



Universidade Católica Dom Bosco
Curso de Bacharelado em Engenharia de Computação

**Segmentação de Faces e Mãos utilizando o
Modelo de Mistura de Gaussianas**

Breno de Paula Fernandes

Orientador: Prof. Dr. Hemerson Pistori

*Relatório Final submetido como um dos requisitos
para a obtenção do título de Bacharel em Engenharia
de Computação.*

UCDB - Campo Grande, MS
Novembro/2006

Dedicatória

À minha mãe, Lúcia Helena Paula do Canto, uma mulher batalhadora, por ter me dado força, carinho e tido compreensão, não só durante a minha graduação mas em toda a minha vida.

Ao meu irmão, Ênio de Paula Fernandes, pela sua sincera amizade e pela compreensão nos momentos em que eu mais precisava de sua ajuda.

Agradecimentos

A Deus, que com sua graça me concedeu o dom da vida.

À minha querida avó, Eunice Paula do Canto, pelo carinho e exemplo de vida com seus 88 anos transmitindo esperança, pela sua força, pelo seu sorriso e otimismo de vida. Aos meus tios, tias, primos e primas, pelo afeto e apoio em todos os momentos da minha vida.

A Katiane Arce Cáceres, por todo seu carinho, compreensão, cumplicidade, confiança, apoio e força para a realização do projeto

Ao meu orientador, Hemerson Pistori, pela orientação e incentivo constante desde a iniciação científica, dividindo o seu profundo conhecimento e sua sabedoria para a realização deste projeto. Ao professor Amaury Antônio de Castro Júnior, que sempre me ajudou quando precisava e à todos os professores de Engenharia de Computação que de uma forma ou de outra me conduziram para a finalização de mais uma etapa da minha vida.

Ao meu amigo Lucas de Moura Bueno pelas baladas durante o curso, ao Kleber Padovani de Souza por ter me ajudado na implementação, à todos os amigos e amigas da sala que jamais negaram ajuda e as amizades do curso adquiridas durante a minha graduação.

Resumo

Diversos pesquisadores têm desenvolvido sistemas computacionais que interagem com seres humanos utilizando imagens capturadas pela *webcam* para detectar, segmentar e rastrear faces e/ou mãos. Para isso é necessário utilizar técnicas de segmentação. Uma das técnicas utilizadas e objeto deste projeto é o modelo de mistura de gaussianas (MMG) no qual é possível segmentar mãos e faces de seres humanos com mais precisão, eliminando ruídos. Com o estudo desse modelo foi possível desenvolver um módulo capaz de segmentar mãos e faces em imagens utilizando o modelo de mistura de gaussianas que será incluído no projeto SIGUS e também fazer comparações com outras técnicas já implementadas nesse projeto como subtração de fundo, subtração de fundo adaptativa e modelo gaussiano.

Abstract

Many researchers have developed computational systems that interact with human beings using images captured for webcam of faces and/or hands to detect, to segment and to track. For this it is necessary to use techniques segmentation. One of the used techniques and object of this project it is the gaussian mixture of model (GMM) in which is possible segmented hands and faces of human beings with more precision, eliminating noises. With the study of this model it was possible to develop a module capable to segment hands and faces in images using the gaussian mixture of model that will be enclosed in project SIGUS and also to make comparisons with others implemented techniques already in this project as subtraction of deep, subtraction of deep adaptative and gaussian model.

Conteúdo

1	Introdução	8
2	Fundamentação Teórica	11
2.1	Principais Características da Imagem	11
2.2	Espaços de Cores	11
2.2.1	Formato de Cor RGB	12
2.2.2	Formato de cor HSV	12
2.3	Segmentação	14
2.4	Conceitos Estatísticos	16
2.4.1	Vetor de Médias	16
2.4.2	Matriz de Covariância	17
3	Trabalhos Correlatos	19
4	Modelo de Mistura de Gaussianas (MMG)	25
5	Desenvolvimento	28
6	Considerações Finais	32
	Referências	33
A	Tabelas	36

Lista de Figuras

2.1	Representação gráfica e numérica dos pixels	12
2.2	Exemplos de diferentes luminâncias.	13
2.3	Exemplos de diferentes crominâncias.	13
2.4	Espaço de cor RGB.	14
2.5	Espaço de cor RGB.	14
2.6	Espaço de cor HSV.	15
2.7	Diferentes cores de pele.	15
2.8	Variações da iluminação em ambientes.	16
3.1	(a) Imagem original. (b) Histograma da imagem.	19
3.2	Sistema de reconhecimento do alfabeto de línguas de sinais . .	20
3.3	(a) Imagem original da íris. (b) Resultado final da segmentação.	21
3.4	Seqüência de quadros de movimentos da mão e da face	22
3.5	Resultado do rastreamento utilizando o MMG	22
3.6	Histograma da somas das 3 gaussianas antes do algoritmo EM.	23
3.7	Histograma da somas das 3 gaussianas depois do algoritmo EM.	24
4.1	Gaussiana simples.	26
4.2	Mistura de 2 Gaussianas.	26
5.1	Protótipo <i>Segmentator</i>	28
5.2	Configurador das técnicas do <i>Segmentator</i>	29
5.3	(a) Amostra da face. (b) Amostra da mão.	31
5.4	(a) Histograma da amostra da face. (b) Histograma da amostra da mão.	31

Capítulo 1

Introdução

O processo para detectar um objeto em uma imagem por meio da informação da cor e da escala de cinzas tem sido há muito tempo estudado por diversos autores [GONZALEZ and WOODS 2000]. Para analisar uma imagem digital, um passo importante a ser realizado é a segmentação, que consiste basicamente em dividir a imagem em partes ou objetos constituintes, com o objetivo de abstrair da imagem características que identifiquem os diferentes sinais. Por exemplo, em uma imagem contendo uma mão, face ou ambos, pode ser interessante extrair informações sobre: posição e os ângulos relativos entre as pontas dos dedos, o contorno, a posição e direção do centro da mão ou face e os “momentos da imagem”. Para extrair essas informações, no entanto, é necessário primeiro identificar, na imagem original, a mão ou face. Diversas técnicas de segmentação da mão e face humana utilizam informações relacionadas com a cor e textura da pele.

Durante anos diversos pesquisadores, como [GEJGUS et al. 2004] e [YANG and AHUJA 1999], têm desenvolvido sistemas computacionais que interagem com seres humanos utilizando imagens capturadas pela *webcam* para detectar, segmentar e rastrear faces e/ou mãos. A fase da detecção é um passo importante por que verifica as imagens que contêm faces ou mãos. Essa etapa é um pouco complicada devido às variações presentes nas imagens como, por exemplo, ruídos.

Para que não ocorra muita variação nas imagens as técnicas que utilizam a cor da pele são usadas para obter uma detecção e segmentação mais robusta e rastrear pessoas em imagens dinâmicas. As cores oferecem muitas vantagens em relação às informações geométricas para problemas como oclusão parcial, rotação em profundidade, troca de escala e resolução [MCKENNA et al. 1998]. São utilizados alguns formatos de cores para rastrear as faces e mãos em imagens dinâmicas, como o RGB e o HSV, entre outros, cada um representando informações de maneiras dife-

rentes [CHAI and NGAN 1999].

Como existem dificuldades para segmentar as faces e/ou as mãos são utilizadas técnicas de segmentação junto com os espaços de cores. Para isso, a técnica do modelo de mistura gaussiana tem sido usado para fazer segmentação das faces e/ou das mãos em imagens como citados em McKenna; Raja e Gong [MCKENNA et al. 1998] e Rasmussen [RASMUSSEN 2000]. A vantagem de se utilizar o modelo de mistura de gaussianas está no fato de que esta técnica determina uma função simples que implementa o mapeamento entrada-saída desejado [SIMÕES and COSTA 2002].

A Plataforma de Apoio ao Desenvolvimento de Sistemas para Inclusão Digital de Pessoas com Necessidades Especiais (SIGUS)¹ é um ambiente computacional que tem como objetivo aumentar o número de programas computacionais destinados a essas pessoas [PISTORI et al. 2006]. Já existem diversos aplicativos que já foram desenvolvidos, como: um editor de símbolos alfabéticos da Língua Brasileira de Sinais (LIBRAS) e um jogo da velha, que pode ser acionado por meio de movimentos da face [PISTORI 2003]; um simulador de cadeira de rodas guiadas por movimentos faciais [FERNANDES et al. 2005] e uma bateria virtual [PISTORI 2006].

Esse trabalho tem como objetivo acrescentar um módulo de segmentação de faces e mãos utilizando o modelo de mistura de gaussiana para que seja possível segmentar apenas essas características humanas. Foi feito um levantamento bibliográfico da segmentação de faces e mãos e do modelo de mistura gaussiana para que o módulo de segmentação pudesse ser desenvolvido para o projeto SIGUS. O módulo que está sendo implementado utiliza a linguagem de programação JAVA e os programas: ImageJ², para processamento digital de imagens; Weka³, para aprendizagem automática; e o Java Media Framework ⁴ (JMF), um pacote utilizado para capturar imagens em tempo real. O ImageJ e Weka são programas com a filosofia de *software livre*.

Os experimentos do módulo foram feitos com imagens que contem faces e mãos retiradas do banco de imagens do projeto SIGUS. As imagens foram selecionadas manualmente e após a escolha, o algoritmo desenvolvido segmentou as mãos e faces em cada quadro utilizando o modelo de gaussiana.

O texto está estruturado da seguinte maneira. O capítulo 2 descreve a fundamentação teórica sobre as principais características da imagem, os espaços de cores RGB e HSV, técnicas de segmentação de objetos e conceitos estatísticos como o vetor média e matriz de covariância. O Capítulo 3 mostra os trabalhos correlatos utilizando técnicas de segmentação de objetos. O

¹Disponível em: <http://www.gpec.ucdb.br/sigus>.

²<http://rsb.info.nih.gov/ij/>

³<http://www.cs.waikato.ac.nz/ml/weka/>

⁴<http://java.sun.com/products/java-media/jmf/>

Capítulo 4 dispõe de informações sobre o modelo de mistura gaussiana. O Capítulo 5 retrata a implementação do módulo de segmentação utilizando o modelo de mistura gaussiana. O Capítulo 6 mostra os experimentos do módulo implementado e no capítulo 7 são feitas as considerações finais.

Capítulo 2

Fundamentação Teórica

Neste capítulo serão apresentados uma visão geral sobre as principais características da imagem como os *pixels*, luminância e cromaticidade; os espaços de cores vermelha, verde e azul (RGB - sigla inglesa de *Red*, *Green* e *Blue*) e matiz, saturação e valor (HSV - sigla inglesa de (*Hue*, *Saturation* e *Value*) e, por último, serão explicados os conceitos estatísticos como o vetor de médias e a matriz de covariância utilizados no modelo de mistura de gaussianas (MMG).

2.1 Principais Características da Imagem

As principais características de uma imagem são representadas pelo *pixel*, pela luminância e pela cromaticidade. A primeira característica, o pixel, é um elemento importante e representa numericamente cada ponto de uma imagem. Essa representação é mostrada na Figura 2.1. A luminância está relacionada com o brilho da cor, ou seja, com o quanto ela é mais clara ou mais escura. A Figura 2.2 mostra um exemplo da luminância. Por último, a cromaticidade é a característica que diferencia as cores, por exemplo, o azul do vermelho. A Figura 2.3 mostra a diferença entre duas imagens com diferente cromaticidade.

2.2 Espaços de Cores

Espaços de cores são formas de se representar cores como vetores de um espaço vetorial, normalmente, tridimensional. Operações de mudança de base nesses espaços vetoriais geram diferentes representações para as cores de uma imagem, permitindo assim que determinadas características interessantes, presentes em uma imagem, sejam destacadas. Dois dos espaços de cores mais citados na literatura são o RGB e o HSV.



Figura 2.1: Representação gráfica e numérica dos pixels

2.2.1 Formato de Cor RGB

O espaço de cor RGB é representado por três cores primárias: vermelho, verde e azul. Esse formato é bastante utilizado em processamento de imagens digitais. A Figura 2.4 mostra como é representado o espaço de cor RGB. Cada cor primária possui 256 tonalidades, totalizando mais de 16 milhões de cores possíveis. Pode-se observar na Figura 2.5 que a utilização de valores numéricos para cada cor primária resulta em outras cores.

2.2.2 Formato de cor HSV

No espaço HSV, cada cor possui três componentes: matiz, saturação e luminância. A matiz está relacionada com o comprimento de onda espectral que produz a percepção da cor. A saturação está relacionada com a pureza da cor. Uma cor pura é aquela que tem em sua composição poucas combinações de diferentes cores, são as chamadas cores vivas. Cores esbranquiçadas e acizentadas são cores pouco saturadas. As cores vermelho e rosa, por exemplo, possuem a mesma matiz, mas o rosa é menos saturado que o vermelho, pois combina o vermelho com o branco. A luminância está relacionada com o brilho da cor, ou seja, com o quanto ela é mais clara ou mais escura. O HSV é viável quando se trabalha com cores de pele humana em imagens. A

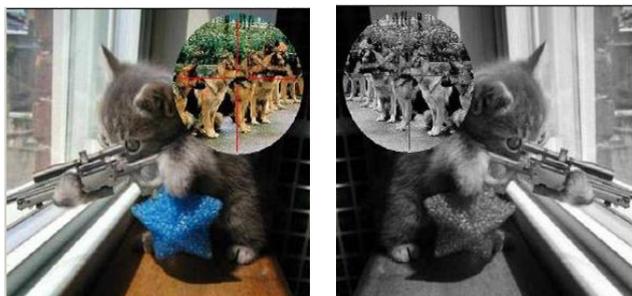


Figura 2.2: Exemplos de diferentes luminâncias.



Figura 2.3: Exemplos de diferentes crominâncias.

Figura 2.6 mostra a representação do formato de cor HSV. A componente H diferencia o azul do vermelho por exemplo. O S representa a saturação e o V representa o brilho da cor.

Nos sistemas computacionais que utilizam os formatos de cores, às vezes, é necessário fazer a conversão dos formatos. A Equação 2.1 mostra a conversão do espaço de cor RGB para HSV.

$$\begin{aligned}
 H &= \arccos \frac{\frac{1}{2}((R - G) + (R - B))}{\sqrt{((R - G)^2 + (R - B) * (G - B))}} & (2.1) \\
 S &= 1 - 3 \frac{\min(R, G, B)}{R + G + B} \\
 V &= \frac{1}{3}(R + G + B)
 \end{aligned}$$

A Tabela 2.2.2 mostra três exemplos de transformações do formato de cor RGB para o formato de cor HSV. A coluna “Componente” mostra as componentes do formato de cor separadamente. As colunas “Exemplo I”, “Exemplo II” e “Exemplo III” mostram os valores dos pixels correspondentes de cada formato de cor.

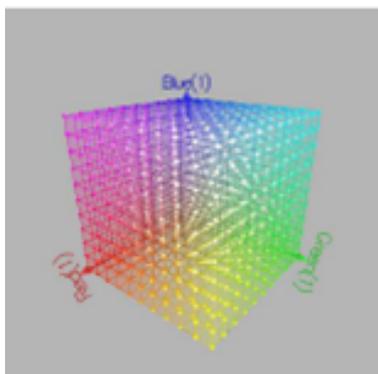


Figura 2.4: Espaço de cor RGB.

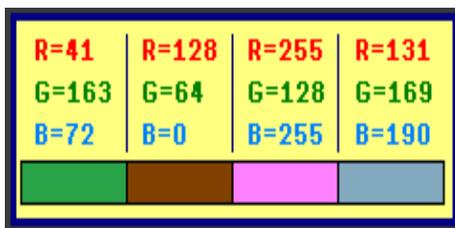


Figura 2.5: Espaço de cor RGB.

Tabela 2.2.2 - Transformações dos formatos de cores separadamente

Componente	Exemplo I	Exemplo II	Exemplo III
R	1	0.5	0
G	0	1	0.5
B	0	0.5	0
H	0	120	240
S	1	1	1
V	1	1	0.5

2.3 Segmentação

Quando se trabalha com análise de imagens um passo importante é a segmentação, que tem como premissa a idéia de segmentar algum objeto dividindo a imagem em partes ou blocos sendo finalizado quando os objetos de interesse tiverem sido separados [GONZALEZ and WOODS 2000]. Porém existem diversos fatores que prejudicam a segmentação como diferentes et-

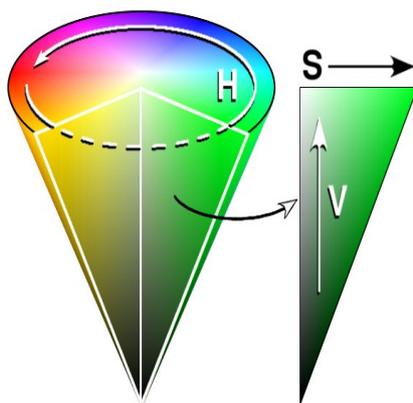


Figura 2.6: Espaço de cor HSV.

nias e variações de iluminação nos ambientes [BEZERRA 2005]. As Figuras 2.7 e 2.8 mostram diferentes tipos de pele e variações de iluminação respectivamente.



Figura 2.7: Diferentes cores de pele.

Existem diversas técnicas para segmentar algum objeto e dependendo da análise obtida dessas técnicas é que poderemos verificar se temos uma boa ou má qualidade na análise das imagens [GONZALEZ and WOODS 2000].

Algumas das técnicas de segmentação que podemos citar são: a subtração do fundo, os modelos de cores, binarização, as redes neurais como os mapas auto-organizáveis (SOM do inglês Self-Organizing Maps) de Kohonen e a Transformada da Distância Euclidiana (TDE),

O SOM é um modelo de redes neurais com características de aprendizagem competitiva e treinamento não supervisionado, sendo capaz de mapear

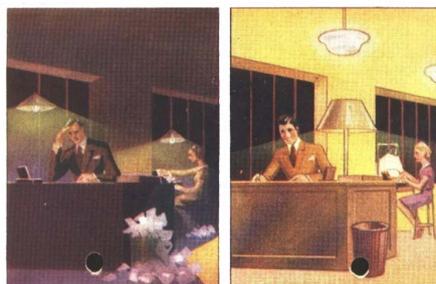


Figura 2.8: Variações da iluminação em ambientes.

um conjunto de dados. O SOM é um dos modelos de redes neurais mais usados na atualidade das redes neurais [VEZHNEVETS et al. 2003].

A Transformada da Distância Euclidiana (TDE) é uma operação básica que se aplica em diversos problemas de processamento de imagens que tem como função converter os pixels de uma imagem digital binária em uma matriz sendo que cada célula possui um valor correspondente à mínima distância do fundo por meio de uma função de distância [COSTA et al. 2005]. A Fórmula da TDE é representada pela Equação 2.2

$$d_e(p, q) = \sqrt{(x - u)^2 + (y - v)^2} \quad (2.2)$$

onde, x e y representam as coordenadas do *pixel* p e u e v representam as coordenadas do *pixel* q .

2.4 Conceitos Estatísticos

Em um estudo sobre Modelo de Mistura Gaussiana são utilizadas informações estatísticas para fornecer dados retirados, por exemplo, de uma imagem. As informações utilizadas são o vetor de médias e a matriz de covariância. A seguir essas medidas serão explicadas separadamente.

2.4.1 Vetor de Médias

O vetor de médias é representado pela Equação 2.3 onde o símbolo μ_i é o vetor de médias, o n é o total de *pixels* de uma região e c_j um vetor de amostras. Para explicar a fórmula, cita-se um exemplo utilizando os formatos de cores. Para cada *pixel* do vetor de amostra são representados pelo seguinte vetor $c_j = \{R_j, G_j, B_j\}$. Em uma situação utilizando as variáveis R e o G do formato de cor RGB têm-se uma média para o R (μ_H) e uma média para

o G (μ_G). Dessa maneira pode-se calcular a média de várias variáveis dos formatos de cores separadamente.

$$\mu_i = \frac{1}{n} \sum_{j=1}^n c_j \quad (2.3)$$

2.4.2 Matriz de Covariância

As propriedades da matriz de covariância são: (a) é simétrica, (b) os valores na diagonal principal são chamados de variância e só pode ter valores positivos e (c) os valores que não são da diagonal são chamados de covariância e podem ter valores positivos ou negativos. A matriz de covariância é representada pela Equação 2.4 onde o símbolo Σ_i é a matriz de covariância, o n representa a quantidade de *pixels*, c_j um vetor de amostras, μ_i o vetor de médias e T significa que a matriz é transposta. Por razões computacionais, em alguns casos assume-se que a matriz de covariância é igual ao valor das variâncias (σ_i^2) das componentes do formato de cor, sendo que elas são independentes como descrito na Equação 2.5.

$$\Sigma_i = \frac{1}{n-1} \sum_{j=1}^n (c_j - \mu)(c_j - \mu)^T \quad (2.4)$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ 0 & \sigma_2^2 & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \sigma_i^2 \end{bmatrix} \quad (2.5)$$

Para calcular a matriz de covariância é necessário ter uma amostra da imagem, por exemplo, uma região de um certo objeto. No exemplo a seguir, X representa os valores de uma região de um certo objeto mostrado na Tabela 2.6. Essa tabela tem cinco observações sendo uma em cada linha com três variáveis cada amostra sendo a primeira o comprimento, a segunda a largura e a terceira a altura, por exemplo. Primeiramente calcula-se o vetor de médias para cada variável sendo representado pela Tabela 2.7. Em seguida, com os valores do objeto e o vetor de médias calculados é possível calcular a matriz de covariância utilizando a Fórmula 2.4. A Tabela 2.8 mostra o resultado da matriz de covariância.

Amostra	Comprimento	Largura	Altura
1	4.0	2.0	0.60
2	4.2	2.1	0.59
3	3.9	2.0	0.58
4	4.3	2.1	0.62
5	4.1	2.2	0.63

(2.6)

$$\mu = [4.10 \quad 2.08 \quad 0.604] \quad (2.7)$$

$$\Sigma = \begin{bmatrix} 0.025 & 0.0075 & 0.00175 \\ 0.0075 & 0.007 & 0.00135 \\ 0.00175 & 0.00135 & 0.00043 \end{bmatrix} \quad (2.8)$$

Sendo assim, o valor 0.025 representa a variância da variável comprimento, 0.0075 o valor da covariância entre o comprimento e a largura, 0.00175 a covariância entre a largura e o altura, 0.007 a variância da largura, 0.00135 a covariância entre a largura e a altura e 0.00043 a variância da variável altura.

Capítulo 3

Trabalhos Correlatos

A modelagem utilizando processos gaussianos para regressão de funções e predição foi testada por [SIMÕES and COSTA 2002] em um ambiente de classificação de cores de *pixels* de imagens monocromáticas visando à segmentação. No total foram feitos 64 exemplos nos quais algumas amostragens foram interpretadas como ruído pelo fato de que a subamostragem de algumas áreas da imagem não obtiveram resultados por causa dos grupos de cores de *pixels* e o desenvolvimento de uma função suave para o tipo dessa imagem. Com isso, na utilização desse processo observou-se que a vantagem está no fato de que uma função simples implementa o mapeamento entrada-saída desejado, ao contrário de redes neurais, porém foram encontrados problemas como: a) a função a ser determinada depende da escolha de uma função covariância adequada na qual não se tem conhecimento dela e b) dificuldade de convergência dos hiperparâmetros que são dependentes da condição inicial. A Figura 3.1a mostra a imagem de uma borracha que não houve qualquer tratamento prévio e a Figura 3.1 mostra o histograma da imagem.

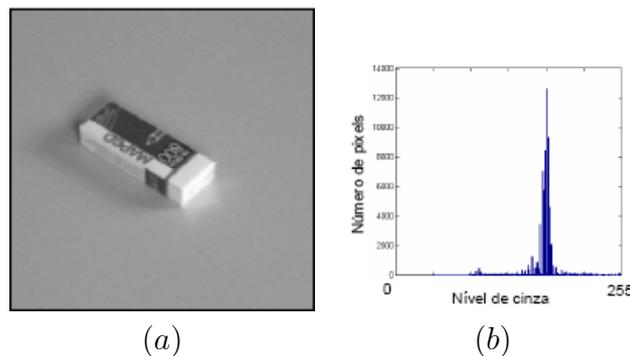


Figura 3.1: (a) Imagem original. (b) Histograma da imagem.

[GEJGUS et al. 2004] utilizaram o Modelo de Mistura Gaussiana para segmentar peles humanas nas imagens em um sistema para ajudar estudantes na aprendizagem do alfabeto de línguas de sinais. Com o modelo, utilizaram um vetor de médias de duas dimensões por causa dos espaços de cores. O sistema captura a imagem da mão por uma câmera CCD, faz o pré-processamento da imagem, detecta os contornos e por último faz o reconhecimento da letra. Para a aprendizagem da letra, o usuário faz o gesto e o sistema faz o treinamento deste. Depois dessas etapas, o usuário faz o gesto da letra e em seguida aparece a letra correspondente no monitor. A Figura 3.2 mostra o sistema de segmentação. Depois de vários testes, a segmentação mostrou bons resultados.



Figura 3.2: Sistema de reconhecimento do alfabeto de línguas de sinais

[JEDYNAK et al. 2003] utilizaram vários modelos estatísticos para a detecção da pele e concluíram que o método da árvore de Bethe conduz a um algoritmo mais eficiente na segmentação da imagem. A comparação entre quatro esquemas da detecção da pele em circunstâncias reais, alterando-se constantemente a iluminação, não obteve bons resultados, segundo relatam [MARTINKAUPPI et al. 2003]. Entretanto, na pesquisa realizada por [HUYNH-THU et al. 2002], nas imagens com fundo complexo e com variação da iluminação utilizando a técnica de limiarização automática e adaptável, observou-se que o método é eficiente na detecção da face humana.

Os resultados preliminares da pesquisa de [TERRILLON et al. 1998] mostram “a eficiência da combinação da segmentação da cor e de momentos invariantes em determinar faces com uma grande variedade de poses contra fundos relativamente complexos”. Brown, Craw e Lewthwaite [BROWN et al. 2001] relataram que introduziram um mapa de organização próprio (SOM - sigla inglesa de Self Organizing Map) como modelo estatístico, que pudesse ser comparado com um histograma convencional e

técnicas de modelo misto. Com isso, eles obtiveram uma exatidão de 94% de tom da pele facial.

Kapur [KAPUR 1997], em sua pesquisa, fez uma combinação da morfologia matemática e limiarização para obter as características que indicassem a presença da face, cuja imagem foi detectada na escala cinza que contém somente áreas detectadas da pele. Segundo Kapur (1997), os resultados dos testes realizados foram confiáveis.

Foi proposto por [COSTA et al. 2005] um método eficiente e rápido para a segmentação da íris por meio da Transformada da Distância Euclidiana. Foram utilizados na pesquisa 756 imagens de 108 olhos sendo que para cada olho foram capturadas 7 imagens. Desse total, 601 imagens foram segmentadas com sucesso que representa 79,5% e o restante das imagens que corresponde a 20,5% (155 imagens) obteve algum erro na segmentação. A Figura 3.3(a) mostra a imagem original da íris e 3.3(b) mostra o resultado final do método da segmentação da íris.

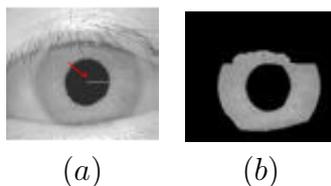


Figura 3.3: (a) Imagem original da íris. (b) Resultado final da segmentação.

Yang e Ahuja [YANG and AHUJA 1999] utilizaram o MMG com duas gaussianas e adicionaram o algoritmo Expectation-Maximization (EM) para obter os parâmetros estimados. Nessa pesquisa foram usadas 2.447 imagens de faces humanas de diferentes grupos de etnias as quais foram convertidas do formato de cor RGB para o formato de cor CIE LUV sendo descartado o valor do brilho. Com isso, obtiveram mais do que 70% dos *pixels* de uma região classificados como cor de pele. A Figura 3.4 mostra uma seqüência de quadros de imagens do movimento da mão e da face antes do rastreamento e a Figura 3.5 mostra o rastreamento de cada quadro da seqüência.

Na pesquisa de Conde [CONDE 2005] foram feitas três análises. A primeira foi a utilização do modelo de mistura gaussiana para definir a cor do objeto. A segunda foi usar esse mesmo modelo para que se adaptasse a iluminação do ambiente e na última foi feita a associação das duas primeiras. Nos experimentos foram usadas quatro gaussianas para modelar a superfície do fundo, e a média de acertos foi de 81,31%. Uma outra análise feita foi segmentar a mão e a face separadamente. Com isso, a taxa média de acertos da mão e da face foram 82,46% e 84,24%, respectivamente.

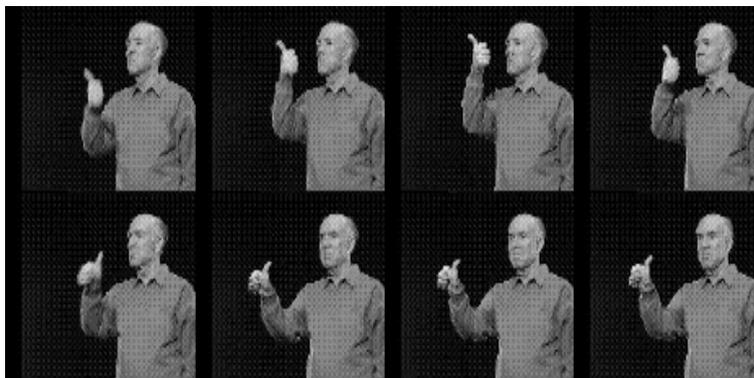


Figura 3.4: Sequência de quadros de movimentos da mão e da face



Figura 3.5: Resultado do rastreamento utilizando o MMG

[JONES and REHG 2002] utilizaram o modelo de mistura de gaussianas para separar as classes consideradas pele das não-pele sendo que foram utilizadas dezesseis gaussianas (kernel) para cada uma das classes e utilizado em paralelo o algoritmo de Expectativa-Maximização (EM) para calcular os valores estatísticos (vetor de médias e matriz de covariância) e o espaço de cor utilizado foi o RGB. As Tabelas 1A e 2A do Anexo A mostram os valores da média, desvio-padrão e o peso das classes considerada pele e não-pele.

Na pesquisa de [POWER and SCHOONEES 2002] foi utilizado o modelo de mistura de gaussianas para segmentação. Os valores utilizados dos parâmetros do modelo foram: a amostra $X = 0, 1, \dots, 255$; a quantidade de gaussianas $k = 3$, sendo uma gaussiana para a superfície da frente e as duas últimas para a de trás; o peso de cada gaussiana $w_k = 0.2, 0.2, 0.6$; o vetor de médias $\mu_k = \{180, 100, 200\}$ e por último a matriz de covariância

$\Sigma_k = \{20, 5, 10\}$ utilizando apenas a variância. Como esse valores não são precisos foi utilizado o algoritmo EM. As Figuras 3.6 e 3.7 mostram o histograma da soma das 3 gaussianas antes e depois da utilização do algoritmo EM, respectivamente.

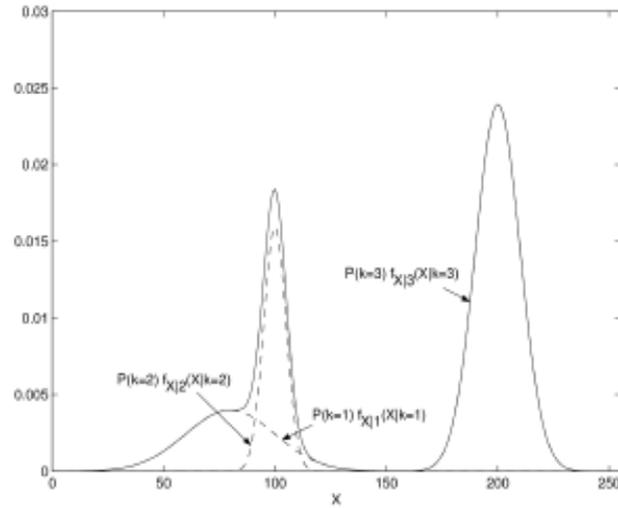


Figura 3.6: Histograma da somas das 3 gaussianas antes do algoritmo EM.

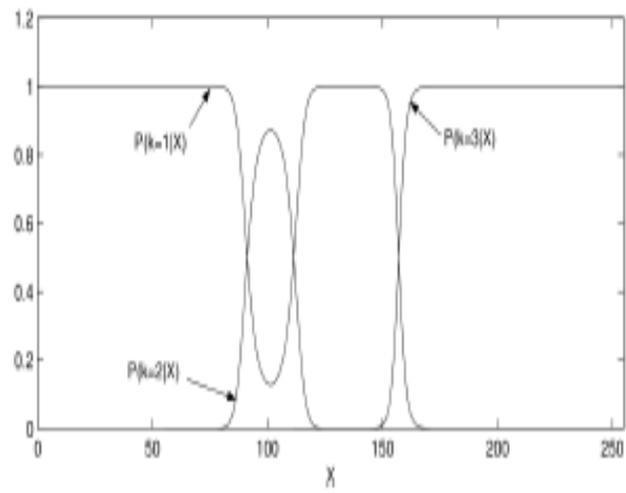


Figura 3.7: Histograma da somas das 3 gaussianas depois do algoritmo EM.

Capítulo 4

Modelo de Mistura de Gaussianas (MMG)

O Modelo de Mistura de Gaussianas é um modelo muito sofisticado capaz de definir formas complexas [VEZHNEVETS et al. 2003]. O MMG tem sido bastante usado para calcular estimativas probabilísticas da cor da pele, cor das roupas e fundos das imagens. Esse modelo pode ser usado para detectar, segmentar e rastrear pessoas, faces e/ou mãos em vídeos [MCKENNA et al. 1998]. A quantidade de componentes, ou seja, o número de gaussianas usados por diferentes pesquisadores, variam significativamente [VEZHNEVETS et al. 2003].

O MMG pode ser utilizado em diversos sistemas usuário-máquina fundamentados em visão computacional que necessitem realizar segmentação de mãos e faces [RIBEIRO and GONZAGA 2006]. A utilização desse modelo ao invés de utilizar a técnica de gaussiana simples é porque a precisão daquele é melhor do que este [YANG and AHUJA 1999]. Por exemplo, se deseja saber quantas pessoas tem pele morena ou pele branca de uma determinada região, com a mistura de gaussianas é possível ter valores multimodal, ou seja, obtêm-se diversos valores que ocorrem com mais frequência sendo possível obter uma maior precisão. Já com a gaussiana simples não teria um resultado satisfatório quando se tem mais de uma característica (ex.: mãos e faces) pelo fato de que teria apenas um valor de pele.

A Fórmula de uma gaussiana simples é representada pela Equação 4.1 onde c é um vetor de cores, k é a probabilidade de ser, por exemplo, a cor da pele, d o número de dimensões que representam os formatos de cores, c_j os valores da amostra e os símbolos μ_k e Σ_k as informações estatísticas que representam o vetor de médias e a matriz de covariância, respectivamente. A Figura 4.1 mostra graficamente como é a representação em uma dimensão.

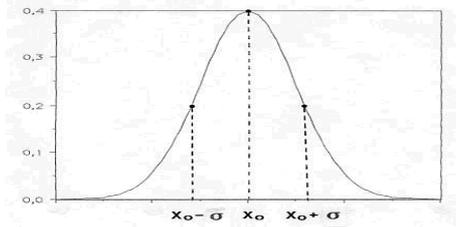


Figura 4.1: Gaussiana simples.

$$p(c|k) = \sqrt{\frac{1}{(2\pi)^d |\Sigma_k|}} \exp \frac{-1}{2} (c_j - \mu_k)^T \Sigma_k^{-1} (c_j - \mu_k) \quad (4.1)$$

A Fórmula 4.2 mostra quando se quer trabalhar com mistura de gaussianas. O símbolo g representa a quantidade de gaussianas, w_i representa o peso de cada gaussiana e a soma delas deve estar de acordo com a equação 4.3 e $p_i(c|k)$ representa a probabilidade normal de cada gaussiana. A Figura 4.2 mostra graficamente duas gaussianas.

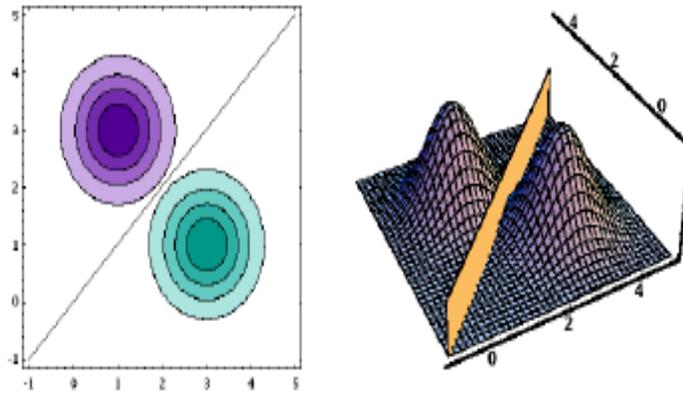


Figura 4.2: Mistura de 2 Gaussianas.

$$p(c|k) = \sum_{i=1}^g w_i * p_i(c|k) \quad (4.2)$$

$$\sum_{i=1}^g w_i = 1; \quad (4.3)$$

Para cada gaussiana é necessário calcular os valores do vetor média e a matriz de covariância, porém na prática os valores verdadeiros deles são desconhecidos. Para resolver esse problema e obter um melhor desempenho pode-se utilizar a técnica do algoritmo de Expectativa-Maximização (EM, sigla inglesa de Expectation-Maximization) no qual é necessário ter o valor de g , ou seja, o número de gaussianas.

O algoritmo EM é utilizado para obter um melhor desempenho das estimativas de parâmetros (μ_i , Σ_i e w_i) calculados pelo modelo de mistura de gaussianas. Esse algoritmo é separado em duas etapas: a primeira é a expectativa (E) que consegue obter um conjunto de estimativas sendo um método iterativo e a segunda etapa é chamada de Maximização (M) que obtêm os valores mais prováveis das estimativas de parâmetros [GEJGUS et al. 2004].

Capítulo 5

Desenvolvimento

Para a desenvolvimento do módulo de segmentação foi utilizado um computador Pentium IV, 2.8 GHz com 256 de memória RAM e uma *webcam* EasyCam Pro. Também estão sendo utilizados o sistema operacional Linux e a linguagem de programação JAVA. Os programas que foram utilizados para a implementação do módulo são: ImageJ, que é um programa de processamento de imagens; o WEKA para aprendizagem automática e o Java Media Framework (JMF) que é um pacote para capturar imagens em tempo real. Estão sendo utilizadas imagens do banco de imagens do projeto SIGUS. Uma ferramenta auxiliar chamada JAMA desenvolvida em JAVA para manipulação de matrizes também foi utilizada.

Foi utilizado um protótipo do projeto SIGUS chamado *Segmentator* no qual é possível capturar as amostras da cor da pele. Nesse protótipo já existem outras técnicas de segmentação que são: a técnica baseado na cor, a técnica de subtração do fundo, a técnica do fundo adaptativo e a técnica de gaussiana. A Figura 5.1 mostra o protótipo *Segmentator*.



Figura 5.1: Protótipo *Segmentator*

O primeiro passo do *Segmentator* é escolher manualmente um vídeo do

banco de imagens. Nesse exemplo foi utilizado um vídeo onde o sinal significava a palavra cérebro. Em seguida, no “Diretório destino” é escolhido um diretório onde armazena a saída da segmentação. Depois é possível escolher as técnicas de segmentação clicando no botão Método e em seguida obter as amostras e por último clicando no botão Tentar para verificar a segmentação.

A implementação do módulo de segmentação de faces e mãos utilizou-se da técnica de mistura de gaussianas. A Figura 5.2 mostra o método dessa técnica, onde o “Número de classes” representa a quantidade de gaussianas, a “Classe atual” representa as classes classificadas como pele (valor 0) ou como não-pele (valor 1). A “Dimensão” é a quantidade de componentes do formato de cor e por último o “Espaço de cor” onde é possível escolher quais componentes utilizar como as componentes HS ou RG, por exemplo.

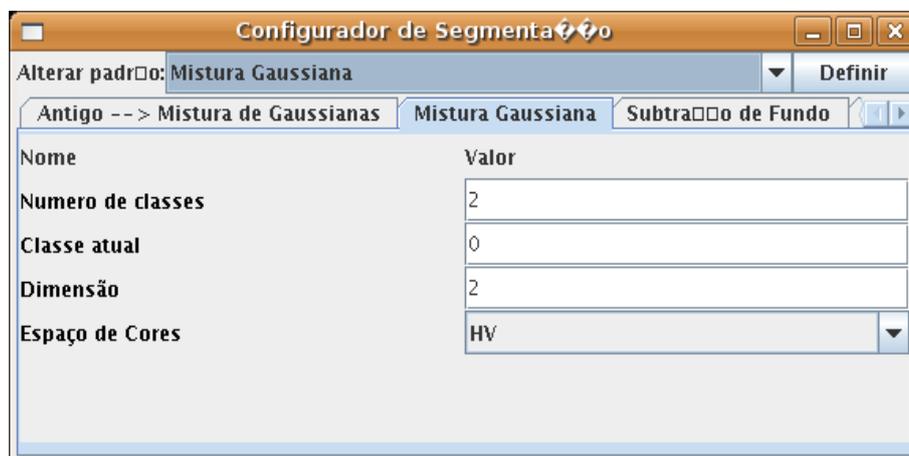


Figura 5.2: Configurador das técnicas do *Segmentator*

Com o modelo implementado no *Segmentator* foi possível extrair as amostras (retângulo amarelo) da cor da pele como mostrado nas Figuras 5.3a, 5.3b sendo que cada amostra representa uma gaussiana. Os histogramas das amostras da face e da mão são mostrados nas Figuras 5.4a e 5.4b. Os pixels extraídos da imagem original estão no formato RGB e foram convertidos para o formato HSV sendo descartado a componente V (valor) que prejudica a segmentação porque representa a luminância.

Em seguida, com os valores das amostras obtidos foi feito um pré-cálculo do vetor de médias apenas das componentes H e S do formato HSV e da matriz covariância. A implementação do vetor de médias e da matriz de covariância são mostrados nos Algoritmos 1 e 2, respectivamente. Esses dois algoritmos já tinham sido implementados.

Algoritmo 1 Vetor de médias

entrada: Quantidade de componentes do formato de cor qtd , total de pixels da amostra n e a matriz de amostras $sample$.

saída: Vetor μ_i contendo a média de cada variável do formato de cor.

```
para  $i = 1$  até  $qtd$  faça
  para  $j = 1$  até  $n$  faça
     $\mu[j] = \mu[j] + sample[j][i]$ 
  fim para
 $\mu[i] = \mu[i]/n$ 
fim para
```

Algoritmo 2 Matriz de covariância

entrada: Vetor de médias μ , matriz da amostra $sample$, total de pixels da amostra n e Quantidade de componentes do formato de cor qtd .

saída: Matriz covariância (Σ).

```
 $double[][] co = newdouble[col][col];$ 
 $Matrix c = newMatrix(co);$ 
 $Matrix temp;$ 
para  $i = 1$  até  $n$  faça
   $temp = newMatrix(linearizar(sample[i]));$ 
   $temp = temp.minus(M);$ 
   $c = c.plus(temp.times(temp.transpose()));$ 
fim para
para  $i = 1$  até  $qtd$  faça
  para  $j = 1$  até  $qtd$  faça
     $co[i][j] = c.get(i, j)/row;$ 
  fim para
fim para
```



Figura 5.3: (a) Amostra da face. (b) Amostra da mão.

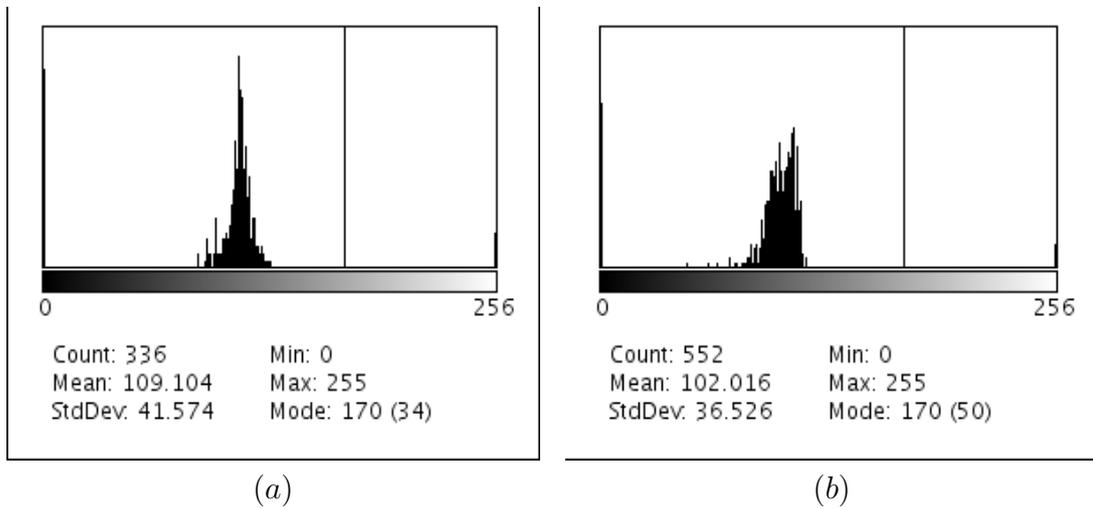


Figura 5.4: (a) Histograma da amostra da face. (b) Histograma da amostra da mão.

Capítulo 6

Considerações Finais

Diversos pesquisadores têm desenvolvido sistemas computacionais que interagem com seres humanos utilizando imagens capturadas pela *webcam* para detectar, segmentar e rastrear faces e/ou mãos. Nesse projeto foi apresentado um levantamento bibliográfico sobre segmentação de faces e mãos e do modelo de mistura de gaussianas. Com isso foi possível implementar um módulo de segmentação de faces e mãos utilizando o modelo de mistura de gaussianas para a plataforma SIGUS, que é um ambiente computacional e tem como objetivo aumentar o número de programas computacionais destinados à pessoas com necessidades especiais. Já existem diversos aplicativos que já foram desenvolvidos na plataforma SIGUS, como: um editor de símbolos alfabéticos da Língua Brasileira de Sinais (LIBRAS) e um jogo da velha, que pode ser acionado por meio de movimentos da face; um simulador de cadeira de rodas guiadas por movimentos faciais e uma bateria virtual.

O módulo utilizou o protótipo *Segmentator* do projeto SIGUS para que fosse possível extrair amostras das características humanas como as faces e mãos. Essas amostras foram retiradas do banco de imagens do projeto SIGUS. Em seguida os pixels foram convertidos do formato RGB para o formato HSV sendo descartado a componente V (valor) que prejudica a segmentação porque representa a luminância. O módulo tem como característica ser *Open Source*, ou seja, poderá ser alterado a qualquer momento para melhorias futuras.

Referências Bibliográficas

- [BEZERRA 2005] BEZERRA, M. A. S. F. (2005). Segmentação de pele em imagens utilizando redes neurais. *I Workshop de Visão Computacional, WVC'2005*.
- [BROWN et al. 2001] BROWN, D., CRAW, I., and LEWTHWAITE, J. (2001). A som based approach to skin detection with application in real time systems.
- [CHAI and NGAN 1999] CHAI, D. and NGAN, K. N. (1999). Face segmentation using skin-color map in videophone applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(4).
- [CONDE 2005] CONDE, F. M. (2005). Segmentação de múltiplos objetos utilizando mistura gaussiana. Graduação. Engenharia de Computação.
- [COSTA et al. 2005] COSTA, R. M. D., RODRIGUES, E. B., and GONZAGA, A. (2005). Um método eficiente de segmentação de íris baseado na transformada da distância euclidiana. *I Workshop de Visão Computacional, WVC'2005*.
- [FERNANDES et al. 2005] FERNANDES, B. d. P., SILVA, V. A. S. d., and PISTORI, H. (2005). Protótipo de um simulador para cadeiras de rodas guiadas por expressões faciais: Estudos preliminares. *3º Congresso Catarinense de Software Livre - SOLISC*.
- [GEJGUS et al. 2004] GEJGUS, P., PLACEK, J., and SPERKA, M. (2004). Skin color segmentation method based on mixture of gaussians and its application in learning system for finger alphabet. In *CompSysTech '04: Proceedings of the 5th international conference on Computer systems and technologies*, pages 1–6, New York, NY, USA. ACM Press.
- [GONZALEZ and WOODS 2000] GONZALEZ, R. C. and WOODS, R. E. (2000). *Digital Image Processing*. Addison-Wesley Publishing Company, INC., Reading, MA.

- [HUYNH-THU et al. 2002] HUYNH-THU, Q., MEGURO, M., and KANEKO, M. (2002). Skin-color extraction in images with complex background and varying illumination. In *WACV '02: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision*, page 280, Washington, DC, USA. IEEE Computer Society.
- [JEDYNAK et al. 2003] JEDYNAK, B., ZHENG, H., and DAOUDI, M. (2003). Statistical models for skin detection.
- [JONES and REHG 2002] JONES, M. J. and REHG, J. M. (2002). Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96.
- [KAPUR 1997] KAPUR, J. P. (1997). Face detection in color images. Washington, DC, USA.
- [MARTINKAUPPI et al. 2003] MARTINKAUPPI, B., SORIANO, M., and PIETIKÄINEN, M. (2003). Detection of skin color under changing illumination: A comparative study. In *ICIAP '03: Proceedings of the 12th International Conference on Image Analysis and Processing*, page 652, Washington, DC, USA. IEEE Computer Society.
- [MCKENNA et al. 1998] MCKENNA, S. J., RAJA, Y., and GONG, S. (1998). Tracking and segmenting people in varying lighting conditions using colour.
- [PISTORI 2003] PISTORI, H. (2003). *Tecnologia Adaptativa em Engenharia de Computação: Estado da Arte e Aplicações*. Phd tese, Universidade de São Paulo, São Paulo, Brasil.
- [PISTORI 2006] PISTORI, H. (2006). Computer vision and digital inclusion of persons with special needs: Overview and state of art. *CompIMAGE - Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications*.
- [PISTORI et al. 2006] PISTORI, H., MARTINS, P. S., PEREIRA, M. C., CASTRO JR., A. A., and NETO, J. J. (2006). Sigus - plataforma de apoio ao desenvolvimento de sistemas para inclusão digital de pessoas com necessidades especiais. *IV Congresso Iberdiscap: Tecnologias de Apoio a Portadores de Deficiência.*, pages 20–22.
- [POWER and SCHOONEES 2002] POWER, W. P. and SCHOONEES, J. A. (2002). Understanding background mixture models for foreground

- segmentation. In *Proceedings Image and Vision Computing New Zealand*, pages 267–271.
- [RASMUSSEN 2000] RASMUSSEN, C. E. (2000). *The Infinite Gaussian Mixture Model*. in *Advances in Neural Information Processing Systems*, pages 554–560.
- [RIBEIRO and GONZAGA 2006] RIBEIRO, H. L. and GONZAGA, A. (2006). Reconhecimento de gestos de mão usando o algoritmo (GMM) e vetor de características de momentos de imagem. *II Workshop de Visão Computacional, WVC2006*.
- [SIMÕES and COSTA 2002] SIMÕES, A. S. and COSTA, A. H. R. (2002). Utilizando processos gaussianos para a segmentação de imagens monocromáticas. In *Anais do IX SEMINCO - Seminário de Computação da Universidade Regional de Blumenau - FURB*, pages 177–188, Blumenau, SC.
- [TERRILLON et al. 1998] TERRILLON, J. C., DAVID, M., and AKAMATSU, S. (1998). Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments. In *FG '98: Proceedings of the 3rd. International Conference on Face & Gesture Recognition*, page 112, Washington, DC, USA. IEEE Computer Society.
- [VEZHNEVETS et al. 2003] VEZHNEVETS, V., SAZONOV, V., and ANDREEVA, A. (2003). A survey on pixel-based skin color detection techniques. In *Proc. Graphicon 2003*.
- [YANG and AHUJA 1999] YANG, M.-H. and AHUJA, N. (1999). Gaussian mixture model for human skin color and its application in image and video databases. *Conf. on Storage and Retrieval for Image and Video Databases (SPIE 99)*, vol. 3656.:p. 458–466.

Anexo A

Tabelas

As Tabelas 1A e 2A utilizaram o formato de cor RGB para calcular o valor das médias, da matriz de covariância e do peso das classes consideradas pele e não-pele.

Tabela 1A - Valores do vetor de médias, da matriz covariância da classe considerada pele.

Kernel	Média	Matriz Covariância	Peso
1	(73.53, 29.94, 17.76)	(765.40, 121.44, 112.80)	0.0294
2	(249.71, 233.94, 217.49)	(39.94, 154.44, 396.05)	0.0331
3	(161.68, 116.25, 96.95)	(291.03, 60.48, 162.85)	0.0654
4	(186.07, 136.62, 114.40)	(274.95, 64.60, 198.27)	0.0756
5	(189.26, 98.37, 51.18)	(633.18, 222.40, 250.69)	0.0554
6	(247.00, 152.20, 90.84)	(65.23, 691.53, 609.92)	0.0314
7	(150.10, 72.66, 37.76)	(408.63, 200.77, 257.57)	0.0454
8	(206.85, 171.09, 156.34)	(530.08, 155.08, 572.79)	0.0469
9	(212.78, 152.82, 120.04)	(160.57, 84.52, 243.90)	0.0956
10	(234.87, 175.43, 138.94)	(163.80, 121.57, 279.22)	0.0763
11	(151.19, 97.74, 74.59)	(425.40, 73.56, 175.11)	0.1100
12	(120.52, 77.55, 59.82)	(330.45, 70.34, 151.82)	0.0676
13	(192.20, 119.62, 82.32)	(152.76, 92.14, 259.15)	0.0755
14	(214.29, 136.08, 87.24)	(204.90, 140.17, 270.19)	0.0500
15	(99.57, 54.33, 38.06)	(448.13, 90.18, 151.29)	0.0667
16	(238.88, 203.08, 176.91)	(178.38, 156.27, 404.99)	0.0749

Fonte: Jones, M. J. e Rehg, J. M. (2002). Statistical color models with application to skin detection. International Journal of Computer Vision, 81-96.

Tabela 2A - Valores do vetor de médias, da variância e o peso da classe considerada não-pele.

Kernel	Vetor de médias	Matriz covariância	Peso
1	(254.37, 254.41, 253.82)	(2.77, 2.81, 5.46)	0.0637
2	(9.39, 8.09, 8.52)	(46.84, 33.59, 32.48)	0.0516
3	(96.57, 96.95, 91.53)	(280.69, 156.79, 436.58)	0.0864
4	(160.44, 162.49, 159.06)	(355.98, 115.89, 591.24)	0.0636
5	(74.98, 63.23, 46.33)	(414.84, 245.95, 361.27)	0.0747
6	(121.83, 60.88, 18.31)	(2502.24, 1383.53, 237.18)	0.0365
7	(202.18, 154.88, 91.04)	(957.42, 1766.94, 1582.52)	0.0349
8	(193.06, 201.93, 206.55)	(562.88, 190.23, 447.28)	0.0649
9	(51.88, 57.14, 61.55)	(344.11, 191.77, 433.40)	0.0656
10	(30.88, 26.84, 25.32)	(222.07, 118.65, 182.41)	0.1189
11	(44.97, 85.96, 131.95)	(651.32, 840.52, 963.67)	0.0362
12	(236.02, 236.27, 230.70)	(225.03, 117.29, 331.95)	0.0849
13	(207.86, 191.20, 164.12)	(494.04, 237.69, 533.52)	0.0368
14	(99.83, 148.11, 188.17)	(955.88, 654.95, 916.70)	0.0389
15	(135.06, 131.92, 123.10)	(350.35, 130.30, 388.43)	0.0943
16	(135.96, 103.89, 66.88)	(806.44, 642.20, 350.36)	0.0477

Fonte: Jones, M. J. e Rehg, J. M. (2002). Statistical color models with application to skin detection. International Journal of Computer Vision, 81-96.