UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

RAFAEL SACHETT MEDEIROS

Detecção de Pele Humana Utilizando Modelos Estocásticos Multi-Escala de Textura

Dissertação apresentada como requisito parcial
para a obtenção do grau de
Mestre em Ciência da Computação

Prof. Dr. Jacob Scharcanski
Orientador

Porto Alegre, março de 2013

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Medeiros, Rafael Sachett

Detecção de Pele Humana Utilizando Modelos Estocásticos Multi-Escala de Textura / Rafael Sachett Medeiros. – Porto Alegre: PPGC da UFRGS, 2013.

79 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2013. Orientador: Jacob Scharcanski.

1. Segmentação de imagens. 2. Fusão estocástica de regiões. 3. Reconhecimento de texturas. 4. Detecção de pele humana. 5. Segmentação de gestos de mão. I. Scharcanski, Jacob. II. Título.

"Sonhar é acordar-se para dentro."
— MÁRIO QUINTANA

"There are no ordinary moments."
— DAN MILLMAN

"Desde que me cansei de procurar, aprendi a encontrar.
Desde que o vento me opõe resistência, velejo com todos os ventos."
— FRIEDRICH NIETZSCHE

AGRADECIMENTOS

Agradeço, primeiro de tudo, a meus pais, Carmem e Luiz. Foram eles que me incentivaram e me deram força durante toda a minha vida. Obrigado por tudo, principalmente pelo amor, carinho e ensinamentos que recebi de vocês.

A meu irmão, Rodrigo, que de alguma forma inexplicável conseguiu passar para mim parte do espírito guerreiro. Muito obrigado pelo exemplo de perseverança e humildade, que me incentivou a não desistir em qualquer dificuldade. Obrigado pela fraternidade e camaradagem que sempre teve comigo.

Devo também um agradecimento muito especial à minha namorada, Camila. Adendo recente, ela me acompanhou apenas nos últimos (e mais difíceis) momentos desta minha jornada, mas com certeza foi alguém que fez toda a diferença nesses momento, assim como fará em toda a minha vida daqui em diante. *Sweetheart*, muito obrigado pela compreensão e apoio nesses momentos.

A todos os membros da minha família, que sempre souberam me apoiaram nas minhas decisões e questioná-las quando necessário. Em especial, às minhas avós, Aeriovalda e Lia, que sempre foram exemplo de positivismo e disposição.

Ao meu amigo e companheiro Tainã, que acompanhou durante grande parte de meu percurso, desde os tempos da faculdade em Pelotas até hoje. Muito obrigado pela amizade em todos esses anos.

Ao Prof. Jacob, que me orientou nesses últimos dois anos. Tanto nos períodos de calmaria, quando parecia que nada ia lugar nenhum, quanto nas épocas de correria, quando se aproximavam os *deadlines*, sempre me ajudou a atingir meus objetivos, e me colocou na direção correta. Muito obrigado pela paciência e ensinamentos.

Muito obrigado também a todos os meus colegas do Lab-PIVC que, entre uma conversa fiada e reclamações sobre o ar condicionado, sempre estavam dispostos a colaborar com algumas linhas de código, compartilhar alguma notícia chocante, ou alguma paródia do Gangnam Style.

Um agradecimento à parte para João Guedes, meu mestre de Kung Fu e Sanda, que diversas vezes teve que interromper suas meditações e afazeres burocráticos para dar atenção ou conselhos (de todo o tipo) para esse aluno meio maluco. Muito obrigado por todo o conhecimento e sabedoria que recebi de ti, não somente em relação à artes marciais, integridade e saúde física, mas todos os aspectos da minha vida. Agradeço também aos colegas de treino e companheiros de combate da Academia Impacto e da Equipe João Guedes, que me ajudaram nessa jornada, com diversos momentos insensatos, talvez até insanos, passando por treinos, campeonatos, treinos, competições, treinos, churrascos, treinos, comemorações, treinos e treinos. Muito obrigado por terem me acompanhado nesse período.

Aos meus colegas de professores do Estúdio de Dança Paulo Pinheiro. Vocês me

ensinaram a me divertir e aproveitar a as coisas boas da vida com muito mais intensidade! E são momentos como esses que fazem a vida valer a pena. Tudo isso enquanto me ensinavam a dançar (ou pelo menos tentavam)! Muito obrigado!

A todos os professores do curso de Ciência da Computação da UFpel, sou grato por todo o conhecimento e por todas as lições que aprendi com vocês, que farão de mim um profissional melhor e mais preparado. Em especial aos meus orientadores da graduação, Prof. Lucas e Prof. Ricardo, sou grato pelas lições e ensinamentos que recebi de vocês, essenciais à minha formação, tanto acadêmica quanto para a vida.

Obrigado também aos meus demais colegas, tanto da UFPel quando da UFRGS, pela amizade, pelo conhecimento e pelos links compartilhados durante todos esses anos.

A todas as outras pessoas que não lembrei de mencionar, mas que cruzaram a minha vida ou de alguma forma contribuíram para a conclusão desta jornada,

Muito Obrigado!

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS

IHC	interação humano-computador
3D	tridimensional
f.d.p.	função densidade de probabilidade
GMM	*Gaussian Mixture Model* (Modelo de Mistura de Gaussianas)
SVM	*Support Vector Machine* (Máquina de Vetores de Suporte)
GLCM	*Gray Level Co-occurrence Matrix* (Matriz de Co-ocorrência)
CIELAB	Espaço de cores CIE L*a*b*
RGB	*red* (vermelho), *green* (verde) e *blue* (azul), espaço de cores
SDC	*Skin Detection for Communication and Perceptual Interfaces*, dataset de imagens
BSDS300	*Berkeley Segmentation Dataset*, dataset de imagens
VC	Visão computacional
ROI	Região de interesse
rgb	espaço de dores RGB normaizado
DoG	Diferença de Gaussianas
MR8	*Maximum response 8*, banco de filtros
LoG	Laplaciano de Gaussiana
MR4	*Maximum response 4*, banco de filtros
STR	*Stochastic texture representation*, representação estocástica de texturas
MAP	Máxima propabilidade a posteriori
MRF	*Markov Random Fields*, campos aleatórios de Markov
TDR	*True Detection Rate*, taxa de verdadeiros positivos
FDR	*False Detection Rate*, taxa de falsos positivos
CTMS	*Color texture meansurement segmentation*
CPGS	*Combined patch and gradient segmentation*
SRM	*Stochastic region mergeing*

LISTA DE FIGURAS

LISTA DE TABELAS

RESUMO

A detecção de gestos é uma etapa importante em aplicações de interação humano-computador. Se a mão do usuário é detectada com precisão, tanto a análise quanto o reconhecimento do gesto de mão se tornam mais simples e confiáveis. Neste trabalho, descrevemos um novo método para detecção de pele humana, destinada a ser empregada como uma etapa de pré-processamento para segmentação de gestos de mão em sistemas que visam o seu reconhecimento. Primeiramente, treinamos os modelos de cor e textura de pele (material a ser identificado) a partir de um conjunto de treinamento formado por imagens de pele. Nessa etapa, construímos um modelo de mistura de Gaussianas (GMM), para determinar os tons de cor da pele e um dicionário de textons, para textura de pele. Em seguida, introduzimos um estratégia de fusão estocástica de regiões de texturas, para determinar todos os segmentos de diferentes materiais presentes na imagem (cada um associado a uma textura). Tendo obtido todas as regiões, cada segmento encontrado é classificado com base nos modelos de cor de pele (GMM) e textura de pele (dicionário de textons). Para testar o desempenho do algoritmo desenvolvido realizamos experimentos com o conjunto de imagens SDC, projetado especialmente para esse tipo de avaliação (detecção de pele humana). Comparado com outras técnicas do estado-da-arte em segmentação de pele humana disponíveis na literatura, os resultados obtidos em nossos experimentos mostram que a abordagem aqui proposta é resistente às variações de cor e iluminação decorrentes de diferentes tons de pele (etnia do usuário), assim como de mudanças de pose da mão, mantendo sua capacidade de discriminar pele humana de outros materiais altamente texturizados presentes na imagem.

Palavras-chave: Segmentação de imagens, fusão estocástica de regiões, reconhecimento de texturas, detecção de pele humana, segmentação de gestos de mão.

Skin Detection for Hand Gesture Segmentation via Multi-scale Stochastic Texture Models

ABSTRACT

Gesture detection is an important task in human-computer interaction applications. If the hand of the user is precisely detected, both analysis and recognition of hand gesture become more simple and reliable. This work describes a new method for human skin detection, used as a pre-processing stage for hand gesture segmentation in recognition systems. First, we obtain the models of color and texture of human skin (material to be identified) from a training set consisting of skin images. At this stage, we build a Gaussian mixture model (GMM) for identifying skin color tones and a dictionary of textons for skin texture. Then, we introduce a stochastic region merging strategy, to determine all segments of different materials present in the image (each associated with a texture). Once the texture regions are obtained, each segment is classified based on skin color (GMM) and skin texture (dictionary of textons) model. To verify the performance of the developed algorithm, we perform experiments on the SDC database, specially designed for this kind of evaluation (human skin detection). Also, compared with other state-of-the-art skin segmentation techniques, the results obtained in our experiments show that the proposed approach is robust to color and illumination variations arising from different skin tones (ethnicity of the user) as well as changes of pose, while keeping its ability for discriminating human skin from other highly textured background materials.

Keywords: Image segmentation, stochastic region merging, texture recognition, human skin detection, hand gesture segmentation.

1 INTRODUÇÃO

O uso de gestos é uma forma natural e intuitiva para comunicação entre pessoas, sendo caracterizada por movimentos significativos das mãos, braços, cabeça, rosto ou do corpo como um todo (YU et al., 2010). Esses movimentos podem ser empregados, dentre as diversas possibilidades, para auxílio a deficientes auditivos, reconhecimento de linguagens de sinais, monitoramento do estado emocional de um indivíduo, navegação e manipulação em ambientes virtuais, e outras formas de interação humano-computador (IHC) (MITRA; ACHARYA, 2007).

Segundo a Organização Mundial da Saúde há cerca de $360.000.000$ de pessoas com problemas auditivos no mundo (WHO FACT SHEETS, 2013), e para essas pessoas (principalmente para as que possuem deficiência auditiva) as linguagens de sinais tornam-se a forma natural de comunicação. Embora a linguagem de sinais seja uma ferramenta eficiente de comunicação para portadores de deficiência auditiva, é necessário que ambos, locutor e interlocutor, saibam utilizá-la. Logo, é comum que ocorram dificuldades na comunicação por linguagem de sinais com pessoas que não possuem o treinamento necessário para compreender seus gestos. Para facilitar a utilização de linguagens de sinais nesses casos, pode-se utilizar soluções tecnológicas que realizem o reconhecimento automatizado dos gestos, que podem então ser convertidos para texto ou voz (ZHOU et al., 2010; PAULRAJ et al., 2011).

Tão complexas quanto as linguagens faladas, as linguagens de sinais possuem milhares de sinais (que equivalem a palavras), e a menor mudança na posição, formato ou movimento da mão pode gerar um sinal com outro significado (KOSMIDOU; PETRANTONAKIS; HADJILEONTIADIS, 2011). Além disso, gestos podem variar de uma pessoa para outra, da mesma forma que seus significados dependem do contexto em que estão inseridos (MITRA; ACHARYA, 2007). Todos esses fatores contribuem para tornar o reconhecimento automatizado de gestos um problema complexo de ser resolvido. Em soluções baseadas em visão computacional, esse problema torna-se ainda mais complicado, pois, além de todas as dificuldades já mencionadas, é necessário detectar a posição da mão no espaço (MITRA; ACHARYA, 2007).

Esta tarefa torna-se bastante difícil e desafiadora devido às enormes variações de aparência que o ambiente e a mão em si podem sofrer. Resumidamente, as características do ambiente, como plano de fundo e a iluminação da cena, podem dificultar tanto o processo de identificação da mão na imagem tanto quanto as diferentes características que a mão em si pode assumir (aberta, fechada, em algum formato específico, etc.) e também as diferentes localizações dessa mão (próxima ou longe do corpo, acima ou abaixo dos ombros, etc.). A Figura 1.1 ilustra algumas das variações possíveis no cenário em que os gestos podem ocorrer.

Outra aplicação que pode ser explorada por meio de técnicas de reconhecimento de

(a) (b) (c) (d)

(e) (f) (g) (h)

Figura 1.1: Exemplos de diferentes situações e desafios na detecção de pele. (1.1a) e (1.1b) variações na ilumiação; (1.1c) etnia do usuário; (1.1d) variações na posição do usuário; (1.1e), (1.1f), (1.1g) e (1.1h) materiais no fundo da imagem;

gestos está na área de interação homem-máquina. Da mesma forma como os gestos de uma linguagem de sinais podem ser traduzidos em palavras, formando sentenças de linguagens naturais para comunicação entre pessoas, também é possível utilizá-los para controlar e interagir com ambientes virtuais.

Com o avanço e popularização da tecnologia e dos computadores digitais, a IHC se tornou presente em diversos aspectos da vida cotidiana. Nesse contexto, os mecanismos clássicos para interação, como *mouse* e teclado, nem sempre constituem a forma mais intuitiva de realizar essa tarefa. Em sistemas de auxilio à educação e aprendizado, por exemplo, essas interfaces dificilmente são capazes de fornecer a sensibilidade ou destreza desejada (RAUTARAY; AGRAWAL, 2012). Nesses sistemas de iteração com ambientes virtuais, o uso de gestos de mão pode contribuir para tornar a interface com a máquina muito prática e intuitiva. Com essa alternativa alcança-se o desempenho necessária para um melhor aproveitamento das capacidades destes ambientes dinâmicos.

Ao permitir que o usuário interaja livremente, sem a necessidade de contato físico com dispositivos periféricos, explorando os fenômenos instintivos de comunicação e manipulação, as técnicas de IHC alcançam um novo patamar (nível). Tornando-se ainda mais adequadas para metodologias de ensino, onde o uso de gestos é um grande facilitador. Também em termos de aprendizado, comandos complexos de aplicações dinâmicas podem ser executados de forma muito mais simples e intuitiva através de gestos de mão. Dentre os exemplos de aplicações que se beneficiam desse tipo de sistema, podemos citar a navegação em coleções de imagens, controlar apresentações de PowerPoint com uso intuitivo de gestos de mão, sem a necessidade de um estudo detalhado dos comandos envolvidos, ou mesmo contato físico com os dispositivos de captura de movimentos.

Para que um usuário possa interagir de forma eficaz e eficiente nestes ambientes virtuais dinâmicos, o uso destas interfaces deve ser natural e imersivo. Isso significa

que o uso das ferramentas de IHC deve minimizar a carga cognitiva e maximizar o sucesso do objetivo a ser atingido, seja em relação à precisão ou ao esforço para completa-lo (O'HAGAN; ZELINSKY, 2000). Em termos práticos, essas interfaces satisfazem esses critérios quando são intuitivas (usabilidade natural), poderosas (em termos de versatilidade e amplitude de aplicações) e de fácil aprendizado.

Algumas das soluções disponíveis atualmente utilizam dispositivos acoplados ao corpo do usuário, tais como acelerômetros, luvas de dados (LIANG; OUHYOUNG, 1998), ou ainda marcadores coloridos (EL-SAWAH; GEORGANAS; PETRIU, 2008). Com o auxílio desses dispositivos, o reconhecimento de gestos da mão torna-se muito mais simples, fornecendo uma localização espacial e estimativa tridimensional (3D) muito mais precisas. Entretanto, essas técnicas resultam em soluções pouco práticas e inconvenientes para o usuário.

Mesmo que o uso desses dispositivos periféricos permita uma maior flexibilidade na interação, quando comparados com os tradicionais teclado e *mouse*, o seu uso é permeado de vários fatores negativos, como estranheza, rigidez, falta de intuitividade e a tendência à distorção em relação aos ambientes reais. Além disso, esses dispositivos possuem um custo elevado, o que dificulta uma possível popularização do seu uso. Comparando esses dispositivos, podemos observar que ambos compartilham de alguns fatores que interferem negativamente em seu uso. No que diz respeito a flexibilidade e naturalidade dos movimentos, por exemplo, o peso e formato destes equipamentos geram limitações na destreza dos movimentos e nas poses de mão que o usuário é capaz de efetuar.

Tendo em vista estes aspectos, abordagens baseadas em visão computacional se tornam uma solução muito mais adequada para resolver a tarefa de determinar os gestos desempenhados pelo usuário. A vantagem mais saliente nesse tipo de sistema é o baixo custo para sua utilização, requerendo unicamente uma câmera. Além disso, uma vez que a interação ou comunicação do usuário é capturada sem a necessidade de contato com dispositivos periféricos, seu uso se torna mais natural e intuitivo, além de proporcionar mais liberdade na movimentação do usuário.

O reconhecimento de gestos baseado em visão computacional geralmente é composto de uma etapa inicial de localização espacial (chamada detecção) do gesto e posteriormente realiza-se o reconhecimento do mesmo a partir de algum conjunto de características globais da região encontrada. Uma vez que a posição da mão do usuário é detectada, pode-se estimar as características que compõem os aspectos físicos da pose em uma cena real, como orientação global e os ângulos das articulações. Além disso, ao incorporar a informação temporal presente em imagens dinâmicas (vídeos), pode-se estimar a posição em relação ao corpo e a trajetória desempenhada, obtendo as informações que constituem gestos dinâmicos.

Efetuar esta estimativa das características, no entanto, não é uma tarefa trivial, já que ela possui uma forte dependência de uma etapa de processamento prévia: a detecção dos gestos na cena. Essa etapa constitui uma tarefa complexa e desafiadora, primeiramente porque, ao fornecer ao usuário a liberdade de movimento desejada, também aumenta a variedade de poses que podem ser desempenhadas. Além disso, é necessário lidar com a influência de fatores do ambiente e iluminação, além das variações intra e inter-usuário.

Tanto no caso de imagens estáticas quanto de vídeos (sequências de imagens), o objetivo da detecção é determinar a posição espacial da mão do usuário. Algoritmos de detecção e rastreamento tradicionais (HAN; AWAD; SUTHERLAND, 2009) geralmente delimitam uma área (*bounding box*) dentro da qual o objeto se encontra. Entretanto, algumas abordagens mais recentes (PHUNG; BOUZERDOUM; CHAI, 2003; JIANG; YAO;

JIANG, 2007) buscam realizar uma detecção mais precisa, identificando com precisão a forma das mãos do usuário. Ao introduzir técnicas de segmentação mais precisas, o reconhecimento dos gestos torna-se mais confiável, pois a própria forma da mão passa a ser uma feição que pode ser utilizada, ampliando a quantidade de informações disponível para a classificação. A ideia principal desse tipo de solução é que o usuário possa interagir sem a necessidade de qualquer contato físico com dispositivos periféricos, com acelerômetros, luvas, ou outro aparato de rastreamento. Sendo assim, sistemas de visão computacional para rastreamento dos gestos costumam utilizar características como cor e textura da pele humana (MITRA; ACHARYA, 2007) para localizar a mão do usuário.

Na literatura existem diversos métodos que propõem a detecção de pele a partir de características de cor (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007). Alguns utilizam modelos universais e estáticos em diferentes espaços de cores, tais como RGB (KOVAC; PEER; SOLINA, 2003), HSV (DARDAS; GEORGANAS, 2011), ou YCbCr (YU et al., 2010). Outros tentam obter modelos de cor de pele através de algoritmos de aprendizado de máquina supervisionados, como classificador de Bayes (YANG; LU; WAIBEL, 1997; KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007) ou *support vector machines* (máquinas de vetores de suporte, SVM) (HAN; AWAD; SUTHERLAND, 2009).

Mesmo com algoritmos de classificação robustos, as técnicas baseadas em cor de pele são propensas a confundirem pele com outros materiais que frequentemente exibem uma tonalidade similar, como areia ou madeira. Para atenuar esse fator, foram propostos métodos que utilizam feições de textura para detectar as regiões de pele (CROSS; JAIN, 1983). Esses métodos atingem taxas de acerto maiores ao considerar a informação presente na vizinhança, ao invés do valor dos *pixels* somente.

Finalmente, há ainda abordagens que sugerem o uso de segmentação de imagens orientada a bordas juntamente com as detecções por cor (PHUNG; BOUZERDOUM; CHAI, 2003) e textura (JIANG; YAO; JIANG, 2007). Esses trabalhos mostram que, ao introduzir uma etapa de segmentação da imagem, a detecção de pele torna-se capaz de identificar os limites das regiões de pele com muito mais precisão. Métodos com essa característica tornam-se muito interessantes para sistemas de reconhecimento de gestos de mão, pois permitem que a própria forma do gesto seja utilizada para a classificação.

Embora os métodos mencionados consigam atingir resultados interessantes pelo uso de descritores de textura, trabalhos recentes demostraram que uma representação ainda mais precisa das texturas pode ser obtida por meio de descritores estocásticos locais de texturas (LIU; FIEGUTH, 2012). Além disso, esses descritores podem ser combinados com um modelo de aparência global das texturas (VARMA; ZISSERMAN, 2009) para obter uma representação ainda mais rica e detalhada.

Por outro lado, os resultados dos trabalhos disponíveis na literatura (PHUNG; BOUZERDOUM; CHAI, 2003) também indicam que a detecção de pele também torna-se mais precisa quando introduzimos técnicas de segmentação que consideram as bordas presentes na imagem (JIANG; YAO; JIANG, 2007). Especialmente ao utilizar técnicas de crescimento de regiões baseadas nos gradientes da imagem, vemos que a forma externa das regiões de pele é respeitada mais facilmente.

Da mesma forma como os descritores de texturas, novos métodos para segmentação de imagens também têm sido propostos nos últimos anos. Em especial estratégias baseadas em fusão de regiões (NOCK; NIELSEN, 2004; WONG; SCHARCANSKI; FIEGUTH, 2011) têm demonstrado desempenho considerável, mesmo ao empregar características simples, como intensidade ou cor média dos *pixels* na região.

Motivado pelos ganhos obtidos em representações de textura, bem como pelas abordagens de fusão de regiões, e os relatos dos demais métodos que empregam técnicas de segmentação de imagens para detecção de pele, este trabalho propõe uma nova estratégia para a detecção de regiões de pele.

Primeiramente, buscamos definir uma forma ampla e robusta de descrever as texturas encontradas nas imagens. Para isso, fazemos uso de feições estocásticas multi-escala de texturas (LIU; FIEGUTH, 2012) descrevendo características em baixo nível, combinados com modelos de aparência globais (VARMA; ZISSERMAN, 2009) para descrição em alto nível. A partir dessa representação de textura, este trabalho propõe um método para realizar a segmentação das regiões de textura na imagem.

Inspirado nas novas abordagens de segmentação por estratégias de fusão de regiões a partir de critérios adaptativos (estatísticos (NOCK; NIELSEN, 2004) ou estocásticos (WONG; SCHARCANSKI; FIEGUTH, 2011)), propomos um método com o potencial para diferenciar entre as variações locais abruptas nas regiões altamente texturadas, das variações de cor e iluminação que ocorrem ao longo da imagem.

Finalmente, propomos uma técnica para detectar as regiões de pele segmentadas na última etapa, a partir de medidas de similaridade com os modelos de pele baseados em cor, obtido pelo treinamento de um modelo de mistura de gaussianas (GMM) e textura, obtido conforme o modelo de textura definido na etapa de treinamento.

1.1 Objetivos

Definimos como objetivo geral deste trabalho a elaboração e descrição de uma técnica de segmentação de regiões de pele apropriada para sistemas de reconhecimento de gestos de mão. Para isso, é necessário que a técnica aqui apresentada possua as seguintes características:

- robustez às variações locais de regiões altamente texturadas;
- robustez às variações de tom de pele intra e inter-usuário, bem como às variações geradas por diferentes fontes e ângulos de iluminação;
- capacidade de diferenciação entre pele e materiais de coloração similar como, por exemplo, areia, madeira, chamas ou pelos de animais amarelos;
- capacidade de identificar corretamente os limites das regiões de pele, delimitando com precisão sua forma externa.

Para atingir esse objetivo subdividimos estas tarefas dentre os seguintes objetivos específicos:

I Elaborar uma representação das texturas presentes nas imagens.

II Desenvolver uma técnica de segmentação de imagens baseada nas características de textura nela apresentadas.

III Combinar a representação de texturas com uma representação de cor para determinar um modelo que descreva adequadamente as características da pele humana. O modelo obtido deve ser independente de fatores como a etnia do usuário, fonte e posição da iluminação da cena e demais objetos presentes no fundo da imagem.

1.2 Organização do Trabalho

Este trabalho está estruturado em 7 capítulos, como descrito a seguir. No Capítulo 2 são descritos brevemente os principais trabalhos relacionados ao problema que nos propusemos a solucionar. Esses trabalhos compõem o estado-da-arte da técnicas de detecção de pele, dentre os quais iremos comparar nossos resultados.

No Capítulo 3, realizamos uma breve revisão dos conceitos teóricos de reconhecimento de padrões e segmentação de imagens na medida necessária para a compreensão do método a ser desenvolvido. Dentre os tópicos selecionados, contam alguns modelos estatísticos de reconhecimento de padrões, tais como classificador bayesiano, K-means e mistura de gaussianas. Estão também inclusas algumas técnicas de segmentação, como limiarização e fusão de regiões.

Analogamente, o Capítulo 4 trata dos conceitos de representação de imagens, que se fazem necessários para o entendimento da solução proposta. Para tal, estudamos alguns espaços de cores, usados para representar a informação cromática da imagem, assim como os principais métodos de decomposição em múltiplos níveis de detalhes, como a pirâmide laplaciana, diferenças de gaussianas e *wavelets*. Por fim, discutimos algumas formas representação de texturas, como matrizes de coocorrência, bancos de filtros, janelas da imagem e dicionários de textons.

O Capítulo 5 destina-se à descrição detalhada do método de detecção e segmentação de pele proposta neste trabalho. Inicialmente são apresentados as feições e modelos de textura utilizados e, em seguida, a estratégia de fusão estocástica de regiões para segmentar a imagem. Por último, expomos a técnica de classificação utilizada para identificação dos segmentos de pele.

O Capítulo 6 apresenta e discute alguns resultados obtidos com o a metodologia proposta, bem como os efeitos de diferentes configurações de seus parâmetros. Primeiramente, realizaremos uma avaliação da técnica de segmentação de texturas proposta neste trabalho como uma ferramenta para segmentação de propósito geral. Para isso, realizamos testes em dois conjuntos de imagens, Prague (imagens sintéticas de texturas) e BSDS300 (imagens naturais), projetados especificamente para avaliar métodos de segmentação de texturas. Também são apresentadas neste capítulo comparações, quantitativas e visuais, com outros métodos do estado-da-arte em segmentação de imagens. Em seguida, realizamos uma avaliação da técnica de identificação de regiões de pele. Os experimentos relatados são referentes a testes realizados com o conjunto de imagens SDC, que foi projetado especialmente para avaliar o desempenho de técnicas de detecção de pele. Neste capítulo, também realizamos comparações, quantitativas e visuais, com outros métodos selecionados do estado-da-arte.

Finalmente, no Capítulo 7, são apresentadas as conclusões dos estudos e experimentos obtidas neste trabalho. Também neste capítulo, colocamos algumas propostas e sugestões de trabalhos futuros. Como fechamento deste trabalho, apresentamos as publicações obtidas e submetidas durante o desenvolvimento do trabalho de pesquisa aqui proposto.

2 TRABALHOS RELACIONADOS

Ao longo dos anos, foram propostas diversas abordagens para detecção e segmentação de gestos de mão a partir do uso de modelagem estatística das características visuais da pele humana. Alguns trabalhos utilizam modelos universais de cor de pele para identificar a região onde o gesto ocorre na cena. Os espaços de cores RGB (KOVAC; PEER; SOLINA, 2003), HSV (DARDAS; GEORGANAS, 2011) ou YCbCr (YU et al., 2010) são mais comumente utilizados nesse processo, mas também é possível utilizar alguns outros espaços de cores (ZHENGMING; TONG; JIN, 2010). Por meio de combinações entre os valores dos canais de cromaticidades, luminância e brilho dessas representações de cores, é possível obter relações que classificam cada *pixel* entre pele ou fundo (*background*).

Estes métodos geralmente utilizam uma série de testes lógicos que determinam a classe a partir da magnitude e das diferenças entre os valores de cor do *pixel*, Dessa forma determina-se um subespaço de cores capaz de identificar a pele humana independentemente da etnia do usuário. Esses métodos oferecem como grande vantagem a velocidade de execução, sendo possível utilizá-los facilmente em tempo real. Entretanto, essa baixa complexidade computacional possui um grande custo de eficácia. Dada a simplicidade dessa técnica, a taxa de acertos encontrada tende a ser muito baixa.

Na detecção de pele por meio de limiares de valores RGB (KOVAC; PEER; SOLINA, 2003), isso ocorre porque essa representação de cores, embora prática para visualização, não é a mais adequada para análise de imagens. Nela, os valores dos *pixels* sofrem uma influência muito grande da iluminação presente na cena, uma vez que essa informação está distribuída entre todos os canais de cor. Já nas abordagens que utilizam modelos universais em outros espaços tais como HSV (DARDAS; GEORGANAS, 2011), YCbCr (YU et al., 2010) ou RGB normalizado (ZHENGMING; TONG; JIN, 2010), é possível reduzir a influência desse fator. Nestas representações, o fator de iluminação encontra-se separado do aspecto cromático. Assim, é possível detectar a pele através das características de cor de forma invariante à iluminação do ambiente.

Uma forma de tornar o modelo de cor de pele mais preciso consiste em, além de selecionar uma representação de cores adequada ao problema, também empregar técnicas de classificação supervisionada para construir o modelo de cor de pele. Esses classificadores geralmente buscam minimizar uma função de custo de erro, assim a função de classificação passa a se ajustar de forma mais precisa ao conjunto de dados de treinamento.

Como detalhado por Kakumanu et al. (2007) pode-se utilizar classificadores bayesianos para tomar essa decisão. Esses métodos tentam encontrar uma função densidade de probabilidade (f.d.p.) que se ajuste às amostras de treinamento de cada classe. Uma vez que as f.d.p.s tenham sido estimadas, a classe de cada teste é obtida aplicando o teorema de Bayes (WEBB, 2002).

Há na literatura várias abordagens disponíveis para estimação da f.d.p. de um con-

junto de dados. Uma forma simples e intuitiva é utilizar histogramas de cores, $2D$ ou $3D$, que geralmente permitem estimativas precisas e estáveis, não sendo afetados por oclusões, mudanças de vista, podendo serem usados para diferenciar um grande número de objetos (YANG; LU; WAIBEL, 1997).

Outra possibilidade para obter a f.d.p. dos dados é estimar os parâmetros de uma função de probabilidade conhecida, como uma gaussiana, ou ainda, em uma abordagem mais elaborada uma mistura de gaussianas (GMM) (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007). Embora um histograma de cores geralmente seja uma estimativa mais robusta da f.d.p. dos tons de cor de pele, a vantagem de utilizar modelos paramétricos reside na sua capacidade de serem gerados confiavelmente com um conjunto de treinamento menor.

Outros trabalhos que utilizam aprendizado supervisionado para obter um modelo que determina o tom de cor da pele fazem uso de técnicas estatísticas ainda mais refinadas, como SVM (HAN; AWAD; SUTHERLAND, 2009). Esse modelo matemático de classificação busca encontrar os vetores de suporte que definem um hiperplano de separação entre as classes (VAPNIK, 1995). Assim como nos métodos que empregam classificadores paramétricos, também é necessário realizar uma etapa prévia de treinamento da SVM, em que são determinados os vetores de suporte.

Embora estes métodos consigam alcançar resultados interessantes, o uso de feições de cor isoladamente é capaz de fornecer apenas uma representação pobre das características da pele. Isso ocorre principalmente devido ao fato de que os materiais presentes no fundo da imagem frequentemente exibem tons de cor similares aos da pele humana e acabam sendo confundidos pelos classificadores. Ao realizar a detecção de pele com base em características pontuais de cor (valor do *pixel*), estamos determinando regiões do espaço de cores (seja qual for), em que o resultado do classificador informa entre as classes "pele"ou "fundo". Essa análise, por considerar as informações dos *pixels* de forma muito simples, embora intuitiva, frequentemente causa sobreposição da classe pele com outros tipos de materiais que, mesmo possuindo uma aparência muito distinta, possuem tons de cor similares, como areia, chamas, flores, pelos amarelados ou madeira, citando apenas alguns exemplos.

Para contornar essas complicações, alguns trabalhos propõem que a detecção de pele seja realizada a partir de características de textura. Diferentemente das feições de cor, que consideram o valor de cada *pixel* individualmente, a ideia principal das feições de textura é utilizar os valores dos *pixels* vizinhos juntamente com o do *pixel* analisado para realizar a classificação (CROSS; JAIN, 1983). Dessa forma, mesmo que apenas características locais sejam utilizadas, ainda assim a classificação torna-se mais robusta, já que passamos a considerar aspectos de aparência. Mesmo que observados apenas localmente, para cada *pixel* separadamente, essas feições permitem identificar uma gama muito maior de materiais.

Algumas abordagens disponíveis na literatura propõem que seja utilizada uma combinação de feições de cor e textura. A abordagem discutida em Wang et al. (2011) emprega uma combinação das representações de cor em RGB e YCgCb, e matrizes de coocorrência de tons de cinza (*gray level co-occurrence matrix*, GLCM) como representação das características de textura. Esse trabalho utiliza estatísticas no espaço de cores RGB para determinar um modelo de cor através de limiares estáticos universais. Já para o espaço YCgCb, o modelo é definido como um círculo com centro e raio fixados em relação aos componentes de cromaticidades. A GLCM representa as texturas a partir da probabilidade conjunta de ocorrência simultânea de dois tons de cinza em *pixels* das posições

(x, y) e $(x + \Delta_x, y + \Delta_y)$, estimada a partir de um histograma bi-variado. Com os valores observados nesse histograma, são calculadas algumas características da textura, tais como contraste, momentos angulares, entropia, correlação e homogeneidade. Esse método consegue atingir resultados interessantes, pois mostra que o uso conjunto de cor e textura fornece uma classificação mais robusta, mesmo com modelos simples de cor e textura. A GLCM, entretanto, é uma técnica muito limitada para representação de texturas. Além de ser sensível à escala e rotação da textura, seu custo computacional tanto em tempo de computação quanto em memória é muito elevado. Isso torna seu uso pouco prático.

Outra forma de utilizar a informação textural presente na imagem para segmentação de pele é através de filtros de Gabor. No trabalho de Jiang et al. (2007), esses filtros são usados para decompor a imagem em vários níveis de detalhes (diferenças locais), através de sucessivas filtragens gaussianas aplicadas em várias direções. Para realizar a segmentação de pele, primeiramente é utilizado um classificador ingênuo de Bayes (*naive Bayes classifier*) a partir do histograma de cor dos *pixels*. Sob a suposição de que as regiões de pele apresentam transições suaves, esse método detecta pele nas regiões que possuem baixas diferenças na decomposição. Assim, a imagem é segmentada por crescimento de regiões utilizando a técnica morfológica *watershed*, adaptada para utilizar os filtros de Gabor como diferenças locais da imagem. Ao utilizar um aprendizado supervisionado para obter o modelo de cor, bem como um conjunto de características de textura mais eficiente, esse método consegue atingir resultados superiores aos de métodos que analisam *pixels* individualmente. Outra conclusão interessante desse trabalho é que o uso de técnicas de segmentação permite que o resultado obtido seja mais preciso. Além disso, ao empregar uma técnica baseada em gradientes para essa tarefa, a detecção de pele se torna mais propensa a respeitar as bordas da imagem. Com isso, as regiões de pele são segmentadas com mais precisão em relação ao seu formato, o que é muito desejável para o reconhecimento de gestos, como discutido anteriormente.

Com toda a atenção recebida da literatura ao longo das anos, os métodos até então propostos para detecção de pele exploraram com profundidade apenas as informações de cor presentes nas imagens. Dentre os poucos métodos que empregam informações texturais para realizar essa tarefa, notamos claramente que os descritores de textura neles empregados mostram-se defasados em comparação com o estado da arte em reconhecimento de texturas. Trabalhos recentes, por exemplo, demostraram que descritores de textura muito mais poderosos são obtidos ao se utilizar janelas de imagens (VARMA; ZISSERMAN, 2009) ou feições estocásticas de texturas (LIU; FIEGUTH, 2012), em especial quando combinados com descritores de aparência global (LEUNG; MALIK, 2001).

Da mesma forma, os métodos do estado-da-arte em detecção de pele também mostram que uma detecção muito mais precisa pode ser obtida ao se considerar as informações de bordas (gradientes locais) presentes nas imagens. Mas, novamente, os já realizados estudos nesse sentido apenas empregam técnicas de segmentação de imagens que já não pertencem ao estado da arte. Nos últimos anos, por exemplo, foram propostos diversos algoritmos para segmentação por fusão de regiões que alcançam segmentações muito mais robustas utilizando critérios estatísticos (NOCK; NIELSEN, 2004) ou estocásticos (WONG; SCHARCANSKI; FIEGUTH, 2011; MEDEIROS; SCHARCANSKI; WONG, 2012).

3 RECONHECIMENTO DE PADRÕES

Reconhecimento de padrões é uma área de pesquisa que tem por objetivo a classificação de objetos (padrões) em um número de categorias ou classes através de um conjunto de propriedades ou características. O termo padrão pode ser usado para denotar as p-dimensões de dados do vetor $\mathbf{x} = (x_1, ..., x_p)^T$, cujos componentes x_i são medidas ou valores das características do objeto. Assim, as características são variáveis especificadas pelo investigador sendo importantes para a classificação. Na discriminação, assume-se que exista C grupos ou classes, denotados por $\omega_1, ..., \omega_C$, e associados a cada padrão $\mathbf{x}$ que é uma variável categórica z que indica o membro da classe ou grupo pertencente, isto é, se $z = i$, então o padrão pertence a ω_i, $i \in \{1, ..., C\}$ (WEBB, 2002).

Exemplos de padrões são medições de uma forma de onda acústica no problema de reconhecimento de fala; medições feitas em um paciente a fim de identificar uma doença (diagnóstico); medições em pacientes, no sentido de identificar o resultado mais provável (prognóstico); medidas meteorológicas para previsões de tempo; e uma imagem digitalizada para o reconhecimento de caracteres (WEBB, 2002).

Existem dois principais tipos de classificação: classificação supervisionada e classificação não-supervisionada. Na classificação supervisionada temos um conjunto de dados amostrais (cada um consistindo de medições em um conjunto de variáveis) rotulados pelo tipo de classe que são usados como exemplos para o projeto de classificador. Na classificação não-supervisionada, os dados não são rotulados e buscamos encontrar grupos de dados e características que distinguem um grupo do outro (WEBB, 2002).

Neste capítulo, veremos os conceitos básicos das técnicas de reconhecimento de padrões, assim como seu emprego na área de visão computacional. Dentre os assuntos tratados, veremos técnicas de abordagens estatísticas para separação e agrupamento de dados e como essas técnicas são aplicadas a imagens. Finalmente, também veremos brevemente o conceito de segmentação de imagens e algumas formas como os algoritmos de reconhecimento de padrões podem ser utilizados para essa tarefa.

3.1 Visão Computacional e Segmentação de Imagens

Visão computacional (VC) é entendida como o conjunto de técnicas para adquirir, processar, analisar e compreender dados através de métodos que buscam reproduzir as características do nosso sistema visual (JAHNE; HAUSSECKER; GEISSLER, 1999). Uma tarefa comum em VC consiste em realizar a identificação e padronização das complexidades presentes nos dados capturados do mundo real. Em visão computacional, o que tentamos fazer é descrever o mundo que vemos pelo uso de imagens e da reconstrução de suas propriedades, tais como forma, iluminação, e distribuição de cor (SZELISKI, 2010).

Por esse motivo diversas técnicas de reconhecimento de padrões são amplamente apli-

cadas em sistemas de visão computacional. Como exemplo dessas combinações, entre as duas áreas, podemos mencionar a identificação e detecção de materiais, reconhecimento de objetos e rastreamento de feições em vídeos (imagens dinâmicas). Para uma correta identificação dos padrões desejados em imagens, muitas vezes é necessário realizar uma etapa prévia à análise dos dados que consiste em segmentar a imagem (SONKA; HLAVAC; BOYLE, 2007), para obter partições homogêneas da mesma. Após realizar esse processo, as informações no interior de cada região são utilizadas para representar suas feições e essas são então analisadas para reconhecer os padrões visuais encontrados na imagem.

Segmentação de imagens é um dos tópicos mais antigos e desafiadores na área de visão computacional, tendo sido amplamente estudado ao longo dos anos (SZELISKI, 2010). Essas técnicas lidam com o particionamento de uma imagem em um conjunto de segmentos (ou regiões) disjuntos, de forma que *pixels* com características similares sejam agrupados em segmentos homogêneos. Em um nível de análise mais elevado do resultado do processo de segmentação de imagens, temos também como objetivo que as regiões encontradas possuam elevada correlação com os objetos ou áreas do mundo real presentes na imagem (SONKA; HLAVAC; BOYLE, 2007).

Do ponto de vista estatístico, este problema é tratado como análise de agrupamentos (*cluster analysis*) e tem sido amplamente estudado, resultando em uma extensa série de diferentes algoritmos (KAUFMAN; ROUSSEEUW, 1990; JAIN; DUIN; MAO, 2000; JAIN; DUBES, 1988; JAIN et al., 2004).

As abordagens mais clássicas para segmentação de imagens em geral utilizam técnicas baseadas em regiões, tanto por divisão quanto por fusão de regiões. Os métodos de divisão buscam determinar as regiões a partir de sucessivas subdivisões da imagem até que todas as regiões restantes sejam homogêneas. Já os algoritmos de fusão de regiões realizam o processo inverso, ou seja, começando com pequenas regiões, que podem ser consideradas trivialmente homogêneas, encontram-se os segmentos finais da imagem a partir de sucessivas fusões entre as regiões nela presentes inicialmente. Essas técnicas correspondem a algoritmos divisivos e aglomerativos da literatura de técnicas de agrupamento (SZELISKI, 2010).

3.1.1 Limiarização

A técnica de limiarização geralmente é tida como a forma mais simples de realizar a segmentação de imagens. Muitas vezes, os objetos ou regiões de interesse (ROI) presentes na imagem podem ser identificados pelo nível de refletividade ou absorção luminosa de sua superfície. Nesses casos, uma constante de brilho ou limiar pode ser determinado para segmentar os objetos desejados (SONKA; HLAVAC; BOYLE, 2007).

Para uma dada imagem f, podemos definir a segmentação desta imagem R, como um conjunto finito de regiões $R_1, ..., R_S$,

$$R = \bigcup_{i=1}^{S} R_i, \qquad R_i \bigcap R_j = \emptyset, \qquad i \neq j. \tag{3.1}$$

Em cenas simples, esse resultado pode ser obtido por meio de limiarização. Essa técnica pode ser definida como uma transformação da imagem f para uma imagem binária g (segmentação), da seguinte maneira:

$$g(i,j) = \begin{cases} 1 & \text{if } f(i,j) \geq T, \\ 0 & \text{if } f(i,j) < T, \end{cases} \tag{3.2}$$

onde T é o valor do limiar, $g(i,j) = 1$ para *pixels* do objeto ou ROI, e $g(i,j) = 0$ para *pixels* do fundo (ou vice-versa).

Se os objetos não se tocam (i.e. não possuem *pixels* adjacentes) e se os níveis de cinza são claramente distintos dos tons de cinza do fundo, essa técnica pode ser considerada adequada. Por ser uma técnica muito simples, a limiarização é rápida e computacionalmente barata, sendo amplamente utilizada em aplicações simples (SONKA; HLAVAC; BOYLE, 2007).

A escolha adequada (correta) do valor de T, entretanto, torna-se crucial para que essa técnica obtenha sucesso na segmentação. Ainda assim, a limiarização simples pode ser capaz de segmentar corretamente uma imagem apenas sob circunstâncias muito incomuns. Mesmo em imagens muito simples há a possibilidade das variações de níveis de cinza dos objetos e do fundo se sobreporem devido à iluminação não uniforme, aos parâmetros do equipamento de aquisição, dentre outros fatores (SONKA; HLAVAC; BOYLE, 2007). Esses fatores podem ser atenuados (embora dificilmente eliminados) pela realização de um pré-processamento na imagem, de forma que os valores dos *pixels* nas regiões de interesse se tornem diferentes dos demais.

3.1.2 Fusão de Regiões

Uma forma muito natural e intuitiva de segmentar uma imagem, ao menos do ponto de vista de VC, é tentar agregar *pixels* vizinhos que possuam características semelhantes, assim formando os segmentos da imagem. Esses métodos costumam ser relativamente robustos em imagens ruidosas ou, que por algum outro fator, possuem bordas (fronteiras entre as regiões) difíceis de ser detectadas (SONKA; HLAVAC; BOYLE, 2007).

Constando dentre os métodos mais antigos de visão computacional (SZELISKI, 2010), nos métodos baseados em regiões busca-se atingir a segmentação da imagem de forma que o critério de *homogeneidade* de cada região seja satisfeito. Homogeneidade é uma propriedade importante das regiões e é usado como o principal critério de segmentação no crescimento de regiões, cuja idéia básica é dividir a imagem em zonas de máxima homogeneidade (SONKA; HLAVAC; BOYLE, 2007).

O critério de homogeneidade a ser satisfeito pode ser baseado no nível de intensidade luminosa, tom de cor, textura, formato, modelo de informação semântica, para citar alguns exemplos. Dessa forma, as propriedades escolhidas para descrever as regiões irão influenciar a forma, complexidade e quantidade de informação *a priori* necessária para segmentar as regiões de interesse especificadas (SONKA; HLAVAC; BOYLE, 2007).

A partir da Equação (3.1), que define os requisitos básicos de uma segmentação de imagens, podemos determinar que as regiões obtidas devem satisfazer as seguintes condições:

$$H(R_i) = \text{VERDADEIRO}, \qquad i = 1, 2, \cdots, S \tag{3.3}$$

$$H(R_i \cap R_j) = \text{FALSO}, \qquad i \neq j, \qquad R_i \text{ é adjacente à } R_j, \tag{3.4}$$

onde S é número total de regiões na imagem e $H(R_i)$ é a avaliação binária de homogeneidade da região R_i. Dessa forma, as regiões encontradas na segmentação devem ser ao mesmo tempo homogêneas e 'máximas', no sentido de que o critério de homogeneidade não mais seria satisfeito após fundir quaisquer duas regiões adjacentes (SONKA; HLAVAC; BOYLE, 2007).

Baseando-se nestas propriedades, uma forma de segmentar a imagem é agregar regiões adjacentes sucessivamente por meio de algum critério de homogeneidade. Baseado

nesses conceitos podemos definir as técnicas de crescimento de regiões pela seguinte estrutura geral (SONKA; HLAVAC; BOYLE, 2007):

1. Definir uma inicialização para a segmentação da imagem, dividindo-a em pequenas regiões satisfazendo a Equação (3.3).

2. Definir um critério de homogeneidade $H(R_i, R_j)$ para unir regiões adjacentes R_i e R_j.

3. Fundir todas as regiões adjacentes que satisfaçam $H(R_i, R_j)$.

4. Se não houver quaisquer duas regiões que possam se unidas mantendo a condição da Equação (3.3), pare, senão volte para o passo 3.

Esse algoritmo representa uma abordagem geral para fusão de regiões. Métodos específicos diferem quanto à forma de inicialização das regiões ou quanto ao critério de união. O resultado dessa segmentação, todavia, depende fortemente da ordem em que os pares de regiões são analisados e, consequentemente, unidos. Isso significa que conjuntos diferentes de regiões serão obtidos se o processo iniciar, por exemplo, no canto superior esquerdo ou no canto inferior direito da imagem. Isso pode ocorrer por que, ao se comparar duas regiões similares e adjacentes R_1 e R_2, uma fusão anterior de R_1 pode fazer com que suas novas características a tornem incompatíveis com R_2. Note que, caso a fusão acontecesse em outra ordem, essa fusão poderia ter sido realizada (SONKA; HLAVAC; BOYLE, 2007).

A forma mais simples de inicialização das regiões é dividir a imagem em pequenos blocos de tamanho 2×2, 4×4 ou 8×8 *pixels*. Dessa forma, podemos utilizar estatísticas, como um histograma dos níveis de cinza ou a cor média das regiões, para descrevê-las de forma mais robusta. Logo, ao realizar o processo de fusão, duas regiões adjacentes são unidas se suas características correspondem entre si (ultrapassam um limiar).

Pode-se ainda definir outras formas de inicialização do mapa de regiões, tais como utilizar cada *pixel* como uma região separada (SZELISKI, 2010; WONG; SCHARCANSKI; FIEGUTH, 2011) ou empregar alguma técnica de segmentação como *superpixels* (MORI et al., 2004) que encontrem pequenos segmentos, mesmo não sendo semanticamente significativos, apresentem homogeneidade interna, satisfazendo a Equação (3.3).

Outros métodos utilizam critérios mais elaborados para realizar a fusão das regiões. Nock e Nielsen (2004) demonstraram que uma segmentação significativa pode ser obtida utilizando um critério estatístico para $H(R_i, R_j)$. Nessa abordagem, o critério de união das regiões relaciona não somente a similaridade entre seus descritores (estatísticas das regiões), mas também informações relativas ao tamanho dessas regiões. De maneira similar, Wong et al. (2011) propôs um método que incorpora a diferença e área da imagem ocupada pelas regiões analisadas, porém de maneira estocástica. Assim, o resultado das fusões se torna mais robusto à possíveis ruídos de ordenação dos pares de regiões e ruído presentes na imagem, o que torna a segmentação menos propensa a ser guiada para máximos locais.

Finalmente, trabalhos recentes demostraram que segmentações ainda mais significativas e robustas podem ser obtidas combinando uma fusão estocástica de regiões com feições de textura que representem eficientemente a aparência das regiões (MEDEIROS; SCHARCANSKI; WONG, 2012). Além disso, o trabalho de Medeiros et al. (2012; 2013a) também mostra que o uso de uma ordenação adequada dos pares de regiões leva a uma melhora significativa da segmentação obtida.

3.2 Modelos Estatísticos

Para o reconhecimento de padrões, podemos usar a estatística para projetar classificadores e realizar a separação dos dados. Esses modelos estatísticos podem ser construídos a partir das informações de médias, dispersão (covariância) e distribuições probabilísticas dos padrões a serem reconhecidos, a fim de inferir sobre a qual classe i a variável observada pertence. Os modelos de reconhecimento de padrões estatísticos usados neste trabalho são apresentados nas seções a seguir.

3.2.1 Classificador de Bayes - Regra de Decisão de Bayes

O estimador de Bayes é um estimador ou regra de decisão que mapeia um dado observado para uma classe apropriada (WEBB, 2002).

Considere C classes, $\omega_1, ..., \omega_C$ com probabilidades a *priori* $p(\omega_1), ..., p(\omega_C)$ (probabilidade de ocorrência de cada classe) conhecidas (WEBB, 2002). Se desejarmos minimizar a probabilidade de erro de classificação entre as classes e não temos outras informações acerca do objeto observado que não seja a distribuição probabilística de cada classe, então nós atribuíamos um objeto a classe ω_j se:

$$p(\omega_j) > p(\omega_k)\ para\ k = 1, ..., C; k \neq j. \tag{3.5}$$

Isto classifica todos os objetos em uma das classes (WEBB, 2002). No entanto, se nós temos um vector de observação $\boldsymbol{x}$ e desejamos atribuir $\boldsymbol{x}$ a umas das C classes. A regra de decisão de Bayes (WEBB, 2002) baseia-se em atribuir $\boldsymbol{x}$ a classe ω_j se a probabilidade da classe ω_j dado a observação $\boldsymbol{x}$, $p(\omega_j|\boldsymbol{x})$, é maior sobre todas as classes $p(\omega_1), ..., p(\omega_C)$. Isto é, atribui-se $\boldsymbol{x}$ a classe ω_j se:

$$p(\omega_j|\boldsymbol{x}) > p(\omega_k|\boldsymbol{x})\ para\ k = 1, ..., C; k \neq j. \tag{3.6}$$

Esta regra de decisão (WEBB, 2002) divide o espaço calculado dentro de C regiões $\Omega_1, ..., \Omega_C$ tal que se $\boldsymbol{x} \in \Omega_j$ então $\boldsymbol{x}$ pertence a classe ω_j.

A probabilidade a *posteriori* (WEBB, 2002) $p(\omega_j|\boldsymbol{x})$ pode ser expressada em termos da probabilidade a *priori* e da função de densidade da classe-condicional $p(\boldsymbol{x}|\omega_j)$ usando o teorema de Bayes como

$$p(\boldsymbol{x}|\omega_j) = \frac{p(\omega_j|\boldsymbol{x})p(\omega_j)}{p(\boldsymbol{x})} \tag{3.7}$$

e assim a regra de decisão pode ser escrita como: atribui-se $\boldsymbol{x}$ a ω_j se

$$p(\boldsymbol{x}|\omega_j)p(\omega_j) > p(\boldsymbol{x}|\omega_k)p(\omega_k)\ para\ k = 1, ..., C; k \neq j \tag{3.8}$$

sendo conhecida como regra de Bayes, que possui uma interessante propriedade de minimização de erro. Ao empregar essa regra, o classificador de Bayes minimiza o erro da função de classificação, assim como também minimiza o custo de uma classificação errônea (WEBB, 2002).

Em sua aplicação para imagens, costumeiramente os dados a serem classificados são as feições (cor, textura, gradiente, etc.) dos *pixels* presentes na imagem e os resultados são as possíveis classes de cada objeto. Para tal, é necessário um conjunto de exemplos de treinamento em que os vetores de características de cada *pixel* são utilizados para estimar as f.d.p.s $p(x)$ e $p(\omega_j)$.

Diversas formas de modelagem estatística podem ser utilizadas, tanto paramétricas (i.e., parametrização de funções conhecidas que descrevem a densidade de probabilidade dos dados), quanto não-paramétricas (estimação de histogramas).

Ao construir um modelo paramétrico que estime a função de densidade de probabilidade dos dados que se deseja classificar, buscamos obter o vetor de parâmetros que definem tal função. Um dos modelos de função mais comumente utilizados é de uma função gaussiana (também chamada *normal*). Esse é um modelo estatístico amplamente conhecido e tem sido extensivamente estudado ao longo dos anos, não somente por suas aplicações em técnicas computacionais, mas também na estatística em geral.

Um modelo gaussiano é uma f.d.p. descrita por uma função gaussiana, que em sua versão multivariada, possui a seguinte forma:

$$\mathcal{G}(x;\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}\left|\Sigma\right|^{1/2}} exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{3.9}$$

onde μ é o vetor de médias, que indica o centroide de densidade dos dados, Σ é a matriz de covariância, que indica a distribuição dos dados em torno da média, d é a dimensão dos dados analisados e $P(x)$ é a probabilidade de ocorrência do vetor de dados x.

Ao empregar os conceitos da regra de Bayes, juntamente com modelos de funções de densidade paramétricas, podemos definir técnicas para identificar os agrupamentos presentes em conjuntos de dados multimodais (SZELISKI, 2010). Nas próximas seções veremos algumas técnicas para agrupamento (*clustering*) de dados.

3.2.2 Modelagem pelas Médias: Algoritmo *K-Means*

O K-Means é um algoritmo para separar N objetos (baseados em atributos) em K aglomerados (*clusters*) ou partições, onde $K < N$. Nesse algoritmo, trabalha-se com a suposição de que a densidade dos dados pode ser determinado pela sobreposição de um pequeno número de distribuições mais simples (como gaussianas, por exemplo).

Assumindo que os atributos formam um vetor no espaço e que todas as distribuições são esfericamente simétricas (i.e. suas matrizes de covariância são o produto de um escalar por uma matriz identidade) (BISHOP, 2006), o objetivo é minimizar a variância intra-cluster, ou seja, minimizar o erro quadrático dado por

$$V = \sum_{i=1}^{K} \sum_{x_j \in S_i} (x_j - \mu_i)^2 \tag{3.10}$$

onde existem K agrupamentos $S_i, i = 1, 2, ..., K$, e μ_i é o centróide ou ponto médio de todos os pontos $x_j \in S_i$.

O algoritmo de agrupamento K-Means foi criado em 1956 (BISHOP, 2006) e sua forma mais comum utiliza uma heurística de refinamento iterativa conhecida como algoritmo de Lloyd (LLOYD, 1982). Esse algoritmo inicia particionando os pontos de entrada em K conjuntos iniciais aleatórios. A seguir calcula-se o ponto médio (ou centroide) para cada conjunto. É construída então uma partição associando cada ponto com o centroide mais próximo. Os centroides são recalculados para os novos agrupamentos e o algoritmo reinicia até que haja convergência, ou seja, os pontos não mais troquem de classe (ou os centroides não se alterem).

O algoritmo de Lloyd e o K-Means são frequentemente referenciados como sinônimos. Entretanto, o algoritmo de Lloyd é apenas uma heurística para a solução do problema do K-Means. E, uma vez que esse processo é dependente dos pontos e centroides iniciais, o algoritmo de Lloyd pode não convergir para a resposta correta. Apesar disso, essa técnica permanece popular graças à alta velocidade que costuma convergir.

3.2.3 Modelagem pela Mistura de Gaussianas

Devido a um modelo de distribuições (esféricas simétricas) assumido pelo algoritmo K-Means, ele reduz o problema de encontrar os agrupamentos à minimização da distância euclidiana entre cada vetor de dados e o centroide (média) da classe correspondente (BISHOP, 2006).

Já no modelo de mistura de gaussianas (*Gaussian Mixture Model*, GMM), cada centroide é definido não apenas por um vetor de médias μ_k, mas também por uma matriz de covariância Σ_k, cujos valores são re-estimados a partir das amostras correspondentes (SZELISKI, 2010) e indicam a distribuição de cada agrupamento como uma gaussiana multivariada. Como método de estimação de densidade, os modelos de mistura são mais flexíveis do que os modelos baseado em normais simples, fornecendo uma melhor discriminância em alguns casos (WEBB, 2002).

Oriundo do modelo de misturas finitas, o GMM tem sido explorado cada vez mais, tendo sua aplicação em modelos de distribuição onde as medições surgem para separação de grupos de dados cujos membros individuais são desconhecidos. As aplicações de modelos de mistura incluem detecção de falhas têxteis, classificação de forma de onda e classificação da ROI (WEBB, 2002).

Um modelo de mistura finita (YANG; AHUJA, 1999) é uma distribuição da forma

$$p(\boldsymbol{x}) = \sum_{j=1}^{g} \gamma_j p(\boldsymbol{x}; \Theta_j) \tag{3.11}$$

onde g é o número de componentes de mistura; $\gamma_j \geq 0$ são as proporções de mistura ($\sum_{j=1}^{g} \gamma_j = 1$); e $p(\boldsymbol{x}; \Theta_j)$, $j = 1, ..., g$, são as funções de densidade de probabilidade que dependem do vetor de parâmetros Θ_j. Existem três conjuntos de parâmetros a serem estimados: os valores de γ_j, os componentes Θ_j e o valor de g. Os componentes de densidade de probabilidade podem ser de diferentes formas paramétricas e são especificados usando um conjunto de dados conhecidos, se disponível. No modelo de mistura de gaussianas $p(\boldsymbol{x}; \Theta_j) = \mathcal{G}(x; \Theta_j)$ é uma distribuição normal multivariada, com $\Theta_j = \{\mu_j, \Sigma_j\}$ (WEBB, 2002).

Dado um conjunto de n observações $(x_1, ..., x_n)$, o objetivo do GMM pode ser definido como maximizar a função de verossimilhança (WEBB, 2002) é dada por

$$L(\varphi) = \prod_{i=1}^{n} \sum_{j=1}^{g} \gamma_j p(\boldsymbol{x}; \Theta_j) \tag{3.12}$$

onde φ denota o conjunto de parâmetros $\{\gamma_1, ..., \gamma_g; \Theta_j, ..., \Theta_g\}$ e $p(x|\Theta_j)$, a dependência dos componentes de densidade sobre os seus parâmetros como.

Em geral, podemos solucionar o problema de encontrar os parâmetros do modelo de mistura finita empregando um método iterativo. Uma abordagem para maximizar a verossimilhança $L(\varphi)$ é usar uma classe geral de processo iterativo conhecido como algoritmo EM (*expectation maximisation*) (WEBB, 2002). Assim como ocorria no algoritmo K-Means, o EM também possui dois estágios que são repetidos sequencialmente até que a modelagem probabilistica dos dados convirja, isto é, nenhuma amostra troque de classe.

1. No estágio de *expectativa* (passo E), são estimadas as responsabilidades:

$$z_{ik} = \frac{1}{Z_i} \gamma_k \mathcal{G}(x|\mu_k, \Sigma_k) \qquad \text{com} \qquad \sum_k z_{ik} = 1 \tag{3.13}$$

que são as estimativas de quão bem cada amostra x_i foi gerada pelo k-ésimo agrupamento gaussiano, ou seja, a determinação de quais amostras são associadas a cada centroide.

2. No estágio de *maximização* (passo M) os parâmetros μ_k, Σ_k e γ_k de cada gaussiana são re-estimados a partir das amostras:

$$\mu_k = \frac{1}{N_k} \sum_i z_{ik} x_i, \tag{3.14}$$

$$\Sigma_k = \frac{1}{N_k} \sum_i z_{ik} (x_i - \mu_k)(x_i - \mu_k)^T, \tag{3.15}$$

$$\gamma_k = \frac{N_k}{N} \tag{3.16}$$

onde

$$N_k = \sum_i z_{ik}. \tag{3.17}$$

é a estimativa de quantos pontos da amostra foram associados para cada gaussiana (agrupamento).

As estimativas iniciais podem ser obtidas através de uma sub-amostragem aleatória dos dados analisados, assim como no K-Means. Dessa forma, a principal dificuldade do método GMM é como determinar o número de componentes g, sendo que essa é a única informação sobre o conjunto de dados que o usuário deve fornecer. É possível que diversos testes tenham que ser realizados para que esse valor seja determinado de maneira satisfatória (WEBB, 2002).

Outro problema é que podem haver muitos mínimos locais para a função de verossimilhança e pode ser necessário que diversas configurações iniciais sejam testadas até que o modelo se ajuste corretamente aos dados. De qualquer forma, vale a pena tentar várias inicializações, uma vez que vários resultados semelhantes indicam uma melhor representatividade da solução escolhida.

4 EXTRAÇÃO DE FEIÇÕES E REPRESENTAÇÃO DE IMAGENS

Atualmente, existem diversas abordagens para representar as informações presentes em diferentes regiões de uma imagem, usando feições como intensidade, tom de cor, textura, etc. Ao observar uma dada imagem, seres humanos conseguem distinguir facilmente dentre diferentes objetos e materiais nela presentes, pelo uso de características como tom de cor e intensidade luminosa. Entretanto, essas informações podem não fornecer a riqueza de detalhes necessária para que métodos computacionais automatizados possam realizar com sucesso a mesma tarefa.

O critério mais adequado para mensurar as características da imagem dependem fortemente do contexto em que ela se encontra (por exemplo, exames médicos, imagens de satélite, imagens naturais, etc.) e do nível de detalhes que se deseja obter na representação da imagem (CHEN et al., 1999).

Este capítulo tem como objetivo detalhar o embasamento teórico necessário para fundamentar a extração de feições e representação de imagens proposta neste trabalho. Para isso, primeiramente serão vistos os conceitos básicos de representação de cor em imagens. Em seguida, serão revisados os conceitos acerca das técnicas de decomposição e representação multi-escala de imagens. Finalmente, também serão analisados os fundamentos relacionados às características e modelagem de texturas, tanto em aspectos locais (para cada *pixel* individualmente), quanto em aspectos globais (por região).

4.1 Representações de Cores

Ao analisar imagens de qualquer tipo, a escolha do espaço de cores pode ser considerada como o primeiro passo do processamento realizado. Como a maioria das técnicas de visão computacional, os equipamentos utilizados na aquisição de imagens geralmente utilizam sensores que buscam reproduzir o sistema de visão humano.

Na estrutura do olho humano, a caracterização de *cor* se dá pela energia luminosa distribuída ao longo de diferentes faixas de comprimento de onda (GONZALEZ; WOODS, 2006). Dessa forma, as cores são percebidas a partir de combinações da decomposição da luz em três fatores (agrupamentos dos comprimentos de onda), que formam as cores primárias: vermelho, verde e azul (do inglês, *red*, *green* e *blue*, RGB).

Empregando esta estrutura de formação de cores, é definido o espaço de cores RGB, sendo esse sistema o mais comumente utilizado, tanto por equipamentos de aquisição quanto para armazenamento digital de imagens. Nele, as cores de um *pixel* são representadas por três valores, que indicam a magnitude da energia de cada uma das cores primárias, O arranjo dos valores de uma determinada cor primária para cada *pixel* da

imagem é chamado canal de cor e apresenta-se na forma de uma matriz bidimensional. Uma imagem é então formada combinando as três matrizes (canais de cor) (GONZALEZ; WOODS, 2006).

Embora seja próxima da representação utilizada pela visão humana, esta representação de cores frequentemente é ineficiente quando utilizada para análise de imagens. Isso ocorre principalmente devido ao fato de esse sistema não representar a luminosidade da cena eficientemente, ao contrário, o fator do iluminante encontra-se distribuído entre os três canais (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007).

Existem ainda diversos outros espaços de cores, que descrevem as cores sob aspectos perceptuais (HSV), ortogonais (YCbCr) ou perceptualmente uniformes (CIE L*a*b*). A partir da descrição em RGB é possível obter qualquer outro espaço de cores, aplicando transformações lineares ou não-lineares.

O espaço de cores CIE L*a*b* (também chamado CIELAB), por exemplo, representa as cores a partir de uniformidade perceptual, ou seja, a partir da forma como a aparência das cores difere para um observador humano (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007). Pelo uso de uma extensiva série de transformações não lineares, esse sistema tenta isolar o fator de iluminação, representado no canal de luminância (L), dos componentes cromáticos, representados nos canais *a* e *b*. Ao realizar essa separação dos aspectos cromáticos (cores) e acromáticos (luminosidade), esse sistema de cores permite determinar os padrões da imagem separadamente para cada dos componentes, o que o torna uma representação adequada para métodos de reconhecimento de padrões e segmentação de imagens de uma maneira geral.

Outra abordagem proposta para dissociar a informação de luminância da informação cromática presente nos canais RGB consiste em normalizar os componentes desse espaço de cores. No espaço de cores RGB normalizado (*rgb*), a normalização busca fazer com que os componentes normalizados tenham soma unitária. Para isso, são utilizadas as seguintes transformações:

$$r = \frac{R}{R+G+B}, \qquad g = \frac{G}{R+G+B}, \qquad b = \frac{B}{R+G+B}, \tag{4.1}$$

onde R, G, e B são os valores dos componentes RGB, e r, g, e b são suas respectivas normalizações. Diferentemente do espaço de cores CIELAB, no *rgb* a informação de luminância é dissociada dos componentes de cromaticidades, eliminando esse fator da representação. Além disso, como os componentes desse sistema possuem somatório unitário, o terceiro componente não agrega qualquer informação relevante e costuma ser descartado para obter uma redução de dimensionalidade dos dados (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007).

Particularmente no que diz respeito à detecção de pele a partir das características do tom da pele, alguns trabalhos observaram que a diferença entre *pixels* com cor de pele advinda das condições de iluminação da cena e etnia são grandemente reduzidas neste espaço de cores (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007). Além disso, as aglomerações de cor de pele no espaço *rgb* (normalizado) possuem variância relativamente baixa em comparação com os seus equivalentes em RGB (não-normalizado). E por esse motivo costumam ser consideradas adequadas para modelagem e detecção da cor de pele (YANG; LU; WAIBEL, 1997; YANG; AHUJA, 1999). Todos esses fatores tornam o *rgb* uma escolha adequada para realizar a detecção de pele (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007).

4.2 Decomposição em Múltiplos Níveis de Detalhes

As representações de imagens da seção anterior denotam o aspecto luminoso de cada *pixel*, que são necessários nos processos de aquisição e exibição das imagens. Mas é possível ver uma imagem como um sinal bidimensional e, como tal, podemos analisar o conteúdo deste sinal em diferentes resoluções, podendo-se observar os padrões da imagem em diferentes escalas (níveis de detalhamento) (SZELISKI, 2010).

O método mais comumente utilizado para representar imagens desta forma é o chamado decomposição piramidal laplaciana (SZELISKI, 2010). Para construir a pirâmide, a imagem é primeiramente borrada com uma filtragem passa-baixas gaussiana e sub-amostrada por um fator de dois, resultando no próximo nível da pirâmide.

A pirâmide é então computada (para cada nível) reconstruindo o seu passa-baixas por interpolação do nível inferior, e subtraindo o mesmo da imagem original. Essa diferença é na realidade uma aproximação do passa banda laplaciano (BURT; ADELSON, 1983). A pirâmide resultante permite uma reconstrução perfeita, isto é, as imagens dos laplacianos de gaussianas são suficientes para reconstruir a imagem original. Também é possível criar uma pirâmide, tomando diretamente as diferenças entre dois passa-baixas de tamanhos diferentes. Nesse caso, evitamos uma série de operações de sub-amostragem e interpolação, mas a reconstrução da imagem resultará em perda de detalhes. Isso acontece porque a decomposição por diferenças de gaussianas (DoG) computa aproximações dos laplacianos de gaussianas (LoG), ao invés dos valores exatos (BURT; ADELSON, 1983; SZELISKI, 2010).

4.2.1 Decomposição *wavelet*

Uma alternativa as pirâmides que tem sido amplamente explorada é a decomposição da imagem através da transformada *wavelet*. O conjunto de filtros *wavelet* alcança uma localização adequada do sinal (imagem) tanto no domínio espacial (*pixels*) quanto no domínio frequência (espectros de frequência do sinal original) e são definidos por meio de escalas hierárquicas. Dessa forma, podemos decompor um sinal em componentes de frequência, sem particionar a imagem em blocos, e que também utilizam uma estrutura piramidal (SZELISKI, 2010).

Enquanto as pirâmides tradicionais são sobrecompletas, isto é, utilizam mais *pixels* do que a imagem original para representar a decomposição, as *wavelet*s são capazes de realizar a decomposição mantendo o número *pixels* (coeficientes da transformada), como mostrado na Figura 4.1. Ainda assim, algumas famílias de *wavelets* escolhem propositalmente apresentar um número de coeficientes mais elevado do que a imagem original, o que pode se dar tanto pela resolução de cada nível (facilidade de localizar os coeficientes relativos a cada *pixel*) quanto pela quantidade de bases (direções dos filtros) em que a imagem é decomposta (MALLAT; ZHONG, 1992; SCHARCANSKI, 2007). Nesse sentido, podemos ainda dizer que *wavelets* são mais seletivas quanto à orientação do que outros tipos de representações considerando filtros passa-bandas regulares (SZELISKI, 2010).

No caso geral, podemos obter a decomposição *wavelet* utilizando um filtro passa-baixas ϕ e um conjunto de filtros passa-banda ψ_s^θ, que são as chamadas "*wavelets*", sendo geradas a partir da variação de escala s e orientação θ de uma única função *wavelet* mãe $\psi(t)$ (SONKA; HLAVAC; BOYLE, 2007). Enquanto o filtro passa-baixas ϕ é responsável por fornecer uma aproximação da imagem, os passa-bandas ψ_s^θ são responsáveis por detectar os detalhes em diferentes níveis (escalas) e orientações. Determinar quais serão os filtros utilizados é uma questão que vem sendo estudada há mais de duas décadas e

(a) (b)

Figura 4.1: Exemplo de decomposição de imagens pela transformada *wavelet*. 4.1a imagem original; 4.1b resultado da transformada considerando 3 níveis de detalhes.

a resposta depende fortemente da aplicação a ser desenvolvida (SZELISKI, 2010). Para obter esse conjunto de filtros, podemos combinar sucessivas aplicações de ϕ, intercaladas com operações de sub-amostragem. Dessa forma, podemos obter os detalhes em diferentes escalas aplicando $\psi^\theta = \psi_0^\theta$ às diferentes escalas de aproximação. Finalmente, para que o resultado da transformada possua exatamente o mesmo número de coeficientes da representação espacial, utiliza-se uma sub-amostragem de fator 2^s (que produz uma pirâmide octagonal) e três filtros passa-altas ψ^h (horizontal), ψ^v (vertical) e ψ^d (diagonal obtido combinando os dois primeiros). Nesse caso, ψ^h e ψ^v são operadores de gradiente orientado (MALLAT; ZHONG, 1992; DO; VETTERLI, 2002).

Ao aplicar todos esses filtros a uma imagem obtemos quatro novas imagens: uma contendo a aproximação e três contendo os detalhes nas direções vertical, horizontal e diagonal. Para realizar uma análise multiescala, a aproximação obtida nesse nível é subamostrada na metade de seu tamanho original, e a decomposição descrita anteriormente é repetida. Um exemplo dessa decomposição, segundo a metodologia proposta por (MALLAT, 1989), é ilustrado na Figura 4.1. Nessa imagem, cada nível da decomposição está representado por uma aproximação (apresentada no quadrante superior esquerdo de cada nível), e gradientes nas direções horizontal, vertical e diagonal (respectivamente nos quadrantes superior direito e inferiores esquerdo e direito). Note que, como cada nível é extraído da aproximação da nível anterior, os coeficientes de cada filtro passa-baixas são substituídos pelo nível seguinte da decomposição.

Uma abordagem alternativa é proposta por Mallat e Zhong (1992). Em seu trabalho é proposta uma forma de realizar a decomposição *wavelet* que, em cada nível, ao invés de subamostrar a imagem, os filtros ϕ e ψ^θ são dilatados como segue:

$$\phi_s(x, y) = \frac{1}{s^2}\phi(\frac{x}{s}, \frac{y}{s}), \tag{4.2}$$

$$\psi^{\theta}{}_{s}(x,y) = \frac{1}{s^2}\psi^{\theta}(\frac{x}{s},\frac{y}{s}) \tag{4.3}$$

onde x e y são, respectivamente, as posições horizontais e verticais na mascara de convolução do filtro, s é a escala atual, ϕ é o filtro passa baixas na primeira escala e ϕ_s é dilatação desse filtro na escala s, ψ^{θ} é o filtro de detecção de bordas na direção θ e $\psi^{\theta}{}_{s}$ é a dilatação deste filtro na estala s. Dessa maneira podemos analisar as texturas presentes na imagem, em diferentes fatores de escala, de forma consistente e sem a necessidade de subamostragem dos dados.

Novamente, embora haja mais coeficientes *wavelet* do que *pixels* na imagem de entrada, as bandas de frequência adicionais ou maior quantidade de orientações fazem com que esse tipo representação mais detalhada seja preferível em tarefa como análise de texturas (DO; VETTERLI, 2002; SCHARCANSKI, 2007; VERDOOLAEGE; SCHEUNDERS, 2011) ou remoção de ruído (PORTILLA et al., 2003).

4.3 Feições de Textura

Particularmente em imagens que apresentam regiões altamente texturadas, isto é, com grande variação de intensidade e cor, utilizar diretamente esses valores para analisar ou segmentar a imagem tende a ser ineficaz. Por esse motivo, ao invés de considerar apenas características pontuais de luminância ou cor, uma descrição mais abrangente pode ser utilizada para representar regiões da imagem através do uso de feições de textura (MEDEIROS; SCHARCANSKI; WONG, 2013a), buscando extrair informações referentes à superfície ou estrutura de objetos (SONKA; HLAVAC; BOYLE, 2007). Essa representação permite uma maior discriminação das características da imagem quando utilizada para avaliação de similaridade das regiões (CHEN et al., 1999). Ainda que essa noção seja intuitiva e óbvia, devido à sua ampla variabilidade, não existe uma definição precisa de *textura* que seja universalmente aceita na literatura, seja em processamento de imagens ou visão computacional. Mesmo assim, grande parte dos trabalhos na área compartilham, ainda que subjetivamente, do mesmo conceito geral. A ideia principal desse conceito caracteriza textura como um padrão local, que pode ser identificado visualmente e se repete sobre alguma área da imagem. Nesse contexto, considera-se que os elementos que definem esse padrão visual são chamados *primitivas da textura* ou *textels* (CHEN et al., 1999; FAUGERAS; PRATT, 1980; FRANCOS; MEIRI; PORAT, 1993; CROSS; JAIN, 1983; SONKA; HLAVAC; BOYLE, 2007) e podem ocorrer tanto de forma periódica como estocástica.

Ao longo dos anos muitas abordagens foram propostas para obter a representação das texturas, tais como gradientes (FOWLKES; MARTIN; MALIK, 2003; MARTIN; FOWLKES; MALIK, 2004), filtros de Gabor (HOANG; GEUSEBROEK; SMEULDERS, 2005), decomposição pela transformada *wavelet* (JAIN; FARROKHNIA, 1990), representações multi-escala (BONET; VIOLA, 1998), janelas de imagens (*image patches*) (VARMA; ZISSERMAN, 2009; LIU; FIEGUTH, 2012; MEDEIROS; SCHARCANSKI; WONG, 2012, 2013a), para citar alguns.

4.3.1 Matrizes de Coocorrência

Matrizes de coocorrência (GLCM) é um método de descrição de textura baseado na contagem das ocorrências de alguma configuração de tons de cinza. Nessa representação, supoõe-se que essa configuração varia rapidamente em regiões de textura fina (por exemplo, grama) e mais lentamente nas regiões homogêneas ou de textura grosseira (por

exemplo, céu limpo claro) (SONKA; HLAVAC; BOYLE, 2007).

A GLCM representa as texturas a partir da probabilidade conjunta de ocorrência simultânea de dois tons de cinza nos *pixels* das posições (x, y) e $(x + \Delta_x, y + \Delta_y)$ estimada a partir de um histograma bivariado. A partir dos valores observados neste histograma são calculadas algumas características da textura, tais como (WANG; ZHANG; YAO, 2011; SONKA; HLAVAC; BOYLE, 2007):

- contraste (medida de variação local),
- energia (segundo momento angular),
- entropia (medida de caos ou ordenação dos dados),
- correlação (medida de linearidade da imagem),
- probabilidade máxima,
- momento da diferença inversa,
- homogeneidade.

Dessa forma, cada região é representada por um vetor de características de textura, obtidos por meio de análises estatísticas da sua GLCM. Para obter uma descrição robusta das texturas, é necessário utilizar diversas matizes, uma para cada combinação de Δ_x e Δ_y, que indicam as diferentes configurações entre os vizinhos analisados.

A matriz de coocorrência é um método que, ao empregar estatísticas de segunda ordem da imagem, é capaz de representar adequadamente uma ampla variedade de texturas. Dentre suas propriedades, podemos destacar como benéficas a representação da relação espacial entre os *pixels* e invariância a transformações monotônicas dos níveis de cinza (SONKA; HLAVAC; BOYLE, 2007). Entretanto, é uma técnica que exibe fortes limitações. Primeiramente, ela não considera o formato das primitivas, o que a torna pouco recomendável para texturas com primitivas de alta resolução. Além de serem sensíveis à escala e à rotação da textura, cada matriz armazena apenas a relação entre uma configuração de *pixels*, sendo necessário utilizar várias matrizes, com diferentes valores de Δ_x e Δ_y para representar adequadamente uma textura. Finalmente, para que informações de cores sejam consideradas, é necessário um número pelo menos 3 vezes maior de matrizes, o que eleva seu custo computacional tanto em tempo de computação quanto em memória (SONKA; HLAVAC; BOYLE, 2007).

4.3.2 Bancos de Filtros

Enquanto as GLCM descrevem texturas por meio de frequências espaciais, também é possível utilizar as frequências de bordas para essa tarefa (SONKA; HLAVAC; BOYLE, 2007). Algumas abordagens para representação de texturas identificam os padrões texturais usando diferenças locais, como gradientes ou derivadas de gaussianas. Essas características detectam mudanças nos fatores de brilho, luminância ou cor. Então ao computar essas diferenças em diversas direções e escalas, é possível obter feições de textura robustas (VARMA; ZISSERMAN, 2009).

Ao observar o comportamento das texturas sob esta ótica de variação de local de intensidades, é possível determinar características de textura por meio de representações multiescala de imagens, tais como decomposições piramidais (SZELISKI, 2010), diferenças

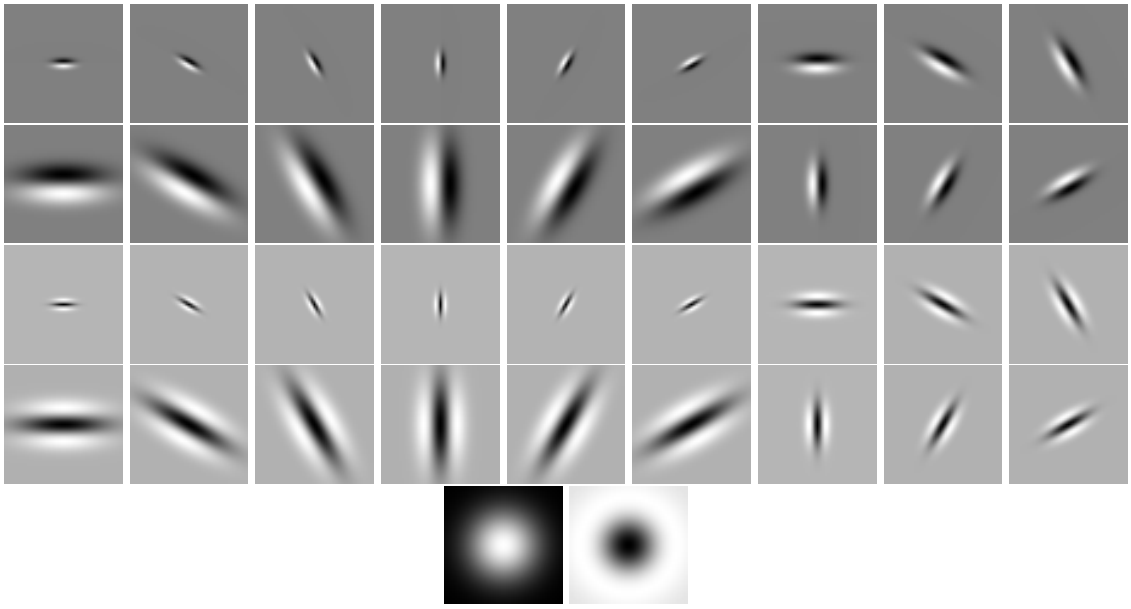


Figura 4.2: Exemplo de um banco de filtros utilizado para extração de feições de textura (VARMA; ZISSERMAN, 2005).

de gaussianas (BURT; ADELSON, 1983) ou transformadas *wavelet* (SCHARCANSKI, 2007; VERDOOLAEGE; SCHEUNDERS, 2011), descritas na Seção 4.2.

Em um exemplo desse tipo de aplicação, Scharcanski (2007) propôs o uso de transformadas *wavelets* diádicas para análise de texturas em materiais industriais. Esse trabalho mostra que certos tipos de textura podem ser identificados a partir da estimação de densidade de probabilidade dos coeficientes *wavelet*, o que pode ser obtido através da parametrização de uma f.d.p. gaussiana. Esse trabalho, assim como outros (VERDOOLAEGE; SCHEUNDERS, 2011), demonstram que é possível obter representações de texturas através do simples uso operadores de gradientes em duas direções que descrevem a relação entre *pixels* vizinhos através de diferenças locais.

Similarmente, outros métodos propõem a obtenção das feições de textura de imagem pelo uso de um conjunto de filtros de Gabor (JAIN; FARROKHNIA, 1990; JIANG; YAO; JIANG, 2007). Nesses conjuntos, estão presentes filtros de diferentes tamanhos (em geral escala 2^i, com $i = 1, 2, \cdots, N$, $N \geq 4$) e direções (geralmente em quantidade múltipla de 4). Por esse motivo, alguns autores propõem que a representação de imagens com filtros de Gabor pode ser vista como uma decomposição *wavelet* quase ortogonal (JAIN; FARROKHNIA, 1990). Além disso, é possível que os coeficientes menos significativos na identificação das texturas sejam descartados (JAIN; FARROKHNIA, 1990) para se atingir um melhor desempenho das técnicas de reconhecimento de padrões (como discutido no Capítulo 3). Isso é especialmente comum no caso dos filtros passa-baixas que nem sempre agregam informações relevantes para identificação de texturas.

A partir desta ideia Varma e Zisserman (2005) propuserem a obtenção das feições de textura a partir de um banco de filtros, chamado MR8. O MR8 é formado por um conjunto de 38 filtros (veja Figura 4.2), mas apenas 8 respostas desses filtros são utilizadas. As filtragens presentes nesse banco empregam uma gaussiana e uma laplaciana de gaussiana (LoG) ambas de mesma escala, filtros de detecção de bordas (derivada primeira) em 6 orientações e 3 escalas, e filtros de laplacianas (derivada segunda), também em 6 orientações e 3 escalas. A reposta dos filtros isotrópicos (gaussiana e LoG) são usados diretamente.

Nos filtros orientados, contudo, apenas a resposta máxima entre todas as orientações de cada escala é utilizada. Dessa forma as 8 respostas dos filtros tornam-se independentes da orientação da textura, o que assegura uma representação de textura invariante à rotação. Nesse trabalho, também é discutido o banco de filtros MR4, que introduz invariância à escala (além de rotação) selecionando apenas a resposta máxima dentre todas as orientações e escalas computadas.

Outra área de foco em representação de texturas é o uso do gradiente de brilho, cor e textura para obter as feições de texturas (FOWLKES; MARTIN; MALIK, 2003; MARTIN; FOWLKES; MALIK, 2004). No trabalho de (MARTIN; FOWLKES; MALIK, 2004), por exemplo, as feições são capturadas através do conceito de filtros de energia orientada, no qual uma combinação de filtragens gaussianas e uma transformada de Hilbert em múltiplas orientações são usadas para extrair as características de textura.

Tendo como objetivo a segmentação das regiões de diferentes texturas na imagem, um histograma de gradientes em uma janela ao redor de cada *pixel* é usado para computar a as variações locais de cor, brilho e textura que definem a aparência de cada região. Esse método apresenta resultados interessantes no processo de segmentação, embora não trate diretamente da similaridade entre texturas, mostra que é possível obter elevada discriminação entre texturas adjacentes por meio de histogramas contendo poucos *pixels*.

4.3.3 Pedaços de Imagens (*image patches*)

Enquanto as feições de texturas por meio de diferenças locais são capazes de fornecer representatividade adequada e distinção entre diferentes materiais, trabalhos recentes demonstraram que é possível atingir uma precisão ainda maior com uso de características muito mais simples e intuitivas (VARMA; ZISSERMAN, 2009; LIU; FIEGUTH, 2012).

Na representação de características de textura por meio de *wavelets* (SCHARCANSKI, 2007; VERDOOLAEGE; SCHEUNDERS, 2011), filtros de Gabor (JAIN; FARROKHNIA, 1990; JIANG; YAO; JIANG, 2007) ou outros bancos de filtros (VARMA; ZISSERMAN, 2005; MARTIN; FOWLKES; MALIK, 2004), cada *pixel* possuiu um vetor de feições de textura, que são as respostas dos filtros utilizados.

Varma e Zisserman (2009) observaram que essas respostas são obtidas pelo produto interno entre vetores de pedaços da imagem (*image patch*) com a matriz de filtragem (máscara do filtro). Logo, as respostas desses filtros são na realidade projeções dos pedaços da imagem em um sub-espaço linear de menor dimensionalidade. É possível então, a representação de texturas por pedaços da imagem. Para cada *pixel* é extraído um vetor de características de dimensão N^2, obtido tomando uma janela (vizinhança quadrada) de $N \times N$ *pixels* em torno do *pixel* analisado, redimensionada na forma de um vetor $N^2 \times 1$. No seu trabalho original, Varma e Zisserman (2009) compararam o uso direto dos pedaços de imagem com o banco de filtros MR8 (VARMA; ZISSERMAN, 2005) e demonstraram que, mesmo com pequenos blocos (3×3, 5×5 ou 7×7), essas feições de textura são ao menos tão eficientes quanto as diferenças locais dos filtros.

A vantagem principal de uma abordagem tão direta é que a extração dos pedaços de imagem é muito mais simples e, mesmo em sua forma natural, oferece uma representação fiel das texturas. Isso ocorre porque, similarmente às GLCM, os pedaços de imagens também indicam com precisão as relações entre *pixels* vizinhos, mas em uma dimensão muito mais elevada, isto é, com mais detalhes.

4.3.4 Dicionário de Textons

Originalmente projetados para uso em conjunto com um banco de filtros (LEUNG; MALIK, 2001), o uso de dicionários de textons permitem uma das representações de texturas mais eficientes e fiéis disponíveis na literatura (VARMA; ZISSERMAN, 2009; LIU; FIEGUTH, 2012; MEDEIROS; SCHARCANSKI; WONG, 2012), obtidos através da análise estatística dos vetores de características extraídos da imagem.

Tanto os bancos de filtros, quanto os pedaços de imagens descritos nas seções anteriores, tentavam incorporar as relações entre *pixels* vizinhos na forma de vetores. Logo, texturas semelhantes deveriam possuir vetores semelhantes e vice-versa. Embora pudessem fornecer uma descrição pontual adequada das texturas da imagem, Leung e Malik (2001) demonstraram que uma representação de textura muito mais robusta pode ser obtida observando as probabilidades de ocorrência desses vetores em regiões inteiras.

Uma vez que as texturas apresentam, por definição (veja Seção 4.3), propriedades de repetição espacial, espera-se que suas características sejam respostas de filtros ou pedaços da imagem, similares, embora não necessariamente iguais, ao longo da textura. Isso ocorre porque os vetores de características possuem redundâncias na descrição da imagem, mesmo que vetores maiores (com mais detalhes) sejam usados (LEUNG; MALIK, 2001).

Essa suposição de que há redundância de informações em vetores da mesma textura, permite que os vetores de características sejam agrupados em um pequeno conjunto de protótipos de feições de textura, atuando como as primitivas da textura. Esses protótipos são chamados *textons*. e podem ser obtidos utilizando algum algoritmo de agrupamento, por exemplo K-Means (LIU; FIEGUTH, 2012), em que o cada centroide será um texton. O conjunto de todos os textons utilizados na representação das texturas é chamado *dicionário de textons*.

Embora os protótipos possam ser obtidos com qualquer técnica de agrupamento, grande parte dos trabalhos utilizam o próprio K-Means (LEUNG; MALIK, 2001; VARMA; ZISSERMAN, 2009; LIU; FIEGUTH; KUANG, 2011a; MEDEIROS; SCHARCANSKI; WONG, 2012) (ou alguma variação dele como *fuzzy c-means* (SONG et al., 2007)), porque seu único parâmetro é a quantidade de agrupamentos desejados. Nesse caso, isso é equivalente a determinar o tamanho (número de entradas) do dicionário, o que é uma propriedade interessante do ponto de vista de análise de texturas.

Uma vez que o dicionário seja obtido, podemos estimar a distribuição dos vetores de características de textura utilizando um histograma, em que cada *bin* contabiliza a probabilidade de ocorrência de um dos protótipos do dicionário. Para isso, é necessário realizar a quantização vetorial das feições de cada *pixel* em um dos textons. Esse processo geralmente é realizado utilizando a menor distância euclidiana entre as feições e os textons (VARMA; ZISSERMAN, 2009). Dessa forma, cada textura passa a ser representada por um histograma. Logo uma análise de similaridade entre texturas com essa representação envolve a comparação de seus histogramas. Diversos trabalhos da literatura sugerem que a similaridade seja calculada com a medida de sobreposição de histogramas χ^2 (VARMA; ZISSERMAN, 2009; LIU; FIEGUTH, 2012; AHERNE; THACKER; ROCKETT, 1998):

$$\chi^2(\widehat{p},\widehat{q}) = \frac{1}{2}\sum_i \frac{(\widehat{p}_i - \widehat{q}_i)^2}{\widehat{p}_i + \widehat{q}_i} \tag{4.4}$$

onde $\widehat{p}$ e $\widehat{q}$ são dois histogramas de qualquer distribuição.

É interessante notar que, como essa representação de texturas considera a distribui-

ção de probabilidade das feições, ao invés das feições em si, esse modelo atinge uma representação muito mais precisa das relações espaciais entre os *pixels*, principalmente, quando combinada com feições de janelas da imagem. De certa forma, pode-se dizer que, enquanto os bancos de filtros e janelas de imagens descrevem apenas feições "pontuais"da textura, o método de dicionário de textons cria uma descrição da *aparência* da textura (LIU; FIEGUTH, 2012).

Embora poderoso, a principal desvantagem deste descritor de texturas é que, por estimar a frequência com que cada protótipo ocorre, apenas regiões com um certo número de *pixels* serão corretamente representadas (MEDEIROS; SCHARCANSKI; WONG, 2012). O histograma de uma região com poucos *pixels* possuirá uma descrição imprecisa e instável, insto é, que varia muito facilmente com a translação sobre uma mesma textura, o que pode ocasionar baixa similaridade entre duas amostras da mesma textura. Já regiões que possuem apenas um *pixel* não podem ser representadas dessa maneira, pois seu histograma seria comparável apenas a outra região exatamente igual, ou seja, outro vetor que tenha sido quantizado para o mesmo texton.

Por outro lado, essa propriedade também indica que conforme são adicionadas amostras (*pixels*) a uma textura, seu descritor de aparência se torna mais robusto, de forma que adquira estabilidade (mantenha-se igual para diferentes amostras da textura) se for utilizada uma área (numero de *pixels*) grande o suficiente. Essa característica torna essa representação adequada para técnicas de segmentação de imagens (MEDEIROS; SCHARCANSKI; WONG, 2013a).

Além disso, uma desvantagem adicional deste descritor é que a comparação entre texturas diferentes deve utilizar o mesmo dicionário. Entretanto, para que uma textura seja representada adequadamente, é necessário que um número mínimo de textons relacionados a ela esteja presente no dicionário. Isso pode ser obtido através da concatenação de diversos dicionários, todos de mesmo tamanho, cada um oriundo de uma textura diferente, sendo a metodologia usual para reconhecimento de texturas (VARMA; ZISSERMAN, 2009; MEDEIROS; SCHARCANSKI; WONG, 2013b).

Trabalhos mais recentes, no entanto, sugerem que é possível obter um dicionário representativo extraindo quantidades diferentes de textons em cada textura, sob a suposição de que as mesmas feições serão encontradas em texturas distintas, apenas com probabilidades de ocorrência diferentes (MEDEIROS; SCHARCANSKI; WONG, 2012, 2013b). Essa suposição é intuitiva, uma vez que as texturas são representadas por histogramas, sendo muito conveniente para o emprego desse descritor em técnicas de segmentação não supervisionada de imagens.

5 METODOLOGIA PROPOSTA PARA SEGMENTAR TEXTURAS E DETECTAR REGIÕES DE PELE

Os métodos mencionados no Capítulo 2, embora não apresentem resultados ótimos, eles demonstram diversas características interessantes da detecção de pele. Particularmente, notamos que o uso de feições de textura combinados com técnicas de crescimento de regiões permitem uma segmentação ainda mais precisa.

Os trabalhos encontrados na literatura, entretanto, representam as texturas por meio de métodos como GLCM ou bancos de filtros. Como discutido no Capítulo 4, estudos mais recentes (VARMA; ZISSERMAN, 2009; MEDEIROS; SCHARCANSKI; WONG, 2013b) mostraram que uma descrição mais robusta e eficiente é obtida utilizando pedaços de imagem e dicionários de textons. Esses descritores têm sido bastante empregados nas técnicas do estado da arte em reconhecimento de texturas e materiais. Em especial, Liu e Fieguth et al. (2012) demonstraram que essa descrição de texturas pode ser aperfeiçoada pela introdução de um fator estocástico na extração das feições da imagem.

Além disso, os estudos em detecção de pele disponíveis na literatura também revelam que as regiões de pele podem ser mais precisamente delimitadas com o auxílio de técnicas de segmentação que consideram as bordas presentes na imagem (PHUNG; BOUZERDOUM; CHAI, 2003; JIANG; YAO; JIANG, 2007). Especialmente ao empregar técnicas de crescimento de regiões, a forma externa das regiões de pele tende a ser identificada mais facilmente. Os trabalhos do estado da arte em segmentação de imagens mostram que há abordagens mais eficazes para segmentação por fusão de regiões, especialmente pela introdução de relações estatísticas (NOCK; NIELSEN, 2004) ou estocásticas (WONG; SCHARCANSKI; FIEGUTH, 2011; MEDEIROS; SCHARCANSKI; WONG, 2012).

Motivados pelos ganhos obtidos com as técnicas mais recentes de reconhecimento de texturas, segmentação por fusão de regiões e principalmente pelos benefícios de combinar essas duas técnicas para detecção de pele, este capítulo descreve a proposta deste trabalho. Combinando as teorias sobre feições de imagem descritas no Capítulo 4 (cor, decomposição multi-escala, textura) com os conceitos de reconhecimento de padrões e segmentação de imagens vistos no Capítulo 3, almejamos desenvolver um método de reconhecimento de regiões de pele para segmentação de gestos de mão por meio de modelos de textura estocásticos multi-escala.

O algoritmo proposto neste trabalho pode ser resumido da seguinte maneira. No estágio inicial, utilizamos métodos estatísticos de reconhecimento de padrões para treinar os modelos de cor (GMM) e textura (histogramas de textons) de pele, obtidos a partir de um conjunto de imagens de pele (exemplos de treinamento). Em seguida utilizamos uma estratégia de segmentação de texturas baseada em fusão de regiões. Essa etapa tem como objetivo identificar e particionar todos os segmentos das diferentes texturas conti-

das na imagem. Finalmente, o último estágio utiliza os modelos de cor de textura obtidos anteriormente para identificar as regiões de pele entre os segmentos obtidos na etapa de segmentação.

5.1 Representação e Modelagem de Texturas

A abordagem de modelagem estocástica multi-escala de texturas proposta para representar a textura contida em uma determinada região pode ser resumida em quatro passos. Primeiramente, todas as cores na imagem, que tradicionalmente estão representadas no espaço RGB, são mapeadas para o espaço de cores CIE L*a*b*. Conforme discutido no Capítulo 4, essa representação de cores minimiza a correlação entre os aspectos acromáticos (luminância) e cromáticos (cores) nos canais representados (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007). Segundo, a imagem é então representada em diferentes escalas (níveis de detalhamento) por meio de uma decomposição no espaço-escala de filtros bilaterais, que separam e preservam o detalhes da imagem ao incorporar as diferenças fotométricas juntamente com as espaciais. Terceiro, feições estocásticas de textura são extraídas por meio de pequenas amostras da imagem. O fator estocástico é introduzido por meio de projeções aleatórias (PA), que reduzem a dimensão das características ao mesmo tempo que preservam a informação textural (LIU; FIEGUTH, 2012). Quarto, todos os vetores de características extraídos são empregados na construção de um dicionário de textons. Como discutido no Capítulo 4, essa representação de texturas permite que a textura em uma determinada região seja descrita na forma de um histograma de textons.

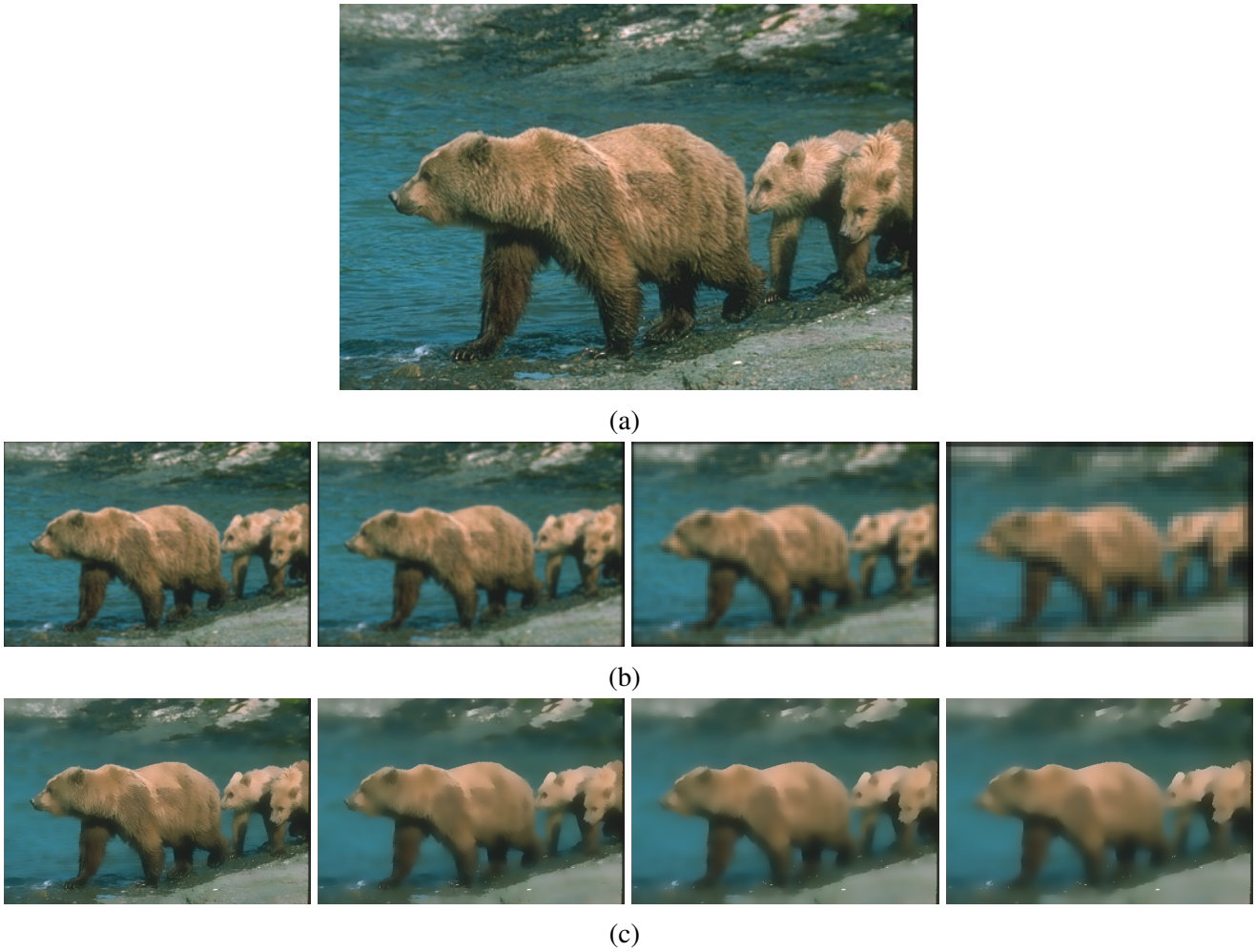
5.1.1 Representação Multi-Escala Bilateral da Imagem Textura

Na metodologia proposta, analisa-se as características de textura da imagem não apenas nos canais de luminância e cor (por exemplo, um conjunto de características de textura para cada um dos canais CIE L*a*b*), mas também em escalas diferentes, isto é, feições de textura desde escalas grosseiras (menos detalhadas) até escalas refinadas (mais detalhadas).

O uso dessas representações de imagens já foi discutido (veja Capítulo 4) e diversas abordagens existem na literatura para realizar essa tarefa por meio de aplicação de filtragem de detecção ou remoção de detalhes em diversas escalas ou resoluções. Todavia, as propostas atualmente existentes necessitam de cálculos de gradientes em diversas orientações para se obter os detalhes relevantes (que são mantidos em todas as escalas), gerando redundância de informação (JAIN; FARROKHNIA, 1990; VARMA; ZISSERMAN, 2005; VERDOOLAEGE; SCHEUNDERS, 2011). Além disso, são ainda usados filtros isotrópicos que executam o processo inverso, eliminando os detalhes em todas as direções.

Portanto, para permitir a análise e a representação de texturas em diferentes escalas e aprimorar o poder de discriminação das feições de textura extraídas, desejamos que, em cada escala, sejam eliminados os detalhes menos significativos, enquanto os mais relevantes sejam mantidos. Para isso, a imagem é decomposta em múltiplas escalas, baseado no conceito de espaço-escala de filtros bilaterais (WONG et al., 2009), em que a imagem é representada utilizando diversos níveis que indicam a escala em que observamos os detalhes da imagem.

Esse espaço foi projetado para decompor os detalhes da imagem baseado não somente na localização espacial, mas também nas diferenças fotométricas. Isso resulta em uma de-



(a)

(b)

(c)

Figura 5.1: Comparação entre as aproximações da pirâmide gaussiana e a decomposição bilateral proposta. Imagem original (5.1a) representada em 4 níveis de detalhes por uma pirâmide gaussiana (5.1b) e pela decomposição bilateral proposta (5.1c). Ambas as decomposições estão representando a imagem em 4 níveis de detalhes (da esquerda para direita, do mais preciso ao mais grosseiro).

composição multi-escala não-linear, no qual os detalhes da imagem em cada escala não somente encontram-se bem separados, mas também são bem localizados e bem preservados.

Seja c a notação de um canal no espaço de cores $\{L, a, b\}$. Para imagem em um dado canal $f_c(j)$, onde j denota a posição de um *pixel*, a representação multi-escala dessa imagem ${f'}_{c,i}$ construída usando o espaço-escala de filtros bilaterais pode ser definida por uma família de derivadas de imagem ${F'}_{c,i}$:

$$ {f'}_{c,i}(j) = \frac{\sum\limits_{\mathcal{N}} w_p\big(j, \mathcal{N}_q\big) w_s\big(j, \mathcal{N}_q\big) {f'}_{c,i-1}\,(s)}{\sum\limits_{\mathcal{N}} w_p\big(j, \mathcal{N}_q\big) w_s\big(j, \mathcal{N}_q\big)} \tag{5.1} $$

onde ${f'}_{c,0} = f_c$, i indica a escala, $\mathcal{N}_q$ define a localização do *pixel* em uma vizinhança local $\mathcal{N}$, e w_p e w_s denotam os pesos fotométricos e espaciais da gaussiana em j, respectivamente,

$$w_p(j,\mathcal{N}) = \exp\left[-\frac{1}{2}\left(\frac{\left\|f'_{c,i-1}(\underline{x}) - f'_{c,i-1}(\mathcal{N})\right\|}{\sigma_p}\right)^2\right] \tag{5.2}$$

$$w_s(j,\mathcal{N}) = \exp\left[-\frac{1}{2}\left(\frac{\|j-\mathcal{N}\|}{\sigma_s}\right)^2\right]. \tag{5.3}$$

A Figura 5.1 mostra uma comparação entre a técnica proposta de decomposição bilateral de imagens e as aproximações de uma pirâmide gaussiana, ambas em 4 níveis. Visivelmente, a representação proposta é capaz de manter as bordas mais relevantes em diferentes níveis, enquanto a pirâmide gaussiana descarta essas informações.

A partir da representação multi-escala nós desejamos construir modelos e descritores de textura em cada escala, assim cada escala da decomposição fornecerá uma descrição única da textura.

5.1.2 Representação em Pedaços Estocásticas de Imagens de Textura

Como discutido no Capítulo 4, algumas abordagens representam os padrões das texturas usando diferenças locais, tais como gradientes ou derivadas de gaussianas. Essas feições detectam as variações de brilho ou cor que, ao serem computadas em diversas direções e escalas, resultam em feições de textura com robustez elevada (MEDEIROS; SCHARCANSKI; WONG, 2013a).

Ao analisar as convoluções usadas para obter tais diferenças, observamos que essas feições são projeções de pequena dimensionalidade de um pedaço da imagem, isto é, uma pequena vizinhança (pedaço) em torno do *pixel*. Por consequência, ao invés de computar todas das diferenças multi-escala e multi-orientadas entre *pixels*, podemos utilizar diretamente os pedaços da imagem como características de textura. Essas feições são muito mais simples de serem extraídas e estudos recentes comprovam que mesmo em sua forma natural (os próprios valores do *pixels*), esses pedaços oferecem uma representação fiel das texturas (VARMA; ZISSERMAN, 2009).

Embora simplesmente tomar os pedaços de imagem em torno de cada *pixel* como um vetor de características forneça uma representação confiável de textura, os trabalhos do estado da arte mostram que essa representação ainda pode ser melhorada visando propósitos de eficiência e robustez (LIU; FIEGUTH; KUANG, 2011b,a). Neste trabalho, em particular nós empregamos o conceito de projeção aleatória para de capturar as informações mais relevantes em cada pedaço de textura (LIU; FIEGUTH, 2012), podendo ser descrito como segue.

De acordo com os estudos de Liu e Fieguth (2012), os pedaços de imagem podem ser vistas como sinais esparsos, para o propósito de redução de dimensionalidade. Logo, neste trabalho propomos o uso de uma projeção (ou transformação) linear dos vetores de características de textura. Ao aplicar essa projeção em um pedaço de textura $x \in \mathbb{R}^n$ estamos, na realidade, representando o vetor $y \in \mathbb{R}^m$ em um espaço de menor dimensionalidade. Portanto

$$y = \Phi x \tag{5.4}$$

onde $\Phi \in \mathbb{R}^{m\times n}$, $m < n$, é a matriz de projeção. Idealmente, essa matriz de projeção Φ deve garantir que a *informação* contida no sinal original seja preservada na nova representação. Isso significa que a distância entre dois vetores quaisquer desse espaço

deve permanecer aproximadamente a mesma tanto antes quanto depois da projeção (LIU; FIEGUTH; KUANG, 2011b).

Trabalhos anteriores demonstraram que a propriedade de preservação de informação está assegurada se a matriz de projeção $\Phi_{m\times n} = \{r_{ij}\}$ for definida de uma maneira aleatória como (ACHLIOPTAS, 2003)

$$r_{ij} = \begin{cases} 0 \text{ com probabilidade } \frac{1}{2} \\ 1 \text{ com probabilidade } \frac{1}{2}. \end{cases} \tag{5.5}$$

Como havíamos suposto que as informações originais em x continham informação relevante (esparsa), e idealmente $m \ll n$, esse vetor pode agora ser representado por y com uma dimensão muito menor usando uma transformação aleatória Φ, que possui a capacidade de preservar as informações dos dados, mantendo as distâncias entre os vetores após a projeção. Uma vez que essa projeção considera a matriz aleatória Φ, o vector resultante y será uma *representação estocástica* do sinal original.

Transportando essa teoria para o domínio do problema deste trabalho (extração das características de textura), consideramos cada pedaço da imagem como um sinal esparso. Portanto, podemos aplicar a projeção aleatória às características de textura capturadas em cada pedaço de imagem, gerando uma representação estocástica das texturas (*stochastic texture representation*, STR) (MEDEIROS; SCHARCANSKI; WONG, 2012). Outra vantagem da representação estocástica de texturas é uma melhora considerável na eficiência da análise das texturas, dado que essa abordagem resulta em uma significante redução na dimensionalidade das feições.

Infelizmente, a representação estocástica das feições de textura não é robusta à rotação. Para torná-la uma representação mais robusta, a propriedade de invariância à rotação pode ser adicionada ao método. É possível obter tal propriedade da representação de texturas estimando a direção do gradiente dominante em cada pedaço, ou ainda adicionando amostras (pedaços) rotacionadas ao construir os modelos de textura (VARMA; ZISSERMAN, 2005, 2009). Na prática, ambos os métodos ainda apresentam problemas. Estimativas da orientação dominante de cada pedaço tendem a ser pouco confiáveis para algumas regiões de textura (principalmente as altamente texturizadas) e a adição de amostras rotacionadas torna o processo de obtenção dos textons (utilizados para descrever texturas em regiões, veja Seção 4.3.4) muito mais custosa computacionalmente.

Neste trabalho, portanto, à invariância a rotação é alcançada ordenando os valores nos pedaços da imagem (rearranjados na forma de um vetor $N^2 \times 1$) antes de computar a STR, de modo que obtenhamos as feições de cada pedaço por:

$$Y = \Phi\ ordenar(x). \tag{5.6}$$

Como o resultado da operação de ordenação é independente do posicionamento original dos elementos em x, as feições projetadas em Y serão invariantes à rotação da textura. Colocando essa formulação em termos de pedaços de imagem em diferentes canais de cor e diferentes escalas, podemos definir um vetor de características de textura para cada *pixel* p em um canal c e escala i , dada por

$$v_{c,i}(p) = \Phi\ ordenar(\mathcal{N}_{c,i}(p)), \tag{5.7}$$

onde $\mathcal{N}_{c,i}(p)$ é a vizinhança do *pixel* p encontrada no canal c e na escada i e $v_{c,i}(p)$ é o vetor de feições STR extraídas dentro da vizinhança daquele *pixel* no canal c e na escala i.

5.1.3 Modelos de Aparência da Textura

Até então, definimos as feições estocásticas de textura extraídas das proximidades de cada *pixel* em um canal e escala, em particular, como um vetor STR, onde uma projeção aleatória é aplicada aos valores ordenados dos *pixels* em um pequeno pedaço da imagem. Essas características representam a informação textural na localização de um único *pixel*. No entanto, como o objetivo deste trabalho é segmentar e posteriormente analisar confiavelmente os segmentos de texturas contidos em uma imagem, também precisamos definir uma representação para a informação textural encontrada em uma região podendo conter vários *pixels*.

Portanto, ao invés de apenas feições locais baseadas em valores de intensidade ou cor, descritores de texturas mais abrangentes podem ser utilizados para alcançar uma segmentação mais precisa para representar não apenas as características locais, mas também características de aparência global (LIU; FIEGUTH, 2012).

Conforme visto no Capítulo 4, há um senso comum, na literatura de análise de texturas, de que uma textura é definida como um padrão, periódico ou estocástico, repetido sobre alguma área. Sob tal suposição, em nossa proposta para modelagem estocástica de texturas representamos a informação textural em uma região reunindo todos os vetores de feições de textura disponíveis, formando um dicionário de textons (VARMA; ZISSERMAN, 2009; LIU; FIEGUTH, 2012), no qual são obtidos protótipos dos vetores de características.

Na abordagem proposta, em um canal e escala específicos, o conjunto de todos os vetores de características são agrupados de maneira não-supervisionada para construir um dicionário de textons (LEUNG; MALIK, 2001; VARMA; ZISSERMAN, 2005), no qual cada texton é um centroide encontrado no estágio de agrupamento. Neste trabalho, empregamos o algoritmo K-Means para realizar o agrupamento dos vetores de feições (HARMER et al., 2011), pois esse método é capaz de criar um dicionário utilizando como parâmetro apenas o tamanho (número de textons) desejado.

Um dicionário com representatividade adequada é obtido incluindo protótipos de textura de todos os materiais (classes de textura) conhecidos (VARMA; ZISSERMAN, 2009; LIU; FIEGUTH, 2012). Essa abordagem para treinamento é clássica em métodos de aprendizado supervisionado de modelos e, em geral, apresenta resultados muito bons quando aplicada em contextos em que apenas materiais conhecidos são analisados.

Entretanto, podem ocorrer situações em que apenas algumas das classes de textura presentes na imagem são conhecidas, como por exemplo, no problema central tratado neste trabalho, que é a detecção de segmentos de pele em imagens naturais. Nesse cenário, apenas os exemplos de treinamento oriundos do material que buscamos encontrar formam um conjunto com representatividade amostral adequada. Outro caso ainda mais complexo e sensível a esse problema é quando não se possui qualquer informação a priori sobre quaisquer das classes envolvidas, como ocorre na segmentação automatizada de imagens. Em ambos os casos, como não há conhecimento *a priori* sobre as classes de textura que podem ocorrer na imagem, qualquer tentativa de predição das texturas desses materiais na imagem de teste iriam, limitar o método a um conjunto de texturas típicas de um ambiente específico ou produzir uma representação imprecisa e ineficaz de algumas texturas, devido à má representatividade dos textons.

Portanto, para resolver este problema, os textons provenientes de classes desconhecidas são obtidos da própria imagem de teste agrupando todos os vetores STR extraídos dela, independentemente de suas classes de textura. Nessa abordagem, estamos supondo que textons distintos irão formar agrupamentos distintos no espaço das feições, o que é a

ideia principal por trás do uso de técnicas de agrupamento. Além disso, texturas diferentes podem fornecer um número diferente de textons para o dicionário. Estudos recentes, contudo, mostraram que com o uso de um número razoável de textons é possível construir um dicionário que consegue representar todas as características texturais da imagem de forma confiável e eficiente (MEDEIROS; SCHARCANSKI; WONG, 2013a,b). Nos cenários em que há exemplos de texturas conhecidas a serem identificadas na imagens, o dicionário final é obtido adicionando aos textons obtidos desta maneira os textons determinados de materiais específicos (MEDEIROS; SCHARCANSKI; WONG, 2013b).

Uma vez que todos os textons componentes do dicionário tenham sido obtidos, um modelo de aparência de textura T para uma região arbitrária R pode ser determinado através das probabilidades de ocorrência dos textons em diferentes canais e escalas. Assim:

$$T(R) = \{H_{c,i}(R) | 1 \leq c \leq 3, 1 \leq i \leq N\}, \tag{5.8}$$

onde $H_{c,i}(R)$ é o histograma das ocorrências de textons da região R para o canal c e escala i. Em outra palavras, a textura de uma região é modelada por um conjunto de $3N$ histogramas, um para cada combinação de canal e escala.

5.1.4 Similaridade de Texturas por Modelos de Aparência

Dado o modelo proposto para aparência de regiões de textura, para poder empregá-lo em métodos tanto de segmentação de imagens quanto da identificação de materiais, é preciso determinar uma medida quantitativa para avaliar a dissimilaridade entre duas texturas. Dado que o modelo proposto é composto de um conjunto de histogramas de ocorrência de textons, a dissimilaridade entre duas texturas pode ser computada quantitativamente pela comparação dos histogramas entre as regiões.

Estimar se duas distribuições diferem ou são consistentes é um problema que surge frequentemente. Como discutido anteriormente (veja Capítulo 4, Seção 4.3.4), a medida de sobreposição de histogramas χ^2 (Equação (4.4)) tem sido largamente empregada para comparação de histogramas em técnicas de classificação de texturas. Entretanto a medida de sobreposição χ^2 possui algumas limitações. A partir de sua definição (Equação (4.4)) constata-se, por exemplo, que o teste χ^2 apresenta problemas sempre que *bins* vazios (com valor zero) são comparados ou ainda quando os histogramas apresentam distribuições muito diferentes (AHERNE; THACKER; ROCKETT, 1998). Para contornar tais dificuldades com o teste χ^2, neste trabalho, usamos a distância de Bhattacharyya para mensurar a dissimilaridade entre dois histogramas de ocorrência de textons. A distância de Bhattacharyya é robusta (não tendenciosa) e se tornou comumente utilizada para a comparação de duas f.d.p. e é expressa por (AHERNE; THACKER; ROCKETT, 1998)

$$D_B(p, q) = -ln\left(\int \sqrt{p(x)q(x)}\right) \tag{5.9}$$

onde $p(x)$ e $q(x)$ são duas funções densidade de probabilidade. Como um histograma pode ser visto como um estimador discreto de uma f.d.p. de uma distribuição desconhecida qualquer, a distância de Bhattacharyya pode ser reescrita numa versão discreta para computar a dissimilaridade entre dois histogramas $\widehat{p}$ e $\widehat{q}$. Assim,

$$D_B(x, y) = -ln\left(\sum_i \sqrt{\widehat{p}_i \widehat{q}_i}\right). \tag{5.10}$$

Como a informação da textura em uma região é representada por um conjunto de histogramas, neste trabalho, propomos que cada histograma seja normalizado (previamente à comparação pela distância de Bhattacharyya) dividindo a contagem de ocorrências em cada *bin* pelo total de amostras naquele histograma. Dessa forma, garantimos que $\sum_i \widehat{p}_i = 1$ e $0 \leq \widehat{p}_i \leq 1$, e cada *bin* passa a indicar a probabilidade de um texton do dicionário ocorrer em um dado pedaço de textura em cada canal e escala.

Finalmente, atribuindo um peso para cada canal $W_g = [w_L, w_a, w_b]$, a dissimilaridade de textura d_T entre duas regiões R_a e R_b é definida como

$$d_T(R_a, R_b) = \frac{W_g \begin{bmatrix} \sum_{i=1}^{N} D_B(H_{L,i}(R_a), H_{L,i}(R_b)) \\ \sum_{i=1}^{N} D_B(H_{a,i}(R_a), H_{a,i}(R_b)) \\ \sum_{i=1}^{N} D_B(H_{b,i}(R_a), H_{b,i}(R_b)) \end{bmatrix}}{\sum W_g}. \tag{5.11}$$

5.2 Segmentação de Regiões Texturadas via Fusão Estocástica de Texturas

Na seção anterior foi descrita a abordagem proposta para caracterizar e comparar as características texturais entre regiões (ou modelos). Uma vez que o nosso modelo de textura tenha sido definido, nesta sessão, focaremos em integrar esse modelo a uma estratégia de fusão estocástica de texturas com o propósito de segmentar imagens que potencialmente (mas não obrigatoriamente) contenham regiões altamente texturizadas (com alta variação de intensidade ou cor).

5.2.1 Formalização do Problema e Resumo do Algoritmo

Seja J uma grade discreta onde a imagem é definida com $j \in J$ uma posição nessa grade, referente à localização de um *pixel* seja também $F = \{f_j | j \in J\}$ a textura observada na posição j da imagem e $R = \{r_s | s \in S\}$ um rótulo para a área em uma região de textura, então a segmentação da imagem pode ser formulada como um problema de estimação da *máxima probabilidade a posteriori* (*maximum a posteriori probability*, MAP) (YU; CLAUSI, 2008; WONG; SCHARCANSKI; FIEGUTH, 2011; MEDEIROS; SCHARCANSKI; WONG, 2012). Portanto,

$$\hat{r} = \arg\max_{r} f\{P(r|f)\} \tag{5.12}$$

onde $P(r|f)$ é a probabilidade *a posteriori*. Pelo teorema de Bayes (veja Equação(3.7) noCapítulo 3), resolver a Equação (5.12) é equivalente a obter:

$$\hat{r} = \arg\max_{r} f\{P(f|r)P(r)\}, \tag{5.13}$$

onde $P(f|r)$ é a verossimilhança, e $P(r)$ é o rótulo do pixel *a priori* (classe de textura mais provável). Embora essa seja uma formulação simples para o problema, pode ser difícil resolvê-lo diretamente, devido aos termos que são geralmente desconhecidos $P(r)$ e $P(f|r)$.

Estudos recentes demonstraram que o problema de MAP pode ser abordado combinando campos aleatórios de Markov (*Markov random fields*, MRF) e algoritmos de fusão de regiões (YU; CLAUSI, 2008). Na realidade, a probabilidade *a priori* $P(r)$ pode ser estimada implicitamente (LI, 2009) e a verossimilhança $P(f|r)$ pode ser determinada

usando uma função de verossimilhança de fusão estocástica de texturas, como proposto neste trabalho.

Nesta técnica de fusão estocástica de texturas, todos os pares de regiões adjacentes no mapa de inicialização são unidos sucessivamente usando um critério probabilístico. Para manter a consistência do processo e torná-lo mais robusto a variações locais (como em regiões altamente texturizadas), os pares de regiões são arranjados em ordem crescente de dissimilaridade. Finalmente, a segmentação pode ser refinada por meio de iterações do processo de fusão.

5.2.2 Inicialização do Mapa de Regiões de Textura

Uma maneira eficaz de lidar com variações locais de luminância, cor e ruídos (artefatos presentes na imagem) é empregar um modelo com campos aleatórios de Markov (MRF) para determinar a probabilidade *a priori* $P(r)$ (LI, 2009). Similarmente aos descritores de textura, nessa abordagem as texturas na imagem são modeladas em termos de relações entre *pixels* vizinhos.

Ao iniciar a segmentação de texturas de uma dada imagem I, com $N \times M$ *pixels*, o método proposto não assume qualquer conhecimento prévio nem faz suposição alguma sobre as regiões de textura presentes na imagem, logo não há qualquer informação a respeito de quantas texturas distintas estão presentes na imagem. Antes de agregar *pixels* sucessivamente formando regiões maiores, assumimos uma distribuição uniforme para $P(r)$ (todas as texturas têm a mesma probabilidade de ocorrência). Como não sabemos quantas texturas estão presentes na imagem, um rótulo único de região de textura $R = \{1, \cdots, N \times M\}$ é associado para cada *pixel*.

No primeiro estágio do processo de segmentação iterativa, construímos um grafo $G = (R, E)$ representando o estado atual dos segmentos. Nesse grafo, cada vértice representa uma região R e as arestas E conectando os vértices correspondem às dissimilaridades locais entre vértices indicando regiões de textura vizinhas. No estágio inicial da segmentação, esse grafo de adjacências é configurado de maneira que cada *pixel* p represente uma única região e p esteja associado um único vértice. Em um estágio qualquer do processo de segmentação, cada vértice (região de textura) está conectado a quatro outros vértices (regiões adjacentes de textura). Por esse motivo, em todos os testes e experimentos relatados, neste trabalho, foi utilizada uma vizinhança-4.

5.2.3 Fusão Estocástica das Regiões de Textura

Realizamos a fusão de uma região R_a com uma região adjacente R_b, com uma probabilidade $\alpha(R_a, R_b)$, em que essa probabilidade de fusão substitui $P(f|r)$ na Equação (5.13) e é obtida a partir de uma função de verossimilhança. Neste trabalho, usamos uma função de verossimilhança de regiões de textura que estende o critério de fusão estocástica de regiões proposto por Wong et al. (2011). Então

$$\alpha(R_a, R_b) = \exp\left[-\frac{d_T(R_a, R_b)}{\Lambda(R_a, R_b)}\right] = \frac{1}{e^{\frac{d_T(R_a,R_b)}{\Lambda(R_a,R_b)}}} \tag{5.14}$$

onde $d_T(R_a, R_b)$ é a dissimilaridade de textura entre R_a e R_b, como definido na Equação (5.11) e Λ é uma penalidade estatística de fusão, definida como:

$$\Lambda(R_a, R_b) = \frac{D_f^2}{2Q}\left[\frac{ln(\Psi(f)^2)}{\Psi(R_a)} + \frac{ln(\Psi(f)^2)}{\Psi(R_b)}\right], \tag{5.15}$$

onde $\Psi(R)$ é o número de elementos (*pixels*) na região R, consequentemente $\Psi(f)$ é o número de *pixels* da imagem, D_f representa a faixa de valores positivos em f e Q é um termo de regularização, que determina a verossimilhança da união.

Para determinar se duas regiões devem ser unidas, a função de verossimilhança é comparada a uma variável aleatória u, gerada com distribuição uniforme em $[0, 1]$. As regiões são unidas se o valor da função de verossimilhança satisfizer o seguinte predicado:

$$\mathcal{P}(R_a, R_b) = \begin{cases} 1 & \text{se } u \leq \alpha(R_a, R_b) \\ 0 & \text{caso contrário.} \end{cases} \tag{5.16}$$

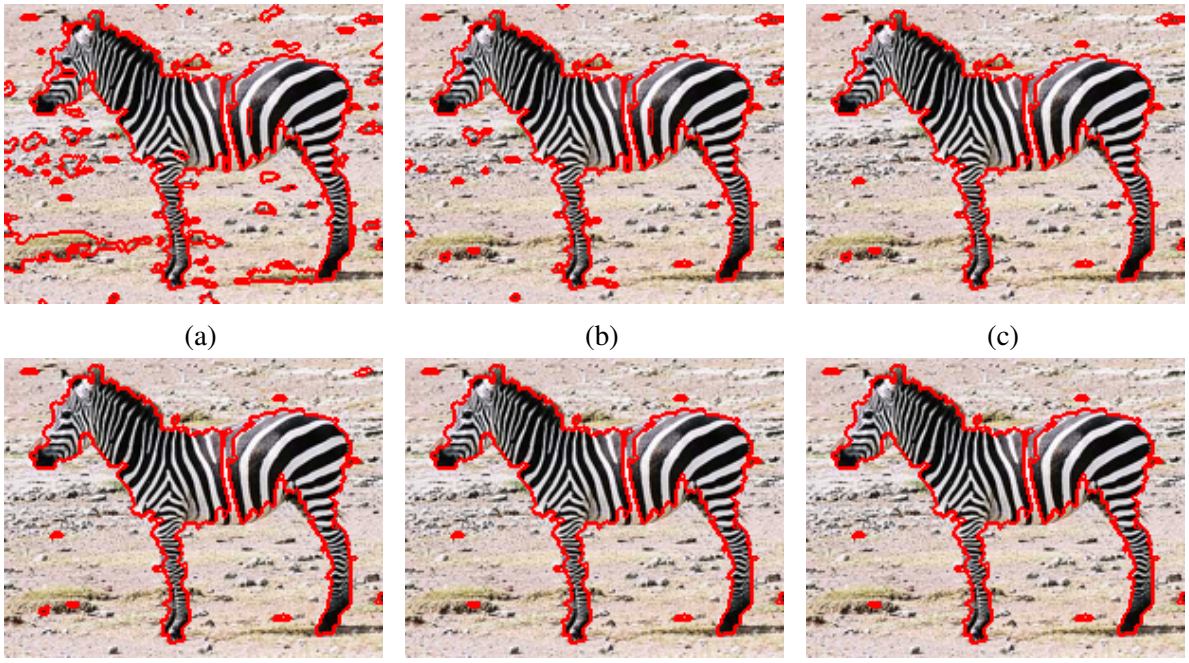
Comparado a fusão estocástica de regiões proposta por Wong et al. (2011), o critério de fusão proposto neste trabalho é capaz de manipular as feições de texturas mais robustamente por permitir o balanço entre os canais da imagem. O critério de fusão proposto por Wong et al. (2011) é limitado ao uso de características muito simples, tais como tom de cor ou luminância, em que a dissimilaridade pode ser facilmente medida pela diferença entre o valor esperado as regiões. Com a formulação proposta na Equação (5.15), o critério de fusão passa a ser mais adequado para a representação de texturas usada aqui, permitindo uma descrição mais robusta das feições de texturas, tanto locais como globais.

A seguir, detalhamos alguma vantagens de empregar o modelo de textura estocástico multi-escala proposto em conjunto com a estratégia de fusão estocástica de regiões. As representações de textura de regiões pequenas, com apenas poucos *pixels*, tendem a ser estatisticamente pobres, conforme discutido na Seção 4.3.4 e isso pode ocasionar que regiões com a mesma textura sejam consideradas como diferentes. Contudo, utilizando a Equação (5.14), a probabilidade de união aumenta conforme o tamanho das regiões de textura diminuem; então, as menores regiões são as que possuem maior probabilidade de serem fundidas. Isso não apenas ajuda no tratamento de ruído (artefatos), mas também torna o processo mais robusto à sobre-segmentação (quando as menores regiões segmentadas tendem a serem fundidas nos estágios iniciais). Para lidar ainda melhor com pequenas regiões de textura, em especial nos casos em que a região possui apenas um *pixel*, sempre que uma região possuir menos de dez *pixels*, o modelo de textura para essa região é substituído pelo modelo de textura da mesma região dilatada com uma estrutura de vizinhança-8.

Outrossim, quanto mais similares forem duas regiões de textura maior será o valor de α, aumentando a homogeneidade interna das regiões de textura segmentadas, mesmo quando as regiões são pequenas. Da mesma maneira, valores maiores de Q tendem a produzir menores valores de α, reduzindo a probabilidade de fusão e aumentando o número de segmentos de texturas encontrados na imagem. Por outro lado, reduzir o valor de Q aumenta a probabilidade de fusão, reduzindo a chance de obter sobre-segmentação.

Como visto no Capítulo 3, Seção 3.1.2, um dos fatores que mais influenciam no resultado dos algoritmos de fusão de regiões é a ordem em que os pares são avaliados. A fim de garantir que as regiões de texturas sejam unidas consistentemente, as arestas do grafo de adjacência são ordenadas e adicionadas a uma fila de prioridade. Cada par de regiões adjacentes está associado a uma única aresta do grafo de adjacência; essa fila de prioridade é usada em todos os testes de união, em que as arestas são inseridas em ordem decrescente de seus pesos (dissimilaridade entre as regiões).

Uma vez que regiões menores são mais propensas a serem fundidas, a ordem no processo de fusão de regiões é importante. Na inicialização, cada *pixel* $p \in I$ é associado a uma única região R no grafo de adjacências e, dessa forma, os pares de regiões são ordenados baseado em suas diferenças *pixel*-a-*pixel*. Em regiões com apenas um *pixel*, a



(a) (b) (c)

(d) (e) (f)

Figura 5.2: Resultados parciais da segmentação estocástica das regiões de textura proposta entre diferentes iterações de uma mesma segmentação. 5.2a 1 iteração $Q = 300$, 5.2b 2 iteração $Q = 273.576$, 5.2c 3 iteração $Q = 227.067$, 5.2d 4 iteração $Q = 209.957$, 5.2e 5 iteração $Q = 203.663$, 5.2f 6 iteração $Q = 201.348$.

representação de textura de regiões torna-se imprecisa para quantificar a dissimilaridade das texturas com Equação (5.11); então, nós inicializamos os pesos das arestas com gradientes locais da imagem. Para incorporar alguma informação textural, esses gradientes são computados na representação muti-escala de imagem I' como diferenças *pixel*-a-*pixel*. Uma vez que essa representação da imagem descreve cada *pixel* por estatísticas de múltiplas ordens, ela produz uma representação mais rica do que os valores dos *pixels* em sua forma natural.

Durante o processo de fusão, sempre que um par de regiões é analisado, ele é removido da fila. Toda vez que ocorre uma fusão, o grafo de adjacências é atualizado e a fila de prioridade é modificada para refletir tais mudanças. Quando todos os pares de regiões adjacentes tiverem sido analisados, a fila de prioridades torna-se vazia e o grafo de adjacências conterá o resultado da segmentação.

5.2.4 Estratégia de Segmentação Iterativa

Antes do processo de fusão ser iniciado, cada *pixel* é atribuído a uma única região, então, cada vértice no grafo de adjacências G está associado a uma região. Conforme os pares de regiões são unidos e a fila de prioridade é esvaziada, G é modificado para refletir o novo conjunto de regiões gerado pela fusão das regiões. Após a fila de prioridades tornar-se vazia, o grafo de adjacências passa a representar a segmentação das texturas da imagem. O processo de segmentação é então iterado, repetindo o processo de fusão estocástica de regiões de textura, produzindo resultados mais confiáveis.

As regiões obtidas na primeira iteração são agora usadas para construir um novo grafo de adjacências e os novos pares de regiões adjacentes são inseridos na fila de prioridade,

que é novamente ordenada em ordem ascendente de dissimilaridade das regiões. Diferentemente de antes, quando todas as regiões possuíam apenas um *pixel*, as regiões no mapa de inicialização são agora maiores, permitindo que a fila de prioridade seja ordenada de acordo com a dissimilaridade de textura dos pares de regiões. O processo de fusão é então repetido utilizando um novo valor para o termo de regularização. O valor de Q é decrementado exponencialmente em cada iteração de acordo com

$$Q_k = (Q_1 - Q_{min}) * \exp(1 - k) + Q_{min} \tag{5.17}$$

onde $k \geq 1$ é a iteração atual, Q_1 é o valor do termo de regularização na primeira iteração (valor inicial), $Q_{min} = 200$ é o valor mínimo para o termo de regularização e Q_k é o valor efetivamente utilizado na k-ésima iteração do processo de fusão.

Ao final de cada iteração, espera-se que o número de regiões seja menor do que na iteração anterior, tornando a segmentação mais precisa. Essas iterações são realizadas repetidamente até que a segmentação da imagem convirja, isto é, até que a mudança no valor de Q em uma iteração seja suficiente para produzir qualquer alteração no grafo de adjacências. Nessas iterações, podemos evitar a ocorrência de sub-segmentação ao escolher um valor alto para Q (termo de regularização inicial), com um baixo risco de sobre-segmentação (MEDEIROS; SCHARCANSKI; WONG, 2013a). Na Figura 5.2, são mostrados alguns resultados parciais obtidos em cada iteração da técnica de segmentação proposta, aplicada a imagens naturais.

5.3 Detecção de Segmentos de Pele

Nas Sessões 5.1 e 5.2 definimos, respectivamente, a metodologia para representação e modelagem estocástica de texturas e a estratégia para segmentação de texturas que são utilizadas na solução dos objetivos específicos deste trabalho. Agora, com esses conceitos definidos, podemos descrever os detalhes do método buscando atingir o nosso objetivo geral. Como comentado no início deste capítulo, o algoritmo aqui proposto possui quatro etapas. São elas: 1) treinamento do modelo de cor de pele; 2) treinamento do modelo de textura de pele; 3) segmentação da imagem em regiões de textura com homogeneidade consistente; 4) classificação dos segmentos encontrados. Esta sessão destina-se a detalhar os processos envolvidos em cada um destes estágios, dos quais a Figura 5.3 apresenta um esboço visual das informações nelas extraídas.

5.3.1 Modelagem de Cor

Neste trabalho, nós propomos uma combinação de modelos de cor e textura para representar e identificar regiões de pele. Primeiramente empregamos um modelo de mistura de gaussianas para descrever o tom de cor da pele. Como descrito anteriormente (veja Capítulo 3, Seção 3.2.3), a GMM é um modelo estatístico que descreve a densidade de probabilidade de um conjunto de dados na forma de uma soma ponderada de um número finito de funções gaussianas. Aqui, utilizamos o algoritmo EM (*expectation maximisation*) (DEMPSTER; LAIRD; RUBIN, 1977) para obter os parâmetros Θ_i de uma mistura de η componentes gaussianas $\mathcal{G}(r, b; \Theta_i)$, $i = 1, ..., \eta$, que constituem nosso modelo de cor $\rho(r, b)$:

$$\rho(r, b) = \sum_{i=1}^{\eta} \gamma_i \cdot \mathcal{G}(r, b; \Theta_i) \tag{5.18}$$

(a) (b) (c) (d) (e)

Figura 5.3: Resultados preliminares do método se detecção de pele proposto neste trabalho. (5.3a) imagem original; (5.3b) resultado da segmentação de texturas; (5.3c) similaridade de cor; (5.3d) similaridade de textura (regiões mais claras possuem maior similaridade); (5.3e) resultado final da classificação.

onde $\mathcal{G}(r, b; \Theta_i)$ é uma gaussiana bivariada (conforme descrito na Equação (3.9)), Θ_i denota seus parâmetros e γ_i é o peso do componente na mistura ($\sum_{i=1}^{\eta} \gamma_i = 1$ e $\gamma_i \geq 0, \forall i$;) finalmente, r e b são os componentes de cromaticidades vermelho e azul da representação de cores rgb (RGB normalizado), como descritos no Capítulo 4, Seção 4.1. (veja Equação (4.1)). Essa representação de cromaticidades foi escolhida porque podemos obter uma discriminação mais robusta entre pele e outros materiais do que no espaço de cores RGB (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007).

5.3.2 Modelagem de Textura

Para alcançar uma segmentação de pele ainda mais robusta, também construímos um modelo de textura de pele. Esse modelo é obtido utilizando a abordagem de representação e modelagem estocástica de texturas proposta anteriormente. Todos os estágios da extração dos vetores de feições STR são aplicados da maneira como foram descritos na Seção 5.1. Posteriormente, na construção do dicionário de textons, incluímos textons de dois exemplos de treinamento que representarão as classes de pele e de fundo da imagem.

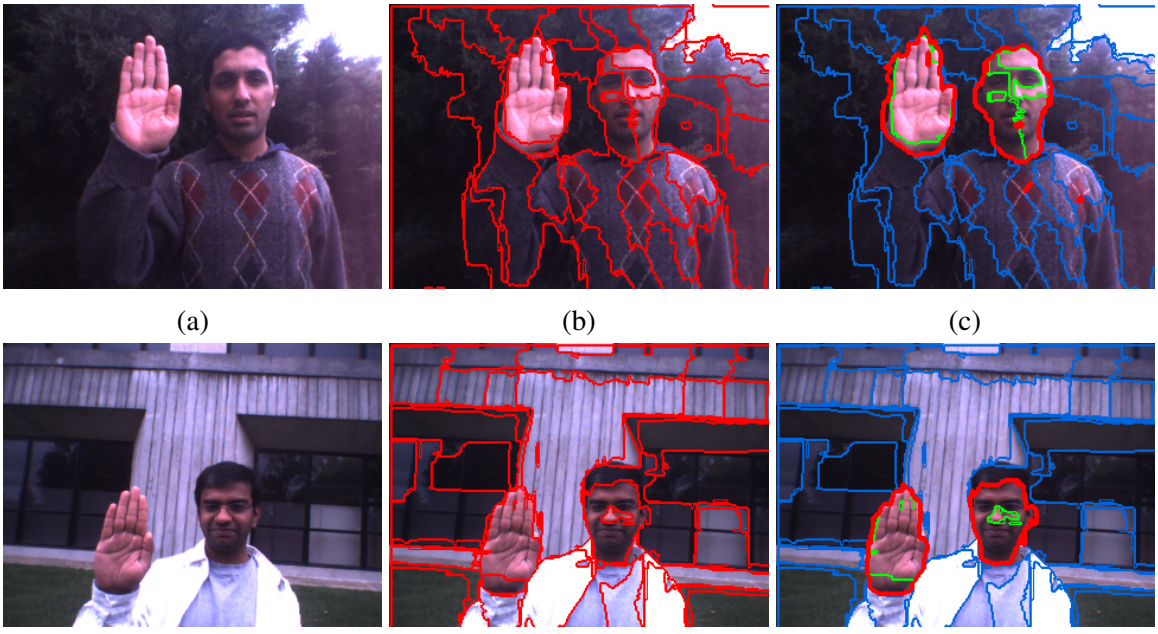
Os protótipos das feições de pele são obtidos a partir das amostras de pele em um banco de imagens (*database* ou *dataset*) $\mathcal{D}$. Como um estágio de treinamento do método proposto, obtemos não apenas K textons de pele, mas também um conjunto de modelos de pele. O modelo de textura que será utilizada para classificação será então uma coleção de conjuntos de histogramas $\mathcal{S}$, definido como

$$\mathcal{S} = \{T(S_i) | S_i \in \mathcal{D}\}, \tag{5.19}$$

onde S_i é uma imagem com exemplos de pele contida no dataset $\mathcal{D}$.

Para cada imagem no banco de dados de treinamento, um exemplo de textura (série de histogramas de ocorrência de textons de pele) é obtido a partir de todas as regiões de pele na imagem.

Como discutido em anteriormente, no cenário proposto, não possuímos qualquer conhecimento *a priori* sobre as classes de textura que podem ocorrer na imagem, logo, qualquer tentativa de predição das texturas desses materiais seria limitante ou insuficiente para algum cenário (veja Seção 5.1.3). Para resolver esse problema, os textons da classe fundo podem ser obtidos a partir da própria imagem de teste (MEDEIROS; SCHARCANSKI; WONG, 2013b). Mesmo que esses textons ocorram em quantidades diferentes para texturas diferentes, estudos recentes demonstraram que um número suficiente K' de centroides é capaz de representar adequadamente todas as texturas na imagem (MEDEI-



(a) (b) (c)

(d) (e) (f)

Figura 5.4: Análise de segmentos de textura encontrados pela fusão de regiões (5.4a), (5.4d) imagem original; (5.4b), (5.4e) resultado da segmentação de texturas; (5.4c), (5.4f) detecção de pele obtida ao final do processo. Os contornos vermelhos correspondem as fronteiras entre os segmentos no final da etapa, os contornos azuis e verdes indicam, respectivamente, fronteiras entre regiões de fundo e pele, que foram eliminadas após a etapa de classificação.

ROS; SCHARCANSKI; WONG, 2013a). Uma vez que esses textons são obtidos sob demanda para cada imagem de teste, eles não são utilizados para computar os modelos de histogramas de ocorrência de pele.

5.3.3 Segmentação das Regiões de Textura

Para alcançar uma segmentação mais precisa das regiões de pele, que permita uma melhor identificação do formato da região, nossa abordagem para detecção de pele considera regiões de texturas (conjuntos de *pixels*) ao invés de detectar *pixels* de pele de maneira pontual (cada *pixel* individualmente). Para obter esse resultado, empregamos o processo de fusão estocástica de regiões de textura (MEDEIROS; SCHARCANSKI; WONG, 2012).

Aqui, executamos todo o processo de fusão conforme descrito na Seção 5.2. Para realizar as comparações entre as texturas das regiões no grafo de adjacências e mensurar sua dissimilaridade (veja Equação (5.11)), computamos seus histogramas usando o dicionário de textons obtido na última etapa (treinamento dos modelos de textura), que é composto de K feições de pele e K' feições da própria imagem.

A única modificação em relação ao método de segmentação descrito anteriormente é em relação às etapas de aprimoramento iterativo da segmentação, que não são realizadas. A razão que embasa dessa alteração é que no caso de alguma fusão indevida ocorrer (sub-segmentação), o desempenho da etapa final de classificação seria prejudicado. No entanto, caso alguma possível fusão devida deixe de ser realizada, se a textura em ambas

estiver adequadamente representada, o classificador será capaz de avaliá-las corretamente.

Pela definição dos objetivos primários de segmentação de imagens dada no Capítulo 3, Seção 3.1.2, as regiões obtidas nesse processo devem apresentar simultaneamente homogeneidade interna (veja Equação (3.3)) e área máxima (veja Equação (3.4)). Nessa técnica, a estratégia de fusão estocástica de regiões (WONG; SCHARCANSKI; FIEGUTH, 2011) garante que o critério de homogeneidade interna seja satisfeito. Já as iterações asseguram que as texturas obtidas possuam o maior tamanho possível (MEDEIROS; SCHARCANSKI; WONG, 2012).

Como neste trabalho pretendemos classificar cada segmento independentemente dos seus vizinhos, basta que seja satisfeita a condição de homogeneidade interna dos segmentos. Não é necessário garantir que os segmentos de textura tenham tamanho máximo porque, se os segmentos adjacentes pertencem à mesma classe (entre pele ou fundo), o critério de homogeneidade garante que na próxima, e última, etapa desse método, eles são classificados corretamente. A Figura 5.4 ilustra como essa propriedade se verifica na prática.

5.3.4 Identificação de Pele nas Regiões de Textura

Uma vez que os segmentos de textura tenham sido obtidos por meio do algoritmo modificado de fusão estocástica de texturas (*stochastic texture merging*, STM), a detecção de pele é aplicada a cada região independentemente, pelo uso dos descritores de cor e textura. Como discutido na sessão anterior, a identificação de pele assume que os segmentos de textura obtidos são homogêneos no sentido de que uma única textura e que essa esteja bem representada pelo modelo de aparência de textura.

Para rotular como pele, a textura de uma dada região R encontrada no grafo final de adjacências G, os critérios de similaridade de cor e textura devem ser satisfeitos. Visto que o modelo de cor de pele é representado por um GMM $\rho(r, b)$, escolhemos o valor esperado dos canais vermelho e azul do rbg normalizado em R, denotado $E[R]$, como o descritor de cor da região. A similaridade de cor $SKIN_C(R)$ é então computada pela probabilidade de $E[R] \equiv (r, b)_{E[R]}$, correspondendo a um tom de pele. Assim,

$$SKIN_C(R) = \rho(E[R]). \tag{5.20}$$

A similaridade de textura de R com cada exemplo de textura S_i é computada como a soma ponderada das diferenças de histogramas, definidas como uma variação da Equação (5.11). Portanto

$$d'_T(R, S_i) = \frac{W'_g \begin{bmatrix} \sum_{i=1}^{N} \chi^2(H_{L,i}(R), H_{L,i}(S_i)) \\ \sum_{i=1}^{N} \chi^2(H_{a,i}(R), H_{a,i}(S_i)) \\ \sum_{i=1}^{N} \chi^2(H_{b,i}(R), H_{b,i}(S_i)) \end{bmatrix}}{\sum W'_g} \tag{5.21}$$

onde W'_g é um vetor de pesos, assim como W_g é na Equação (5.11), e χ^2 é o teste de sobreposição de histogramas chi-quadrado, como na Equação (4.4). essa medida de similaridade foi escolhida porque é limitada ao intervalo $[0, 1]$, facilitando o uso de discriminantes lineares (o que não é o caso de outras medidas de similaridade populares, como a distância de Bhattacharya).

Usando a Equação (5.21), a similaridade de textura para cada exemplo S_i é estimada e a amostra mais similar a R é estimada por $SKIN_T(R)$ como segue:

$$SKIN_T(R) = \max_{S_i} d'_T(R, S_i). \tag{5.22}$$

Dada uma região arbitrária R, obtida pela etapa de segmentação de texturas, classificamos essa região como pele se $SKIN_c \geq \beta_c$ e $SKIN_T \geq \beta_T$, onde β_c é um limiar de similaridade de cor e β_T é um limiar de similaridade de textura. Isso significa que, para ser classificado como pele, R deve possuir aparência em termos de cor, tanto quanto em termos de textura. Esse processo de classificação é ilustrado na Figura 5.5.

É interessante observarmos também que um segmento do fundo da imagem poderia ser incorretamente rotulado como pele baseado no critério de cor (ou textura) independentemente, mas se o critério de textura (ou cor) não é satisfeito simultaneamente o segmento de fundo não é rotulado (erroneamente classificado) como um segmento de pele. Outra observação importante é que se ambos os limiares são definidos com valores mais baixos, já que ambas as medições são efetuadas independentemente uma da outra, a ocorrência de falsos positivos/negativos será menor do que seria obtido considerando apenas um dos critérios mencionados anteriormente.

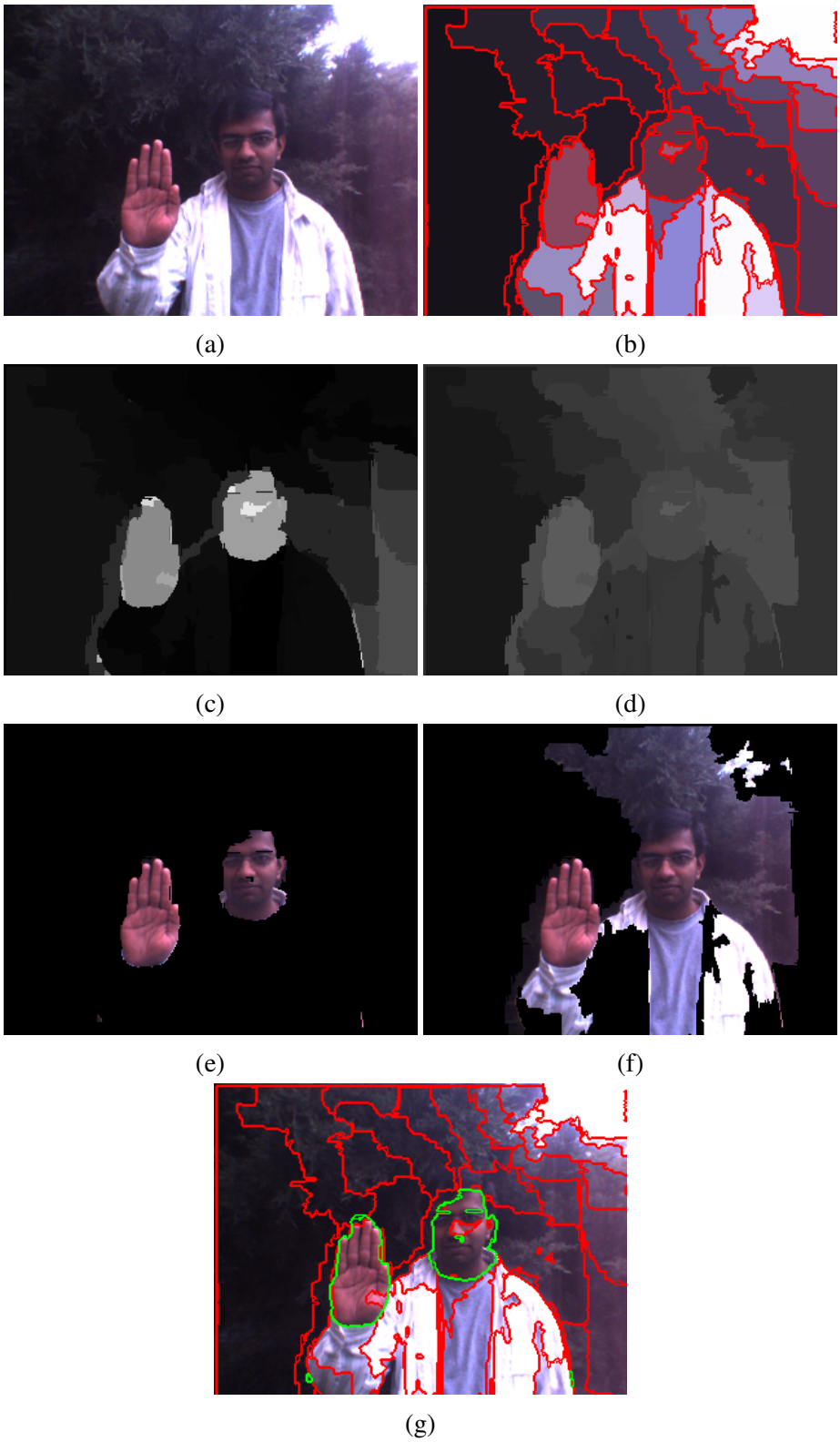


Figura 5.5: Comparação entre as classificações por pele e textura. (5.5a) imagem original; (5.5b) segmentos de textura, representados pela cor média da região. (5.5c) similaridade de cor; (5.5d) similaridade de textura; (5.5e) detecção de pele por cor; (5.5f) detecção de pele por textura; (5.5g) resultado final, indicando as regiões de pele com contornos verdes, e regiões unidas pela classificação com os vermelhos.

6 RESULTADOS EXPERIMENTAIS

Neste capítulo, iremos avaliar e discutir os resultados obtidos com a metodologia proposta por meio de uma série de testes e comparações com outros métodos selecionados do estado-da-arte e com os *ground truths*[1] fornecidos pelos bancos de imagens utilizadas. Primeiramente avaliaremos os resultados obtidos pelo método de fusão estocástica de regiões texturadas utilizando dois bancos de imagens de texturas. Com base nos resultados desses experimentos, definiremos a configuração de parâmetros que será empregada na etapa final dos experimentos, relativos à técnica de detecção de pele. Nessa segunda etapa verificaremos a eficácia do método proposto, considerando comparações com métodos selecionados do estado da arte em detecção de pele, executados em um conjunto de imagens de pele.

6.1 Avaliação do Método na Fusão Estocástica de Regiões de Textura

Para atingir o objetivo geral proposto neste trabalho (detecção de pele), definimos como objetivo específico o desenvolvimento de uma técnica de segmentação de texturas em imagens naturais, que foi utilizada como ferramenta no processo de detecção das regiões de pele. Esse método de segmentação de texturas, entretanto, constitui uma ferramenta de propósito geral, que pode ser aplicada aos mais diversos problemas e cenários. Nesta seção, discutiremos os resultados da técnica de segmentação de imagens naturais, como apresentada no Capítulo 5, por meio de fusão estocástica (veja Seção 5.2) de texturas descritas por feições estocásticas multi-escala (veja Seção 5.1).

Para avaliar a qualidade dos resultados obtidos pelo método de segmentação estocástico de texturas proposto, foram realizados testes em dois bancos de imagens: *Prague* (HAINDL; MIKEŠ, 2008) e *BSDS300* (MARTIN et al., 2001). Ambos os bancos de imagens estão publicamente disponíveis e foram projetados especialmente para a comparação de métodos de segmentação de textura. Em ambos os bancos de imagens, *Prague* e *BSDS300*, utilizamos os mesmos parâmetros de segmentação como descrito a seguir. Para a primeira etapa do processo de segmentação, na qual as características de textura são extraídas da imagem, a representação multi-escala utiliza $N = 7$ níveis e os pedaços de imagem são extraídos com largura $W = 5$ *pixels* em torno de cada *pixel*. Cada pedaço de imagem é então transformado com a STR, reduzindo seu tamanho para $M = 10$ elementos.

Como discutido na Seção 5.1, o tamanho do dicionário de textons depende do número de classes de texturas no problema (imagem a ser segmentada). Uma representação precisa de textura geralmente requer um grande número K de textons no dicionário (GRAF;

[1] *Ground truth*: Resultado esperado da segmentação.

LUSCHGY, 2000), especialmente quando não temos qualquer conhecimento prévio sobre o número de texturas na imagem. No entanto, conforme K aumenta, a complexidade da segmentação também cresce.

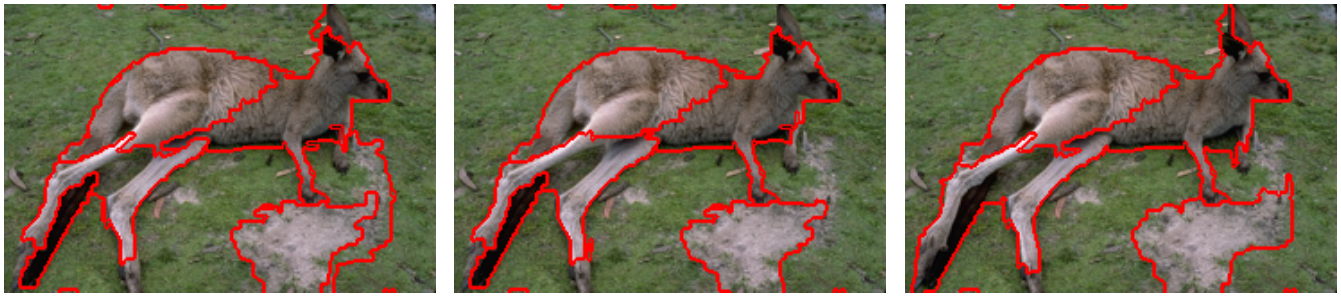
De fato, nosso modelo estocástico de textura não usa diretamente os textons para representar as texturas, ao invés disso, usamos a probabilidade de ocorrência de todos os textons do dicionário para representar uma amostra de textura. Desse modo, mesmo se uma região de textura compartilhe textons com outras regiões de textura, ou se ela contribuir com poucos textons na construção do dicionário, essa representação ainda pode discriminar eficientemente as amostras de textura contanto que a taxa de ocorrência desses textons não seja a mesma em ambas. Portanto, a escolha do valor de K deveria oferecer um balanço entre o erro de quantização na representação de texturas (inversamente proporcional a K), e o custo computacional de um maior dicionário de textons (proporcional a K). Baseado na literatura atual de dicionário de textons (LIU; FIEGUTH, 2012; VARMA; ZISSERMAN, 2009) determinamos empiricamente que $K = 30$ textons permite um dicionário de representatividade satisfatório para a segmentação da imagem a um custo computacional aceitável.

A melhor configuração dos parâmetros pode variar para diferentes bancos de imagens, e testes mostraram que os parâmetros da extração e modelagem da texturas têm uma forte dependência da escala da textura analisada. Se a escala aumenta, o tamanho das primitivas da textura tende a aumentar também, e uma nova configuração de parâmetros pode ser necessária, uma vez que o dicionário de textons necessário seria diferente. Mas, para os bancos de imagem usados em nossos experimentos, a configuração proposta mostrou-se eficiente.

Deve ser observado que a representação multi-escala de textura utilizada neste trabalho fornece simultaneamente estatísticas locais de múltiplas ordens, e a textura é então descrita em vários níveis de detalhes. No entanto, esse processo não é invariante à escala, porque quando variamos a escala das texturas (tamanho das primitivas) estatísticas diferentes são capturadas. Então, nossa representação multi-escala é sensível ao tamanho das primitivas da textura, o que significa que, mesmo se estatísticas de múltiplas ordens são empregadas na extração de feições, a representação da textura não é invariante à escala e precisará de uma configuração diferenciada para segmentar texturas em diferentes escalas.

O banco de imagens de texturas *Prague* é composto de blocos de textura gerados por computador. Cada bloco contém texturas escolhidas aleatoriamente a partir de um conjunto de 114 imagens de textura, com 512×512 *pixels*. Em nossos testes usamos 30 blocos, cada um contendo nove texturas escolhidas aleatoriamente. Para reduzir a complexidade computacional, essas texturas foram primeiramente redimensionadas para 78×78 *pixels* e então agregadas em blocos de 234×234 *pixels*. Nesses conjuntos de blocos justapostos, as fronteiras entre as texturas são conhecidas e cada imagem desse banco de imagens tem um *ground truth* objetivo (resultado ideal para segmentação).

O banco de imagens *BSDS300* consiste de 300 imagens naturais a cores de tamanho 481×321 *pixels*. Esse banco de imagens contém 200 imagens que compõem um conjunto de treinamento e as 100 imagens restantes formam o conjunto de teste. Lembre-se que em nossa abordagem (para segmentação de texturas com propósito geral) não há fase de treinamento, então o dicionário de textons é construído apenas baseado nas 100 imagens do conjunto de testes (ao segmentar cada imagem). Essas imagens foram redimensionadas para 241×161 para reduzir a complexidade computacional. Diferentemente do banco de imagens *Prague*, que possui apenas mosaicos agregados artificialmente, em imagens



(a) (b) (c)

Figura 6.1: Resultados da segmentação com diferentes valores de W_G em imagens naturais. (6.1a) $W_G = [1, 1, 1]$; (6.1b) $W_G = [1, 1.5, 1.5]$; (6.1c) $W_G = [2, 1, 1]$.

naturais a segmentação ideal das texturas pode ser muito subjetiva, por isso o *BSDS300* inclui um conjunto com até 10 segmentações, feitas a mão por diferentes indivíduos, com o *ground truths* para cada imagem.

Como mencionado anteriormente, existem dois parâmetros que controlam a etapa de fusão estocástica de regiões: (a) o vetor de pesos W_g, e (b) o termo de regularização Q. Inspirado nos experimentos e conclusões dos trabalhos de Nock e Nielsen (2004) e Wong et. al. (2011), os experimentos analisados neste trabalho avaliam configurações em que Q assume valores (discretos) no intervalo $[200, 400] \subset \mathbb{N}$ e $W_g^c \in [1, 2] \subset \mathbb{R}$.

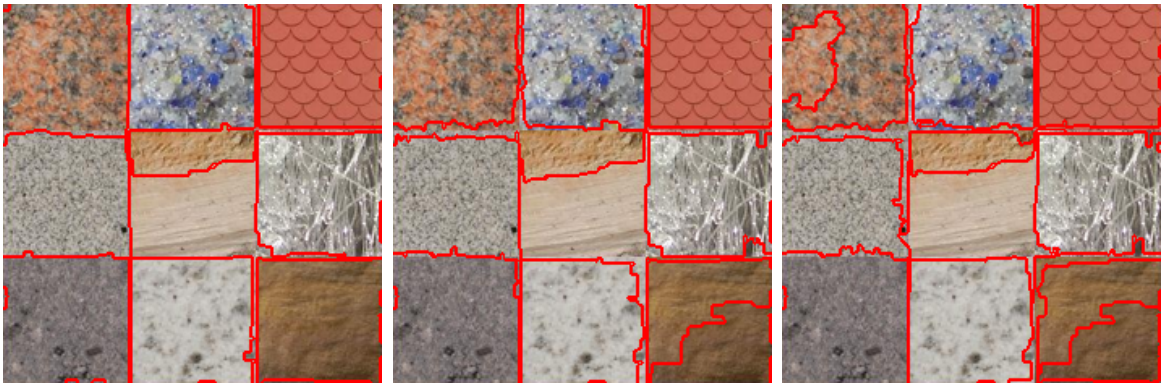
Como podemos observar na Figura 6.1, o vetor de pesos W_g regula o equilíbrio entre as feições provenientes dos canais de cor e luminância, e mudanças em W_g afetam quanto das informações de cor ou de luminância são levadas em conta ao mensurar a dissimilaridade entre regiões. Na prática, o resultado de alterar os valores em W_g tem um impacto na segmentação, e fusões podem (ou não) ocorrer, dependendo de como esse parâmetro é definido.

O termo de regularização Q, por outro lado, permite controlar a quantidade de segmentos de textura no grafo adjacências G ao final do processo de segmentação. Ao aumentar o valor de Q, causamos redução dos valores de α na Equação (5.14), o que diminui a probabilidade das fusões e, consequentemente, leva à descoberta de mais regiões na segmentação final. Em contrapartida, utilizar valores mais baixos de Q, terá o efeito oposto sobre a segmentação da imagem, gerando menos regiões segmentadas. A Figura 6.2 mostra alguns resultados comparando o uso de diferentes valores de Q.

As iterações do método de fusão estocástica de regiões de textura contribuem para a convergência de segmentação (veja Equação (5.17)), uma vez que oferecem a todos os pares de regiões adjacentes de serem unidos novamente. Além disso, devido à diminuição no valor de Q por um fator exponencial, permitimos que os maiores segmentos mantenham a estabilidade entre as iterações (já que menos fusões serão realizadas). Ainda assim, é possível que um maior número de regiões de textura seja obtido ao final da segmentação, basta utilizar um valor maior de Q, como podemos observar na Figura 6.2.

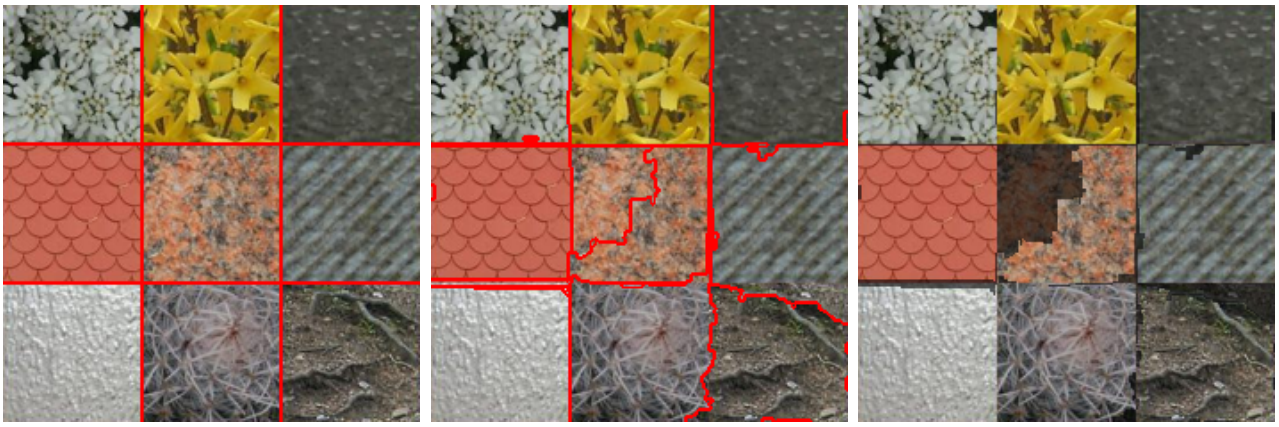
6.1.1 Medida Quantitativa do Erro de Segmentação

É de conhecimento comum que avaliar resultados de segmentação de imagens é uma tarefa muito desafiadora. Até mesmo para imagens sintéticas, comparar dois resultados de segmentação diferentes para a mesma imagem pode ser bastante subjetivo. Mesmo que uma avaliação quantitativa da qualidade de segmentação de imagens possa não refletir integralmente a eficácia do método testado para imagens, ainda vale a pena perseguir



(a) (b) (c)

Figura 6.2: Resultados da segmentação final (após as todas as iterações) utilizando valores crescentes de Q em imagens sintéticas. (6.2a) $Q = 200$; (6.2b) $Q = 300$; (6.2c) $Q = 400$.



(a) (b) (c)

Figura 6.3: Exemplo da medida de acurácia de segmentação proposta. Os contornos vermelhos indicam fronteiras entre as texturas na imagem. (6.3a) o *ground truth* para a segmentação de texturas; (6.3b) resultado da segmentação utilizando o método proposto; (6.3c) erro de segmentação (os *pixels* mais claros indicam as áreas segmentadas erroneamente em relação a (6.3a)) A acurácia computada nesse caso é de $\mathrm{ACC}_G^S = 91.95\%$.

esse objetivo. Neste trabalho, utilizamos uma avaliação quantitativa do resultado da segmentação que mede a semelhança entre o mapa de regiões obtido e o *ground truth*.

Para uma dada imagem I, seja S o mapa resultante da segmentação, formado por um conjunto de regiões não sobrepostas rotuladas como $S_i \in \{1, \cdots, k\}$ e seja G o *ground truth* para a segmentação de I, também formada por um conjunto de regiões não sobrepostas rotuladas como $G_j \in \{1, \cdots, n\}$ então, para medir o erro de segmentação, cada região de textura G_j no *ground truth* é comparada espacialmente (*pixel*-a-*pixel*, para obter o grau de sobreposição), com uma região de textura em $\{S\}$ que satisfaça

$$\hat{S}_j = \arg\max_{S_i} |G_j \cap S_i|, \tag{6.1}$$

onde $\hat{S}_j$ é o pedaço de textura que se encontra na mesma posição espacial que a região de textura G_j. A acurácia da segmentação é estimada com base no número de *pixels* segmentados que correspondam ao tamanho de cada região de textura G_j. Assim, definimos

Tabela 6.1: Desempenho da segmentação no banco de imagens de texturas *Prague*

Método	Acurácia média. (%)	σ (%)
CTMS	86.83	5.98
JSEG	87.46	3.90
CPGS	83.72	6.42
SRM	82.60	3.62
Método Proposto ($Q = 200$)	**92.76**	**3.27**
Método Proposto ($Q = 300$)	89.13	3.90
Método Proposto ($Q = 400$)	85.74	3.53

a acurácia da segmentação, dado o mapa da região $\{S\}$, e o conjunto de *ground truths* $\{G\}$. Portanto,

$$\mathrm{ACC}_G^S = \frac{1}{\Phi(I)} \sum_{i=1}^{n} |G_i \cap \hat{S}_i|, \tag{6.2}$$

onde ACC_G^S indica a proporção de *pixels* da imagem que foram corretamente atribuídos ao rótulo indicado em $\{G\}$.

Para lidar com casos de sub-segmentação, se um segmento (em $\{R\}$) for associado a mais do que uma região de textura (em $\{G\}$), consideramos apenas o segmento com o rótulo de textura que possui mais *pixels* classificadas corretamente. Em casos de sobre-segmentação, a precisão da segmentação diminui naturalmente uma vez que cada região de textura em G só é parcialmente correspondida. Assim, em segmentações incorretas, as regiões de textura serão apenas parcialmente correspondidas, ou não serão associadas à qualquer textura em G. A Figura 6.3 mostra um exemplo da medida de acurácia de segmentação descrita, aplicada a um mapa de segmentação Figura (6.3b), para um dado *ground truth* (6.3a). Essa avaliação da segmentação mede a semelhança entre a segmentação de mapa S e do *ground truth* G estimando a sobreposição das regiões correspondentes em ambos os mapas. Assim, quanto mais semelhante $\{S\}$ é a G, maior é ACC_G^S, uma vez que a área das regiões sobrepostas em $\{S\}$ e $\{G\}$ aumenta.

6.1.2 Comparação com o Estado-da-Arte

Nesta seção, comparamos a segmentação produzida pela técnica proposta com resultados de segmentação obtidos por outros métodos do estado-da-arte, tais como o método de segmentação por medidas de texturas de cor (*color texture measurement segmentation*, CTMS) (HOANG; GEUSEBROEK; SMEULDERS, 2005), JSEG (DENG; MANJUNATH; SHIN, 1999), segmentação por gradientes (*gradient segmentation*, CPGS) (FOWLKES; MARTIN; MALIK, 2003), e fusão estocástica de regiões (*stochastic region merging*, SRM) (WONG; SCHARCANSKI; FIEGUTH, 2011). Embora a SRM não use informações de textura na segmentação de imagens, incluímos esse método em nossa comparação porque o método proposto é uma extensão do SRM, e porque estamos interessados em métodos do estado-da-arte em segmentação de imagens coloridas e algoritmos de fusão de regiões. Nessas comparações, foram utilizados os parâmetros especificados na respectiva literatura de cada método. O critério de avaliação é baseado no método descrito na Equação (6.2).

A Tabela 6.1 mostra a comparação dos métodos para o banco de imagens *Prague*, que contém 30 imagens sintéticas, por meio da média da acurácias de cada imagem no banco de imagens, e da dispersão desses valores (desvio padrão). Nessa tabela, os testes

Tabela 6.2: Desempenho da segmentação no banco de imagens de texturas *BSDS300*.

Método	Acurácia média. (%)	σ (%)
CTMS	78.17	12.84
JSEG	75.86	13.21
SRM	75.41	11.12
Proposto ($Q = 200$)	**80.55**	**13.36**
Proposto ($Q = 300$)	75.44	13.90
Proposto ($Q = 400$)	69.35	14.74

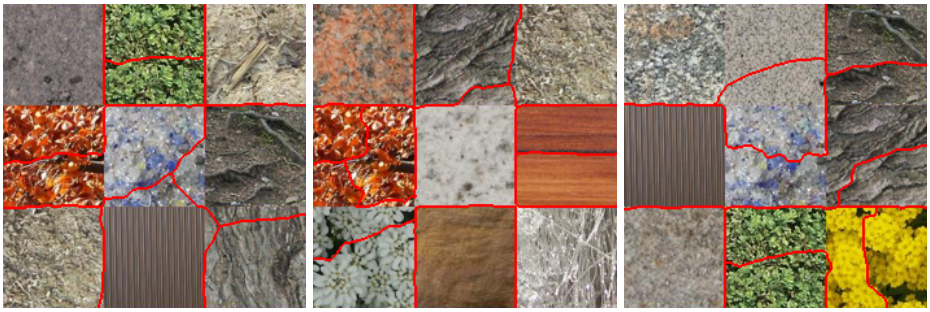
utilizaram o parâmetro W_G foi configurado como $W_G = [2, 1, 1]$, que fornece o melhor resultado (de acordo com a equação 6.2) e foram utilizados três valores de Q (200, 300 e 400). Em média, o SMTR (método proposto) demonstrou uma maior precisão do que os métodos CTMS, JSEG, CPGS e SRM, com menor desvio padrão (σ). Com base em nosso critério de avaliação (veja Equação (6.2)), o mapa de segmentação obtido pelo método proposto corresponde ao ground truth em $92,76\%$ dos *pixels* (em média), enquanto os outros métodos geram segmentações que são menos semelhantes ao groud truth. Além disso, uma vez que o método proposto tem uma menor dispersão do erro de segmentação, ele se mostra adaptável a diferentes tipos de textura.

Uma ilustração dos resultados obtidos pode ser visto na Figura 6.4. O nosso método tende a proporcionar uma segmentação mais precisa, enquanto os outros métodos tendem a sobre-segmentar texturas complexas, tais como folhas, flores e pedras com elevada variação de luminância. No entanto, quando texturas semelhantes são colocados lado-a-lado, situação em que o resultado mais comum é a sub-segmentação da imagem, o método proposto também apresenta resultados melhores do que os demais métodos.

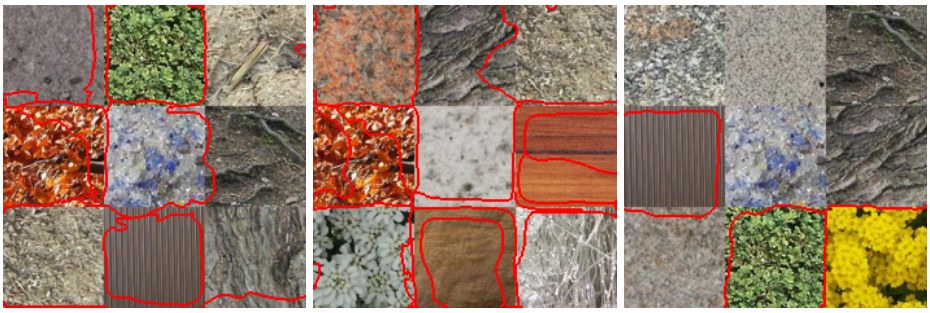
Para o banco de imagens *BSDS300*, que contém 100 imagens naturais, mostramos a comparação do desempenho geral na Tabela 6.2, com a mesma configuração de W_G e Q descrita anteriormente. Para imagens naturais a qualidade de uma segmentação é muito subjetiva, não somente pela falta de critérios de comparação (como acontece nas imagens sintéticas), mas também porque podem existir várias possíveis segmentações corretas, dependendo do indivíduo e dos critérios de avaliação utilizados. Logo, uma medida quantitativa pode não representar totalmente a eficácia da segmentação. Como o conjunto de dados *BSDS300* fornece um conjunto de segmentações feitas a mão para cada imagem de teste, foram comparados os resultados da segmentação com todos os *ground truths* fornecidos para cada imagem e escolhemos a mais similar segundo a medida de precisão proposta (veja Equação (6.2)) e, então, calculamos a acurácia média e desvio padrão de erro para todas as imagens no banco de imagens.

Os resultados quantitativos mostram que o método de segmentação de texturas proposto neste trabalho tem um desempenho significativamente melhor do que os métodos selecionados do estado-da-arte. Embora a precisão nas imagens naturais tendem a ser mais baixas do que no caso das imagens sintéticas, observamos que a queda da precisão ocorreu igualmente em todos os métodos testados. Ainda assim, o SMTR alcança uma maior precisão em comparação com os outros métodos, com um aumento muito pequeno do desvio padrão do erro.

Alguns resultados visuais para o banco de imagens *BSDS300* podem ser visualizados na Figura 6.5. Tal como no caso de imagens sintéticas, a comparação visual sugere que o método proposto tende a proporcionar melhores resultados do que os métodos de segmentação de cor e textura selecionados do estado-da-arte. Em alguns casos, outros

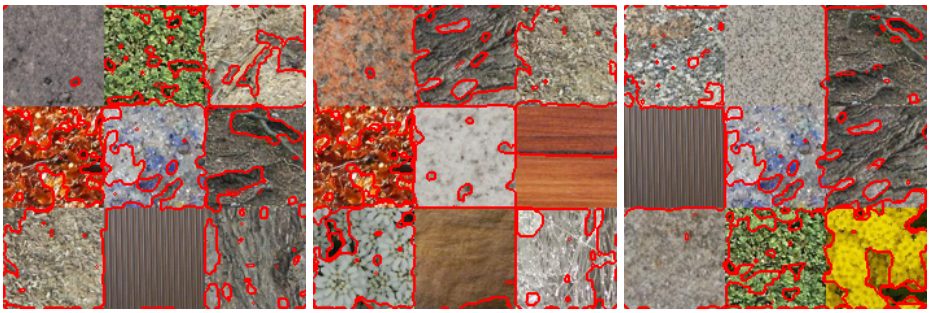


(a)

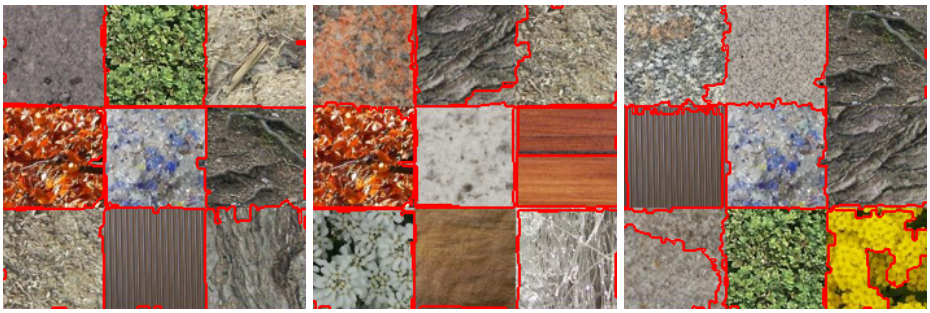


(b)

(c)

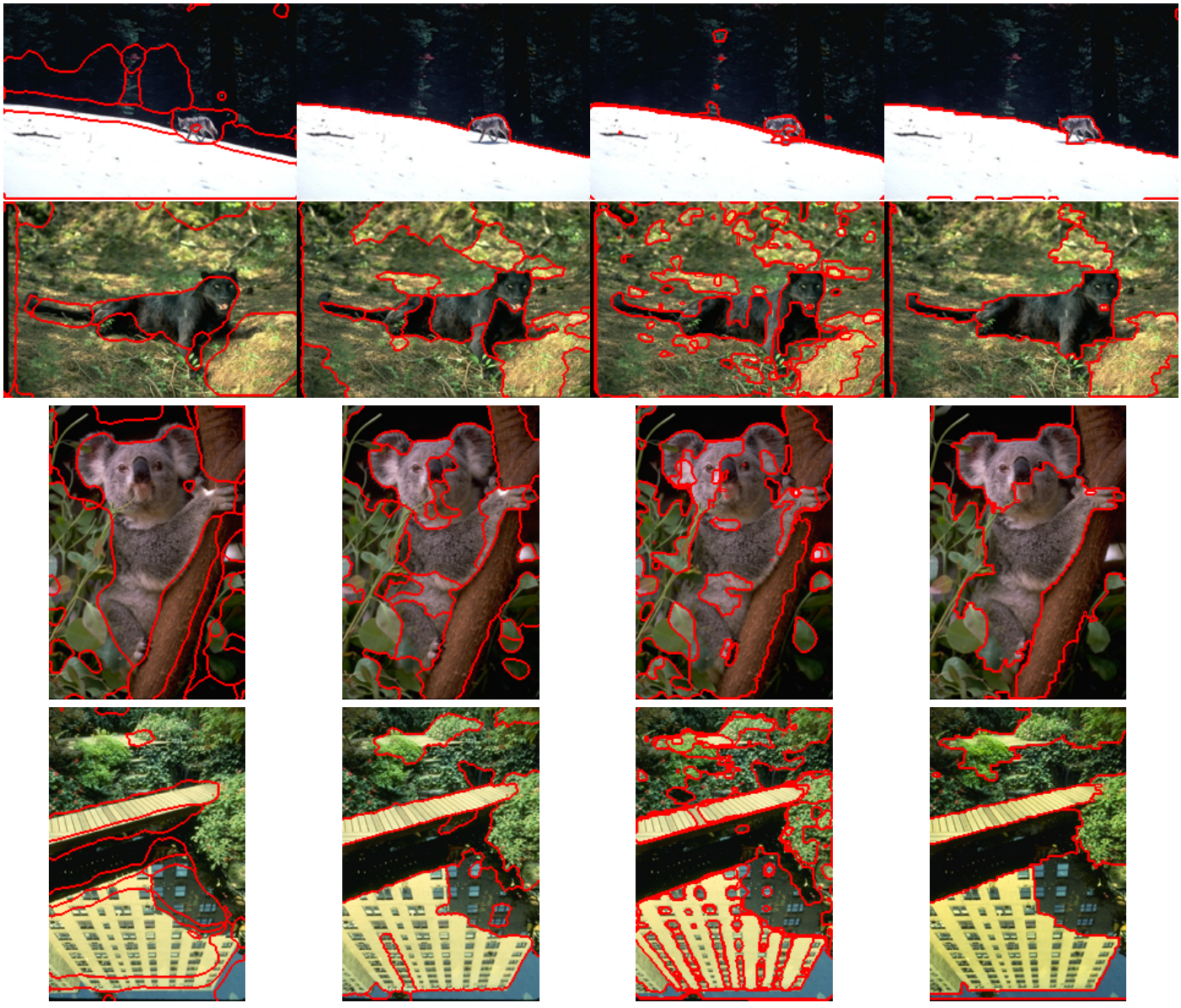


(d)



(e)

Figura 6.4: Comparação entre o método de segmentação proposto contra os métodos do estado-da-arte para algumas imagens no banco de imagens *Prague*. (6.4a) CPGS; (6.4b) CTMS; (6.4c) JSEG; (6.4d) SRM; (6.4e) método proposto.



(a) (b) (c) (d)

Figura 6.5: Comparação entre a técnica de segmentação proposta contra os métodos do estado-da-arte para algumas imagens no banco de imagens *BSDS300*. (6.5a) CTMS; (6.5b) JSEG; (6.5c) SRM; (6.5d) método proposto.

métodos tendem a sobre-segmentar texturas com variações suaves de cores e luminância ou ainda que são altamente texturizadas. Por exemplo, o método CTMS tende a segmentar incorretamente regiões próximas a bordas ou pequenas regiões, já o SRM e o JSEG tendem a gerar sobre-segmentação, devido a variações locais dos gradientes de imagem. O método proposto, por outro lado, lida de forma eficiente com a maioria das texturas do nosso conjunto de teste.

Esta capacidade de identificar de forma robusta as variações intra-texturais pode ser atribuída a algumas características do nosso método. Na etapa de extração das feições de textura, a combinação das representações multi-escala com com a extração estocástica dos pedaços de imagem para obter as características textura e a abordagem de dicionário textons para representar as texturas no interior de cada região fornecem representações consistentes das texturas (que se tornam mais precisas conforme as regiões de textura crescem em tamanho). Além disso, a fusão estocástica de regiões de textura evita ótimos locais (mínimos ou máximos), e ordenar a fila de prioridade de acordo com as diferenças regionais (gradiente da decomposição multi-escala da imagem) faz com que o processo de fusão seja mais confiável, especialmente para regiões menores.

Em resumo, a vantagem principal da segmentação proposta sobre outros métodos, tais como o CTMS, JSEG e CPG, é um aumento da robustez a variações locais no interior das regiões de textura. Comparado ao SRM, o uso de texturas de cores ao invés de apenas cores demonstrou alcançar uma representação mais adequada para as feições da imagem, evitando sobre-segmentação das texturas, além de ser uma melhor maneira de identificar as variações locais de gradiente dentro das regiões de textura.

6.2 Avaliação do método de Segmentação de Pele

Para avaliar o desempenho do nosso método de detecção de pele, realizamos experimentos com o banco de imagens de imagens SDC (SCHMUGGE et al., 2007). Este banco de imagens é publicamente disponível e consiste de 34 imagens de 320×240 *pixels*, cada uma acompanhada por seu respectivo *ground truth* de detecção de pele, que indica quais *pixels* pertencem a regiões de pele. Essas imagens são projetadas especificamente para avaliar técnicas de segmentação de pele para reconhecimento de gestos e foram adquiridas de múltiplos cenários, tanto de ambientes de interior quanto ao ar livre, sob variados fatores de iluminação, posse da mão e tons de pele (etnia do usuário).

Em nossos experimentos, o modelo de cor de pele $\rho(r, b)$ é construído como uma mistura de $\eta = 4$ gaussianas. Para extrair as feições de textura STR, tanto no treinamento do modelo quanto nas imagens de teste, a representação multi-escala da imagem utiliza $N = 7$ níveis de decomposição e os pedaços da imagem são extraídas com largura $W = 5$ *pixels* em torno de cada *pixel*. Cada pedaço é então reduzida pela projeção aleatória para um vetor de $M = 10$ elementos. Esses parâmetros foram determinados com base nos experimentos relatados na seção anterior.

Como discutido no capítulo anterior, o tamanho do dicionário de textons depende do número de classes (texturas) a serem reconhecidas. Uma representação precisa das texturas geralmente necessita de um número grande de textons $K_d = K + K'$ no dicionário (GRAF; LUSCHGY, 2000), especialmente quando não temos conhecimento prévio do número de texturas na imagem. Quando todos os materiais são conhecidos, o dicionário pode ser obtido com poucos protótipos ($K = 10$) de cada classe (VARMA; ZISSERMAN, 2009; LIU; FIEGUTH, 2012). Contudo, conforme K' aumenta, enfrentamos maior complexidade da técnica de segmentação, visto que no método deste trabalho apenas uma

Tabela 6.3: Desempenho da detecção de pele no banco de imagens SDC

Method	TDR	TDR_σ	FDR	FDR_σ
RGB	2.94%	17.15%	100.00%	0.00%
HSV	25.26%	30.00%	98.78%	2.05%
Bayes	58.33%	20.98%	93.28%	8.28%
Método Proposto ($W'_g = [1, 1, 1]$)	84.34%	19.33%	75.46%	13.19%

classe é conhecida (pele).

Nos experimentos descritos na Seção 6.1 determinamos que $K' = 30$ textons permite um dicionário de representatividade satisfatória para as texturas da imagem a um custo de computação aceitável. Logo, para permitir uma distribuição justa entre as classes da imagem, estabelecemos que o dicionário de textons terá 30 protótipos de feições por classe, isto é, haverá $K = 30$ textons de pele, obtidos na etapa de treinamento e $K' = 30$ textons de fundo, obtidos da própria imagem de teste.

Como mencionado anteriormente, existem dois parâmetros que controlam a etapa de fusão estocástica de regiões: (a) o vetor de pesos W_g, e (b) o termo de regularização Q. O vector de pesos W_g determina o balanço entre cor e luminância na medida de dissimilaridade. Já o termo de regularização Q, controla o número final de segmentos. Na medida em que esse valor aumenta, mais segmentos são encontrados devido ao seu efeito na probabilidade de fusão α. Ambos os parâmetros são discutidos de forma mais detalhada na Seção 6.1. Tendo em vista os fatos acima observados, na etapa de fusão estocástica das regiões, definimos o termo de regularização como $Q = 400$ e o vetor de pesos como $W_g = [2, 1, 1]$.

Finalmente, na etapa de detecção da pele, o vector pesos da comparação textura é definida como $W'_g = [1, 1, 1]$ e os limiares de similaridade de cor e textura da pele são $\beta_c = 0.54$ e $\beta_T = 0.2$. Novamente, ao exigirmos que ambos os critérios de cor e textura sejam satisfeitos, podemos usar valores mais baixos para esses limiares, sem perda de precisão. Todos os parâmetros do método proposto foram escolhidos experimentalmente e com base na literatura atual de modelagem de cor da pele, extração de características de textura e segmentação de imagem por fusão de regiões.

Para medir o desempenho da técnica de detecção de pele proposta neste trabalho, foi utilizado uma estratégia de validação-cruzada (*cross-validation*) de 10 partições. Dessa forma, as imagens contidas em cada partição são classificadas a partir do modelo treinado com todas as imagens restantes nas outras 9 partições (KOHAVI, 1995). Em seguida, para cada imagem testada, é calculada a taxa de detecção verdadeira (*true detection rate*, TDR) e da taxa de detecção falsa (*false detection rate*, FDR), definidos como

$$TDR = \frac{VP}{VP + FP}, \qquad FDR = \frac{VN}{VN + FN} \tag{6.3}$$

onde VP e FP indicam a quantidade de verdadeiros e falsos positivos, respectivamente, e VN e FN indicam a quantidade de verdadeiros e falsos negativos.

Como uma estimativa de desempenho geral para o conjunto de dados inteiro, usamos a média e desvio padrão dessas medições de qualidade. Como uma estimativa da desempenho global para todo o banco de imagens, usamos a média e desvio padrão dessas medidas de avaliação de qualidade.

Figura 6.6: Comparação visual do método proposto de detecção de pele. (6.6a) imagem original, (6.6b) limiarização HSV (DARDAS; GEORGANAS, 2011), (6.6c) classificador de Bayes (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007), (6.6d) método proposto, (6.6e) *ground truth*.

Neste trabalho, comparamos os nossos resultados com outros métodos do estado da arte, tais como limiares no RGB (KOVAC; PEER; SOLINA, 2003) e HSV (DARDAS; GEORGANAS, 2011) e o classificador de Bayes (KAKUMANU; MAKROGIANNIS; BOURBAKIS, 2007). Os resultados comparativos são mostrados na Tabela 6.3, em termos de TDR e FDT (média e desvio padrão), e indicam que o método de segmentação de pele proposto leva a melhores resultados quando comparado com os métodos do estado-da-arte, isto é, valores mais elevados de TDR e valores mais baixos de FDR.

A Figura 6.6 apresenta uma comparação visual dos resultados obtidos. O método proposto demonstra ser mais sensível às variações locais no interior dos segmentos de textura, e a combinação de textura e cores parece representar melhor as características da pele. Isso permite uma maior poder discriminação das regiões de pele e fundo, tratando eficientemente as variações de locais no interior das regiões de textura, e usar essas informações para identificar as classes pele e fundo nos segmentos da imagem. Podemos atribuir essa robustez, não somente à representação e modelagem estocástica multi-escala de texturas, mas também ao emprego da fusão estocástica de regiões, que tornam a segmentação mais confiável, conforme discutido em mais detalhes na Seção 6.1.

Além disso, a etapa de segmentação da texturas proporciona uma extração mais precisa dos limites das regiões da pele, tornando-se uma grande vantagem quando se aplica a técnica aqui descrita em sistemas de reconhecimento de gesto, em que uma identificação precisa da forma da mão permite uma utilização mais segura dos descritores de forma para classificação e análise de gesto.

7 CONCLUSÕES

Este trabalho objetivou a elaboração de uma técnica de segmentação de regiões de pele adaptada a sistemas de reconhecimento de gestos de mão. Para tal, o algoritmo desenvolvido deveria ser robusto às possíveis variações de iluminação e tom de pele do usuário. A técnica proposta também é ser capaz de distinguir entre pele e outros materiais similares como, por exemplo, areia, madeira, chamas ou pelo amarelo. Outrossim, a detecção de pele é capaz de identificar corretamente os limites das regiões de pele, delimitando com precisão sua forma externa.

Tendo em vista este objetivo principal, foram elaborados três objetivos específicos, que em conjunto constituem a solução proposta para o problema. A fim de obter uma melhor diferenciação entre as regiões de pele e os demais materiais na imagem, o primeiro marco é elaborar uma representação de texturas presentes na imagem. Para garantir que os limites das regiões de pele sejam corretamente identificados (i.e. extração correta de formato), o segundo marco neste trabalho é desenvolver um algoritmo de segmentação de imagens que identifique as fronteiras entre os diferentes materiais (texturas) nela contidos. Finalmente, para que a correta identificação de todos os segmentos encontrados na imagem, o terceiro e último marco deste trabalho consiste em combinar a representação de texturas com uma representação de cor, que juntos modelem corretamente as características de aparência da pele, independentemente da etnia do usuário, pose da mão, fonte ou posição de iluminação da cena e demais objetos na imagem.

Para efetuar a descrição de texturas, este trabalho propôs uma representação estocástica multi-escala das imagens. Para tal, primeiramente construímos uma representação multi-escada da imagem que, inspirado no conceito de espaço-escala de filtros bilaterais, descreva a imagem em vários níveis de detalhamento. Em seguida, extraímos da imagem os vetores de feições de textura na forma de pedaços de imagem estocásticas ordenadas. A partir desses vetores, utilizamos a abordagem de dicionário de textons para descrever a aparência das texturas. Para manter a discriminação da regiões de pele, usamos protótipos das feições de pele, enquanto o restante dos textons é obtido da própria imagem de teste. Além disso, armazenamos modelos de textura de pele relativos a todas as imagens utilizadas no treinamento, para posterior comparação das imagens de teste.

Já para segmentar a imagem, utilizamos uma estratégia de fusão de regiões a partir de um critério estocástico, que por sua vez é baseado na dissimilaridade das texturas (descritas pelo modelo proposto anteriormente) e tamanho das regiões. Cada *pixel* constitui inicialmente uma região, as quais são unidas sucessivamente até que seja obtida a segmentação final da imagem. A ordem dessas fusão é determinada pelo gradiente da representação multi-escala da imagem.

Uma vez obtida a segmentação da imagem, classificamos cada região individualmente baseado em critérios de similaridade de cor e textura. A textura é modelada pela descrição

de aparência descrita anteriormente e a similaridade com esse modelo é calculada a partir da Equação (5.11) (dissimilaridade de texturas). Já para modelar a cor de pele, utilizamos uma GMM ρ, logo calculamos a similaridade com esse modelo pela probabilidade da cor média da região corresponder a um tom de pele.

Para avaliar o desempenho da técnica proposta, realizamos experimentos com o banco de imagens SDC. Para mensurar a qualidade dos resultados obtidos empregamos uma estratégia de validação cruzada em 10-partições. Os resultados experimentais obtidos com esse banco de imagens mostraram que o método proposto realiza uma identificação mais precisa das regiões da pele do que os outros métodos disponíveis na literatura.

Visualmente, a técnica proposta neste trabalho mostrou-se mais resistente às variações locais da imagem de entrada, além de ser capaz de discriminar pele de outros materiais, mesmo se eles são semelhantes em cor. Além disso, a estratégia de segmentação de imagem via a STM produz uma delimitação precisa das fronteiras das texturas de pele na imagem. Outrossim, as regiões da pele detectadas fornecem não somente o seu posicionamento espacial, mas também permitem uma forte estimativa da sua forma, o qual é um recurso muito desejável em uma técnica de segmentação de pele quando empregada em sistemas de reconhecimento de gestos.

Analogamente, também testamos e avaliamos o desempenho da segmentação estocástica de texturas proposta nesse trabalho, como uma ferramenta genérica de segmentação de imagens. Para isso, propomos uma medida quantitativa de erro, que quantifica a similaridade de um mapa de regiões com o *ground truth*. Utilizando essa medida, mensuramos a qualidade da segmentação nos bancos de imagens Prague, composto de imagens sintéticas, e BSDS300, composto de imagens naturais. Além disso, também comparamos esses resultados com alguns métodos selecionados do estado da arte em segmentação de texturas e fusão de regiões, indicando que o método desenvolvido neste trabalho é mais preciso, segundo a medida utilizada.

A comparação visual com os outros métodos testados também demostrou uma facilidade maior em identificar as variações de cor e luminância contidas no interior das regiões de texturas, enquanto isso, os outros métodos tendem a sobre-segmentar esse tipo de região. Adicionalmente, a técnica proposta também diferencia essas variações intra-texturais das verdadeiras fronteiras entre as regiões adjacentes, mesmo que similares. Outros métodos testados, no entanto, são visivelmente mais propensos à sub-segmentar esse tipo de imagem, unindo regiões de texturas distintas.

7.1 Trabalhos Futuros

Como trabalhos futuros, investigaremos a eficácia da técnica proposta em sistemas de reconhecimento de gestos utilizando sequências de vídeo. A detecção da pele será utilizada como uma etapa de pré-processamento do algoritmo de classificação para identificar e interpretar o gesto realizado. Para isso, pretendemos investigar o uso de outros modelos de cor. Visto que a cor média da região é uma estimativa estatística simples, podem haver outras abordagens que apresentem maior riqueza e detalhamento das informações cromáticas contidas na região que possam descrever a cor da região de uma maneira mais rica e detalhada.

Também sugerimos que o método de segmentação utilizado aqui, grande responsável pela elevada precisão da forma externa das regiões de pele encontradas, seja modificado a fim de incorporar informações temporais. Algoritmos de segmentação de imagens tradicionalmente possuem complexidade computacional proporcional à qualidade de seus

resultados. Logo, é pouco comum que métodos mais eficazes, como o proposto neste trabalho, possam ser executados em tempo real, pelo menos sem otimizações de baixo nível. Por esse motivo, modificar o método para usufruir da informação temporal disponível em vídeos (sequências de imagens) adquire grande importância em trabalhos futuros.

Outro aspecto que pode ser investigado é o uso de métodos de aprendizados de máquina mais elaborados para realizar a classificação dos segmentos. Embora o simples uso de regras de limiarização tenha demonstrado relativa eficiência, ainda é possível que resultados mais confiáveis sejam obtidos ao empregar técnicas mais complexas. Deve-se manter em mente, contudo, que na metodologia desenvolvida neste trabalho, o treinamento dos modelos requer apenas exemplos de pele e essa característica deve permanecer independente do classificador utilizado.

Finalmente, para verificar o real desempenho da técnica de segmentação de pele apresentada neste trabalho, pretendemos empregá-lo em um sistema de reconhecimento e interpretação de gestos de mão. Apenas ao realizar testes práticos, poderemos demonstrar a relevância da extração de forma atingida neste trabalho para identificação de gestos.

Além da informação de forma da região, a modelagem de texturas proposta neste trabalho é uma feição que também pode ser útil nessa tarefa (reconhecimento de gestos) (MITRA; ACHARYA, 2007). Assim sendo, outra possibilidade a ser explorada é verificar o desempenho da descrição estocástica que modela a aparência das regiões, que tem o potencial de atingir uma discriminação adequada entre as diferentes poses que a mão humana pode efetuar.

REFERÊNCIAS

ACHLIOPTAS, D. Database-friendly random projections: johnson-lindenstrauss with binary coins. **Journal of Computer and System Sciences**, [S.l.], v.66, n.4, p.671 – 687, 2003.

AHERNE, F. J.; THACKER, N. A.; ROCKETT, P. I. The Bhattacharyya metric as an absolute similarity measure for frequency coded data. **Kybernetika**, [S.l.], v.32, p.1–7, 1998.

BISHOP, C. M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

BONET, J. S. D.; VIOLA, P. Texture Recognition Using a Non-parametric Multi-Scale Statistical Model. In: IN PROC. IEEE COMPUTER VISION AND PATTERN RECOGNITION. **Anais...** [S.l.: s.n.], 1998. p.641–647.

BURT, P.; ADELSON, E. The Laplacian Pyramid as a Compact Image Code. **Communications, IEEE Transactions on**, [S.l.], v.31, n.4, p.532 – 540, apr 1983.

CHEN, C. et al. **Handbook of Pattern Recognition and Computer Vision**. [S.l.]: World Scientific, 1999.

CROSS, G. R.; JAIN, A. K. Markov Random Field Texture Models. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.PAMI-5, n.1, p.25 –39, jan. 1983.

DARDAS, N.; GEORGANAS, N. Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques. **Instrumentation and Measurement, IEEE Transactions on**, [S.l.], v.60, n.11, p.3592 –3607, nov. 2011.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B**, [S.l.], v.39, n.1, p.1–38, 1977.

DENG, Y.; MANJUNATH, B.; SHIN, H. Color image segmentation. In: COMPUTER VISION AND PATTERN RECOGNITION, 1999. IEEE COMPUTER SOCIETY CONFERENCE ON. **Anais...** [S.l.: s.n.], 1999. v.2, p.2 vol. (xxiii+637+663).

DO, M.; VETTERLI, M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. **Image Processing, IEEE Transactions on**, [S.l.], v.11, n.2, p.146 –158, feb 2002.

EL-SAWAH, A.; GEORGANAS, N.; PETRIU, E. A Prototype for 3-D Hand Tracking and Posture Estimation. **Instrumentation and Measurement, IEEE Transactions on**, [S.l.], v.57, n.8, p.1627 –1636, aug. 2008.

FAUGERAS, O. D.; PRATT, W. K. Decorrelation Methods of Texture Feature Extraction. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.PAMI-2, n.4, p.323 –332, july 1980.

FOWLKES, C.; MARTIN, D.; MALIK, J. Learning affinity functions for image segmentation: combining patch-based and gradient-based approaches. In: COMPUTER VISION AND PATTERN RECOGNITION, 2003. PROCEEDINGS. 2003 IEEE COMPUTER SOCIETY CONFERENCE ON. **Anais...** [S.l.: s.n.], 2003. v.2, p.II – 54–61 vol.2.

FRANCOS, J.; MEIRI, A.; PORAT, B. A unified texture model based on a 2-D Wold-like decomposition. **Signal Processing, IEEE Transactions on**, [S.l.], v.41, n.8, p.2665 –2678, aug 1993.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.

GRAF, S.; LUSCHGY, H. **Foundations of Quantization for Probability Distributions**. [S.l.]: Springer Berlin Heidelberg, 2000. v.1730.

HAINDL, M.; MIKEŠ, S. Texture Segmentation Benchmark. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, ICPR 2008, 19., Los Alamitos. **Proceedings...** IEEE Computer Society, 2008. p.1–4.

HAN, J.; AWAD, G.; SUTHERLAND, A. Automatic skin segmentation and tracking in sign language recognition. **Computer Vision, IET**, [S.l.], v.3, n.1, p.24 –35, march 2009.

HARMER, P. K. et al. Using Differential Evolution to Optimize 'Learning From Signals' and Enhance Network Security. In: ANNUAL CONFERENCE ON GENETIC AND EVOLUTIONARY COMPUTATION, 13., New York, NY, USA. **Proceedings...** ACM, 2011. p.1811–1818. (GECCO '11).

HOANG, M. A.; GEUSEBROEK, J. M.; SMEULDERS, A. W. M. Color Texture Measurement and Segmentation. **Signal Processing**, [S.l.], v.85, n.2, p.265–275, 2005.

JAHNE, B.; HAUSSECKER, H.; GEISSLER, P. **Handbook of Computer Vision and Applications. Volume 1. Sensors and Imaging**. [S.l.]: Academic Press, 1999.

JAIN, A.; DUIN, R.; MAO, J. Statistical pattern recognition: a review. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.22, n.1, p.4 –37, jan 2000.

JAIN, A. et al. Landscape of clustering algorithms. In: PATTERN RECOGNITION, 2004. ICPR 2004. PROCEEDINGS OF THE 17TH INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2004. v.1, p.260 – 263 Vol.1.

JAIN, A.; FARROKHNIA, F. Unsupervised texture segmentation using Gabor filters. In: SYSTEMS, MAN AND CYBERNETICS, 1990. CONFERENCE PROCEEDINGS., IEEE INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 1990. p.14 –19.

JAIN, A. K.; DUBES, R. C. **Algorithms for clustering data**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.

JIANG, Z.; YAO, M.; JIANG, W. Skin Detection Using Color, Texture and Space Information. **Fuzzy Systems and Knowledge Discovery, Fourth International Conference on**, Los Alamitos, CA, USA, v.3, p.366–370, 2007.

KAKUMANU, P.; MAKROGIANNIS, S.; BOURBAKIS, N. A survey of skin-color modeling and detection methods. **Pattern Recognition**, [S.l.], v.40, n.3, p.1106 – 1122, 2007.

KAUFMAN, L.; ROUSSEEUW, P. J. **Finding groups in data**: an introduction to cluster analysis. New York: John Wiley and Sons, 1990.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: ARTIFICIAL INTELLIGENCE - VOLUME 2, 14., San Francisco, CA, USA. **Proceedings...** Morgan Kaufmann Publishers Inc., 1995. p.1137–1143. (IJCAI'95).

KOSMIDOU, V.; PETRANTONAKIS, P.; HADJILEONTIADIS, L. J. Enhanced Sign Language Recognition Using Weighted Intrinsic-Mode Entropy and Signer's Level of Deafness. **IEEE Transactions on Systems, Man, and Cybernetics, Part B**, [S.l.], v.41, n.6, p.1531–1543, 2011.

KOVAC, J.; PEER, P.; SOLINA, F. Human skin color clustering for face detection. In: EUROCON 2003. COMPUTER AS A TOOL. THE IEEE REGION 8. **Anais...** [S.l.: s.n.], 2003. v.2, p.144 – 148 vol.2.

LEUNG, T.; MALIK, J. Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons. **Int. J. Comput. Vision**, Hingham, MA, USA, v.43, n.1, p.29–44, June 2001.

LI, S. Z. **Markov Random Field Modeling in Image Analysis**. 3rd.ed. [S.l.]: Springer Publishing Company, Incorporated, 2009.

LIANG, R.-H.; OUHYOUNG, M. A real-time continuous gesture recognition system for sign language. In: AUTOMATIC FACE AND GESTURE RECOGNITION, 1998. PROCEEDINGS. THIRD IEEE INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 1998. p.558 –567.

LIU, L.; FIEGUTH, P. Texture Classification from Random Features. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.34, n.3, p.574 –586, march 2012.

LIU, L.; FIEGUTH, P.; KUANG, G. Combining sorted random features for texture classification. In: IMAGE PROCESSING (ICIP), 2011 18TH IEEE INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2011. p.833 –836.

LIU, L.; FIEGUTH, P.; KUANG, G. Compressed Sensing for Robust Texture Classification. In: KIMMEL, R.; KLETTE, R.; SUGIMOTO, A. (Ed.). **Computer Vision Đ ACCV 2010**. [S.l.]: Springer Berlin / Heidelberg, 2011. p.383–396. (Lecture Notes in Computer Science, v.6492).

LLOYD, S. Least squares quantization in PCM. **Information Theory, IEEE Transactions on**, [S.l.], v.28, n.2, p.129 – 137, mar 1982.

MALLAT, S. A theory for multiresolution signal decomposition: the wavelet representation. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.11, n.7, p.674 –693, jul 1989.

MALLAT, S.; ZHONG, S. Characterization of signals from multiscale edges. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.14, n.7, p.710 – 732, jul 1992.

MARTIN, D. et al. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In: INT'L CONF. COMPUTER VISION, 8. **Proceedings...** [S.l.: s.n.], 2001. v.2, p.416–423.

MARTIN, D.; FOWLKES, C.; MALIK, J. Learning to detect natural image boundaries using local brightness, color, and texture cues. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.26, n.5, p.530 –549, may 2004.

MEDEIROS, R. S.; SCHARCANSKI, J.; WONG, A. Stochastic Region Merging Approach to Image Segmentation using Multi-scale Stochastic Texture Models. **IEEE Transactions on Image Processing**, [S.l.], o 2012. Trabalho não publicado.

MEDEIROS, R. S.; SCHARCANSKI, J.; WONG, A. Natural Scene Segmentation Based on a Stochastic Texture Region Merging Approach. In: ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP), 2013 IEEE INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2013.

MEDEIROS, R. S.; SCHARCANSKI, J.; WONG, A. Multi-scale Stochastic Color Texture Models for Skin Region Segmentation and Gesture Detection. In: MULTIMEDIA AND EXPO (ICME), 2013 IEEE INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2013. Trabalho não publicado.

MITRA, S.; ACHARYA, T. Gesture Recognition: a survey. **Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on**, [S.l.], v.37, n.3, p.311 –324, may 2007.

MORI, G. et al. Recovering human body configurations: combining segmentation and recognition. In: COMPUTER VISION AND PATTERN RECOGNITION, 2004. CVPR 2004. PROCEEDINGS OF THE 2004 IEEE COMPUTER SOCIETY CONFERENCE ON. **Anais...** [S.l.: s.n.], 2004. v.2, p.II–326 – II–333 Vol.2.

NOCK, R.; NIELSEN, F. Statistical region merging. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.26, n.11, p.1452 –1458, nov. 2004.

O'HAGAN, R.; ZELINSKY, A. Visual Gesture Interfaces for Virtual Environments. In: FIRST AUSTRALASIAN USER INTERFACE CONFERENCE, Washington, DC, USA. **Proceedings...** IEEE Computer Society, 2000. p.73–. (AUIC '00).

PAULRAJ, M. et al. A phoneme based sign language recognition system using 2D moment invariant interleaving feature and Neural Network. In: RESEARCH AND DEVELOPMENT (SCORED), 2011 IEEE STUDENT CONFERENCE ON. **Anais...** [S.l.: s.n.], 2011. p.111 –116.

PHUNG, S.; BOUZERDOUM, A.; CHAI, D. Skin segmentation using color and edge information. In: SIGNAL PROCESSING AND ITS APPLICATIONS, 2003. PROCEEDINGS. SEVENTH INTERNATIONAL SYMPOSIUM ON. **Anais...** [S.l.: s.n.], 2003. v.1, p.525 – 528 vol.1.

PORTILLA, J. et al. Image denoising using scale mixtures of Gaussians in the wavelet domain. **Image Processing, IEEE Transactions on**, [S.l.], v.12, n.11, p.1338 – 1351, nov. 2003.

RAUTARAY, S.; AGRAWAL, A. Design of gesture recognition system for dynamic user interface. In: TECHNOLOGY ENHANCED EDUCATION (ICTEE), 2012 IEEE INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2012. p.1 –6.

SCHARCANSKI, J. A Wavelet-Based Approach for Analyzing Industrial Stochastic Textures With Applications. **Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on**, [S.l.], v.37, n.1, p.10 –22, jan. 2007.

SCHMUGGE, S. J. et al. Task-based evaluation of skin detection for communication and perceptual interfaces. **J. Vis. Comun. Image Represent.**, Orlando, FL, USA, v.18, n.6, p.487–495, Dec. 2007.

SONG, E. et al. Boundary Refined Texture Segmentation Based on K-Views and Datagram Methods. In: COMPUTATIONAL INTELLIGENCE IN IMAGE AND SIGNAL PROCESSING, 2007. CIISP 2007. IEEE SYMPOSIUM ON. **Anais...** [S.l.: s.n.], 2007. p.19 –23.

SONKA, M.; HLAVAC, V.; BOYLE, R. **Image Processing, Analysis, and Machine Vision**. [S.l.]: Thomson-Engineering, 2007.

SZELISKI, R. **Computer Vision**: algorithms and applications. 1st Edition..ed. [S.l.]: Springer, 2010.

VAPNIK, V. N. **The nature of statistical learning theory**. New York, NY, USA: Springer-Verlag New York, Inc., 1995.

VARMA, M.; ZISSERMAN, A. A Statistical Approach to Texture Classification from Single Images. **Int. J. Comput. Vision**, Hingham, MA, USA, v.62, n.1-2, p.61–81, Apr. 2005.

VARMA, M.; ZISSERMAN, A. A Statistical Approach to Material Classification Using Image Patch Exemplars. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.31, n.11, p.2032 –2047, nov. 2009.

VERDOOLAEGE, G.; SCHEUNDERS, P. Geodesics on the Manifold of Multivariate Generalized Gaussian Distributions with an Application to Multicomponent Texture Discrimination. **International Journal of Computer Vision**, [S.l.], v.95, p.265–286, 2011.

WANG, X.; ZHANG, X.; YAO, J. Skin color detection under complex background. In: MECHATRONIC SCIENCE, ELECTRIC ENGINEERING AND COMPUTER (MEC), 2011 INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2011. p.1985 –1988.

WEBB, A. R. **Statistical Pattern Recognition, 2nd Edition**. [S.l.]: John Wiley & Sons, 2002.

WHO Fact Sheets. 300.ed. [S.l.]: World Health Organization, 2013.

WONG, A. et al. Intervertebral Disc Segmentation and Volumetric Reconstruction From Peripheral Quantitative Computed Tomography Imaging. **IEEE Trans. Biomed. Engineering**, [S.l.], v.56, n.11, p.2748–2751, 2009.

WONG, A.; SCHARCANSKI, J.; FIEGUTH, P. Automatic Skin Lesion Segmentation via Iterative Stochastic Region Merging. **Information Technology in Biomedicine, IEEE Transactions on**, [S.l.], v.15, n.6, p.929 –936, nov. 2011.

YANG, J.; LU, W.; WAIBEL, A. Skin-Color Modeling and Adaptation. In: THIRD ASIAN CONFERENCE ON COMPUTER VISION-VOLUME II, London, UK, UK. **Proceedings...** Springer-Verlag, 1997. p.687–694. (ACCV '98).

YANG, M. hsuan; AHUJA, N. Gaussian Mixture Model for Human Skin Color and Its Applications in Image and Video Databases. In: ITS APPLICATION IN IMAGE AND VIDEO DATABASES.Ó PROCEEDINGS OF SPIE Õ99 (SAN JOSE CA. **Anais...** [S.l.: s.n.], 1999. p.458–466.

YU, C. et al. Vision-Based Hand Gesture Recognition Using Combinational Features. In: INTELLIGENT INFORMATION HIDING AND MULTIMEDIA SIGNAL PROCESSING (IIH-MSP), 2010 SIXTH INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2010. p.543 –546.

YU, Q.; CLAUSI, D. IRGS: image segmentation using edge penalties and region growing. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.30, n.12, p.2126 –2139, dec. 2008.

ZHENGMING, L.; TONG, Z.; JIN, Z. Skin detection in color images. In: COMPUTER ENGINEERING AND TECHNOLOGY (ICCET), 2010 2ND INTERNATIONAL CONFERENCE ON. **Anais...** [S.l.: s.n.], 2010. v.1, p.V1–156 –V1–159.

ZHOU, Y. et al. Adaptive Sign Language Recognition With Exemplar Extraction and MAP/IVFS. **IEEE Signal Processing Letters**, [S.l.], v.17, p.297–300, Mar. 2010.

APÊNDICE A PUBLICAÇÕES E CONTRIBUIÇÕES

No processo de desenvolvimento do método apresentado neste trabalho, foram produzidos artigos para publicação em eventos e periódicos da área, relatando os resultados atingidos em etapas intermediárias específicas deste trabalho. Nesta sessão, iremos descrever resumidamente o conteúdo destas publicações.

ARTIGOS PUBLICADOS EM CONFERÊNCIAS:

1. *Natural Scene Segmentation Based on a Stochastic Texture Region Merging Approach*. Com: MEDEIROS, R. S.; SCHARCANSKI, J.; WONG, A.. Em: **38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)**.

   (MEDEIROS; SCHARCANSKI; WONG, 2013a) Resumo do trabalho:
   Este artigo apresenta uma abordagem para a segmentação de cenas naturais com base nas características elementares de textura usando uma estratégia de fusão estocástica de regiões. Modelos de textura regionais são construídos a partir de feições estocásticas de textura baseadas em pedaços de imagem utilizando uma abordagem de aprendizado de dicionário de textons. Finalmente, uma estratégia de fusão estocástica de regiões realiza a segmentação de imagens baseada na verossimilhança das regiões de textura. Comparado com outros métodos do estado-da-arte em segmentação de texturas, nossos resultados experimentais sugerem que a nossa abordagem tem o potencial para reagir melhor às regiões altamente texturizadas comumente encontrados em cenas naturais, e também pode ser mais robusto a variações de cores e iluminação.

ARTIGOS SUBMETIDOS PARA PUBLICAÇÃO EM CONFERÊNCIAS:

1. *Multi-scale Stochastic Color Texture Models for Skin Region Segmentation and Gesture Detection*. Com: MEDEIROS, R. S.; SCHARCANSKI, J.; WONG, A.. Em: **2013 IEEE International Conference on Multimedia and Expo (ICME 2013)**. Trabalho não publicado.

   (MEDEIROS; SCHARCANSKI; WONG, 2013a) Resumo do trabalho:
   Detecção de gestos é uma tarefa importante em aplicações de interação humano-computador. Se a mão é detectada com precisão, análise e reconhecimento do gesto de mão torna-se mais simples e confiável. Este trabalho apresenta um

novo método para detecção de pele como uma etapa de pré-processamento para segmentação de gestos (de mão). Primeiro, os modelos de cor e textura da pele são obtidos a partir de um conjunto de treinamento de imagens de pele, onde um modelo de mistura de gaussianas (GMM) e um dicionário de textons são construídos. Em seguida, uma estratégia de fusão estocástica de regiões é usada para segmentar as regiões de textura da imagem, a partir da qual cada segmento é classificado com base nos modelos de cor de pele e textura de pele. Comparado com outras técnicas do estado-da-arte em segmentação de pele, nossos resultados experimentais sugerem que a nossa abordagem pode lidar com variações de cor e iluminação decorrentes de tom de pele e alterações de pose, mantendo sua capacidade de discriminar pele de outros materiais de fundo altamente texturizadas.

ARTIGOS SUBMETIDOS PARA PUBLICAÇÃO EM PERIÓDICOS:

1. *Stochastic Region Merging Approach to Image Segmentation using Multi-scale Stochastic Texture Models*. Com: MEDEIROS, R. S.; SCHARCANSKI, J.; WONG, A.. Em: **IEEE Transactions on Image Processing**. Trabalho não publicado.

   (MEDEIROS; SCHARCANSKI; WONG, 2012) Resumo do trabalho:
   Neste trabalho, um novo método para a segmentação de imagens baseado em modelos de textura estocásticos multi-escala é apresentado. Primeiramente uma representação multi-escala da imagem é construída utilizando uma decomposição iterativa no espaço-escala de filtros bilaterais. Feições locais de textura baseadas em pedaços da imagem são então extraídas por meio de projeções aleatórias para gerar representações estocásticas multi-escala de textura. Um dicionário de textons é então construído utilizando as representações de textura e usado para representar o modelo aparência global textura baseado na probabilidade de ocorrência dos textons. Finalmente, com base nos modelos de textura, um algoritmo de fusão estocástica de regiões é usado para realizar a segmentação de imagens com base na verossimilhança das regiões de textura. Os resultados experimentais, utilizando os bancos de imagens de segmentação texturas Prague e BDSD300 mostraram que o método proposto proporciona um melhor tratamento das variações de cor e luminância, bem como um desempenho consistente na segmentação para imagens com regiões de interesse altamente texturizadas quando comparado a uma série de métodos anteriores.