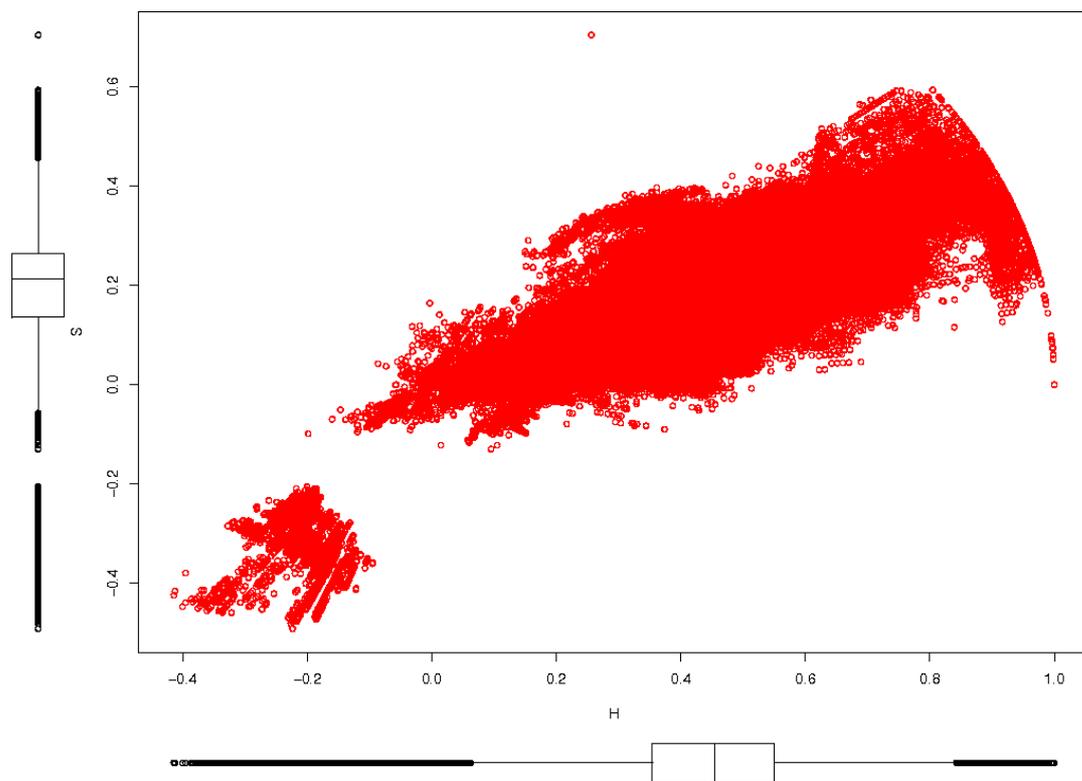


Figura 4.9: Scatter Plot dos dados de treinamento projetados no plano HS

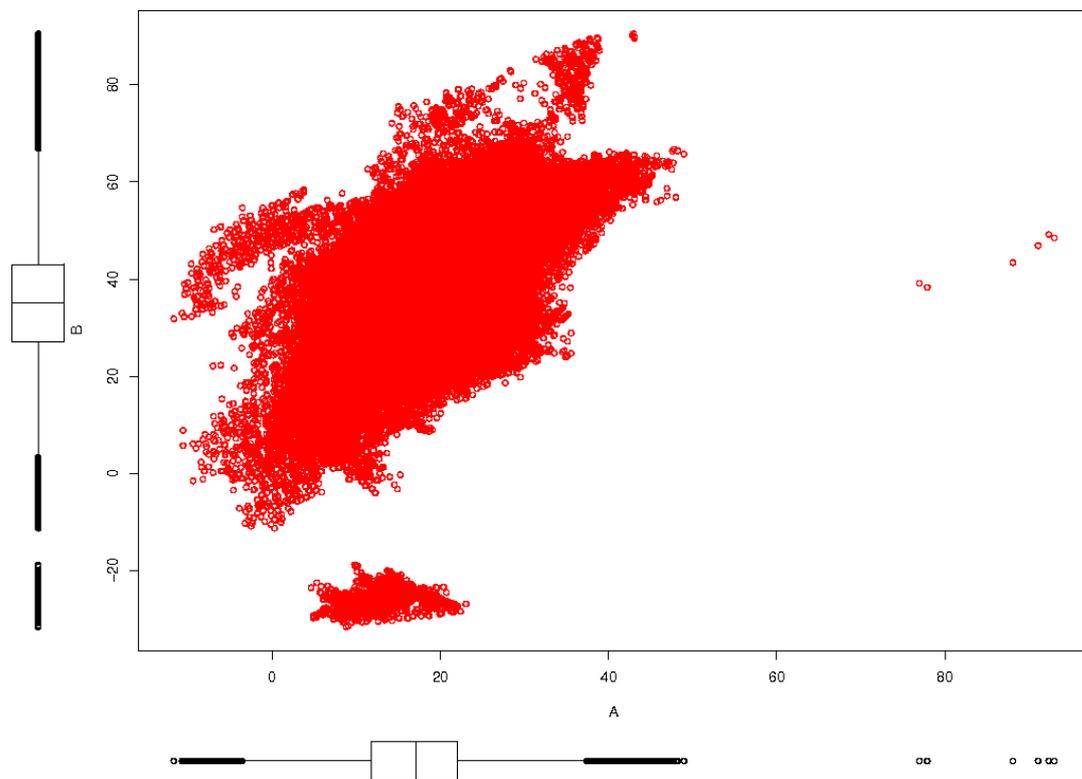


Figura 4.10: Scatter Plot dos dados de treinamento projetados no plano LA

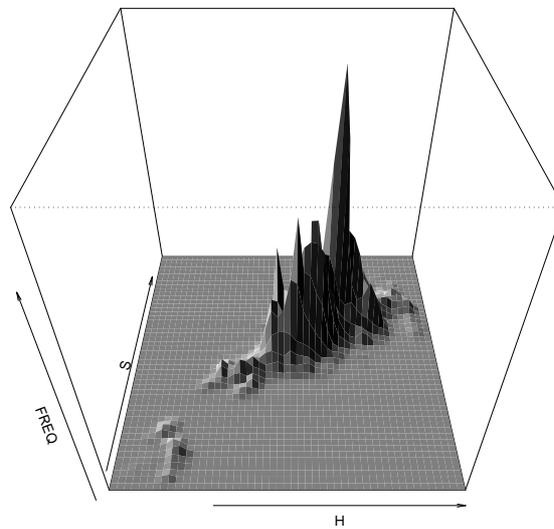


Figura 4.11: Frequências de ocorrência dos pixels (plano HS)

a imagem foi adquirida e das características do dispositivo de aquisição da imagem (câmera fotográfica).

## 4.4 Correção de Cor

Considerando que a cor é uma característica útil para identificação de objetos, a sua aparência deve permanecer constante quando o objeto for visualizado em diferentes situações, como por exemplo em diferentes iluminações (luz do dia, lâmpada fluorescente e lâmpada incandescente) para que uma boa desempenho na identificação desses objetos seja alcançada.

As cores de imagens adquiridas por câmeras fotográficas dependem não só das propriedades da superfície, como também das condições de iluminação (geometria e cor da iluminação) e das características da câmera. Por exemplo, uma imagem digital de um pedaço de papel branco parecerá amarela quando submetida a uma iluminação incandescente

ou azulada/esverdeada quando submetida a uma iluminação fluorescente. Sob diferentes condições de iluminação, o mesmo papel parecerá branco ou não pois o valor medido de cada um dos três canais de cor será diferente para diferentes condições de iluminação.

A característica do sistema de visão dos seres humanos de preservar aproximadamente a cor original em diferentes iluminações, é chamada de “adaptação cromática”. Infelizmente, diferentemente do sistema humano de visão, as câmeras fotográficas não conseguem adaptar suas respostas espectrais a fim de compensar as condições de iluminação e manter os atributos de cor constantes; como resultado, a imagem adquirida por câmeras fotográficas poderá conter um desvio indesejável no seu espectro de cor. Na Figura 4.12 vemos claramente a imagem com coloração azulada provocada por iluminação fluorescente.

Este efeito é indesejável na detecção de pele pois desloca os *pixels* do agrupamento de pele para fora dele (Martinkauppi et al. 2003). Este efeito é melhor visualizado na Figura 4.9 onde todos os *pixels* marcados como pele pertencentes à Figura 4.12 pertencem ao agrupamento inferior à esquerda. Dessa forma, os modelos de detecção de pele não irão detectar os *pixels* de pele como sendo pele (erro de omissão). Percebe-se que os dados dessa imagem podem ser considerados *outliers*. Os *outliers* podem contaminar a inferência baseada no modelo, caso essa imagem pertença à base de dados de treinamento, ou podem fazer com que a imagem de onde eles foram extraídos seja mal classificada, caso seja submetida a um sistema de detecção automática de pele.

Segundo Rein-Lien et al. (2002) a cor da pele depende da iluminação, independente do espaço de cor, sendo portanto necessária a utilização dos modelos de correção de cor; estes, por sua vez, são chamados de “modelos de adaptação cromática”.

Os métodos estudados nesse trabalho são baseados nos modelos encontrados na literatura e não consideram adaptações nos atributos de aparência como Matiz, Saturação e Brilho; eles provêm uma transformação no espaço tricromático de uma determinada condição de visualização para outra. Esses modelos são baseados na hipótese de Von



Figura 4.12: Imagem com um desvio no seu espectro de cor causado por iluminação fluorescente

Kries (Gasparini & Schettini 2003) que se sugere a correção de cor no espaço de cromaticidade aplicando um ganho de regulação em cada um dos três canais de cor através de três diferentes coeficientes:

$$L_a = K_L L \quad M_a = K_M M \quad S_a = K_S S,$$

onde L, M e S representam os valores iniciais de energia incidente obtidos pelos sensores sintonizados nos canais de cor de baixa, média e alta frequência como descrito na Seção 2.1, e  $K_L$ ,  $K_M$  e  $K_S$  são os coeficientes de correção de cor que produzem um efeito de escala no sinal original gerando os valores corrigidos  $L_a$ ,  $M_a$  e  $S_a$ .

Diversos métodos de correção de cor estão presentes na literatura como os apresentados por Land (1977), Brainard & Wandell (1986), Gasparini & Schettini (2003), Rizzi et al. (2002) e Finlayson et al. (2001). Iremos tratar as duas técnicas amplamente conhecidas pela comunidade de visão computacional: (i) *Gray World Assumption* (ii) *White Patch*.

A primeira técnica supõe que dada uma imagem com um número de variações de cor suficiente, o valor médio dos componentes R,G e B da imagem será um valor médio de cinza (Rizzi et al. 2002), ou seja, a reflectância média de todas as superfícies da cena é acromática (Finlayson et al. 2001). Dessa forma, uma imagem adquirida através de uma câ-

mera em um determinado ambiente de iluminação terá sua coloração corrigida ao aplicar a transformação de Von Kries com os seguintes coeficientes:

$$K_R = \frac{\text{GrayR}}{R_{\text{avg}}} \quad K_G = \frac{\text{GrayG}}{G_{\text{avg}}} \quad K_B = \frac{\text{GrayB}}{B_{\text{avg}}},$$

onde  $R_{\text{avg}}$ ,  $G_{\text{avg}}$  e  $B_{\text{avg}}$  correspondem aos valores médios dos três componentes RGB e GrayR, GrayG e GrayB representam o cinza de referência. Uma boa aproximação para o cinza de referência é o valor médio dos canais RGB  $((R_{\text{avg}} + G_{\text{avg}} + B_{\text{avg}})/3)$  da imagem em questão.

A segunda técnica consiste em tomar como referência um ponto de uma região da cena real que é considerado como branco, ou seja, nessa técnica supõe-se que sempre existe uma região branca que pode ser utilizada como referência. Um fator determinante para a qualidade da técnica é a identificação correta deste ponto de referência. Uma maneira simples de implementar a técnica é calcular a razão entre o ponto branco de referência (WhiteR, WhiteG e WhiteB) e o valor máximo de cada canal ( $R_{\text{Max}}$ ,  $G_{\text{Max}}$  e  $B_{\text{Max}}$ ), calculando os coeficientes de Von Kries:

$$K_R = \frac{\text{WhiteR}}{R_{\text{Max}}} \quad K_G = \frac{\text{WhiteG}}{G_{\text{Max}}} \quad K_B = \frac{\text{WhiteB}}{B_{\text{Max}}}$$

Um trabalho que faz uso do *White Patch* de uma maneira bastante interessante é o apresentado por Rein-Lien et al. (2002). Neste trabalho, para encontrar o “branco de referência”, eles utilizam os *pixels* que apresentam os 5% maiores valores de luminância da imagem; para isto, eles utilizam o canal (Y) do espaço de cor YCbCr.

Caso uma imagem possua um número suficiente de *pixels* localizados nesse percentil, os componentes R, G e B da imagem são ajustados de forma que o cinza médio desses *pixels* seja linearmente escalado para 255. Para Rein-Lien et al. (2002), uma imagem que não possua um número suficiente desses *pixels* ou que a cor média dos *pixels* seja similar à cor de pele humana não são submetidas à correção de cor. A hipótese utilizada pelos autores desse trabalho é que as imagens que eles traba-

lham usualmente apresentam um “branco real”, pois algumas regiões de interesse (como a região dos olhos) sempre apresentam branco, e que a cor dominante geralmente aparece como o “branco real”.

Essa hipótese é verificada na aplicação do trabalho de Rein-Lien et al. (2002) mas não é necessariamente válida para a nosso caso, dado que imagens mais gerais (não apenas as que contém face) como algumas fotos bem próximas ao corpo humano podem não apresentar essas características.

Algumas características da técnica de correção de cor proposta por Rein-Lien et al. (2002) devem ser ressaltadas como: (i) imagens onde os *pixels* detectados como “branco de referência” são realmente brancos ( $RGB = (255, 255, 255)$ ), a técnica não irá modificar a imagem; (ii) imagens onde os *pixels* detectados como “branco de referência” pertecem à pele humana, a técnica irá aumentar o brilho desses *pixels*, os deixando praticamente brancos; portanto fora do agrupamento de pele; (iii) imagens que apresentam um fundo próximo a cor da pele como paredes, sofás e detalhes em madeira, por exemplo, e esses *pixels* são detectados como “branco de referência”, a técnica irá corrigir esses *pixels* para uma coloração mais próxima à branca e poderá separar bem a pele humana evitando o erro de comissão.

Um exemplo para cada uma dessas características está apresentado na Figura 4.13. As Figuras 4.13(a) e 4.13(b), mostram um exemplo para a característica (i) onde a imagem praticamente não foi alterada pela correção de cor de Rein-Lien et al. (2002); nas Figuras 4.13(c) e 4.13(d) temos uma imagem onde a cor da pele foi alterada para a cor branca e nas Figuras 4.13(e) e 4.13(f), temos a correção de cor correta, onde a classe pele foi realçada do fundo (parede).

A heurística utilizada por Rein-Lien et al. (2002) não produz resultados bons para muitas imagens presentes em nossa base de treinamento pois não necessariamente temos a presença da cor branca para detectarmos o “branco de referência”. A aplicação do *Gray World Assumption* também não produziu bons resultados para todas as imagens pois apenas imagens que possuam uma grande variação de cores apresentam



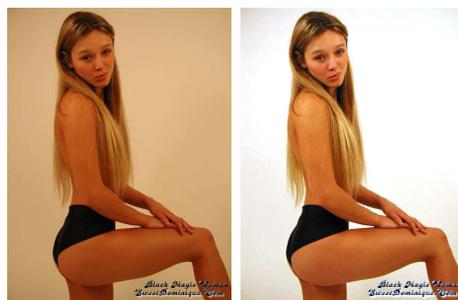
(a) Original

(b) Correção de Cor



(c) Original

(d) Correção de Cor



(e) Original

(f) Correção de Cor

Figura 4.13: Exemplos da correção de cor proposta por Rein-Lien et al. (2002)

uma boa correção de cor. Estes fatos nos remeteram a adaptação da heurística de Rein-Lien et al. (2002).

Na nossa heurística, uma combinação dos dois algoritmos foi avaliada de modo que em algumas imagens, seja aplicado o algoritmo *Gray World Assumption*, em outras o algoritmo *White Patch* e, quando não se fizer necessário, não seja aplicada correção de cor. A heurística foi criada avaliando as possibilidades de falha de cada uma das técnicas, chegando à seguinte regra:

$$\text{Correção de Cor} = \begin{cases} \text{GrayWorld,} & \text{se } N_{\text{white}} < 3\% \\ \text{WhitePatch} & \text{se } (N_{\text{white}} + 5L > 120) \\ \text{sem Correção,} & \text{caso contrário,} \end{cases}$$

onde  $N_{\text{white}}$  é o percentual de *pixels* detectados como branco seguindo a regra de Rein-Lien et al. (2002) e  $5L$  é o valor da componente L do espaço de cor CIELAB para o percentil 95%.

Esta heurística sugere que a aplicação da técnica *Gray World Assumption* para os casos em que a quantidade de pontos brancos de referência seja bastante pequenena, de *White Patch* para os casos em que tem-se uma quantidade suficiente de *pixels* representando o branco de referência e que o nível de iluminação desse conjunto de *pixels* seja alta, e nenhuma técnica caso as regras anteriores não sejam satisfeitas.

Vale ressaltar que essa heurística não tem por objetivo melhorar a aparência geral de apresentação da imagem, dado que os exemplos sem sucesso apresentados na Figura 4.13, podem acontecer. Essa heurística apenas minimiza a ocorrência dos problemas e também evita aplicar a correção de cor no caso em que não existam *pixels* suficientes para detectar o “branco de referência”.

O objetivo desta técnica é simplesmente deslocar os *pixels* de pele de volta para o agrupamento modelado.

Essas técnica são utilizadas em máquinas digitais para minimizar o efeito de variação da coloração dos objetos quando sujeitos a diferentes fontes de iluminação. Essa característica é conhecida no jargão da fo-

tografia como *white balance* e é baseada no método apresentado aqui como *White Patch* e, em alguns casos, no método *Gray World*.

Infelizmente nem todas as câmeras digitais corrigem bem a variação de iluminação ou nem todos os usuários dessas câmeras utilizam essa funcionalidade da maneira mais correta. A forma mais comum de operação é o modo automático, que não necessariamente faz a correção a contento. Este fato faz com que uma grande quantidade de fotos com problemas de iluminação estejam presente na Web.

Para imagens controladas, ou seja, imagens onde o *white balance* é utilizado da forma correta, não se faz necessária a utilização de técnicas de correção de cor para detecção de pele; em contrapartida, para o propósito que o nosso sistema está desenvolvido, essa suposição não é verdadeira pois na Web existe uma variedade muito grande fotos que foram capturadas sem os devidos cuidados. Storrington et al. (1999) apresentam uma discussão bastante completa sobre o deslocamento que o agrupamento de pele sofre no plano cromático com a variação de iluminação.

Detectar se uma imagem deve ou não passar por algum desses algoritmos de correção de cor, ou propor uma técnica genérica para qualquer tipo de imagem é uma tarefa ainda não bem resolvida pela comunidade de visão computacional sendo que diversos pesquisadores têm procurado propor técnicas com Land (1977) o pioneiro, e outros como Forsyth (1990), Rein-Lien et al. (2002), Gasparini & Schettini (2003), Rizzi et al. (2002), Ebner (2003) e Finlayson et al. (2001) que, dentre outros vários, buscam solucionar este problema.

Na próxima seção discutiremos as técnicas de detecção de pele avaliadas nessa dissertação.

## **4.5 Técnicas de Detecção de Pele**

Apesar da detecção de pele ser reportada como um problema solucionado pela comunidade de visão computacional (Bosson et al. 2002),

quando levamos em consideração que as imagens digitais apresentam características como cores de pele de várias etnias, vários tipos de iluminação, foco e ângulos de visão variados, técnicas que sejam robustas às variações desses parâmetros devem ser escolhidas para compor o módulo detecção de pele humana.

As principais técnicas supervisionadas para detecção de pele em imagens da Web encontradas na literatura (ver Laurent et al. 2003) são classificadas em três diferentes tipos: (i) regras geométricas (Gomez & Morales 2002), (ii) estatísticas paramétricas (Caetano & Barone 2001, Caetano et al. 2002) e (iii) baseadas em histogramas (Jones & Rehg 2002).

As regras que definem a primeira são bastante simples e baseadas na regra do paralelogramo (Schalkoff 1989), ou obtidas através de *data-mining*; elas são de implementação simples. Essas técnicas são de regra fixa, ou seja, sem limiar de controle, e não permitem ajustes na qualidade da detecção.

Técnicas paramétricas são susceptíveis a desvios do modelo suposto e a dados anômalos, conhecidos por *outliers*.

Técnicas não-paramétricas, como as baseadas em histograma, tornam-se confiáveis quando se dispõe de uma grande quantidade de dados de amostra de treinamento como apresentado no trabalho de Jones & Rehg (2002), onde mais de  $10^9$  *pixels* estão presentes na base de treinamento.

Para o nosso trabalho, fizemos a avaliação das seguintes técnicas: duas diferentes técnicas de limiar de classificação (seção 4.5.1), três variações da técnica histograma de cores (seção 4.5.2, página 80) e duas diferentes técnicas baseadas em estatística paramétrica (seção 4.5.3, página 84). Nem todas as técnicas avaliadas estão mostradas no Capítulo 5, apenas as que apresentaram os melhores índices de avaliação ou aquelas que apresentaram características que fossem relevantes para os resultados desta dissertação.

### 4.5.1 Limiar de Classificação

As técnicas baseadas em limiar de classificação usam a propriedade que as cores de pele humana formam um agrupamento compacto em cada dimensão do espaço de cor. Este agrupamento pode ser modelado por intervalos em cada componente de cor,  $[C_{\min}^i, C_{\max}^i]$  onde  $i$  representa a dimensão do espaço de cor  $C$ . Desta maneira, as cores associadas a pele podem ser modeladas de várias formas, ou como sendo as contidas no paralelepípedo definido pelo produto cartesiano desses intervalos ou por alguma outra regra apta a descrever a região. Feito isso, isto é, fornecido o modelo para a região do espaço de cores que descreve cores de pele, pode utilizar-se uma máscara binária  $J$  a partir da imagem original onde,  $J(h(s)) = 1$ , se  $h(s)$  pertence à região associada às cores.

A regra, assim, é bastante simples e fácil de implementar, sendo a principal dificuldade a determinação da região.

Este tipo de técnica foi implementada para ser utilizada como comparação com as outras técnicas empregadas nesse trabalho. Para o tal foram utilizadas duas regras que descrevem os agrupamentos dos *pixels* em dois espaços de cor diferentes, o RGB e o YCbCr.

A primeira regra utilizada foi:

$$\begin{aligned} & R > 95 \quad \wedge \quad G > 40 \quad \wedge \\ & B > 20 \quad \wedge \quad \max\{R,G,B\} - \min\{R,G,B\} > 15 \quad \wedge \\ & |R - G| > 15 \quad \wedge \quad R > G \quad \wedge \quad R > B \end{aligned}$$

Esta é a regra apresentada no trabalho de Kovac et al. (2003). Este método vem atraindo vários pesquisadores como Albiol & Torres (2001) pela simplicidade, facilidade de implementação e por apresentar um baixo custo computacional.

A segunda técnica de limiar de classificação utilizada foi baseada no trabalho de Rein-Lien et al. (2002). Para encontrar a regra de classificação, Rein-Lien et al. (2002) utilizaram a comparação feita por Terrillon et al. (2000) entre diversos espaços de cor para detecção de pele com o

objetivo de detecção de face. Terrillon et al. (2000) constatam que o espaço rg normalizado não é a melhor escolha para detecção de pele com o objetivo de detecção de face. Baseado nos estudos de Terrillon et al. (2000), Rein-Lien et al. (2002) utilizaram uma transformação não-linear no espaço de cor YCbCr, visando obter um plano de crominância Cr'Cb' que seja independente da componente Y (Luma) para o agrupamento de pele obtido na sua base de treinamento. Após isto, Rein-Lien et al. (2002) verificaram que a projeção dos pontos do agrupamento de pele nesse novo plano Cr'Cb', forma um agrupamento bastante compacto, podendo ser descrito por uma elipse. Com isto, uma simples equação de uma elipse descreve o agrupamento de pele e é utilizada para classificar os *pixels* de uma imagem. A equação da elipse proposta no trabalho de Rein-Lien et al. (2002) está apresentada nas Equações (4.1) e (4.2).

$$\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} = 1, \quad (4.1)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta & \text{sen } \theta \\ -\text{sen } \theta & \cos \theta \end{bmatrix} \begin{bmatrix} C'_b - c_x \\ C'_r - c_y \end{bmatrix}, \quad (4.2)$$

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 no plano  $C'_b C'_r$ . Estes dados foram obtidos a partir do experimento de Rein-Lien et al. (2002) para detecção de faces. Os resultados da avaliação deste classificador estão mostrados no Capítulo 5. A transformação do espaço YCbCr para o espaço YCb'Cr' está detalhada em Rein-Lien et al. (2002).

A seguir, descreveremos outra classe de técnicas de detecção de pele: os classificadores estatísticos.

### 4.5.2 Estatístico - Histograma de Cor

Os modelos estatísticos são frequentemente propostos na literatura através de duas técnicas diferentes. A primeira consiste em projetar um modelo de cor de pele construindo dois histogramas de cor (Jones &

Rehg 2002): um histograma para as cores da pele e um histograma para as cores que não fazem parte do agrupamento de pele. Ambos os histogramas são obtidos através de marcação manual dos *pixels* como mostrado na Seção 4.2 na página 59.

As contagens dos histogramas são convertidas em distribuição de probabilidade discretas fazendo :

$$P(c | pele) = \frac{pele[c]}{T_p}$$

$$P(c | \overline{pele}) = \frac{n[c]}{T_n};$$

onde  $pele[c]$  indica a contagem do histograma das cores de pele associado a cor  $c$ ,  $n[c]$  indica a contagem do histograma das cores de não-pele associado a cor  $c$  e  $T_p$  e  $T_n$  correspondem às contagens totais obtidas dos histogramas das cores de pele e não-pele, respectivamente. Os valores de  $P(c | pele)$  e  $P(c | \overline{pele})$  indicam as probabilidades da cor  $c$  ser pele e não-pele, respectivamente.

Os valores dessas probabilidades são obtidos a partir das contagens dos histogramas montados através da marcação manual utilizando a amostra de treinamento do modelo.

Uma vez montados os histogramas de cores de pele e não-pele, a construção do classificador de *pixel* de pele utiliza a razão de verossimilhança mostrada na Equação (4.3) (Jones & Rehg 2002):

$$\frac{P(c | pele)}{P(c | \overline{pele})} \geq \Theta, \quad (4.3)$$

onde  $0 \leq \Theta \leq 1$  é um limiar de classificação. O limiar  $\Theta$  pode ser ajustado seguindo uma relação de compromisso entre detecções corretas e falsos positivos. Uma propriedade interessante que pode ser derivada da variação do parâmetro  $\Theta$  na Equação (4.3) é a curva ROC (*Receiver Operating Characteristic*) que foi amplamente utilizada em todas as técnicas supervisionadas presentes nesse trabalho para uma comparação quantitativa entre elas. As curvas ROC dos classificadores apresenta-

dos neste trabalho, estão apresentadas no Capítulo 5.

Esta técnica foi utilizada com o espaço de cor RGB, onde foi montada uma tabela de consulta com aproximadamente 400 mil *pixels* marcados como pele e 9 milhões de *pixels* marcados como não-pele, de acordo com o apresentado na Seção 4.2. As entradas dessa tabela foram convertidas para um histograma de intervalo de largura 1. Utilizando a Equação (4.3) calcula-se a verosimilhança de cada *pixel* e a compara com o valor do limiar  $\Theta$ , classificando-o como pele ou não-pele.

Ao utilizarmos a técnica descrita acima, encontramos um comportamento interessante: ao variarmos o limiar  $\Theta$ , o percentual de *pixels* classificados corretamente não ultrapassava um valor em torno de 60%. Este fato pode ser explicado com a ajuda da Figura 4.14.

Nesta figura temos um histograma de apenas uma dimensão, com 8 intervalos; são observados “buracos” entre os intervalos, correspondentes aos valores de cor que estão presentes na amostra de testes e não estão presentes na amostra de treinamento. Ao variarmos o valor de  $\Theta$  entre  $[\Theta_0, \Theta_9]$ ,  $\Theta \in \mathbb{R}_+$ , o percentual de acertos corresponde à área hachurada apresentada na Figura 4.14. Vemos claramente que os “buracos” presentes no histograma farão com que a área hachurada seja bem menor do que a área sobre a linha pontilhada correspondente ao limiar  $\Theta$ . Este fato faz com que, mesmo usando o valor  $\Theta_9$  como limiar de classificação (valor onde o limiar é o menos exigente possível), o percentual de acerto não ultrapassasse ao valor da área hachurada que pode ser muito menor do que a área do retângulo formado por  $\Theta$  e o eixo das abcissas. Essa área será muito menor, por exemplo do que a área sob uma curva gaussiana, pois as informações faltantes contribuem com área zero. Podemos extrapolar as características enfatizadas aqui para um histograma em três dimensões como o que foi apresentado na técnica descrita acima.

Como alternativa ao problema descrito anteriormente, utilizamos a técnica de suavização do histograma e o aumento do intervalo para 64. As duas técnicas procuram preencher as informações faltantes com valores estimados. Na primeira, o valor faltante vai ser estimado pela

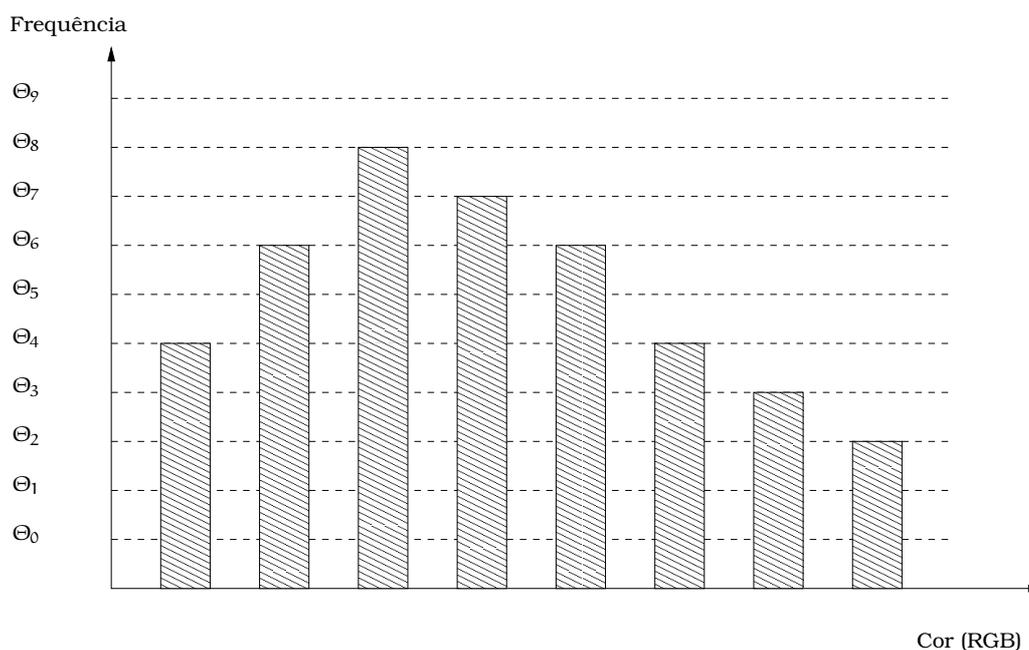


Figura 4.14: Limite da taxa de acerto pelo método do histograma

curva que descreve o histograma e na segunda, o valor vai ser estimado considerando que o agrupamento de pele é compacto e que os *pixels* vizinhos servem como uma boa estimativa da probabilidade do *pixel* faltante. A utilização de uma base de *pixels* de treinamento muito maior pode diminuir este problema mas nada garante que o evite totalmente.

Na suavização do histograma, um estimador não-paramétrico do histograma foi calculado através do sistema R. Para essa suavização, os dados em três dimensões foram reorganizados através de uma função  $g: \mathbb{R}^3 \rightarrow \mathbb{R}$ . Esta reorganização foi realizada através da ordenação lexicográfica dos valores de cada eixo do “cubo” RGB, ou seja, a ordenção inicia com os valores de R, depois de G e por fim de B. Por exemplo, o ponto  $(255, 0, 0)$  é anterior ao ponto  $(255, 1, 0)$ , e assim sucesivamente.

Foi utilizada a função *bkde2d* do sistema R com a finalidade de encontrar uma curva que melhor descrevesse os histogramas das cores de pele e não-pele. Os resultados estão apresentados nas Figuras 4.15 e 4.16. É válido observar que os “buracos” não apareceram nestas figuras por causa da escala apresentada. Nestas figuras, a linha vermelha é a

função que descreve o histograma suavizado. Dessa forma, os valores que não estão presentes (“buracos”) são estimados.

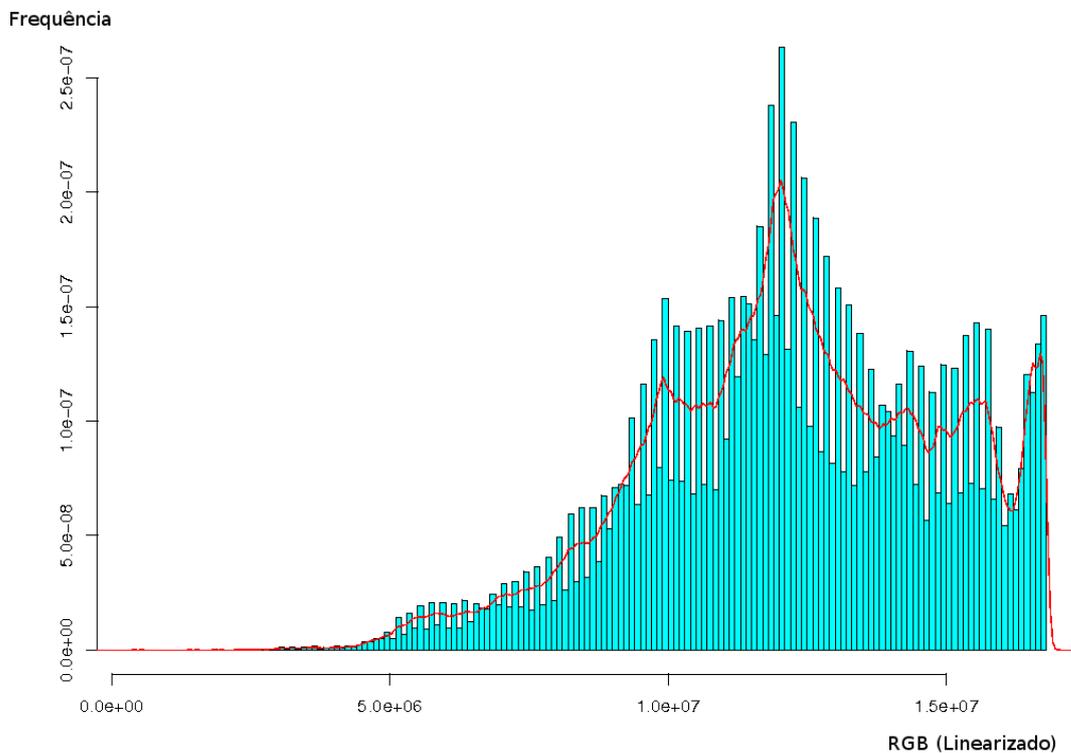


Figura 4.15: Suavização do histograma de pele

Tanto a suavização do histograma quanto o aumento do tamanho do intervalo para 64, amenizaram o problema dos buracos no histograma e uma comparação quantitativa entre essas técnicas está apresentada no Capítulo 5.

### 4.5.3 Estatístico - Paramétrico

A segunda categoria de métodos estatísticos consiste em utilizar uma distribuição de probabilidade gaussiana para modelar um conjunto de *pixels* manualmente marcado como pele. Após a modelagem Gaussiana,

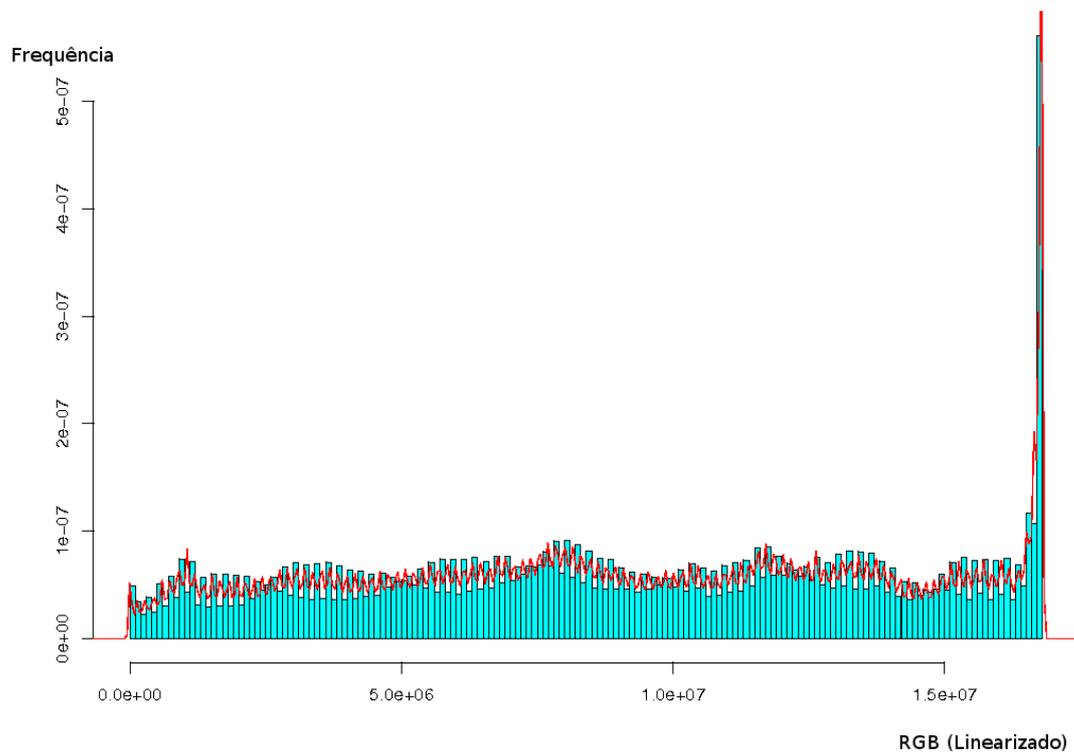


Figura 4.16: Suavização do histograma de não-pele

classificar um *pixel* como pele se reduz à computação da distância de Mahalanobis (Caetano & Barone 2001) entre a cor do *pixel*  $h(s)$  e a média da distribuição. O quadrado da distância de Mahalanobis de  $x$  para  $\mu$  é definido por:

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

$x$  corresponde a cor do *pixel* e  $\mu$  o vetor de médias. Para este método foi utilizado um limiar de classificação  $\Theta$ , e montadas as curvas ROC para comparação dos classificadores.

Nesta técnica, foi utilizada uma 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 TCP), onde a função de distribuição de probabilidades é dada por:

$$P(c | pele) = \sum_{i=1}^k w_i P_i(c | pele), \quad (4.4)$$

onde  $w_i$  denota o peso (número total de *pixels* da classe dividido pelo total de *pixels* da amostra) de cada uma das gaussianas que compõem a mistura. Cada uma das componentes da mistura possui a seguinte função de distribuição de probabilidades:

$$P(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \frac{z}{2(1-\rho^2)}, \quad (4.5)$$

onde:

$$z = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2},$$

$\mu_1$  e  $\mu_2$  são as médias,  $\sigma_1$  e  $\sigma_2$  são os desvios padrões de cada uma das variáveis  $x_1$  e  $x_2$  e  $\rho$  é o coeficiente correlação entre as variáveis  $x_1$  e  $x_2$ .

As variáveis  $x_1$  e  $x_2$  representam os valores de H e S, A e B ou  $Cp_1$  e  $Cp_2$ .

Para a mistura de gaussianas, foi utilizado o limiar de probabilidade

ao invés da distância de Malahanobis para a implementação do classificador.

A projeção dos pontos HSV no plano HS dos dados de treinamento estão apresentados nas Figuras 4.9 e 4.11, os dados do agrupamento de pele da base de treinamento projetados plano AB estão apresentados na Figura 4.10 e os mesmos dados projetados no plano  $C_{p_1}$  e  $C_{p_2}$  estão apresentados na Figura 4.8.

Foi utilizado o sistema R, com o pacote *Mclust* para determinar qual a quantidade de classes que a “nuvem” de dados apresenta.

O pacote *Mclust* do sistema R usa *Estimation-Maximization (EM)* para descobrir o número de classes da amostra. Para que o método seja computacionalmente viável, foi colhida uma sub-amostra de 1%, uniformemente, dos aproximadamente 400 mil pontos da amostra de treinamento e foi obtido o valor de nove classes distintas, onde cada grupo de *pixels* é modelado por cada uma das gaussianas que compõem a mistura da Equação (4.4). Os grupos foram encontrados com o algoritmo EM para os dados de entrada referentes a cada uma das dimensões do plano HS do espaço de cor HSV para a Figura 4.17 e para as duas primeiras componentes principais do cubo RGB, denotadas CP1 e CP2, como mostrado na Figura 4.19.

O motivo pelo qual foi retirada uma amostra com 1% dos dados é que a plataforma R rodada em um computador Pentium IV 2 GHz com 512 MB de memória se mostrou inviável para o cálculo do algoritmo de *Estimation-Maximization*.

Foi montado um modelo com misturas de nove gaussianas bivariadas como apresentado nas Equações (4.4) e (4.5). Foi avaliado também um modelo com misturas de duas gaussianas bivariadas com as classes apresentadas na Figura 4.18 para o plano HS e na Figura 4.20 para as duas primeiras componentes principais do cubo RGB.

Uma avaliação de desempenho dos classificadores gaussianos está apresentada no Capítulo 5.

No próximo capítulo apresentamos os resultados do experimento.

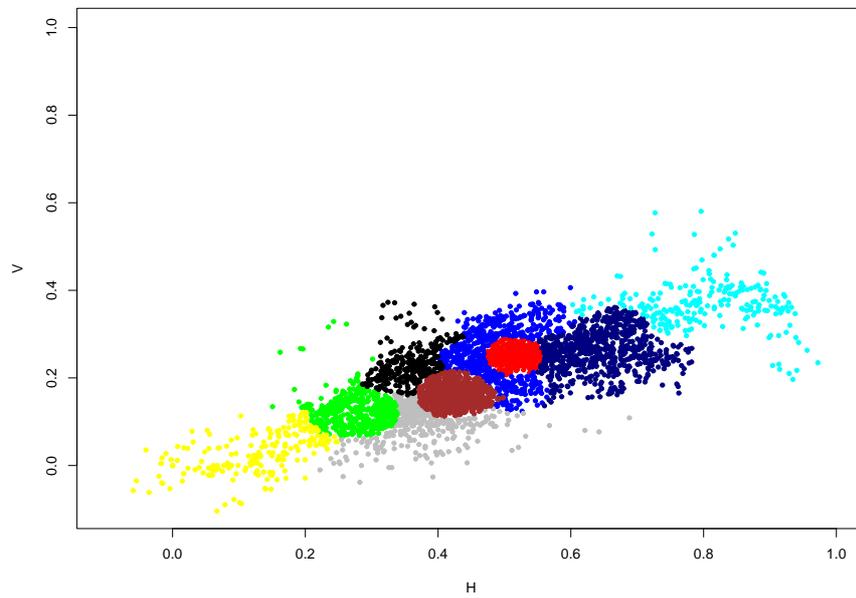


Figura 4.17: Separação dos *pixels* no plano HS em 9 classes distintas

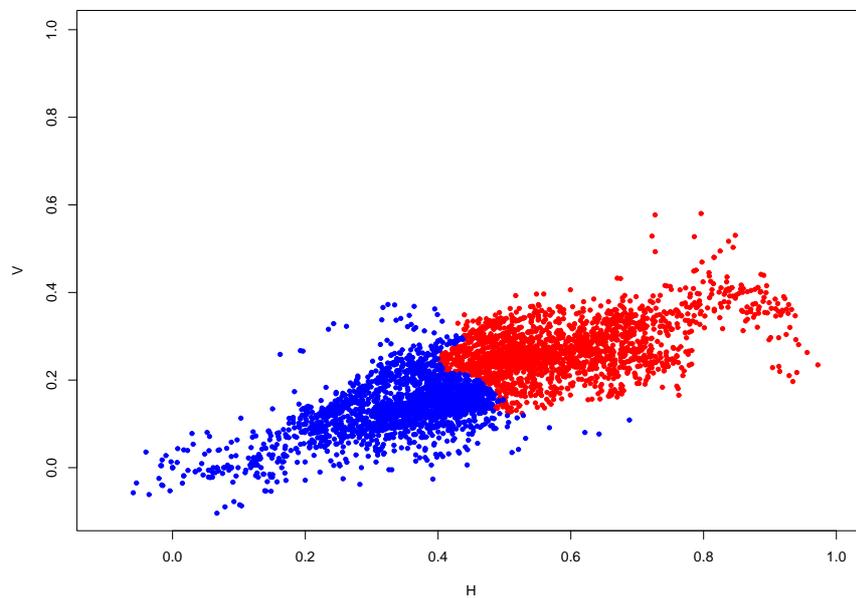


Figura 4.18: Separação dos *pixels* no plano HS em 2 classes distintas

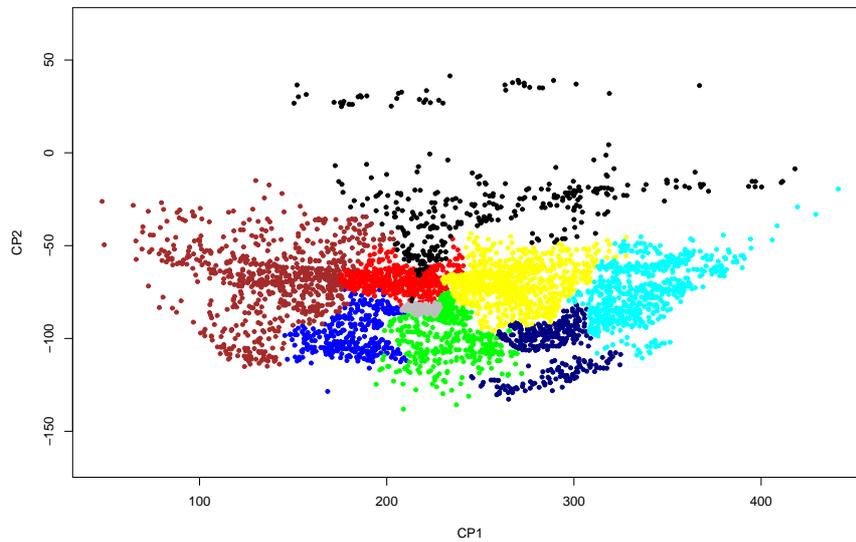


Figura 4.19: Separação dos *pixels* no plano CP1,CP2 em 9 classes distintas

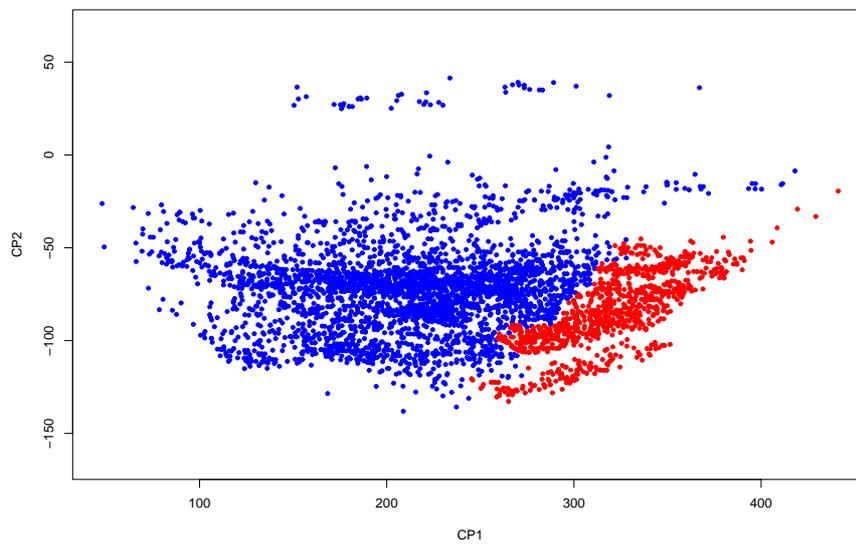


Figura 4.20: Separação dos *pixels* no plano CP1,CP2 em 2 classes distintas

# Capítulo 5

## Resultados

Foi montado um experimento para a avaliação das diversas técnicas apresentadas nesse trabalho e para a escolha de qual método irá compor o sistema de detecção de pele. Um resumo das técnicas utilizadas e dos dados de entrada está apresentado na Tabela 5.1. Nesta tabela, a entrada “–” indica que a técnica não foi avaliada com o dado de entrada correspondente.

A análise dos resultados se dá através de forma qualitativa e quantitativa. A análise qualitativa se dá apenas comparando visualmente a detecção de pele, através de imagens binárias<sup>1</sup> onde a cor branca indica que o *pixel* foi detectado como pele e a cor preta indica que o *pixel* não foi detectado como pele.

A análise quantitativa dos resultados é baseada na comparação de indicadores numéricos calculados a partir de informações extraídas após a aplicação de cada uma das técnicas de detecção de pele. A metodologia utilizada para avaliação das técnicas de detecção de pele está apresentada na Seção 2.4 página 25 e também será utilizada para avaliar a heurística proposta para correção de cor. Para este propósito, foi montado um experimento onde, primeiramente, comparamos as técnicas de detecção de pele —as que obtiveram mais sucesso dentre as apresentadas na Seção 4.5, sem a utilização de correção de cor, e posteriormente,

---

<sup>1</sup>Imagens que apresentam apenas duas cores, em geral preto e branco.

Tabela 5.1: Técnicas Avaliadas e seus dados de entrada

<b>Técnica</b>	<b>Dados de Entrada</b>			
	<b>RGB</b>	<b>HS</b>	<b>Cp1Cp2</b>	<b>C'rC'b</b>
Limiar de Classificação de Kovac et al. (2003)	✓	-	-	-
Limiar de Classificação de Rein-Lien et al. (2002)	-	-	-	✓
Histograma de cor - largura 1	✓	-	-	-
Histograma de cor - largura 64	✓	-	-	-
Histograma de cor suavizado	✓	-	-	-
Mistura de 2 Gaussianas bivariadas	-	✓	✓	-
Mistura de 9 Gaussianas bivariadas	-	✓	✓	-

---

comparamos as mesmas técnicas utilizando a correção de cor proposta. Finalmente, comparamos as quatro técnicas com mais sucesso, com e sem a correção de cor proposta. Para todas as técnicas, foi avaliado o desempenho através dos indicadores das matrizes confusão, das medidas derivadas dela e das curvas ROC.

Neste experimento, utilizamos a notação apresentada na Tabela 5.2 para identificação das técnicas utilizadas.

O Apêndice A mostra as matrizes de confusão referentes às técnicas de detecção de pele discutidas nesse trabalho seguindo a notação apresentada na Tabela 5.2. Nas matrizes de confusão estão apresentados os valores discutidos na Seção 2.4. Todos os classificadores avaliados fazem parte da categoria dos *Classificadores Supervisionados*, sendo dois classificadores de regra fixa (sem limiar de controle) apresentados por Kovac et al. (2003) e por Rein-Lien et al. (2002) e os outros são classificadores onde se utiliza o limiar de controle.

As matrizes de confusão apresentadas no Apêndice A foram geradas através das taxas de erro e de acerto descritas na Seção 2.4 quando utilizados os classificadores de regra fixa, e através das mesmas taxas de erro e de acerto para os outros classificadores, sendo que utilizando o valor do limiar de classificação  $\Theta$  que gerou a menor distância euclidiana (ver Equação (2.5)), ponderada com  $W = 0,8$ , ou seja, priorizando a classe pele.

O valor da precisão  $P$  para este mesmo limiar de classificação também está apresentado junto com as matrizes de confusão. Dessa forma, temos duas medidas quantitativas para comparar os classificadores utilizados.

Junto com as matrizes de confusão também estão apresentados os Verdadeiros Positivos (TP) e os Falsos Positivos (FP); esses valores são calculados diretamente dos valores apresentados na matriz de confusão mas estão apresentados para propiciar uma leitura mais direta.

Para evitar possíveis erros de avaliação (uma discussão mais aprofundada é encontrada em Hamilton et al. 2002), também utilizamos as curvas ROC. Três curvas ROC são apresentadas nas Figuras 5.1, 5.2 e

Tabela 5.2: Notação utilizada no experimento de avaliação das técnicas de detecção de pele

<b>Notação</b>	<b>Técnica</b>
<b>geoPeer</b>	Regra geométrica baseada no trabalho de Kovac et al. (2003)
<b>geoPeercc</b>	Regra geométrica baseada no trabalho de Kovac et al. (2003), com correção de cor
<b>geoHsu</b>	Regra geométrica baseada no trabalho de Rein-Lien et al. (2002)
<b>geoHsucc</b>	Regra geométrica baseada no trabalho de Rein-Lien et al. (2002), com correção de cor
<b>bin64</b>	Histogramas de cor com intervalo 64
<b>bin64cc</b>	Histogramas de cor com intervalo 64, com correção de cor
<b>spline</b>	Histograma de cor suavizado
<b>splinecc</b>	Histograma de cor suavizado com correção de cor
<b>pca2x</b>	Mistura de duas gaussianas bivariadas, utilizando PCA
<b>pca2xcc</b>	Mistura de 2 gaussianas bivariadas, utilizando PCA e correção de cor
<b>pca9x</b>	Mistura de 9 gaussianas bivariadas, utilizando PCA
<b>pca9xcc</b>	Mistura de 9 gaussianas bivariadas, utilizando PCA e correção de cor
<b>hs2x</b>	Mistura de 2 gaussianas bivariadas, utilizando o plano HS
<b>hs2xcc</b>	Mistura de 2 gaussianas bivariadas, utilizando o plano HS e correção de cor
<b>hs9x</b>	Mistura de 9 gaussianas bivariadas, utilizando o plano HS
<b>hs9xcc</b>	Mistura de 9 gaussianas bivariadas, utilizando o plano HS e correção de cor

5.3. Essa curva é formada pelos pares (TP,FP) para diversos valores de limiar de classificação  $\Theta$  para o caso dos classificadores onde que utilizam o limiar de controle e um ponto para os classificadores de regra fixa.

Os classificadores que utilizam o limiar de controle estão representados através de curvas com suas respectivas cores, e os classificadores de regra fixa estão representados através de um ponto com seus respectivos símbolos. Para o caso dos classificadores que utilizam o limiar de controle, foi destacado o ponto de menor distância euclidiana, representado por um pequeno círculo ao longo da curva.

Os resultados apresentados no Apêndice A, referentes aos classificadores que não utilizaram a correção de cor, indicam que os classificadores que apresentam os melhores indicadores são: (i) o classificador estatístico que utiliza a mistura de duas gaussianas bivariadas com dados de entrada extraídos das duas primeiras componentes principais dos dados RGB (*pca2x*), por apresentar a menor distância euclidiana (0.115782) e a maior precisão (0.66415), (ii) o classificador geométrico apresentado por Kovac et al. (2003) (*geoPeer*), com valores da distância euclidiana e de precisão bastante aproximados aos apresentados pela técnica anterior (0.122678 e 0.660195) e (iii) o classificador estatístico que utiliza a mistura de nove gaussianas bivariadas com dados de entrada extraídos das duas primeiras componentes principais dos dados RGB (*pca9x*), pelo menos motivo (0.119328 e 0.692130).

A escolha do melhor classificador a ser utilizado, pode ser auxiliada pela curva ROC apresentada na Figura 5.1. Nessa Figura temos uma comparação entre todos os classificadores apresentados na Tabela 5.2. Observando os três classificadores que apresentaram os melhores indicadores presentes no Apêndice A, quando não é utilizada a correção de cor, percebemos que o segundo classificador (*pca2x*), mostrado através da curva verde, apresenta valores superiores aos outros dois, representados pela curva azul para o classificador *pca9x*, e o pelo ponto  $\Delta$  para o classificador geométrico de Kovac et al. (2003), em quase todas as regiões.

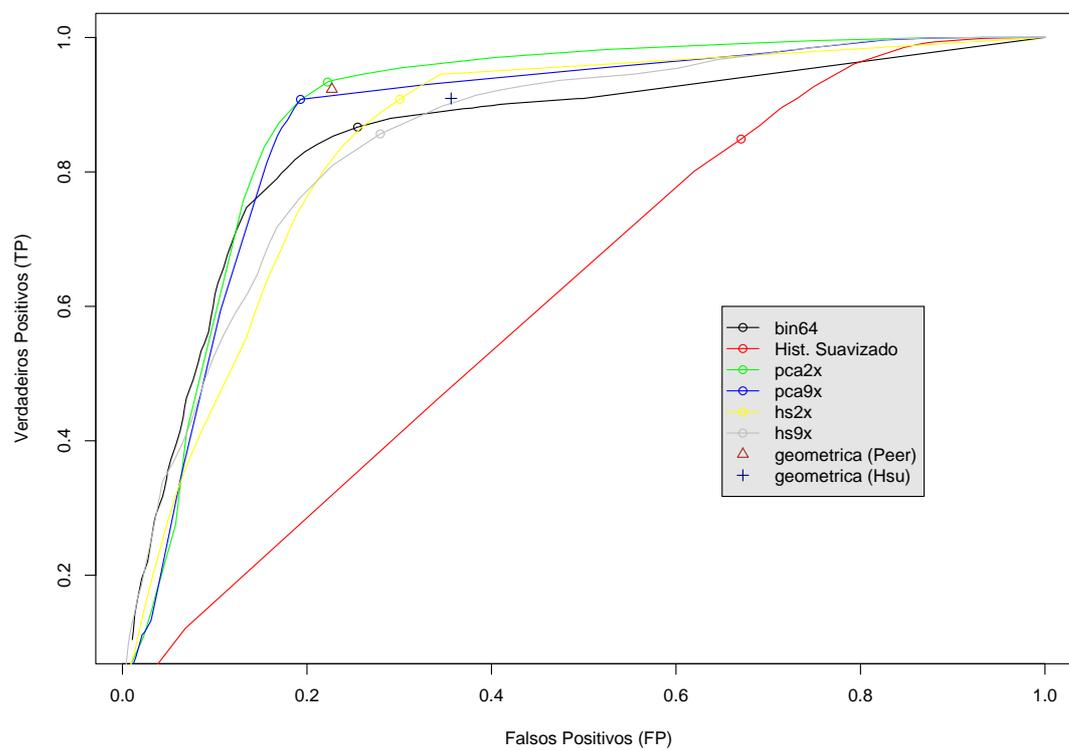


Figura 5.1: Curva ROC para os métodos analisados sem a correção de cor

---

A análise dos resultados apresentados no Apêndice A para os classificadores que utilizam como dados de entrada a imagem após a correção de cor, indica que os melhores classificadores são: (i) o classificador geométrico apresentado por Kovac et al. (2003) (*geoPeercc*), por apresentar a menor distância euclidiana (0.099711) e a precisão (0.708873), (ii) o classificador estatístico que utiliza a mistura de duas gaussianas bivariadas com dados de entrada extraídos das duas primeiras componentes principais dos dados RGB (*pca2xcc*), com valores da distância euclidiana e de precisão bastante aproximados aos apresentados pela técnica anterior (0.101641 e 0.720415) e (iii) o classificador estatístico que utiliza a mistura de nove gaussianas bivariadas com dados de entrada extraídos das duas primeiras componentes principais dos dados RGB (*pca9xcc*), por apresentar uma distância euclidiana semelhante às outras duas técnicas e a maior precisão (0.106334 e 0.758655).

Novamente temos os três classificadores que obtiveram os melhores indicadores, quando os aplicamos às imagens sem correção de cor, como sendo os que apresentam os melhores indicadores ao aplicá-los em imagens com a correção de cor proposta.

Para o caso de correção de cor, tivemos que o classificador que apresenta menor distância euclidiana e o que apresenta maior precisão, não coincidem, não tendo assim o melhor classificador e o pior classificador dentre os três. Para facilitar a escolha entre os três, podemos analisar as curvas ROC apresentadas na Figura 5.2. As mesmas cores e símbolos são utilizados para representar os classificadores avaliados; temos o classificador *pca2x* (curva verde) novamente superando os outros dois em quase todas as regiões da ROC.

Para avaliar a correção de cor proposta, utilizamos a Figura 5.3, onde temos quatro classificadores (*bin64*, *pca2x*, *pca9x* e *geoPeer*) aplicados às imagens originais e às imagens após a correção de pele. Vale observar que em todos os casos, para a amostra de testes, a correção de cor proposta melhorou os indicadores dos classificadores. Essa afirmativa pode ser verificada tanto na curva ROC da Figura 5.3, quanto nos dados do Apêndice A.

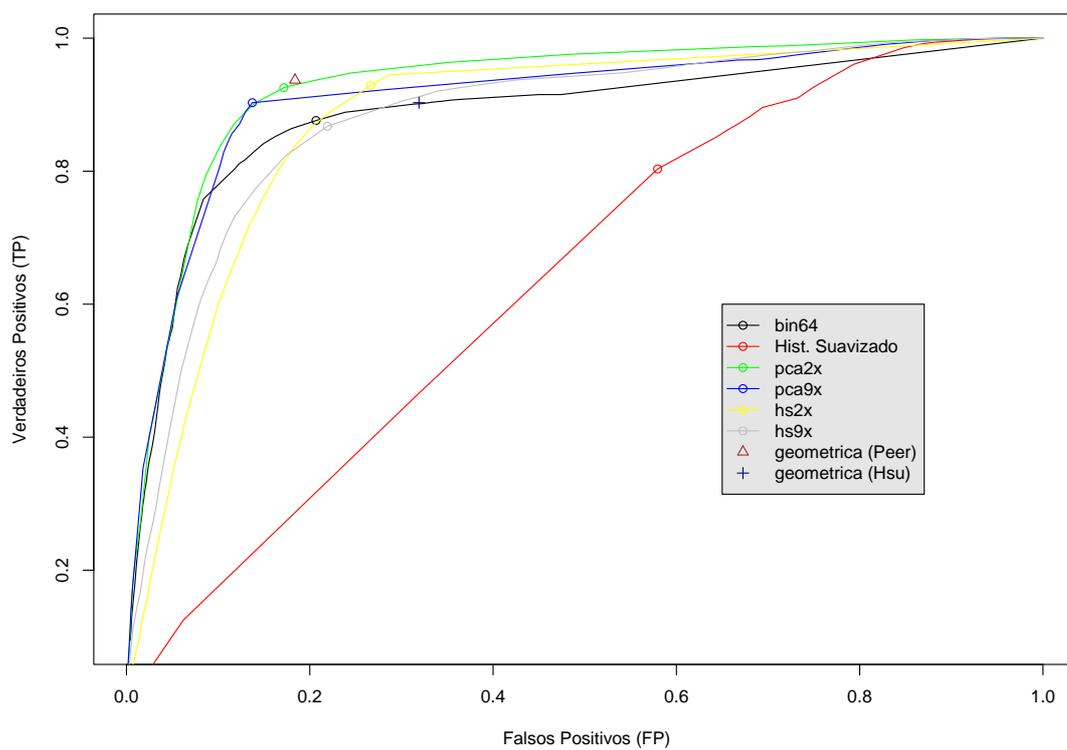


Figura 5.2: Curva ROC para os métodos analisados após a correção de cor

---

Através da análise acima, escolhemos o classificador estatístico que utiliza a mistura de duas gaussianas bivariadas com dados de entrada extraídos das duas primeiras componentes principais dos dados RGB (*pca2x*) para fazer a avaliação qualitativa, onde apresentamos as imagens presentes no Apêndice B. Nessas imagens, temos na primeira coluna as imagens originais, na segunda coluna as imagens após a correção de cor proposta, na terceira coluna a detecção de pele utilizando o classificador *pca2x* nas imagens originais e, na quarta coluna, a detecção de pele utilizando o classificador *pca2x* nas imagens após a correção de pele. As imagens apresentadas nas terceiras e quartas colunas codificam na cor preta os *pixels* que não foram classificados como pele e na cor branca os *pixels* que foram classificados como pele.

No Apêndice B temos diversas imagens coletadas na Web que pertencem à base de testes e outras que não pertencem à base de testes. É fácil perceber que, de uma forma geral, a correção de cor proposta diminui a classificação errônea de paisagens, paredes e objetos que, ao serem submetidos à iluminação inadequada ficam com coloração semelhante à cor da pele e provocam erro de comissão; ver as Figuras B.01, B.05, B.06, B.07, B.09, B.17, B.18, B.22, B.24, B.29 e B.30.

Todas as técnicas de detecção de pele avaliadas, sem utilizar correção de cor, independente do espaço de cor em uso, apresentaram problemas com a variação de iluminação. Em alguns casos, a correção de cor diminui o erro de comissão ao custo de aumentar o erro de omissão (ver Figuras B.04, B.05, B.06, B.09, B.12, B.14, B.15, B.18, B.20, B.21 e B.27). Para esses casos, a utilização da correção de cor pode não melhorar a qualidade geral da classificação. Por este motivo, a heurística proposta na Seção 4.4 visa evitar a aplicação da correção de cor em todas as imagens.

A correção de cor poderá tanto diminuir os erros de comissão quanto aumentar os erros de omissão na mesma imagem. Este fato está destacado nas Figuras que foram citadas para os dois casos (Figuras B.05, B.06, B.09 e B.18). Em algumas situações, a correção de cor não provoca grandes alterações na imagem, não modificando de forma signi-

ficativa a detecção de pele (ver Figuras B.02, B.03, B.10, B.11, B.13, B.16, B.19, B.23, B.25, B.26 e B.28); este fato pode ser facilmente justificado quando o algoritmo encontra todos os pontos mais brilhantes da imagem como sendo o branco RGB = (255, 255, 255).

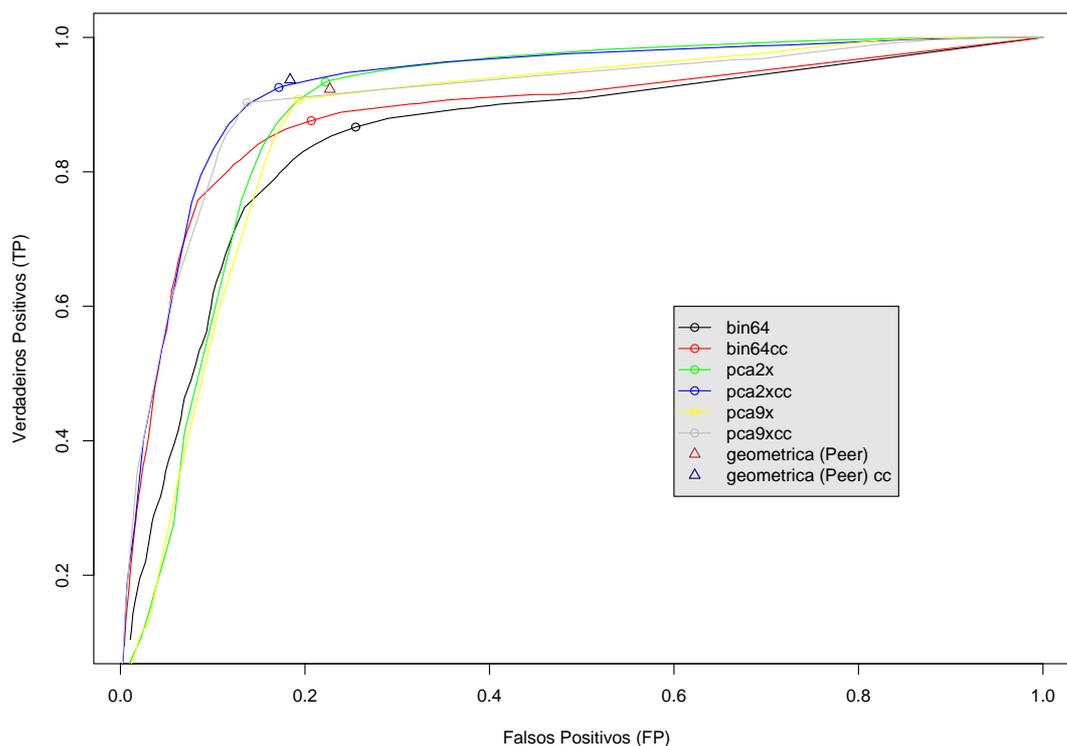


Figura 5.3: Curva ROC para os métodos analisados com e sem correção de cor

Os dois exemplos subsequentes ilustram a influência da variação de iluminação para detecção de pele. Eles estão destacados do Apêndice B por apresentarem resultados bastante marcantes com relação à variação entre a iluminação incandescente e a iluminação fluorescente.

As Figuras 5.4 e 5.5 mostram dois exemplos de detecção de pele com e sem correção de cor. Na Figura 5.4(a) e 5.5(a), temos as imagens originais, nas Figuras 5.4(b) e 5.5(b) temos o resultado da correção de cor, utilizando a heurística proposta, aplicada à imagem original. Nas Figu-

ras 5.4(c) e 5.5(c) temos a detecção de pele utilizando *pca2x* aplicado à imagem original. Nas Figuras 5.4(d) e 5.5(d) temos a detecção de pele, utilizando a mesma técnica na imagem com a correção de cor baseada na heurística proposta (*pca2xcc*).

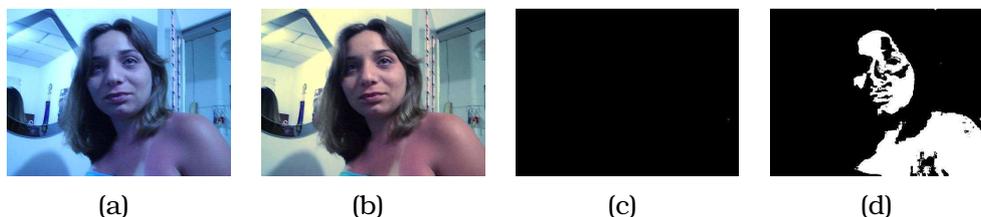


Figura 5.4: Detecção de pele em uma imagem azulada (iluminação fluorescente)

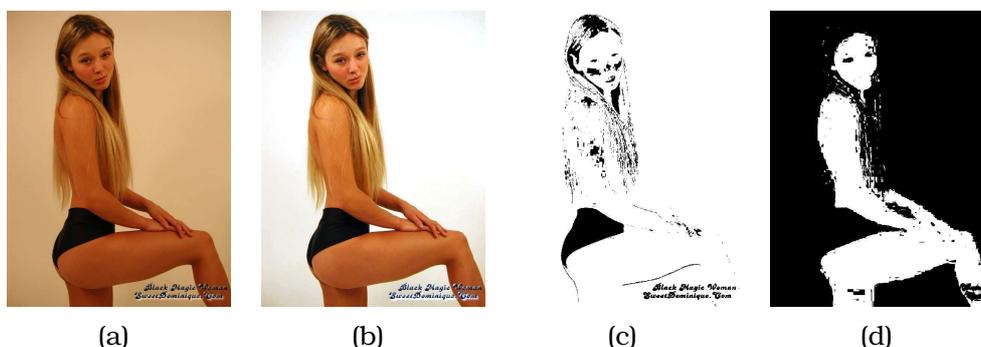


Figura 5.5: Detecção de pele em uma imagem amarelada (iluminação incandescente)

Percebe-se que a correção de cor baseada na heurística proposta pode melhorar consideravelmente a detecção de pele corrigindo erros de comissão e erros de omissão: na Figura 5.4(c) nenhum *pixel* foi classificado como pele (erro de omissão) devido a aparência azulada provocada pela iluminação fluorescente e a Figura 5.4(d) apresenta uma detecção de pele muito mais eficiente após a aplicação da correção de cor com a heurística apresentada. Da mesma forma, a Figura 5.5(c) mostra a detecção de pele com um número exagerado de *pixels* marcados como pele (erro de comissão) devido à coloração amarelada provocada pela iluminação incandescente e na Figura 5.5(d) temos uma classificação

mais eficiente após a utilização da correção de cor baseada na heurística proposta.

No próximo capítulo apresentaremos as conclusões do trabalho e sugestões para trabalhos futuros.

# Capítulo 6

## Conclusões e Trabalhos Futuros

### 6.1 Conclusões

A detecção de pele em imagens veiculadas na Web foi o objetivo principal deste trabalho. Pelo fato da maioria das imagens encontradas na Web apresentar informação extremamente variável em termos de resolução, iluminação e conteúdo, o trabalho aqui proposto não pode beneficiar-se de abordagens que são muito bem sucedidas em domínios restritos ou controlados. Foi, portanto, necessário desenvolver uma técnica geral apta a lidar com uma grande diversidade de situações. Os trabalhos de Rein-Lien et al. (2002), Bretzner et al. (2002) e Albiol & Torres (2001), entre outros, apresentam restrições intrínsecas da aplicação fazendo com que as imagens analisadas sejam mais específicas do que as imagens presentes na Web, pois supõem a presença de face humana e de mãos humanas, por exemplo. No trabalho de Phung et al. (2005) uma base de imagens que utiliza faces e pessoas também fazem parte de um universo mais restrito pois podemos supor a presença de pessoas nas imagens. As imagens depositadas nos servidores Web apresentam diferentes resoluções, tipos de iluminação e qualidade o que torna a detecção de pele difícil. Além disso, as pessoas presentes nessas imagens apresentam diversas etnias, aparecem em diferentes poses e, em geral, aparecem sobre uma paisagem complexa a respeito da qual nada

se sabe a priori.

Com todas as características apresentadas, o desafio desse trabalho foi realizar a especificação, desenvolvimento e avaliação de um sistema de detecção de pele apto a ser acoplado a outros sistemas. O sistema aqui desenvolvido poderá integrar-se a sistemas de detecção de nudez. Um sistema desse tipo está sendo atualmente desenvolvido, mas as ferramentas desenvolvidas para detecção de pele, que formam o cerne desta dissertação, poderão ser acopladas a diversos outros sistemas como fora apresentado na Seção 3.2.

As conclusões resultantes do desenvolvimento desta dissertação serão agora apresentadas seguindo a divisão funcional do sistema apresentado na Seção 4.1.

Uma importante conclusão que podemos apresentar após o desenvolvimento desse trabalho é que, para imagens para as quais não há controle sobre a qualidade e nem sobre o tipo de iluminação utilizada, uma técnica consistente de correção de cor se faz necessária para que uma boa classificação seja alcançada. Negligenciar esta etapa poderá redundar em erros grosseiros de classificação. Este fato foi apresentado exhaustivamente na Seção 4.4.

A adaptação cromática presente no sistema de visão dos seres humanos faz com que as cores se preservem aproximadamente constantes mesmo quando há variação da iluminação, mas essa característica não é reproduzida com a mesma eficiência pelos sistemas computacionais. Observou-se que determinadas técnicas de correção de cor podem melhorar consideravelmente a detecção de pele quando aplicadas a algumas imagens, mas que podem piorar drasticamente a detecção de pele em outros casos. Dessa forma, algumas técnicas foram avaliadas neste trabalho, e uma heurística foi proposta para que fosse feita uma seleção de qual imagem deve ser submetida às técnicas de correção de cor. O resultado obtido para esta etapa é que a heurística proposta neste trabalho sugere que a correção de cor seja efetuada por uma das técnicas *White Patch* ou *Gray World*, ou nenhuma técnica. O uso desta heurística provocou uma melhora considerável na etapa de detecção de pele,

quando aplicada à base de treinamento montada para a construção do classificador. Testes envolvendo esse método de correção de cor devem ser exaustivamente aplicados em bases de testes maiores para verificar a escalabilidade do método com relação a uma diversidade maior de imagens.

A cor, para os seres humanos, é descrita em espaços tridimensionais e alguns desses espaços fazem uma distinção entre luminância e crominância. Baseados nisso, diversos autores (como por exemplo, Caetano & Barone 2001, Storrington et al. 1999) utilizam esse tipo de representação para a cor e descartam a informação de luminância com o propósito de tornar a cor da pele independente da iluminação presente na cena onde a imagem foi adquirida. Essa abordagem não se mostrou suficientemente robusta.

Storrington et al. (1999) apresentam uma discussão bastante completa sobre o deslocamento que o agrupamento de cores de pele sofre no plano cromático com a variação de iluminação e Rein-Lien et al. (2002) e Shin et al. (2002) sugerem que a utilização do descarte da dimensão de luminância não apresenta os resultados esperados. A utilização das chamadas “cores puras” (rg normalizado) também é bastante popular com o mesmo objetivo, porém Terrillon et al. (2000) indicam que o uso dessas cores não apresenta bons resultados para a detecção de pele com o intuito de reconhecimento de faces. Esse resultado pode ser estendido para a detecção de pele que estamos interessados.

Uma técnica alternativa à mudança de espaço de cor foi apresentada na Seção 4.3, e foi aplicada aos dados RGB da amostra de cores de pele colhida na base de treinamento: a transformação de componentes principais. Os modelos que utilizam essas informações como dados de entrada apresentaram os melhores indicadores de qualidade de classificação ao detectarem pele nas imagens de nosso interesse. Esse fato sugere que a transformação de componentes principais preserva mais informações sobre a amostra de cores de pele, mesmo ao desprezarmos a terceira componente (de menor variância), quando comparada ao descarte da dimensão de luminância dos espaços de cor avaliados. Este

fato reforça a hipótese que cor da pele é dependente da luminância e que a redução de dimensionalidade apenas diminui o custo computacional despendido para a implementação do modelo e melhora a tratabilidade dos modelos matemáticos para a detecção de pele, mas não torna a cor da pele mais independente da iluminação da cena.

Com a correção de cor apresentada e a redução de dimensionalidade através de transformações de componentes principais, temos informações suficientemente confiáveis para montar um modelo matemático que descreva bem a amostra de cores de pele. Para esse fim, diversas técnicas foram avaliadas e apresentados os resultados no Capítulos 5.

A técnica de histograma de cores não apresentou resultados tão bons quanto outras técnicas, porém pode ser melhorada aumentando o tamanho das amostras empregadas; isso foi detalhado na Seção 4.5. O trabalho de Jones & Rehg (2002) utilizou uma base de dados da ordem de  $10^9$  *pixels* (bases de teste e treinamento) e os autores relatam que os resultados foram satisfatórios. Nenhuma correção de cor foi proposta neste trabalho o que torna esse classificador susceptível às variações de iluminação.

O modelo apresentado por Kovac et al. (2003) apresenta uma regra geométrica que resultou em indicadores de qualidade semelhantes aos da técnica proposta neste trabalho, quando aplicada junto com a correção de cor aqui proposta. Os classificadores estatísticos paramétricos avaliados nesse trabalho como modelo misturas de gaussianas. Apesar da técnica de *Expectation Maximization* sugerir uma mistura composta de nove gaussianas bivaradas, o classificador gaussiano baseado em uma mistura de duas gaussianas bivaridas apresenta melhores resultados. Este fato pode ter ocorrido devido a um “sobre-treinamento” do modelo com relação à amostra de treinamento. Dessa forma, o classificador baseado em uma mistura de duas gaussianas bivariadas apresentou os melhores indicadores de desempenho de acordo com o explanado no Capítulo 5.

Basedo a discussão acima, os resultados do presente trabalho mostram:

- ☞ que a utilização de correção de cor para a detecção de pele em imagens veiculadas na Web é imprescindível para alcançar boas taxas de acerto na detecção de pele;
- ☞ que é suficiente usar apenas as duas primeiras componentes principais como características de entrada no classificador, pois há nelas informações suficientes para a criação de modelos confiáveis de detecção de pele;
- ☞ que o classificador estatístico paramétrico que utiliza uma mistura de duas gaussianas (*pca2xcc*) produz resultados que atendem os propósitos deste trabalho.

## 6.2 Trabalhos Futuros

O presente trabalho não está esgotado. Durante a sua execução identificamos diversos trabalhos que poderão dar-lhe continuidade. A heurística para a utilização da correção de cor proposta neste trabalho é bastante restritiva de modo que diversas imagens poderiam apresentar resultados de detecção de pele melhores. Em virtude da possibilidade de piorar a detecção de pele em outros casos, optou-se por deixar a heurística assim restritiva. Todavia, uma técnica de correção de cor mais geral que possa ser aplicável a todas as imagens, apenas mudando significativamente a imagem quando se fizer necessário, é extremamente desejável, ficando como a primeira sugestão para um trabalho futuro.

Há ainda várias possíveis técnicas de identificação de pele que podem ser aplicadas uma vez feita a correção de cor a contento; dentre elas podemos citar

- outros valores de  $k$  no modelo de mistura de gaussianas (neste trabalho foram utilizadas duas e nove);
- o uso da envolvente convexas dos dados como regra geométrica (a função *chull* da plataforma R fornece calcula em  $\mathbb{R}^2$ );

- regras baseadas em entropia (Jedynak et al. 2002) e redes neurais (Brown et al. 2001, Phung et al. 2005).

A implementação do sistema de detecção de pornografia proposto nessa dissertação, bem como uma avaliação de desempenho criteriosa como a apresentada para a detecção de pele consiste em outra indicação de trabalho futuro. Nessa etapa, técnicas de extração de atributos e reconhecimento de padrões serão utilizadas para destacar as características que irão ser apresentadas ao módulo qualificador. Algumas características foram extraídas das imagens classificadas, e o seu potencial discriminador se mostrou promissor.

A aplicação do sistema de detecção de pele aqui proposto em outras áreas como, por exemplo, nos servidores de hospedagem, servidores de e-mail, servidores *proxys*, servidores MMC e nos sistemas de auditoria, serviria como uma excelente avaliação da escalabilidade da solução aqui proposta.

## Apêndice A

# Matrizes de confusão para as Técnicas de Detecção de Pele Avaliadas

Tabela A.1: Matriz de confusão para o classificador – Regra geométrica de Kovac et al. (2003) sem correção de cor (geoPeer)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>6.722.064</b>	<b>1.974.254</b>
	Positivo	<b>319.950</b>	<b>3.835.771</b>
Distância Euclidiana ( $AD_d$ )		0.122678	
Verdadeiros Positivos (TP)		0.923010	
Falsos Positivos (FP)		0.227022	
Precisão (P)		0.660195	

Tabela A.2: Matriz de confusão para o classificador – Regra geométrica de Kovac et al. (2003) com correção de cor (geoPeercc)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>7.097.240</b>	<b>1.599.078</b>
	Positivo	<b>262.002</b>	<b>3.893.719</b>
Distância Euclidiana ( $AD_d$ )		0.099711	
Verdadeiros Positivos (TP)		0.936954	
Falsos Positivos (FP)		0.183880	
Precisão (P)		0.708873	

Tabela A.3: Matriz de confusão para o classificador – Regra geométrica de Rein-Lien et al. (2002) sem correção de cor (geoHsu)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>5.598.219</b>	<b>3.098.099</b>
	Positivo	<b>376.867</b>	<b>3.778.854</b>
Distância Euclidiana ( $AD_d$ )		0.178781	
Verdadeiros Positivos (TP)		0.909314	
Falsos Positivos (FP)		0.356254	
Precisão (P)		0.549493	

Tabela A.4: Matriz de confusão para o classificador – Regra geométrica de Rein-Lien et al. (2002) com correção de cor (geoHsucc)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>5.920.866</b>	<b>2.775.452</b>
	Positivo	<b>403.363</b>	<b>3.752.358</b>
Distância Euclidiana ( $AD_d$ )		0.167058	
Verdadeiros Positivos (TP)		0.902938	
Falsos Positivos (FP)		0.319153	
Precisão (P)		0.574823	

Tabela A.5: Matriz de confusão para o classificador – Histograma sem correção de cor (bin64)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>6.479.411</b>	<b>2.216.907</b>
	Positivo	<b>554.138</b>	<b>3.601.583</b>
Distância Euclidiana ( $AD_d$ )		0.164990	
Verdadeiros Positivos (TP)		0.866657	
Falsos Positivos (FP)		0.254925	
Precisão (P)		0.618989	

Tabela A.6: Matriz de confusão para o classificador – Histograma com correção de cor (bin64cc)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>6.897.991</b>	<b>1.798.327</b>
	Positivo	<b>514.742</b>	<b>3.640.979</b>
Distância Euclidiana ( $AD_d$ )		0.144313	
Verdadeiros Positivos (TP)		0.876137	
Falsos Positivos (FP)		0.206792	
Precisão (P)		0.669383	

Tabela A.7: Matriz de confusão para o classificador – Histograma suavizado sem correção de cor (spline)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>2.864.761</b>	<b>5.831.557</b>
	Positivo	<b>628.110</b>	<b>3.527.611</b>
Distância Euclidiana ( $AD_d$ )		0.328953	
Verdadeiros Positivos (TP)		0.848857	
Falsos Positivos (FP)		0.670578	
Precisão (P)		0.376915	

Tabela A.8: Matriz de confusão para o classificador – Histograma suavizado com correção de cor (splinecc)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>3.656.957</b>	<b>5.039.361</b>
	Positivo	<b>817.189</b>	<b>3.338.532</b>
Distância Euclidiana ( $AD_d$ )		0.313200	
Verdadeiros Positivos (TP)		0.803358	
Falsos Positivos (FP)		0.579482	
Precisão (P)		0.398493	

Tabela A.9: Matriz de confusão para o classificador – Mistura Gaussiana Bivariada (PCA) sem correção de cor (pca2x)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>6.762.787</b>	<b>1.933.531</b>
	Positivo	<b>275.601</b>	<b>3.880.120</b>
Distância Euclidiana ( $AD_d$ )		0.115782	
Verdadeiros Positivos (TP)		0.933682	
Falsos Positivos (FP)		0.222339	
Precisão (P)		0.667415	

Tabela A.10: Matriz de confusão para o classificador – Mistura Gaussiana Bivariada (PCA) com correção de cor (pca2xcc)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>7.203.662</b>	<b>1.492.656</b>
	Positivo	<b>309.546</b>	<b>3.846.175</b>
Distância Euclidiana ( $AD_d$ )		0.101641	
Verdadeiros Positivos (TP)		0.925513	
Falsos Positivos (FP)		0.171642	
Precisão (P)		0.720415	

Tabela A.11: Matriz de confusão para o classificador – Mistura Gaussiana 9x Variada (PCA) sem correção de cor (pca9x)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>7.018.098</b>	<b>1.678.220</b>
	Positivo	<b>382.881</b>	<b>3.772.840</b>
Distância Euclidiana ( $AD_d$ )		0.119328	
Verdadeiros Positivos (TP)		0.907867	
Falsos Positivos (FP)		0.192981	
Precisão (P)		0.692130	

Tabela A.12: Matriz de confusão para o classificador – Mistura Gaussiana 9x Variada (PCA) com correção de cor (pca9xcc)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>7.502.622</b>	<b>1.193.696</b>
	Positivo	<b>403.409</b>	<b>3.752.312</b>
Distância Euclidiana ( $AD_d$ )		0.106334	
Verdadeiros Positivos (TP)		0.902927	
Falsos Positivos (FP)		0.137265	
Precisão (P)		0.758655	

Tabela A.13: Matriz de confusão para o classificador – Mistura Gaussiana Bivariada (HS) sem correção de cor (hs2x)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>6.082.202</b>	<b>2.614.116</b>
	Positivo	<b>381.972</b>	<b>3.773.749</b>
Distância Euclidiana ( $AD_d$ )		0.157578	
Verdadeiros Positivos (TP)		0.908085	
Falsos Positivos (FP)		0.300600	
Precisão (P)		0.590768	

Tabela A.14: Matriz de confusão para o classificador – Mistura Gaussiana Bivariada (HS) com correção de cor (hs2xcc)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>6.379.087</b>	<b>2.317.231</b>
	Positivo	<b>295.491</b>	<b>3.860.230</b>
Distância Euclidiana ( $AD_d$ )		0.135074	
Verdadeiros Positivos (TP)		0.928895	
Falsos Positivos (FP)		0.266461	
Precisão (P)		0.624889	

Tabela A.15: Matriz de confusão para o classificador – Mistura Gaussiana 9x Variada (HS) sem correção de cor (hs9x)

		Classificação	
		Negativo	Positivo
Real	Negativo	<b>6.266.460</b>	<b>2.429.858</b>
	Positivo	<b>596.286</b>	<b>3.559.435</b>
Distância Euclidiana ( $AD_d$ )		0.179122	
Verdadeiros Positivos (TP)		0.856514	
Falsos Positivos (FP)		0.279412	
Precisão (P)		0.594300	

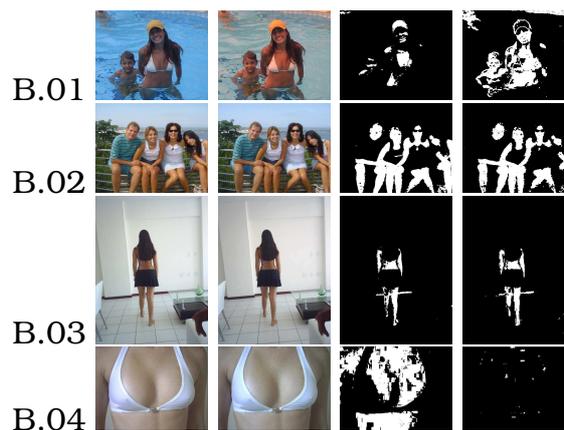
Tabela A.16: Matriz de confusão para o classificador – Mistura Gaussiana 9x Variada (HS) com correção de cor (hs9xcc)

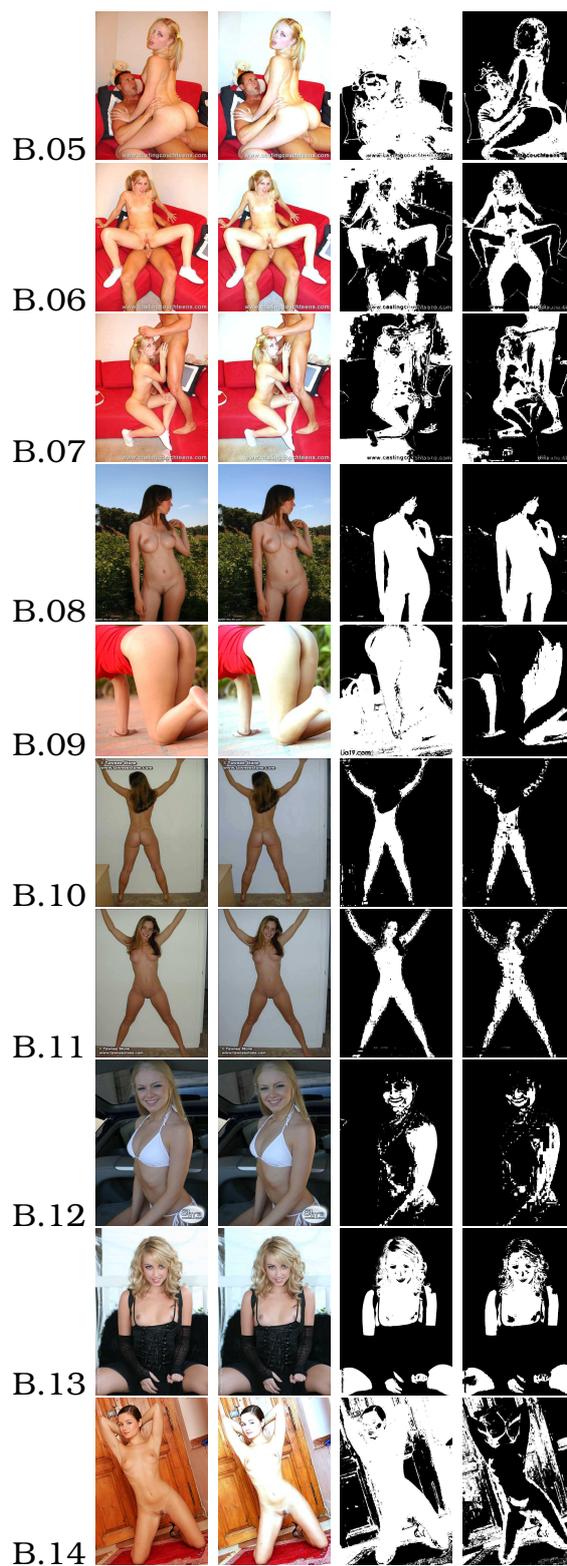
		Classificação	
		Negativo	Positivo
Real	Negativo	<b>6.789.657</b>	<b>1.906.661</b>
	Positivo	<b>550.517</b>	<b>3.605.204</b>
Distância Euclidiana ( $AD_d$ )		0.153796	
Verdadeiros Positivos (TP)		0.867528	
Falsos Positivos (FP)		0.219249	
Precisão (P)		0.654081	

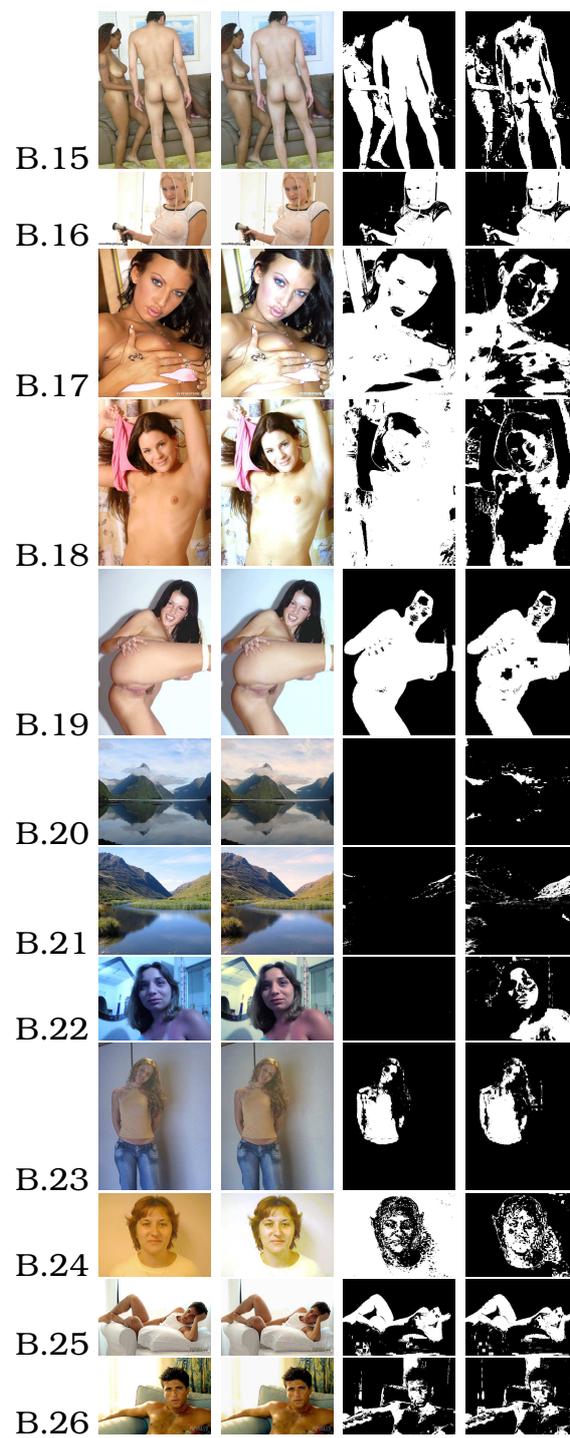
## Apêndice B

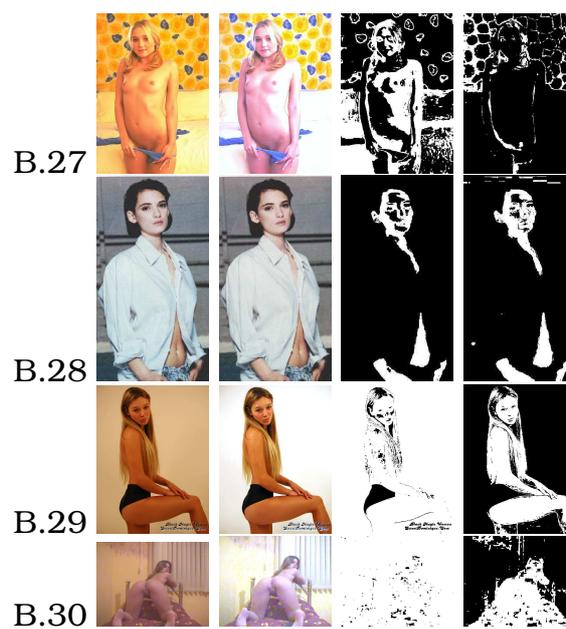
# Exemplos de Detecção de Pele em Imagens Tipicamente Encontradas na Web

Neste apêndice, temos vários exemplos de detecção de pele utilizando mistura de duas gaussianas bidimensionais. Este modelo utiliza as duas primeiras componentes principais como dados de entrada. Na primeira coluna temos a imagem original, na segunda coluna temos a imagem gerada após a correção de cor, na terceira coluna temos a detecção de pele utilizando a imagem original e na quarta e última coluna temos a detecção de pele com a imagem após a correção de cor.









## Referências Bibliográficas

- Albiol, A. & Torres, L. (2001), An unsupervised color image segmentation algorithm for face detection applications, *in* 'Proc. of International Conference on Image Processing', Vol. 2, IEEE, pp. 681–684.
- Bosson, A., Cawley, B., Chan, Y. & Harvey, R. (2002), Non-retrieval: Blocking pornographic images, *in* 'Proceedings of the International Conference on Image and Video Retrieval', Springer-Verlag, pp. 50–60.
- Brainard, D. H. & Wandell, B. A. (1986), 'Analysis of the retinex theory of color vision', *Journal of Optical Society* **3**, 1651–1661.
- Bretzner, L., Laptev, I. & Lindeberg, T. (2002), Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering, *in* 'FGR '02: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition', IEEE Computer Society, Washington, DC, USA, p. 423.
- Brown, D., Craw, I. & Lewthwaite, J. (2001), A SOM based approach to skin detection with application in real time systems, *in* 'The British Machine Vision Conference'.
- Bustos, O. H. & Frery, A. C. (1992), 'Reporting Monte Carlo results in statistics: suggestions and an example', *Revista de la Sociedad Chilena de Estadística* **9**(2), 46–95.

- Cabral, M. & Frery, A. C. (2003), 'Mascaramento de informações: histórico, definições e aplicações', *Revista Eletrônica de Iniciação Científica* **III**(1), 1–20.
- Caetano, T. & Barone, D. (2001), A probabilistic model for the human skin color, in 'Proceedings of 11th International Conference on Image Analysis and Processing ICIAP'01', IEEE, pp. 279–284.
- Caetano, T., Olabarriaga, S. & Barone, D. (2002), Performance evaluation of single and multiple-Gaussian models for skin color modeling, in 'Proceedings XV Brazilian Symposium on Computer Graphics and Image Processing SIBGRAPI', IEEE, pp. 275–282.
- Chan, Y., Harvey, R. & Smith, D. (1999), Building systems to block pornography, in 'Challenge of Image Retrieval, BCS Electronic Workshops in Computing series', British Computer Society, pp. 34–40.
- Duda, R. O., Hart, P. E. & Stork, D. G. (2001), *Pattern classification*, 2 edn, John Wiley & Sons, New York.
- Ebner, M. (2003), Combining white-patch retinex and the gray world assumption to achieve color constancy for multiple illuminants, in 'Lecture Notes on Computer Science: Pattern Recognition', Vol. 2781/2003, Springer-Verlag, pp. 60–67.
- Finlayson, G. D., Hordley, S. D. & Hubel, P. M. (2001), 'Color by correlation: A simple, unifying framework for color constancy', *IEEE Transactionis on Pattern Analysis and Machine Intelligence* **23**(11), 1209–1221.
- Fleck, M., Forsyth, D. & Bregler, C. (1996), Finding naked people, in 'European Conference on Computer Vision', Vol. II, pp. 592–602.
- Forsyth, D. (1990), 'A novel algorithm for color constancy', *International Journal of Computer Vision* **5**(1), 5–36.

- Forsyth, D. & Fleck, M. (1996), Identifying nude people, in 'Proceedings of 3th IEE Workshop on Applications of Computer Vision', IEEE Press, pp. 103–108.
- Fukunaga, K. (1972), *Introduction to Statistical Pattern Recognition*, Electrical Science Series, Academic, New York.
- Gasparini, F. & Schettini, R. (2003), Color correction for digital photographs, in 'Proceedings of 12th International Conference on Image Analysis and Processing (ICIAP03)', Vol. 2, IEEE Computer Society, pp. 646–651.
- Gomes, J. & Velho, L. (1994), *Computação Gráfica: Imagem*, Série Computação e Matemática, IMPA/SBM, Rio de Janeiro, RJ, Brazil.
- Gomez, G. & Morales, E. (2002), Automatic feature construction and a simple rule induction algorithm for skin detection, in 'Proceedings of the ICML Workshop on Machine Learning in Computer Vision', Vol. 1, Sydney, Australia, pp. 31–38.
- Hamilton, H., Gurak, E., Findlater, L. & Olive, W. (2002), 'Computer science 831: Knowledge discovery in databases', Página Web. Última visita em Janeiro de 2006.
- INPE (2006), 'SPRING: Transf. por principais componentes', Página Web. Última visita: Janeiro de 2006.
- Jedynak, B., Zheng, H., Daoudi, M. & Barret, D. (2002), Maximum entropy models for skin detection, in 'Third Indian Conference On Computer Vision, Graphics and Image Processing, ICVGIP', pp. 276–281.
- Jones, M. & Rehg, J. (2002), Statistical color models with application to skin detection, in 'International Journal Of Computer Vision', Kluwer, pp. 81–96.

- Kovac, J., Peer, P. & Solina, F. (2003), Human skin colour clustering for face detection, *in* 'EUROCON 2003 - International Conference on Computer as a Tool', Vol. 2, IEEE, pp. 144–148.
- Krzanowski, W. J. (1995), *Recent Advances in Descriptive Multivariate Analysis*, Vol. 2 of *Royal Statistical Society Lecture Note Series*, Clarendon Press, Oxford.
- Land, E. (1977), 'The retinex theory of color vision', *Scientific American* **237**, 108–129.
- Laurent, C., Laurent, N. & Bodo, Y. (2003), A human skin detector combining mean shift analysis and watershed algorithm, *in* 'Proc. of International Conference on Image Processing', Vol. III, IEEE, pp. 1025–1028.
- Lira Chávez, J. (2002), *Introducción al tratamiento digital de imágenes*, 1 edn, Instituto Politécnico Nacional, Centro de investigación en Computación, México.
- Martinkauppi, B., Soriano, M. & Pietikainen, M. (2003), Detection of skin color under changing illumination: A comparative study, *in* '12th International Conference on Image Analysis and Processing (ICIAP'03)', p. 652.
- Miller, K. S. (1964), *Multidimensional Gaussian Distributions*, The SIAM Series in Applied Mathematics, John Wiley and Sons, New York.
- Phung, S. L., Bouzerdoum, A. & Chai, D. (2005), 'Skin segmentation using color pixel classification: Analysis and comparison', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(1), 148–154.
- Poynton, C. (2002), 'Color faq - frequently asked questions color', Página Web. Última visita em outubro de 2005.

- R Development Core Team (2005), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Rein-Lien, H., Abdel-Mottaleb, M. & Jain, A. K. (2002), 'Face detection in color images', *IEEE Trans. Pattern Analysis and Machine Intelligence* **24**(5), 696–706.
- Resnick, P. & Miller, J. (1996), 'Pics: Internet access controls without censorship', *Communications of the ACM* **39**(10), 87–93.
- Richards, J. A. & Xiuping, J. (1999), *Remote Sensing Digital Image Analysis: An Introduction*, 2 edn, Springer, Berlin.
- Rizzi, A., Gatta, C. & Marini, D. (2002), Color correction between gray world and white patch, in 'Proceedings of SPIE', Vol. 4662 - Human Vision and Electronic Imaging, SPIE, pp. 367–376.
- Russell, S. & Norvig, P. (1995), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ.
- Schalkoff, R. (1989), *Digital Image Processing and Computer Vision*, John Wiley and Sons.
- Shin, M. C., Chang, K. I. & Tsap, L. V. (2002), Does colorspace transformation make any difference on skin detection?, in 'WACV '02: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision', IEEE Computer Society, Washington, DC, USA, p. 275.
- Skarbek, W. & Koschan, A. (1994), Colour image segmentation: a survey, Technical report, University of Berlin.
- Sleepycat (2005), 'Sleepycat software: Berkeley DB database, native XML database, native Java database', Página Web. Última visita em julho de 2005.

- Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A. & Jain, R. (2000), 'Content-based image retrieval at the end of the early years', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(12), 1349–1380.
- Smith, G. J. H. (2001), *Internet Law and Regulation*, 3 edn, Sweet & Maxwell Ltd, Lambs Conduit St., London, UK.
- Snort Team (2004), 'The open source network intrusion detection system', Página Web. Última visita em julho de 2005.
- Squid (2005), 'Squid web proxy cache', Página Web. Última visita em julho de 2005.
- SquidGuard (2005), 'An ultrafast and free filter, redirector and access controller for Squid', Página Web. Última visita em julho de 2005.
- Storring, M., Andersen, H. & Granum, E. (1999), Skin colour detection under changing lighting condition, in H. Araujo & J. Dias, eds, '7th Symposium on Intelligent Robotics System', Coimbra, Portugal, pp. 187–195.
- Terrillon, J., Shirazi, M. N., Fukamachi, H. & Akamatsu, S. (2000), Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, in 'IEEE International Conference on Face and Gesture Recognition', pp. 54–61.
- Vaughan-Nichols, S. J. (2003), 'Saving private e-mail', *IEEE Spectrum* pp. 40–44.
- Wyszecki, G. (1986), *Handbook of perception and human performance*, Vol. 1: Sensory processes and perception, John Wiley & Sons, chapter 9: Color appearance.
- Yi, J., Park, J., Kim, J. & Choi, J. (2003), Robust skin color segmentation using a 2D plane of RGB color space, in 'Lecture Notes in Computer Science', Vol. 2869, ISICIS 2003, pp. 413–420.

Zarit, B. D., Super, B. J. & Quek, F. K. H. (1999), Comparison of five color models in skin pixel classification, *in* 'RATFG-RTS '99: Proceedings of the International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems', IEEE Computer Society, Washington, DC, USA, pp. 58–63.

Zheng, H., Daoudi, M. & Jedynek, B. (2004), 'Blocking adult images based on statistical skin detection', *Electronic Letters on Computer Vision and Image Analysis* **4**(2), 1–14.

Este documento foi preparado utilizando  $\text{\LaTeX}$ .  
As referências bibliográficas foram administradas com  $\text{\BibTeX}$ .  
O fonte do texto é Bookman e o das equações Euler, em 12 pontos.

