

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E  
INFORMÁTICA INDUSTRIAL

DIOGO ROSA KUIASKI

**SEGMENTAÇÃO DE PELE EM IMAGENS DIGITAIS PARA A  
DETECÇÃO AUTOMÁTICA DE CONTEÚDO OFENSIVO**

DISSERTAÇÃO DE MESTRADO

CURITIBA

2010

**DIOGO ROSA KUIASKI**

**SEGMENTAÇÃO DE PELE EM IMAGENS DIGITAIS PARA A  
DETECÇÃO AUTOMÁTICA DE CONTEÚDO OFENSIVO**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Mestre em Ciências” – Área de Concentração: Informática Industrial.

Orientador: Prof. Dr. Hugo Vieira Neto

**CURITIBA**

**2010**

Ficha catalográfica elaborada pela Biblioteca da UTFPR – Campus Curitiba

K96s Kuiaski, Diogo Rosa  
Segmentação de pele em imagens digitais para a detecção automática de conteúdo ofensivo / Diogo Rosa Kuiaski. – 2010.  
77 f. : il. ; 30 cm

Orientador: Hugo Vieira Neto  
Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial. Área de Concentração: Informática Industrial, Curitiba, 2010  
Bibliografia: f. 75-7

1. Processamento de imagens – Técnicas digitais. 2. Detecção de pele. 3. Banco de dados de imagens. 4. Pixel. 5. Reconhecimento de padrões. 6. Mineração de dados (Computação). I. Vieira Neto, Hugo, orient.  
II. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial. Área de Concentração em Informática Industrial. III. Título.

CDD 621.3



**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**  
Campus Curitiba



*Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial*

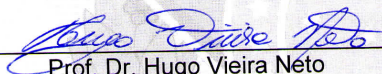
**Título da Dissertação Nº 517:**

**“Segmentação de Pele em Imagens Digitais para a  
Detecção Automática de Conteúdo Ofensivo”**

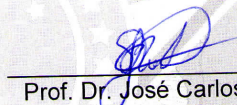
por

**Diogo Rosa Kuiaski**


Esta dissertação foi apresentada, às 09h30min do dia 25 de fevereiro de 2010, como requisito parcial para a obtenção do grau de MESTRE EM CIÊNCIAS – Área de Concentração: Informática Industrial, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial – CPGEI –, na Universidade Tecnológica Federal do Paraná – UTFPR. O trabalho foi aprovado pela Banca Examinadora, composta pelos professores:



Prof. Dr. Hugo Vieira Néto  
(Orientador - UTFPR)




Prof. Dr. José Carlos Ferreira da Rocha  
(UEPG)



Prof. Dr. Marcelo Victor Wüst Zibetti  
(UTFPR)

Visto da coordenação:



Prof. Dr. Humberto Remígio Gamba  
(Coordenador do CPGEI)

## **AGRADECIMENTOS**

Agradeço a meus familiares pelo apoio, aos colegas de laboratório pela ajuda e incentivo, e aos amigos pelo suporte nesta etapa de minha carreira.

Ainda, agradeço o apoio financeiro recebido da CAPES e UOL.

O projeto de pesquisa desta dissertação teve o apoio do programa UOL Bolsa Pesquisa - processos 20080129153400 e 20090213081100.

(<http://bolsapesquisa.uol.com.br>)

*Habe nun, ach! Philosophie,  
Juristerei und Medizin,  
Und leider auch Theologie  
Durchaus studiert, mit heißem Bemühn.  
Da steh ich nun, ich armer Tor!  
Und bin so klug als wie zuvor;*

Faust - Goethe.

Então, ah! Filosofia,  
Direito e Medicina,  
e, infelizmente, também a Teologia  
Estudando calorosamente.  
Estou eu aqui, pobre tolo!  
Tão inteligente quanto antes;

Faust - Goethe.

## RESUMO

KUIASKI, Diogo R.. Segmentação de Pele em Imagens Digitais para a Detecção Automática de Conteúdo Ofensivo. 79 f. Dissertação – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2010.

O presente trabalho tem como objetivo estudar meios de efetuar a detecção automática de conteúdo ofensivo (pornografia) em imagens digitais. Para tal estudou-se largamente segmentação de pixels de pele, espaços de cor e descritores de conteúdo. Esse trabalho tem um foco maior na segmentação de pele, pois é a etapa primordial nos trabalhos envolvendo detecção de conteúdo ofensivo. Testou-se quatro métodos de segmentação de pixels de pele e foi construído um banco de dados estruturado para o estudo de segmentação de pele, com meios de anotação de imagens para auxiliar na estruturação e no controle das características das imagens do banco. Com o auxílio das metainformações do banco de imagens, foram conduzidos estudos envolvendo as condições de iluminação e a segmentação de pele. Por fim, foi implementado um algoritmo de extração de características em sistemas de classificação pelo conteúdo de imagens (CBIR) para detecção de conteúdo ofensivo.

**Palavras-chave:** detecção de pele; segmentação de cor; anotação de imagens; banco de dados de imagens; CBIR.

## ABSTRACT

KUIASKI, Diogo R.. Skin Segmentation on Digital Images for Automatic Offensive-content Detection. 79 f. Dissertação – Programa de Pós-graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná. Curitiba, 2010.

This work presents a study of suitable approaches for automatic detection of offensive content (pornography) in digital images. Extensive experiments were conducted for skin pixel segmentation, colour spaces and content descriptors. This work focus its efforts on skin pixel segmentation, since this segmentation is the pre-processing stage for almost every content-based offensive image classification methods in the literature. Four skin skin segmentation methods were tested in six colour spaces. Also, a structured image database was built to help improve studies in skin segmentation, with the possibility of adding meta-information to the images in the database, such as illumination conditions and camera standards. With the help of meta information from the image database, experimets involving illumination conditions and skin colour segmentation were also done. Finally, some feature extraction algorithms were implemented in order to apply content-based image retrieval (CBIR) algorithms to classify offensive images.

**Keywords:** skin detection; color segmentation; image annotation; image databases; CBIR.



## LISTA DE FIGURAS

FIGURA 1 – ETAPAS DA DETECÇÃO DE CONTEÚDO. ....	20
FIGURA 2 – EXEMPLO DE SEGMENTAÇÃO DE PIXELS DE PELE. ....	24
FIGURA 3 – DIAGRAMA DOS VALORES DE TP, FN, TN, FP. ....	25
FIGURA 4 – CURVAS ROC PARA O CLASSIFICADOR JONES & REHG. ....	40
FIGURA 5 – CURVAS ROC PARA O CLASSIFICADOR <i>NAÏVE</i> BAYES. ....	41
FIGURA 6 – CURVAS ROC PARA O CLASSIFICADOR POR SIMPLES GAUSSIANA. ....	42
FIGURA 7 – CURVAS ROC PARA O CLASSIFICADOR MOG. ....	43
FIGURA 8 – IMAGENS DO BANCO SKINDB. ....	47
FIGURA 9 – DIAGRAMA EM ÁRVORE XML ....	48
FIGURA 10 – CURVAS ROC PARA O CLASSIFICADOR JONES & REHG. ....	51
FIGURA 11 – CURVAS ROC PARA O CLASSIFICADOR <i>NAÏVE</i> BAYES. ....	51
FIGURA 12 – CURVAS ROC PARA O CLASSIFICADOR MOG. ....	52
FIGURA 13 – CURVAS ROC PARA O CLASSIFICADOR POR SIMPLES GAUSSIANA. ....	52
FIGURA 14 – CURVAS ROC PARA O CLASSIFICADOR JONES & REHG UTILIZANDO IMAGENS EM AMBIENTES ABERTOS. ....	54
FIGURA 15 – CURVAS ROC PARA O CLASSIFICADOR JONES & REHG UTILIZANDO IMAGENS EM AMBIENTES FECHADOS. ....	55
FIGURA 16 – CURVAS ROC PARA O CLASSIFICADOR <i>NAÏVE</i> BAYES UTILIZANDO IMAGENS EM AMBIENTES ABERTOS. ....	56
FIGURA 17 – CURVAS ROC PARA O CLASSIFICADOR <i>NAÏVE</i> BAYES UTILIZANDO IMAGENS EM AMBIENTES FECHADOS. ....	56
FIGURA 18 – CURVAS ROC PARA O CLASSIFICADOR MOG UTILIZANDO IMA- GENS EM AMBIENTES ABERTOS. ....	57
FIGURA 19 – CURVAS ROC PARA O CLASSIFICADOR MOG UTILIZANDO IMA- GENS EM AMBIENTES FECHADOS. ....	58
FIGURA 20 – CURVAS ROC PARA O CLASSIFICADOR SIMPLES GAUSSIANA UTI- LIZANDO IMAGENS EM AMBIENTES ABERTOS. ....	59
FIGURA 21 – CURVAS ROC PARA O CLASSIFICADOR SIMPLES GAUSSIANA UTI- LIZANDO IMAGENS EM AMBIENTES FECHADOS. ....	59
FIGURA 22 – PROCEDIMENTO PARA DETECÇÃO DE CONTEÚDO OFENSIVO .	61
FIGURA 23 – DIVISÃO DA IMAGEM PARA O EHD ....	64
FIGURA 24 – MATRIZES DE SUPER-PIXELS ....	64
FIGURA 25 – DIVISÃO DA IMAGEM PARA O CD ....	65
FIGURA 26 – PROJEÇÕES DE BLOCOS DE PELE ....	66

## LISTA DE TABELAS

TABELA 1	–	RESULTADOS PARA O CLASSIFICADOR JONES & REHG.	.....	39
TABELA 2	–	RESULTADOS PARA O CLASSIFICADOR NAÏVE BAYES.	.....	41
TABELA 3	–	RESULTADOS PARA O CLASSIFICADOR POR SIMPLES GAUS- SIANA.	.....	42
TABELA 4	–	RESULTADOS PARA O CLASSIFICADOR MOG.	.....	43
TABELA 5	–	CURVAS ROC PARA OS CLASSIFICADORES EDSR.	.....	44
TABELA 6	–	RESULTADOS PARA O CLASSIFICADOR JONES & REHG.	.....	55
TABELA 7	–	RESULTADOS PARA O CLASSIFICADOR NAÏVE BAYES	.....	57
TABELA 8	–	RESULTADO PARA O CLASSIFICADOR MOG.	.....	58
TABELA 9	–	RESULTADOS PARA O C4.5	.....	69
TABELA 10	–	MATRIZ DE CONFUSÃO C4.5	.....	69
TABELA 11	–	RESULTADOS PARA O ADABOOST	.....	70
TABELA 12	–	MATRIZ DE CONFUSÃO ADABOOST	.....	70
TABELA 13	–	RESULTADOS PARA O BAGGING	.....	71
TABELA 14	–	MATRIZ DE CONFUSÃO BAGGING	.....	71

## LISTA DE SIGLAS

CBIR	<i>Content-based image retrieval</i> - Busca de imagens com base no conteúdo
PNG	<i>Portable network graphics</i> - Gráficos de rede portavel
KNN	<i>K-nearest neighbors</i> - K vizinhos mais próximos
VPP	Valor de predição positiva
GT	<i>Ground-Truth</i> - Padrão-ouro
ROC	<i>Receiver Operating Characteristics</i> - Características operacionais do receptor
MoG	<i>Mixture of Gaussians</i> - Mistura de Gaussianas
EM	<i>Expectation Maximization</i> - Maximização do valor esperado
EDSR	<i>Explicitly Defined Skin Regions</i> - Regiões de pele explicitamente definidas
XML	<i>eXtensible Markup Language</i> - Linguagem de marcação extensível
URL	<i>Uniform resource locator</i> - Localizador uniforme de recursos
ROI	<i>Region of interest</i> - Região de interesse
CSV	<i>Comma Separated Values</i> - Valores separados por vírgulas
SCD	<i>Scalable Colour Descriptor</i> - Descritor escalável de cor
EHD	<i>Edge Histogram Descriptor</i> - Descritor histograma de bordas
CD	<i>Compactness Descriptor</i> - Descritor de compactação
WEKA	<i>Waikato Environment for Knowledge Analysis</i> - Ambiente de Waikato para a análise de conhecimento
REPTree	<i>Regression Tree</i> - Árvore de regressão

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
1.1	DELIMITAÇÃO DO TEMA	16
1.2	JUSTIFICATIVA	17
1.3	ESTRUTURAÇÃO DO TRABALHO	17
<b>2</b>	<b>REVISÃO DOS MÉTODOS</b>	<b>19</b>
2.1	DETECÇÃO DE CONTEÚDO OFENSIVO	19
2.1.1	Vetores de Características e Principais Métodos da Literatura	20
2.1.2	Avaliação dos Resultados	21
2.2	SEGMENTAÇÃO DE PELE	22
2.2.1	Curvas ROC	23
<b>3</b>	<b>SEGMENTAÇÃO DE PELE</b>	<b>26</b>
3.1	ESPAÇOS DE COR	27
3.1.1	Espaços de Cor Aditivos (RGB e RG normalizado)	28
3.1.2	Espaços de Cor Perceptuais (HSI e HS)	29
3.1.3	Espaços de Cor Ortogonais (YCbCr e CbCr)	29
3.2	MODELOS PROBABILÍSTICOS	30
3.2.1	Gaussianas	31
3.2.2	Mistura de Gaussianas	31
3.2.3	Algoritmo EM	32
3.2.4	Histogramas de Cor	34
3.2.5	Histograma de Cor <i>naïve</i> Bayes	35
3.3	CLASSIFICADORES DE PIXELS DE PELE	35
3.3.1	Regra de Regiões	36
3.3.2	Regra de Jones & Rehg	37
3.3.3	Comparação Direta	38
3.4	RESULTADOS	38
3.4.1	Classificador Jones & Rehg	39
3.4.2	Classificador <i>naïve</i> Bayes	40
3.4.3	Classificador por Simplex Gaussiana	42
3.4.4	Classificador MoG	43
3.4.5	Classificadores EDSR	44
3.5	DISCUSSÕES	44
<b>4</b>	<b>EFEITOS DA ILUMINAÇÃO AMBIENTE</b>	<b>46</b>
4.1	BANCO DE IMAGENS ESTRUTURADO - SKINDB	46
4.1.1	Anotação das Imagens	47
4.2	ILUMINAÇÃO AMBIENTE	49
4.3	MÉTODOS UTILIZADOS	50
4.4	RESULTADOS	50
4.4.1	Classificador Jones & Rehg	53
4.4.2	Classificador <i>naïve</i> Bayes	55
4.4.3	Classificador MoG	57

4.4.4	Classificador por Simples Gaussiana .....	58
4.5	DISCUSSÕES .....	59
<b>5</b>	<b>DETECÇÃO DE CONTEÚDO OFENSIVO .....</b>	<b>61</b>
5.1	CONJUNTO DE CARACTERÍSTICAS .....	62
5.1.1	Remoção do Plano de Fundo .....	62
5.1.2	Descritor Escalável de Cor .....	63
5.1.3	Descritor Histograma de Bordas .....	63
5.1.4	Descritor de Compactação .....	65
5.1.5	Projeções de Pele .....	65
5.2	ALGORITMOS DE CLASSIFICAÇÃO .....	66
5.2.1	C4.5 .....	66
5.2.2	AdaBoost .....	67
5.2.3	Bagging .....	67
5.3	RESULTADOS E DISCUSSÕES .....	68
5.4	DISCUSSÕES .....	70
<b>6</b>	<b>CONCLUSÕES .....</b>	<b>73</b>
6.1	SEGMENTAÇÃO DE PELE .....	73
6.2	ILUMINAÇÃO AMBIENTE E BANCO DE DADOS CONTROLADO .....	73
6.3	DETECÇÃO DE CONTEÚDO OFENSIVO .....	74
6.4	TRABALHOS FUTUROS .....	75
6.4.1	Segmentação de Pele .....	75
6.4.2	Iluminação Ambiente e Banco de Dados Controlado .....	75
6.4.3	Detecção de Conteúdo Ofensivo .....	75
	<b>REFERÊNCIAS .....</b>	<b>77</b>

## 1 INTRODUÇÃO

As informações contidas na rede mundial de computadores são de fácil acesso, aparentemente sem restrições. A detecção de conteúdo ofensivo representa um papel importante para limitar o acesso a informações indesejáveis a usuários da rede. Entretanto, o fator motivador desta pesquisa é prover uma ferramenta aos mantenedores de sítios de uma rede. Essa ferramenta mostraria se em alguma máquina, sob a supervisão do mantenedor, está armazenado algum tipo de conteúdo ofensivo, o qual é especificamente a pornografia.

A detecção automática de sítios de conteúdo ofensivo é uma tarefa relativamente fácil quando se tem acesso a todas as informações do sítio. As informações contidas no sítio são na maioria textuais. Palavras-chave indicam se existe ou não conteúdo ofensivo em um sítio da rede. A proposta do presente trabalho é encontrar imagens de conteúdo ofensivo, sem utilizar outras informações além das contidas na própria imagem.

Trabalhar com os elementos de uma imagem e a partir deles conseguir definir o contexto de uma cena é uma tarefa desafiadora. Escolher um método, implementar e testar com um padrão-ouro é um trabalho que demanda muito tempo e poder computacional. Foram feitos alguns testes para alguns modelos, mas ainda há muito o que ser feito para a detecção coerente de conteúdo ofensivo em imagens digitais. Os desafios da pesquisa se encontram na diversidade de métodos encontradas na literatura e na falta de padrões para comparação do método apresentado na presente pesquisa com os propostos por Jones e Rehg (1999); Fleck, Forsyth e Bregler (1996); e Shih, Lee e Yang (2007); por exemplo.

A maior contribuição desse trabalho está na realização de uma pesquisa relativa à etapa segmentação de pixels de pele. O presente trabalho investigou questões ainda pouco abordadas pela literatura, relativas ao impacto que a variação da cor dos pixels de pele exerce no desempenho da segmentação. A cor de pele varia principalmente com o tipo de iluminação ambiente e isso reflete nas taxas de acerto utilizadas na segmentação dos pixels de pele. Para investigar essa influência, os métodos da literatura foram testados exaustivamente para vários modelos de espaços de cor e iluminações ambiente.

## 1.1 DELIMITAÇÃO DO TEMA

Delimitou-se o tema na investigação de metodologias de detecção de conteúdo ofensivo em imagens. As metodologias aqui estudadas são, em sua totalidade, algoritmos que utilizam informações contidas em imagens digitais, sem que sejam utilizadas análises textuais ou outras fontes de informação. Como resultado do estudo pretende-se testar e discutir o desempenho das metodologias citadas por autores da área.

O objetivo primordial nesta dissertação é estudar metodologias de detecção de conteúdo ofensivo, mas outros estudos se fazem necessários para realizar tal tarefa. Entre os passos intermediários da pesquisa estão estudar a segmentação de pele, espaços de cor e técnicas de *Content-based image retrieval* - Busca de imagens com base no conteúdo (CBIR).

Será dada uma atenção especial a métodos de segmentação automática de pele humana. O foco foi posto desta forma pois a segmentação de pele é utilizada na maioria dos algoritmos de detecção de conteúdo ofensivo (SOLAR et al., 2005; JONES; REHG, 1999; FLECK; FORSYTH; BREGLER, 1996). Durante a pesquisa sobre segmentação automática de pele foram realizados alguns estudos com o intuito de otimizar a segmentação na tentativa de melhorar taxas de acerto do classificador de conteúdo ofensivo. Esses estudos limitaram-se a espaços de cor, bancos de imagens e efeitos da iluminação ambiente.

Na segmentação de pixels são utilizados modelos estatísticos para representar matematicamente a cor de pele. Os modelos, paramétricos ou não-paramétricos, foram a base para a montagem de classificadores, sendo todos eles baseados em cor de pixel. Foram realizados testes para seis espaços de cor e quatro classificadores. Os resultados dos classificadores são comparados utilizando uma mesma base de dados a fim de se encontrar o classificador de melhor resultado. Após encontrado o melhor classificador, esse foi selecionado para a aplicação nos classificadores de conteúdo ofensivo.

A literatura descreve que a iluminação ambiente tem forte influência sobre a segmentação de pixels de pele (VEZHNEVETS V.; SAZONOV, 2003; JAYARAM et al., 2004; KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007). Para averiguar esse fato, foi proposto um estudo no qual somente imagens sob um tipo conhecido de iluminação são apresentadas ao classificador de pixels de pele. Os resultados desse estudo mostram que existe mesmo uma influência da iluminação ambiente na segmentação de pele. No estudo dos efeitos da iluminação ambiente foram testadas imagens de um banco de dados próprio para os mesmos classificadores de pixels de pele implementados nesse trabalho.

Para a detecção de conteúdo ofensivo foram testados alguns classificadores e vetores de

características. Os objetivos foram: descobrir quais características das imagens melhor representam conteúdo ofensivo e encontrar um classificador ótimo para a mesma tarefa.

## 1.2 JUSTIFICATIVA

A tarefa principal em reconhecimento de conteúdo através de características visuais está em se retirar informações relevantes e concisas dos elementos que constituem uma imagem.

Em uma cena, quando um observador tenta identificar seres humanos, existem características marcantes que o levam a reconhecer uma pessoa (VEZHNEVETS V.; SAZONOV, 2003). Essas características são reconhecidas através da união de um conhecimento prévio, ou conhecimento *top-down*, do observador, e de informações visuais trazidas pela imagem, ou conhecimento *bottom-up* (PALMER, 1999). Com o auxílio dessa estrutura, um observador humano discrimina regiões de interesse em imagens.

Entretanto, quando não se trata de um observador humano, mas de um computador, é preciso utilizar técnicas que permitam reproduzir esta estrutura de conhecimento. Características *top-down* ainda não são tão bem representadas por modelos matemáticos, deixando para as características *bottom-up* os resultados de uma boa ou má classificação. Esse fato foi acompanhado e gerou variações drásticas de resultados durante o decorrer do trabalho e será melhor discutido nos capítulos que seguem.

## 1.3 ESTRUTURAÇÃO DO TRABALHO

O trabalho está dividido em seis partes: a contextualização do tema, a revisão dos métodos, a segmentação de pixels de pele, a anotação de imagens, o desenvolvimento do banco de dados controlado e a classificação de conteúdo ofensivo. Nessa estrutura, ambos os conhecimentos, *top-down* e *bottom-up*, estão presentes tanto na segmentação de pixels de pele quanto na classificação de imagens ofensivas. As demais partes apresentam estudos feitos para melhorar os resultados das etapas de classificação e entender melhor o processo como um todo.

No capítulo 2 é feita a revisão dos métodos mais utilizados nas áreas de reconhecimento de conteúdo ofensivo, reconhecimento de padrões e mineração de dados. Isso fundamentou a adoção dos modelos e métodos utilizados no decorrer do trabalho. Ainda, o capítulo mostra como todas as etapas do trabalho estão conectadas, além de uma visão geral sobre os algoritmos e métodos dessa pesquisa.

Como a segmentação de pixels de pele representa muito na detecção de conteúdo ofen-



sivo, o capítulo 3 é exclusivamente dedicado à revisão, resultados e conclusões tiradas da segmentação automática de pixels de pele. Nesse capítulo são comparados métodos de segmentação de pixels de pele. A análise dos resultados mostra qual classificador obteve os melhores resultados.

Alguns experimentos feitos durante a etapa da segmentação de pixels de pele geraram resultados significativos para a pesquisa, sendo esses os estudos dos efeitos da iluminação ambiente. No capítulo 4 é explicado como a iluminação ambiente influencia na detecção de pixels de pele e, ainda, como foi construído o banco de dados de estudos de padrões em pele humana (SkinDB).

No capítulo 5 são descritos os métodos utilizados na detecção de conteúdo ofensivo, com os resultados e as considerações acerca do que esperar de uma boa ou má classificação. E, finalmente, no capítulo 6 são tiradas as conclusões sobre todas as etapas da presente pesquisa.

## 2 REVISÃO DOS MÉTODOS

A seguir serão mostrados alguns dos métodos mais utilizados para a detecção de conteúdo ofensivo, o que inclui etapas intermediárias como a segmentação de pele e problemas descritos pelos autores, que são abordados nos capítulos que seguem.

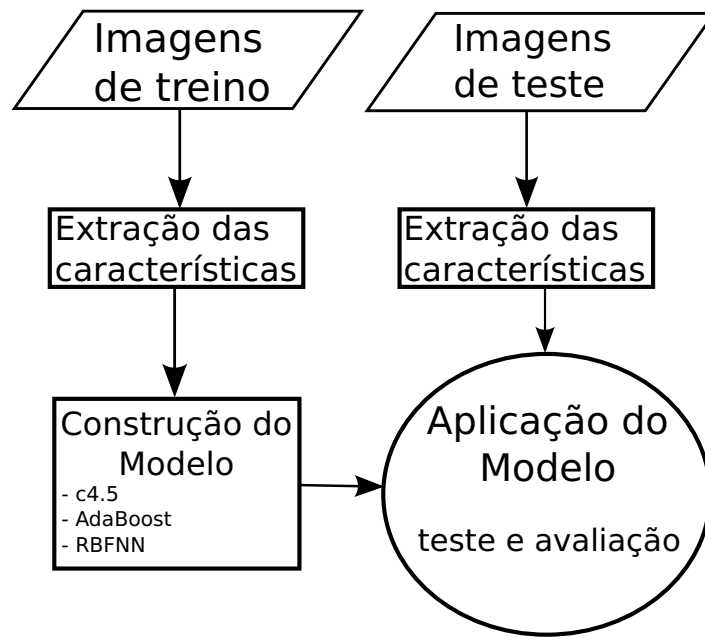
### 2.1 DETECÇÃO DE CONTEÚDO OFENSIVO

As metodologias de detecção de conteúdo ofensivo encontrada na literatura são relativamente semelhantes. O procedimento de classificação é simples e em sua maioria realizado da seguinte forma: Segmentação de pele, construção do vetor de características e classificação. Os classificadores mais comuns necessitam de um conhecimento prévio para serem construídos. A estrutura adotada pela maioria dos trabalhos na literatura é representada na figura 1, a qual mostra ao lado esquerdo a construção do modelo através de imagens com conteúdo previamente conhecido e ao lado direito a análise do modelo em imagens teste.

Tal estrutura pode ser vista nos trabalhos de Jones e Rehg (1999), Shih, Lee e Yang (2007), entre outros. Mesmo tendo a mesma estrutura, os métodos propostos pelos autores se diferenciam em como cada uma das etapas é feita.

Para todas as etapas deste trabalho, utilizou-se o banco de imagens conhecido como CompaqDB (JONES; REHG, 1999), criado por Jones e Rehg (1999). O banco é composto por aproximadamente 5000 imagens contendo pixels de pele, e outras 8000 com imagens sem a presença de pessoas. Das 5000 imagens em que aparecem pessoas na cena, aproximadamente um terço são imagens de conteúdo ofensivo.

Para facilitar o entendimento sobre quais imagens foram utilizadas para cada etapa do presente trabalho, foram criadas seleções do banco CompaqDB. A primeira seleção é o CompaqDBpng, na qual todas as imagens foram convertidas para o formato *Portable network graphics* - Gráficos de rede portavel (PNG) e separadas em dois grupos, imagens com pixels de pele e imagens sem pixels de pele. A partir da seleção CompaqDBpng foi criada a



**Figura 1: Modelo genérico para a classificação de imagens quanto à presença de conteúdo ofensivo.**

CompaqDBeasy500, um conjunto de 500 imagens aleatórias para cada um dos dois grupos do CompaqDBpng. A seleção CompaqDBContent foi criada separando o CompaqDBpng em dois grupos iguais de 1349 imagens: um com conteúdo ofensivo e outro de conteúdo não ofensivo. Na seleção que não apresenta conteúdo ofensivo foram igualmente selecionadas imagens de pessoas e outras imagens sem a presença de pessoas.

Com o objetivo de ter uma visão geral sobre os métodos, cada uma dessas etapas será explorada. Os métodos estudados e a discussão sobre a escolha para a implementação e simulações estão na subseção 2.1.1 e na seção 2.2.

### 2.1.1 Vetores de Características e Principais Métodos da Literatura

Os trabalhos de detecção de conteúdo ofensivo mais significativos têm como trabalho base os estudos feitos por Fleck, Forsyth e Bregler (1996), nos quais foram utilizadas características geométricas, como a presença de rostos e a posição relativa entre partes do corpo. A classificação foi feita pela presença ou ausência das características geométricas nas imagens.

Jones e Rehg (1999) utilizam características geométricas de tamanho relativo de áreas de pixels de pele, quantidade de pixels de pele conectados e número de regiões de pele conectadas. A classificação é feita através do algoritmo de construção de árvores de decisão C4.5 (QUINLAN, 1996).

Já nos trabalhos de Shih, Lee e Yang (2007), são utilizadas características de cor, geometria e textura para descrever o conteúdo. O método de classificação utilizado é o algoritmo de *K-nearest neighbors* - K vizinhos mais próximos (KNN). Nesse método, o vetor de característica é comparado com vetores de característica de classe conhecida e a classe da maioria absoluta dos *K* primeiros é a classe a qual pertence o vetor em análise. A métrica utilizada foi a diferença simples ou distância L1, aplicada entre o vetor de características da imagem que se deseja avaliar e 100 vetores conhecidos gerados pela extração de características de imagens previamente classificadas entre conteúdo ofensivo e conteúdo não-ofensivo.

Para os trabalhos de Solar et al. (2005) os descritores também são geométricos e aplicados sobre os maiores blocos de pele encontrados na imagem. São descritores de excentricidade, posição, alongação e área do blocos de pele. Uma novidade trazida pelos autores é fazer um estudo de caso numa região específica da internet: somente sítios chilenos.

A maioria dos métodos utiliza descritores geométricos. No presente trabalho, foi utilizado o método de extração de características proposto por Shih, Lee e Yang (2007) e aplicado a alguns modelos de CBIR. Foi proposta tal metodologia, pois as taxas de acerto dos trabalhos de Shih, Lee e Yang (2007) foram as melhores encontradas na literatura, além da descrição das características ser clara e coerente com o contexto das imagens de conteúdo ofensivo.

Dentre outros trabalhos da literatura pode-se citar os de Hammami, Chahir e Chen (2006), Hu et al. (2007) e Lee e Luh (2008), os quais utilizam informações além das encontradas nos pixels das imagens, como análise de textos, *tags* HTML entre outros. A maioria dos trabalhos encontrados na literatura apresenta uma taxa de acerto de aproximadamente 90% de classificação para imagens de conteúdo ofensivo.

Entretanto, não existe nesses trabalhos um banco de dados padrão que pudesse ser utilizado para comparar os resultados obtidos neste trabalho com os outros da literatura. Os bancos de imagens encontrados na literatura são em sua maioria imagens capturadas diretamente da internet sem passar pela aferição de observadores e não disponível para testes de outros pesquisadores. O método utilizado para a captura das imagens da internet utiliza um procedimento conhecido como *web crawling*.

### 2.1.2 Avaliação dos Resultados

A análise de resultados para a detecção de conteúdo ofensivo é feita através de valor de predição positiva (*precision*) e sensibilidade (*recall*) (THEODORIDIS; KOUTROUMBAS, 2006). O Valor de predição positiva (VPP) representa quantas imagens são realmente imagens

de conteúdo ofensivo dentre todas as imagens classificadas como tal, ou seja, a soma das corretamente e incorretamente classificadas. A sensibilidade representa a quantidade de imagens corretamente classificadas dentre todas as imagens que realmente são imagens de conteúdo ofensivo.

O melhor valor da classificação ocorre quando ambas as taxas são próximas de 1, pois o classificador acertou o maior número de imagens ofensivas e não errou com imagens não ofensivas. Também foi utilizada a matriz de confusão, a qual mostra quantas imagens foram classificadas em quais classes e, ainda, se a classificação foi correta ou não. Foram utilizadas duas classes para detecção de conteúdo ofensivo: imagens de conteúdo ofensivo (classe conteúdo ofensivo) e imagens sem conteúdo ofensivo (classe conteúdo não-ofensivo). Utilizando 2 classes a matriz de confusão apresenta 2 colunas por 2 linhas. A matriz de confusão pode ser vista na figura 3 aplicada à análise dos dados na seção 2.2.1

## 2.2 SEGMENTAÇÃO DE PELE

A cor de pele é uma característica marcante (VEZHNEVETS V.; SAZONOV, 2003) e por isso os métodos de segmentação de pele nesse trabalho realizam a classificação tendo como base somente a informação de cor. Além do fato de se utilizar a cor, os métodos para a segmentação de pele aqui descritos são feitos pixel a pixel, como sugerem a maioria dos trabalhos na área (JONES; REHG, 1999; PHUNG; CHAI; BOUZERDOUM, 2001; FLECK; FORSYTH; BREGLER, 1996). Outra opção de segmentação de pele é a utilização de informações de regiões de pixels (KRUPPA; BAUER; SCHIELE, 2002; JEDYNAK; ZHENG; DAOUDI, 2003; YANG; AHUJA, 1998), mas foi verificado junto à literatura que os métodos de classificação pixel a pixel são mais robustos e rápidos (VEZHNEVETS V.; SAZONOV, 2003; KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007), enquanto apresentam taxas de acerto e erro parecidas com os métodos por regiões.

Durante os estudos de segmentação de pele e para a construção dos modelos utilizou-se, também, as seleções CompaqDBpng e CompaqDBeasy500. Cada imagem do banco CompaqDB contém uma imagem binária correspondente que indica através da cor quais são pixels de pele (pixels brancos) e quais não são (pixels pretos). Essas imagens binárias são, também, conhecidas como máscaras binárias e o conjunto de máscaras constitui o *Ground-Truth* - Padrão-ouro (GT).

Utilizou-se o CompaqDB como banco de imagens, pois os principais trabalhos encontrados na literatura foram feitos sobre um banco de imagens próprio, o qual na maioria dos casos não

é aberto ou de difícil acesso para outros pesquisadores, como por exemplo os bancos ECU e M2VTS (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007), dois bancos de imagens utilizados para a detecção de faces sem o padrão-ouro para pele. O fato de conciliar os métodos de segmentação de pele com um único banco de imagens assegura que a comparação entre eles seja mais justa. Para resolver o problema de um banco padrão, foi iniciada a construção de um banco de imagens específico para estudos envolvendo a segmentação de pele. O banco construído neste trabalho é chamado de skinDB e as formas de aquisição das imagens são explicadas em maiores detalhes no capítulo 4.

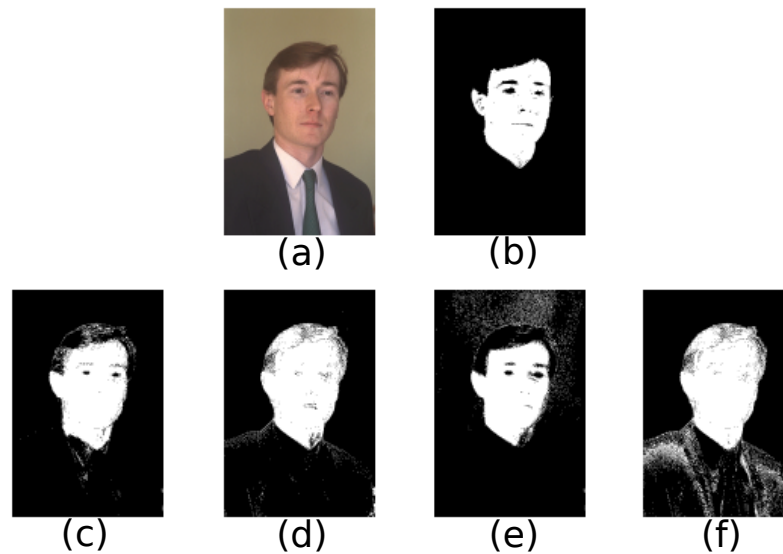
A metodologia utilizada para a validação dos métodos de segmentação de pele foi construir os modelos matemáticos de cor de pele utilizando a seleção CompaqDBeas500 e testar os modelos de segmentação de pele pixel a pixel encontrados na literatura no conjunto total de imagens, o banco CompaqDBpng. Os modelos foram aplicados a uma série de espaços de cor buscando qual conjunto espaço-modelo seria ótimo para a segmentação de pixels de pele.

Os resultados são analisados de duas formas: na forma de curvas de *Receiver Operating Characteristics* - Características operacionais do receptor (ROC) (FAWCETT, 2004) e na forma de mapas de pele. Mapas de pele são resultados primários dos experimentos de classificação de pixels. Tais mapas são imagens binárias de tamanho igual à imagem a ser segmentada, ou imagem de entrada, os quais contêm a predição sobre quais pixels foram ou não foram considerados pele. A predição positiva (pixel de pele) gera no mapa de pele um pixel de cor branca, caso contrário esse pixel será de cor preta. A figura 2 apresenta um exemplo de uma imagem da seleção CompaqDBeas500 juntamente com o padrão-ouro e alguns mapas de pele gerados através de diferentes modelos probabilísticos e espaços de cor.

### 2.2.1 Curvas ROC

Uma das formas de análise dos resultados da segmentação de pixels de pele é feita através de curvas ROC. Trata-se de uma análise comparativa e efetuada entre classificadores. Assim, é possível comparar o desempenho dos métodos utilizados. As curvas foram concebidas como ferramenta de decisão estatística (FAWCETT, 2004). Foram largamente aplicadas durante a 2ª Guerra Mundial para análise de imagens de radar e atualmente são muito utilizadas no auxílio a diagnóstico médico e são também aplicadas para a avaliação de classificadores binários (formados somente por duas classes) como é o caso neste trabalho.

Os eixos de uma curva ROC correspondem a taxas de acerto e erro. No eixo vertical estão as taxas de acerto ou de verdadeiros positivos (TPR - *True Positive Rate*) enquanto no eixo horizontal estão as taxas de erro, ou de falsos positivos (FPR - *False Positive Rate*). As taxas



**Figura 2: Exemplo de segmentação: (a) Imagem original; (b) Padrão-ouro; (c) Resultado do classificador Jones & Rehg e espaço de cor CbCr; (d) Resultado do classificador Jones & Rehg e espaço de cor RG normalizado; (e) Resultado do classificador naïve Bayes e espaço de color CbCr; (f) resultado do classificador naïve Bayes e espaço de cor RG normalizado.**

são adquiridas comparando cada pixel do padrão-ouro com o resultado do classificador. A comparação é feita utilizando o operando lógico “e” (**and**), portanto surgem quatro valores possíveis:

1. Verdadeiros-positivos (TP), quando GT é pele **and** resultado é pele;
2. Falsos-negativos (FN), quando GT é pele **and** resultado é não-pele;
3. Falsos-positivos (FP), quando GT é não-pele **and** resultado é pele;
4. Verdadeiros-negativos (TN), quando GT é pele-não **and** resultado é não-pele;

As taxas TPR e FPR são obtidas através do número de ocorrências de TP, FP, TN e FN para todos os dados analisados. As equações 1 e 2 mostram como são calculadas essas taxas.

$$TPR = \frac{nTP}{nTP + nFN} \quad (1)$$

$$FPR = \frac{nFP}{nTN + nFP} \quad (2)$$

onde  $nTP$  é o número de verdadeiros positivos,  $nFP$  o número de falsos positivos,  $nFN$  o número de falsos negativos e  $nTN$  o número de verdadeiros negativos.

A figura 3 mostra a tabela de contingência (Matriz de confusão).

		Padrão-ouro (GT)	
		p	n
Resultado da Classificação	p	Verdadeiro Positivo	Falso Positivo
	n	Falso Negativo	Verdadeiro Negativo

**Figura 3: Diagrama ilustrativo do que são os valores de TP, FN, FP, TN. Onde  $p$  são pixels de pele e  $n$  pixels de não pele.**

Para se levantar uma curva ROC são necessários vários pontos (FPR,TPR). Tais valores são conseguidos através da variação de parâmetros do classificador. Para a detecção de pele, o parâmetro utilizado para construção das curvas ROC é um valor de limiar, dado pela regras de decisão do que se considera pele e do que se rejeita como sendo pele. Isso é melhor explicado no capítulo 3. O valor de limiar que apresenta melhor resultado é o ponto (FPR,TPR) que contém menor valor de FPR e, ao mesmo tempo, o maior TPR. O valor resultante corresponde ao ponto conhecido como “joelho” da curva.



### 3 SEGMENTAÇÃO DE PELE

O trabalho de segmentação de pele consiste de estudar, implementar e testar as metodologias mais utilizadas da literatura, com o objetivo de escolher, dentre os métodos, qual apresenta o melhor desempenho. Tal método de segmentação é aplicado na detecção de conteúdo ofensivo.

A segmentação foi feita utilizando informações de cor de cada pixel, que é um ponto em um espaço de cor  $n$ -dimensional, definido por um vetor  $c = [c_1, c_2, \dots, c_n]$ . As componentes do vetor  $c$  representam as informações de luminância e de cromaticidade do pixel. Espaços de cor são geralmente tridimensionais,  $c = [c_1, c_2, c_3]$ , como o RGB, HSI, YCbCr, YIQ, SCT, cieLUV e cieLAB, entre outros. Entretanto, para este trabalho foram utilizados somente espaços de cor com 2 e 3 dimensões. No vetor estão representadas as informações de luminância e cromaticidade, que são propriedades da luz incidente no sensor de uma câmera. Normalmente, a luminância é representada por uma das três dimensões, deixando as outras duas para a informação de cromaticidade. Nesta dissertação foram conduzidos experimentos de redução da dimensionalidade dos espaços de cor, ou seja, três dos seis espaços de cor investigados são bidimensionais (somente cromaticidade) e os outros três são tridimensionais (cromaticidade e luminância).

A representação do conjunto de cores de pixel que compõe a pele humana é feita através de modelos matemáticos. Para montar um modelo de cor da pele humana (STÖRRING, 2004) existem dois possíveis caminhos: (a) Fixar uma região no espaço de cor e assumir que todos os pixels contidos dentro desse subespaço são pixels de pele; (b) Estabelecer modelos probabilísticos, nos quais probabilidades de cor da pele indicam a chance do pixel ser de pele.

A classificação de pixels de pele é feita aplicando ao modelo matemático uma regra de decisão. Uma regra de decisão é feita dividindo o espaço de cores em regiões que são consideradas pele e regiões que são consideradas não-pele. Por exemplo, quando é definida a região no espaço de cor essa regra é: está dentro da região é pele senão não é. Para os modelos probabilísticos, os quais somente mostram as probabilidades de um determinado pixel ser um pixel

de pele, é necessário estabelecer outras regras de decisão.

No presente capítulo serão discutidas as metodologias implementadas para a segmentação de pixels de pele e os resultados atingidos. Na seção 3.1 são mostrados os espaços de cor utilizados ao decorrer do estudo. Os modelos matemáticos para a cor de pele estão detalhados na seção 3.2, enquanto na seção 3.3 são mostrados alguns classificadores baseados nos modelos de cor de pele. Por fim, resultados e discussões são mostrados nas seções 3.4 e 3.5.

### 3.1 ESPAÇOS DE COR

A maioria dos métodos de segmentação de pele, e por consequência os de detecção de conteúdo ofensivo, trabalham sobre um espaço de cor específico. Espaços de cor são formas de se representar características físicas da luz. A luz refletida em objetos atinge as células que compõem a retina do ser humano, onde se encontram, dentre outras, dois tipos de células, os cones e os bastonetes. Existem três tipos diferentes de cones e cada um deles responde ao estímulo de uma determinada faixa de comprimento de onda da luz incidente. A composição do estímulo desses três tipos de cores propicia a visão em cores (PALMER, 1999).

Câmeras digitais utilizam um princípio parecido com o da retina para a aquisição e codificação das propriedades da luz em arquivos de imagem. Existem duas das propriedades da luz que interessam na presente pesquisa: luminância e cromaticidade. A primeira, luminância, é o número de fótons incidentes no sensor da câmera, ou seja, a energia da fonte de luz. A cromaticidade é composição espectral das ondas luminosas, ou seja, os comprimentos de onda que formam a luz (STÖRRING, 2004).

As imagens utilizadas são digitais, logo, quantizadas e amostradas em espaços discretos (pixels). Tal quantização, aliada aos avanços da computação gráfica, compressão de dados, transmissão de sinal de vídeo, fez surgir inúmeras formas de representar as propriedades da cor, formas essas chamadas de espaços de cor (POYNTON, 1995), cada qual com certas propriedades e peculiaridades. Dentre todos os espaços utilizados na literatura para a segmentação de pele foram estudados seis espaços de cor: RGB, RG normalizado, HSI, HS, YCbCr e CbCr. A escolha dos espaços de cor se deve principalmente ao fato de que os espaços são amplamente utilizados na literatura e é possível trabalhar com 3 e 2 dimensões por um mesmo espaço de cor. Poucos trabalhos fogem a um dos espaços de cor citados e utilizam os espaços cieLUV, cieLAB e YQQ.

### 3.1.1 Espaços de Cor Aditivos (RGB e RG normalizado)

O espaço de cor RGB tem como base os três tipos de foto-receptores (cones) que estão presentes na retina de um ser humano normal. A representação de cor desse modelo consiste na combinação aditiva de três cores primárias: o vermelho (*red-R*), o verde (*green-G*) e o azul (*blue-B*). Originalmente tal modelo foi concebido para monitores com tubo de raios catódicos, mas ainda é muito utilizado em algoritmos de processamento de imagens principalmente por razões históricas relacionadas à codificação de arquivos de imagem (POYNTON, 1995).

Entretanto, a escolha do espaço RGB para a classificação de pixels de pele não parece ser a melhor escolha, pois outros espaços de cor apresentam melhores valores de TPR (SHIN; CHANG; TSAP, 2002). Esse espaço de cor é conhecido por apresentar uma alta correlação entre os três canais e não existe uma separação clara entre as informações de luminância e crominância, ou seja, a luminância encontra-se “diluída” entre os três canais de cor (VEZHNEVETS V.; SAZONOV, 2003). O modelo RGB também é notoriamente vulnerável a variações de iluminação ambiente (KUIASKI et al., 2009; STÖRRING, 2004). Mas existem trabalhos, como o de Albiol, Torres e Delp (2001), que mostraram resultados nos quais o espaço RGB apresenta os mesmos valores de classificação de outros espaços de cor, como o HSI e o YCbCr.

Em uma tentativa de reduzir os efeitos da iluminação ambiente, removeu-se a informação de luminância (JONES; REHG, 1999). Esse experimento foi feito para todos os espaços de cor contidos nesse trabalho. Particularmente para o RGB, que não apresenta claramente a separação de luminância e crominância, é feita uma transformação de espaço mostrada nas equações (3) e (4) em uma tentativa de desacoplar a luminância da crominância (CAETANO; BARONE, 2001).

$$r = \frac{R}{R + G + B}, \quad (3)$$

$$g = \frac{G}{R + G + B}, \quad (4)$$

onde,  $R$ ,  $G$  e  $B$  são os valores dos canais vermelho, verde e azul, respectivamente, e  $r$  e  $g$  são valores conhecidos como  $R$  e  $G$  normalizados.

A componente azul normalizado ( $b$ ) não precisa ser calculada, pois a informação de cor já se encontra nas outras duas componentes  $r$  e  $g$ . Haja visto que a componente  $b$  é uma combinação linear de  $r$  e  $g$ , ou seja,  $r + g + b = 1$ .

### 3.1.2 Espaços de Cor Perceptuais (HSI e HS)

O espaço de cor HSI (Matiz -  $H$ , Saturação -  $S$  e Intensidade -  $I$ ) tenta reproduzir as cores numericamente como um ser humano o faria perceptualmente. Matiz representa a cor primária de um objeto (exemplo: verde, vermelho, azul, laranja), saturação representa quão pura é a cor (vermelho é a cor pura, rosa é vermelho com adição de uma certa quantidade de branco) e intensidade representa a informação de luminância. As componentes  $H$  e  $S$  representam juntas a informação de crominância. Para este estudo, como foi feito para o modelo RGB é possível retirar a componente  $I$  em uma tentativa de reduzir efeitos da iluminação ambiente.

Todas as imagens utilizadas são codificadas em RGB e por tal motivo é necessário realizar a transformação para o espaço HSI. O transformação é realizada de acordo com as equações (5), (6) e (7).

$$I = \frac{R + G + B}{3}, \quad (5)$$

$$S = 1 - \frac{\min(R, G, B)}{I}, \quad (6)$$

$$H = \arctan \left[ \frac{\sqrt{3}(G - B)}{2R - G - B} \right], \quad (7)$$

onde,  $R$ ,  $G$  e  $B$  são os canais de cor do espaço RGB normalizados no intervalo  $[0, 1]$ , e  $H$ ,  $S$  e  $I$  são as componentes do espaço HSI respectivamente.

### 3.1.3 Espaços de Cor Ortogonais (YCbCr e CbCr)

O espaço de cor YCbCr é largamente utilizado nos sistemas de televisão europeus, e consiste em uma transformação linear do espaço de cor RGB (VEZHNEVETS V.; SAZONOV, 2003).  $Y$  representa *luma* ou luminância,  $Cb$  representa o chroma azul e  $Cr$  o chroma vermelho. Por causa da decorrelação entre luminância e crominância, esse espaço de cor é chamado, também, de espaço de cor ortogonal (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007). Por tal motivo é esperado que a remoção da componente de luminância  $Y$  resulte em bons resultados de segmentação de pixels de pele.

A transformação de RGB para YCbCr é feita segundo as equações (8), (9) e (10).

$$Y = 0.299R + 0.587G + 0.114B, \quad (8)$$

$$Cb = -0.168736R - 0.331264G + 0.5B, \quad (9)$$

$$Cr = 0.5R - 0.418688G - 0.081312B, \quad (10)$$

onde  $K$  é o valor do limiar que gera o valor de FPR desejado. onde,  $R$ ,  $G$  e  $B$  são as componentes do espaço RGB, e  $Y$ ,  $Cb$  e  $Cr$  são as componentes do espaço YCbCr respectivamente.

### 3.2 MODELOS PROBABILÍSTICOS

Modelos probabilísticos buscam, utilizando a teoria da probabilidade, quais das cores de um espaço têm mais chance de representar um determinado objeto. Aqui, são apresentados métodos probabilísticos de modelamento de cor da pele. Esses modelos podem ser paramétricos, os quais apresentam funções densidade de probabilidade analíticas que usam parâmetros (média e desvio padrão, por exemplo) ou não-paramétricos, nos quais as funções de probabilidade são construídas por meio de inúmeras observações de pixels contendo cor de pele. Os métodos utilizados neste trabalho apenas considerada a informação da cor de um pixel sem utilizar informações de pixels vizinhos.

Foram implementados quatro diferentes modelos probabilísticos: Simples Gaussiana, *Mixture of Gaussians* - Mistura de Gaussianas (MoG), Histogramas de Cor e Histogramas *naïve* Bayes. Destes, os dois primeiros são métodos paramétricos e os outros dois são não-paramétricos. A informação existente nos modelos é a função densidade de probabilidade de um pixel de cor  $c$  dado que este pertence à classe pele ou à classe não-pele, ou seja,  $p(c|pele)$  e  $p(c|não-pele)$ , respectivamente.

Modelos não-paramétricos parecem representar melhor a cor de pele. Entretanto tais métodos dependem muito do banco de dados que gerou a função densidade de probabilidade. Uma saída para generalizar a função densidade de probabilidade é a utilização de modelos paramétricos conhecidos. Quer para o presente trabalho, quer para os de Caetano e Barone (2001) o modelo paramétrico considerado é uma distribuição Gaussiana multidimensional. Outro modelo paramétrico mais robusto que busca uma distribuição genérica desconhecida é o método da Mistura de Gaussianas, que tem sido bem aceito na literatura (CAETANO; BARONE, 2001; TERRILLON et al., 2000; JONES; REHG, 1999).

### 3.2.1 Gaussianas

A função densidade de probabilidade da cor de pele pode ser modelada como sendo uma Gaussiana multidimensional, a qual segue a equação (11).

$$p(c|pele) = \frac{1}{2\pi^{d/2} \cdot |\Sigma|^{1/2}} e^{-\frac{1}{2}(c-\mu)^T \Sigma^{-1} (c-\mu)} \quad (11)$$

onde  $d$  é o número de dimensões do vetor  $c$  e os parâmetros  $\mu$  e  $\Sigma$  representam a média e matriz de covariância, respectivamente. Para estimá-los foram utilizados todos os pixels de pele da seleção CompaqDBeas500, aplicando as equações (12) e (13).

$$\mu = \frac{1}{N} \cdot \sum_{x=1}^N c_x \quad (12)$$

$$\Sigma = \frac{1}{N} \cdot \sum_{x=1}^N (c_x - \mu) \cdot (c_x - \mu)^t \quad (13)$$

onde  $N$  é o número de pixels de pele do banco e  $c_x$  é o  $x$ -ésimo pixel do banco.

### 3.2.2 Mistura de Gaussianas

Uma forma mais sofisticada de se modelar cor de pele é a utilização da soma ponderada de distribuições Gaussianas. O modelo conhecido como mistura de Gaussianas, ou MoG (*Mixture of Gaussians*) tem como objetivo compor a função densidade de probabilidade de pele através dessa soma ponderada. A equação (14) mostra a soma ponderada por pesos  $\omega_i$  de  $k$  densidades de probabilidades que foram geradas pelas Gaussianas representadas na equação (15).

$$p(c|pele) = \sum_{i=1}^k \omega_i \cdot p_i(c|pele) \quad (14)$$

$$p_i(c|pele) = \frac{1}{2\pi^{d/2} \cdot |\Sigma_i|^{1/2}} e^{-\frac{1}{2}(c-\mu_i)^T \Sigma_i^{-1} (c-\mu_i)} \quad (15)$$

onde  $\omega_i$ ,  $\mu_i$  e  $\Sigma_i$  são, respectivamente, o peso, a média e a matriz de covariância para cada Gaussiana  $i$  de um total de  $k$  que compõem a função densidade de probabilidade total.

Supondo que os dados das observações são dados incompletos de probabilidade, é necessário estimar os valores de  $\omega_i$ ,  $\mu_i$  e  $\Sigma_i$ . Isso é feito através do algoritmo iterativo de *Expectation Maximization* - Maximização do valor esperado (EM) (THEODORIDIS; KOUTROUMBAS, 2006).

### 3.2.3 Algoritmo EM

O algoritmo EM é um método que aproxima uma distribuição qualquer a outra paramétrica através de sucessivas iterações. Seja uma variável  $y \in Y$  onde  $Y$  é um subconjunto multi-dimensional dos números reais. Essa variável representa um conjunto de dados com todas as observações e uma função densidade de probabilidade  $p_y(y; \theta)$ , onde  $\theta$  é um vetor de parâmetros desconhecidos. Entretanto, quando  $y$  não pode ser observado completamente, devido a necessidade de infinitas observações necessárias para determinar a distribuição, é possível observar outra variável que é função de  $y$ , como por exemplo  $x = g(y)$ . Pode-se então aplicar o algoritmo EM para se encontrar uma aproximação dos parâmetros de  $p_y(y; \theta)$ .

Os cálculos dos parâmetros têm como ponto de partida a função densidade de probabilidade  $p_x(x; \theta)$ . Sendo o conjunto  $Y(x)$  um subconjunto de  $Y$ , utiliza-se esse método para estimar  $p_y(y; \theta)$  a partir do conjunto de dados observáveis  $x$ . O algoritmo é iniciado com a definição de uma função densidade da variável incompleta  $x$ ,  $p_x(x; \theta)$ , em função da desejada  $y$ . A definição de  $p_x(x; \theta)$  é dada pela equação (16).

$$p_x(x; \theta) = \int_{Y(x)} p_y(y; \theta) dy \quad (16)$$

onde  $\theta$  é o vetor de parâmetros da função densidade.

Para estimar os parâmetros  $\theta$  para  $y$  utiliza-se o algoritmo de máxima verossimilhança (*maximum likelihood*). A equação (17) mostra que os parâmetros  $\theta_{ML}$  são os parâmetros  $\theta$  quando se dá a resolução da igualdade. Entretanto, quando faltam os dados  $y$ , os parâmetros são estimados a partir dos dados incompletos  $x$  utilizando o algoritmo EM. O EM maximiza iterativamente a expectativa da função de verossimilhança logarítmica (*log-likelihood*), representada na equação (17) como sendo a somatória das derivadas parciais do logaritmo de  $p_y$  em relação a  $\theta$ .

$$\theta_{ML} : \sum_{j=1}^n \frac{\partial \ln(p_y(y_j; \theta))}{\partial \theta} = 0 \quad (17)$$

onde  $\theta_{ML}$  é o vetor de parâmetros de máxima verossimilhança estimado.

O EM apresenta dois passos: o de estimação (passo-E); e o de maximização (passo-M).

**PASSO-E:** Para o cálculo da  $(t + 1)$ -ésima iteração, dado que  $\theta(t)$  é conhecido, é primeiramente calculado o valor esperado da função de verossimilhança logarítmica, como mostra a equação (18).

$$Q(\theta; \theta(t)) \equiv E \left\{ \sum_{j=1}^n \ln(p_y(y_j; \theta | X; \theta(t))) \right\} \quad (18)$$

onde  $Q(\theta; \theta(t))$  é o valor esperado da função de máxima verossimilhança para a iteração  $(t)$ .

**PASSO-M:** Maximiza-se o valor esperado de  $Q(\theta; \theta(t))$ . Os valores máximos serão os próximos parâmetros,  $\theta(t+1)$ . A equação (19) mostra que a derivada parcial da função de máxima verossimilhança deve ser zero, denotando o valor de máximo da função.

$$\theta(t+1) : \frac{\partial Q(\theta; \theta(t))}{\partial \theta} = 0 \quad (19)$$

A aplicação do algoritmo necessita um estado inicial para os parâmetros  $\theta$ ,  $\theta(0)$ , e uma condição de convergência,  $\|\theta(t+1) - \theta(t)\| < \varepsilon$ . No caso em estudo, o vetor  $\theta$  é composto pelos valores de  $\omega$ ,  $\mu$  e  $\Sigma$ . Especificou-se como valor de convergência  $\varepsilon = 10^{-4}$  e para os valores iniciais de  $\omega$ ,  $\mu$  e  $\Sigma$ , dividiu-se o banco de imagens igualmente em  $k$  conjuntos e foram calculados os parâmetros segundo as equações (20), (21) e (22).

$$\omega_i^0 = \frac{1}{k} \quad (20)$$

$$\mu_i^0 = \frac{1}{n_{ab}} \sum_a^b c \quad (21)$$

$$\Sigma_i^0 = \frac{1}{n_{ab}} \sum_a^b (c - \mu_i^0) \cdot (c - \mu_i^0)^T \quad (22)$$

onde o intervalo  $[a, b]$ , dependente de  $i$ , representa o  $k$ -ésimo conjunto de imagens dentro do total de imagens do banco. E  $n_{ab}$  é o número pixels nas imagens entre  $a$  e  $b$ .

A resolução algébrica do algoritmo EM aplicado para o caso de mistura de distribuições Gaussianas resulta nas equações (23), (24), (25) e (26) (THEODORIDIS; KOUTROUMBAS, 2006).

$$\omega_i^{t+1} = \frac{1}{n} \sum_{j=1}^n P^t(i|x_j) \quad (23)$$

$$\mu_i^{t+1} = \frac{\sum_{j=1}^n P^t(i|x_j) \cdot x_j}{\sum_{j=1}^n P^t(i|x_j)} \quad (24)$$



$$\Sigma_i^{t+1} = \frac{\sum_{j=1}^n P^t(i|x_j) \cdot \|x_j - \mu^{t+1}\|^2}{\sum_{j=1}^n P^t(i|x_j)} \quad (25)$$

$$P^t(i|x_j) = \frac{p^t(x_j|i) \cdot \omega_i^t}{p^t(x_j)} \quad (26)$$

onde  $p^t(x_j|i)$  e  $p^t(x_j)$  são os resultados das equações (14) e (15), respectivamente,  $n$  é o número total de pixels de pele,  $t$  é a iteração atual e  $t + 1$  a próxima iteração.

### 3.2.4 Histogramas de Cor

O modelo histograma de cor tenta fazer uma aproximação da função densidade de probabilidade  $p(c|pele)$ . Como se desconhece a função densidade de probabilidade da cor de pele humana, são feitas inúmeras observações para cada possível valor de cor de pele. Esse processo é feito para cada pixel de pele do banco CompaqDB e dividindo-se o resultado pelo número total de pixels, como mostra a equação (27), é obtida uma aproximação da função densidade de probabilidade  $p(c|pele)$ .

$$p(c|pele) = \frac{1}{N_{pele}} \cdot \sum_{j=1}^N c_{pele} \quad (27)$$

onde  $c_{pele}$  é um pixel de cor de pele e  $N_{pele}$  é o número de pixels de pele do banco utilizado para construir o modelo.

As vantagens do modelo histograma de cor estão na facilidade de implementação e em conseguir representar melhor a distribuição densidade de probabilidades que os modelos paramétricos estudados. Entretanto, desvantagens incluem a dependência do banco de imagens utilizado (falta de um modelo genérico) e a quantidade de memória utilizada para o armazenamento de dados de probabilidade esparsos.

Para os estudos realizados, foram construídos histogramas de cor quantizado em 256 posições para todos os espaços de cor citados acima. Uma prática encontrada na literatura para auxiliar na quantidade de memória utilizada é quantizar a as componentes de cor e criar histogramas de 128, 64, ou 32 posições sem alterar as taxas de acerto ou erro. Uma comparação entre a quantização e espaços de cor foi feita por Shin, Chang e Tsap (2002), cujos estudos mostraram que as taxas TPR e FPR variam para cada espaço de cor e para cada valor de quantização, de forma que nesta pesquisa utilizou-se somente histogramas de 256 posições.

Uma outra abordagem feita em histogramas de cor é supor que a função densidade de

probabilidade  $p(c|pele)$  depende somente das componentes de cor  $c_i$  e não de cada possível valor de  $c$ . Tal suposição foi testada para auxiliar no armazenamento dos dados esparsos e verificar a independência entre as componentes de cor. O modelo gerado por essa suposição é chamado no presente trabalho de Histograma de cor *naïve* Bayes.

### 3.2.5 Histograma de Cor *naïve* Bayes

O método consiste em tratar cada dimensão do vetor de características  $c$  como sendo uma variável independente, o que possibilita calcular histogramas separados para cada componente de cor  $c_i$ . Os histogramas obtidos separadamente representam a função densidade de probabilidade de cada componente de cor dado que aquela componente é uma componente de pele,  $p(c_i|pele)$ . Para calcular a probabilidade de pele  $p(c|pele)$  é utilizada a equação (28).

$$p(c|pele) = \prod_{i=1}^d p(c_i|pele), \quad (28)$$

onde  $i$  é a componente de cor do vetor  $c$  e  $d$  é o número de dimensões de  $c$ .

A utilização do modelo histograma de cor *naïve* Bayes implica uma redução drástica de utilização de memória. Ao invés de utilizar histogramas com  $256^3$  ou  $256^2$  posições são utilizados histogramas de  $3 \times 256$  ou  $2 \times 256$ .

## 3.3 CLASSIFICADORES DE PIXELS DE PELE

Associando a um modelo de cor de pele uma regra de classificação obtém-se um classificador de pixels de pele. Durante o presente trabalho foram implementados cinco classificadores que atuam pixel-a-pixel. Cada pixel é classificado em uma de duas categorias: pele ou não-pele. A decisão é feita através de uma regra como exemplifica a equação (29).

$$classe(c) = \begin{cases} \text{pele, se satisfaz a regra de decisão} \\ \text{não-pele, caso contrário} \end{cases} \quad (29)$$

onde  $c$  é a cor de um pixel de entrada.

As subseções 3.3.1, 3.3.2 e 3.3.3 mostram as regras de decisão aplicadas no presente trabalho. No total, unindo os modelos matemáticos e regras de decisão, foram implementados cinco classificadores de pixels de pele. Os classificadores gerados foram: o classificador Jones & Rehg (Histograma de cor / regra de Jones & Rehg), o classificador *naïve* Bayes (Histograma de cor *naïve* Bayes / regra de Jones & Rehg), o classificador por mistura de Gaussianas (Mis-

tura de Gaussianas / comparação direta), o classificador por simples Gaussianas (Gaussianas / comparação direta) e o classificador *Explicitly Defined Skin Regions* - Regiões de pele explicitamente definidas (EDSR).

### 3.3.1 Regra de Regiões

Esse método consiste em fixar uma região dentro de um espaço de cor. Isso normalmente é feito através de regras lógicas comparando os valores do pixel estudado com valores pré-definidos. Utilizou-se duas metodologias utilizando essa regra de decisão para os espaços de cor RGB e YCbCr, propostas por Kovac, Peer e Solina (2003) e Garcia e Tziritas (1999), respectivamente. Para o classificador de Kovac, o qual utiliza o espaço de cor RGB, as regras que definem se um pixel é de pele são mostradas na equação (30).

$$classe(c) = \begin{cases} \text{pele, caso} & \left\{ \begin{array}{l} R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and} \\ \max(R, G, B) - \min(R, G, B) > 15 \text{ and} \\ |R - G| > 15 \text{ and } R > G \text{ and } R > B \end{array} \right. \\ \text{não-pele, caso contrário} \end{cases} \quad (30)$$

onde  $R$ ,  $G$  e  $B$  são os valores das componentes de cor do espaço RGB e **and** é o operador lógico “e”.

Para o classificador de Garcia, o qual utiliza o espaço YCbCr, as regras podem ser vistas na equação (31).

$$classe(c) = \begin{cases} \text{pele, caso} & \left\{ \begin{array}{l} Cr \geq \max \begin{pmatrix} -2(Cb + 24); \\ -(Cb + 17); \\ -4(Cb + 32); \\ 2,5(Cb + \theta_1); \\ \theta_3; \\ 0,5(\theta_4 - Cb) \end{pmatrix} \\ \text{and} \\ Cr \leq \min \begin{pmatrix} (220 - Cb)/6; \\ 4(\theta_2 - Cb)/3 \end{pmatrix} \end{array} \right. \\ \text{não-pele, caso contrário} \end{cases} \quad (31)$$

onde  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$  e  $\theta_4$  estão representados na equação (32).

$$\begin{aligned}
\theta_1 &= \begin{cases} -2 + (256 - Y)/16 & , \text{ se } Y > 128 \\ 6 & , \text{ caso contrário} \end{cases} \\
\theta_2 &= \begin{cases} 20 - (256 - Y)/16 & , \text{ se } Y > 128 \\ 12 & , \text{ caso contrário} \end{cases} \\
\theta_3 &= \begin{cases} 6 & , \text{ se } Y > 128 \\ 2 + Y/32 & , \text{ caso contrário} \end{cases} \\
&\text{e} \\
\theta_4 &= \begin{cases} -8 & , \text{ se } Y > 128 \\ -16 + Y/16 & , \text{ caso contrário} \end{cases}
\end{aligned} \tag{32}$$

As vantagens desse método são não demandar tempo para a criação do modelo e redução de recursos computacionais. Entretanto, as regras não são flexíveis, ou seja, não apresentam parâmetros que possam ser alterados. Essa ausência de parâmetros faz com que os classificadores gerem somente um ponto na análise ROC, de maneira que a comparação com os outros métodos é feita através desse único ponto com o melhor resultado dos outros métodos citados a seguir.

### 3.3.2 Regra de Jones & Rehg

Essa regra de decisão foi proposta por Jones e Rehg (1999) e é feita aplicando-se um valor de limiar sobre  $p(\text{pele}|c)$ , como mostra a equação (33). A transformação do valor presente no modelo  $p(c|\text{pele})$  para o requerido pela regra de decisão é feita pela equação (34).

$$\text{classe}(c) = \begin{cases} \text{pele, se } p(\text{pele}|c) > \Theta \\ \text{não-pele, caso contrário} \end{cases} \tag{33}$$

$$p(\text{pele}|c) = \frac{p(c|\text{pele})p(\text{pele})}{p(c)} \tag{34}$$

A regra de decisão aplicada sobre  $p(\text{pele}|c)$  é invariante às probabilidades *a priori*  $p(\text{pele})$  e  $p(\text{não-pele})$ , as quais afetam somente a escolha do valor de limiar  $\Theta$ . Então, a equação (34) pode ser reescrita como uma relação entre  $p(c|\text{pele})$  e  $p(c|\text{não-pele})$  como mostra a equação (35). A transformação para converter o valor de  $\Theta$  utiliza a equação (36).

$$\frac{p(c|\text{pele})}{p(c|\text{não-pele})} < \Theta \tag{35}$$

$$\Theta = K \frac{1 - p(\text{pele})}{p(\text{pele})}, \quad (36)$$

onde  $K$  é um parâmetro no intervalo  $[0, 1]$  que é utilizado para normalizar  $\Theta$ .

Aplicando-se a regra para cada um dos modelos probabilísticos descritos na seção 3.2, tem-se dois classificadores: classificador Jones & Rehg (histograma completo  $n$ -dimensional) e o classificador *naïve* Bayes ( $n$  histogramas parciais unidimensionais).

### 3.3.3 Comparação Direta

Como fora feita a aproximação das distribuições de pele para Gaussianas tentando generalizar a distribuição de cor de pixels de pele, para aplicar a regra de decisão Jones & Rehg necessita-se de uma generalização das cores de não-pele. Uma generalização plausível para a distribuição de todas as cores que não fossem pele seria uma distribuição uniforme.

Uma distribuição uniforme apresenta a mesma probabilidade para cada valor possível do conjunto. Logo, o valor da distribuição de pixels de pele é a máxima probabilidade 1 sobre o número total de possíveis valores de cor, como para o presente trabalho as cores foram quantizadas em 256 valores, a probabilidade de cada cor é de  $1/256^d$ , onde  $d$  é o número de dimensões do espaço de cor.

Da mesma forma que para  $p(\text{pele}|c)$  o valor de  $p(c|\text{não-pele})$  só afeta a escolha do valor de limiar. Logo, foram reescritas as equações da regra Jones & Rehg utilizando como valor de probabilidade  $p(c|\text{não-pele})$  e  $p(\text{não-pele})$  como sendo 1. Isso, gera uma comparação direta entre os valores de  $p(c|\text{pele})$  com valores de limiar variando de 0 a  $p(\text{pele})$ .

## 3.4 RESULTADOS

Os resultados primários estão na forma de mapas de pele, como já visto na figura 2 do capítulo 2, onde são apresentadas algumas imagens do banco CompaqDB com os devidos padrões-ouro e alguns mapas de pele gerados por dois classificadores e dois espaços de cor. Entretanto, os resultados mais significativos são mostrados na forma de curvas ROC.

Na análise dos dados do presente trabalho as curvas ROC são de grande valia, pois mostram em valores numéricos a eficiência do classificador para diversos valores de limiar. Além de mostrar valores de desempenho, as curvas são utilizadas para selecionar o valor o limiar que mais se ajusta à aplicação utilizada. Por exemplo, caso a percentagem de erro seja um fator crítico utiliza-se o menor valor de FPR mesmo que prejudicando o valor de TPR.

Para avaliar cada classificador foram geradas curvas ROC utilizando todas as imagens do banco CompaqDB. Com uma grande quantidade de imagens de diversas fontes é possível analisar o real comportamento do classificador, pois efeitos do acaso são minimizados, ou seja, isso evita que a análise por curvas ROC seja tendenciosa e que imagens coletadas para a análise acidentalmente apresentem resultados muito bons ou muito ruins.

As curvas ROC foram montadas utilizando 100 valores para cada um dos classificadores utilizados. Cada ponto na curva representa um valor de limiar,  $\Theta$  ou  $K$ . Os 100 valores de limiar foram igualmente espaçados no intervalo  $[0, MAX\_LIMIAR]$ , onde  $MAX\_LIMIAR$  é o maior valor de limiar aceito pelo classificador. Somente para os classificadores EDSR não foi seguido tal procedimento. Isso pois, não existem parâmetros que possam variar para o referido método, haja visto, que existe somente uma regra fixa. Os valores desses classificadores são apresentados na forma de tabela com um único valor de TPR e FPR.

Para se efetuar uma comparação entre os métodos, como feito nos trabalhos de Phung, Chai e Bouzerdoum (2001), estabeleceu-se um valor de FPR e averiguou-se o valor de TPR correspondente. Esse é um valor empírico de  $15\% \pm 0,2\%$ . As seções a seguir apresentam os resultados específicos de cada espaço de cor para cada classificador.

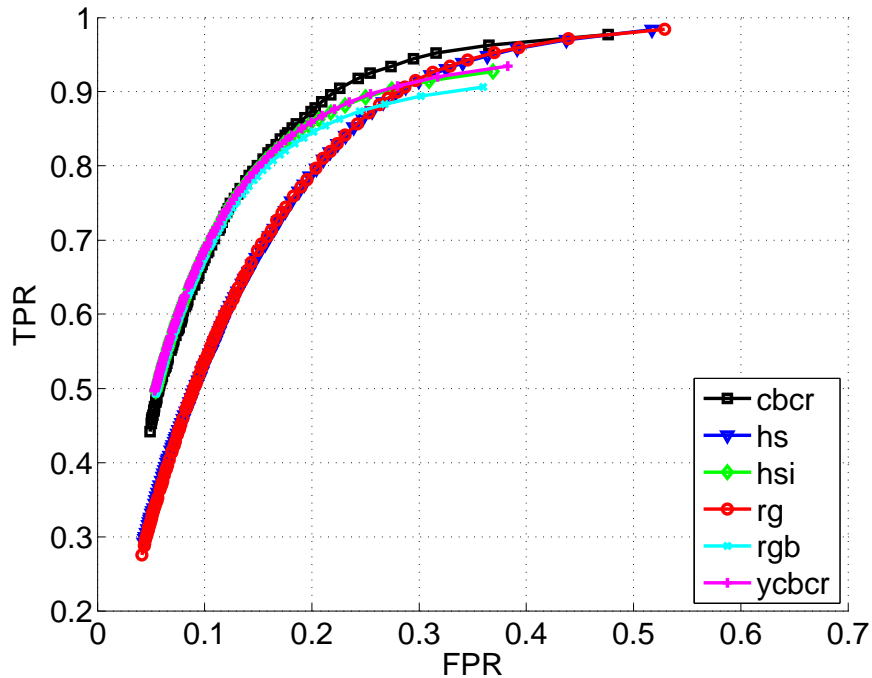
#### 3.4.1 Classificador Jones & Rehg

Para o classificador Jones & Rehg as taxas de acerto são muito parecidas para todos os espaços de cor utilizados no presente trabalho. Espaços de cor 2D obtiveram resultados piores do que espaços 3D, com a exceção do espaço CbCr, o qual em alguns pontos da curva apresentou os melhores resultados. Para melhor visualização foi feita a comparação entre as curvas representadas na figura 4.

Comparando os espaços de cor com taxa FPR próximas a 15% os espaços de cor CbCr e HSI superam os demais. Os resultados utilizando a taxa FPR fixa são mostrados na tabela 1, onde  $K$  é o valor do limiar que gera o valor de FPR desejado.

**Tabela 1: Valores de TPR e FPR para o classificador Jones & Rehg para os espaços de cor analisados.**

Espaço de cor	TPR	FPR	$K$
RG normalizado	0,686	0,149	0,29
HS	0,686	0,152	0,29
CbCr	<b>0,801</b>	0,149	0,23
RGB	0,786	0,149	0,15
HSI	<b>0,801</b>	0,152	0,15
YCbCr	0,799	0,150	0,16



**Figura 4: Curvas ROC para o classificador Jones & Rehg.**

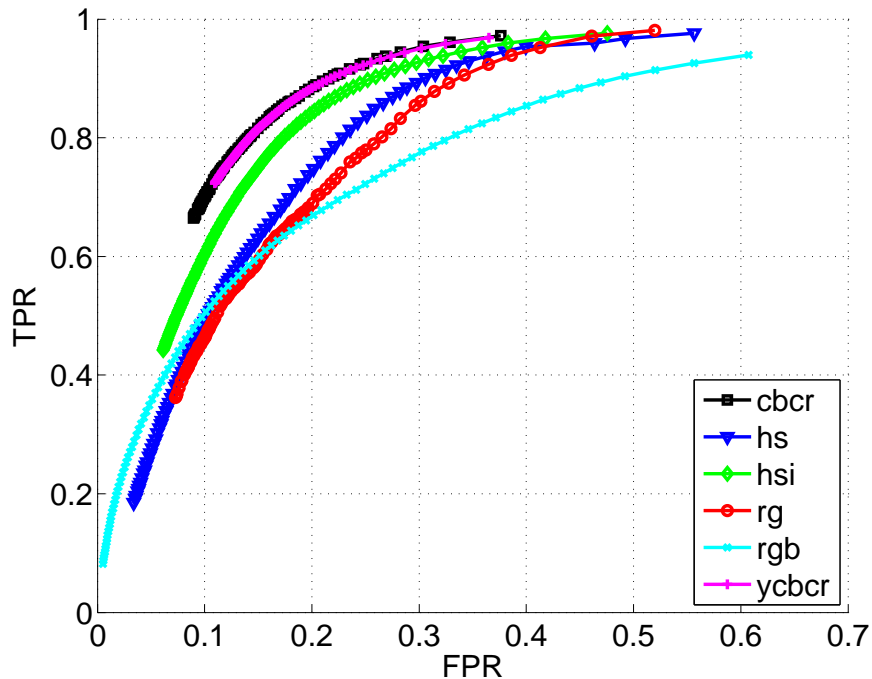
Dentre os espaços de cor utilizados o que obteve melhor resultado é o CbCr, pois até mesmo o espaço de cor HSI que apresenta o mesmo valor para TPR, apresenta um valor levemente mais baixo de FPR (0,03%). Ainda, sem a componente de luminância o CbCr obteve resultados próximos e até melhores a todos os espaços 3D, o que leva a dizer que o CbCr teve os melhores resultados para o classificador Jones & Rehg.

Os valores conseguidos de taxa de acerto parecem baixos a uma primeira vista, quando comparados fixando-se FPR em 15%. Apesar de se conseguir valores mais altos de TPR aceitando valores também maiores de FPR, a segmentação de pele deve ser conduzida de forma conservadora, ou seja, prezando pelo acerto e não pelo erro, haja vista que como o objetivo é reconhecer pessoas, não é interessante deixar margem para outros objetos além de pessoas aparecerem nas predições do classificador.

### 3.4.2 Classificador *naïve* Bayes

Os espaços YCbCr e CbCr são os que apresentaram o melhor desempenho para o classificador *naïve* Bayes. Quando comparado com o classificador Jones & Rehg, as curvas para os espaços de cor YCbCr e CbCr permanecem, com um ligeiro incremento, praticamente as mesmas, enquanto para os outros espaços a taxas de acerto caem. Isso é devido à suposição de que as componentes de cor são independentes. Para os espaços de cor ortogonais isso é válido,

mas para os outros espaços o mesmo não ocorre. Para melhor visualizar os resultados foram traçadas as curvas contidas na figura 5, sobre as quais efetuou-se a comparação entre os espaços de cor.



**Figura 5: Curvas ROC para o classificador *naïve* Bayes.**

Comparando os espaços de cor com taxa FPR próximas a 15% o espaço de cor YCbCr supera os demais. Tais resultados são mostrados na tabela 2.

**Tabela 2: Valores de TPR e FPR para o classificador *naïve* Bayes para os espaços de cor analisados.**

Espaço de cor	TPR	FPR	$K$
RG normalizado	0,595	0,151	0,46
HS	0,630	0,150	0,36
CbCr	0,815	0,149	0,36
RGB	0,601	0,151	0,37
HSI	0,749	0,152	0,37
YCbCr	<b>0,816</b>	0,150	0,47

Como pode-se notar na tabela 2 e na figura 5, a remoção da componente  $Y$  não prejudica em quase nada a classificação, indicando que as componentes de crominância são as mais relevantes para o processo.



### 3.4.3 Classificador por Simples Gaussiana

Na figura 6 é possível ver que à medida que se altera o valor de limiar, as curvas ROC não apresentam a configuração esperada, ou seja, sem um crescimento constante das variáveis de TPR e FPR para o aumento do valor de limiar. A falta de um crescimento constante nos valores de TPR e FPR pode mostrar que talvez a distribuição de pixels de pele para os espaços HS e HSI não se assemelha a uma distribuição Gaussiana. Entretanto, dependendo do valor de limiar que é utilizado, a aproximação à distribuição Gaussiana é válida.

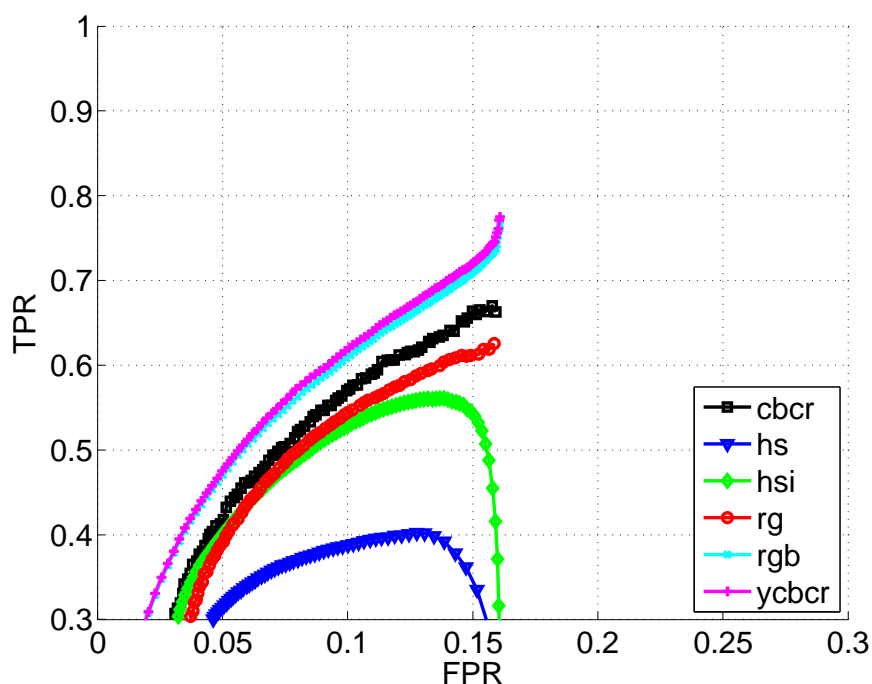


Figura 6: Curvas ROC para o classificador por simples Gaussiana.

Comparando os espaços de cor com taxa de FPR próximas a 15% o espaço de cor CbCr supera os demais. Tais resultados são mostrados na tabela 3.

Tabela 3: Valores de TPR e FPR para o classificador por simples Gaussiana para os espaços de cor analisados.

Espaço de cor	TPR	FPR	$K$
RG normalizado	0,612	0,149	0,95
HS	0,335	0,152	0,95
CbCr	0,664	0,150	0,94
RGB	0,710	0,150	0,82
HSI	0,545	0,149	0,87
YCbCr	<b>0,722</b>	0,151	0,83

### 3.4.4 Classificador MoG

A figura 7 mostra os resultados encontrados para aplicação de um valor de limiar sobre a aproximação de  $p(c|skin)$ . O espaço de cor HSI obteve os melhores resultados que os outros espaços de cor para todos os valores de limiar. Diferentemente do classificador por simples Gaussiana, o classificador MoG supõe que a função distribuição de probabilidade é composta pela soma de 7 Gaussianas ponderadas. O número de Gaussianas foi arbitrado e utilizado para todos os espaços de cor, provando-se satisfatório para o espaço HSI, enquanto para ou outros espaços a representação por simples Gaussiana conseguiu melhores resultados.

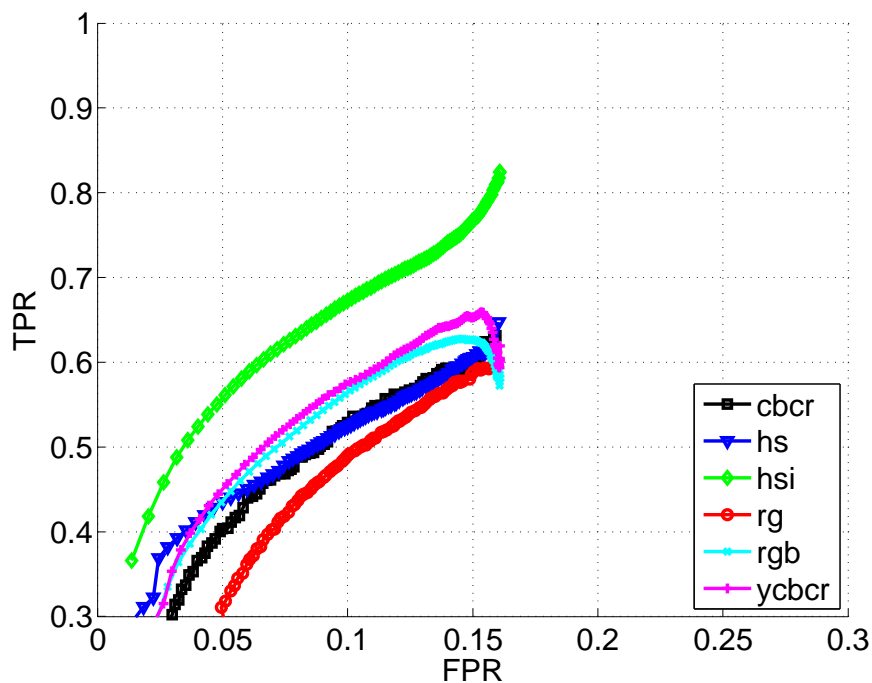


Figura 7: Curvas ROC para o classificador MoG.

Os valores de TPR para uma taxa de FPR próxima de 15% estão mostrados na tabela 4. Entre os valores encontrados, o melhor foi obtido para o espaço de cor HSI, com 77% de taxa acerto.

Tabela 4: Valores de TPR e FPR para o classificador MoG para os espaços de cor analisados.

Espaço de cor	TPR	FPR	$K$
RG normalizado	0,589	0,150	0,91
HS	0,606	0,150	0,88
CbCr	0,596	0,151	0,92
RGB	0,627	0,150	0,67
HSI	<b>0,769</b>	0,151	0,77
YCbCr	0,653	0,150	0,70

Quando comparados os resultados encontrados para o classificador por Simples Gaussiana com os encontrados pelo MoG, é possível notar que para os espaços HSI e HS são melhor representado pelo MoG. A distribuição do HSI e do HS, mesmo não sendo uma única Gaussiana, pode ser representada pela soma ponderada de Gaussianas. Apesar do número suficiente de gaussianas para modelar a distribuição de cor para cada espaço de cor não ser bem definido, são necessárias mais pesquisas sobre esse método para definir um número que melhor se adapte a distribuição dos espaços de cor.

### 3.4.5 Classificadores EDSR

Os classificadores EDSR, apesar de não apresentarem a melhor classificação, têm um desempenho satisfatório e apresentam um uso muito reduzido de recursos de máquina. O classificador EDSR Kovac apresenta uma taxa de acerto baixa, mas é similar as taxas dos classificadores probabilísticos para a mesma taxa de erro. Esse dado pode ser observado nas curvas ROC. Classificadores EDSR prezam pela taxa de acerto, ou seja, são conservadores.

A tabela 5 mostra os resultados dos classificadores EDSR. O classificador proposto por Kovac, Peer e Solina (2003) apresentou os melhores resultados utilizando o espaço RGB.

**Tabela 5: Valores de TPR e FPR para os classificadores EDSR para os espaços de cor específicos.**

Classificador EDSR	TPR	FPR
Kovac - RGB	0,465194	0,0269292
Garcia - YCbCr	0,325298	0,0727721

## 3.5 DISCUSSÕES

A cor da pele é uma característica visual muito atrativa para se trabalhar com detecção de pele, pois a segmentação utilizando a cor tem um processamento rápido e robusto.

Os classificadores que utilizam modelos de pele não-paramétricos obtiveram melhores taxas de acerto e erro para os espaços de cor ortogonais (YCbCr e CbCr). Isso reforça a hipótese de que eles sejam os espaços de cor ótimos para a segmentação de pele. A grande variedade de etnias e condições de iluminação leva o compaqDB a ser um banco de dados útil para se trabalhar com a segmentação de pele, mesmo com as deficiências de informação sobre a condição específica de cada imagem.

A utilização de espaços de cor 2D onde é suprimida a componente de luminância (direta ou indiretamente) trouxe benefícios no processamento e recursos de máquina. Além disso, quando

retirada a componente de luminância as taxas de desempenho se mantiveram relativamente as mesmas. Observou-se que existe certa independência entre canais de cor para espaços de cor ortogonais, haja vista que quando comparados os resultados gerados pelo classificador Jones & Rehg e *naïve* Bayes, esses são semelhantes. Os resultados obtidos pelo classificador *naïve* Bayes mostraram que existe uma maior correlação entre as componentes de cor para os espaços de cor não ortogonais, pois nesses casos as taxas de desempenho do classificador *naïve* Baye quando comparadas com o classificador Jones & Rehg foram mais baixas.

Os classificadores paramétricos obtiveram resultados que levaram a questionar a função densidade de probabilidade da cor de pele. A partir de valores específicos de limiar as curvas ROC começam a não obedecer a tendência esperada, como notado pelo comportamento irregular da curva ROC. Isso mostra que áreas de pele nos espaços de cor não fazem intersecção com as supostas distribuições paramétricas: Gaussianas ou a mistura de 7 Gaussianas. Mais estudos são necessários para averiguar se existe um número de Gaussianas que seja ótimo para cada espaço de cor. Ainda é possível tentar modelar as distribuições através de outras distribuições paramétricas.

Apesar da cor ser considerada a melhor característica para ser utilizada para detectar pele, somente com a informação contida na cor não se pode aumentar as taxas de acerto e erro, haja vista que para o mesmo método probabilístico, independente da transformação feita as taxas variam pouco. Logo, existe a necessidade de se adicionar informação caso se queira aumentar estas taxas. Outras saídas que podem ser utilizadas para a detecção de pixels de pele são a informação de cor da vizinhança dos pixels, misturar informações entre espaços de cor, textura de regiões de pele e o conhecimento sobre as taxas de acerto e erro para cada um dos métodos utilizados.

Como o objetivo maior desse trabalho é detectar conteúdo ofensivo, a detecção de pixels de pele não necessita ser a mais precisa. Entretanto, as informações que serão retiradas das regiões encontradas têm que apresentar informações suficientes para classificar o conteúdo como ofensivo ou não.

Apesar da literatura ter apontado que a iluminação é um fator de erro para os algoritmos de classificação, não se pode confirmar essa afirmativa sem devidos estudos sobre a relação entre a iluminação ambiente, espaços de cor e taxas de acerto e erro. A interferência da iluminação ambiente nas taxas de classificação dos métodos utilizados na presente pesquisa são melhor estudados no capítulo 4.

## 4 EFEITOS DA ILUMINAÇÃO AMBIENTE

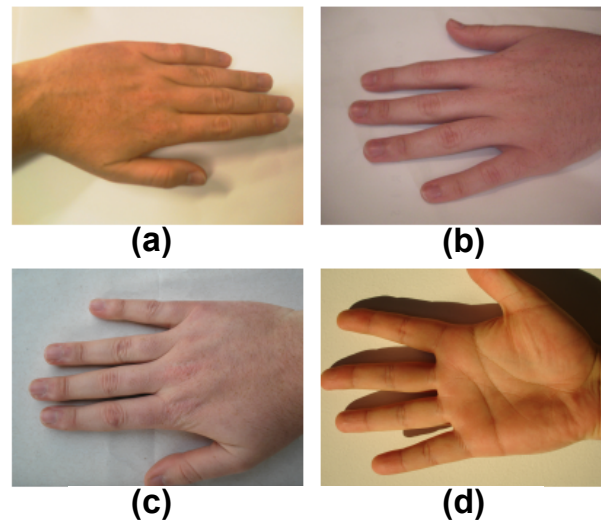
A iluminação ambiente é constantemente citada na literatura como sendo o maior problema na segmentação de pixels de pele (STÖRRING, 2004). Por consequência foi proposto um estudo relacionando a iluminação ambiente e a detecção de pixels de pele. O objetivo desse estudo é verificar se existe um espaço de cor que apresente melhores taxas de acerto e erro para diferentes tipos de iluminação ambiente.

O estudo da iluminação foi realizado da seguinte forma: separação das imagens conforme os tipos de iluminação, aplicação dos métodos e análise das taxas TPR e FPR por tipo de iluminação. Para tal foram aplicados os métodos descritos no capítulo 3 aos seis espaços de cor definidos e a quatro tipos de iluminação ambiente.

Devido ao fato do banco CompaqDB utilizado nas experiências de segmentação de pixels de pele não apresentar informações da iluminação ambiente para a realização do estudo proposto foi necessária a criação de um novo banco de imagens contendo as informações de iluminação necessárias a respeito de cada imagem adicionada ao banco. O banco de imagens criado foi chamando de skinDB.

### 4.1 BANCO DE IMAGENS ESTRUTURADO - SKINDB

Com o intuito de se estudar melhor a segmentação de pele, foi criado um banco de dados estruturado e contendo metadados. Chamado de skinDB, o banco de imagens gerou publicações (KUIASKI et al., 2008, 2009) e é composto atualmente por 120 imagens igualmente distribuídas entre quatro iluminações ambiente: sol, sombra, incandescente e fluorescente. As imagens capturadas são de voluntários de diversas etnias e de partes específicas do corpo humano, como de ante-braços, costas e mãos. Assim como o compaqDB, o skinDB apresenta máscaras de pele manualmente geradas para cada uma das 120 imagens do banco, formando, assim, o padrão-ouro de pele para o banco. Todas as imagens estão codificadas no padrão JPEG, fotos digitais, e os parâmetros das câmeras foram mantidos os mesmos. Exemplos de imagens do skinDB sob diferentes iluminações podem ser vistas na figura 8.



**Figura 8: Imagens do banco skinDB: (a) Imagem na luz fluorescente ; (b) Imagem na luz incandescente; (c) Imagem na sombra; (d) Imagem na luz do sol.**

O banco skinDB é aberto e está disponível juntamente com os respectivos metadados em: <http://www.labiem.cpgei.cefetpr.br/Members/diogo>

Quando foi feita a proposta de disponibilizar publicamente o banco, a intenção era chamar novos pesquisadores a utilizá-lo e atualizá-lo. Por atualizar o banco entende-se facilitar a inclusão de novas imagens, deixar o banco mais estruturado e com isso incentivar as pesquisas sobre segmentação de pixels de pele. Pensando nisso foram propostas formas de acrescentar mais informações além da iluminação ambiente. Até o presente momento, as informações referem-se a dados do indivíduo que está presente na imagem e algumas estatísticas sobre a cor da região de pele da imagem. A técnica utilizada para armazenar informações a respeito das imagens foi a anotação de imagens (KUIASKI et al., 2008).

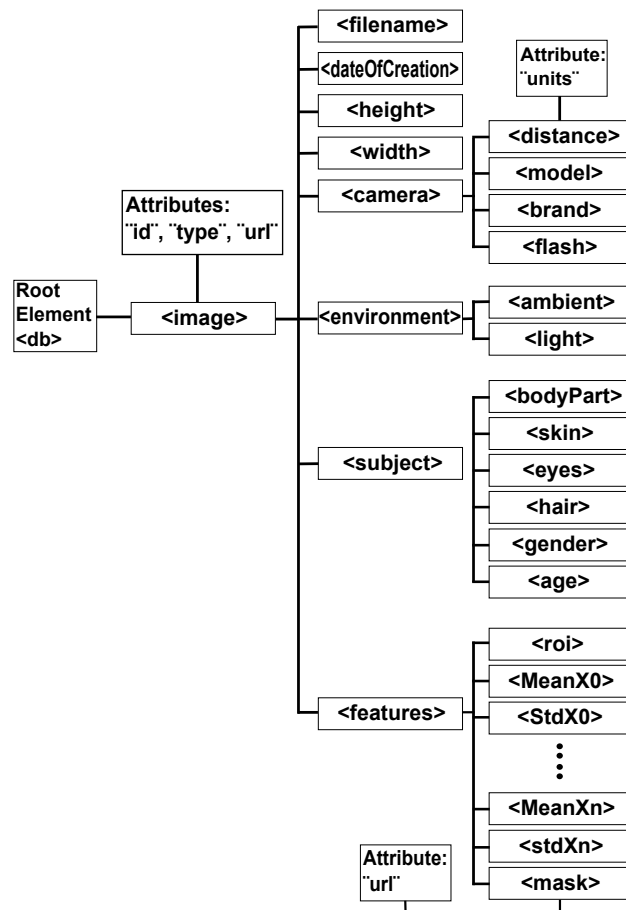
#### 4.1.1 Anotação das Imagens

Esse método prevê a adição de informações não pictoriais a imagens, como por exemplo a forma de algum objeto na imagem, o nome desse objeto, classe semântica ou qualquer informação que seja relevante ao estudo feito. A anotação de imagens pode ser tanto automática quanto manual. Para o skinDB, as informações foram manualmente inseridas, pois fazem parte do padrão-ouro.

As metainformações inseridas no skinDB foram armazenadas utilizando *eXtensible Markup Language* - Linguagem de marcação extensível (XML). Para facilitar a escrita e leitura das im-

agens do banco, desenvolveu-se um aplicativo chamado *skinChooser*. O aplicativo apresenta a visualização reduzida da imagem e uma série de campos a serem preenchidos, os quais contém as metainformações referentes a uma imagem.

As metainformações que são propostas no presente trabalho são listadas na figura 9, a qual mostra que cada elemento *image* apresenta os atributos *id*, *type* e *url*, que são respectivamente o número de identificação no banco, a que classe pertence a imagem, e o *Uniform resource locator* - Localizador uniforme de recursos (URL) da imagem. Dentro do elemento de imagem, os elementos *filename*, *dateIdCreation*, *width* e *height*, não possuem filhos e representam, respectivamente, o nome do arquivo, data de criação e dimensões da imagem em pixels.



**Figura 9: Diagrama em árvore XML desenvolvido.**

**Fonte: (KUIASKI et al., 2008)**

Outros elementos de mesmo nível hierárquico que apresentam filhos são responsáveis pelas demais informações relevantes da imagem. Em *camera* é possível armazenar as informações de distância (*distance*), presença de *flash* (*flash*), modelo (*model*) e marca da câmera (*brand*). Dentro de *environment* é onde se encontram as informações sobre a iluminação ambiente *ambient*

e *light* que são respectivamente o tipo de ambiente e a iluminação do mesmo. Para *ambient* sugere-se que contenham um dos seguintes valores: *indoor* (ambientes fechados) ou *outdoor* (ambientes abertos). O elemento *light* apresenta como valores um dos quatro tipos de iluminação ambiente utilizados no presente estudo.

O elemento *subject* apresenta informações sobre o indivíduo do qual a imagem fora capturada. Dentre os campos ali presentes estão *bodyPart* e *skin*, nos quais são armazenadas a parte do corpo presente na foto e a cor de pele do indivíduo. Neste último sugere-se a utilização das seguintes categorias: *white*, *light brown*, *dark brown* ou *asiatic*.

Por fim, no elemento *feature* podem ser anotadas características de cor média e o desvio padrão dos pixels de pele contidos em regiões ora representadas por máscaras binárias ora por uma *Region of interest* - Região de interesse (ROI) retangular. Os elementos responsáveis por armazenar as informações de cor média e desvio padrão são respectivamente *meanX* e *stdX*, onde *X* é o espaço de cor representado. O elemento *roi* apresenta as coordenadas dessa região retangular, enquanto o elemento *mask* armazena a localização e nome da imagem utilizada como padrão-ouro (GT). Ambos os elementos *roi* e *mask* não são mutuamente excludentes, entretanto é altamente recomendado que pelo menos um seja completado.

## 4.2 ILUMINAÇÃO AMBIENTE

A escolha dos tipos de iluminação foi feita observando quais são os ambientes mais corriqueiros que contém pessoas. A escolha das iluminações foi arbitrária e ainda está sujeita a mudanças. Entretanto, como o resultado do estudo mostra que há variação na cor de pele ocorre para os ambientes escolhidos, estes servem bem ao propósito. A variação da cor de pele em fotos tiradas sob as quatro condições de iluminação pode ser vista na figura 8, na qual todas as imagens pertencem ao mesmo indivíduo.

No presente trabalho foram utilizadas quatro formas de iluminação ambiente: sol (*sun*), sombra (*shadow*), incandescente (*incandescent*) e fluorescente (*fluorescent*). As imagens pertencentes tanto sob a iluminação do sol quanto sob sombra foram retiradas em ambientes abertos, ou ambientes *outdoor*, enquanto para as outras duas iluminações as fotos foram capturadas em ambientes fechados, ou ambientes *indoor*.

A metodologia utilizada na captura das imagens realizada em ambientes abertos foi capturar imagens em diversos horários do dia sob a iluminação direta ou indireta da luz solar, iluminações sob sol e sombra respectivamente. A captura das imagens para ambientes fechados foi feita a partir do tipo de lâmpada presente no ambiente, fluorescente ou incandescente.



### 4.3 MÉTODOS UTILIZADOS

Por questão de coerência foram utilizados alguns dos classificadores apresentados no capítulo 3. São eles o classificador *naïve* Bayes, Jones & Rehg, MoG e por simples Gaussiana. Os classificadores EDSR não foram utilizados no teste por não apresentarem flexibilidade de variação nos parâmetros que os compõem, pois as comparações são feitas utilizando curvas ROC e o único resultado comparável gerado pelos EDSR é um único ponto no plano  $TPR \times FPR$ .

A etapa de geração dos modelos foi feita utilizando a seleção CompaqDBeasy500, pois o banco contém uma grande diversidade de imagens de diversas etnias e sob as mais diversas iluminações ambiente. Manter o modelo de pele gerado pela CompaqDBeasy500, e não reconstruir um novo a partir de parte do skinDB, tem como objetivo não utilizar um modelo tendencioso de cor de pele.

### 4.4 RESULTADOS

Os resultados que são mostrados nesta seção são, exclusivamente, curvas ROC discriminadas por classificador e por iluminação. Entretanto, antes de entrar no mérito da avaliação dos efeitos da iluminação na classificação de pele, é abordado o desempenho geral dos classificadores para o banco skinDB.

A avaliação foi feita utilizando todas as imagens do banco skinDB para os quatro classificadores (Jones & Rehg, *naïve* Bayes, MoG e simples Gaussiana) e para todos os seis espaços de cor (RGB, RG normalizado, HSI, HS, YCbCr e CbCr), apresentados anteriormente. As figuras 10 e 11 mostram as curvas ROC obtidas para o classificador Jones & Rehg e as obtidas para o classificador *naïve* Bayes. As figuras 12 e 13 mostram as curvas ROC obtidas para o classificador MoG e as obtidas para o classificador por simples Gaussiana.

As avaliações gerais dos classificadores mostradas nas figuras 10, 11, 12 e 13 indicam que, na maioria dos casos, espaços de cor bidimensionais apresentam melhores resultados, ou seja, que somente a informação de cromaticidade é suficiente para classificar pixels de pele. Os melhores resultados foram obtidos para os espaços de cor RG normalizado e HS para o classificador Jones & Rehg, e HSI e HS para o *naïve* Bayes. Os resultados ora apresentados quando comparados com os resultados obtidos por outros pesquisadores, como para Phung, Bouzerdoum e Chai (2005) e Jones e Rehg (2002), parecem contraditórios, pois para esses pesquisadores os espaços de cor tridimensionais atingiram melhores resultados que os bidimensionais em todos os pontos da curva ROC. Uma hipótese para essas discrepâncias é que o banco skinDB é rel-

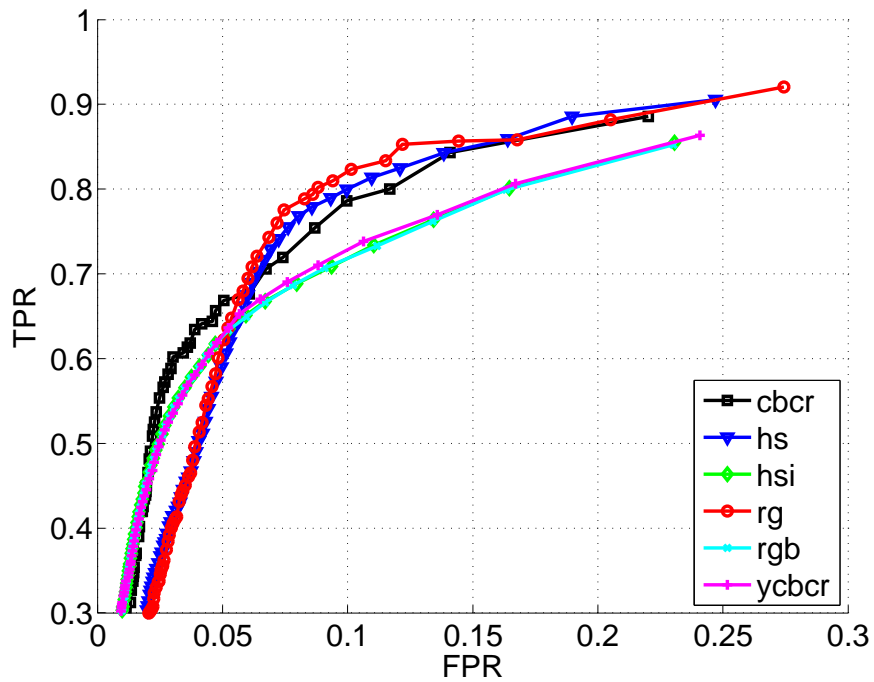


Figura 10: Curvas ROC para o classificador Jones & Rehg para todos os espaços de cor citados.

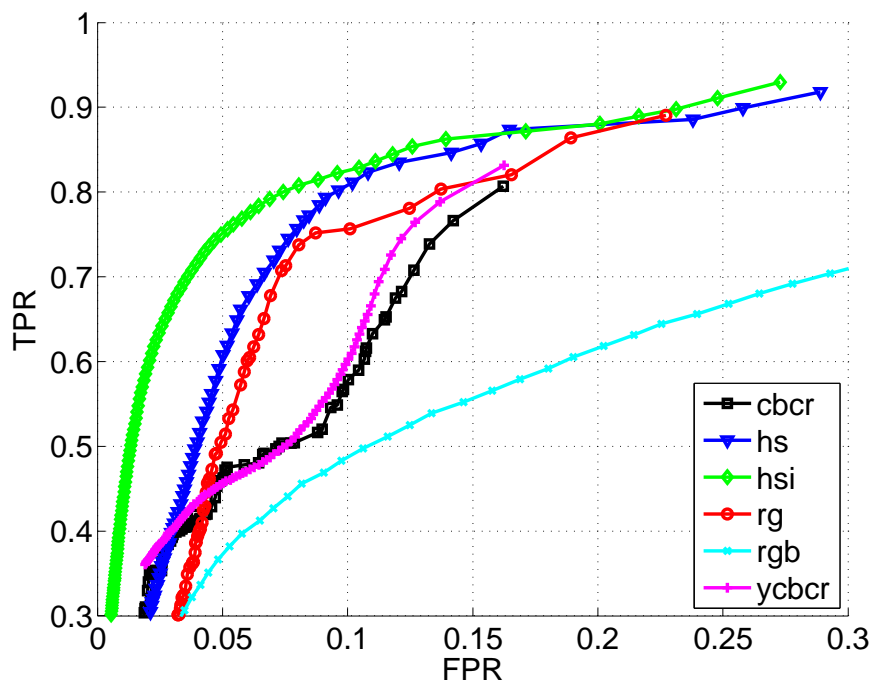


Figura 11: Curvas ROC para o classificador *naïve* Bayes para todos os espaços de cor citados.

ativamente pequeno, especialmente feito com imagens adquiridas de forma controlada com o propósito de estudar segmentação de pele.

Outro resultado importante, retirado a partir das curvas ROC do classificador Jones & Rehg,

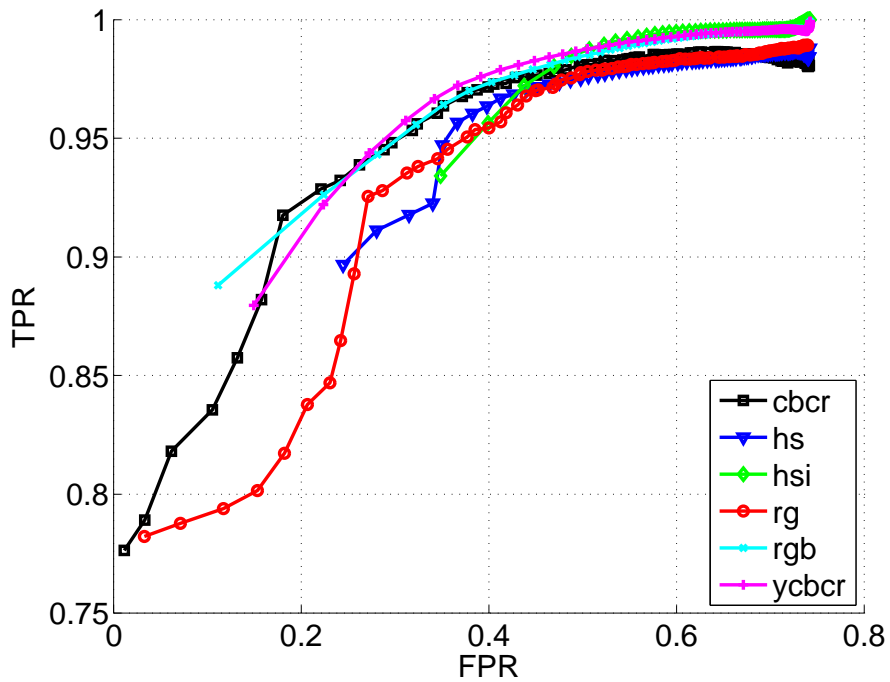


Figura 12: Curvas ROC para o classificador MoG para todos os espaços de cor citados.

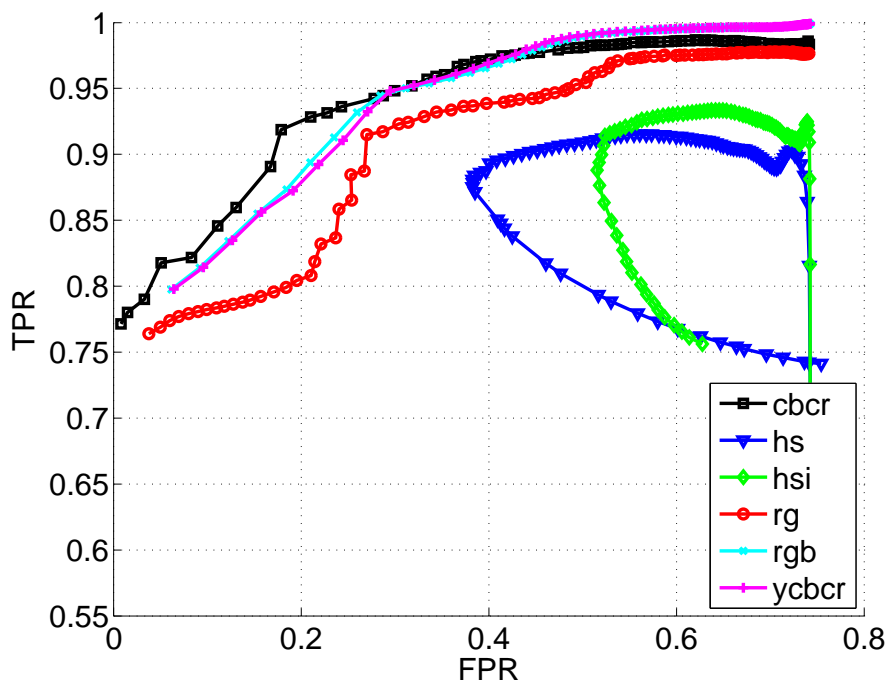


Figura 13: Curvas ROC para o classificador por simples Gaussiana para todos os espaços de cor citados.

está no fato de que todos os espaços de cor tridimensionais obtiveram resultados similares (figura 10). Isso indica que não há perda significativa de informação pela transformação do espaço RGB para outro espaço tridimensional, como esperado.

Quando são analisadas as curvas do classificador *naïve* Bayes, o único espaço de cor que mantém resultados semelhantes aos encontrados para o classificador Jones & Rehg é o HS (figura 11), enquanto que para o HSI os resultados foram melhores. Tais resultados para os espaços HS e HSI podem ser explicados pois esses espaços apresentam uma alta independência entre as componentes de cor, ou seja, a probabilidade de um pixel corresponder à classe pele não depende da probabilidade total, mas somente das probabilidades independentes de cada componente de cor. Entretanto, comparando os resultados do HS e HSI com os resultados obtidos no capítulo 3, para os mesmos classificadores, pode-se afirmar que o banco skinDB ainda não apresenta a mesma diversidade de cores de pele do compaqDB, haja visto que para o compaqDB a independência entre canais não ocorreu para esses espaços de cor.

Algo semelhante ocorre com os espaços de cor YCbCr e CbCr. Os resultados são piores do que os encontrados para o classificador Jones & Rehg, mas as curvas permanecem próximas uma da outra, o que nos leva a concluir que existe uma independência entre os canais do YCbCr e que a componente *Y* não afeta significativamente o resultado. Finalmente, os resultados para o espaço RGB mostram que esse apresenta uma alta correlação entre os canais de cor como esperado dada a total dissimilaridade entre as curvas ROC, diferentemente do que foi observado para o HSI e o YCbCr.

A avaliação dos classificadores paramétricos é um pouco diferenciada. Sem muitas variações entre os resultados para diferentes espaços de cor e para os mesmos valores de FPR tanto o classificador MoG (figura 12) como o por simples Gaussiana (figura 13) apresentam valores mais elevados de TPR. Isto se deve muito ao banco skinDB, o qual não contém toda a variação de cores de pele que está presente no banco CompaqDB. Com uma quantidade reduzida de cores é possível mais facilmente classificá-las nos modelos mais complexos gerados pela CompaqD-Beasy500. O classificador MoG nos espaço de cor HSI e HS obtiveram os piores resultados, embora tenham obtido melhores resultados quando feita a análise de todo o banco CompaqDB.

A análise dos resultados foi dividida por classificador e por iluminação ambiente. Primeiramente, a influência da iluminação é analisada para o classificador Jones & Rehg, cujos resultados estão nas figuras 14 e 15. Após, é analisada a influência da iluminação para o classificador *naïve* Bayes, cujo os resultados estão nas figuras 16 e 17. Igualmente, os resultados para o classificador MoG e por simples Gaussiana estão nas figuras 18 e 19 e 20 e 21, respectivamente.

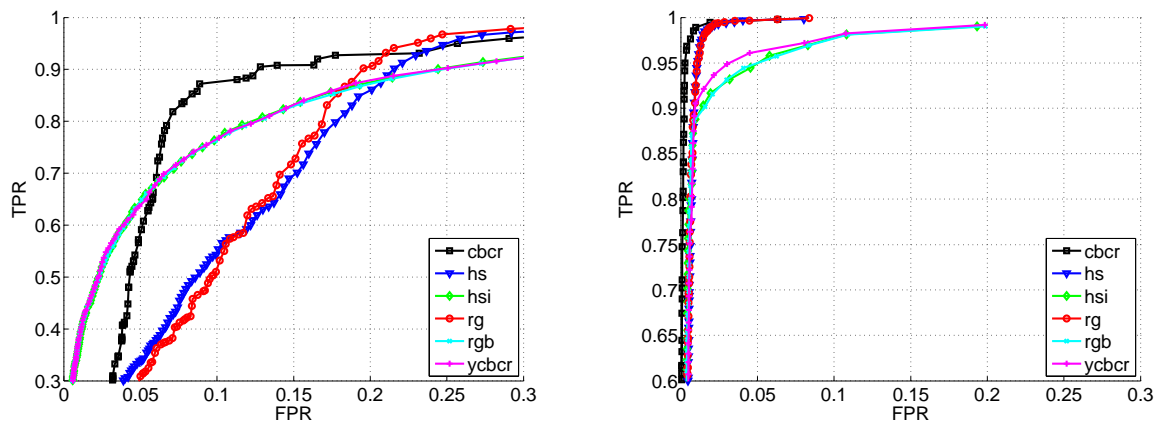
#### 4.4.1 Classificador Jones & Rehg

Os resultados do classificador Jones & Rehg para diferentes iluminações evidencia que espaços de cor tridimensionais atingem melhores taxas de classificação para uma faixa de val-

ores de FPR próxima da origem. Para valores de FPR menores que 5% e sob as iluminações sol e incandescente, espaços de cor tridimensionais apresentam melhores resultados do que espaços de cor bidimensionais. Entretanto, se for admissível uma FPR maior, espaços de cor bidimensionais são uma melhor escolha. Sob a iluminação sombra o mesmo acontece, mas para um valor menor de FPR (aproximadamente 2,5%). As curvas ROC dos espaços de cor tridimensionais, independentemente da iluminação, continuam similares.

Para as imagens na luz do sol os resultados (figura 14 - esquerda) mostram que o espaço de cor CbCr apresenta os melhores resultados para essa iluminação, 86% TPR para 10% FPR. A iluminação da luz do sol apresentou os resultados mais divergentes, especialmente para o espaço de cor RG normalizado e para o HS. Os resultados divergentes podem ser influência do horário do dia em que foram feitas as imagens, pois a composição de cores muda pelo efeito de refração da luz solar na atmosfera. Como por exemplo a ocorrência de sombra nas imagens devida ao ângulo de incidência da luz, o que pode ser visto na figura 8.

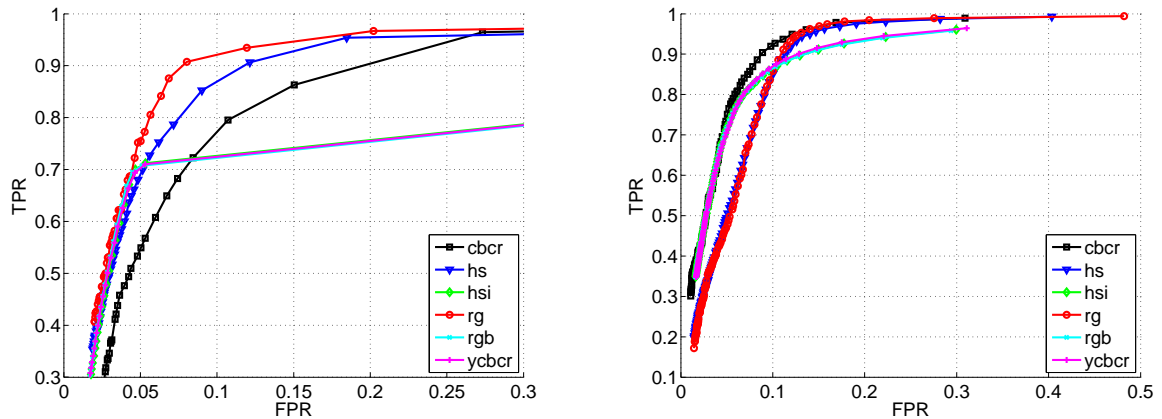
Os melhores resultados do trabalho foram obtidos para o classificador Jones & Rehg sob a iluminação sombra (figura 14 - direita). Todos os espaços de cor utilizados atingiram valores de TPR maiores de 95% para valores de FPR inferiores a 2,5%. Acredita-se que essas taxas se devem à pouca influência do controle de equilíbrio de branco (*white balance*) da câmera na iluminação sombra, conservando a cor da pele.



**Figura 14:** Curvas ROC para o classificador Jones & Rehg utilizando imagens em ambientes abertos: (esquerda) sol e (direita) sombra.

As imagens adquiridas na luz incandescente (figura 15 - direita) obtiveram resultados entre os obtidos sob a sombra e os sob luz solar. A hipótese de que o equilíbrio de branco automático da câmera altera significativamente as cores da imagem é reforçada, pois a luz incandescente tem um espectro semelhante ao da luz solar, entretanto de menor intensidade e por esta razão pode ser responsável pelas alterações nas taxas TPR e FPR.

Os melhores resultados de classificação, 92,5% TPR para 10% FPR, para a luz incandescente foram conseguidos utilizando o espaço CbCr. Para o estudo da influência da iluminação optou-se por fixar o valor de FPR em 10%, pois os resultados do skinDB são significativamente melhores que os encontrados para o CompaqDB e reduzindo o valor de comparação a variação de TPF é maior.



**Figura 15:** Curvas ROC para o classificador Jones & Rehg utilizando imagens em ambientes fechados: (esquerda) fluorescente e (direita) incandescente.

A tabela 6 resume as taxas de classificação obtidas para o classificador Jones & Rehg para todas as iluminações e espaços de cor. Nessa tabela foi fixado o valor de FPR em 10%.

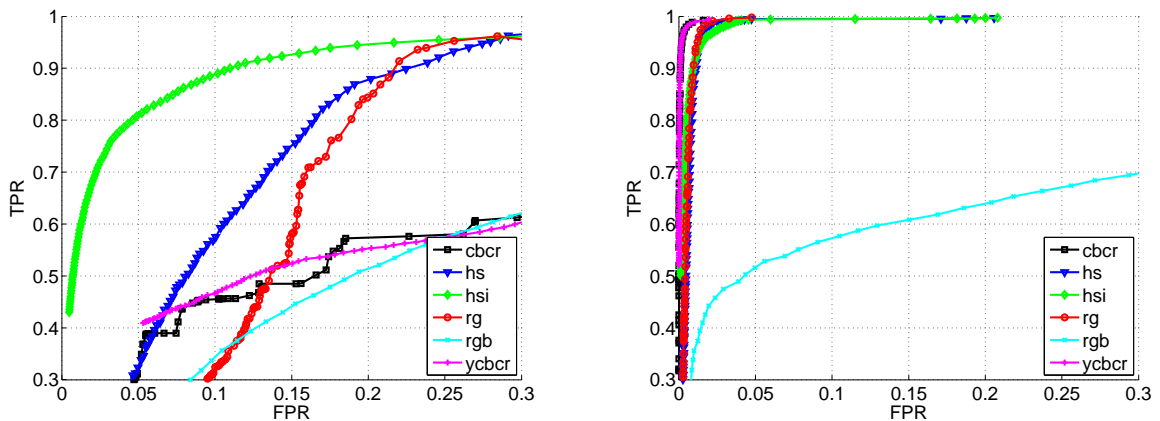
**Tabela 6:** Valores de TPR em porcentagem do classificador Jones & Rehg para 10% de FPR

	Sol	Sombra	Incandescente	Fluorescente
RGB	76,3%	96,5%	86,3%	75,5%
RG	52,1%	99,6%	85,2%	92,1%
HSI	76,5%	96,5%	86,3%	75,5%
HS	54,7%	99,5%	84,7%	87,0%
YCbCr	75,5%	96,4%	85,6%	75,5%
CbCr	87,1%	99,0%	91,9%	77,3%

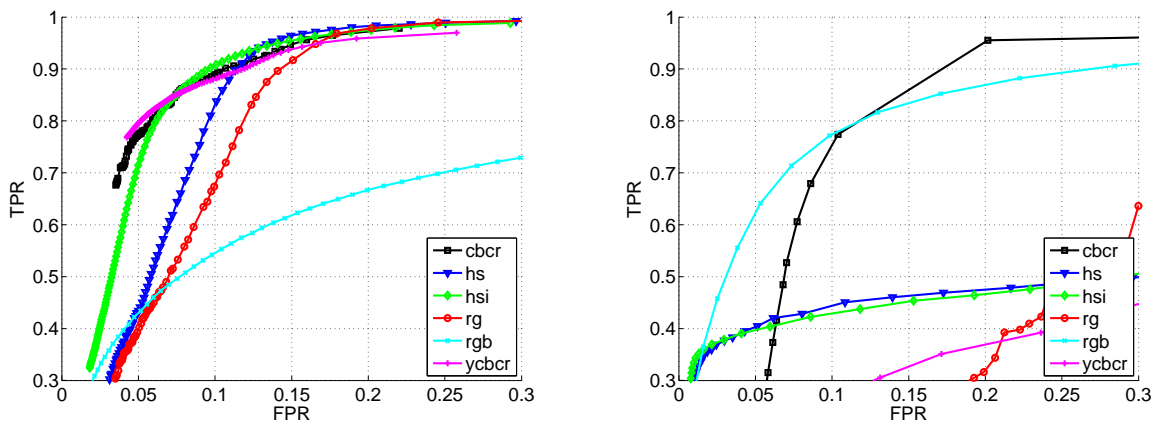
#### 4.4.2 Classificador *naïve* Bayes

O classificador supõe que as componentes de cor são estatisticamente independentes umas das outras. Apesar de se saber que, na maioria dos casos, a independência estatística não se mantém, o classificador foi utilizado para investigar qual espaço de cor apresenta uma maior independência estatística entre as componentes. Isso economiza recursos computacionais (memória e processamento) e pode mostrar para qual espaço de cor pode-se retirar a componente de luminância sem que sejam alteradas as taxas de detecção.

Como afirmado anteriormente, os canais do espaço RGB são os que apresentam a maior dependência entre as componentes de cor. Esse efeito pode ser visto no desempenho do classificador *naïve* Bayes como mostram as figuras 16 - esquerda e 17 - esquerda, nas quais o espaço de cor RGB obteve as menores taxas de acerto. Por razões ainda desconhecidas, sob luz fluorescente o espaço de cor RGB apresenta resultados relativamente bons em relação aos demais espaços de cor, como mostra a figura 17 - direita. Entretanto, no geral os resultados para luz fluorescente são piores do que para outros tipos de iluminação.



**Figura 16:** Curvas ROC para o classificador *naïve* Bayes utilizando imagens em ambientes abertos: (esquerda) sol e (direita) sombra.



**Figura 17:** Curvas ROC para o classificador *naïve* Bayes utilizando imagens em ambientes fechados: (esquerda) incandescente e (direita) fluorescente.

Para as imagens obtidas na luz solar (figura 16 - esquerda), os melhores resultados foram atingidos pelos espaços de cor HSI, HS e RG normalizado, atingindo 90% TPR para 20% FPR.

Na sombra (figura 16 - direita), os resultados foram similares aos obtidos para o classificador Jones & Rehg, com a exceção do espaço RGB., atingindo taxas de acerto maiores que 95% para taxas de erro inferiores a 5% para o classificador *naïve* Bayes. Pde se dizer que

pelo fato do RGB apresentar resultados inferiores a todos os outros espaços de cor, este espaço apresenta uma grande correlação estatística entre canais.

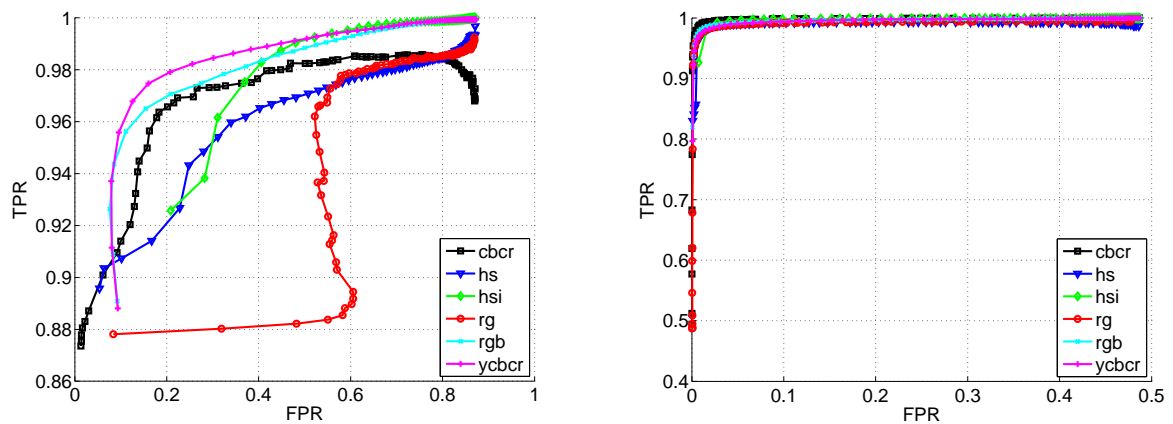
A tabela 7 resume os valores de TPR obtidos para FPR fixo em 10% para todos os espaços de cor e condições de iluminação utilizados nesse trabalho para o classificador *naïve* Bayes.

**Tabela 7: Valores em porcentagem de TPR para o classificador *naïve* Bayes para 10% de FPR**

	Sol	Sombra	Incandescente	Fluorescente
RGB	23,3%	42,9%	45,5%	77,4%
RG	36,2%	95,2%	43,4%	<20%
HSI	71,5%	99,5%	85,2%	42,9%
HS	54,5%	99,8%	74,2%	44,4%
YCbCr	42,4%	99,6%	88,9%	26,0%
CbCr	41,4%	99,5%	89,0%	75,3%

#### 4.4.3 Classificador MoG

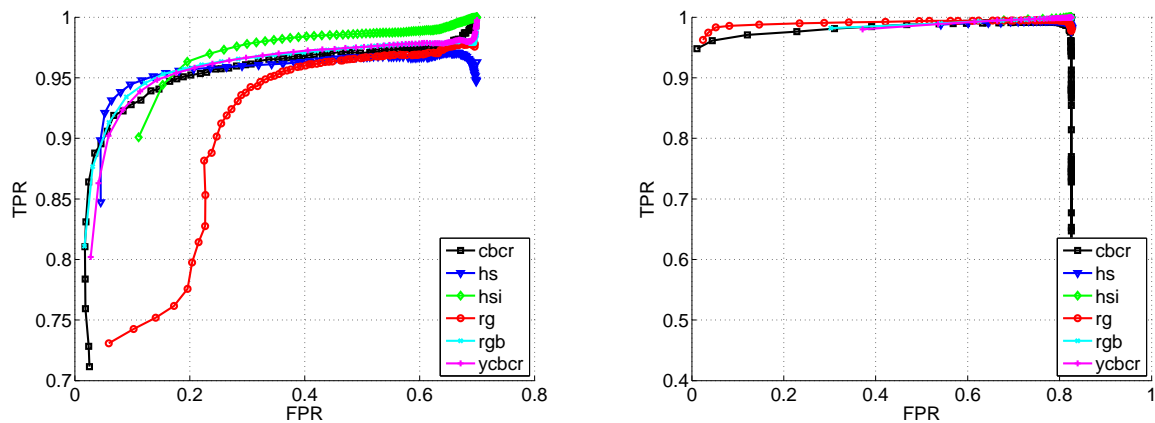
No caso do classificador MoG, as curvas para as iluminações sob a sombra e sob a luz incandescente foram consistentes com os resultados encontrados para os outros classificadores e pouca diferença foi encontrada entre os resultados para os diferentes espaços de cor. Os resultados do classificador MoG para as diferentes iluminações pode ser visto nas figuras 18 e 19.



**Figura 18: Curvas ROC para o classificador MoG utilizando imagens em ambientes abertos: (esquerda) sol e (direita) sombra.**

Nas curvas ROC que representam as iluminações fluorescente e sol fogem do padrão encontrado para outros classificadores desse trabalho. Entretanto quando comparamos os resultados encontrados para o skinDB com os do CompaqDB pode-se observar que a anomalia citada no capítulo 3 para os espaços HSI e HS não se mantém para o skinDB. É possível que as





**Figura 19: Curvas ROC para o classificador MoG utilizando imagens em ambientes fechados: (esquerda) incandescente e (direita) fluorescente.**

distribuições densidade de probabilidade do skinDB sejam melhor representadas pelo modelo criado pelo classificador MoG. Apesar as curvas irregulares, as curvas ROC para o classificador MoG apresentam tanto taxas de acerto quanto taxas de erro muito elevadas quando comparadas a dos classificadores Jones & Rehg, *naïve* Bayes e com os próprios classificadores quando analisadas as imagens do banco CompaqDB.

Para as imagens adquiridas na luz solar (figura 18 - esquerda), os melhores resultados foram atingidos pelos espaços de cor YCbCr, CbCr e RGB, atingindo mais de 95% TPR para 20% FPR. Na sombra (figura 18 - direita), os resultados foram os melhores resultados de todos os testes realizados, superando até mesmo os resultados do classificador Jones & Rehg.

A tabela 8 resume os valores de TPR (10% FPR) obtidos para todos os espaços de cor e condições de iluminação utilizados nesse trabalho para o classificador MoG.

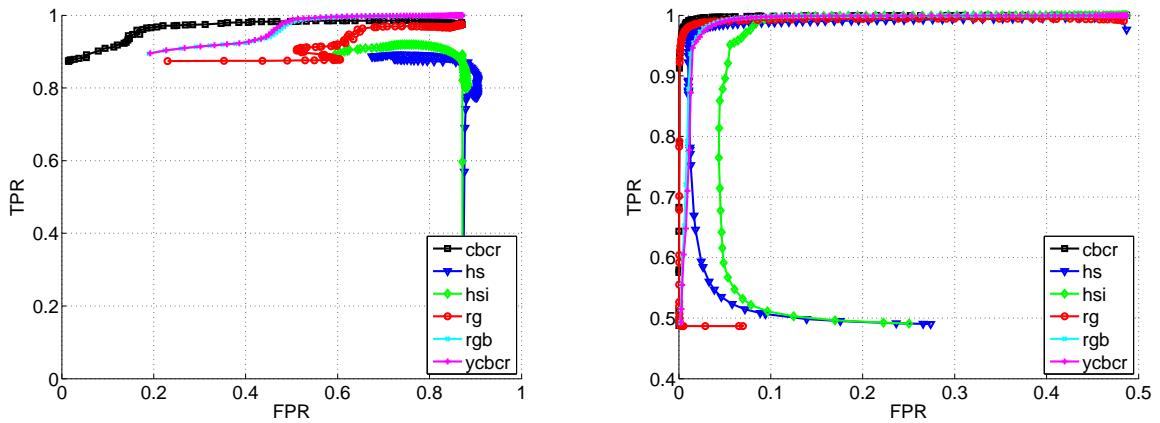
**Tabela 8: Valores em porcentagem de TPR para o classificador MoG para 10% de FPR**

	Sol	Sombra	Incandescente	Fluorescente
RGB	99,5%	99,5%	94,0%	97,0%
RG	88,0%	99,1%	94,0%	99,0%
HSI	92,0%	99,8%	90,0%	98,0%
HS	90,5%	99,2%	94,5%	96,0%
YCbCr	95,5%	99,4%	93,5%	95,5%
CbCr	91,0%	99,8%	93,0%	97,5%

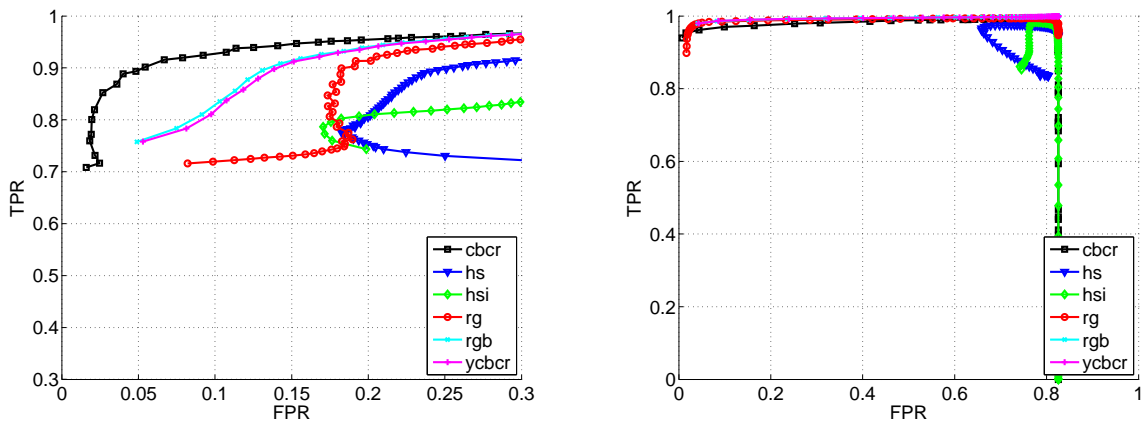
#### 4.4.4 Classificador por Simples Gaussiana

O classificador por simples Gaussiana não apresentou resultados semelhantes a nenhum dos classificadores estudados. Apesar disso, é possível ver que a estranha anomalia na curva ROC

encontrada na análise dos resultados para todas as imagens do skinDB para esse classificador é reduzida sob a sombra (figura 20 - direita). Analisando as curvas pode-se notar que as curvas do CbCr apresentam melhores resultados, o que leva a afirmar que a distribuição de pixels de pele do o skinDB para o espaço CbCr seja a mais próxima de uma Gaussiana.



**Figura 20:** Curvas ROC para o classificador simples Gaussiana utilizando imagens em ambientes abertos: (esquerda) sol e (direita) sombra.



**Figura 21:** Curvas ROC para o classificador simples Gaussiana utilizando imagens em ambientes fechados: (esquerda) incandescente e (direita) fluorescente.

#### 4.5 DISCUSSÕES

Apesar da cor de pele ser uma característica muito atrativa para a segmentação de pixels de pele, os efeitos da iluminação apresentam forte influência sobre esses métodos de segmentação. As condições de iluminação e o controle de equilíbrio de branco (*white balance*) da câmera alteram a cor de pele adquirida pelo sensor da câmera, o que levou a se investigar qual espaço de cor minimizaria os efeitos no desempenho da classificação. Um ponto importante sobre

o banco é que para nenhuma imagem foi desabilitado o controle de equilíbrio de branco da câmera. Acredita-se que tal fato possa ter gerado algumas variações nas taxas de classificação.

Os experimentos feitos mostram que espaços de cor bidimensionais (RG norma., HS e CbCr) em geral apresentam melhores resultados descartando as informações de luminância e por consequência reduzindo efeitos da iluminação. Os resultados dos espaços bidimensionais mostram que a utilização de somente informação de crominância é suficiente para uma classificação de pixels de cor razoavelmente boa. No geral o melhor classificador é o por simples Gaussiana, para o espaço de cor CbCr para o skinDB. O melhor resultado geral se mostrou muito diferente dos encontrados para o CompaqDB, levando a concluir que o skinDB ainda não contém toda a variedade de cores do CompaqDB.

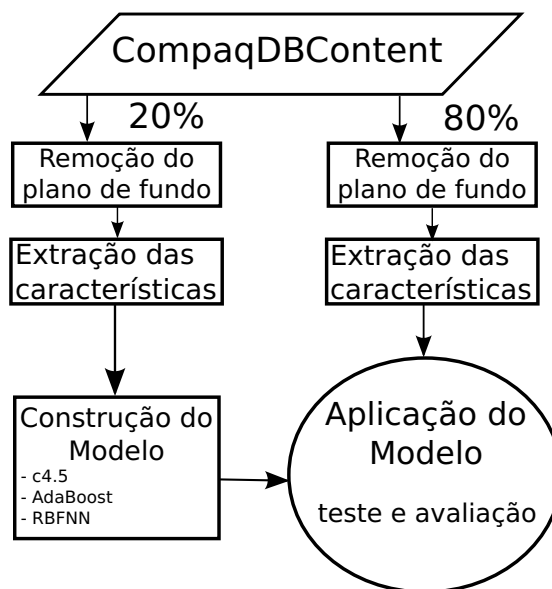
Entretanto, mesmo após analisar seis espaços de cor, não foi possível identificar um que tenha as melhores taxas de acerto para todas as iluminações. Outros espaços de cor devem ser testados para que se comprove a existência de um espaço ótimo para a detecção de pixels de pele. Aqueles resultados para os classificadores de modelos paramétricos que foram muito bons devem ser olhados com desconfiança. O banco skinDB, apesar de terem sido tomados os devidos cuidados para manter a diversidade de imagens, ainda é pequeno e provavelmente tendencioso na escolha das imagens e na composição de cor de pele.

Pesquisas futuras são necessárias para esclarecer resultados contraditórios entre os experimentos do presente trabalho e os de outros (PHUNG; BOUZERDOUM; CHAI, 2005; JONES; REHG, 2002). Trabalhos futuros incluem melhorar o banco de imagens skinDB utilizando outros parâmetros de controle, como o controle de equilíbrio de branco da câmera, tipo de fundo nas fotos, identificação anônima das pessoas fotografadas e distância da câmera. Esse cenário mais controlado pode contribuir para deixar mais marcante quais variações afetam mais a cor de pele.

Algoritmos de constância de cores (FINLAYSON; SCHIELE; CROWLEY, 1998) também devem ser pesquisados com o intuito de tentar aproximar as cores a um modelo próximo do observado na iluminação sombra, isso pois tal iluminação obteve os melhores resultados para todos os modelos e espaços de cor.

## 5 DETECÇÃO DE CONTEÚDO OFENSIVO

O presente trabalho teve como proposta estudos sobre a detecção de conteúdo ofensivo e consequente análise da viabilidade da criação de um aplicativo para tal. Dentre os estudos de detecção de conteúdo, foram realizadas análises estatísticas de métodos de CBIR e de conjunto de características. A detecção de conteúdo ofensivo foi feita sobre as imagens do CompaqDB separadas por conteúdo, a seleção compaqDBContent. Foram extraídas características de cada uma das imagens e o vetor de características é aplicado a um classificador. Os modelos de conteúdo são montados com 20% do total de vetores de características e sobre os outros 80% são aplicados os classificadores e geradas as estatísticas de acerto e erro. O diagrama em blocos desse procedimento pode ser visto na figura 22.



**Figura 22: Diagrama do procedimento feito para o estudo da detecção de conteúdo ofensivo.**

Antes da extração de características é feita a remoção do plano de fundo sobre todas as imagens da seleção compaqDBContent. A remoção do plano de fundo é uma técnica que tem como objetivo restringir a cena a ação exercida por pessoas. Feita a remoção do plano de fundo é feita extração do vetor de características e armazenamento das informações em um arquivo *Comma Separated Values* - Valores separados por vírgulas (CSV). Nesse arquivo cada vetor

de características é acompanhado pelo nome da classe a qual pertence, o conjunto de vetores gera o padrão-ouro para a detecção de conteúdo. Tanto a remoção do plano de fundo quanto a extração de algumas características é feita sobre um mapa de pele. Para gerar o mapa de pele, o método de segmentação de pele utilizado para gerar mapas de pele foi o classificador Jones & Rehg para o espaço RGB com limiar de 0,11.

## 5.1 CONJUNTO DE CARACTERÍSTICAS

O conjunto de características foi escolhido durante os estudos do trabalho de Shih, Lee e Yang (2007). Esses autores propõem uma abordagem utilizando a remoção do plano de fundo e descritores de cor, textura e forma. Os descritores usados pertencem às normas MPEG-7 (MANJUNATH, 2002). O descritor de cor utilizado foi o descritor escalável de cor, o de textura foi o descritor histograma de bordas e o de forma foi descritor de compactação. Além dessas características foram incorporadas as dimensões da imagem e a projeção das regiões de pele na horizontal e vertical. Maiores detalhes são apresentados nas subseções a seguir.

### 5.1.1 Remoção do Plano de Fundo

O objeto de interesse em imagens de conteúdo ofensivo são pessoas, razão pela qual é feita a remoção do plano de fundo. Esse método tenta abstrair o cenário e concentrar o reconhecimento da cena nas pessoas da imagem e nas ações que elas estejam executando.

A remoção é feita sobre o mapa de pele de uma imagem submetida a segmentação de pixels de pele pelo classificador Jones & Rehg. Sobre o mapa é aplicado o operador morfológico da abertura, para agrupar os pixels de pele, então o mapa é dividido em blocos de  $32 \times 32$  pixels. Se a quantidade de pixels de pele de cada bloco for maior do que 50% do número de pixels total do bloco, esse é considerado um bloco de pele. Feito isso, a maior região conectada de blocos de pele é considerada como sendo a região de interesse. A imagem resultante da remoção do plano de fundo é a formada pelo menor retângulo capaz de englobar toda a região de interesse.

Caso a imagem seja menor que  $64 \times 64$  pixels, ou seja o dobro do tamanho do bloco, ou o número de regiões de pele seja inferior a dois, não é executada a remoção do plano de fundo. A extração das características é feita sobre a imagem inteira.

### 5.1.2 Descritor Escalável de Cor

O *Scalable Colour Descriptor* - Descritor escalável de cor (SCD) é um histograma de cor de dimensões variáveis. Ou seja, cada componente de cor será quantizada novamente em um número menor que 256 posições, ou 8 bits. O grande mérito do método é que as componentes não precisam necessariamente apresentar a mesma quantização, logo o tamanho do vetor do descritor SCD é variável. Esse tipo de abordagem traz benefícios para a detecção de conteúdo, pois pode-se dar mais importância a uma determinada componente de cor, a qual mais se discrimina no contexto da detecção.

No detector de conteúdo ofensivo são utilizados os espaços de cor RGB, HSI e YCbCr para formar o SCD. O vetor resultante é composto por 256 ou 512 posições em diferentes escalas. As escalas utilizadas servem para a avaliação da influência do espaço de cor na detecção de conteúdo e a proporção entre canais de cor para esse trabalho podem ser (4:4:4) e (16:4:4). A distribuição entre cores foi colocada de forma que a cor mais discriminante em imagens de conteúdo ofensivo fosse a cor de pele, a qual contém uma quantidade maior de vermelho na composição. O histograma é global e feito para a imagem após a remoção do plano de fundo. O vetor **sdc** resultante é dado na seguinte forma:

$$\mathbf{sdc} = [s[1], s[2], \dots, s[n]],$$

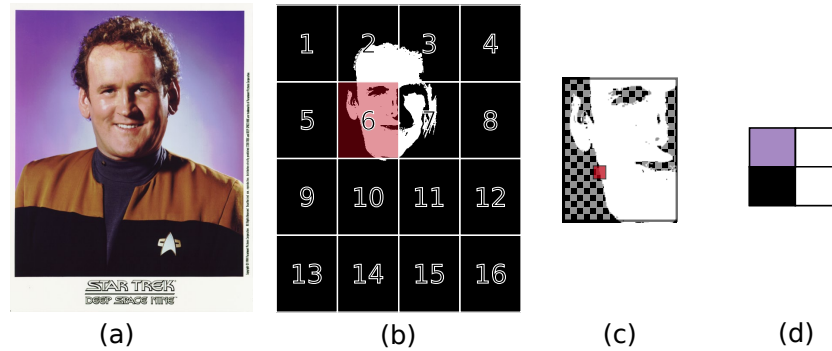
onde  $s[i]$ ,  $i \in [1, n]$  é a probabilidade da componente de cor  $i$  e  $n$  o número de posições do vetor.

### 5.1.3 Descritor Histograma de Bordas

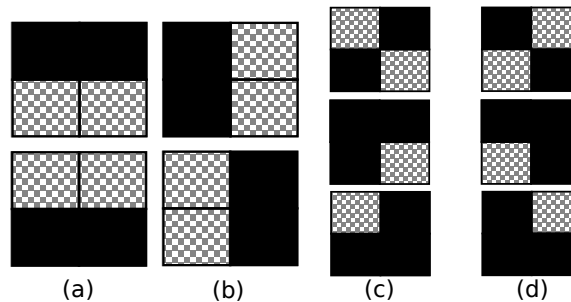
As normas MPEG-7 descrevem o *Edge Histogram Descriptor* - Descritor histograma de bordas (EHD) como sendo um descritor de textura. O histograma é construído dividindo as bordas entre cinco categorias. As categorias dizem respeito a orientação da borda: vertical, horizontal, 45 graus, 135 graus e demais bordas.

As bordas são detectadas sobre a imagem em tons de cinza gerada pela média dos 3 canais de cor e utilizando o algoritmo de Sobel (GONZALEZ; WOODS, 2006). Após feita a detecção de bordas sobre a imagem, essa é igualmente dividida em 16 sub-imagens retangulares e não sobrepostas. Cada uma das sub-imagens gera um histograma de cinco posições, para gerar um histograma de uma sub-imagem é necessário dividir novamente cada sub-imagem, cada nova divisão é conhecida como blocos de imagem. O tamanho de cada bloco de imagem depende do tamanho total da imagem em questão. As normas MPEG-7 recomendam que o número de blocos seja próximo a 1100 para que se tenha um boa descrição de textura. O algoritmo de Sobel é aplicado a imagem antes da divisão, mas para facilitar a visualização das divisões em

sub-imagens e os blocos de imagem é mostrado na figura 23 como foi realizado o processo sobre o mapa de pele. Cada bloco de imagem é tratado como sendo uma matriz  $2 \times 2$  de super pixels e a formação dos super pixels gera um valor no histograma de bordas. A figura 24 mostra quais formações da matriz de super pixels pertencem a quais características.



**Figura 23:** Imagem dividida em 16 sub-imagens, cada sub-imagem dividida em blocos e para cada bloco feita a matriz de super pixels.



**Figura 24:** Matrizes de super-pixels: (a) borda horizontal; (b) borda vertical; (c) borda a 45 graus e (d) borda a 135 graus.

Para facilitar a implementação, haja visto que as dimensões de cada imagem são diferentes, os super-pixels foram fixados no tamanho  $4 \times 4$  pixels. Logo, o número de blocos de imagem é variável e em torno de 400 devido ao tamanho médio das imagens do banco CompaqDB.

O histograma para cada sub-imagem representa características locais. Para melhor representar a imagem foram adicionados ao final do vetor de características mais cinco valores que representam o histograma de bordas global da imagem. O vetor **ehd** resultante é dados na seguinte forma:

$$\mathbf{ehd} = [g_1[1], g_1[2], \dots, g_1[5], g_2[1], g_2[2], \dots, g_{16}[5], g_e[1], g_e[2], \dots, g_e[5]],$$

onde  $g_j[i]$ ,  $j \in [1, 16]$  e  $i \in [1, 5]$ , é a quantidade de bordas na posição  $i$  para sub-imagem  $j$ . O descritor global é  $g_e[i]$ .

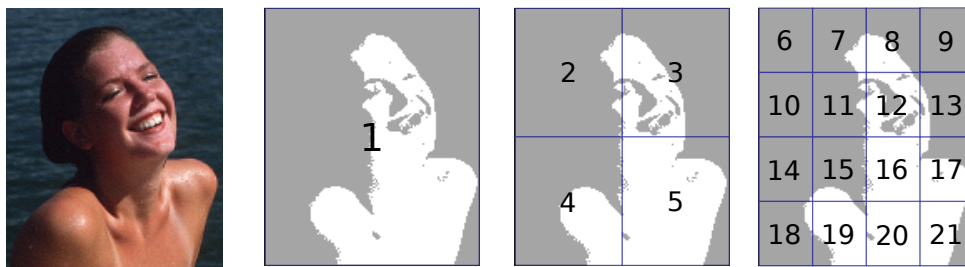
### 5.1.4 Descritor de Compactação

O *Compactness Descriptor* - Descritor de compactação (CD) mede a distribuição espacial de regiões de interesse em uma imagem, nesse caso regiões de pele. O objetivo desse descritor é representar a forma com que as pessoas aparecem na cena. O descritor é calculado sobre o mapa de pele resultante da imagem sem o plano de fundo.

Os valores do descritor são calculados tomando o número de pixels de pele de uma região da imagem e dividindo pelo número total de pixels dessa região. As regiões são em número de 21. A primeira região consiste da imagem inteira, para as demais regiões a imagem é dividida em quatro e 16 regiões iguais e não sobrepostas, como mostra a figura 25. O vetor CD é dado por:

$$\mathbf{cd} = [c[1], c[2], \dots, c[21]],$$

onde  $c[i]$ ,  $i \in [1, 21]$ , é a área relativa ao tamanho da região  $i$ .



**Figura 25: Imagem dividida para o cálculo do descritor de Compactação.**

### 5.1.5 Projeções de Pele

Alguns autores da literatura utilizaram como característica o maior número de pixels de pele conectados. Foi proposta uma abordagem parecida para buscar essa característica, chamada de projeções de pele, que consiste de histogramas para linhas e colunas contendo o número de blocos de pele.

O mapa de pele, após feita a remoção do plano de fundo, é dividido em 100 sub-imagens retangulares e não sobrepostas. Se o número de pixels de pele de uma sub-imagem for maior que 50% do total de pixels, a sub-imagem é considerada como contendo pele. O descritor é calculado somando o número de sub-imagens de pele contidos em cada linha e armazenando o valor em um vetor  $h$ , como mostra a figura 26. O mesmo é feito para cada coluna e o resultado é armazenado no vetor  $v$ . Concatenando os dois vetores obtém-se a projeção de pele na forma de um vetor **proj**:



$$\mathbf{proj} = [h[1], h[2], \dots, h[10], v[1], v[2], \dots, v[10]],$$

onde  $h[i]$  e  $v[i]$ ,  $i \in [1, 10]$ , são a quantidade de blocos de pele na  $i$ -ésima divisão horizontal ou vertical, respectivamente.



Figura 26: Imagem dividida para o cálculo das projeções de pele.

## 5.2 ALGORITMOS DE CLASSIFICAÇÃO

Foram aplicados três métodos de classificação para a detecção de conteúdo ofensivo. Para tal, foi utilizado o software *Waikato Environment for Knowledge Analysis* - Ambiente de Waikato para a análise de conhecimento (WEKA) que possui implementados os mais diversos algoritmos de agrupamento (*clustering*) e classificação existentes na literatura. No presente trabalho foram utilizados os algoritmos de classificação C4.5 (J48), AdaBoost (AdaBoost.M.1) e Bagging.

### 5.2.1 C4.5

O C4.5 é um método de criação de árvores de decisão, que utiliza a máxima entropia de Shannon para separar cada valor dos vetores de características entre as classes (QUINLAN, 1996). O referido algoritmo segue o mesmo princípio para a construção da árvore de decisão que o ID3, desenvolvido por Quinlan (1996).

O modelo é construído para um banco de dados conhecido e previamente classificado (GT),  $S = S_1, S_2, \dots, S_n$ , onde  $S_i$  é o  $i$ -ésimo vetor de características e  $n$  é o número de vetores de características. Cada vetor  $S$  é formado por  $m$  características,  $S_i = X_1, X_2, \dots, X_m$ , onde  $X_j$  é uma característica representada por um número real. A cada vetor  $S$  está associada uma classe, que para esse trabalho são duas: conteúdo ofensivo e conteúdo não-ofensivo.

Para todos os atributos de todos os vetores de características de uma classe é calculada a entropia de Shannon e para o atributo que apresenta o menor valor de entropia é criado um nó

na árvore de decisão. A implementação utilizada nesse trabalho para o classificador C4.5 foi a J48 presente no WEKA.

### 5.2.2 AdaBoost

Abreviação de *Adaptive Boosting* (impulsão adaptativa) o algoritmo AdaBoost utiliza varias instâncias ponderadas de um classificador fraco para gerar um classificador robusto (FREUND; SCHAPIRE, 1997). O classificador fraco é um classificador que sozinho não executa bem a separação entre as classes. Entretanto quando utilizado em conjunto com outros classificadores ou repetidas vezes, classificador robusto, consegue melhorar os resultados de classificação. Para construir o modelo AdaBoost o classificador fraco foi utilizado repetidas vezes sobre o padrão-ouro. Para cada vetor de característica do banco de dados utilizado na construção do modelo é aplicado o classificador em sendo o resultado uma falsa classificação, o peso desse vetor de características é incrementado para forçar uma classificação correta.

Repetidas vezes se aplica o classificador fraco no conjunto de dados, e a cada vez são incrementados os valores dos pesos. Ao final da série de aplicações do classificador fraco tem-se uma função distribuição de probabilidade das quais as amostras do espaço de características são mais prováveis que sejam de outra classe daquela em que foram classificadas.

O classificador fraco utilizado é o de *Regression Tree* - Árvore de regressão (REPTree) (BREIMAN, 1996). Também presente no WEKA, a árvore de regressão é um classificador em forma de árvore de decisão semelhante ao C4.5 utilizando ao invés da entropia a variância para a separação dos nós.

### 5.2.3 Bagging

Bagging, ou *bootstrap aggregating*, é uma técnica para melhorar as taxas de acurácia e estabilidade de um classificador (BREIMAN, 1996). Da estatística *bootstrapping* é um método utilizado para realizar inferências estatísticas. A inferência é feita calculando parâmetros, como a variância, para vários sub-conjuntos das observações. Cada sub-conjunto é formado por reamostragens das observações, buscando padrões de distribuições. Esse método pode se aplicado para reconhecimento de padrões, então, surge o classificador Bagging.

Técnicas de Bagging são usualmente utilizadas com classificadores de árvores de decisão. A partir do banco de dados utilizado são criados vários outros bancos contendo amostras do banco original. Cada novo banco contém um número diferente de amostras que podem se repetir, para enfatizar as características mais marcante ou menosprezar as menos marcantes.

Sobre cada novo banco é construído um classificador. Amostras a serem classificadas são expostas a todos os classificadores fracos e a maioria absoluta dos resultados para todos os classificadores é a classe à qual pertence o novo dado de entrada. O mesmo classificador fraco utilizado pelo AdaBoost foi utilizado para o classificador Bagging.

### 5.3 RESULTADOS E DISCUSSÕES

Foram feitos 14 testes para cada classificador, seis diferentes vetores SCD e quatro composições do vetor de características. Os vetores de características utilizados foram criados concatenando os descritores citados na seção 5.1. As quatro composições são:

1. [ehd, cd, dimensões da imagem];
2. [ehd, cd, dimensões da imagem, projeções de pele];
3. [ehd, scd, cd, dimensões da imagem];
4. [ehd, scd, cd, dimensões da imagem, projeções de pele];

As diferentes variações do vetor de características SCD foram feitas utilizando os espaços de cor RGB, HSI e YCbCr, cada qual com duas escalas de cor: (8:8:8) e (16:4:4). Para o YCbCr utilizou-se a escala (4:4:16), ao invés da (16:4:4), para reforçar a cor vermelha, a qual representa melhor a pele de pessoas em uma cena. Para a análise dos resultados é feita a associação da composição ao número descrito acima. Por exemplo a composição 1 é a [ehd, cd, dimensões da imagem] e assim por diante até a composição 4.

Durante os estudos para a escolha dos classificadores foram realizados experimentos a respeito da segmentação de pele. Para esse estudo utilizou-se o classificador de pele Jones & Rehg para alguns valores de limiar. A segmentação de pele tem influência sobre a detecção de conteúdo ofensivo, mas ela podem ser minimizadas se as taxas de erro estiverem no intervalo de 0,05 até 0,15.

Para o classificador C4.5 foram obtidos os piores resultados. A tabela 9 mostra os valores de predição positiva (VPP) e sensibilidade (Sens.) para o classificador C4.5. Esses valores são calculados para cada classe.

Entre as composições utilizadas para o classificador C4.5 o melhor resultado foi obtido para a composição 1, a tabela 10 mostra quantas imagens utilizadas para teste (80% da seleção CompaqDBContent) foram corretamente classificadas. Aparentemente a classificação de conteúdo

**Tabela 9: Valores de predição positiva e sensibilidade para o classificador C4.5 utilizando todas as composições do vetor de características**

Composição	Espaço de cor	Conteúdo Ofensivo		Conteúdo não-Ofensivo	
		VPP	Sens.	VPP	Sens.
1	sem o SCD	0,736	0,748	0,736	0,723
2	sem o SCD	0,721	0,755	0,734	0,698
3	RGB (8:8:8)	0,702	0,689	0,685	0,699
	RGB (16:4:4)	0,692	0,681	0,676	0,687
	HSI (8:8:8)	0,688	0,673	0,670	0,685
	HSI (16:4:4)	0,696	0,654	0,664	0,705
	YCbCr (8:8:8)	0,699	0,717	0,700	0,682
	YCbCr (4:4:16)	0,727	0,699	0,701	0,729
	4	RGB (8:8:8)	0,721	0,755	0,734
RGB (16:4:4)		0,721	0,755	0,734	0,698
HSI (8:8:8)		0,721	0,755	0,734	0,698
HSI (16:4:4)		0,721	0,755	0,734	0,698
YCbCr (8:8:8)		0,721	0,755	0,734	0,698
YCbCr (4:4:16)		0,721	0,755	0,734	0,698

ofensivo é mais suscetível a descritores geométricos e de textura. Descritores de cor não apresentam melhorias nos valores de VPP e sensibilidade.

**Tabela 10: Matriz de confusão para o melhor resultado do classificador C4.5, composição 1**

		Classificação	
		Conteúdo Ofensivo	Conteúdo não-Ofensivo
Padrão-ouro	Conteúdo Ofensivo	820	276
	Conteúdo não-Ofensivo	294	768

O classificador C4.5 na composição 1 atingiu um total de acerto de 73,6% das imagens e 26,4% de imagens incorretamente classificadas. Diferentemente dos valores relatados na literatura, os resultados obtidos ficaram abaixo dos 90% de acerto. Pelos resultados próximos em termos de VPP e sensibilidade tanto para o C4.5, quanto para o AdaBoost leva a concluir que as características retirada das imagens não são condizentes com as imagens do banco CompaqDB.

Os resultados obtidos utilizando o AdaBoost foram melhores que os obtidos para o C4.5. A tabela 11 mostra os valores VPP e sensibilidade para o classificador AdaBoost.

O melhor resultado encontrado para o classificador AdaBoost foi utilizando a composição 3 utilizando o espaço de cor RGB (8:8:8) para o SCD. A tabela 12 mostra o número de imagens classificadas corretamente e incorretamente para cada uma das classes. Para o classificador AdaBoost utilizando a composição 3, RGB (8:8:8), o número de imagens corretamente classificadas aumenta para 74,2%. Apesar do aumento, as taxas de classificação continuam parecidas umas com as outras. Então, o aumento na quantidade de imagens corretamente classificadas

deve-se ao classificador e não às características utilizadas.

**Tabela 11: Valores de predição positiva e sensibilidade para o classificador AdaBoost utilizando todas as composições do vetor de características**

Composição	Espaço de cor	Conteúdo Ofensivo		Conteúdo não-Ofensivo	
		VPP	Sens.	VPP	Sens.
1	sem o SCD	0,778	0,673	0,704	0,802
2	sem o SCD	0,752	0,728	0,728	0,752
3	RGB (8:8:8)	0,763	0,715	0,724	0,77
	RGB (16:4:4)	0,756	0,725	0,728	0,758
	HSI (8:8:8)	0,747	0,739	0,733	0,741
	HSI (16:4:4)	0,759	0,749	0,745	0,755
	YCbCr (8:8:8)	0,761	0,685	0,705	0,778
	YCbCr (4:4:16)	0,761	0,716	0,724	0,767
4	RGB (8:8:8)	0,752	0,728	0,728	0,752
	RGB (16:4:4)	0,752	0,728	0,728	0,752
	HSI (8:8:8)	0,752	0,728	0,728	0,752
	HSI (16:4:4)	0,752	0,728	0,728	0,752
	YCbCr (8:8:8)	0,752	0,728	0,728	0,752
	YCbCr (4:4:16)	0,752	0,728	0,728	0,752

**Tabela 12: Matriz de confusão para o melhor resultado do classificador AdaBoost, composição 3 utilizando o SCD RGB (8:8:8)**

		Classificação	
		Conteúdo Ofensivo	Conteúdo não-Ofensivo
Padrão-ouro	Conteúdo Ofensivo	784	312
	Conteúdo não-Ofensivo	244	818

Os melhores resultados foram obtidos para o classificador Bagging utilizando a composição 3, YCbCr (4:4:16). Para esse classificador o descritor de cor aumentou um pouco os valores de VPP e sensibilidade, como pode-se observar na tabela 13.

O número de imagens correta e incorretamente classificadas para o classificador Bagging para o melhor caso, são mostradas na tabela 14. O número de imagens corretamente classificadas correspondem a 77,16% das imagens do banco.

## 5.4 DISCUSSÕES

Conforme os resultados obtidos, as características mais marcantes em detecção de conteúdo ofensivo parecem ser características geométricas, pois os resultados para todos os classificadores utilizados no presente trabalho mostram que a cor auxilia pouco na detecção. As projeções de pele também auxiliam pouco e em certos casos, como para o classificador C4.5,

**Tabela 13: Valores de predição positiva e sensibilidade para o classificador Bagging utilizando todas as composições do vetor de varacterísticas**

Composição	Espaço de cor	Conteúdo Ofensivo		Conteúdo não-Ofensivo	
		VPP	Sens.	VPP	Sens.
1	sem o SCD	0,762	0,769	0,760	0,752
2	sem o SCD	0,757	0,782	0,767	0,741
3	RGB (8:8:8)	0,768	0,756	0,752	0,764
	RGB (16:4:4)	0,752	0,804	0,782	0,727
	HSI (8:8:8)	0,744	0,768	0,752	0,727
	HSI (16:4:4)	0,748	0,756	0,745	0,736
	YCbCr (8:8:8)	0,760	0,784	0,769	0,745
	YCbCr (4:4:16)	0,759	0,806	0,786	0,736
	4	RGB (8:8:8)	0,757	0,782	0,767
RGB (16:4:4)		0,757	0,782	0,767	0,741
HSI (8:8:8)		0,757	0,782	0,767	0,741
HSI (16:4:4)		0,757	0,782	0,767	0,741
YCbCr (8:8:8)		0,757	0,782	0,767	0,741
YCbCr (4:4:16)		0,757	0,782	0,767	0,741

**Tabela 14: Matriz de confusão para o melhor resultado do classificador Bagging, composição 3 utilizando o SCD YCbCr (4:4:16)**

		Classificação	
		Conteúdo Ofensivo	Conteúdo não-Ofensivo
Padrão-ouro	Conteúdo Ofensivo	883	213
	Conteúdo não-Ofensivo	280	782

atrapalham a classificação. Mesmo assim, a informação contida nas projeções de pele não representa bem ações de pessoas nas imagens.

O melhor classificador foi o Bagging com a composição 3, ou seja, sem projeções de pele e com descritor de cor RGB (8:8:8). Dar preferência à cor vermelha no descritor de cor não obteve melhoria nos resultados. Apesar da cor não trazer melhorias significativas nos resultados, os melhores foram encontrados para descritores de cor igualmente escalados e utilizando o espaço de cor RGB.

Os resultados encontrados na presente pesquisa apresentam valores abaixo dos encontrados na literatura. O melhor resultado de classificação foi de 72% para imagens de conteúdo ofensivo e 77% para imagens sem conteúdo ofensivo. Na literatura os resultados de classificação para conteúdo ofensivo e não-ofensivo são respectivamente de: 99,54% e 83,49% para Shih, Lee e Yang (2007); 79,3% e 88,7% para Fleck, Forsyth e Bregler (1996); e 85,8% e 92,5% para Jones e Rehg (1999). Especula-se que a diferença entre valores se deve principalmente ao vetor de características e ao banco de dados utilizados em cada um dos trabalhos citados. Um modelo

ótimo de detecção de conteúdo ofensivo ainda não foi encontrado.

Como trabalho futuro, a presente pesquisa de detecção de conteúdo ofensivo deve se aprofundar no estudo de descritores geométricos e em descritores que indiquem as quantidades de blocos de pele conectados. Descritores de cor podem ser úteis em outras funções na detecção de conteúdo ofensivo, tais como avaliação do plano de fundo ou discriminação através de cores incomuns nas cenas.

## 6 CONCLUSÕES

### 6.1 SEGMENTAÇÃO DE PELE

A cor é uma boa característica para a segmentação de pixels de pele. Sem a necessidade de técnicas complexas e muitos recursos computacionais, um simples vetor de dois ou três elementos pode segmentar pixels de pele. Para o classificador Jones & Rehg o espaço de cor CbCr apresenta uma classificação equivalente à obtida com espaços de três dimensões. Para o mesmo classificador, espaços de três dimensões apresentaram resultados semelhantes, mostrando que não há perda de informação nas transformações de espaços de cor.

A distribuição de probabilidade da cor de pele pode ser aproximada para uma Gaussiana para os espaços YCbCr e CbCr, mas não para os demais. Entretanto, não se pode afirmar que a distribuição densidade de probabilidade de cor de pele não seja paramétrica, pois o número de observações apesar de muito elevado não abrange todos os casos possíveis, o que justificaria descartar a ideia de uma distribuição paramétrica.

Para aumentar as taxas de classificação de pele é necessário que seja agregada mais informação à cor de um pixel. Segundo os autores que defendem a segmentação de pele através de regiões e não apenas pixels (JEDYNAK; ZHENG; DAOUDI, 2003), a pele em imagens digitais aparece conectada em grandes blocos. Isso foi verificado ao se estudar os trabalhos de detecção de conteúdo ofensivo, onde a remoção do plano de fundo é feita sobre os maiores blocos de pele conectados. Uma abordagem para aumentar as taxas de classificação está em utilizar juntamente com a cor do pixel outras informações, como cor dos pixels da vizinhança e textura de pele.

### 6.2 ILUMINAÇÃO AMBIENTE E BANCO DE DADOS CONTROLADO

A iluminação ambiente influencia, e muito, a segmentação de pixels de pele. O controle de equilíbrio de branco do dispositivo de aquisição é outro fator que tem influência na cor de pele. Como os algoritmos no presente trabalho utilizam a classificação pixel-a-pixel, não há a informação necessária para corrigir a cor antes da classificação.



O banco de dados controlado foi utilizado nos estudos sobre a iluminação ambiente e taxas de classificação. Os resultados encontrados não poderiam ter sido encontrados sem as imagens do banco skinDB, apesar do mesmo apresentar relativamente menos imagens que o CompaqDB e de necessitar de mais variáveis a serem controladas.

### 6.3 DETECÇÃO DE CONTEÚDO OFENSIVO

Foram estudadas as mais populares metodologias para a detecção de conteúdo ofensivo. O melhor conjunto de características citado na literatura foi implementado e testado para alguns algoritmos de classificação, mas os resultados não foram similares aos relatados para nenhum algoritmo de classificação. Especula-se que tais divergências se devem, principalmente, ao banco de dados que foi utilizado, indicando a possibilidade de que a detecção seja altamente influenciada pelo banco que está sendo estudado.

Para comparar resultados de uma forma não polarizada é necessário criar um padrão de conteúdo ofensivo. Esse padrão pode ser um banco de imagens próprio para que vários pesquisadores possam usar de referência e com um padrão-ouro que tenha sido aprovado por vários observadores. Dessa forma, seria minimizada a chance de obtenção de resultados polarizados inadvertidamente.

O fato dos três classificadores resultarem em desempenhos similares entre si leva a concluir que as características não são as melhores para descrever o conteúdo do banco compaqDB. Os trabalhos encontrados na literatura apresentam, cada qual, diferentes vetores de características, métodos de classificação e banco de dados. Sem um banco de dados padrão ou um estudo de caso mais específico, conclusões sobre a detecção de conteúdo ofensivo devem ser feitas com cautela sob pena de se fazer uma generalização incorreta.

A comparação entre o método proposto no presente trabalho com o trabalho de Jones e Rehg (1999) pode ser feita pois foi utilizado o mesmo banco de dados em ambos os casos. As características utilizadas por Jones e Rehg (1999) são melhores para demonstrar o conteúdo ofensivo no banco CompaqDB especificamente. Apesar do banco CompaqDB ser um padrão para detecção de conteúdo ofensivo, contém certas peculiaridades que são tendenciosas ao conjunto de características utilizado pelos autores Jones e Rehg (1999), tais proporção de blocos de pele em relação ao tamanho da imagem e pessoas em primeiro plano.

## 6.4 TRABALHOS FUTUROS

### 6.4.1 Segmentação de Pele

Para a sequência do estudo de segmentação de pele propõe-se a busca de um espaço de cor próprio para a segmentação de pele, haja vista que a cor de pele é um subconjunto específico de cores. Ainda, deve-se buscar uma transformação que não seja dependente de um banco de dados para treino e que apresente as mesmas taxas TPR e FPR para modelos diferentes. Isso é possível se forem analisadas as propriedades físicas da luz utilizada como iluminante e quantizadas devidamente em um espaço de cor próprio.

O efeito da iluminação pode ser reduzido obtendo informações sobre a distribuição de cores da imagem e através dessa informação fazer a correção das cores para que seja mantido o padrão da cor de pele. Foram testados, preliminarmente, algoritmos de correção da constância de cor (FINLAYSON; SCHIELE; CROWLEY, 1998) para o classificador Jones & Rehg, mas ainda sem sucesso. Logo, mais pesquisas devem ser feitas para comprovar a efetividade da correção de cor.

Propõe-se buscar outras características, além da cor para a segmentação de pele. A textura, após alguns estudos preliminares, mostrou-se capaz de detectar certas regiões de pele em imagens em tons de cinza. Um trabalho possível de ser feito é de unir a segmentação através de cor com a segmentação por textura na tentativa de aumentar as taxas TPR, reduzindo as taxas FPR.

### 6.4.2 Iluminação Ambiente e Banco de Dados Controlado

O banco de dados controlado skinDB ainda apresenta poucas imagens e precisa ter mais variáveis controladas. Entre as variáveis estão: aumentar o número de imagens, fazer controle das cores do plano de fundo, fotografar utilizando iluminações de várias cores de composição conhecida e conhecer a distância da câmera para a foto, além de desabilitar controle automático do equilíbrio de branco do dispositivo de captura.

### 6.4.3 Detecção de Conteúdo Ofensivo

Para se realizar uma comparação mais apropriada, deve-se fazer uma pesquisa que compare métodos da literatura utilizando o mesmo banco de dados. Com o mesmo banco de dados, diferentes classificadores e diferentes conjuntos de características podem ser comparados através dos valores de VPP e sensibilidade. Nesse cenário seria possível indicar qual classificador e qual conjunto de características expressa melhor o conteúdo ofensivo.

Com o objetivo de aumentar as taxas de acerto e superar as encontradas na literatura, novas características devem ser pesquisadas. Realizar a detecção de faces e reconhecimento de feições podem representar bem características para a detecção de conteúdo ofensivo.

## REFERÊNCIAS

- ALBIOL, A.; TORRES, L.; DELP, E. J. Optimum color spaces for skin detection. In: **International Conference on Image Processing**. [S.l.: s.n.], 2001. v. 1, p. 122–124.
- BREIMAN, L. Bagging predictors. **Machine Learning**, Kluwer Academic Publishers, Hingham, MA, USA, v. 24, p. 123–140, August 1996. ISSN 0885-6125.
- CAETANO, T.; BARONE, D. A probabilistic model for the human skin color. In: **Proceedings of the 11th International Conference on Image Analysis and Processing**. [S.l.: s.n.], 2001. p. 279–283.
- FAWCETT, T. **ROC Graphs: Notes and Practical Considerations for Researchers**. 2004. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.9777>>. Acesso em: 8 de maio de 2009.
- FINLAYSON, G. D.; SCHIELE, B.; CROWLEY, J. L. Comprehensive colour image normalization. In: **ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume I**. London, UK: Springer-Verlag, 1998. p. 475–490. ISBN 3-540-64569-1.
- FLECK, M.; FORSYTH, D.; BREGLER, C. Finding naked people. In: **European Conference on Computer Vision**. [S.l.]: Springer-Verlag, 1996. v. 2, p. 592–602.
- FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. **Journal of Computer and System Science**, Academic Press, Inc., Orlando, FL, USA, v. 55, n. 1, p. 119–139, 1997. ISSN 0022-0000.
- GARCIA, C.; TZIRITAS, G. Face detection using quantized skin color regions merging and wavelet packet analysis. In: **IEEE Transactions on Multimedia**. [S.l.: s.n.], 1999. p. 264–277. ISSN 1520-9210.
- GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X.
- HAMMAMI, M.; CHAHIR, Y.; CHEN, L. Webguard: A web filtering engine combining textual, structural, and visual content-based analysis. **IEEE Transactions on Knowledge and Data Engineering**, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 18, n. 2, p. 272–284, 2006. ISSN 1041-4347.
- HU, W. et al. Recognition of pornographic web pages by classifying texts and images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, IEEE Computer Society, Washington, DC, USA, v. 29, p. 1019–1034, June 2007. ISSN 0162-8828.
- JAYARAM, S. et al. Effect of colorspace transformation, the illuminance component and color modeling on skin detection. In: . [S.l.: s.n.], 2004. p. II: 813–818.

- JEDYNAK, B.; ZHENG, H.; DAOUDI, M. Maximum entropy models for skin detection. In: **Energy Minimization Methods in Computer Vision and Pattern Recognition**. [S.l.: s.n.], 2003. p. 180–193.
- JONES, M.; REHG, J. Statistical color models with application to skin detection. In: **Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition**. [S.l.: s.n.], 1999. v. 1, p. 274–280.
- JONES, M.; REHG, J. Statistical color models with application to skin detection. **International Journal of Computer Vision**, v. 46, n. 1, p. 81–96, January 2002.
- KAKUMANU, P.; MAKROGIANNIS, S.; BOURBAKIS, N. A survey of skin-color modeling and detection methods. **Pattern Recognition**, v. 40, n. 3, p. 1106–1122, March 2007.
- KOVAC, J.; PEER, P.; SOLINA, F. Human skin color clustering for face detection. In: **Proceedingd of the EUROCON 2003**. [S.l.: s.n.], 2003. v. 2, p. 144–148 vol.2.
- KRUPPA, H.; BAUER, M.; SCHIELE, B. Skin patch detection in real-world images. In: **Annual Symposium for Pattern Recognition of the DAGM**. [S.l.]: Springer, 2002. p. 109–117.
- KUIASKI, D. et al. Estrutura para anotação de imagens para a avaliação de algoritmos de detecção de pele humana. In: **Anais do IV Congresso Brasileiro de Engenharia Biomédica**. Salvador, Bahia, Brasil: [s.n.], 2008.
- KUIASKI, D. et al. A study of the effect of illumination conditions and color spaces on skin segmentation. In: **Proceedings of the XXII Brazilian Symposium on Computer Graphics and Image Processing**. Rio de Janeiro, RJ, Brazil: [s.n.], 2009.
- LEE, L.-H.; LUH, C.-J. Generation of pornographic blacklist and its incremental update using an inverse chi-square based method. **Information Processing and Management**, Pergamon Press, Inc., Tarrytown, NY, USA, v. 44, n. 5, p. 1698–1706, 2008. ISSN 0306-4573.
- MANJUNATH, B. S. **Introduction to MPEG-7, Multimedia Content Description Interface**. [S.l.]: John Wiley and Sons, Ltd., 2002.
- PALMER, S. **Vision Science: Photons to Phenomenology**. [S.l.]: MIT Press, 1999.
- PHUNG, S. L.; BOUZERDOUM, A.; CHAI, D. Skin segmentation using color pixel classification: Analysis and comparison. **IEEE Transactions on Pattern Analysis Machine Intelligence**, IEEE Computer Society, Washington, DC, USA, v. 27, n. 1, p. 148–154, January 2005. ISSN 0162-8828.
- PHUNG, S. L.; CHAI, D.; BOUZERDOUM, A. Skin colour based face detection. In: **Proceedings Intelligent Information Systems Conference The Seventh Australian and New Zealand 2001**. [S.l.: s.n.], 2001. p. 171–176.
- POYNTON, C. **Frequently Asked Questions about Color**. 1995. Disponível em: <<http://www.poynton.com/ColorFAQ.html>>. Acesso em: 23 de novembro de 2008.
- QUINLAN, J. R. Improved use of continuous attributes in c4.5. **Journal of Artificial Intelligence Research**, AI Access Foundation, v. 4, p. 77–90, March 1996. ISSN 1076-9757.

SHIH, J.-L.; LEE, C.-H.; YANG, C.-S. An adult image identification system employing image retrieval technique. **Pattern Recognition Letters**, Elsevier Science Inc., New York, NY, USA, v. 28, n. 16, p. 2367–2374, 2007. ISSN 0167-8655.

SHIN, M.; CHANG, K.; TSAP, L. Does colorspace transformation make any difference on skin detection? In: **IEEE Workshop on Applications of Computer Vision**. [S.l.: s.n.], 2002. p. 275–279.

SOLAR, J. R. del et al. Characterizing objectionable image content (pornography and nude images) of specific web segments: Chile as a case study. In: **LA-WEB '05: Proceedings of the Third Latin American Web Congress**. Washington, DC, USA: IEEE Computer Society, 2005. ISBN 0-7695-2471-0.

STÖRRING, M. **Computer Vision and Human Skin Colour**. Tese (Doutorado) — Faculty of Engineering and Science, Aalborg University, Denmark, 2004.

TERRILLON, J.-C. et al. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In: **Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000**. Washington, DC, USA: IEEE Computer Society, 2000. p. 54–. ISBN 0-7695-0580-5.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. [S.l.]: Elsevier, 2006.

VEZHNEVETS V.; SAZONOV, V. A. A. A survey on pixel-based skin color detection techniques. In: **Proceedings of GRAPHICON 2003**. [S.l.]: GRAPHICON 2003, 2003. p. 85–92.

YANG, M.-H.; AHUJA, N. Detecting human faces in color images. In: **Proceedings of the International Conference on Image Processing ICIP 98**. [S.l.: s.n.], 1998. v. 1, p. 127–130.