

UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Análise de Características para Detecção de Nudez em Imagens

Clayton André Maia dos Santos

Manaus - Amazonas
Março de 2012

Clayton André Maia dos Santos

Análise de Características para Detecção de Nudez em Imagens

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Informática.

Área de concentração: Redes de Computadores.

Orientador: Prof. Dr. Eduardo James Pereira Souto
Co-Orientadora: Prof^a. Dra. Eulanda Miranda dos Santos

© 2012, Clayton André Maia dos Santos.

Todos os direitos reservados

Ficha Catalográfica
(Catalogação realizada pela Biblioteca Central da UFAM)

S237a Santos, Clayton André Maia dos

Análise de características para detecção de nudez em imagens/Clayton André Maia dos Santos. – Manaus, AM: UFAM, 2012.

87f.: il. color.

Dissertação (Mestrado em Informática) – Universidade Federal do Amazonas, 2012.

Orientador: Prof Dr. Eduardo James Pereira Souto

Co- Orientador: Prof^a Dr^a Eulanda Miranda dos Santos

1. Informática- Zoneamento de Imagens 2. Informática- Detecção de nudez- Imagens 3. Inteligência Computacional I. Souto, Eduardo James Pereira. (Orient.) II.Santos, Eulanda Miranda dos (Co-Orient.) III. Universidade Federal do Amazonas IV. Título

CDU (1997) 004.383.5(043.3)



PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA



FOLHA DE APROVAÇÃO

“Análise de Características para Detecção de Nudez em Imagens”

CLAYTON ANDRÉ MAIA DOS SANTOS

Dissertação defendida e aprovada pela banca examinadora
constituída pelos Professores:

PROF. EDUARDO JAMES PEREIRA SOUTO – PRESIDENTE

PROFa. ÂNGELA VILHENA DIAS – MEMBRO

PROFa. EULANDA MIRANDA DOS SANTOS – MEMBRO

PROF. JOÃO MARCOS BASTOS CAVALCANTI – MEMBRO

UNIVERSIDADE FEDERAL DO AMAZONAS
DCC/ICE/UFAM
CONFERE COM O ORIGINAL
Assinatura:
Matrícula SIAPE: 2011208

Manaus, 30 de março de 2012.

À minha família.

Agradecimentos

A Deus, pela vida, saúde, proteção, iluminar meus passos e oportunidade de vivenciar este momento.

Aos meus pais, por todo amor, atenção e pela excelente educação que me foi dada e, que apesar de distantes sempre me apoiaram e não me deixaram sentir sozinho, seja através de orações ou palavras de incentivo.

Aos meus irmãos, Cleto Júnior, Clívia e Marcos, que sempre me apoiaram.

Ao meu irmão Alex Santos, por acreditar em minha capacidade e oferecer toda a atenção ao cumprimento do alcance de mais um objetivo em minha vida.

A minha namorada, Brenda Nunes, pelo companheirismo, confiança, amor, carinho e, principalmente, ter acompanhado, mesmo que distante todos os momentos difíceis que passei.

Aos meus orientadores, Eduardo Souto e Eulanda Miranda dos Santos, pela oportunidade, total incentivo, paciência, grandes conhecimentos compartilhados e, principalmente, a confiança na capacidade de realização deste trabalho.

Aos amigos Juan Collona e Polianny pela amizade, compreensão e aprendizado compartilhado. Aos colegas Professor JB, Viviane, João, Davi, Joacir, Mirlem, Janainny e muitos outros pela companhia, solidariedade, ajuda e momentos de diversão, fundamentais para meu fortalecimento nesta caminhada.

Aos meus vizinhos Jorge Fernando, Camilla e Neiva, pela imensa solidariedade e carinho dispensados.

Ao PPGI, pela oportunidade e a CAPES pelo apoio financeiro.

A todos aqueles que ajudaram de alguma forma na realização deste trabalho, o meu mais profundo agradecimento.

“Para tudo o que almejamos na vida sempre será exigido, no mínimo, um sacrifício”.

Clayton Santos

Resumo

Com a popularização do acesso a Internet, instituições e pais têm encontrado sérios problemas para evitar o acesso de funcionários e crianças a conteúdos impróprios como páginas pornográficas. Na *Web*, este tipo de conteúdo pode estar disponibilizado em forma de vídeos, sons, texto e, principalmente em forma de imagens. Como a maior parte do conteúdo impróprio está disponível através de imagens, faz-se necessária a utilização de mecanismos que possibilitem analisar o conteúdo da imagem para combater este tipo de abuso. Nesse contexto, a detecção de nudez em imagens é normalmente uma etapa importante.

Diversas abordagens aplicam detecção de pele como passo fundamental para a detecção de nudez. Esta tarefa não é trivial, uma vez que há a necessidade de uso de filtros de pele robustos a variações de tonalidades ocasionadas por luminosidade. Adicionalmente, é utilizada uma combinação de características baseadas em cor, textura e formas, que podem ocasionar um aumento indesejado na complexidade e no tempo de processamento dos algoritmos de detecção. Apesar dessa desvantagem, em muitos trabalhos disponíveis na literatura, não é realizada uma análise da relevância das características envolvidas no processo de classificação das imagens.

O objetivo desta dissertação é investigar as principais características para descrição de imagens de nudez e selecionar as que obtiverem maior relevância em função da precisão do classificador. A análise das características é feita através de um conjunto de séries de experimentos que representam diferentes cenários de comparação. Primeiramente, são feitas comparações entre características extraídas sem aplicação de filtro de pele, denominadas propriedades globais. Em seguida, características extraídas a partir da aplicação do filtro de pele são também comparadas. Essas características são denominadas propriedades locais. Em uma terceira série de experimentos, um algoritmo de zoneamento é utilizado para que seja analisado o impacto das características, tanto locais como globais, em cada zona da imagem. Em todas as séries de experimentos, cada característica é analisada tanto de forma isolada, quanto em subconjuntos, para que seja determinado o melhor compromisso entre o conjunto de características e o desempenho do classificador. Para isso, é utilizada uma arquitetura denominada ANDImage (*Architecture for Nude Detection in Image*), que permite a inclusão e exclusão desses diferentes módulos.

Palavras-chave: imagens, análise de características, classificação, detecção de nudez.

Abstract

The popularization of Internet access has lead institutions and parents to face serious problems on preventing employees, as well as children, to have access to inappropriate content, such as pornographic pages. This kind of content is available in different forms, including videos, sounds, text, and especially images, on the Web. Since most of this inappropriate content is provided as images, it is necessary to employ strategies which allow the analysis of image content in order to control access to inappropriate content. In this context, nudity detection in images plays an important role.

Several approaches apply skin detection as a key step toward nudity detection. Skin detection is a difficult task due to the fact that it is necessary to use skin filters robust to shade variations caused by light. In addition, these methods employ a combination of features based on color, texture and shape, which may increase the complexity and time processing of detection algorithms. In despite of this drawback, feature analysis, or selection, is not carried out in most of the work available in the literature,

The objective of this work is to investigate the features most frequently used in the literature for the description of nude images, as well as to select the most relevant subset of features taking into account classification accuracy. The feature analysis is carried out through three series of experiments focusing on investigating different scenarios of comparison. In the first series, we compare features extracted without applying skin filter, called global evidence in this work. In the second series, features extracted after skin filter are also compared. These features are called local evidence. Finally, in the third series of experiments, a zoning algorithm is used in order to allow us to analyze the impact of both local and global features in each area of the image. In all series of experiments, each feature is analyzed individually and all subsets of features are tested so as to determinate the best tradeoff between feature set and classification accuracy. In addition, an architecture called ANDImage (Architecture for Nude Detection in Image) is proposed. ANDImage allows that different modules may be used in order to provide the possibility of dealing with different scenarios of features comparison.

Keywords: images, feature analysis, classification, nudity detection.

Sumário

1. Introdução.....	1
1.1. Detecção de Nudez.....	3
1.2. Motivação.....	4
1.3. Justificativa	6
1.4. Objetivos	7
1.5. Organização do Documento.....	7
2. Conceitos Básicos e Trabalhos Relacionados.....	9
2.1. Imagens Digitais	9
2.1.1. Espaço de Cores	12
2.1.2. Propriedades em Imagens	16
2.1.2.1. Propriedades Globais.....	17
2.1.2.2. Propriedades Locais.....	21
2.2. Aprendizagem de Máquinas.....	23
2.2.1. SVM.....	24
2.3. Métricas de Avaliação	28
2.4. Trabalhos Relacionados	29
3. A Arquitetura ANDImage.....	35
3.1. Componente Normalizador.....	36
3.2. Filtro de Pele.....	37
3.3. Zoneamento em Imagens	37
3.4. Extração de Características.....	39
3.4.1. Extração de Características de Cor	39

3.4.2. Extração de Características de Textura	40
3.4.3. Extração de Características de Formas	41
3.4.1. Extração de Características Baseadas no Filtro de Pele.....	42
3.5. Classificação	43
4. Experimentos e Análise dos Resultados	45
4.1. Protocolo Experimental.....	46
4.1.1. Base de Dados.....	46
4.1.2. Desenvolvimento da Arquitetura ANDImage.....	48
4.1.3. Ajuste de Parâmetros de SVM	51
4.2. Resultados	53
4.2.1. Análise das Características Globais	53
4.2.2. Análise das Características Locais	58
4.2.3. Combinação das Abordagens Global e Local através do Zoneamento em Imagens	59
4.2.4. Análise de Tempo de Execução.....	65
5. Conclusões e Trabalhos Futuros	67
5.1. Trabalhos Futuros.....	69
5.2. Publicações obtidas	69
Referências.....	70

Lista de Figuras

2.1. Representação de uma imagem digital.....	9
2.2. Resultado de Segmentação: (a) imagem original b) segmentação através de detecção de borda.....	11
2.3. Imagens que demonstram o resultado da aplicação do zoneamento, onde (a) representa o zoneamento em uma imagem de nudez com orientação vertical, (b) o zoneamento em imagem de nudez com orientação horizontal	11
2.4. Espaço de cor RGB e cores produzidas a partir de cores primárias	13
2.5. Espaço de cor <i>HSV</i>	14
2.6. Espaço de cor <i>CIELab</i>	16
2.7. Terminologia das propriedades em imagens	17
2.8. Construção da matriz vizinhança	19
2.9. Exemplo do processo de extração de características do EHD, onde (a) corresponde a uma imagem binarizada após a aplicação do operador Sobel e; (b) uma imagem segmentada e particionada em blocos fixos de 4 x 4	20
2.10. Máscara de processamento 3x3.....	21
2.11. Extração de características locais.....	22
2.12. Exemplificação de características extraídas a partir da borda, (a) excentricidade e; (b) retangularidade	23
2.13. Modelo Genérico de aprendizagem de máquina.....	24
2.14. Hiperplano de separação ótima para um problema com duas classes.....	26
2.15. Mapeamento do espaço de entrada via função <i>kernel</i>	26
3.1. Arquitetura geral para sistemas de detecção de nudez.....	35
3.2. Visão Geral da Arquitetura ANDImage	36
3.3. Máscaras para extração de formas: (a) linhas 135°; (b) linhas horizontais; (c) linhas verticais; (d) linhas 45°; (e) pontos	42

3.4. Um classificador recebe o vetor de características como entrada e produz uma classe na saída.....	43
4.1. Exemplos de imagens de nudez	47
4.2. Exemplos de imagens que não contém nudez.....	47
4.3. Divisão da base de dados usada nos experimentos	48
4.4. Exemplos de amostras de pele utilizadas na obtenção dos valores de cromaticidade	49
4.5. Imagens que demonstram o resultado do filtro de pele, onde (a) representa uma imagem original de nudez, (b) o resultado do filtro de pele na imagem de nudez, (c) imagem de pessoas com roupa, (d) o resultado do filtro de pele para pessoas com roupa, (e) imagem original que não contém pele e, (f) o resultado do filtro na imagem que não contém pele.....	50
4.6. Imagens que demonstram o resultado da aplicação do zoneamento, em que (a) representa o zoneamento em uma imagem de nudez com orientação vertical, (b) o resultado do zoneamento em imagem de nudez na vertical após a aplicação do filtro de pele, (c) o zoneamento em imagem de nudez com orientação na horizontal, (d) o resultado do zoneamento em imagens de nudez na horizontal após o filtro de pele	51
4.7. Imagens que demonstram o resultado da aplicação do zoneamento, em que (a) representa o zoneamento em imagens de face, (b) o resultado do zoneamento em imagem de face após a aplicação do filtro de pele, (c) o zoneamento em imagens de nudez, (d) o resultado do zoneamento em imagens de nudez após o filtro de pele	62

Lista de Tabelas

2.1. Funções de <i>kernel</i> mais utilizadas com SVM	27
2.2. Exemplificação de uma matriz de confusão	28
2.3. Sumarização dos modelos de detecção de nudez.....	34
4.1. Treinamento e ajuste do parâmetro de penalidade (C) em SVM para análise da quantidade de vetores de suporte e influência desse parâmetro na precisão do classificador	51
4.2. Treinamento e ajuste do parâmetro γ em SVM para análise da quantidade de vetores de suporte e influência de tais parâmetros na precisão do classificador.....	52
4.3. Resultados obtidos na validação cruzada durante o treinamento.....	53
4.4. Resultados para classificação de imagens de nudez usando diferentes quantizações de cor.....	54
4.5. Resultados obtidos por SVM na base de teste com as características de cor. As características foram extraídas sem a aplicação do módulo de zoneamento em imagens.....	55
4.6. Análise de características de textura. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, sem a aplicação do zoneamento em imagens.....	56
4.7. Análise de características de formas. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, sem a aplicação do zoneamento em imagens.....	57
4.8. Análise de todas as características globais. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, sem a aplicação do zoneamento em imagens.....	57
4.9. Análise das características locais. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, sem a aplicação do zoneamento em imagens.....	59
4.10. Análise de características de cor. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, com a aplicação do zoneamento em imagens	59
4.11. Análise de características de textura. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, com a aplicação do zoneamento em imagens	60
4.12. Análise de características de formas. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, com a aplicação do zoneamento em imagens	61

4.13. Análise das características locais. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, com a aplicação do zoneamento em imagens	62
4.14. Análise da combinação de todas as características e seleção do melhor conjunto para descrição de imagens de nudez através do zoneamento em imagens.....	63
4.15. Análise da eficiência da arquitetura	65

Capítulo 1

Introdução

A massificação de dispositivos pessoais como câmeras digitais, celulares e PDAs (*Personal Digital Assistants*) tem contribuído para o aumento de conteúdo multimídia na Internet. Com isso os usuários da Internet têm facilidade para inserção de tipos de conteúdos variados como a publicação de texto e imagens em *blogs* e redes sociais como *Blogspot*¹ e *Facebook*², fotos em álbuns virtuais como *Fotolog*³ ou no *Flickr*⁴, arquivos de vídeos e de músicas em sites de compartilhamento como *YouTube*⁵ e *MySpace*⁶, e conteúdos diversificados em redes *peer-to-peer* (P2P) como *bittorrent*⁷ e *emule*⁸.

Logo, a *Word Wide Web* é considerada um enorme centro de dados, permitindo que seus usuários no mundo todo possam compartilhar e trocar informações. O sucesso da Internet se dá pela liberdade de expressão assegurada pela falta de controle das informações disponibilizadas, fato que contribui para o aumento dos benefícios proporcionados por páginas com conteúdo lícito. Por outro lado, essa falta de controle permite também que muito conteúdo considerado ofensivo na *Web*, como vídeos e imagens contendo nudez, possa ser involuntariamente acessado por crianças, ou indevidamente armazenado em instituições onde o acesso a esse tipo de conteúdo não é autorizado [Kherfi *et al.* 2004]. Outro fator importante é a segurança. Em relatório recente, a Symantec [2010] mostra que existe uma relação direta entre pornografia e a difusão de *malware*. De acordo com esse relatório, do total de ataques provenientes da Internet, 49% são feitos por meio de sites pornográficos.

O conteúdo ofensivo, neste caso, está relacionado à nudez humana ou à pornografia, e está presente na *Web* em diferentes formatos e mídias. Este tipo de conteúdo pode ser visualizado

¹ <http://www.blogspot.com>
² <http://www.facebook.com>
³ <http://www.fotolog.com>
⁴ <http://www.flickr.com>
⁵ <http://www.youtube.com>
⁶ <http://www.myspace.com>
⁷ <http://www.bittorrent.com>
⁸ <http://www.emule.com>

na forma de texto. Porém, uma pessoa precisa ler o texto para conseguir observar o conteúdo ofensivo. Também está disponível na forma de vídeos e animações, cujo acesso depende de visualizadores específicos [Belém 2006]. Por fim, o conteúdo ofensivo também está disponível na forma de imagens, mídia que tende a ser a forma mais usada devido ao fato de atualmente todos os navegadores *Web* apresentarem a capacidade de exibição de imagens e, à facilidade de acesso e de visualização de imagens.

Uma forma de evitar o acesso a conteúdo ofensivo é a utilização de filtros *Web*. A maioria dos métodos de filtragem pode ser classificada em: reconhecimento baseado em texto, reconhecimento baseado em imagens e, reconhecimento baseado na combinação de texto e imagem.

Grande parte dos métodos do primeiro grupo baseia-se na comparação de URLs e endereços IPs. Porém, como os filtros baseados em texto possuem uma limitação em atualizar listas, faz-se necessário o desenvolvimento de métodos que busquem analisar imagens, uma vez que a Internet é muito dinâmica e os sites estão sempre mudando.

A detecção de imagens com conteúdo ofensivo é considerada por muitos pesquisadores como um problema de classificação [Belém 2006], cujo principal obstáculo está relacionado com o fato das imagens estarem inseridas em páginas, onde, além de informações que podem ser extraídas da própria imagem, também há informações textuais. Como consequência, alguns métodos de filtragem utilizam as duas fontes de informação. Entretanto, elevadas taxas de detecção podem ser obtidas com a utilização de informações extraídas somente de imagens [Jeong *et al.* 2004].

A detecção de nudez em imagens é considerada um passo importante no sentido de limitar a proliferação de conteúdo indesejável, como a pornografia. Na literatura, diferentes métodos para filtragem de imagens contendo nudez têm sido considerados, tais como os trabalhos de Fleck *et al.* [1996], Wang *et al.* [1998], Jones & Rehg [2002], Wang *et al.* [2009], entre outros. Neste trabalho, será apresentada uma abordagem de detecção de nudez que utiliza informações extraídas unicamente de imagens, através do uso de técnicas de processamento digital de imagens.

1.1 Detecção de Nudez

Segundo Hu *et al.* [2009], um dos fatores que dificulta a detecção de nudez é própria definição do termo nudez, que geralmente é confundido com pornografia. A pornografia é representada por um determinado grupo de imagens que contém nudez, sem a cobertura de genitálias, em posições totalmente diferentes das habituais ou, contendo sexo explícito. No entanto, quando o assunto é a *Web*, há uma divergência ainda maior sobre as definições dos termos nudez e pornografia, pois uma imagem de nudez artística retirada em uma praia, mesmo com a cobertura de genitálias, pode ser considerada pornográfica por despertar o desejo sexual, tanto de crianças, jovens, adultos, ou por atrair a atenção de um determinado grupo de funcionários de uma instituição. Neste trabalho, o termo nudez é definido como o estado em que a pessoa se encontra totalmente descoberta de roupas ou vestindo pouca roupa, com exposição de alguma parte íntima. Portanto, neste contexto, a imagem de uma criança em uma praia é considerada imagem de nudez.

Outro fator que dificulta o processo de detecção de nudez está relacionado à detecção de pele, uma vez que a presença de muita região de pele é um forte indicativo de nudez em uma imagem digital. A maioria das pesquisas em detecção de pele e nudez está centrada na análise do espectro visual da imagem, que é uma tarefa muito difícil devido à diversidade de tonalidades de cores de pele, tornando a imagem sensível a diversos fatores como, por exemplo:

- *Iluminação*: Uma mudança na distribuição da fonte de luz e no nível de iluminação (*indoor*, *outdoor*, sombras, luzes não-brancas) produz uma mudança na cor da pele na imagem (problema conhecido como constância da cor). A variação de iluminação é o problema mais importante entre os sistemas atuais de detecção de pele [kakumanu *et al.* 2007]. Este problema prejudica seriamente o desempenho dos sistemas;
- *Características da câmera*: Mesmo sob a mesma iluminação, a distribuição da cor da pele para a mesma pessoa difere de uma câmera para outra, dependendo das características do sensor da câmera. As cores reproduzidas por uma câmera dependem da refletância, prevalecendo as condições de iluminação e as sensibilidades do sensor da câmera;
- *Etnia*: A cor da pele também varia em função de diferentes grupos étnicos. Por exemplo, a cor da pele de pessoas pertencentes a grupos como Caucásios e Hispânicos é diferente de grupos como Africanos e Asiáticos;

- *Características individuais*: Características individuais, tais como idade, sexo e partes do corpo também são afetadas pela aparência da pele;

Logo, o passo principal para detecção de pele em imagens é a representação dos *pixels* de pele em um espaço de cor adequado, através da modelagem dos *pixels* de pele e não-pele, utilizando uma distribuição adequada. A escolha do espaço de cor também determina a eficácia da modelagem de um sistema de detecção de pele [kakumanu *et al.* 2007].

Além das informações de pele, a maioria dos trabalhos de detecção de nudez encontrados na literatura utiliza uma combinação de informações como cor, textura e forma. Entretanto, normalmente não é feita uma análise sobre o impacto de cada característica na descrição das imagens. A análise das características permite a identificação das informações mais relevantes para a detecção de nudez em imagens.

Este trabalho apresenta uma análise e a seleção de características com o auxílio de uma arquitetura denominada ANDImage (*Architecture for Nude Detection in Image*). ANDImage utiliza um módulo de detecção de pele e um módulo de zoneamento de imagens, que divide a imagem em regiões para análise e tratamentos independentes, adotando a hipótese de que em imagens de nudez há uma maior predominância de região de pele no centro da imagem. O módulo de zoneamento também permite a comparação entre características obtidas a partir da detecção de pele, propriedades locais, e características obtidas sem detecção de pele, propriedades globais. A detecção de nudez é realizada pela técnica de aprendizagem de máquina, *Support Vector Machines* (SVM) [Burges 1998].

1.2 Motivação

O crescimento da Internet tem proporcionado uma grande mudança nos hábitos da população mundial. Entretanto, toda nova tecnologia traz benefícios e possíveis problemas. Por exemplo, muito conteúdo disponibilizado nas diversas mídias veiculadas na Internet é considerado não aceitável para determinadas empresas ou corporações. Porém, a principal dificuldade para filtrar mídia imprópria está relacionada ao volume de dados disponíveis na Internet, bem como à natureza dinâmica do ambiente.

É possível notar que uma estratégia manual para limitar a veiculação de conteúdo ofensivo seria muito complexa em termos de recursos humanos. Diante dessa situação, estratégias automáticas ou semi-automáticas de filtro poderiam prover uma solução desejável

para o problema em questão. Neste caso, há diversas pesquisas que utilizam a combinação de técnicas de processamento digital de imagens, aprendizagem de máquina, mineração de dados e recuperação da informação, para julgar com sucesso se uma imagem exhibe conteúdo impróprio ou não.

Muitos sistemas de detecção de imagens de nudez utilizam um pré-processamento de detecção de pele. Este trabalho segue a mesma linha de raciocínio e busca desenvolver um módulo de detecção de pele robusto em relação à variação de luminosidade no ambiente. Esse módulo pode ser usado em outras aplicações, como por exemplo, a detecção de faces e reconhecimento de gestos.

Em se tratando de sistema de detecção de nudez, existem diversas aplicações para uso desse tipo de sistema, inclusive com valor comercial, como por exemplo:

- **Servidor Web:** Um servidor pode usar um sistema de detecção de nudez nos casos em que se deseja filtrar o conteúdo armazenado nele. Isso pode ser aplicado em servidores de armazenamento de conteúdo multimídia, como os de álbuns digitais e comunidades virtuais;
- **Servidor de E-mail:** Um dos maiores problemas enfrentados por servidores de e-mail é a proliferação de *spam*, especialmente devido à disseminação de *malwares* através de imagens contendo nudez. Neste caso, o conteúdo indesejado pode ser detectado por um módulo de detecção de nudez e informado ao servidor de e-mail, que por sua vez não faz a entrega da mensagem na caixa postal do usuário;
- **Plug-in:** *Plug-ins* de detecção de nudez podem ser aplicados em navegadores *Web* para bloquear automaticamente imagens consideradas inadequadas para a visualização;
- **Sistemas de Auditoria:** Um sistema que busca encontrar conteúdo ofensivo em estações de trabalho de instituições, tais como empresas e universidades, pode ser acoplado ao servidor *proxy*, através de filtro, para evitar o acesso a conteúdo pornográfico pelos usuários. Também pode ser usado como ferramenta de investigação de crimes digitais, como a pedofilia;
- **Máquinas de busca:** Diante de um conjunto de características apresentadas, as máquinas de busca na *Web* podem utilizar um classificador para filtrar páginas indesejadas nos resultados apresentados aos seus usuários.

1.3 Justificativa

Algumas das abordagens para detecção de nudez [Wang *et al.* 1998] [Jones & Regh 2002] usam uma combinação de características extraídas de imagens de pele humana como cor [Yogarajah *et al.* 2010], textura [Zhu *et al.* 2007], formas [Safar *et al.* 2000] e localização dos *pixels* de pele na imagem [Zheng *et al.* 2004]. Entretanto, o uso de muitas características na descrição de imagens de nudez pode representar um aumento indesejado na complexidade e no tempo de processamento dos algoritmos. Além disso, a maioria dos trabalhos não realiza uma análise qualitativa das características envolvidas no processo de classificação, além de não oferecer informações importantes sobre condições de luminosidade, aquisição da base de imagens, quantidade e etnia dos indivíduos em estudo [Kelly *et al.* 2007]. Essas informações são fundamentais para que qualquer modelo proposto seja replicado e validado.

Para evitar este problema, neste trabalho é utilizado um processo de seleção de características para detecção de nudez. Este processo busca reduzir a dimensão do espaço de características, eliminar os ruídos no classificador ocasionados por dados pouco relevantes e aumentar a precisão do classificador.

Com base na afirmação de diversos autores de que há uma associação clara entre a quantidade de *pixels* de cor de pele e imagens de nudez, este trabalho utiliza um filtro de pele e busca através do algoritmo de zoneamento em imagens, não só analisar a distribuição das características em uma imagem de nudez, mas também confirmar a hipótese de que nessas imagens há uma maior exposição de áreas de pele no centro da imagem.

Diante de um conjunto de características extraídas a partir do conteúdo da imagem, há a necessidade de utilização de alguma ferramenta para detectar nudez na referida imagem. Uma maneira de realizar esta tarefa é utilizar um classificador, ou seja, uma técnica de aprendizagem de máquina, como árvores de decisão, redes neurais e SVM. Através de um processo de treinamento, estes classificadores são capazes de identificar de forma automática e elevada precisão, se uma determinada imagem pertence à classe nudez ou à classe não-nudez. Neste trabalho o classificador SVM foi utilizado por possuir forte fundamentação matemática, implementação disponível em ferramentas públicas e melhores indicadores no processo de classificação de imagens de nudez.

1.4 Objetivos do Trabalho

O objetivo deste trabalho é analisar e selecionar as principais características envolvidas no processo de detecção de imagens de nudez, com o auxílio de uma arquitetura que utiliza um filtro de pele e um módulo de zoneamento em imagens, o qual permite o tratamento independente de características globais e locais. Para isso, técnicas de Processamento Digital de Imagens (PDI) e de aprendizagem de máquina são utilizadas.

Os objetivos específicos desta dissertação são descritos nos itens a seguir:

1. Criar um módulo de detecção de pele robusto a tonalidades de pele e variações de luminosidade devido ao ambiente, que apresente um desempenho adequado para a composição de um sistema de detecção de nudez;
2. Aplicar o algoritmo de zoneamento em imagens objetivando dividir a imagem em regiões para análise e tratamento independentes, adotando a hipótese de que em imagens de nudez há uma maior predominância de região de pele no centro da imagem;
3. Analisar qualitativamente algumas das principais características envolvidas no processo de classificação de imagens de nudez propostas na literatura;
4. Comparar as abordagens globais e locais para descrição de imagens de nudez, em termos de precisão e tempo que o recurso utiliza para analisar cada imagem, de forma a obter um melhor compromisso entre esses dois fatores para aplicação em sistemas *on-line*;
5. Selecionar as características, com o propósito de reduzir a dimensionalidade do espaço de características e ao mesmo tempo, aumentar a taxa de classificação.

1.5 Organização do Documento

O restante desta dissertação está organizado como segue:

- O Capítulo 2 fornece informações necessárias para o entendimento deste trabalho. Serão detalhados os conceitos básicos de imagens digitais, espaços de cores, principais características em imagens, métricas de desempenho, aprendizagem de máquina e, principalmente, o funcionamento básico do classificador SVM. Será discutida também, uma visão geral sobre alguns trabalhos relacionados encontrados na literatura;

- O Capítulo 3 detalha os principais componentes da arquitetura ANDImage, como por exemplo, a normalização, filtro de pele, zoneamento em imagens, e o processo de classificação. No processo de extração de características serão fornecidas definições sobre o emprego de características globais e locais;
- O Capítulo 4 apresenta resultados preliminares obtidos, além de informações sobre a base de dados, desenvolvimento da arquitetura, experimentos. Os resultados obtidos também são discutidos;
- Por fim, no quinto Capítulo, é apresentado um resumo dos resultados obtidos, as conclusões são discutidas e algumas direções futuras são propostas.

Capítulo 2

Conceitos Básicos e Trabalhos Relacionados

Nesta seção são apresentados conceitos básicos sobre imagens digitais, processamento de imagens, uma breve descrição sobre aprendizagem de máquina, métricas de avaliação e, alguns trabalhos relacionados que são de fundamental importância para o entendimento deste trabalho.

2.1 Imagens Digitais

Inicialmente é apresentada uma descrição do conceito de imagens digitais e alguns de seus componentes, bem como algumas técnicas de processamento de imagens.

De acordo com Gonzales & Woods [2008], uma imagem digital é representada por uma função $f(x,y)$ da intensidade luminosa, sendo seu valor, em qualquer ponto de coordenadas espaciais (x,y) , proporcional ao brilho (ou nível de cinza) da imagem naquele ponto, ou seja, uma imagem digital pode ser considerada uma matriz de pontos elementares, em que cada ponto recebe o nome de *pixel*. A Figura 2.1 mostra a representação de uma imagem digital.

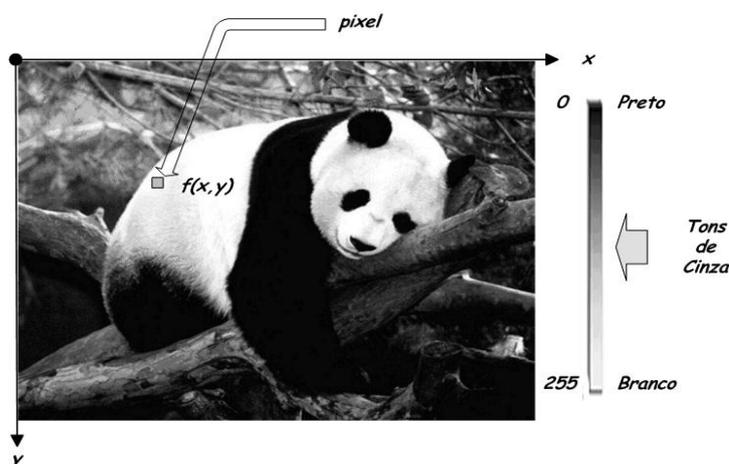


Figura 2.1: Representação de uma imagem digital.

A partir da imagem na Figura 2.1, são destacados os principais conceitos de imagens digitais:

- *Pixel*: É a representação de cada elemento da imagem, sendo que cada *pixel* contém um valor e localização específicos. O conjunto de *pixels* de uma imagem é finito.
- *Níveis de Cinza*: Imagem digital em níveis de cinza é uma imagem em que cada *pixel* varia no espaço de cores entre o preto (0), como menor intensidade, e o branco (255), como maior intensidade.

Algumas técnicas de processamento digital de imagens podem ser aplicadas para possibilitar a manipulação e a extração de atributos relevantes das imagens. As técnicas utilizadas neste trabalho, cujo uso é justificado no Capítulo 3, são descritas a seguir.

- *Análise de Histograma*: O Histograma é apresentado através de um gráfico que mostra a relação entre os níveis de cinza e a quantidade (ou frequência) de *pixels* com esses níveis de cinza na imagem. Do ponto de vista estatístico, o histograma de uma imagem digital representa a função de distribuição de probabilidade (FDP) dos níveis de cinza presentes na imagem. Então, se todos os valores de frequência forem divididos pelo número total de *pixels* da imagem, é possível obter-se a FDP.
- *Normalização*: Processo que consiste em gerar um novo conjunto de amostras para aplicar a um determinado contexto a ser explorado, ou seja, aplicação de técnicas como redução de imagens e conversão do formato original das imagens, são incluídas no processo de normalização, pois dependendo do contexto, faz-se necessário manter as imagens em um mesmo tamanho e mesmo formato, para que as mesmas sejam padronizadas.
- *Segmentação*: Tem por objetivo dividir a imagem em regiões ou objetos, segundo um critério pré-estabelecido. Geralmente, o resultado esperado não é uma nova imagem, mas sim, um conjunto de regiões/objetos. Vale ressaltar que a precisão da fase de segmentação determina o sucesso ou falha dos procedimentos de análise de imagem. A segmentação possui duas estratégias genéricas: descontinuidade, onde a partição de uma imagem ocorre com base nas alterações bruscas de intensidade (detecção de contornos) e; similaridade, em que a partição é efetuada com base na similaridade entre os *pixels*, seguindo um determinado critério, como por exemplo, limiarização, crescimento de regiões, etc. A Figura 2.2 mostra o resultado do processo de segmentação através de detecção de bordas.

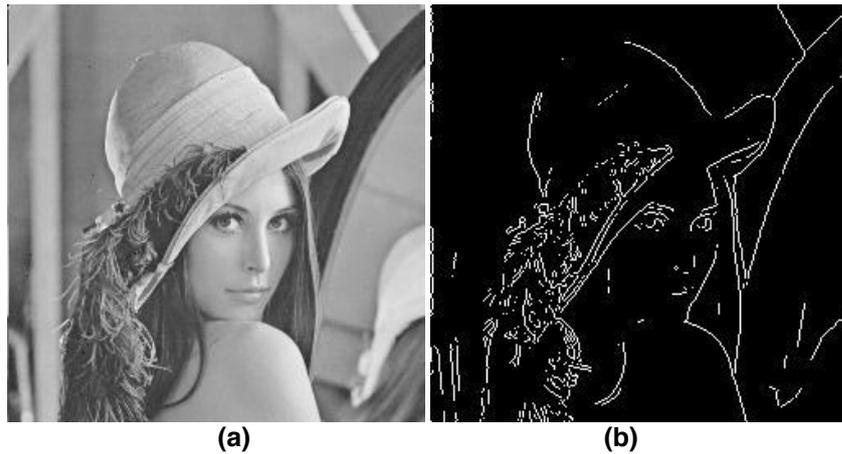


Figura 2.2: Resultado de Segmentação: (a) imagem original (b) imagem segmentada através de detecção de borda.

- *Zoneamento em Imagens*: É uma técnica de segmentação que consiste em definir regiões na imagem e tratá-las de maneira independente, ou seja, os algoritmos de extração de características analisam independentemente cada uma dessas regiões. Essa estratégia baseia-se na hipótese de que a região central das imagens é a mais representativa, enquanto que regiões mais externas representam cenários ou outros elementos menos importantes. A Figura 2.3 exibe dois exemplos de aplicação do zoneamento. As imagens nessa figura estão divididas em três regiões retangulares concêntricas e proporcionais às suas dimensões.

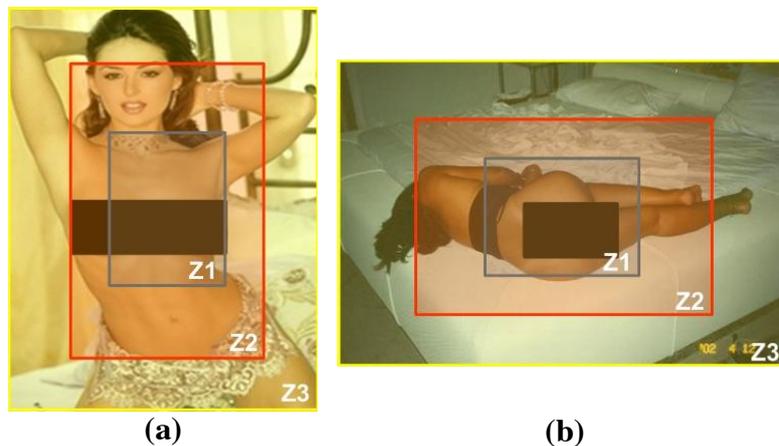


Figura 2.3: Imagens que demonstram o resultado da aplicação do zoneamento, onde (a) representa o zoneamento em uma imagem de nudez com orientação vertical, (b) o zoneamento em imagem de nudez com orientação horizontal.

Considerando-se o fato de uma imagem ser formada por um conjunto de cores, faz-se necessário descrever os principais espaços de cores utilizados na literatura para detecção de nudez. Esses espaços de cores são descritos na próxima seção.

2.1.1 Espaços de Cores

Espaços de cores são estratégias através das quais é possível especificar, criar e visualizar a cor. A cor é geralmente especificada usando-se três coordenadas, ou parâmetros, que descrevem a posição da cor no espaço de cores a ser utilizado.

Cada espaço de cor possui diferentes propriedades e características úteis ou nocivas para cada aplicação. A escolha do espaço de cor pode ser considerada como o passo fundamental para a classificação de cores de pele. Diversos espaços de cores vêm sendo utilizados na literatura em aplicações como Colorimetria, Computação Gráfica, Visão Computacional e Detecção de Pele.

O espaço de cor RGB (do inglês *Red, Green, Blue*) é o padrão mais comumente usado para armazenar e representar imagens digitais, uma vez que os dados capturados por uma câmera são normalmente fornecidos em RGB. Neste espaço, as cores são geradas a partir da combinação de três cores básicas: vermelho, verde e azul. Qualquer outro espaço de cor pode ser obtido a partir de uma transformação linear ou não-linear do espaço RGB.

Como a pele é geralmente utilizada como forte indicativo de presença de nudez em imagens, a transformação de espaço de cor é utilizada, por exemplo, para diminuir a sobreposição entre *pixels* de pele e não-pele, a fim de contribuir para uma classificação correta dos *pixels* de pele através do fornecimento de parâmetros imunes a diferentes condições de luminosidade. Outro fator observado é que os diferentes tons de pele possuem maior intensidade do que crominância. Devido a esses fatores, uma prática comum é isolar a componente de luminância para classificação de pele [Yang *et al.* 1998]. A seguir são descritos os principais espaços de cores utilizados na literatura e suas propriedades.

Espaços de Cores Básicos (RGB, RGB Normalizado e CIE-XYZ)

Como visto na Figura 2.4, e mencionado anteriormente, os valores RGB correspondem as cores primárias vermelho, verde e azul. Para reduzir a dependência de iluminação, os componentes de cor RGB são normalizados de modo que a soma dos componentes normalizados seja uma unidade de valor ($r + g + b = 1$). Sendo a soma desses componentes igual a 1, o terceiro componente não possui qualquer informação significativa e normalmente é descartado de forma a obter uma redução na dimensionalidade. Foi observado em [Yang & Ahuja, 1998] que, sob determinados pressupostos, as diferenças na cor de *pixels* de pele ocasionadas por condições de iluminação e etnia, podem ser reduzidas significativamente no espaço normalizado RGB.

O sistema CIE (*Commission Internationale de l'Eclairage*) adota o espaço de cor XYZ que corresponde aos três estímulos percebidos pelos fotorreceptores do sistema visual humano, descrevendo a cor como componente *Y* (luminância), e dois componentes adicionais *X* e *Z* (crominância). O *CIE-XYZ* foi construído a partir de experimentos da psicofísica e adota algumas premissas tais como: as componentes de cor devem ser positivas; obtenção do maior número possível de cores espectrais com algumas coordenadas de cromaticidade nulas e; duas primárias devem ser nulas [kakumanu *et al.* 2007].

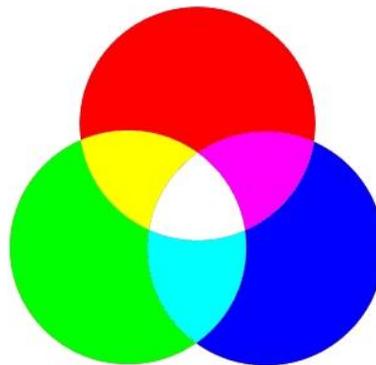


Figura 2.4: Espaço de cor RGB e cores produzidas a partir de cores primárias.

Espaços de Cores Perceptuais (HSI, HSV, HSL e TSL)

O *HSV* (*Hue, Saturation, Value*), também conhecido como *HSB* (*Hue, Saturation, Brightness*), é um espaço de cor não-linear obtido através de uma transformação do espaço *RGB*, podendo ser referido como um espaço de cor perceptual devido a sua semelhança com a percepção humana de cor. *Hue* é um componente que descreve cor pura (por exemplo, amarelo puro, laranja ou vermelho), enquanto *Saturation* fornece uma medida do grau em que uma cor pura é diluída pela luz branca. *Value* são tentativas em representar o brilho ao longo do eixo de cinza (por exemplo, do branco ao preto), mas como o brilho é subjetivo, torna-se difícil de ser medido [Gonzalez & Woods, 2008]. A Figura 2.5 ilustra o espectro de cor representado pelo espaço HSV.

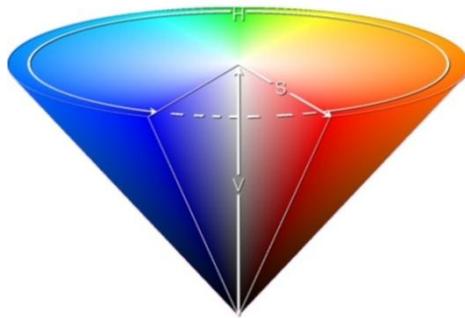


Figura 2.5: Espaço de cor HSV.

A transformação do espaço de cor *RGB* para *HSV* é invariante à alta intensidade de luzes brancas, luz ambiente e orientações de superfície em relação à fonte de luz e, portanto, pode se configurar como uma boa escolha para métodos de detecção de pele. O *HSV* é de fato um dos espaços mais utilizados para a detecção de cor de pele em imagens [Brown *et al.* 2001][Basilio *et al.* 2011].

O *HSI* (*Hue, Saturation, Intensity*), também conhecido como *HSL* (*Hue, Saturation, Luminance*), é outro espaço de cor perceptivo que fornece excelentes resultados para detecção de pele. Como *Value* em *HSV*, a intensidade é outra representação do nível de cinza, mas dissociada dos componentes de cor (*Hue e Saturation*). O espaço *HSI* foi usado por Wang *et al.* [2005] como parte de uma abordagem baseada em conteúdo. Esses autores afirmam que os *pixels* de pele e do fundo da imagem podem ser mais facilmente diferenciados com *HSI* do que com *RGB*.

Outro espaço de cor similar é o *TSL* que define a cor como *Tint* (*Hue* com branco adicionado), *Saturation* e *Lightness*. *TSL* foi usado para detecção de pele por Terrillon *et al.* [1998].

Espaços de Cores Ortogonais (YCbCr, YIQ, YUV, YES)

Os espaços de cores ortogonais reduzem a redundância presente nos canais de *RGB* e representam a cor como componentes estatisticamente independentes. Como a luminância e crominância são explicitamente componentes separados, estes espaços são uma escolha favorável para a detecção de pele.

O espaço *YCbCr*, geralmente associado a vídeos digitais, representa a cor como luminância (*Y*) e crominância (*Cb* e *Cr*). Luminância é uma representação de brilho em uma imagem e crominância define os dois atributos de uma tonalidade de cor e saturação. *Y* é calculada como uma

soma ponderada dos valores RGB, enquanto a cromaticidade (Cb e Cr) é calculada subtraindo-se o componente de luminância dos valores de B (*Blue*) e R (*Red*), sendo obtidas através das equações 2.1 e 2.2:

$$Cb = \frac{B}{R + G + B} \quad (2.1)$$

$$Cr = \frac{R}{R + G + B} \quad (2.2)$$

O espaço $YCbCr$ é uma das opções mais populares para detecção de pele. Esse espaço foi utilizado por Arentz & Ostald [2004], Zheng *et al.* [2004] e, Lee *et al.* [2007]. Uma variante do espaço $YCbCr$, o $YCgCr$ foi usado por Dios & Garcia [2001] e difere do original no uso do componente de cor Cg , no lugar do componente Cb . Os autores indicam uma melhor eficiência desta variante em relação ao $YCbCr$.

Outros espaços de cor nesta categoria são YIQ , YUV e YES , sendo as cores representadas como luminância (Y) e cromaticidade (IQ , UV , ES). Para a detecção de pele, estes espaços foram usados por Jiao *et al.* [2002], Duan *et al.* [2002], Saber & Tekalp [1998], respectivamente.

Espaço de cor perceptualmente uniforme (CIELab e CIELuv)

A percepção de uniformidade é representada como duas cores que diferem em aparência através da observação humana, ou seja, ao alterar a matiz, a saturação ou o brilho, a mudança de cores se apresentará uniforme, como o usuário realmente espera. O espaço de cor uniforme (*Uniform Color Space* – UCS) foi definido de modo que todas as cores são organizadas pela diferença de percepção das cores. No entanto, a uniformidade perceptual nestes espaços de cor é obtida através de rigorosas transformações computacionais. Nestes espaços de cores, o cálculo da luminância (L) e cromaticidade (ab ou uv) é feito através de um mapeamento não-linear das coordenadas XYZ . Para a detecção de pele, o espaço $CIELab$ (vide Figura 2.6) foi utilizado por Cai & Goshtasby [1999], Kawato & Ohya [2000]. O espaço $CIELuv$ foi usado por Yang & Ahuja [1998].

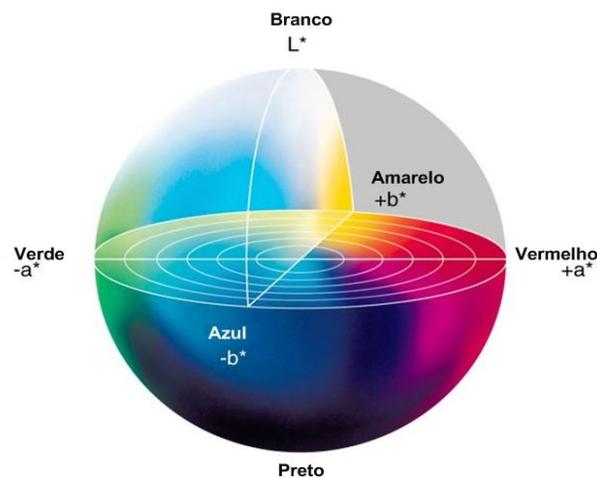


Figura 2.6: Espaço de cor CIE Lab.

Para que a classificação de imagens seja realizada, são extraídas características que representem informações relevantes das imagens. Na próxima seção são descritas as principais propriedades de imagens usadas na literatura, assim como a classificação dessas características em globais e locais.

2.1.2 Propriedades de Imagens

O método mais eficaz para reconhecimento de padrões e consulta em imagens é extrair as informações contidas na própria imagem, ou seja, armazenar em disco os valores dos *pixels* da própria imagem, para que esta seja classificada ou comparada com valores correspondentes de outras imagens.

Atualmente, porém, diversas pesquisas têm sido realizadas no sentido de melhorar os métodos de extração de características, buscando diminuir o espaço em disco e aumentar a precisão no reconhecimento de objetos. Para tal, filtros e transformações podem ser aplicadas à imagem para produzir uma representação mais compacta ou para enfatizar certas propriedades da imagem, por exemplo, o uso de Análise de Componentes Principais (PCA – do inglês *Principal Component Analysis*) para redução de dimensionalidade e a aplicação de filtros Sobel para enfatizar as bordas.

Conforme Figura 2.7, podem-se destacar três grandes grupos de propriedades em imagens: cor, forma e textura. Além disso, há também outras terminologias empregadas no processo de extração de características em imagens que dependem unicamente da forma como as informações são elencadas, ou seja, se a informação é extraída de toda a imagem, pode-se dizer que a informação

é global. Porém, se há necessidade de segmentar ou particionar a imagem para posterior extração das características, o recurso é considerado local.

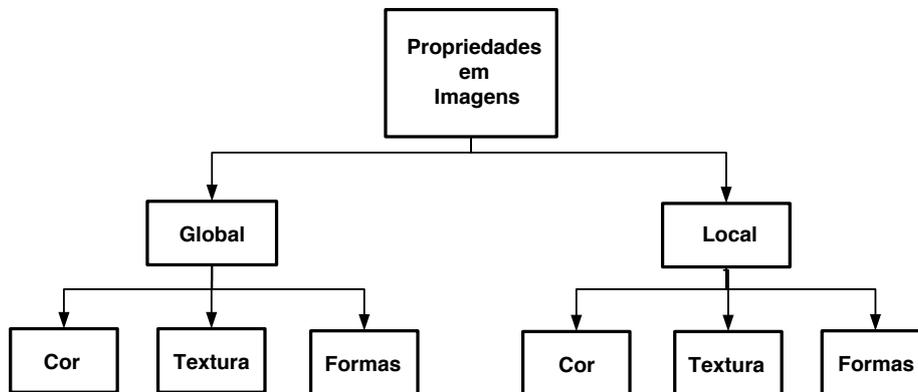


Figura 2.7: Terminologia das propriedades em imagens.

Maiores detalhes acerca dessas propriedades são fornecidos nas próximas seções.

2.1.2.1 Propriedades Globais

Diversos sistemas de reconhecimento de objetos utilizam recursos globais. A maioria dos descritores de textura e de forma se enquadra nesta categoria. As propriedades globais são interessantes porque produzem representações compactas de imagens, em que cada imagem corresponde a um ponto em um espaço de características de alta dimensão. Qualquer classificador pode ser treinado com essas características. Por outro lado, características globais são sensíveis à desorganização e oclusão. Como resultado, é assumido que uma imagem contém apenas um único objeto, ou que apesar de conter diversos objetos, o objeto de interesse possui cor, textura e forma bem agrupadas.

A seguir são dadas melhores definições sobre as propriedades de cor, forma e textura, bem como descritores que serão utilizados nesta dissertação como recursos globais.

Propriedades Globais de Cor

A cor representa uma das propriedades visuais mais importantes identificadas em imagens pela visão humana, além de agregar menos esforço computacional no processo de extração. Diversos trabalhos baseiam-se unicamente neste tipo de propriedade ou realizam a combinação com outros tipos, como textura e forma [Jain & Valaya 1995]. A cor pode determinar semelhança entre

imagens, pois há uma predominância de cores específicas em determinadas classes de imagens. Em pessoas, por exemplo, há uma diversidade de tons de pele que são pouco frequentes em imagens de ambientes diversos, veículos, etc. Esta informação é extremamente relevante para a construção de um classificador de imagens, tanto pelo fato de precisão, quanto pelo custo computacional. Trata-se, portanto, de informação de baixo nível [Arentz & Ostald 2004].

O GCH (do inglês – *Global Color Histogram*) é um dos descritores mais populares na literatura por ser um dos métodos mais antigos e por possuir simples implementação e eficácia aceitável em diversas aplicações. O GCH quantiza o espaço de cor em uma quantidade uniforme de *bins* (arquivo contendo informações da imagem), e percorre a imagem uma única vez calculando a quantidade de *pixels* que pertence a cada *bin*. O tamanho do vetor de características gerado depende diretamente da quantização realizada. Segundo Penatti & Torres [2008], a complexidade do algoritmo de extração de características do GCH é da ordem de $O(n)$ e se configura como um algoritmo que obtém um curto tempo de resposta no caso de uma consulta na *Web*, independentemente dos requisitos de armazenamento gerados na quantificação do espaço de cor. Por esses motivos, o GCH foi escolhido para ser implementado e testado nos experimentos deste trabalho.

Propriedades Globais de Textura

Esse tipo de propriedade é importante devido à representatividade que pode fornecer para distinguir diferentes classes. Baseia-se em indicadores estatísticos representados por descritores globais ou locais para obter padrões de textura.

Para extrair informações da textura, as imagens são inicialmente convertidas para níveis de cinza e submetidas a um operador de textura que computa a relação entre os níveis de cinza e as posições dos *pixels* na imagem. Neste trabalho, as características de textura são extraídas a partir de uma matriz de co-ocorrência (GLCM – do inglês *Gray-Level Co-occurrence Matrix*) proposta por Haralick *et al.* [1973]. A seguir são fornecidas maiores informações sobre o descritor.

Uma GLCM é uma matriz bi-dimensional que representa a probabilidade de ocorrência de conjuntos de níveis de cinza em i e j dentro de uma relação espacial definida na imagem, sendo que i e j são os índices da linha e da coluna da matriz de vizinhança. Na GLCM, o *pixel* vizinho a ser comparado com o *pixel* analisado normalmente é o *pixel* à direita (*default*), embora qualquer *pixel* vizinho possa ser utilizado, inclusive alguns *pixels* com distância de vizinhança maior que 1. Há

também a possibilidade de se trabalhar com mais de um *pixel* vizinho, criando-se uma matriz de vizinhança. A Figura 2.8 mostra como a matriz de vizinhança pode ser construída, sendo que a relação espacial é definida em termos de uma distância d e um ângulo θ .

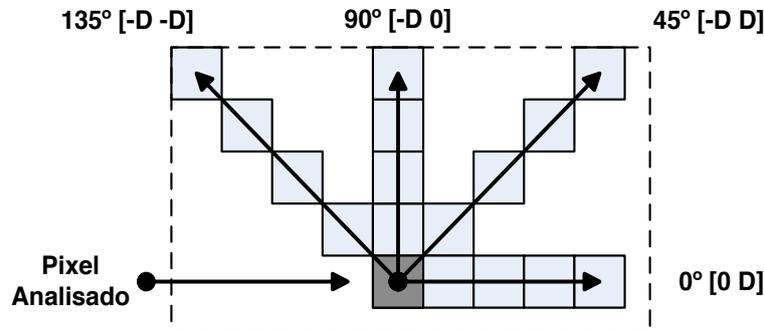


Figura 2.8: Construção da matriz vizinhança.

A partir da GLCM, alguns dados estatísticos podem ser extraídos, tais como: homogeneidade, correlação, contraste e energia.

Propriedades Globais de Formas

Este grupo de propriedade busca identificar formas geométricas simples nas imagens. As arestas (bordas) são definidas como discontinuidades na intensidade do nível de cinza dos *pixels* e, constituem um importante recurso para representar o conteúdo dos objetos contidos na imagem [Won *et al.* 2002]. Portanto, as arestas são úteis para segmentação e identificação dos objetos em uma cena.

Algumas propriedades de forma utilizadas em detecção de nudez baseiam-se em descritores, aplicados após a segmentação de áreas de pele, como por exemplo, os momentos invariantes, proposto por Zheng *et al.* [2004] e SIFT (*Scale-Invariant Feature Transform*), utilizado por Lopes *et al.* [2009]. Outros descritores de formas utilizam informações simples calculadas a partir do conceito de histograma de arestas. Segundo Jain & Valaia [1991], o histograma de arestas é muito útil para a indexação e recuperação de imagens, pois representa a frequência e direção das mudanças de intensidade da imagem.

Neste trabalho é utilizado um descritor denominado *Edge Histogram Descriptor* (EHD) proposto por Won *et al.* [2002] para expressar a distribuição de informações da borda ou do objeto presente na imagem. Considerando o fato de que é importante manter o tamanho do vetor o mais compacto possível para armazenamento eficiente dos metadados, este descritor realiza uma

normalização na imagem de maneira que o vetor de características contenha apenas 80 atributos para descrição da distribuição local na imagem. Assim, para melhorar o desempenho de recuperação, é necessária a utilização da informação de distribuição da borda global.

O EHD basicamente representa a distribuição de 5 tipos de arestas em cada área local da chamada sub-imagem. Como mostrado na Figura 2.9, a sub-imagem é definida pela divisão da imagem em blocos fixos de 4×4 . Assim, a partição da imagem sempre produz 16 sub-imagens de tamanho igual, independentemente do tamanho original da imagem. Para caracterizar a imagem, o histograma de arestas para cada sub-imagem é gerado. As arestas nas sub-imagens são classificadas em cinco tipos: vertical, horizontal, linhas diagonais de 45 graus, linhas diagonais de 135 graus, e linhas não-direcionais (pontos isolados).

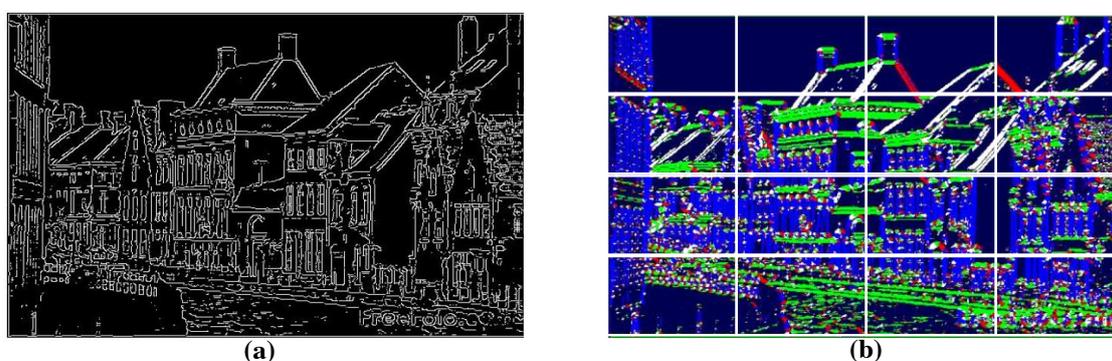


Figura 2.9: Exemplo do processo de extração de características do EHD, onde (a) corresponde a uma imagem binarizada após a aplicação do operador Sobel e; (b) uma imagem segmentada e particionada em blocos fixos de 4×4 .

O histograma para cada sub-imagem representa a frequência relativa de ocorrência dos 5 tipos de arestas na sub-imagem correspondente. Logo, cada atributo corresponde a um dos cinco tipos de borda. Uma vez que existem 16 sub-imagens, um total de $5 \times 16 = 80$ atributos do histograma são necessários para cada imagem. Para localização de arestas nas imagens, o EHD aplica operadores como Canny [Canny 1986] e Sobel [Sobel 1970] que utilizam máscaras de convolução, cujo custo computacional é baixo, em termos de cálculo [Jiao *et al.* 2001].

O processo de extração de características de formas utilizado neste trabalho consiste na aplicação de operadores de Sobel para encontrar bordas, e posteriormente uma série de máscaras para computar o número de formas predefinidas na imagem. Essas características podem ser extraídas através da convolução de máscaras 3×3 , conforme mostrado na equação abaixo:

$$S = w_1z_1 + w_2z_2 + \dots + w_9z_9 = \sum_{i=1}^9 w_i z_i \quad (2.3)$$

onde w_1, w_2, \dots, w_9 , correspondem aos pesos da máscara, e são apresentados na Figura 2.10. Já z_1, z_2, \dots, z_9 são os valores dos níveis de cinza da imagem correspondente a estes pesos, e S é o resultado da operação de convolução, sendo atribuído o valor de S ao pixel que coincide com a posição central (w_5) da máscara, que percorrerá a imagem, desde o seu canto superior esquerdo até seu canto inferior direito.

w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

Figura 2.10: Máscara de processamento 3x3.

2.1.2.2 Propriedades Locais

A utilização de propriedades locais surge como um importante recurso para reconhecimento de objetos e descrição de imagens. Todo o processo é baseado em computar pontos de interesse na imagem, seja através de segmentação ou particionamento da imagem [Lisin *et al.* 2005]. A seguir, serão descritos alguns métodos em que recursos locais são utilizados.

Um dos principais argumentos para utilização de propriedades locais baseia-se no fato de haver inúmeros pontos de interesse em uma imagem, como por exemplo, a determinação de relações geométricas de um objeto, fazendo com que a precisão no reconhecimento seja mais eficaz, porém a um custo computacional maior.

Normalmente, os pontos de interesse são detectados em múltiplas escalas e devem ser repetidos em diferentes pontos de vista de um objeto, como mostrado na Figura 2.11. Os pontos de interesse também são esperados para capturar a essência da aparência do objeto. Logo, o descritor local descreve a imagem em torno de um ponto de interesse.

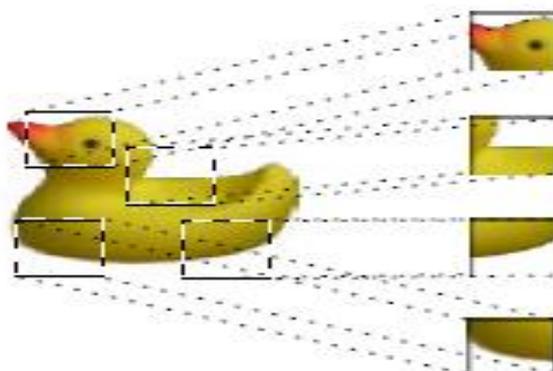


Figura 2.11: Extração de características locais.

Outra vantagem de usar propriedades locais é que elas podem ser usadas para reconhecer o objeto, apesar de haver desordens significativas e oclusão. Também não precisam de uma segmentação do objeto do fundo, como é o caso de particionamento em imagens. O simples fato de dividir a imagens em regiões de blocos fixos para computar outras características globais como cor, textura e forma é considerado segundo Lisin *et al.* [2005] como um recurso local.

Propriedades Locais de Cor

Em aplicações para detecção de nudez, há diversas abordagens que utilizam informações da região de pele como recurso local para descrição de conteúdo ofensivo, por exemplo, o trabalho pioneiro de Forsyth *et al.* [1996] que combina filtro de pele e análise de textura para detecção de pele. Após a detecção de pele, é realizada uma análise geométrica para identificar grupos de regiões que contém pele, e posteriormente, localizar diferentes partes do corpo humano. Tais características são baseadas na quantidade de *pixels* de pele conectados e porcentagem de *pixels* de pele na imagem. Wang *et al.* [1998] desenvolveram um sistema denominado *Wavelet Image Pornographic Eliminator* (WIPE_{TM}), que utiliza uma combinação de filtros de sequência (*Daubechies wavelets*), no qual os filtros mais rápidos são executados primeiro para que as imagens não-ofensivas passem mais rápido, reduzindo o tempo de classificação. Após a passagem da imagem por todos os filtros, o algoritmo cria os vetores de características baseadas no histograma de cor, textura, formas e região de pele, para verificar a similaridade da mesma com um conjunto de imagens de treinamento pré-classificadas.

Outras abordagens, após o filtro de pele, utilizam informações da borda para descrever imagens ofensivas. Lee *et al.* [2007] utilizam informações da distribuição dos *pixels* de pele na

imagem, combinada com informações da textura e após o filtro de pele, extraem informações geométricas como *área* e *perímetro da borda*. Já o trabalho de Zheng *et al.* [2004] utiliza as informações de borda para extrair outras características como, por exemplo, *compacidade*, *excentricidade* e *retangularidade*. A Figura 2.12 fornece um exemplo do conceito de *excentricidade* e *retangularidade*.

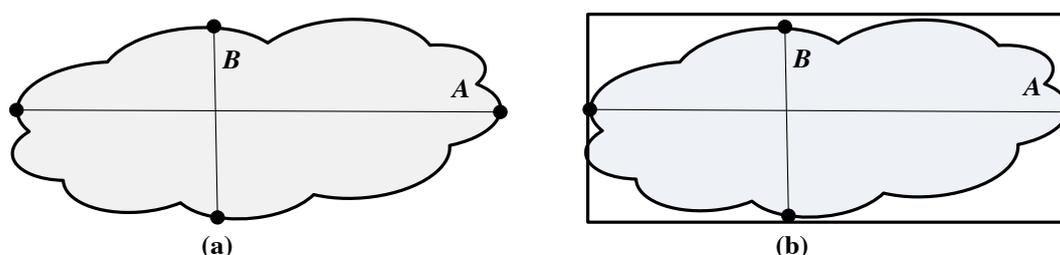


Figura 2.12: Exemplo de características extraídas a partir da borda, (a) *excentricidade* e; (b) *retangularidade*.

As características extraídas são usadas para compor vetores que são submetidos a técnicas de classificação para a tomada de decisão. Técnicas de classificação são algoritmos de aprendizagem de máquina. Essa área de pesquisa é descrita resumidamente na próxima seção.

2.2 Aprendizagem de Máquina

Entidades inteligentes destacam-se pela capacidade de adequação a novos ambientes e de resolução de novos problemas. Um computador pode ser orientado a interpretar informações recebidas de uma forma que melhore gradualmente seu desempenho [Rich & Knight 1991]. Essa é a base na qual a área de pesquisa em aprendizagem de máquina, subárea da inteligência artificial, está fundamentada, isto é, aprendizagem de máquina estuda métodos computacionais capazes de obter um determinado conhecimento específico através de experiências.

Os algoritmos de aprendizagem de máquinas possuem normalmente foco em métodos estatísticos e sua aplicação prática inclui tarefas de reconhecimento de padrões diversos, tais como detecção de intrusão, diagnósticos médicos, dentre outros. Portanto, esses algoritmos podem “aprender” a definir padrões das classes envolvidas em um problema, a partir de exemplos reais obtidos do ambiente. A Figura 2.13 mostra um modelo genérico de aprendizagem de máquina. O ambiente fornece informação para um elemento de aprendizagem que usa essa informação para

fazer melhoramentos em uma base de conhecimento e então, um elemento de desempenho usa essa base para executar sua tarefa.

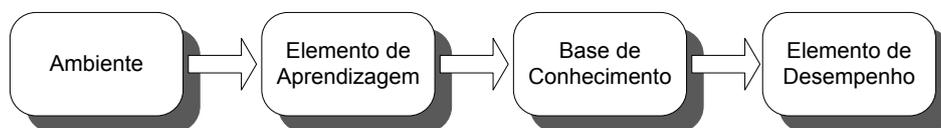


Figura 2.13: Modelo genérico de aprendizagem de máquina.

Aprendizagem de máquina pode ser dividida em dois paradigmas fundamentais: aprendizagem com professor e aprendizagem sem professor [Haykin 2008]. No primeiro, normalmente chamado aprendizagem supervisionada, o sistema precisa conhecer o ambiente. Esse conhecimento é representado por um conjunto de exemplos de pares de entrada-saída que são transmitidos em uma sequência de instruções que o computador seguirá para alcançar o efeito desejado. Na aprendizagem sem professor, não há exemplos rotulados da função a ser aprendida.

Independente do tipo de aprendizagem, normalmente o conjunto de todas as regras definidas para a solução de um problema de aprendizagem é chamado algoritmo de aprendizagem ou classificador. Há uma variedade de algoritmos que oferecem vantagens e desvantagens, dependendo do problema em que são aplicados.

Um classificador é um método de agrupamento de classes baseado nos resultados das características extraídas. Segundo [Kelly *et al.* 2007], técnicas de aprendizado supervisionado possuem forte aplicação em detecção de nudez e, dentre as principais publicações, há uma prevalência do uso de pelo menos uma das quatro técnicas citadas a seguir: Máquinas a Vetores de Suporte (SVM – do inglês *Support Vector Machines*), Redes Neurais (NN – do inglês *Neural Network*), árvore de decisão (DT – do inglês *Decision Trees*) e k-vizinho mais próximo (KNN – do inglês *k-Nearest Neighbor*).

SVM foi escolhido para ser usado neste trabalho por apresentar resultados animadores em problemas de detecção de pele e detecção de nudez [Kelly *et al.* 2007]. Esse classificador é descrito na próxima seção.

2.2.1 SVM

É uma técnica de classificação amplamente aplicada em problemas de reconhecimento de objetos [Santos 2002], classificação de textos [Tong & Koller 2001], detecção de *phishing* [Miyamoto *et al.*

2009], detecção de intrusos [Xiao *et al.* 2007], classificação de sites ofensivos [Hu *et al.* 2009], e detecção de nudez [Lee *et al.* 2007].

A principal vantagem de SVM é a baixa probabilidade de erros de generalização [Vapnik 1995]. Quanto à utilização de SVM para detecção de nudez, por exemplo, há duas vantagens principais. A primeira está relacionada à rapidez do algoritmo em fase de uso, uma vez que o desempenho *online* é de primordial importância para esse tipo de aplicação. A segunda razão é a escalabilidade, pois SVM é relativamente insensível ao número de pontos de dados. Dessa forma, a taxa de precisão na classificação não depende diretamente da dimensão do espaço de características [Mukammala *et al.* 2002]. Entretanto, SVM tem como desvantagem a elevada complexidade computacional na fase de treinamento, fato que pode inviabilizar o uso de SVM em problemas que necessitem de treinamento *online*.

Basicamente, o funcionamento de SVM pode ser descrito da seguinte forma: dadas duas classes e um conjunto de instâncias de treinamento cujas amostras pertencem a essas classes, SVM constrói um hiperplano que divide o espaço de características em duas regiões, maximizando a margem de separação entre as mesmas. Esse hiperplano é conhecido como hiperplano de separação ótima. As amostras desconhecidas (exemplos de teste) são então mapeadas para esse mesmo espaço, e atribuídas a uma das classes [Alpaydim 2010].

A Figura 2.14 mostra o hiperplano de separação ótima (reta separadora) para um problema bidimensional típico e linearmente separável. As retas pontilhadas H_1 e H_2 , paralelas ao hiperplano, constituem o par de hiperplanos que geram a margem máxima pela minimização do vetor peso w . Além disso, $|b|/||w||$ é a distância perpendicular do hiperplano à origem e $||w||$ é a norma Euclidiana de w . Os pontos que estão em um dos hiperplanos H_1 e H_2 são chamados vetores de suporte. Esses pontos, indicados na Figura 2.14 por círculos extras, alteram a solução encontrada caso sejam removidos. Além disso, em fase de uso de SVM, apenas os vetores de suporte são necessários para que dados desconhecidos sejam classificados.

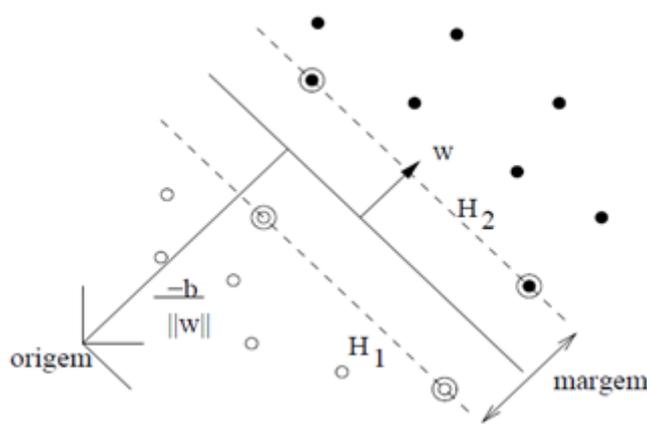


Figura 2.14 - Hiperplano de separação ótima para um problema com duas classes.

O algoritmo original de SVM não encontra a solução desejada quando aplicado a dados não linearmente separáveis, característica presente na maioria dos problemas reais [Alpaydim 2010]. Entretanto, tais problemas podem ser solucionados por SVM através da utilização de funções de separação de dados mais complexas do que funções lineares. Sendo assim, o uso de diferentes funções *kernel* possibilita a construção de SVM com diferentes tipos de superfícies de decisão não-linear no espaço de entrada.

Com o uso de funções *kernel*, as instâncias são inicialmente mapeadas para um espaço de características com dimensão maior do que o espaço de características original, permitindo dessa forma, a classificação em espaços não linearmente separáveis. A Figura 2.15 mostra o processo de transformação de um domínio não linearmente separável, em um problema linearmente separável através do aumento da dimensão, consequência do mapeamento feito por uma função *kernel* $F(x)$.

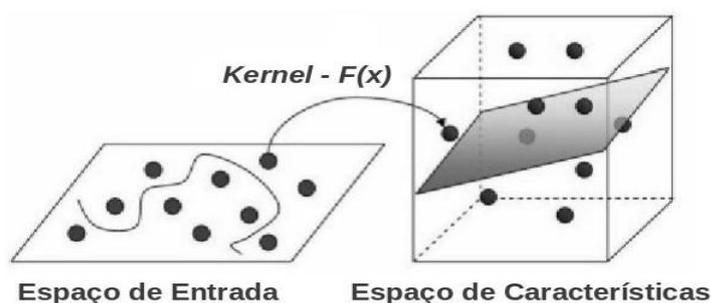


Figura 2.15 - Mapeamento do espaço de entrada via função *kernel*.

Dentre as funções *kernel* mais usadas destacam-se: Polinômios, Funções de Base Radial (RBF) ou Gaussiana e Sigmóide, definindo diferentes máquinas de aprendizagem conforme Tabela 2.1. Nessa tabela, x representa os vetores de suporte, enquanto y representa os dados de teste.

Tabela 2.1: Funções de *Kernel* mais utilizadas com SVM

Tipo de <i>Kernel</i>	Função $K(x, x_i)$	Tipo de Classificador
Polinomial	$[(x * y) + 1]^d$	Máquina de Aprendizagem Polinomial
Gaussiano ou (RBF)	$\exp\left(-\frac{(x - y)^2}{2\sigma^2}\right)$	Rede RBF
Sigmoidal	$\tanh[\beta_0(x * y)] + \beta_1$	Perceptron de duas camadas

Segundo Hsu *et al.* [2003], há diversas razões para que ocorra um melhor desempenho de RBF em relação a outros tipos de *kernel* para aplicações utilizando imagens, podendo-se destacar a baixa dimensionalidade do espaço de características e menor complexidade dos cálculos. Neste caso, há apenas dois parâmetros que necessitam ser ajustados, tais como C e γ .

O parâmetro C funciona como um parâmetro de penalização que permite que SVM tolere erros de treinamento, criando uma margem menos rígida. Quando se aumenta o valor de C , aumenta-se o custo dos pontos de dados distribuídos no espaço de características, forçando a criação de um modelo mais preciso. Já o parâmetro *Gamma* (γ), determina a rigidez do hiperplano de separação ótima, ou seja, quanto maior o valor de *Gamma*, maior será a quantidade de vetores de suporte.

Ainda segundo Hsu *et al.* [2003] a utilização de validação cruzada se configura como uma excelente estratégia para determinar as melhores configurações de parâmetros C e γ no *kernel* RBF. A validação cruzada consiste na divisão do conjunto total de padrões em k partições (*k-folds*) com tamanhos aproximadamente iguais, para que o treinamento possa ser realizado k vezes, sendo que a cada treinamento uma das partições é deixada para teste e, as $k-1$ partições restantes são utilizadas para treinamento. Desta forma, o sistema é treinado exaustivamente, tornando o problema mais próximo de uma situação real.

O desempenho de SVM, assim como outros classificadores, pode ser medido através de diferentes métricas de avaliação. Na próxima seção são descritas as métricas de avaliação utilizadas neste trabalho.

2.3 Métricas de Avaliação

Neste trabalho é realizada a classificação de imagens digitais tipicamente encontradas na *Web* em duas classes, nudez e não-nudez (neutras). O resultado da classificação é medido através de métricas calculadas a partir da matriz de confusão, que é uma das ferramentas mais utilizadas para a avaliação de classificadores [Fukunaga 1990]. A matriz de confusão contém informações sobre os valores reais e os valores encontrados pelos classificadores, ou seja, contém estatísticas sobre os acertos e os erros dos classificadores. Na tabela 2.2, tem-se uma especificação da matriz de confusão para classificadores contendo apenas duas classes.

Tabela 2.2: Exemplo de uma matriz de confusão.

		Classificação	
		Nudez	Neutra
Real	Nudez	A	B
	Neutra	C	D

Nesta tabela as entradas estão associadas aos seguintes valores:

- **A** é o número de classificações corretas de amostras da classe nudez;
- **B** é o número de classificações incorretas de amostras da classe nudez;
- **C** é o número de classificações incorretas de amostras da classe neutra;
- **D** é o número de classificações corretas de amostras da classe neutra.

Há na literatura diversas métricas para avaliar o desempenho de um sistema de detecção de nudez. Este trabalho adota as métricas propostas por Wang *et al.* [1998] que são utilizadas na maioria dos trabalhos descritos na Seção 2.4. Essas métricas são definidas abaixo:

- Especificidade: definida como a razão entre o número de imagens neutras corretamente classificadas (**D**) para o número total de imagens neutras (**C + D**);

$$\text{Especificidade} = \frac{D}{C + D} \quad (2.4)$$

- Sensitividade: definida como a razão entre o número de imagens de nudez classificadas corretamente (**A**) para o número total de imagens de nudez (**A + B**).

$$\text{Sensitividade} = \frac{A}{A + B} \quad (2.5)$$

- Falso Negativo: definida como a razão entre o número de imagens de nudez incorretamente classificadas (B) para o número total de imagens de nudez ($A + B$);

$$Falso\ Negativo = \frac{B}{A + B} \quad (2.6)$$

- Precisão: definida como a razão entre o número de imagens classificadas corretamente ($A + D$) para o número total de imagens ($A + B + C + D$).

$$Precisão = \frac{A + D}{A + B + C + D} \quad (2.7)$$

Alguns trabalhos que utilizam conceitos definidos neste capítulo e relacionados ao trabalho descrito neste documento são descritos na Seção 2.4.

2.4 Trabalhos Relacionados

Nesta seção são discutidos alguns métodos de detecção de nudez em imagens. Vale ressaltar que, os resultados dos experimentos dos métodos são apresentados utilizando os conceitos de sensibilidade e especificidade. Outro fator importante a se destacar é que estes dados são usados apenas como informativos e não como comparação direta, uma vez que os recursos computacionais não são compatíveis, além de muitos trabalhos não disponibilizarem as bases de dados utilizadas. Antes, porém, são descritas estratégias de detecção de nudez em texto e na combinação de imagem e texto.

Conforme mencionado na introdução, a maioria dos métodos para reconhecimento de páginas contendo nudez na *Web* pode ser agrupada em três categorias: reconhecimento baseado em texto, reconhecimento baseado na combinação de texto e imagem e reconhecimento baseado em imagens. Independentemente da categoria, o objetivo principal é a busca por métodos com baixo custo computacional, critério fundamental para o desenvolvimento de sistemas *online* de detecção de nudez [Lopes *et al.* 2009].

Abordagens atuais para reconhecimento de textos com conteúdo ofensivo são baseadas principalmente em listas brancas e listas negras (*Blacklists e Whitelists*), frequência de palavras-chave e classificação de texto. As *blacklists* e *whitelists* são listas de sites que devem ser bloqueados ou permitidos, respectivamente. Listas negras geralmente são criadas por sites da *Web* através de uma análise manual para a classificação de sites como pertencentes a classes proibidas, tais como "Nudez" e "Violência". Sites também podem ser automaticamente incluídos em listas negras se seu domínio possuir palavras-chave como "sexo" ou "xxx". No caso de listas brancas, é gerada uma

lista de sites permitidos. Qualquer site não encontrado na lista é bloqueado. O principal problema com essas listas é o contínuo surgimento de novos sites, sendo difícil construir e manter listas atualizadas.

Algumas pesquisas têm sido desenvolvidas no sentido de combinar informações de texto e imagem para reconhecimento de páginas com conteúdo ofensivo. Belém [2006] apresenta duas abordagens para detecção de conteúdo ofensivo em imagens na *Web*, sendo uma baseada no conteúdo da imagem, mais especificamente nas cores presentes na mesma, e outra baseada nas evidências de texto associadas à imagem. A abordagem baseada em imagem, denominada SNIF (*Simple Nude Image Finder*), combina um algoritmo de extração de propriedades de cor, chamado BIC (*Border Interior pixel Classifier*), juntamente com o classificador SVM. O método proposto obteve 87,6% de especificidade e 98% de sensibilidade em uma base de dados contendo 2.000 imagens de nudez e 2.135 imagens neutras. Esta abordagem possui a vantagem de economizar espaço em disco por utilizar apenas uma característica baseada em cor. Já a abordagem baseada em texto mostrou-se superior à abordagem baseada em imagens, com resultados de sensibilidade e especificidade próximos a 99%. A tentativa de combinação das abordagens não obteve resultados satisfatórios e levou o autor a concluir que, quando houver texto associado a imagens a melhor solução é utilizar a abordagem baseada em texto e, quando houver somente imagens, o SNIF é a melhor alternativa. Além disso, o pouco uso da combinação de informações de texto e imagens se dá pela necessidade de construir dois modelos de classificação, um para cada tipo de fonte de informação.

Por outro lado, o reconhecimento automático de imagens de nudez tem sido alvo de diversas pesquisas. Os atuais métodos podem ser classificados como baseados em modelo (*Model-Based*), baseados em características (*Feature-Based*) e, baseados em região (*Region-based*) [Hu *et al.* 2009]. Os métodos baseados em modelo enfatizam a estrutura simétrica do corpo humano. Pode-se citar a abordagem proposta por Fleck *et al.* [1996], que constrói um modelo a partir da segmentação de *pixels* de pele, usando informações de cor e textura, para em seguida localizar todas as regiões conectadas que são candidatas a troncos e membros do corpo humano, tendo como base variados modelos geométricos do corpo humano. Quando as combinações dessas estruturas geométricas formam um corpo humano, a imagem é tratada como pertencente à classe nudez. Os autores utilizaram nos experimentos um conjunto de dados de 565 imagens de nudez e 4.289 imagens neutras. Os resultados mostram que este modelo obteve 52,2% de sensibilidade e 96,6% de especificidade, denotando que houve uma disparidade muito grande na classificação das amostras.

Já os métodos baseados em características enfatizam a extração de diversas características da imagem. Schettini *et al.* [2003] apresentam um modelo que utiliza características da borda, textura e espaço de cores *HSV*, criando um vetor para classificação usando *CART (Classification on Regression Trees)* e *SVM*. O processo experimental consistiu de 1.500 imagens, sendo 750 de nudez e as 750 restantes, neutras. Os resultados experimentais obtiveram 90,4% de sensibilidade e 88,4% de especificidade. Esse trabalho empregou uma grande combinação de características, sem realizar uma análise qualitativa das mesmas. Além disso, foram utilizadas bases de dados balanceadas, fator pouco comum em problemas reais.

Os métodos baseados em regiões extraem características para reconhecimento de imagens de nudez baseado na localização de regiões de pele. Esses métodos apresentam os melhores resultados para detecção de nudez e por isso são frequentemente usados na literatura.

Jiao *et al.* [2002] utilizam a transformação de *pixels* de *RGB* para o espaço de cores *YUV* e, posteriormente, *YIQ* (*Y* representa a luminância e *UV-IQ* a cromaticidade). O modelo proposto utilizou informações de cor e bordas para criar um vetor de características para posterior classificação das imagens através de *SVM*. O conjunto de dados foi constituído por 1.200 imagens de pessoas e 1.200 imagens diversas, obtendo 89,3% de sensibilidade e 90,6% de especificidade. Porém, apesar de utilizar informações de formas, esse trabalho analisou apenas imagens contendo nudez em que as pessoas estão em posição frontal, tornando o resultado pouco abrangente.

Duan *et al.* [2002] apresentam um novo método de detecção de pele baseado em propriedades da cor. O método primeiramente rotula os *pixels* referentes à pele como exemplos de treinamento para posteriormente, converter os valores de *pixels* de *RGB* para os espaços de cores *YUV* e *YIQ*. Assim, o modelo de distribuição de pele foi obtido e o vetor de características criado para classificação utilizando *SVM*. O conjunto de teste foi composto por 312 imagens de nudez e 710 imagens neutras. O método obteve 80,7% de sensibilidade e 90% de especificidade, usando apenas duas características. Porém, a quantidade de amostras utilizada foi pouco representativa.

Lee *et al.* [2007] propõem um algoritmo para detecção de nudez em imagens que utiliza detecção de pele, extraindo informações referentes à cor e propriedades geométricas e, posteriormente, detecção de faces para garantir a eficiência do modelo. O conjunto de treinamento foi constituído por 508 imagens de nudez e 482 imagens neutras, sendo classificadas utilizando a combinação de classificadores *SVM* e Redes Neurais. Este método obteve 86,4% de sensibilidade e 94,8% de especificidade. Apesar de terem aplicado um conjunto relevante de características, os

autores não realizaram verificação do impacto das características no classificador, fato que pode ocasionar uma taxa de classificação não otimizada. Verifica-se também que foi utilizado um conjunto de imagens de nudez maior que o conjunto de imagens neutras. Este fato não representa a realidade de tráfego na Internet, pois é possível observar que a quantidade de imagens de nudez é inferior à quantidade de imagens sem nudez.

Zhu *et al.* [2007] apresentam um modelo que combina espaço de cores como *HSV*, *YUV* e *YIQ* para controle de luminosidade e filtragem de pele para extração de informações referentes à cor. Em seguida, informações da textura são utilizadas através da matriz de co-ocorrência e da análise da estrutura geométrica dos seres humanos. Essa combinação de características foi usada pelo classificador SVM. Os autores usaram um conjunto de 400 imagens de nudez e 400 imagens diversas, sendo que foram obtidos 88,9% de sensibilidade e 84,6% de especificidade. Apesar da elevada taxa de classificação, não foi realizado um processo de análise de características e o modelo utiliza muitas informações baseadas na forma, que demandam alto esforço computacional.

Cao *et al.* [2002], apresentam um método para detecção de imagens de nudez que combina diversas propriedades de baixo nível como cor, textura, formas e outras informações baseadas no filtro de pele. As informações de cor são extraídas através do espaço de cor HSV. Já as informações de textura são extraídas da matriz de co-ocorrência e as evidências de formas por meio de momentos invariantes. O filtro de pele foi criado através da conversão de RGB para YCbCr. Nesse trabalho foi utilizado um algoritmo de classificação denominado Adaboost, que possui a capacidade de formar uma sequência de classificadores, em que os pesos dos atributos são atualizados dinamicamente de acordo com os erros na aprendizagem anterior. Utilizando uma base de dados que contém cerca de 1.200 imagens, o sistema foi capaz de alcançar uma precisão de até 86% sobre o conjunto de teste. É possível observar que foi realizado um processo de quantização do vetor de características através de um algoritmo denominado LVQ (*Learning Vector Quantization*), e que mesmo com a redução de características, a precisão do modelo permaneceu inalterada. Entretanto, não foi feito um processo de análise para determinar as características mais relevantes para a classificação de imagens de nudez.

Sevmili *et al.* [2011] apresentam um sistema que combina propriedades globais e locais para descrição de imagens de nudez e é composto por 4 componentes. O primeiro componente é a utilização de recurso local como a filtragem de *pixels* de pele, o segundo componente aborda a detecção de faces, o terceiro componente trata da extração de características globais baseadas em

cor, textura e formas e, o quarto componente é a classificação das imagens em estudo através de SVM. Nesse trabalho é analisada a combinação de cada característica global com a filtragem de pele (recurso local). As características baseadas em cor obtiveram os melhores resultados na detecção de nudez, com taxas de detecção de até 92,8% combinada com a filtragem de pele e, 91,4% sem a filtragem de pele, em uma base contendo cerca de 8.510 imagens. Os autores afirmam que o processo de eliminar imagens a partir da detecção de faces contribuiu para o aumento da eficácia na detecção de nudez, porém elevou o esforço computacional, uma vez que somente a detecção de pele é suficiente para obter resultados considerados satisfatórios.

A diferença entre os trabalhos apresentados nesta seção e a abordagem proposta nesta dissertação é que, em se tratando de conteúdo da imagem, os demais trabalhos utilizam normalmente um grande conjunto de características sem que seja realizado um processo de análise qualitativa das mesmas. Porém, a literatura em aprendizagem de máquina indica claramente que a seleção das características mais relevantes para um problema de classificação pode reduzir o espaço de armazenamento das características em disco e, conseqüentemente, o custo computacional para extração e classificação das mesmas, além de possibilitar aumento na taxa de detecção. Outra diferença deste trabalho é a aplicação do algoritmo de zoneamento em imagens para análise e tratamento independentes das características. As características são avaliadas em dois diferentes cenários: (1) sem uso de zoneamento, para verificarmos o impacto de cada característica no desempenho geral do classificador; (2) as características são analisadas em cada zona para podermos indicar as características mais relevantes para cada zona da imagem.

O trabalho de Sevmili *et al.* [2011] é o que mais se assemelha ao trabalho apresentado neste documento. Os autores realizam a combinação de propriedades globais e locais para descrição de imagens de nudez, configurando-se como um modelo baseado em regiões, utilizando o mesmo conjunto de características empregadas neste trabalho. No entanto, não é realizado um processo de análise de características, além da não utilização do algoritmo de zoneamento. Nesse trabalho, 5 (cinco) classes (normal, *topless*, esportes aquáticos, nudez e pornográficos) são investigadas, sendo que nesta dissertação apenas 2 (duas) classes (nudez e neutras) são consideradas.

A Tabela 2.3 sintetiza os métodos de detecção de nudez apresentados nesta seção, que influenciaram de alguma forma os experimentos a serem realizados nesta dissertação, pois se configuram como um guia indicando aspectos promissores e relevantes na área de detecção de nudez utilizando aprendizado de máquina.

Tabela 2.3: Sumarização dos modelos de detecção de nudez.

Publicação	Técnicas de AM	Propriedades	Informações da Base de Dados				Resultados (%)
			Fonte	Etnia	Luminosidade	Composição	
Fleck <i>et al.</i> [1996]	SVM	Cor, forma e textura (Global)	Internet	Caucasianos, africanos e asiáticos	Variada	Nudez: 565 Não-nudez: 4289	Sensitividade: 52,2% Especificidade: 96,6%
Schetinni <i>et al.</i> [2003]	DT e SVM	Cor e Textura (Global) Forma (Local)	Internet	Caucasianos, africanos, indianos	Não Informada	Nudez: 750 Não-nudez: 750	Sensitividade: 90,4% Especificidade: 88,4%
Jiao <i>et al.</i> [2002]	SVM	Cor e Formas (Local)	Internet	Caucasianos, asiáticos	Variada	Nudez: 1200 Não-nudez: 1200	Sensitividade: 89,3% Especificidade: 90,6%
Duan <i>et al.</i> [2002]	SVM	Cor (Local)	Não Informada	Caucasianos, asiáticos, europeus	Variada	Nudez: 312 Não-nudez: 710	Sensitividade: 80,7% Especificidade: 90%
Lee <i>et al.</i> [2007]	SVM e NN	Cor e Formas (Local)	Não Informada	Caucasianos, africanos, asiáticos	Controlada	Nudez: 508 Não-nudez: 482	Sensitividade: 86,4% Especificidade: 94,8%
Sevmili <i>et al.</i> [2011]	SVM	Cor, textura e forma (Global) Cor (Local)	FERET Dataset	Caucasianos, africanos, entre outros	Variada	Nudez: 6808 Não-Nudez: 1702	Precisão de 92,8%
Zhu <i>et al.</i> [2007]	SVM	Cor e formas (local) textura (global)	Internet	Não Informada	Não Informada	Nudez: 400 Não-Nudez: 400	Sensitividade: 88,9% Especificidade: 86%
Cao <i>et al.</i> [2002]	Adaboost	Cor, textura e forma (global) Cor e forma (local)	Internet	Não informada	Não Informada	1200 imagens	Precisão de 86%
Belém <i>et al.</i> [2006]	SVM	Textuais e Cor (Global)	Yahoo!	Caucasianos, africanos, entre outros	Variadas	Nudez: 2000 Não-Nudez: 2135	Sensitividade: 87,8% Especificidade: 98%

Capítulo 3

A Arquitetura ANDImage

Nesta seção é discutida a proposta de uma arquitetura para detecção de nudez denominada ANDImage (*Architecture for Nude Detection in Image*). Na Figura 3.1 é possível observar um modelo de arquitetura comumente usada na maioria dos sistemas de detecção de nudez. Essa arquitetura é geralmente composta por um módulo de detecção de pele, um extrator de características e um classificador responsável pelo julgamento das imagens.

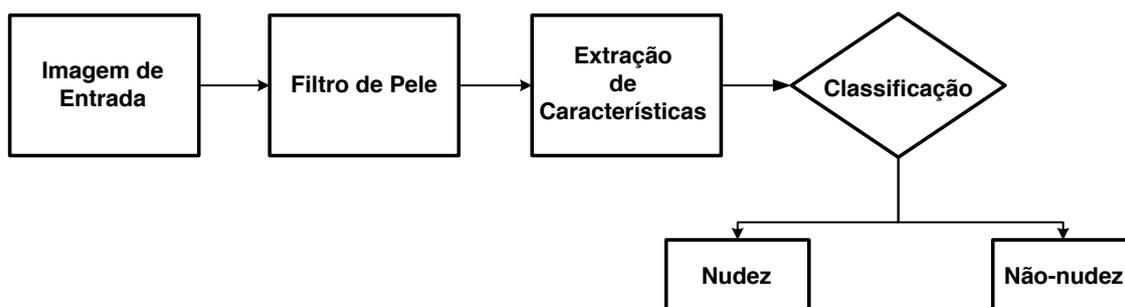


Figura 3.1: Arquitetura geral para sistemas de detecção de nudez.

Diferentemente dos sistemas atuais, a arquitetura ANDImage inclui um módulo de zoneamento. Como mostrado na Figura 3.2, essa arquitetura pode ser utilizada considerando-se os dois grupos de características mencionados no capítulo anterior, isto é, propriedades locais e globais. O sistema realiza a detecção de nudez em imagens através da combinação de características globais quando o módulo filtro de pele não é incluído na arquitetura, ou seja, não ocorre segmentação da imagem. Por outro lado, se o módulo filtro de pele é utilizado, então o sistema de detecção de nudez utiliza características locais (com segmentação na imagem).

A arquitetura ANDImage é basicamente composta por um componente normalizador, responsável pela padronização e análise das imagens obtidas na Internet. No caso de propriedades locais, um módulo de detecção de pele é utilizado para tratar a luminosidade nas imagens através do espaço de cor $YCbCr$, segmentando a imagem em busca de áreas de pele. Em seguida, tanto para a extração de características locais como globais, tem-se um módulo de zoneamento em imagens, que

divide a imagem em regiões para análise e tratamento independentes, adotando a hipótese de que em imagens de nudez há uma predominância de região de pele no centro da imagem. Por fim, um extrator de característica da imagem é gerado a partir de diversos descritores de cor, textura e forma e, posteriormente, usado como parâmetro de entrada para o componente classificador. Este último classifica a imagem em nudez ou neutra (não-nudez).

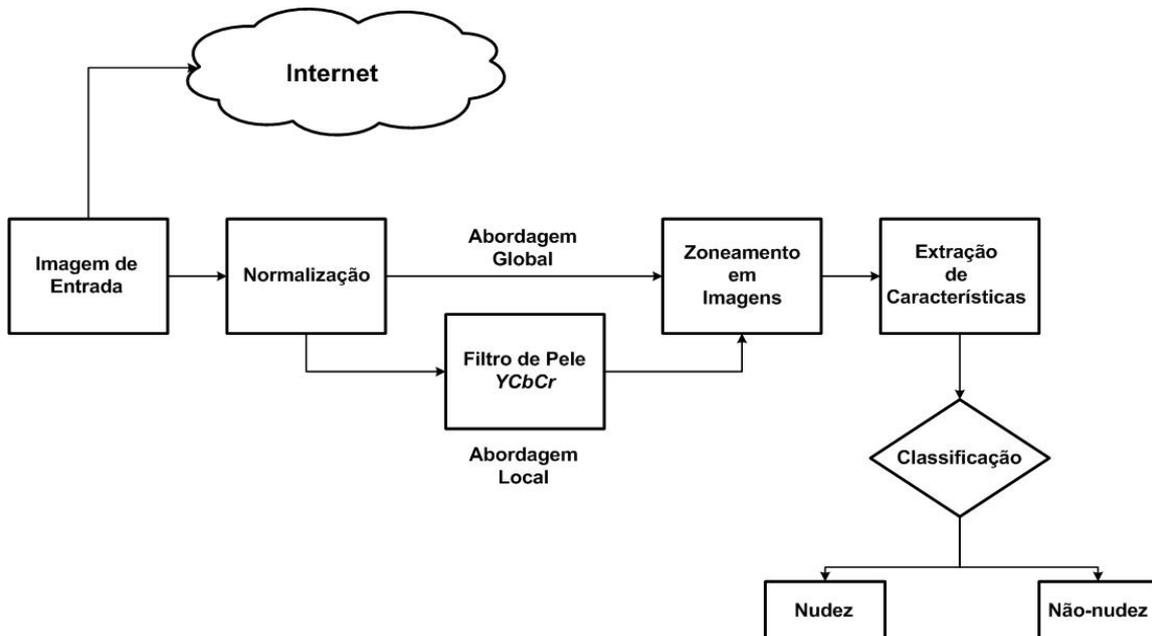


Figura 3.2: Visão geral da Arquitetura ANDImage.

Os principais componentes da arquitetura ANDImage são descritos em mais detalhes nas próximas seções.

3.1 Componente Normalizador

Este componente corresponde a uma camada que torna o sistema independente do formato gráfico a ser tratado, repassando ao módulo seguinte uma imagem totalmente padronizada, uma vez que na Internet as imagens estão dispostas em diversos formatos de arquivos (JPG, GIF, PNG) e resoluções, fato que dificulta o trabalho de análise de imagens e o desenvolvimento de um sistema automático de detecção de nudez. Logo, é necessário normalizar as imagens, ou seja, padronizá-las em uma mesma dimensão e formato. Conforme visto na Seção 2.1, o processo de normalização das imagens consiste em gerar um novo conjunto de amostras para aplicar a um determinado contexto a ser explorado.

3.2 Filtro de Pele

Como mencionado no capítulo anterior, as informações mais importantes em uma imagem estão representadas na cor, pois são relevantes para detecção e localização de objetos [Zhu *et al.* 2007]. Assim, após a obtenção da imagem normalizada, a segunda etapa da arquitetura concentra-se no processo de segmentação das imagens em duas regiões: (1) região de pele; e (2) região de não-pele, para características locais. Logo a detecção de pele é também objeto de estudo desta dissertação.

Como as imagens armazenadas em páginas *Web* podem representar cores de pele de pessoas de várias etnias e diferentes condições de luminosidade, torna-se necessário o desenvolvimento de técnicas robustas a essas propriedades para que o módulo de detecção de pele seja aplicado com sucesso. Segundo Lee *et al.* [2007], a pele possui a característica de absorver a luminosidade devido à melanina (substância que envolve e protege o cabelo e a pele humana). Logo, é importante o uso de um espaço de cor que isole e trate corretamente a luminosidade em imagens. Existem diversos espaços de cores, sendo que cada um contém particularidades e aplicações específicas. Neste trabalho, a detecção de pele ocorre a partir da conversão do espaço de cor *RGB* para o espaço *YCbCr*.

A escolha do espaço de cor *YCbCr* deve-se à propriedade que esse espaço possui de tratar a luminosidade em imagens. Além disso, como a pele humana contém uma gama restrita de valores de *pixels* no espaço de cor, conforme Kelly *et al.* [2007], o uso do espaço *YCbCr* proporciona as melhores taxas de detecção de pele, além de reduzir o custo computacional. Maiores detalhes sobre a implementação do filtro de pele a ser utilizado nesta dissertação são apresentados no Capítulo 4. Vale ressaltar que, a detecção de pele humana em imagens é uma etapa de fundamental importância para filtrar imagens que contenham nudez, pois os módulos subsequentes receberão informações da imagem resultante da aplicação do filtro de pele. Qualquer falha de implementação nesta etapa, será propagada para etapas posteriores, resultando em erros de classificação.

3.3 Zoneamento em Imagens

A região central é a mais representativa em imagens, enquanto que regiões mais externas representam cenários ou outros elementos menos importantes [Kalva *et al.* 2008]. Com base nesse fundamento, foi definida a seguinte hipótese: em imagens de nudez, a região central é composta predominantemente por regiões de pele. A partir desta suposição e para introduzir informação local

no vetor de características a ser gerado, foi adotado o conceito de zoneamento em imagens definido na Seção 2.1.

Neste trabalho, o zoneamento em imagens foi definido como um processo recursivo que divide a imagem toda em k sub-regiões (zonas), conforme descrito a seguir.

Seja I uma imagem, $NZ(I)$ o número total de zonas de I , Z uma certa zona de I , $FE(Z)$ o processo de *extração de características* de Z , $FV(Z)$ o *vetor de características* de Z e CP o processo de *classificação*. Baseado nestas definições, o algoritmo de zoneamento é definido da seguinte forma:

Algoritmo 1: Definição do Zoneamento em Imagens

$NZ(I) = k$;

Para toda *imagem* (I) faça;

Dividir I em NZ ;

Para toda *zona* (Z) faça;

$FE(Z)$;

$FE(I) = FE(I) + FE(Z)$;

Fim;

Fim;

$CP(FE(I))$;

O algoritmo de zoneamento é brevemente explicado como segue: inicialmente, a entrada é a imagem toda. Dependendo da dimensão, o algoritmo atribui um determinado número de zonas $NZ(I)$ para que a imagem seja dividida a partir do centro. A partir daí, o algoritmo auxilia no processo de extração de características para cada zona $FE(Z)$ com o objetivo de obter o vetor de características da imagem $FE(I)$, que é usado como parâmetro para o processo de classificação $CP(FE(I))$. Todo o processo objetiva possibilitar a confirmação da hipótese adotada neste trabalho.

Vale ressaltar que no trabalho de Kalva *et al.* [2008] um algoritmo de zoneamento foi utilizado no processo de busca e classificação de imagens diversas, com propriedades globais, ou seja, a extração de evidências na imagem não requer segmentação. Entretanto, esta técnica ainda não havia sido empregada ao domínio específico de detecção de nudez. Adicionalmente, além da utilização de propriedades globais, também foi utilizada segmentação para filtragem de regiões de pele para descrição de uma imagem de nudez.

3.4 Extração de características

Esta etapa consiste em gerar um vetor de características a partir dos descritores de cor, textura e forma, para posterior processo de seleção de características na classificação de imagens de nudez. Logo, um passo crucial para qualquer algoritmo de reconhecimento de padrões é que exista uma clara distinção entre as características extraídas. Através da análise de alguns trabalhos e de um grande número de imagens, este trabalho realiza a seleção de características de cor, forma e textura, através de uma comparação e combinação de abordagens que utilizam propriedades globais e propriedades locais.

A maioria dos sistemas de reconhecimento de objetos em imagens tende a usar tanto propriedades globais, quanto propriedades locais. A principal diferença entre propriedades globais e locais é a complexidade computacional, ou seja, a forma com que os dados são calculados e o armazenamento de espaço em disco. Dentre os descritores globais geralmente usados para detecção de pele e nudez, destacam-se *Global Color Histogram* (GCH), *Gray-Level Co-occurrence Matrix* (GLCM), *Edge Histogram Descriptor* (EHD). Já entre os descritores locais destacam-se *Local Color Histogram* (LCH) e *Scale Invariant Feature Transform* (SIFT).

3.4.1 Extração de Características de Cor

As características globais de cor utilizadas neste trabalho são baseadas no GCH, descrito na Seção 2.1.2.1, onde a partir dos dados coletados é possível extrair também informações referentes ao momento do histograma, como média e desvio padrão. O histograma de cor se configura como uma ferramenta útil para aplicações de detecção de pele e nudez, devido à sua simplicidade e à velocidade para computar os dados [Jones & Regh 2002]. Outra razão para o uso de evidências baseadas no histograma em detecção de nudez é o fato da pele formar *clusters*, ou seja, os *pixels* de pele tendem a agrupar-se em uma determinada região e quando o espaço de cor é quantizado, as cores também tendem a agrupar-se. Schetinni *et al.* [2003] e Sevmili *et al.* [2011] são exemplos de autores que utilizam essas evidências para auxílio no processo de detecção de nudez. Detalhes sobre essas características são descritos a seguir.

Neste trabalho foram computados histogramas quantizados em 32, 64, 128 e 256 cores. Logo, no histograma quantizado em 64 cores, o vetor de características é composto por 198 elementos, sendo 64 informações da cor, mais a média e desvio padrão para cada banda. As características de cor utilizadas são:

- *Média*: Calculada pelo somatório dos valores de níveis de cinza, nc , de todos os n ($n = nlin * ncol$) *pixels* da imagem divididos pelo número total de *pixels*. É denotada na seguinte fórmula:

$$\mu_i = \frac{\sum_{k=1}^n nc_{ki}}{n} \quad (3.1)$$

- *Desvio Padrão*: É a raiz quadrada da variância e indica a homogeneidade, ou heterogeneidade, de uma imagem digital, sendo denotada pela fórmula:

$$std = \sqrt{\frac{\sum_{k=1}^n (nc_{ki})^2}{n - 1}} \quad (3.2)$$

3.4.2 Extração de Características de Textura

Neste trabalho, as características de textura são extraídas a partir de uma matriz de co-ocorrência (GLCM), descrita na Seção 2.1.2.1. Segundo Zhu *et al.* [2007] a utilização de características de textura baseadas na GLCM para detecção de pele e nudez se justifica pela correção na classificação de *pixels* de pele que, mesmo sendo realizado o controle de luminosidade, poderão ser confundidos com diversos objetos. Portanto, assume-se que a textura da pele é diferente dos demais objetos. Cao *et al.* [2002] e Zhu *et al.* [2007] são exemplos de autores que utilizam a matriz de co-ocorrência para calcular a textura das imagens, através de características como contraste, energia, correlação e homogeneidade, para descrever a imagem que contém nudez. Como visto na Figura 2.7 da Seção 2.1.2.1, é necessário definir uma relação espacial em termos de distância d e ângulo θ para que possa ser construída a matriz de vizinhança. Este trabalho encontrou a melhor distância para $d = 1$, calculada em 8 direções e $\theta = [0, 45, 90, 135, 180, 225, 270, 315]$. Logo o vetor de características de textura é composto por 32 componentes. As quatro métricas obtidas estão listadas abaixo:

1. *Contraste*. Retorna uma medida do contraste entre as intensidades de um *pixel* analisado e do *pixel* vizinho. A comparação é realizada em todos os *pixels* da imagem. Para uma imagem constante, o contraste é 0 (zero).

$$Con = \sum_{i,j} (i - j)^2 P(i, j) \quad (3.3)$$

2. *Correlação*. Retorna uma medida de quão correlacionado está um *pixel* com o seu vizinho. A comparação é feita em todos os *pixels* da imagem. A correlação é 1 para uma imagem totalmente correlacionada.

$$Corr = \frac{\sum_{i,j} (i - \mu_i) (j - \mu_j) P(i, j)}{\sigma_i \sigma_j} \quad (3.4)$$

3. *Energia*. Conhecida também como Segundo Momento Angular calcula a uniformidade da textura na matriz de co-ocorrência de tons de cinza, ou seja, avalia as repetições de pares de *pixels* em tons de cinza de uma imagem. A energia possui valor 1 para uma imagem constante, onde somente *pixels* similares estão presentes na imagem, ou 0 para heterogênea, quando há grande disparidade entre os *pixels*.

$$Enrg = \sum_{i,j} P^2(i, j) \quad (3.5)$$

4. *Homogeneidade*. Retorna um valor que representa a proximidade da distribuição dos elementos em relação à diagonal da matriz de co-ocorrência dos tons de cinza.

$$Hmg = \frac{\sum_{i,j} P(i, j)}{1 + (i - j)} \quad (3.6)$$

3.4.3 Extração de Características de Formas

Neste trabalho é utilizado o descritor *Edge Histogram Descriptor*, proposto por Won *et al.* [2002], que foi descrito na Seção 2.1.2.1. Schetinni *et al.* [2003], Jiao *et al.* [2002] e Sevmili *et al.* [2011] são exemplos de autores que utilizam estas evidências no processo de detecção de nudez. O processo de extração destas características está resumido abaixo:

1. Conversão da imagem de entrada para níveis de cinza;
2. Detecção de bordas através da aplicação de operadores de Sobel;
3. Aplicação de algoritmos de detecção de linhas horizontais e verticais, linhas inclinadas e pontos isolados, para computar as características na imagem, conforme Figura 3.3.

2	-1	-1
-1	2	-1
-1	-1	2

(a)

-1	-1	-1
2	2	2
-1	-1	-1

(b)

-1	2	-1
-1	2	-1
-1	2	-1

(c)

-1	-1	2
-1	2	-1
2	-1	-1

(d)

-1	-1	-1
-1	8	-1
-1	-1	-1

(e)

Figura 3.3 – Máscaras para extração de formas: (a) linhas 135°; (b) linhas horizontais; (c) linhas verticais; (d) linhas 45°; (e) pontos.

Para efeito de esclarecimento, neste trabalho é utilizado apenas o conceito do histograma de arestas, sem que seja feita normalização para extração de características locais. Sendo assim, apenas os recursos globais são usados para descrição da imagem, contendo apenas 5 atributos, sendo que cada atributo armazena a quantidade de ocorrências de cada uma das formas na imagem. O recurso local é utilizado apenas quando o algoritmo de zoneamento é aplicado.

3.4.4 Extração de Características Baseadas no Filtro de Pele

O processo de extração de características locais é realizado a partir da segmentação da imagem obtida através do filtro de pele. Verifica-se que a maioria dos trabalhos mencionados na Seção 2.4 utiliza características baseadas em filtro de pele, como: quantidade de *pixels* de pele conectados e porcentagem de *pixels* de pele na imagem. Tais evidências foram adotadas como exemplos para mensurar a relevância de características locais no processo de classificação de imagens de nudez. A descrição das características é dada abaixo:

- *Quantidade de pixels de pele conectados.* Essa característica fornece informações referentes ao número de *pixels* de pele conectados N_s e pode ser obtida através da seguinte fórmula:

$$N_s = \sum P(L) \quad (3.7)$$

onde $P(L)$ é o número de *pixels* rotulados como pele.

- *Porcentagem de pele.* Essa métrica fornece informações sobre a proporção de segmentos de pele na imagem P_s e pode ser obtida pela seguinte fórmula:

$$P_s = \frac{N_s}{N_t} \quad (3.8)$$

onde N_s é o número de *pixels* de pele e N_t é o número de *pixels* total da imagem.

Vale ressaltar que a extração de informações geométricas da imagem só é necessária quando a imagem não é dividida em zonas. Ao aplicar-se o algoritmo de zoneamento em uma imagem, a região de interesse geralmente é dividida e, portanto, torna-se difícil computar suas informações de borda, área e perímetro. Diante desse contexto, optou-se por não utilizar informações geométricas neste trabalho.

3.5 Classificação

A última etapa da arquitetura ANDImage corresponde ao processo de classificação das imagens que tem como finalidade, a partir da leitura do vetor de características, distinguir as imagens que pertencem à classe nudez das imagens que pertencem à classe não-nudez. Vale ressaltar que o sucesso nas etapas anteriores da arquitetura (normalização, filtro de pele, zoneamento e extração de características) é imprescindível para que a etapa de classificação apresente elevada taxa de acerto. Nesta etapa, busca-se também analisar o potencial de influência de cada característica no desempenho de classificação das imagens. O objetivo é que seja atribuída a relevância de cada característica no processo de detecção de nudez, para que essa informação seja utilizada como referência para aplicações futuras.

Diversos classificadores podem ser utilizados para desempenhar o papel representado na Figura 3.4. Como o processo de classificação de imagens é normalmente supervisionado, ou seja, os dados precisam estar rotulados, classificadores como SVM, k-NN, Redes Neurais e árvores de decisão são candidatos a desempenhar esta função.

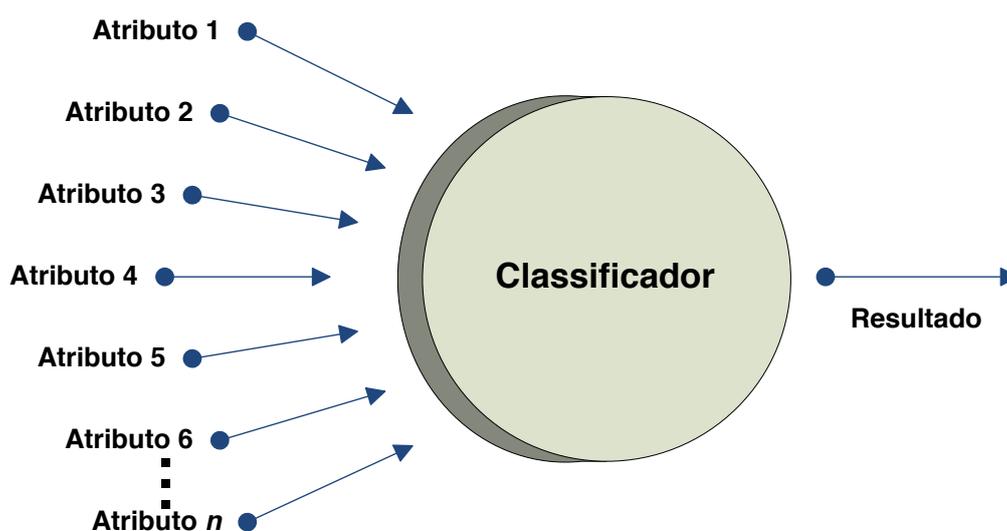


Figura 3.4 – Um classificador recebe o vetor de características como entrada e produz uma classe na saída.

Neste trabalho é utilizada a técnica de classificação SVM, descrita na Seção 2.2.4. Essa técnica foi escolhida por apresentar elevadas taxas de acerto em problemas de detecção de pele e detecção de nudez [Kelly *et al.* 2007]. Os experimentos realizados neste trabalho e os resultados obtidos serão descritos no próximo capítulo. É importante mencionar que o processo descrito no componente classificador foi realizado através do ambiente Weka [2010]. O Weka (*Waikato Environment for Knowledge Analysis*) é uma coleção de algoritmos de aprendizagem de máquina para resolução de problemas de associação, classificação, regressão e *clustering*. Desenvolvido em Java pela universidade de Waikato, Nova Zelândia, é uma ferramenta gratuita (*General Public License*) com interface gráfica simples.

Capítulo 4

Experimentos e Análise dos Resultados

Neste capítulo são apresentados detalhes sobre o desenvolvimento da arquitetura ANDImage, o protocolo utilizado para a realização dos experimentos, incluindo informações sobre as bases de imagens de treinamento e teste, parâmetros e configurações do classificador SVM, bem como os resultados obtidos através da análise do impacto das características.

Primeiramente, informações sobre a base de dados, desenvolvimento da arquitetura ANDImage e ajustes de parâmetros de SVM são descritos no protocolo experimental. Vale ressaltar que os experimentos realizados neste trabalho são baseados na análise de características globais e locais, cujas definições foram apresentadas na Seção 2.1.2. Em seguida, os resultados dos experimentos são apresentados, envolvendo primeiramente as características globais, ou seja, sem filtro de pele nem zoneamento. Posteriormente, são descritos os resultados obtidos com filtro de pele, sem aplicação do zoneamento. Por fim, as características globais e locais são combinadas através do algoritmo de zoneamento para que seja possível, não somente exibir os resultados, mas também, apresentar uma análise minuciosa sobre o comportamento das características em cada zona da imagem. O objetivo é a seleção das características mais relevantes para o processo de detecção de nudez, a fim de reduzir o espaço de características para que o tempo e o custo sejam otimizados, assim como a precisão do modelo.

O processo de análise das características segue os seguintes passos em cada grupo:

- Análise isolada: para que a relevância individual seja obtida;
- Um menos resto: cada característica é retirada para que se possa avaliar a precisão das características restantes;
- Subconjuntos: são testadas todas as combinações possíveis de características. Neste processo, é possível observar se uma determinada característica que não forneceu dados

suficientes para o aprendizado de forma isolada, possui relevância ao ser usada em subconjuntos.

Ao final, os 3 grupos de evidências são combinados. As características que, tanto na análise isolada quanto em subconjuntos, não fornecerem dados suficientes para o aprendizado do classificador são consideradas irrelevantes. Portanto, através do processo de análise de características pretendemos indicar o melhor subconjunto de evidências globais para descrição de imagens de nudez. O melhor subconjunto de características será aquele que obtiver resultados superiores aos produzidos com o conjunto de todas as características ou que obtiver maior aproximação, tendo em vista que ao final será verificado se a redução do espaço influencia no tempo de processamento.

Todos os experimentos foram realizados em uma estação de trabalho Core i3, 2,4 MHz, com 4GB de memória e um disco SATA com 320 GB de espaço de armazenamento. Os algoritmos para normalização, filtro de pele e extração de características foram executados através da ferramenta Matlab. Vale ressaltar que este trabalho adota como métricas de desempenho os conceitos de sensibilidade e especificidade, detalhados na Seção 2.3.

4.1 Protocolo Experimental

Nesta seção são fornecidas informações sobre a aquisição e composição da base de dados, bem como detalhes a respeito do desenvolvimento da arquitetura ANDImage e configuração de ajustes de parâmetros do classificador SVM.

4.1.1 Base de Dados

A base de dados empregada nos experimentos é composta por 5.360 imagens, sendo que 4.135 imagens (2.000 ofensivas e 2.135 não-ofensivas) foram utilizadas nos experimentos da dissertação de mestrado de Belém [2006]. Adicionalmente, 1.225 imagens adquiridas do diretório de Internet Yahoo! cedidas pelo grupo de Banco de Dados e Recuperação da Informação (BDRI) – UFAM, também foram utilizadas. Essas amostras foram adicionadas porque as imagens contidas na base cedida por Belém [2006] estão majoritariamente relacionadas com conteúdo ofensivo, mais abrangente do que somente nudez. Além disso, houve necessidade de aumentar a quantidade de amostras para tornar a base de dados mais representativa. As imagens de nudez da base Yahoo! foram rotuladas manualmente. Esse repositório contém cerca de 103.000 imagens não

categorizadas.

Conforme Figura 4.3, a base de dados usada em nossos experimentos contém 2.250 imagens nudez e 2.510 imagens de não-nudez. As imagens de nudez, conforme descrição feita na Seção 1.2, são imagens que exibem pessoas sem roupas ou com pouca roupa ou, até mesmo com exposição de partes íntimas do corpo humano (vide Figura 4.1). Buscou-se também incluir nesta classe, cenas de nudez que contenham pessoas de diferentes etnias (africanos, caucasianos, asiáticos, indianos, europeus) sob diversas condições de luminosidade (ambientes internos e externos, noite e dia) para validar o módulo de detecção de pele. Já as imagens que não contém nudez incluem animais, carros, motos, aviões, ambientes diversos e pessoas praticando esportes, com vestimentas, conforme Figura 4.2. Logo, todas as imagens foram selecionadas e rotuladas manualmente.



Figura 4.1: Exemplos de imagens de nudez.



Figura 4.2: Exemplos de imagens de não-nudez.

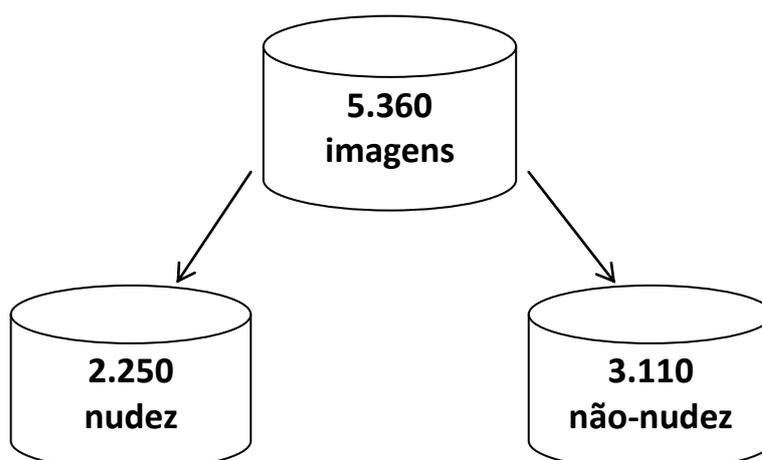


Figura 4.3: Divisão da base de dados usada nos experimentos.

Para análise de desempenho do classificador, a base de dados foi dividida em base de treinamento e base de teste. O conjunto de treinamento foi composto por 2.680 imagens, sendo 1.340 imagens de nudez e 1.340 imagens de não-nudez, totalizando 50% do conjunto inicial de imagens. A base de teste foi constituída pelas 2.680 imagens restantes, sendo 910 imagens ofensivas e 1.770 imagens não-ofensivas. Verifica-se que no conjunto de treinamento, a quantidade de amostras de imagens ofensivas é igual à quantidade de amostras de imagens não ofensivas. A base foi dividida desta forma para que o fenômeno conhecido como aprendizagem viciada seja minimizado. Este fenômeno ocorre quando o algoritmo de aprendizagem adapta-se ao conjunto de treinamento a ponto de memorizar ruídos e peculiaridades deste, ocorrendo uma generalização inferior ao esperado [Santos 2011].

4.1.2 Desenvolvimento da Arquitetura ANDImage

Com a finalidade de desenvolver um sistema de detecção de nudez em imagens, foi projetada e desenvolvida a arquitetura descrita no capítulo 3, que é composta pelos seguintes módulos:

- Componente normalizador;
- Filtro de pele;
- Zoneamento em imagens;
- Extrator de características;
- Componente classificador;

Como as informações do processo de extração de características foram fornecidas no capítulo anterior, nesta seção serão apresentadas apenas informações sobre o processo de normalização, aplicação do filtro de pele e zoneamento.

Normalização

Neste trabalho, foram aplicadas técnicas para redução e padronização do formato de entrada das imagens para estabelecer um parâmetro de geração do vetor de características a ser interpretado pelo classificador. Foram utilizadas diversas configurações para particionar as imagens (32x32, 64x64, 128x128, 256x256). Por fim, as imagens foram padronizadas em um valor aproximado de 256 colunas por 256 linhas. Então, sejam g a imagem original e h a imagem re-dimensionada, (Xg, Yg) e (Xh, Yh) as dimensões das imagens originais e re-dimensionadas, a transformação de g para h pode ser expressa pela seguinte equação:

$$h(x, y) = g\left(\text{INT}\left(\frac{x * Xg}{Xh}\right), \text{INT}\left(\frac{y * Yg}{Yh}\right)\right), \forall x \in \{0, \dots, Xh\}, y \in \{0, \dots, Yh\} \quad (4.1)$$

Filtro de Pele

Para o desenvolvimento de um módulo de detecção de pele robusto a variações de tonalidade de pele devido à luminosidade, foi necessário utilizar um conjunto de amostras de zonas de pele, com dimensões de 30x30 *pixels* cada, extraídas de diversos pontos de faces de pessoas de diversas etnias (caucasianos, africanos, europeus, asiáticos, indianos, entre outros) e sob diferentes condições de luminosidade, conforme exibido na Figura 4.4. As imagens foram convertidas do espaço de cor RGB para $YCbCr$. Em seguida, através da análise de histograma, os valores médios de cromaticidade (componentes Cb e Cr) das imagens foram obtidos. Por fim, após esta análise foi observado que os valores de *pixels* referentes à pele variam em Cb de 77 a 127 e em Cr , de 133 a 173.



Figura 4.4: Exemplos de amostras de pele utilizadas na obtenção dos valores de cromaticidade.

A Figura 4.5 mostra um exemplo de imagens e seus correspondentes resultados após a aplicação do filtro de pele. Os resultados do filtro de pele foram publicados em Santos *et al.* [2011]

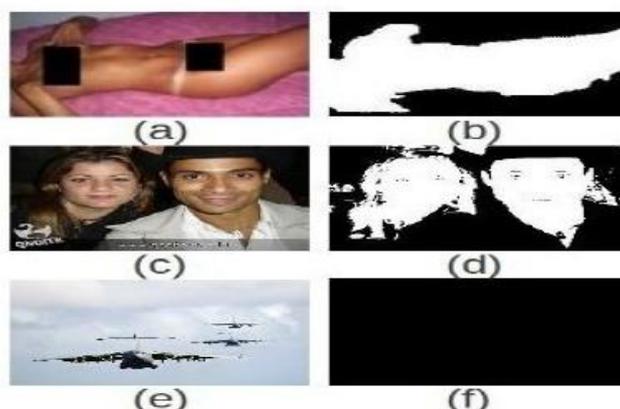


Figura 4.5: Imagens que demonstram o resultado do filtro de pele, onde (a) representa uma imagem original de nudez, (b) o resultado do filtro de pele na imagem de nudez, (c) imagem de pessoas com roupa, (d) o resultado do filtro de pele para pessoas com roupa, (e) imagem original que não contém pele e, (f) o resultado do filtro na imagem que não contém pele.

É importante observar nas imagens da Figura 4.5 que o cabelo é confundido com região de pele, este fato é ocasionado pela concentração de melanina que absorve a luminosidade, substância que é mais abundante em pessoas de raça negra [Jones & Regh 2002].

Zoneamento em Imagens

Conforme definido na Seção 3.3, o algoritmo de zoneamento foi utilizado para auxiliar no processo de análise de características e confirmar a hipótese de que em imagens de nudez há uma predominância de regiões de pele no centro da imagem.

Neste trabalho, a definição do número de zonas está relacionada diretamente com o processo de normalização de imagens. Como as imagens foram padronizadas em um tamanho aproximado de 256 colunas por 256 linhas, o algoritmo de zoneamento atribuiu três zonas para a imagem. Vale ressaltar que, diversos testes foram realizados para a definição do número exato de zonas para cada configuração de imagem, pois quando as imagens são reduzidas a um tamanho de 64x64, há uma considerável perda de dados e uma aplicação do algoritmo de zoneamento é desnecessária. Através da análise de características, será possível observar e constatar tais afirmações, pois em geral, a zona mais externa da imagem, fornece dados imprecisos. Os resultados do zoneamento e análise de características foram submetidos à publicação em Santos *et al.* [2012]

A Figura 4.6 exhibe exemplos de zoneamento obtidos através da divisão da imagem em três regiões retangulares concêntricas e proporcionais à sua dimensão.

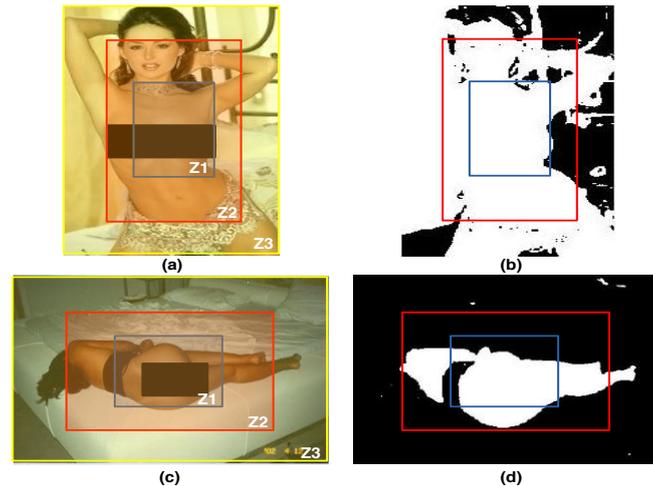


Figura 4.6 – Imagens que demonstram o resultado da aplicação do zoneamento, em que (a) representa o zoneamento em uma imagem de nudez com orientação vertical, (b) o resultado do zoneamento em imagem de nudez na vertical após a aplicação do filtro de pele, (c) o zoneamento em imagem de nudez com orientação na horizontal, (d) o resultado do zoneamento em imagens de nudez na horizontal após o filtro de pele.

4.1.3 Ajuste de Parâmetros de SVM

Como mencionado na Seção 2.2.4, SVM possui diferentes funções de *kernel*, tais como, Polinomial, RBF e Sigmóide, que precisam ser ajustadas durante a fase de treinamento. Neste trabalho, inicialmente foram testados diferentes tipos de *kernel*, porém o *kernel* RBF obteve melhor desempenho.

Seguindo as especificações de Hsu *et al.* [2003], validação cruzada foi utilizada durante a fase de treinamento para que os valores de C e γ fossem ajustados. Os valores de C variaram no intervalo de 10^{-3} a 10^3 . A Tabela 4.1 exibe os resultados de ajuste do parâmetro de penalidade (C) durante a fase de treinamento, e suas respectivas influências na quantidade de vetores de suporte e determinação do modelo SVM. É importante destacar que a Tabela 4.1 mostra a precisão do classificador tanto com o uso de zoneamento, quanto sem zoneamento.

Tabela 4.1 – Treinamento e ajuste do parâmetro de penalidade (C) em SVM para análise da quantidade de vetores de suporte e influência desse parâmetro na precisão do classificador.

Vetores de Suporte		Fator de Penalidade (C)	Precisão	
Sem zoneamento	Com zoneamento		Sem zoneamento	Com zoneamento
2680	2680	0,001	63,23%	76,32%
2680	2680	0,01	67,94%	78,23%
2439	1764	0,1	82,31%	94,10%
1478	769	1	84,85%	96,64%
1048	363	10	87,16%	97,72%
864	207	100	88,05%	98,61%
751	153	1000	87,41%	98,16%

Alpaydim [2010] atenta para o ajuste do fator de penalidade, pois se o valor de C for muito elevado, os pontos de dados não separáveis poderão sofrer uma penalidade maior e, posteriormente, ocasionar aprendizagem viciada. Caso contrário, poderá ocorrer *underfitting*⁹. É possível observar nesta tabela que ao atribuímos um valor muito baixo para o fator de penalidade, a quantidade de vetores de suporte foi elevada, enquanto a precisão do modelo não se mostrou satisfatória. Entretanto, ao aumentarmos o valor de C , foi verificado que tanto a quantidade de vetores de suporte quanto a precisão aumentaram. Diante desses resultados, foi utilizado $C=100$ como o melhor parâmetro. Esse valor está destacado na Tabela 4.1.

Em se tratando de ajustes dos valores de *Gamma* (γ), estes variaram no intervalo de 10^{-2} a 10^2 , conforme protocolo experimental utilizado por Hsu *et al.* [2003]. Uma razão para utilização desta estratégia é evitar buscas exaustivas, pois apenas alguns valores no intervalo são testados. Outra razão é que devido à baixa quantidade de parâmetros que precisam ser ajustados no *kernel* RBF, o tempo computacional é reduzido. Se for observada uma disparidade na precisão em um determinado intervalo, faz-se necessário realizar uma busca mais aprofundada neste intervalo. A Tabela 4.2 exibe os resultados de ajuste do parâmetro *Gamma* (γ) durante a fase de treinamento, e suas respectivas influências na quantidade de vetores de suporte e determinação do modelo SVM. Novamente a precisão do classificador é analisada com e sem zoneamento.

Tabela 4.2 – Treinamento e ajuste de parâmetro *Gamma* (γ) em SVM para análise da quantidade de vetores de suporte e influência de tais parâmetros na precisão do classificador.

Vetores de Suporte		Gamma (γ)	Precisão	
Sem zoneamento	Com zoneamento		Sem zoneamento	Com zoneamento
207	864	0,01	88,05%	98,61%
180	822	0,02	88,47%	98,65%
170	787	0,03	88,95%	98,67%
177	777	0,04	89,66%	98,69%
186	764	0,05	89,81%	98,58%
191	754	0,06	89,92%	98,47%
198	770	0,07	90,07%	98,77%
199	774	0,08	89,81%	98,65%
209	764	0,09	89,66%	98,65%
216	759	0,1	89,62%	98,56%
1078	1366	1	88,35%	97,76%
2679	2674	10	74,25%	70,63%
2678	2674	100	52,38%	51,75%

Durante o ajuste do parâmetro γ foi possível observar que ao aumentarmos os valores do

⁹ *Underfitting* ocorre quando o conjunto de amostras é pouco representativo, ou seja, elementos de grande importância são desconsiderados ou possuem menor peso que o ideal, fazendo com que o classificador cubra uma extensão menor que a adequada.

parâmetro, a quantidade de vetores de suporte também aumentava e o desempenho do classificador reduzia significativamente. Verificou-se também que ao variarmos os valores exponencialmente, houve uma disparidade entre 10^{-2} e 10^{-1} . Logo, foi realizada uma busca mais aprofundada neste intervalo e foi definido $\gamma=0,07$ como melhor valor deste parâmetro. A linha destacada na Tabela 4.2 mostra os resultados obtidos com $\gamma=0,07$.

Foi possível observar durante os experimentos que a quantidade de vetores de suporte afeta diretamente a complexidade de tempo e a precisão na construção do modelo SVM, pois quando há muitos vetores de suporte, o desempenho do classificador aumenta e, conseqüentemente, a velocidade para processar os dados e classificar corretamente as amostras também aumenta. Outra análise constatada é que a quantidade de vetores de suporte está ligada à dimensão do vetor de características, pois nestes experimentos, ao aumentarmos a quantidade de características através da aplicação do algoritmo de zoneamento em imagens, houve uma diminuição significativa da quantidade de vetores de suporte e aumento da precisão do modelo. A Tabela 4.3 exibe o resultado da validação cruzada na fase de treinamento, onde a base foi dividida em 10 *folds*. É importante destacar que a tabela foi obtida utilizando o *kernel* RBF, com $C = 100$ e $\gamma = 0,07$.

Tabela 4.3 – Resultados obtidos na validação cruzada durante o treinamento.

Validação Cruzada		
Partição	Sem zoneamento	Com zoneamento
F1	89,2%	98,2%
F2	90,1%	98,9%
F3	90,7%	98,9%
F4	90,1%	98,9%
F5	90,3%	98,9%
F6	90,7%	99,3%
F7	89,6%	99,3%
F8	90,1%	99,3%
F9	90,1%	97,4%
F10	89,2%	98,6%
Média	90%	98,8%

4.2 Resultados

Nesta seção são discutidos os resultados obtidos na etapa de classificação, com ênfase nas comparações das abordagens globais e locais para detecção de nudez. Além disso, são apresentados os resultados da seleção de subconjuntos de características obtidos através de uma análise qualitativa das características, conhecida na literatura como *Wrapper* [Kohavi & John 1997]. Nesta técnica, a relevância das características é avaliada em função da precisão do classificador. Nestes experimentos, cada característica é analisada isoladamente para que seja determinado o impacto

individual no desempenho do classificador. Em seguida, diversas combinações são testadas para que seja determinado o melhor compromisso entre o conjunto de características e a precisão do classificador. Como consequência, tem-se a redução da dimensão do espaço de entrada e a minimização da influência de ruídos no classificador para que sejam obtidos resultados mais precisos.

Por fim, também foram comparadas as características através da aplicação do algoritmo de zoneamento para que seja analisado o impacto das características em cada zona da imagem. Desta forma, busca-se confirmar a hipótese de que em imagens de nudez há uma predominância de *pixels* de pele no centro da imagem.

4.2.1 Análise das Características Globais

Como visto na Seção 2.1.2, características globais são evidências que descrevem a imagem sem que seja realizada segmentação. Neste trabalho, as características globais foram divididas em 3 (três) grandes grupos: cor, forma e textura.

Características de Cor

Iniciamos o processo de análise com as características de cor. Como mencionado anteriormente, neste trabalho foram utilizadas características baseadas no histograma da imagem. Conforme especificação da Tabela 4.4, foram computados histogramas quantizados em 32, 64, 128 e 256 cores. O histograma quantizado em 64 cores obteve o melhor desempenho, além de reduzir significativamente a dimensão do espaço, uma vez que se faz necessário computar o histograma para cada um dos 3 componentes do espaço RGB. Neste caso específico, o vetor de características é composto por 198 elementos, sendo 64 informações da cor, mais a média e o desvio padrão para cada componente.

Tabela 4.4 – Resultados para classificação de imagens de nudez usando diferentes quantizações de cor.

Quantidade de cores	Sensitividade	Especificidade	Precisão
32 cores	70,2%	89,2%	81,3%
64 cores	82,6%	93,3%	89,7%
128 cores	77,3%	90,6%	84,6%
264 cores	78,7%	91,4%	86,5%

A Tabela 4.5 mostra resultados obtidos por SVM na base de teste ao ser treinado com as características de cor (*histograma, média e desvio padrão*).

Tabela 4.5 – Resultados obtidos por SVM na base de teste com as características de cor. As características foram extraídas sem aplicação do módulo de zoneamento em imagens.

Característica de cor	Sensitividade	Especificidade	Precisão
<i>Histograma</i>	81,4%	93,1%	89,1%
<i>Média</i>	0%	100%	66%
<i>Desvio Padrão</i>	73,4%	89,4%	85,3%
<i>Histograma + Média</i>	81,4%	93,2%	89,2%
<i>Histograma + Desvio Padrão</i>	82,3%	93,2%	89,5%
<i>Média + Desvio Padrão</i>	76,4%	93,2%	87,6%
<i>Todas</i>	82,6%	93,3%	89,7%

Analisando o conjunto de características de cor isoladamente, é possível observar através da Tabela 4.5 que o *histograma* é capaz de produzir precisão de aproximadamente 89,1%. Verifica-se também que a *média* isoladamente não é suficiente para descrever as imagens, pois esta característica produziu elevada classificação incorreta de imagens de nudez. Isso se dá pela pouca representatividade das informações repassadas ao classificador, o qual não é capaz de aprender a distinguir as classes.

A análise de características de cor resultou na seleção do subconjunto composto pela combinação do *histograma* com o *desvio padrão*. Este subconjunto produziu cerca 82,3% de sensibilidade (749 imagens de nudez classificadas corretamente), 93,2% de especificidade (1649 imagens neutras classificadas corretamente) e, precisão de 89,5% (2398 imagens classificadas corretamente). Este desempenho foi semelhante às taxas produzidas pelo conjunto composto por todas as características de cor.

Características de Textura

Em um segundo momento, as características de textura foram analisadas. É importante mencionar que neste trabalho foi adotada a GLCM para extração dessas evidências. A Tabela 4.6 exhibe resultados da aplicação de características de textura definidas na Seção 3.4.2: *energia, contraste, homogeneidade e correlação*.

Tabela 4.6 – Análise das características de textura. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, sem a aplicação do zoneamento em imagens.

Característica de textura	Sensitividade	Especificidade	Precisão
<i>Contraste</i>	0%	100%	66%
<i>Energia</i>	0%	100%	66%
<i>Homogeneidade</i>	50,3%	82,9%	71,8%
<i>Correlação</i>	0%	100%	66%
<i>Contraste + Energia</i>	56,1%	82,8%	73,7%
<i>Contraste + Homogeneidade</i>	56,3%	83,2%	74%
<i>Contraste + Correlação</i>	44,7%	86,7%	72,5%
<i>Energia + Homogeneidade</i>	63,8%	81,4%	75,4%
<i>Energia + Correlação</i>	52,4%	86,4%	74,9%
<i>Correlação + Homogeneidade</i>	70,2%	83,1%	78,7%
<i>Todas</i>	71,6%	84,6%	80,2%

Diante da tabela exibida acima é possível observar que a *homogeneidade* isoladamente descreve as imagens de forma mais discriminante, pois é capaz de produzir aproximadamente 71,8% de precisão, enquanto que as evidências de *energia*, *contraste* e *correlação*, mostraram-se insuficientes para fornecer dados para o aprendizado do classificador. Entretanto, os resultados também mostram que a combinação das evidências de textura ocasiona um aumento na precisão, tal como, a combinação das características de *correlação* e *homogeneidade*. Isso ocorre porque é difícil o classificador conseguir aprender a partir de dados de apenas uma característica. Logo, o aprendizado pode tornar-se mais efetivo quando há um conjunto de características disponível para o treinamento.

Em termos de combinação, o melhor subconjunto de características de textura é composto por apenas duas evidências (*homogeneidade* e *correlação*), cujo desempenho é semelhante ao obtido com todas as características combinadas. Este subconjunto de características obteve 70,2% de sensibilidade (639 imagens de nudez classificadas corretamente), 83,1% de especificidade (1469 imagens neutras classificadas corretamente) e, 78,7% de precisão (2108 imagens classificadas corretamente).

Características de Forma

Por fim, foi feita a análise das características de formas, extraídas a partir do EHD definido na Seção 3.4.3. Vale ressaltar que para tornar possível a configuração do descritor global, a imagem não foi dividida em blocos de 4x4. Semelhantemente as demais características de cor e textura, a escolha das evidências de formas foi baseada nos seguintes critérios: evidências de fácil implementação que possam ser extraídas com baixo custo computacional e que sejam

frequentemente aplicadas em problemas de detecção de nudez. A Tabela 4.7 exibe resultados para aplicação de características de formas (*linhas horizontais, linhas verticais, linhas 45°, linhas 135° e pontos*) na base de teste, obtidos com o classificador SVM.

Tabela 4.7 – Análise das características de formas. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, sem a aplicação do zoneamento em imagens.

Característica de formas	Sensitividade	Especificidade	Precisão
<i>Linhas Horizontais</i>	0%	100%	66%
<i>Linhas Verticais</i>	0%	100%	66%
<i>Linhas 45°</i>	0%	100%	66%
<i>Linhas 135°</i>	0%	100%	66%
<i>Pontos</i>	0%	100%	66%
<i>Horizontais + Verticais</i>	0%	100%	66%
<i>Horizontais + 45°</i>	0%	100%	66%
<i>Horizontais + 135°</i>	0%	100%	66%
<i>Horizontais + Pontos</i>	0%	100%	66%
<i>Verticais + 45°</i>	0%	100%	66%
<i>Verticais + 135°</i>	0%	100%	66%
<i>Verticais + Pontos</i>	0%	100%	66%
<i>45° + 135°</i>	0%	100%	66%
<i>Todas</i>	2%	98%	66,1%

A tabela acima mostra que tanto isoladamente quanto combinadas, as características de formas não são capazes de fornecer dados suficientes para o aprendizado do classificador. Em todas as situações investigadas, SVM classificou todas as imagens como pertencentes à classe não-nudez. Acreditamos que esse resultado ocorreu devido à heterogeneidade das imagens da classe não-nudez, pois a tarefa de localizar objetos na imagem sem que seja realizado um particionamento (zoneamento) ou até mesmo um processo de segmentação é praticamente inviável.

Ao final de todo o processo de análise de características globais, todos os grupos (cor, forma e textura) foram combinados, isto é, tanto características consideradas relevantes quanto características irrelevantes, foram combinadas e o desempenho do classificador foi calculado. A Tabela 4.8 exibe o resultado da combinação de todas as características globais para descrição de imagens de nudez, utilizando o classificador SVM.

Tabela 4.8 – Análise de todas as características globais. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, sem a aplicação do zoneamento em imagens.

Características globais	Sensitividade	Especificidade	Precisão
<i>Todas as características</i>	86,9%	93,6%	91,3%
<i>Histograma + desvio padrão + homogeneidade + correlação</i>	87,5%	93,1%	91,1%

A tabela acima exhibe apenas o resultado da melhor combinação das características globais, uma vez que é desnecessário exhibir uma tabela com todas as combinações. Verificou-se que o descarte de características irrelevantes contribuiu para que o classificador obtivesse uma melhoria na taxa de sensibilidade, ou seja, a classificação correta de imagens de nudez. Isso se deve à eliminação dos ruídos no classificador, ou seja, dados ambíguos. O melhor subconjunto de características é composto por 2 (duas) características de cor, *histograma* e *desvio padrão* e 2 (duas) características de textura, sendo *homogeneidade* e *correlação*.

Ao analisarmos a composição do vetor de características e precisão do modelo, quando se tinha o conjunto de todas as características o vetor era formado de 235 atributos, sendo 196 atributos de cor, 32 atributos de textura e 5 atributos de formas. Com a seleção de características, um novo vetor foi formado com 211 atributos, sendo 193 atributos de cor e 16 de textura. Neste caso, houve uma redução de aproximadamente 10,3% da dimensão do espaço de características.

Desta forma, este trabalho mostra que a utilização de características globais possui forte aplicação e resultados satisfatórios. Em contrapartida, os resultados podem ser melhorados através do zoneamento em imagens, pois as informações podem ser analisadas de forma local, para que seja determinada a relevância de cada característica em cada zona da imagem. É importante mencionar que a aplicação de tais evidências e, os resultados obtidos dependem diretamente do tipo de objeto que se deseja identificar e, evidentemente da base de imagens utilizada.

4.2.2 Análise das Características Locais

A abordagem local usa apenas 2 (duas) características baseadas na cor a partir de uma imagem segmentada: *quantidade de pixels conectados* e *porcentagem de pixels de pele na imagem*. Estas características são usadas na maioria dos trabalhos que realizam detecção de nudez baseada em regiões, ou seja, a partir de informações do filtro de pele. Vale ressaltar que outras características de forma, tais como, *área* e *perímetro*, não foram utilizadas neste trabalho devido não haver aplicabilidade no algoritmo de zoneamento. Isto ocorre porque ao se aplicar o zoneamento em uma imagem segmentada é possível observar que as informações de borda são perdidas, tornando inviável a extração dessas características em conjunto com zoneamento. Um exemplo deste problema pode ser visualizado na Figura 4.3 da Seção 4.1. A Tabela 4.9 exhibe os resultados da análise de características locais para descrição de imagens de nudez, porém sem a aplicação do algoritmo de zoneamento.

Tabela 4.9 – Análise das características locais. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, sem a aplicação do zoneamento em imagens.

Característica locais	Sensitividade	Especificidade	Precisão
<i>Quantidade de pixels de pele</i>	76,9%	84,6%	82%
<i>Porcentagem de pixels de pele</i>	81,9%	84,8%	83,8%
<i>Quantidade + Porcentagem</i>	77,8%	87,2%	84%

É possível notar que as características apresentadas, ao serem utilizadas isoladamente, produzem elevado desempenho do classificador. Adicionalmente, a combinação de ambas contribui para aumentar o desempenho geral do classificador, com exceção da sensitividade. Porém, esse desempenho não superou as taxas obtidas por SVM ao ser treinado com características globais. A solução para melhoria do desempenho pode ser a combinação das características globais e locais. A próxima seção trata dessa combinação através da aplicação do algoritmo de zoneamento.

4.2.3 Combinação das Abordagens Global e Local através do Zoneamento em Imagens

Esta abordagem visa combinar características globais e locais e, ao mesmo tempo, analisar a relevância de cada característica em uma determinada zona da imagem, de maneira a selecionar as características mais relevantes e reduzir a dimensão do espaço de características. Por definição, características globais são informações extraídas da imagem sem a necessidade de pré-processamento, enquanto características locais são evidências extraídas a partir de regiões específicas. Neste trabalho, o termo combinação se dá pelo fato do algoritmo de zoneamento possibilitar a extração de características globais em diferentes zonas da imagem, ou seja, a técnica de particionar a imagem se configura como um processo de extração de informações locais.

A Tabela 4.10 apresenta os resultados da análise de características de cor em cada zona da imagem e a união de todas as zonas da imagem. Nesta tabela, semelhante às análises realizadas anteriormente, as características foram avaliadas isoladamente e em subgrupos.

Tabela 4.10 - Análise das características de cor. As taxas de desempenho foram calculadas na base de teste usando o classificador SVM, com a aplicação do zoneamento em imagens.

Características de Cor	Zona 1		Zona 2		Zona 3		Todas as Zonas	
	Sensit.	Espec.	Sensit.	Espec.	Sensit.	Espec.	Sensit.	Espec.
<i>Histograma</i>	82,1%	91,6%	78,9%	91,6%	0%	100%	89,9%	95,1%
<i>Média</i>	0%	100%	0%	100%	0%	100%	34,4%	85,4%
<i>Desvio Padrão</i>	53%	88,5%	26,8%	87,6%	0%	100%	67,5%	86,8%
<i>Histograma + Média</i>	82,3%	91,6%	79%	91,5%	0%	100%	89,8%	94,7%
<i>Histograma + Desvio Padrão</i>	82,3%	91,9%	79,3%	91,6%	0%	100%	90,3%	95%
<i>Média + Desvio Padrão</i>	60,1%	86,4%	51%	86,2%	0%	100%	71,3%	87,9%
<i>Todas de Cor</i>	81,4%	92%	78,8%	91,8%	0%	100%	90,5%	94,9%

A tabela mostra que ao analisarmos as características de cor isoladamente, o *histograma* é superior às demais características nas zonas 1 e 2. No entanto, verifica-se que na zona 3 nenhuma das características é suficientemente discriminante. Nesta análise, é possível confirmar a hipótese de Kalva *et al.* [2008] que afirma que a região central de imagens representa os objetos de interesse, enquanto regiões mais externas descrevem cenários ou objetos de menor interesse, ou seja, geralmente a zona mais externa é comum a todo tipo de imagem, não fornecendo informações suficientes que caracterizem uma determinada classe. É importante notar também que a *média* apresentou resultados semelhantes aos obtidos na abordagem global, e por isso pode ser considerada insuficiente para descrever imagens de nudez. Pode-se afirmar que o *histograma* isoladamente é suficiente para descrever imagens de nudez, pois a junção de todas as características de cor e a melhor combinação (*histograma + desvio padrão*) produziu resultados cuja diferença não é significativa.

Após a análise das características de cor, foi necessário realizar a análise das características de textura em cada zona da imagem, zonas independentes (*zona 1 + zona 2*, *zona 1 + zona 3*, e assim sucessivamente) e a união de todas as zonas. A Tabela 4.11 exibe os resultados da análise de características de textura no contexto do zoneamento em imagens.

Tabela 4.11 - Análise das características de textura. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, com a aplicação do zoneamento em imagens.

Características de Textura	Zona 1		Zona 2		Zona 3		Todas as Zonas	
	Sensit.	Espec.	Sensit.	Espec.	Sensit.	Espec.	Sensit.	Espec.
<i>Contraste</i>	72,7%	81,4%	30%	95,6%	19,6%	96,9%	81,4%	91,6%
<i>Correlação</i>	72,1%	81,4%	51,5%	88,2%	12,1%	98,5%	70%	89,8%
<i>Homogeneidade</i>	50,4%	89%	39,5%	88,5%	13,3%	97,5%	82,3%	91,8%
<i>Energia</i>	47%	85%	0%	100%	10,3%	99,8%	61,5%	86,7%
<i>Contraste + Correlação</i>	79,3%	87,7%	52,4%	89%	23,1%	95,9%	86,3%	93,4%
<i>Contraste + Homogeneidade</i>	82%	88,3%	66,6%	87,9%	32,1%	93,1%	87,4%	94%
<i>Contraste + Energia</i>	80,7%	87,3%	36,8%	94,5%	16,7%	98,7%	84,2%	92,3%
<i>Energia+ Homogeneidade</i>	74,8%	85,8%	66,6%	86,4%	37,7%	91%	85,2%	92,3%
<i>Energia+ Correlação</i>	66,2%	87,9%	56,8%	86,3%	13,2%	99,4%	80,7%	92%
<i>Homogeneidade + Correlação</i>	77,1%	89,5%	67,5%	87,6%	50,4%	89%	88,4%	93,4%
<i>Todas Textura</i>	82,3%	89,2%	69,7%	86,9%	52,4%	90%	91,4%	95,6%

O desempenho produzido pelas características de textura extraídas em zonas da imagem foi diferente do comportamento observado nos experimentos em que o algoritmo de zoneamento não foi utilizado, ou seja, características consideradas irrelevantes sem a aplicação do zoneamento tornaram-se relevantes com o uso do módulo de zoneamento. Na análise individual, pode-se perceber que ao

alterarmos a zona de aplicação, a relevância da característica varia. Como exemplo pode-se citar o *contraste*, que obteve os melhores desempenhos nas *zonas 1* e *3*, sendo que na *zona 2*, a correlação superou as demais. De forma geral, ou seja, considerando todas as zonas, o melhor desempenho de características de textura foi obtido pela combinação de homogeneidade e correlação, com 88,4% de sensibilidade (804 imagens de nudez classificadas corretamente), 93,4 % de especificidade (1653 imagens neutras classificadas corretamente) e precisão de 91,6% (2457 imagens classificadas corretamente).

Em se tratando de análise de características de formas utilizando o zoneamento em imagens, os experimentos realizados confirmam a hipótese de que a aplicação de evidências de formas se faz necessária somente em regiões específicas da imagem. A Tabela 4.12 exhibe os resultados dessa aplicação.

Tabela 4.12 - Análise das características de formas. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, com a aplicação do zoneamento em imagens.

Características de Formas	Zona 1		Zona 2		Zona 3		Todas as Zonas	
	Sensit.	Espec.	Sensit.	Espec.	Sensit.	Espec.	Sensit.	Espec.
<i>Linhas Horizontais</i>	70,1%	79,1%	73,5%	70,8%	0%	100%	73,6%	81,1%
<i>Linhas Verticais</i>	69,1%	75%	0%	100%	0%	100%	66,5%	80%
<i>Linhas 45°</i>	68%	78,8%	0%	100%	0%	100%	65,2%	81,8%
<i>Linhas 135°</i>	64,6%	80,2%	0%	100%	0%	100%	64,3%	82,2%
<i>Pontos</i>	17,9%	92,1%	0%	100%	0%	100%	59,1%	81,8%
<i>Horizontais + Verticais</i>	73,1%	78%	73,7%	69,9%	0%	100%	74,1%	83,6%
<i>Horizontais + 45°</i>	70,9%	80,3%	73,1%	72,4%	0%	100%	74,8%	84,4%
<i>Horizontais + 135°</i>	70,2%	78,9%	75,7%	71,8%	0%	100%	75,6%	83,6%
<i>Horizontais + Pontos</i>	70,5%	79%	74,7%	69,9%	0%	100%	72,1%	84,7%
<i>Verticais + 45°</i>	68,5%	79,3%	0%	100%	0%	100%	65,8%	81,4%
<i>Verticais + 135°</i>	68%	77%	0%	100%	0%	100%	67,7%	82,4%
<i>Verticais + Pontos</i>	65,6%	77%	0%	100%	0%	100%	67,1%	83,4%
<i>45° + 135°</i>	65,2%	80,4%	0%	100%	0%	100%	66,4%	81,5%
<i>Todas</i>	75,1%	79,3%	75,2%	73,1%	0%	100%	78,4%	91%

Como observado na tabela, em geral as características de formas são relevantes apenas na *zona* mais ao centro da imagem (*zona 1*). Na análise isolada, somente a característica baseada em *linhas horizontais* apresentou-se discriminante nas *zonas 1* e *2*. Entretanto na *zona 3*, nenhuma das características foi suficientemente representativa. Outro ponto a destacar é que na *zona 2*, todas as características ao serem combinadas com *linhas horizontais* mostraram-se relevantes, fato que não ocorreu com as demais combinações. Logo, para efeito de seleção de características, ao juntarmos todas as zonas o melhor desempenho resultou da combinação de *linhas horizontais* e *linhas 135°*, com 75,6% de sensibilidade (687 imagens de nudez classificadas corretamente), 83,6% de

especificidade (1479 imagens neutras classificadas corretamente) e precisão de 80,8% (2166 imagens classificadas corretamente).

A avaliação dos dados referentes à análise de características baseadas no filtro de pele obteve resultados bastante animadores, pois em síntese, confirma a hipótese de que em imagens de nudez há uma maior predominância de *pixels* de pele no centro da imagem, conforme Figura 4.7.

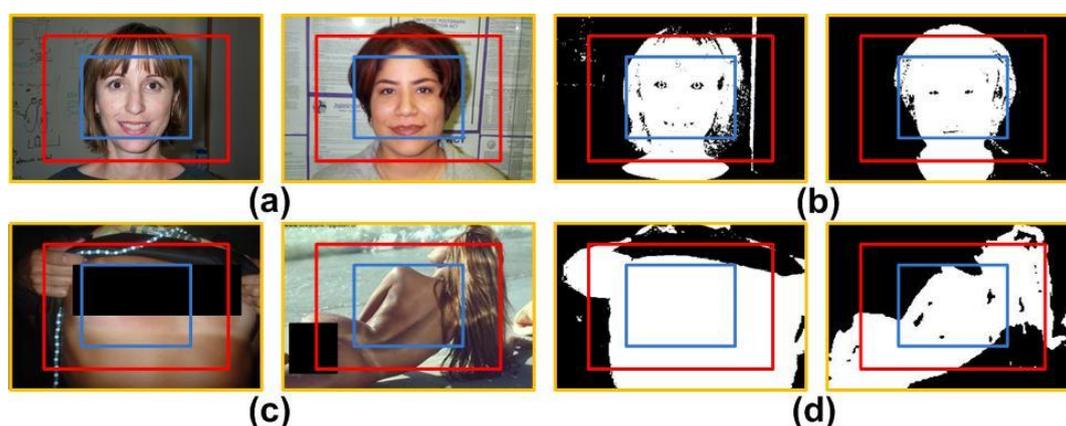


Figura 4.7 – Imagens que demonstram o resultado da aplicação do zoneamento, em que (a) representa o zoneamento em imagens de face, (b) o resultado do zoneamento em imagens de face após a aplicação do filtro de pele, (c) o zoneamento em imagem de nudez, (d) o resultado do zoneamento em imagens de nudez após o filtro de pele.

É possível observar que em imagens de faces também há uma grande quantidade de *pixels* de pele no centro da imagem. Porém, o que diferencia as imagens de face das imagens de nudez é a grande concentração de *pixels* de pele na zona intermediária (zona 2), conforme a Tabela 4.13. Outra característica que possibilita a diferenciação das classes é o baixo valor de contraste presente nas imagens de nudez, uma vez que esta medida de textura extrai informações referentes às bordas das imagens, onde em geral, as imagens de nudez contêm os *pixels* de pele concentrados, enquanto que imagens de faces têm bordas referentes à boca, nariz e olhos. Logo, esta hipótese se configura como uma nova característica para descrição de imagens de nudez e, portanto, uma das contribuições deste trabalho. A Tabela 4.13 apresenta resultados da análise de características baseadas no filtro de pele através do zoneamento em imagens.

Tabela 4.13 - Análise das características locais. As taxas de desempenho foram calculadas na base de teste com uso do classificador SVM, com a aplicação do zoneamento em imagens.

Características Locais	Zona 1		Zona 2		Zona 3		Todas as Zonas	
	Sensit.	Espec.	Sensit.	Espec.	Sensit.	Espec.	Sensit.	Espec.
Quantidade de Pixels de pele	89,3%	92%	88,9%	92,3%	68,1%	86,3%	91,5%	93,7%
Porcentagem de Pixels de pele	89,1%	92%	87,4%	92,8%	69,6%	87,6%	90%	94,3%
Combinação	89,2%	92,8%	88,7%	93,2%	68,1%	89,4%	92%	95,3%

Diante da tabela acima é possível observar que a *quantidade de pixels de pele* no centro da imagem (*zona 1*) é uma característica que possui elevada relevância. Esta característica produziu um ganho de aproximadamente 11% de precisão se comparado com a abordagem sem a utilização do zoneamento, apresentada na Tabela 4.9. É importante destacar que a porcentagem de *pixels* de pele obteve também melhorou os resultados, uma vez que obteve aproximadamente o mesmo ganho obtido pela característica quantidade de *pixels* de pele.

Esta análise mostra que o zoneamento em imagens se faz necessário, não somente para aplicações de detecção de nudez, mas também em aplicações diversas, para facilitar o reconhecimento de objetos em imagens.

Para finalizar, foram realizadas todas as combinações dos grupos de características para possibilitar a análise e a seleção das características relevantes para descrição de imagens de nudez. Isto é, os grupos de cor, textura, forma, juntamente com características baseadas no filtro de pele foram combinados e diferentes subconjuntos de características foram testados. A Tabela 4.14 apresenta o resultado da combinação de todas as características e da seleção do melhor conjunto com a utilização do zoneamento em imagens.

Tabela 4.14 – Análise da combinação de todas as características e seleção do melhor conjunto para descrição de imagens de nudez através do zoneamento em imagens.

Conjunto de Características	Sensitividade	Especificidade	Precisão
<i>Todas as características da Zona 1</i>	95,4%	97,1%	96,4%
<i>Todas as características da Zona 2</i>	93,8%	96,5%	95,6%
<i>Todas as características da Zona 3</i>	57,2%	92%	80,1%
<i>Melhor combinação de características nas 3 zonas</i>	98,4%	99,1%	98,8%
<i>Todas as Características e zonas</i>	96%	98,1%	97,3%

Após realizarmos todas as combinações possíveis e analisarmos a distribuição das características em cada *zona*, podemos afirmar que a melhor combinação resultante na *zona 1* (zona central) é suficiente para caracterizar imagens de nudez, com precisão de 96,4%, aproximadamente. Verificou-se que o melhor conjunto de características na *zona 1* foi composto por: *Histograma*, *Desvio Padrão*, *Homogeneidade*, *Contraste*, *Linhas Horizontais*, *Linhas Verticais*, *Quantidade de Pixels de pele* e *Porcentagem de Pixels de pele*.

Conforme pode ser observado na Tabela 4.14, o desempenho do classificador quando treinado com o melhor subconjunto na zona 2 também foi elevado, esse subconjunto é composto por: *histograma*, *desvio padrão*, *homogeneidade*, *correlação*, *linhas horizontais*, *linhas 135°*,

quantidade de pixels e porcentagem de pixels. Embora o desempenho seja semelhante ao obtido na zona 1, houve redução na taxa de sensibilidade, que representa a taxa de imagens de nudez classificadas corretamente.

Os resultados obtidos na *zona 3* confirmam novamente nossa hipótese de que as informações mais relevantes concentram-se nas áreas centrais das imagens. As características extraídas na zona 3 não são suficientes para o aprendizado do classificador, fato confirmado pela ocorrência de uma grande disparidade entre sensibilidade (57,2%) e especificidade (92%).

Logo, os resultados produzidos pelo melhor subconjunto de características, obtido a partir da combinação de todas as zonas, indicam que a seleção de características pode reduzir a dimensão do espaço de características e melhorar o desempenho do classificador, uma vez que os ruídos são eliminados. A melhor combinação de características foi resultante da análise e seleção do melhor subconjunto da zona 1 e o melhor subconjunto da zona 2, sendo descritos abaixo:

- *Zona 1: Histograma, homogeneidade, contraste, linhas horizontais, linhas verticais, quantidade de pixels de pele e porcentagem de pixels de pele;*
- *Zona 2: Histograma, homogeneidade, correlação, linhas horizontais, linhas verticais, quantidade de pixels de pele e porcentagem de pixels de pele;*

Em resumo, o vetor de características composto somente com as características mais relevantes possibilitou que SVM alcançasse 98,4% de sensibilidade (895 imagens de nudez classificadas corretamente), 99,1% de especificidade (1754 imagens neutras classificadas corretamente) e, 98,8% de precisão (2649 imagens classificadas corretamente). Vale ressaltar que, ao adicionar características da zona 3, o desempenho do classificador reduziu e, portanto, as características desta zona não foram utilizadas. Isso se dá em razão da zona 3 estar diretamente associada ao fundo da imagem, que em geral não possui informação relevante para a classificação.

Em termos de redução de dimensionalidade, um vetor resultante da união de todas as características extraídas nas 3 zonas, é composto por 711 atributos sendo: 198 atributos de cor, 32 atributos de textura, 5 atributos de formas e 2 atributos do filtro de pele, todos extraídos em cada zona. Com a seleção das características, um novo vetor foi formado contendo apenas 424 atributos, sendo: 384 atributos de cor, 32 atributos de textura, 4 atributos de formas e 4 atributos do filtro de pele. Verifica-se então que houve uma redução de aproximadamente 46% na dimensão do espaço de características.

4.2.4 Análise de Tempo de Execução

Também vale destacar a necessidade de análise da eficiência das abordagens investigadas, ou seja, o tempo necessário para que cada imagem seja analisada em cada abordagem. Essa informação é importante para ser verificada a aplicabilidade da arquitetura ANDImage em sistemas *online*. Segundo Zhu *et al.* [2004] o tempo de execução ideal para que uma abordagem seja aplicável em sistemas *online* deve ser inferior a 0,3 segundos por imagem (s/i). A Tabela 4.15 exibe uma comparação da eficiência da arquitetura ANDImage através da análise de características com e sem a utilização do zoneamento para descrição de imagens de nudez.

Tabela 4.15 – Análise da eficiência da arquitetura.

	Sem Zoneamento		Com Zoneamento	
	Todas as características	Melhor subconjunto de características	Todas as características	Melhor subconjunto de características
Normalização	0,07s/i	0,07s/i	0,07s/i	0,07s/i
Extração de características	0,12s/i	0,05s/i	0,15s/i	0,06s/i
Classificação	0,11s/i	0,09s/i	0,19s/i	0,15s/i
Tempo Médio do Processo	0,30s/i	0,21s/i	0,41s/i	0,28s/i
Precisão / Falso Negativo	91,9% / 11,1%	91,7% / 11%	97,3% / 4%	98,8% / 1,6%
Quantidade de Atributos	237	211	711	424

Como visto na tabela acima, o tempo utilizado para normalizar as imagens em dimensões de aproximadamente 256 colunas e 256 linhas em uma base contendo 5.360 imagens é de 0,07 segundos por imagem. Logo após, foi verificado o tempo utilizado para a extração de características sem a utilização do zoneamento, e o tempo gasto para extrair somente as características mais relevantes. Feito isso, foi calculado o tempo para classificação das amostras. O tempo médio do processo é dado pela soma dos tempos utilizados para normalizar as imagens, extrair as características e classificar as imagens. É possível observar que com o processo de seleção de características, sem o zoneamento em imagens, o tempo de execução reduz em aproximadamente 9 (nove) milissegundos, enquanto a precisão fica praticamente inalterada. Quando este mesmo processo é realizado através do zoneamento em imagens, o tempo de execução reduz em até 13 milissegundos. Como o tempo gasto para processar cada imagem foi abaixo de 30 milissegundos (cerca de 28 milissegundos por imagem), pode-se afirmar que nossa arquitetura se configura como um problema para aplicações *online*.

Além disso, a seleção de características possibilitou a redução da dimensão do espaço de características e, conseqüentemente, a eliminação de ruídos no classificador, ocasionando aumento na precisão. Outros fatores que podem ser observados estão relacionados ao aumento da precisão

em aproximadamente 7% e redução percentual de falsos negativos em 9%, obtidos com o uso do módulo de zoneamento. Esses resultados mostram que o algoritmo de zoneamento pode ser aplicado não somente para reconhecimento de imagens de nudez, mas também em outros problemas de classificação de imagens.

Capítulo 5

Conclusões e Trabalhos Futuros

Neste trabalho foi realizada uma análise e seleção de características para detecção de nudez com o auxílio de uma arquitetura que permite o uso de propriedades globais, locais e zoneamento em imagens. A técnica de aprendizagem de máquina SVM foi utilizada para classificar as amostras em nudez ou neutra (não-nudez).

Na abordagem global, as características são extraídas sem a necessidade de realização de segmentação das imagens. Foram analisados três grandes grupos de características: cor, forma e textura, tanto de forma isolada, quanto de forma combinada, para compor um classificador capaz de determinar a que classe (nudez ou neutra) pertence determinada imagem. Utilizando-se o conjunto de todas as características globais, SVM é capaz de alcançar 91,3% de precisão, tendo um vetor composto por 235 atributos. Já com o conjunto formado apenas com as características mais relevantes (*histograma, desvio padrão, homogeneidade e correlação*) SVM alcançou 91,1% de precisão com um vetor formado por 211 atributos. Foi observado que houve uma economia de 10,3% no espaço de armazenamento devido à redução do vetor de características. Estes resultados estão descritos na Seção 4.2.1.

Na Seção 4.2.2 foram apresentados os resultados dos experimentos para uso da abordagem local que utilizou apenas duas características (*quantidade de pixels de pele e porcentagem de pixels de pele na imagem*). Foi verificado que, isoladamente, as características fornecem informações suficientes para SVM alcançar elevada taxa de desempenho em detecção de nudez. Isso mostra a eficiência de modelos de detecção de nudez baseados em regiões, onde o tempo para pré-processar a imagem é compensado pelo tamanho reduzido do vetor de características. Entretanto, para melhorar a precisão desta abordagem, este trabalho propôs a combinação com evidências globais (cor, textura e forma) e aplicação do zoneamento em imagens, para que fosse analisada a distribuição das características em regiões específicas da imagem e confirmada a hipótese de que em imagens de nudez há uma maior concentração de *pixels* de pele no centro da imagem.

Através do zoneamento em imagens e análise de características foi possível observar que:

- As características possuem grande variabilidade, dependendo da região de aplicação;
- As características de cor não possuem relevância na *zona 3*, pois nesta região as características são comuns a todo tipo de imagem;
- As características de textura se configuram como um excelente conjunto de dados para detecção de nudez, pois atuam como o grande diferencial na confusão de *pixels* de pele;
- As características de formas só fornecem dados suficientes para aprendizado do classificador caso as informações sejam extraídas de regiões específicas da imagem, ou seja, caso a imagem seja particionada. Outro fator observado neste conjunto de características é o fato de somente possuir relevância na região central da imagem, ou seja, a região que geralmente compreende os objetos de interesse;
- Nenhuma das características foi suficientemente relevante na *zona 3*, pois nesta região as características são comuns a todos os tipos de imagens, tornando as informações insuficientes para o aprendizado do classificador;
- A *quantidade de pixels de pele* na *zona 1*, ou seja, centro da imagem, configurou-se como uma importante característica para detecção de imagens de nudez, confirmando a hipótese adotada no objetivo do trabalho.

As tentativas de combinação das características globais e locais através do algoritmo de zoneamento produziram elevadas taxas de desempenho, pois foi possível aumentar a precisão em aproximadamente 7%, em relação a não utilização do zoneamento em imagens. A partir da análise de características foi possível selecionar algumas características das *zonas 1* e *2*, resultando na redução de 46% do vetor de características e aumento da precisão devido à eliminação de ruídos no classificador. Além disso, o tempo gasto para processar cada imagem pode ser considerado aplicável em sistemas *online*, sendo abaixo de 30 milissegundos.

Dessa forma, a combinação de propriedades globais e locais, juntamente com a aplicação do zoneamento em imagens se faz necessária para reconhecimento de imagens de nudez, configurando-se como um método aplicável a diversos domínios de reconhecimento de objetos.

5.1. Trabalhos Futuros

O presente trabalho mostra que ainda restam muitas comparações e testes a serem realizados. Durante a execução do projeto foram identificados diversos trabalhos que podem contribuir para a melhoria deste trabalho.

Um possível trabalho futuro seria a comparação de diferentes processos de particionamento da imagem para reconhecimento de imagens de nudez, uma vez que, neste trabalho foi utilizada apenas divisão da imagem em 3 (três) zonas. Outra solução seria a implementação de diversas outras características propostas na literatura, analisando o comportamento em diferentes espaços de cores. Isto poderá contribuir para que tais métodos possam ser aplicáveis em outros domínios de reconhecimento de objetos.

Por fim, deseja-se realizar a implementação prática das abordagens aqui propostas, como por exemplo, através de CBIR (*Content-Based Image Retrieval*) ou, o desenvolvimento de uma espécie de *plugin* para navegadores *Web*, servidores *proxy* e servidores de e-mail em um ambiente organizacional.

5.2. Publicações obtidas

A seguir, segue a lista de todas as publicações obtidas durante o mestrado. Todas elas são artigos publicados em conferências e mostram os resultados obtidos nesta dissertação.

- Santos, C. A. M.; Santos, E. M.; Souto, E. J. P. “Nudity Detection Based on Image Zoning ” In: 11th International conference on Information Science, Signal Processing and their Applications (ISSPA 2012), Montreal, Canadá. (aceito)
- Santos, C. A. M.; Pinto, K. M. B.; Santos, E. M.; Souto, E. J. P. “Análise de Características para Detecção de Pele em Imagens. In: X Congresso Brasileiro de Inteligência Computacional, 2011, Fortaleza - Ceará. Anais do X Congresso Brasileiro de Inteligência Computacional, 2011. (aceito)
- Henke, M; Santos, C. A. M.; Nunan, A. E. ; Feitosa, E. L.; Santos, E. M.; Souto, E. J. P. “Aprendizagem de Máquina para Segurança em Redes de Computadores: Métodos e Aplicações”. In: Minicursos do XI Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg 2011). Brasília. 2011, v. 1, p. 53-103. (aceito)

Referências

- Alpaydim, E. (2010) “Introduction to Machine Learning” *The MIT Press*. Cambridge, Massachusetts, EUA.537 p.
- Arentz, W. A. & Olstad, B. (2004) “Classifying Offensive Sites Based on Image Content”. *Computer Vision and Image Understanding* 94. pp. 295-310.
- Basilio, J. A. M.; Torres, G. A.; Pérez, G. S.; Medina, L. K. T. & Meana, H. M. P. (2011) “Detection of Pornographic Digital Images” *International Journal of Computers*, Vol. 5 (2), pp. 298 – 305.
- Belém, R. J. S. (2006) “Detecção Automática de Conteúdo Ofensivo na Web”, *Dissertação de Mestrado*, Programa de Pós-Graduação em Informática (PPGI), Universidade Federal do Amazonas (UFAM), 51p.
- Brown, D.; Craw, I. & Lewthwaite, J. (2001) “A SOM Based Approach to Skin Detection with Application in Real Time Systems”, *BMVC01*, pp. 491-500.
- Cai, J. & Goshtasby, A. (1999) “Detecting Human Faces in Color Images”. *Image Vision Comput.* Vol. 18, pp. 63–75.
- Cao, L. L.; Li, X. L.; Yu, N. H. & Liu, Z. K.(2002) “Naked People Retrieval Based on Adaboost Learning”, *ICMLC 2002*, pp 1133 - 1138.
- Dios, J. J. & Garcia, N. (2003) “Face Detection Based on a New Color Space YCgCr”, Em *International Conference on Image Processing (ICIP'03.)*, pp. 909-912.
- Duan, L.; Cui, C.; Gao, W. & Zhang, H. (2002) “Adult Image Detection Method Base-on Skin Color Model And Support Vector Machine”. In: *ACCV2002*, Melbourne, Australia, pp. 22-25.
- Fleck, M.; Forsyth, D.A. & Bregler, C. (1996) “Finding naked people”, *Proc. 4th European Conf. On Computer Vision*, vol. 2, pp. 593-602.
- Fukunaga, K. (1990). “Introduction to Statistical Pattern Recognition”, *Academic Press Prof.*, 2nd ed., San Diego, CA, USA.
- Gonzalez, R.C. & Woods, R. E. (2008) “Digital Image Processing”, Terceira Edição, *Addison-Wesley Publishing Company*, 954 pp.
- Haykin, S. (2008) “Neural Networks and Learning Machines” 3rd Edition, *Prentice Hall*.
- Henke, M; Santos, C. A. M.; Nunan, A. E. ; Feitosa, E. L.; Santos, E. M.; Souto, E. J. P. “Aprendizagem de Máquina para Segurança em Redes de Computadores: Métodos e Aplicações”. In: *Minicursos do XI Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg 2011)*. Brasília. 2011, v. 1, p. 53-103.
- Hu, W.; Wu, O.; Chen, Z.; Fu, Z. & Maybank, S. (2009) “Recognition of Pornographic Web Pages by Classifying Texts and Images”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1019-1034.
- Hsu, C-W.; Chang, C-C.; Lin, C-J. (2003) “A Pratical Guide to Support Vector Machines Classification” *Computer and Information Science*, Vol. 1 (1), pp. 1-16.
- Jain, A. K. & Valaya. (1995) “Image Retrieval Using Color and Shape”. *Pattern Recognition*, vol. 29 (8), pp.1233 – 124.

- Jeong, C-Y.; Kim, J-S.; Hong, K-S. (2004) "Appearance-Based Nude Image Detection" *Pattern Recognition, ICPR 200*, vol.4, pp. 467- 470.
- Jiao, F.; Gao, W.; Duan, L. & Cui, G. (2001) "Detecting Adult Image using Multiple Features", *Proc. International Conference Info-tech and Info-net (ICII 2001)*, pp.378-383.
- Jones, M. J. & Rehg, J. M. (2002) "Statistical Color Models with Application to Skin Detection". *Computer Vision*, Vol. 46, pp. 81-86.
- Kalva, P. R.; Enembreck, F. & Koerich, A. L. (2008) "WEB Image Classification using Combination of Classifiers", *IEEE Latin America Transactions*, vol. 6, pp. 661-671.
- Kakumanu, P.; Makrogiannis, S.; Bourbakis, N. (2007) "A Survey of Skin-color Modeling and Detection Methods". Em *Pattern Recognition*. Vol. 40 (3), pp. 1106-1122.
- Kawato, S. & Ohya, J. (2000) "Automatic Skin-Color Distribution Extraction for Face Detection and Tracking, Em 5th International Conference on Signal Processing, Vol. 2, pp. 1415–1418.
- Kelly, W.; Donnellan, A.; & Molloy, D. (2007) "Screening for Objectionable Images: A Review of Skin Detection Techniques". Em *International Conference Machine Vision and Image Processing ICMVIPPA'07*, pp. 310-319.
- Kherfi, M. L.; Ziou, D. & Bernardi, A. (2004) "Image Retrieval from the World Wide Web: Issues, Techniques, and Systems". *ACM Computing Survey*, Vol. 36, pp. 35-67.
- Kohavi, R. & John, G. H. (1997) "Wrappers for Feature Subset Selection", *Artificial Intelligence*, Vol. 97 (1-2), pp. 273-324.
- Lee, J. S.; Kuo, Y. M.; Chung, P.C. & Chen, E. L. (2007) "Naked Image Detection Based on Adaptive and Extensible Skin Colour Model", *Pattern Recognition*, Vol. 40, pp. 2261-2270.
- Lisin, D. A.; Mattar, M. A.; Blaschko, M. B.; Learned-Miller, E. G. & Benfield, M. C. (2005) "Combining Local and Global Image Features for Object Class Recognition" Em *CVPR'05*.
- Lopes, A. P. B.; Avila, S. E. F.; Peixoto, A. N. A.; Oliveira, R. S. & Araújo, A. A. (2009) "A Bag-of-Features Approach based on Hue-Sift Descriptor for Nude Detection". Em *17nd European Signal Processing Conference (EUSIPCO 2009)*, pp. 1552-1556.
- Miyamoto, D.; Hazeyama, H.; & Kadobayashi, Y. (2009) "An Evaluation of Machine Learning-Based Methods for Detection of Phishing Sites" Em *Proceedings of the 15th International Conference on Advances in Neuro-Information Processing*, pp. 539-546.
- Penatti, O.A.B. & Torres, R. S. (2008) "Color Descriptors for Web Image Retrieval: A Comparative Study", *XXI Brazilian Symposium on Computer Graphics and Image Processing, 2008. SIBGRAPI '08*, pp.163-170.
- Rich, E. & Knight, K. (1991) "Artificial Intelligence", *McGraw-Hill*.
- Saber, E. & Tekalp, A. M. (1998) "Frontal-view Face Detection and Facial Feature Extraction Using Color, Shape and Symmetry Based Cost Functions, *Pattern Recognition*, Vol. 17 (8), pp. 669-680.
- Safar, M.; Shahabi, C. & Sun, X. (2000) "Image Retrieval by Shape: A Comparative Study", In: *IEEE International Conference on Multimedia and Expo*. pp. 141–144.
- Santos, C. A. M.; Santos, E. M.; Souto, E. J. P. (2012) "Nudity Detection Based on Image Zoning" In: *11th International conference on Information Science, Signal Processing and their Applications (ISSPA 2012)*, Montreal, Canadá.

- Santos, C. A. M.; Pinto, K. M. B.; Santos, E. M.; Souto, E. J. P. (2011) “Análise de Características para Detecção de Pele em Imagens. In: *X Congresso Brasileiro de Inteligência Computacional*, 2011, Fortaleza - Ceará. Anais do X Congresso Brasileiro de Inteligência Computacional.
- Santos, E. M. (2011) “Ocorrência de Aprendizagem Viciada em Seleção de Conjuntos de Classificadores”, Em *I Encontro Regional de Pesquisa Operacional do Norte, ERPO 2011*.
- Schetini, R.; Brambilla, C.; Cusano, C. & Ciocca, G. (2003) “On the Detection of Pornographic Digital Images”, *VCIP'03*, pp. 2105-2113.
- Sevmili, H.; Esen, E.; Ates, T. K.; Ozan, E. C.; Tekin, M.; Logoglu, B.; Sevinç, A. M.; Saracoglu, A.; Yazici, A. & Alatan, A. A. (2010) “Adult Image Content Classification Using Global Features and Skin Region Detection”, *ISCIS'10*, Springer, Vol. 62, pp. 253-258.
- Sobel, I. E. (1970) “Camera Models and Machine Perception”, *Tese de Doutorado*, Stanford University, Stanford, CA, USA.
- Terrillon, J.-C.; David, M. & Akamatsu, S. (1998) “Detection of Human Faces in Complex Scene Images by Use of a Skin Color Model and of Invariant Fourier-Mellin Moments”, Em *Proceedings 50th International Conference on Pattern Recognition (ICPR'98)*, pp.1350-1355.
- Vapnik, V. N. (1995) “The Nature of Statistical Learning Theory”. *Springer*, New York.
- Xiao, H.; Hong, F.; Zhang, Z. & Liao, J. (2007) “Intrusion Detection Using Ensemble of SVM Classifier”, *FKSD 2007*, pp. 45-49.
- Yang, J.; Lu, W. & Waibel, A. (1998) “Skin-color Modeling and Adaptation”, In *ACCV98*.
- Yang, M. H. & Ahuja, N. (1998) “Detecting human faces in color Images”. In *International Conference on Image Processing (ICIP)*, vol. 1, pp. 127–130.
- Yogarajah, P.; Condell, J.; Curran, K.; Cheddad, A. & McKeivitt, P. (2010) “A Dinamic Threshold Approach for Skin Segmentation in Color Images”, In Proc. *IEEE 17nd International Conference on Image Processing (ICIP 2010)*. Hong Kong, China. pp. 2225-2228.
- Wang, S-L.; Hu, H.; Li, S-H. & Zhang, H. (2005) “Exploring Content-Based and Image-Based Features for Nude Image Detection” Em *FSKD 2005*, Vol. 2, pp. 324-328.
- Wang, J.; Li, J.; Wiederhold & Firschein, G. (1998) “System for Screening Objectionable Images” *Computer Communications*. pp. 20-30.
- Weka. (2010) “Data Mining with Open Source Machine Learnig Software in Java”. University of Waikato. Disponível em: <http://www.cs.waikato.ac.nz/~ml/weka>. Acesso em Outubro de 2010.
- Won, C-S.; Park, D-K. & Park, S-J. (2002) “Efficient Use of MPEG-7 Edge Histogram Descriptor” *ETRI Journal*, vol. 24 (1), pp.23-30.
- Zheng, Z.; Daoudi, M. & Jedynek, B. (2004) “Blocking Adult Images Based on Statistical Skin Detection”. *Electronic Letters on Computer Vision and Image Analysis*, Vol. 4, pp. 1-14.
- Zheng, Q. F.; Zhang, M. J. & Wang, W. Q. (2004) “Shape-based Adult Image Detection”, *ICIG2004*, pp. 150-153.
- Zhu, H.; Zhou, S.; Wang, J. & Yin, Z. (2007) “An Algorithm of Pornographic Image Detection”. In *Fourth International Conference on Image and Graphics (ICIG 07)*. pp. 801-804.
- Zhu, Q.; Wu, C-T.; Cheng, K-T. & Wu, Y-L. (2004) “An Adaptive Skin Model and its Application to Objectionable Image Filtering”, *ACM Multimedia*, pp. 56-63.