

Reconhecimento Facial Baseado em Gabor Wavelets com Foveação

Thiago Oliveira dos Santos

Dissertação submetida ao Programa de Pós-graduação em Informática da Universidade Federal do Espírito Santo como requisito parcial para a obtenção do grau de Mestre em Informática. Aprovada em 15/09/06 por:

Prof. Thomas Walter Rauber, Ph.D -
DI/UFES

Prof. Alberto Ferreira de Souza, Ph.D -
DI/UFES

Prof. Jacques Facon, Ph.D - DI/PUCPR

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Vitória, Setembro de 2006

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

Santos, Thiago Oliveira dos, 1979-
S237r Reconhecimento facial baseado em gabor wavelets com foveação /
Thiago Oliveira dos Santos. - 2006.
116. : il.

Orientador: Thomas Walter Rauber.
Dissertação (mestrado) – Universidade Federal do Espírito Santo,
Centro Tecnológico.

1. Visão por computador. 2. Reconhecimento de padrões.
3. Características físicas - Localização. 4. Wavelets (Matemática). 5. Retina.
I. Rauber, Thomas Walter. II. Universidade Federal do Espírito Santo.
Centro Tecnológico. III. Título.

CDU: 004

Sumário

Lista de Tabelas

Lista de Figuras

Lista de Siglas

Resumo

Abstract

1	Introdução	p. 15
1.1	Definição do Problema	p. 17
1.1.1	Restrição	p. 17
1.1.2	Implementação	p. 18
1.2	Organização do Trabalho	p. 21
2	Trabalhos Anteriores em Reconhecimento Facial	p. 22
2.1	Localização de Faces Baseado em Redes Neurais	p. 22
2.2	Localização de Faces Através do Espaço de Cores	p. 23
2.3	Template-Matching	p. 24
2.4	Eigenfaces	p. 26
2.5	Elastic Bunch Graph Matching	p. 28
3	Foveação (Foveation)	p. 31
3.1	Campo Receptivo da Retina	p. 31

3.1.0.1	Tipos de RFs	p. 32
3.1.0.2	RFs Circulares	p. 32
3.1.1	Função de Perfil Diferença-de-Gaussianas	p. 32
3.2	Tipos de Distribuição	p. 35
3.2.1	Transformação Log-Polar	p. 36
3.2.2	Transformação Log-Polar Suave	p. 38
3.3	Definição da Retina	p. 41
3.3.1	Aplicando a Foveação	p. 42
3.3.2	Região de Interesse	p. 42
4	Representação da Informação	p. 44
4.1	Gabor	p. 45
4.2	Gabor Wavelet	p. 46
4.3	Componente DC	p. 48
4.4	Wavelets com Foveação	p. 48
4.4.1	Processamento com Gabors	p. 49
4.5	Jets	p. 51
4.6	Similaridade entre Dois Jets	p. 52
5	Identificação do Indivíduo	p. 53
5.1	Representação Facial de um Indivíduo	p. 53
5.1.1	Definição dos Parâmetros	p. 56
Especificação da Retina:	p. 57
Especificação das Wavelets:	p. 57
5.2	Identificação	p. 57
6	Localização de Características	p. 60
6.1	Representação das Características	p. 60

6.2	Similaridade	p. 62
6.3	Rotação	p. 62
6.4	Busca	p. 63
6.4.1	Definição dos Parâmetros	p. 67
	Especificação da Retina:	p. 67
	Especificação das Wavelets:	p. 68
7	Experimentos	p. 69
7.1	Testes de Identificação	p. 69
7.1.1	Teste de Expressões	p. 70
7.1.2	Teste de Iluminação	p. 74
7.1.3	Teste de Oclusões Naturais	p. 77
7.1.4	Teste de Rejeição	p. 82
7.1.5	Teste de Erro na Localização	p. 86
7.1.6	Teste de Tamanho da Base de Imagens	p. 88
7.2	Testes de Localização de Características	p. 89
7.2.1	Teste de Localização do Nariz	p. 89
7.2.2	Teste de Localização do Olho Esquerdo	p. 94
7.2.3	Teste de Localização da Boca	p. 101
7.2.4	Teste de Rotação	p. 105
8	Conclusão e Trabalhos Futuros	p. 108
	Referências	p. 111
	Apêndice A – AR Face Database	p. 114
A.1	Motivação para usar o AR Face Database	p. 116

Lista de Tabelas

1	Testes de expressão da primeira sessão	p. 71
2	Testes de expressão da segunda sessão	p. 71
3	Testes de expressão da primeira sessão encontrados na literatura	p. 72
4	Testes de expressão encontrados na literatura	p. 73
5	Testes de iluminação	p. 74
6	Testes de iluminação encontrados na literatura	p. 76
7	Testes de oclusão dos olhos	p. 77
8	Testes de oclusão da boca	p. 78
9	Testes de oclusão dos olhos encontrados na literatura	p. 80
10	Testes de oclusão da boca encontrados na literatura	p. 81
11	Testes de rejeição $\Omega_{rej} = 0,1765$	p. 83
12	Testes de rejeição $\Omega_{rej} = 0,2941$	p. 84
13	Testes de erros de localização	p. 87
14	Testes de tamanho da base de dados	p. 88
15	Testes de localização do nariz 1	p. 90
16	Testes de localização do nariz 2	p. 91
17	Testes de localização do olho esquerdo 1	p. 95
18	Testes de localização do olho esquerdo 2	p. 96
19	Testes de localização da boca	p. 102
20	Testes de rotação	p. 106

Lista de Figuras

1	Sistema de reconhecimento facial genérico	p. 16
2	Sistema de reconhecimento facial	p. 19
3	Fluxo de informação	p. 20
4	Algoritmo utilizado para detecção de faces	p. 23
5	Algoritmo de detecção facial	p. 24
6	Espaço de cores YC_bC_r	p. 24
7	Representação dos templates de uma face	p. 25
8	Modelo de aspecto	p. 27
9	Modelo discriminativo	p. 28
10	Grafo da face	p. 29
11	Exemplo de um Grafo para uma face genérica	p. 30
12	Exemplo de Grafos genéricos e específicos para face	p. 30
13	Centro-periferia (Center-Sorround)	p. 33
14	Modelagem Centro-periferia (Center-Sorround)	p. 33
15	Diferença de Gaussianas - DOG	p. 34
16	Exemplo de alguns RFs do tipo DOG	p. 34
17	Exemplo da redução de dados espaço variante	p. 36
18	Modelo $\log(z)$	p. 38
19	Distribuição radial com transição suave para $f = 9$	p. 40
20	Distribuição de RFs pela retina com 128 pixels de diâmetro e fóvea com 8 pixels de raio	p. 41
21	Exemplo dos pontos da retina com o foco de atenção no nariz.	p. 42

22	Exemplos de diferentes orientações da função Gabor	p. 45
23	Exemplos de diferentes comprimentos de onda da função Gabor	p. 45
24	Exemplos de diferentes fases da função Gabor	p. 45
25	Exemplos de diferentes raios da gaussiana da função Gabor	p. 46
26	Exemplos de diferentes formas da gaussiana da função Gabor	p. 46
27	Exemplo de algumas Wavelets formadas por Gabor	p. 48
28	Exemplo do pós-processamento realizado na informação da retina	p. 50
29	Exemplo de algumas Wavelets formadas por Gabor com Foveação	p. 51
30	Localização dos pontos faciais	p. 54
31	Exemplo da distribuição dos pontos faciais em outras imagens.	p. 54
32	Exemplo da amostragem em cada ponto para identificação	p. 56
33	Exemplo da amostragem em cada ponto para a localização	p. 61
34	Exemplo da amostragem realizada no nariz	p. 61
35	Exemplo dos pontos da retina	p. 62
36	Fluxograma da busca por sacada.	p. 64
37	Exemplo de uma região de amostragem para a sacada	p. 65
38	Exemplo da busca por sacada.	p. 67
39	Exemplo de fotos dos conjuntos AR1, AR2, AR3 e AR4, tiradas na primeira sessão.	p. 71
40	Exemplo de fotos dos conjuntos AR14, AR15, AR16 e AR17, tiradas na segunda sessão.	p. 72
41	Exemplo de fotos dos conjuntos AR1, AR5, AR6, AR7, AR18, AR19 e AR20.	p. 75
42	Exemplo de fotos dos conjuntos AR1, AR8, AR9, AR10, AR21, AR22 e AR23.	p. 78
43	Exemplo de fotos dos conjuntos AR1, AR11, AR12, AR13, AR24, AR25 e AR26.	p. 79

44	Exemplo de fotos dos conjuntos AR1, AR2, AR3, AR4, AR14, AR15, AR16 e AR17.	p. 85
45	Exemplo de fotos dos conjuntos AR1, AR2, AR3 e AR4.	p. 87
46	Exemplo de fotos dos conjuntos AR1 e AR4.	p. 88
47	Exemplo de fotos de treino para localização do nariz.	p. 89
48	Exemplo de narizes localizados erroneamente.	p. 90
49	Exemplo de narizes localizados erroneamente.	p. 91
50	1 Espalhamento dos pontos do nariz localizados em cada I_S	p. 92
51	2 Espalhamento dos pontos do nariz localizados em cada I_S	p. 93
52	Exemplo de fotos de treino para localização do olho esquerdo.	p. 95
53	Exemplo de olhos localizados erroneamente.	p. 96
54	Exemplo de olhos localizados erroneamente.	p. 97
55	1 Espalhamento dos pontos do olho esquerdo localizados em cada ISON-DAGEM	p. 98
56	2 Espalhamento dos pontos do olho esquerdo localizados em cada ISON-DAGEM	p. 99
57	3 Espalhamento dos pontos do olho esquerdo localizados em cada ISON-DAGEM	p. 100
58	Exemplo de fotos de treino para localização da boca.	p. 102
59	Exemplo de bocas localizadas erroneamente.	p. 102
60	1 Espalhamento dos pontos da boca localizados em cada I_S	p. 103
61	2 Espalhamento dos pontos da boca localizados em cada I_S	p. 104
62	Exemplo de fotos de treino do teste de rotação através do nariz.	p. 105
63	Exemplo da rotação de 90° aplicada nas fotos utilizadas durante a sondagem, I_S	p. 106
64	Exemplo de rotações identificadas erroneamente.	p. 106
65	Exemplo de uma Figura com inclinação na cabeça mesmo com a imagem em 0°.	p. 107

66	Exemplo das fotos relativas à primeira sessão	p.115
67	Exemplo das fotos relativas à segunda sessão	p.115

Lista de Siglas

I_S	Imagens de Sondagem
I_T	Imagens de Treino
AFGR	International Conference on Automatic Face and Gesture Recognition
AVBPA	International Conference on Audio and Video-Based Authentication
Bayesian	Probabilidade de Bayes
DOG	Difference-Of-Gaussians
FaceIt	Modulo de reconhecimento baseado em Local Feature Analysis
FP	Funções de Perfil
GW	Gabor Wavelets sem foveação
GWF	Gabor Wavelets com foveação
IC	Identificação Correta
IE	Identificação Errada
IEI	Identificação Errada Impostor
IET	Identificação Errada Trocada
LDA	Linear Discriminant Analysis
LGN	Lateral Geniculate Nucleus
NMF-100+Correlation	Non-Negative Matrix Factorization com dimensão 100 e utilizando distância por Correlação
NMF-100+L1 Norm	Non-Negative Matrix Factorization com dimensão 100 e utilizando distância de Cidade de Blocos
NMF-100+L2 Norm	Non-Negative Matrix Factorization com dimensão 100 e utilizando distância Euclideana
NMF-150+Correlation	Non-Negative Matrix Factorization com dimensão 100 e utilizando distância por Correlação
NMF-150+L1 Norm	Non-Negative Matrix Factorization com dimensão 150 e utilizando distância Euclideana

NMF-150+L2 Norm	Non-Negative Matrix Factorization com dimensão 150 e utilizando distância Euclideana
NMF-50+Correlation	Non-Negative Matrix Factorization com dimensão 50 e utilizando distância por Correlação
NMF-50+L1 Norm	Non-Negative Matrix Factorization com dimensão 50 e utilizando distância de Cidade de Blocos
NMF-50+L2 Norm	Non-Negative Matrix Factorization com dimensão 50 e utilizando distância Euclideana
PCA	Principal Component Analysis
PCA-100+L2 Norm	Principal Component Analysis com dimensão 100 e utilizando distância Euclideana
PCA-150+L2 Norm	Principal Component Analysis com dimensão 150 e utilizando distância Euclideana
PCA-50+L2 Norm	Principal Component Analysis com dimensão 50 e utilizando distância Euclideana
RC	Rejeição Correta
RE	Rejeição Errada
RF	Receptive Field

Resumo

Este trabalho apresenta um sistema de reconhecimento facial que é composto por duas partes: localização de características faciais e identificação facial. Elas são baseadas em Gabor Wavelets (técnica que simula o comportamento de células corticais do cérebro humano). Um dos pontos-chaves do trabalho é a utilização de um método de foveação (processamento realizado pela retina do olho humano) para a compactação da informação a ser analisada pelas Wavelets. O algoritmo de identificação foi desenvolvido de forma a lidar com variações, como diferença temporal, diferença de iluminação, diferentes expressões e oclusões naturais. Ele também é capaz de reconhecer indivíduos não pertencentes ao sistema, isto é, ele consegue rejeitar intrusos. Para avaliar a influência da foveação na representatividade das Gabors Wavelets, fez-se uma comparação entre o algoritmo implementado com informação proveniente de uma imagem foveada com o implementado com informação proveniente da imagem original. Os testes realizados mostraram que os dois lidam bem com as diversidades citadas acima tendo resultados com percentuais de acerto acima 80%. Contudo, em geral, o algoritmo com foveação teve um percentual de acerto ligeiramente melhor (cerca de 3%) do que o sem foveação. A parte de localização de características, foi implementada para localizar três pontos específicos da face (nariz, olho esquerdo e boca). Isso porque o algoritmo de identificação toma essas posições como ponto de partida. Como método de busca ela utilizou a sacada (saltos feitos pelo olho humano para analisar uma região de interesse) a fim de reduzir o esforço computacional. Nos testes realizados a localização não foi muito flexível quanto a variações nas características procuradas. Entretanto, para condições normais os resultados foram muito bons tendo percentuais de acertos acima de 92%. Além da localização que identifica translações das características, introduziu-se um mecanismo para aproveitar a informação foveada para a identificação de uma rotação na região procurada. Os testes realizados mostraram ser possível obter tal informação angular.

Palavras-chave: Reconhecimento Facial; Localização de Características; Foveação; Gabor Wavelets; Sacada; Retina.

Abstract

This work presents a face recognition system composed of two parts: face feature localization and face identification. They are based on Gabor Wavelets (technique that simulates the cortical cells behaviors of the human brain). One of the key points of this work is the use of a foveation method (process accomplished by the retina of the human eye) to compact the information to be analyzed by the Wavelets. The identification algorithm was developed to deal with some variations, like temporal difference, light differences, different expressions and natural occlusions. It also can recognize people that are not in the system and reject intruders. To evaluate the influence of the foveation in the representativeness of the Gabor Wavelets, a comparison was made between the algorithm implemented with the foveated image and the one without foveation. Tests have shown that both deal very well with the variations mentioned above with recognition rates greater than 80%. However, in general, the foveated one has been slight better (about 3%) than the other one. The feature localization algorithm was implemented to find three specific face points (nose, left eye and mouth). It has been made because the identification algorithm needs these positions as parameters. It used a saccadic search (jumps accomplished by the human eye to analyze an interest region) so that the computational effort got reduced. Tests have shown that the localization is not so flexible to deal with variations, but it has been very well to normal conditions with localization rates greater than 92%. Besides the localization that identify features translations, a way to process the foveated information to identify some rotation in the searched region was introduced. Tests have shown that it's possible to get such angular information.

Keywords: Facial Recognition; Feature Localization; Foveation; Gabor Wavelets; Saccadic Search; Retina.

1 *Introdução*

O Reconhecimento Facial é uma tarefa cotidiana para a maioria dos seres humanos sendo necessária a cada vez que uma pessoa se depara com outra, seja pessoalmente, por meio de fotografias ou vídeos.

Essa identificação, aparentemente simples para o cérebro humano, não é igualmente simples de ser realizada pelos computadores fazendo com que muitos pesquisadores dediquem uma atenção especial a esse tipo de aplicação. A confirmação disso é o surgimento de conferências específicas para esse tópico, como: *International Conference on Audio and Video-Based Authentication* (AVBPA) desde 1997 e *International Conference on Automatic Face and Gesture Recognition* (AFGR) desde 1995 (ZHAO et al., 2000).

Uma das motivações para essa crescente dedicação é a vontade de automatizar processos que requerem reconhecimento facial. Zhao et al. (2000) apresenta um resumo das áreas abrangidas por essas aplicações, e destaca:

- **Entretenimento:** vídeo game, realidade virtual, programas de treinamento, interação humano-máquina, etc;
- **Cartões inteligentes:** carteira de habilitação, imigração, identidade, passaporte, etc;
- **Segurança da informação:** login pessoal em equipamentos, aplicações de segurança, banco de dados de segurança, encriptação de arquivos, segurança de intranet, acesso a internet, registros médicos, etc;
- **Execução da lei:** pesquisa de vídeo avançada, análise de pós-eventos, rastreamento e investigação de suspeitos, etc.

Independentemente da aplicação a ser atendida, o processo de reconhecimento facial segue uma estrutura genérica composta por duas tarefas principais: localização da face na imagem e identificação da face localizada. Dependendo do método utilizado para identificação pode ser necessário a localização de certos pontos da face, como por exemplo:

olhos, nariz, boca, sobrancelha, etc. A etapa responsável por encontrar tais pontos é denominada de localização de características e tem como objetivo retornar a posição (x, y) de alguns pontos específicos da face.

A Figura 1 ilustra um sistema genérico de reconhecimento facial. Ele recebe como entrada uma imagem, que também pode ser de vídeo, e gera uma ou várias identificações como saída. Para isso, é feita a localização facial, que pode ocorrer simultaneamente com a localização de características da face, e posteriormente a identificação facial. A etapa de localização de características só será realizada se o método de identificação necessitar da saída gerada por ela. Outras técnicas auxiliares objetivando aumentar a robustez do sistema também podem ser consideradas, como: eliminação de ruídos, normalizações (quanto a cores, tamanho, rotação, ...).

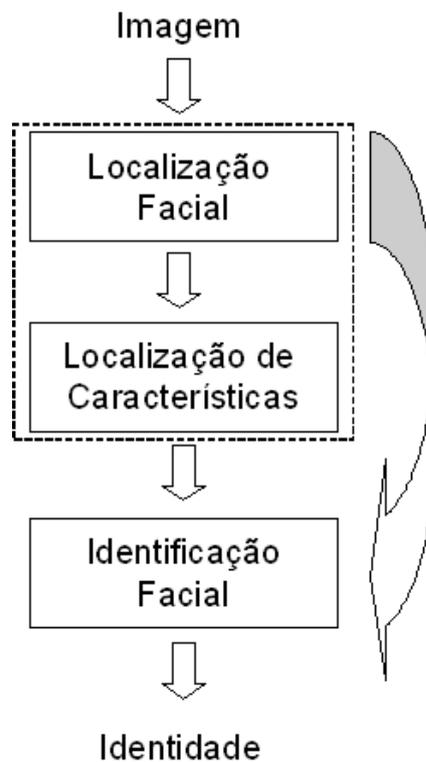


Figura 1: Sistema de reconhecimento facial genérico. O sistema recebe como entrada uma imagem (foto ou vídeo) e gera como saída uma identificação. A localização facial pode ocorrer simultaneamente com a localização de características. Essa, pode ocorrer ou não.

Seguindo essa estrutura, surgiram na literatura variadas técnicas computacionais para resolver, ou ajudar a resolver, o problema de reconhecimento facial, como será descrito na seção 2. Entre elas, existe uma linha de pesquisa dedicada a investigar o comportamento da visão e do cérebro de animais a fim de entender melhor o sistema visual humano e tentar reproduzi-lo computacionalmente, como em (SCHWARTZ, 1977, 1980; WILSON,

1983; SCHWARTZ, 1994a, 1994b, 1985; SCHWARTZ et al., 1983). Um forte motivo para essa analogia é a robustez e a velocidade com que o ser humano resolve essa tarefa.

Esse trabalho tem como objetivo elucidar um problema específico na área de pesquisa em reconhecimento facial e propor uma solução com a unificação de algumas técnicas pré-existentes e baseadas no sistema biológico.

1.1 Definição do Problema

Como apresentado anteriormente, a gama de aplicações que requerem reconhecimento facial é muito grande e cada uma delas pode lidar com diferentes fontes de imagem. De acordo com (ZHAO et al., 2000), é possível separá-las em dois grandes grupos de estudo: os provenientes de fotografias estáticas com formatos controlados; e os provenientes de imagens vídeo com formatos variados e dinâmicos. Quanto menos controlado é a fonte de entrada, maior é a variação entre as imagens adquiridas e mais robusto deve ser o sistema.

Essas diferenças podem ser muitas, como por exemplo: quanto a qualidade, variabilidade do fundo, variabilidade do indivíduo a ser reconhecido (expressões, idade, óculos, rotação, escalabilidade, translação), entre outras que podem surgir de acordo com o tipo de aplicação.

Outros fatores a serem levados em consideração são: o tempo que a aplicação terá para processar o reconhecimento do indivíduo e o espaço necessário para armazenar as informações a respeito deles. Algumas aplicações necessitam de respostas imediatas (sistemas de tempo real), já outras, podem fazer o reconhecimento em um momento posterior. Assim como algumas têm espaço de armazenamento restrito (como as de celulares, Hand Held ou Palm Top) e outras têm espaço de armazenamento consideravelmente grande (como em grandes servidores). Cada um desses fatores, pode ser decisivo no momento de escolha do algoritmo para realizar tal tarefa.

Devido a essa grande variabilidade, a próxima seção dedica-se a restringir o problema a ser abordado nesse trabalho.

1.1.1 Restrição

Neste trabalho reconhecimento é feito a partir de uma imagem de entrada e de um banco de imagens de pessoas a ser previamente conhecido pelo sistema. Elas são todas estáticas, isto é, não há vídeos, e foram obtidas em um ambiente controlado em relação

ao fundo e a posição. Isso significa que as fotos foram tiradas sempre de uma mesma distância e utilizando o mesmo fundo, minimizando a variação desses dois quesitos, embora o método implementado não requeira a segunda restrição. Em algumas dessas imagens há oclusão de partes da face, como: olhos (cobertos por óculos de sol) e boca (coberta por um cachecol).

Não foram impostas restrições de tempo de resposta ou quanto ao espaço de armazenamento da base de dados. Contudo, acredita-se que isso não impede o método de funcionar corretamente em um ambiente restrito, ou seja, que responda em um tempo relativamente pequeno, ou que tenha um espaço de armazenamento pequeno.

1.1.2 Implementação

Para solucionar o problema definido anteriormente implementou-se um sistema de reconhecimento facial dividido em duas partes. Uma responsável por localizar três características da face (nariz, olho esquerdo e boca), e outra responsável por fazer a identificação de indivíduos tendo a localização das três características citadas previamente conhecida. Apesar de terem seguido a mesma linha de pensamento, elas foram desenvolvidas independentemente, para que, caso necessário, fosse possível o uso de métodos e testes diferentes. A Figura 2 ilustra o sistema implementado.

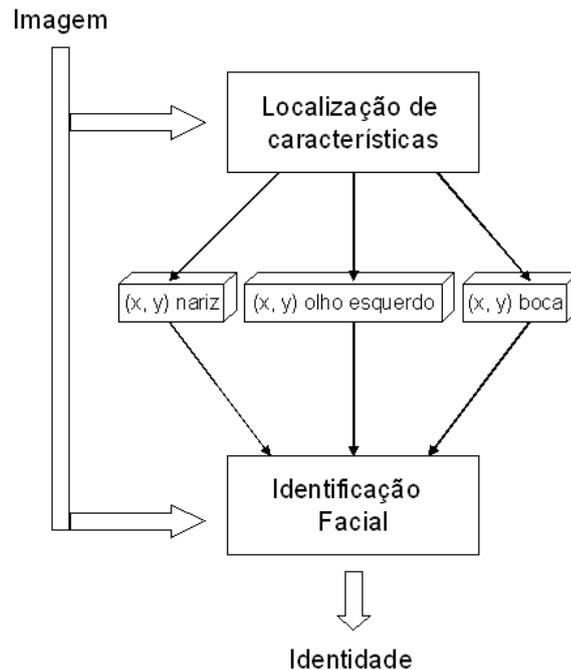


Figura 2: Sistema de reconhecimento facial. O sistema recebe como entrada uma imagem contendo uma face e gera como saída a identificação do indivíduo. Se as coordenadas (x, y) do olho esquerdo, do nariz e da boca forem previamente conhecidas, então não há necessidade do módulo de localização de características.

O sistema baseia-se em técnicas amplamente conhecidas e estudadas, como Gabor Wavelets e Foveação.

- **Gabor Wavelets:** para representar a informação necessária a ser extraída das imagens em forma de orientações e frequências;
- **Foveação:** para realizar uma compactação na quantidade de informação extraída da imagem para ser processada;

O emprego destas técnicas neste trabalho busca reproduzir computacionalmente o comportamento de parte do sistema visual de alguns animais, inclusive dos seres humanos. As Gabor Wavelets modelam a resposta de algumas células corticais responsáveis por certas funções do sistema visual (VALOIS; VALOIS, 1990). Já Foveação, modela a estrutura/processamento da retina visando promover uma compactação sem a perda do campo visual (BERNARDINO, 2004; BOLDUC; LEVINE, 1997, 1998). As duas técnicas são bastante utilizadas no ramo da visão computacional, porém de forma independente. Um dos objetivos desse trabalho é estudar como a unificação das duas afetará a capacidade de identificação dos indivíduos pelo sistema. Para isso, implementou-se a parte de identificação baseada em Gabor Wavelets geradas a partir de imagens foveadas. A Figura 3

ilustra o fluxo da informação a ser extraída em cada região da imagem que seja relevante para a identificação.

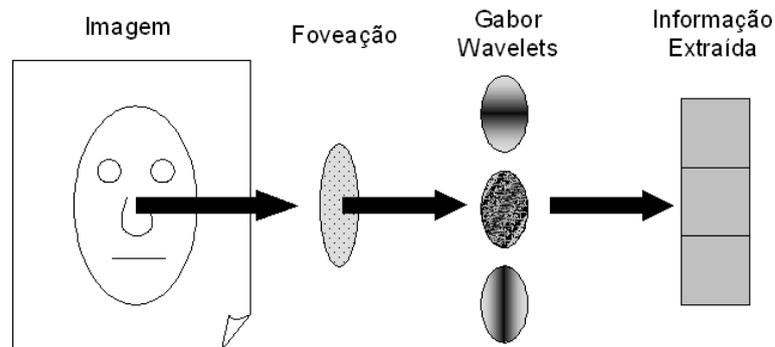


Figura 3: Fluxo de informação. Escolhe-se a região da imagem a partir do qual a informação será extraída. Aplica-se a foveação. Aplica-se as Gabor Wavelets na informação foveada gerando-se a informação que representará a região em questão.

Essa unificação foi motivada pelo fluxo da informação visual, apresentado em (VALOIS; VALOIS, 1990), no qual a retina coleta os dados do campo visual, promove uma compactação e envia os sinais compactados para o Córtex através do nervo ótico.

Seguindo essa mesma abordagem a respeito do fluxo de informação, implementou-se a parte de localização das características da face (nariz, olho esquerdo e boca). Com isso foi possível estudar a representatividade de cada um desses atributos a partir dessa técnica, assim como outras formas de aproveitar a informação proveniente da foveação como, por exemplo, para descobrir uma possível rotação na imagem.

Uma outra técnica utilizada nesse trabalho e inspirada no modelo biológico foi a busca visual baseada em sacadas. Sacadas são saltos rápidos realizados pelo olho humano visando explorar diferentes partes do ambiente. Ela foi associada ao método de localização de características para reduzir o espaço e tempo de busca visual, possibilitando assim, a realização de testes mais consistentes em relação ao método de localização.

Segue abaixo uma lista dos pontos estudados na parte de identificação de indivíduos:

- Comparação da capacidade de reconhecimento do sistema com foveação e sem foveação;
- Verificação da viabilidade do método em diferentes condições (variação de expressões, variação de luz, oclusão ocular, oclusão bucal);
- Avaliação da confiabilidade de resposta do sistema.

Segue abaixo uma lista dos pontos estudados na parte de localização de características:

- Verificação da capacidade do sistema de representar cada característica;
- Verificação da viabilidade do método em diferentes condições (variação de expressões, variação de luz, oclusão ocular, oclusão bucal);
- Verificação da possibilidade de identificação de rotação da característica procurada.

1.2 Organização do Trabalho

Este trabalho está organizado conforme abaixo:

Capítulo 2: Apresenta trabalhos existentes na literatura relacionados ao tema Reconhecimento Facial;

Capítulo 3: Descreve a técnica de foveação e apresenta o modelo de retina utilizada;

Capítulo 4: Descreve as Gabor Wavelets e mostra como gerá-las com ou sem foveação;

Capítulo 5: Descreve a parte do sistema responsável por identificar um indivíduo;

Capítulo 6: Descreve a parte do sistema responsável por localizar as características da face a serem utilizadas pela parte de identificação;

Capítulo 7: Apresenta os experimentos realizados com as técnicas implementadas;

Capítulo 8: Apresenta as conclusões finais e as perspectivas para trabalhos futuros.

2 Trabalhos Anteriores em Reconhecimento Facial

Nos últimos anos, o problema de *Reconhecimento Facial* computacional tem atraído pesquisadores de várias áreas, como processamento de imagem, reconhecimento de padrões, redes neurais, visão computacional, computação gráfica e psicologia, fazendo com que surgissem soluções variadas, cada uma com seus pontos fortes e fracos.

Essa seção é dedicada a apresentar alguns trabalhos existentes na literatura relacionados ao tema Reconhecimento Facial.

2.1 Localização de Faces Baseado em Redes Neurais

Rowley (1999) apresenta um algoritmo, baseado em *Redes Neurais*, para detecção de face em uma imagem. O primeiro componente do sistema é uma rede neural que recebe como entrada uma região da imagem, isto é, uma sub imagem de formato quadrático, e gera como saída um valor de -1 a 1, significando a presença ou não de uma face. Para tornar-se invariante quanto à translação, a rede recebe como entrada todas as combinações possíveis da sub imagem quadrática a imagem original, isto é, retira uma sub imagem de cada parte da imagem a fim de verificar a presença ou não de face. Para tornar-se invariante quanto à mudança de escala, a imagem de entrada é reduzida de tamanho e o procedimento anterior é repetido. O número de reduções varia de acordo com a necessidade de invariância quanto a mudança de escala. Antes da entrada ser passada para a rede faz-se uma correção na luz e uma equalização de histograma.

A rede neural foi implementada de forma que a sub imagem quadrática de entrada fosse dividida em pedaços menores, sendo: quatro sub regiões quadráticas de tamanhos iguais e sem sobreposição, dezesseis sub regiões quadráticas de tamanhos iguais e sem sobreposição, e seis sub regiões retangulares tendo o maior comprimento na direção horizontal e uma sobreposição na direção vertical, como pode ser visto na parte direita da Figura 4. Cada uma dessas sub regiões estará conectada totalmente à camada escondida da rede. O formato delas foi escolhido de maneira a permitir à rede distinguir característi-

cas que possam ser importantes para a detecção da face. As regiões horizontais permitem detectar características como: bocas, ou pares de olhos. Enquanto as regiões quadráticas são responsáveis por olhos individuais, nariz, ou cantos da boca. A Figura 4 ilustra o algoritmo utilizado.

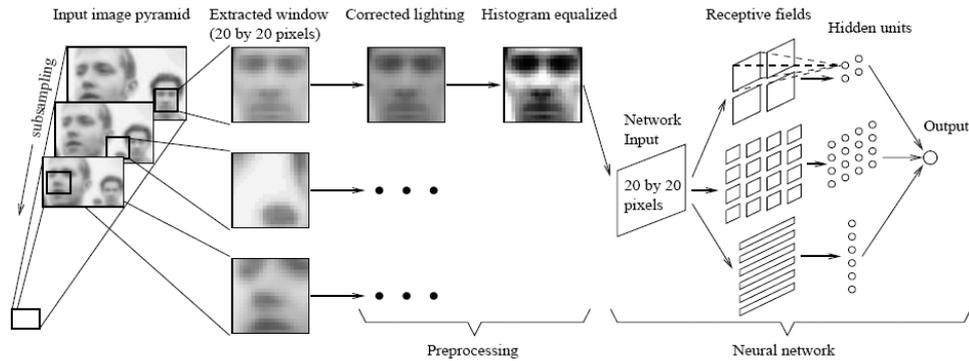


Figura 4: Algoritmo utilizado para detecção de faces. Figura retirada de (ROWLEY, 1999).

2.2 Localização de Faces Através do Espaço de Cores

Uma das formas de detectar faces em uma imagem é utilizar as informações de cores. Senior et al. (2002) propõe um algoritmo para a detecção facial em imagens coloridas na presença de variadas condições de luz e planos de fundo complexos. Ele é baseado em uma técnica de compensação de luz e uma transformação não linear de cor sendo capaz de detectar regiões de pele por toda a imagem, e depois gerar os candidatos a face de acordo com a distribuição espacial desses pedaços de pele. O algoritmo detecta também as regiões de olho, boca, e limites faciais para a verificação de cada candidato a face.

Na Figura 5 pode ser vista uma descrição em alto nível do algoritmo de detecção facial. Ele é composto por dois grandes módulos: (i) Localização facial para achar os candidatos a face, (ii) Localização das características da face para verificação dos candidatos. Primeiro faz-se uma estimativa e correção da cor de acordo com uma técnica de compensação de luz. Depois é feita uma transformação não linear das componentes RGB para o espaço de cores YC_bC_r . Os tons de pele são detectados utilizando-se um modelo elíptico de pele no espaço bi-dimensional obtido a partir do YC_bC_r . A elipse corresponde ao contorno da distância de Mahalanobis constante, assumindo-se uma distribuição gaussiana para os tons de pele, ver Figura 6. Os pixels avaliados com tons de pele são iterativamente segmentados utilizando uma variância local de cor em componentes conectadas. Posteriormente, eles são agrupados em candidatos a face de acordo com a distribuição espacial

dessas componentes e a similaridade de suas cores. O tamanho de um candidato a face pode variar de $13 * 13$ pixels a três quartos da imagem de entrada. O módulo de detecção de características faciais também faz a localização através de espaço de cores e rejeita regiões que não contenham características faciais, tal como: olhos, boca ou limite facial. Ainda na Figura 5 é possível ver uma face localizada.

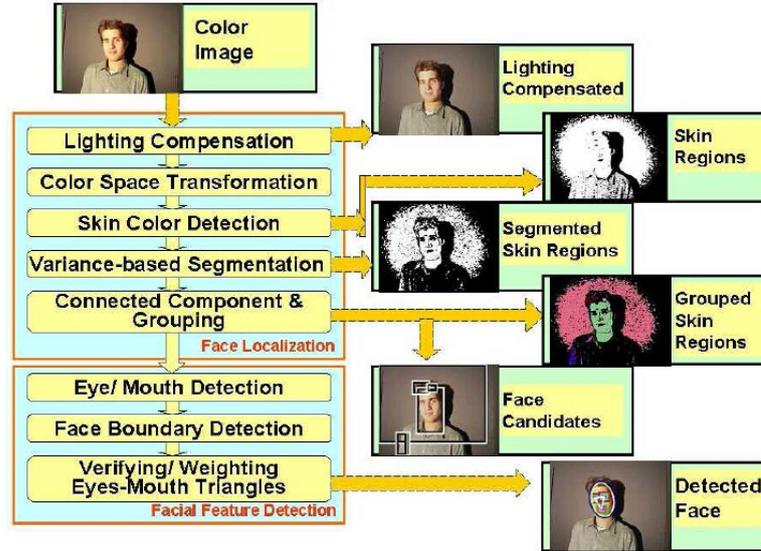


Figura 5: Algoritmo de detecção facial. Figura retirada de (SENIOR et al., 2002).

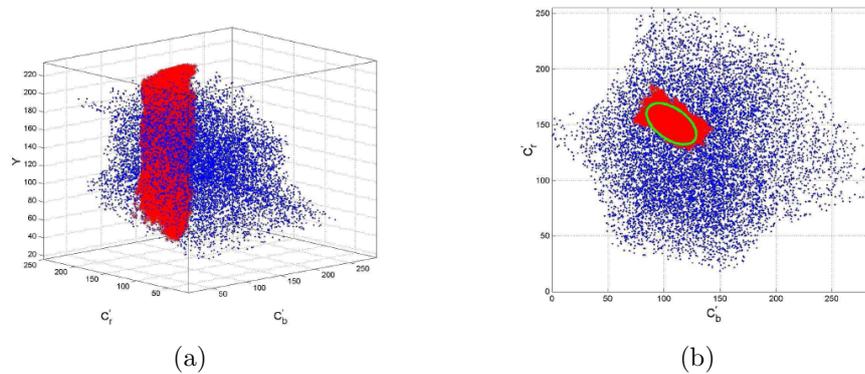


Figura 6: Espaço de cores YC_bC_r . Os pontos em azul representam as cores possíveis de um monitor. Os pontos vermelhos representam as amostras de tons de pele. Os pontos em verde representam a elipse. (a) espaço de cores YC_bC_r . (b) espaço bi-dimensional obtido a partir de YC_bC_r . Figura retirada de (SENIOR et al., 2002).

2.3 Template-Matching

De acordo com Brunelli e Poggio (1993), a versão mais simples dos métodos baseados em casamento de templates (*Template-Matching*) utiliza uma matriz contendo os valores

das intensidades, pixel a pixel, de uma face. A partir disso, gera-se uma galeria com imagens conhecidas com cada face representada por uma matriz dessas. Para a classificação de uma face é feita uma comparação, utilizando a distância euclidiana, do vetor da imagem de teste com cada um da galeria conhecida. O par que possuir a menor distância é eleito como casado.

Porém, existem várias outras maneiras de realizar esse casamento. Uma delas é pre-processar as informações de intensidade de forma a eliminar variações indesejadas. Outra mais complexa e diferente, conhecida como template elástico, é utilizar um template simples juntamente com um modelo qualitativo de como uma face genérica transforma-se sob uma mudança no ponto de vista. O modelo de deformação é posteriormente utilizado na construção de uma métrica para o casamento.

No sistema implementado por Brunelli e Poggio (1993), é feita uma prévia normalização (coeficiente de correlação-cruzada) nas imagens a fim de eliminar variações de posição, escala e rotação. Cada pessoa é representada por quatro máscaras: dos olhos, do nariz, da boca e da face, como pode ser visto na Figura 7. Após a comparação de cada máscara da imagem de teste com as das imagens conhecidas, é eleito como casado o par que apresentar maior similaridade acumulada nos quatro templates.

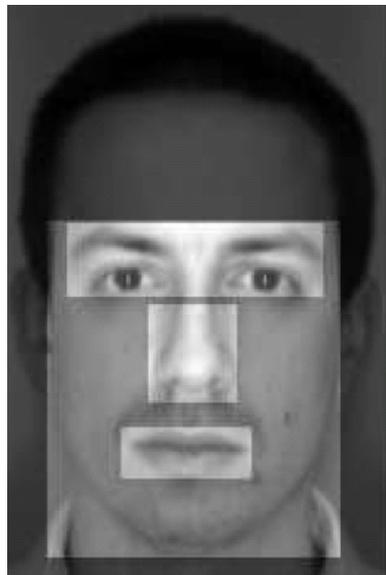


Figura 7: Representação dos templates de uma face. Figura retirada de (BRUNELLI; POGGIO, 1993).

2.4 Eigenfaces

Uma técnica bem conhecida para realizar a identificação de faces humanas é a baseada em Eigenfaces, como proposto por Turk e Pentland (1991). Ela utiliza um método estatístico, conhecido como *Análise de Componentes Principais* (PCA), ou *Karhunen-Loeve*, para gerar um descritor das informações que possuem maior variação entre as faces conhecidas. Com isso, é possível extrair, de uma face, vetores (conhecidos como *eigenvectors*) contendo somente as informações mais relevantes. Essa restrição na qualidade das informações promove uma compactação diminuindo o esforço computacional para uma posterior comparação. O espaço de características descrito por um conjunto desses vetores é chamado de *Espaço de faces*, ou Eigenfaces.

Como pode ser visto em (KEPENEKCI, 2001; NES, 2003; KEAT,) existem duas abordagens principais para a utilização de eigenfaces no reconhecimento de faces: modelo de aspecto (appearance model) e modelo discriminativo (discriminative model).

Modelo de aspecto, ver Figura 8:

1. Uma base de dados de imagens faciais é coletada;
2. Um conjunto de eigenfaces é gerado aplicando a PCA em cada imagem facial. Aproximadamente, 100 eigenvectors são suficientes para codificar uma grande base de dados;
3. Cada face é representada como uma combinação linear de eigenfaces;
4. Uma imagem de teste é aproximada por uma combinação de eigenfaces. A distância medida é utilizada para comparar a similaridade entre duas imagens.

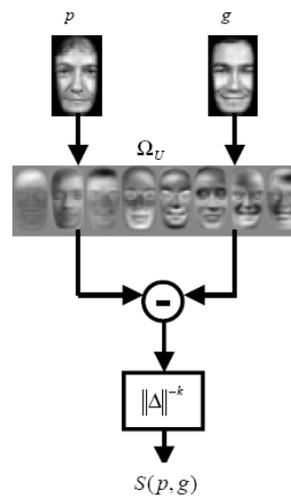


Figura 8: Modelo de aspecto. Figura retirada de (KEPENEKCI, 2001).

Modelo de discriminativo, ver Figura 9:

1. Nesse modelo a diferença de duas imagens pode pertencer a duas categorias dando origem a duas bases de dados Ω_I e Ω_E . Uma através da diferença inter-pessoal (casando duas visões de cada individuo na base de dados) e outra pela diferença extra-pessoal (casando diferentes indivíduos na base de dados), respectivamente;
2. Duas bases de dados são geradas aplicando a PCA em cada uma das classes;
3. Para classificar uma imagem de teste utiliza-se a probabilidade de Bayes a fim de identificar se as variações entre a imagem de teste e uma imagem da base de dados são de origem inter-pessoal ou extra-pessoal.

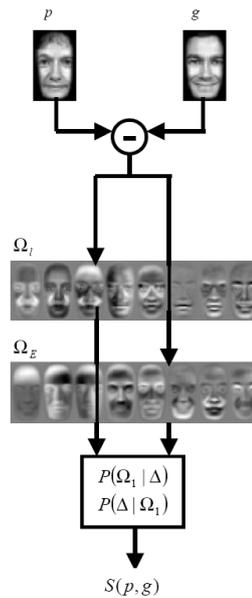


Figura 9: Modelo discriminativo. Figura retirada de (KEPENEKCI, 2001).

Uma variação desse método utiliza, ao invés da face inteira, apenas porções da face tidas como relevante (olhos, nariz e boca) para a geração do espaço de características. Sendo portanto, denominado de *Eigenfeatures*, ao invés de *Eigenfaces*.

2.5 Elastic Bunch Graph Matching

Wiskott et al. (1997) propõe um sistema para reconhecimento de faces humanas sob variações de posição, tamanho, expressão e pose. Ele extrai descrições da face através de grafos, nos quais os vértices estão localizados em pontos representativos da face, como: olhos, nariz, boca e outros. Cada vértice armazena informações de orientação, frequência e fase, obtidas por convoluções entre funções Gabor com diferentes parâmetros, e a região a qual representa. O reconhecimento é obtido através da comparação de cada grafo contido na base de dados com o grafo da imagem de teste.

A Figura 10 ilustra um exemplo do grafo representativo de uma face. Para construí-lo, utiliza-se um *Jet*, que é um conjunto de 40 coeficientes complexos (magnitudes e fases) obtidos pela convolução de um ponto da imagem com a função de Gabor em 5 diferentes frequências e 8 diferentes orientações. Cada face a ser armazenada na base de dados deve ser convertida para um grafo. Para realizar o reconhecimento de uma imagem de teste, gera-se um grafo para ela e calcula-se uma média das similaridades entre cada *Jet* contido do grafo da face testada com os contidos em cada grafo das imagens da base de imagens. O grafo da base que apresentar uma similaridade média maior é o representante da face.

A função de comparação entre os *Jets* pode assumir três formas: considerando somente a magnitude; considerando a magnitude e a fase; ou considerando a magnitude, a fase e um deslocamento estimado através da maximização da função de similaridade.

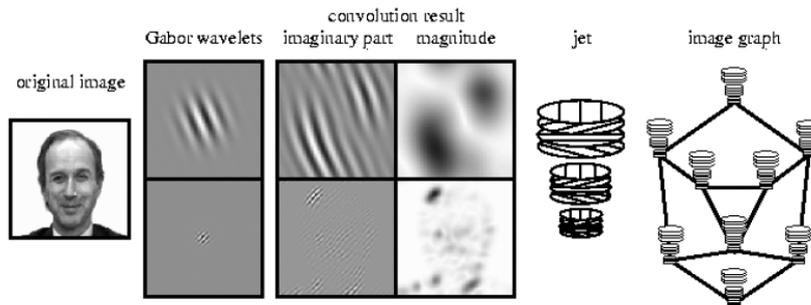


Figura 10: Grafo da face. Da esquerda para direita, vê-se a imagem original, funções de Gabor, a parte imaginária da convolução da imagem original com a função de Gabor, a magnitude da convolução da imagem original com a função de Gabor, um conjunto de informações (magnitude e fase) extraídas de um ponto da imagem, um grafo da face composto por vários conjuntos de informações. Figura retirada de (WISKOTT et al., 1997).

Ainda em (WISKOTT et al., 1997) e utilizando-se a mesma idéia de grafo, é proposto um método para geração automática desses grafos através de um conjunto de *Jets* para cada vértices, ao invés de apenas um. Esse conjunto seria responsável por representar um vértice para uma face genérica. Na Figura 11 é ilustrado um grafo representativo de uma face genérica. Pode-se ver que cada vértice é composto por um conjunto de *Jets*, a fim de que possua informações suficientes para identificar qualquer tipo de face. Por exemplo, um vértice representando um olho deve conter informações de olhos abertos, fechados, femininos, masculinos e assim por diante. Quanto mais informação o vértice possuir, mais representativo ele será.

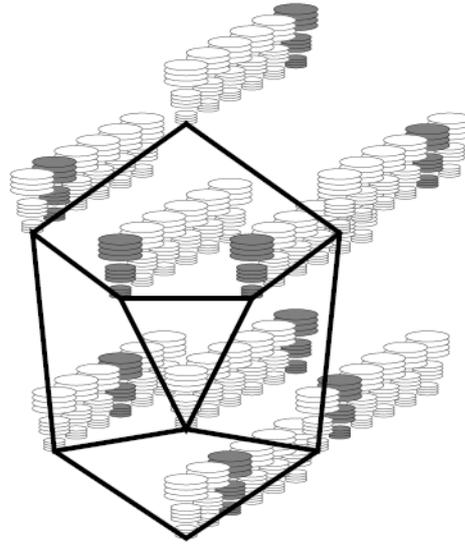


Figura 11: Exemplo de um Grafo para uma face genérica. Cada vértice possui vários *Jets*. Figura retirada de (WISKOTT et al., 1997).

O ponto chave para a geração do grafo da face a partir do grafo da face genérica é achar os vértices que maximizem a função de similaridade. Essa leva em consideração tanto os *Jets*, quanto distância relativa entre os vértices do grafo. Com isso é possível identificar variações na pose, no tamanho e translações. As duas imagens na esquerda da Figura 12 ilustram exemplos de grafos para localização automática e as quatro imagens na direita da Figura 12 ilustram exemplos de grafos da face gerados a partir da localização automática.

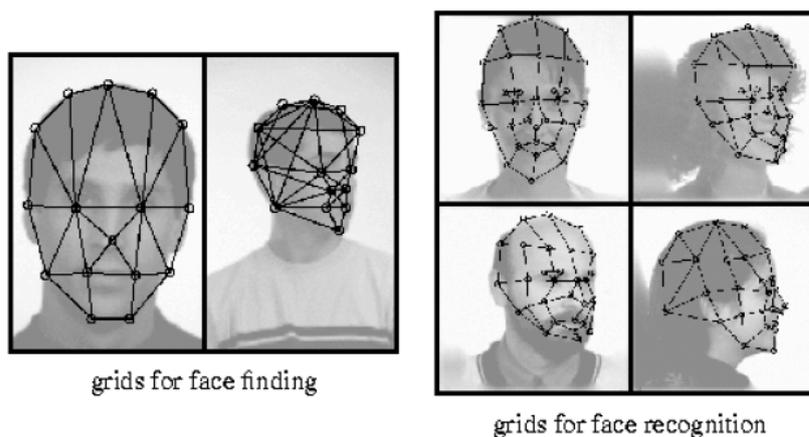


Figura 12: Exemplo de Grafos genéricos e específicos para face. Figura retirada de (WISKOTT et al., 1997).

3 *Foveação (Foveation)*

Em Bernardino (2004) pode-se ver, que tanto o sistema natural quanto o artificial têm que lidar com grandes quantidades de informação advinda do ambiente ao seu redor. Quando uma operação de tempo real é requerida, como acontece com animais e robôs, a aquisição e processamento da imagem devem ser feitos o mais rápido possível, sendo para isso necessário algum tipo de compactação na imagem. Porém, ao comprimi-la surgem dois problemas: perda de resolução e/ou diminuição do campo visual.

Alguns animais resolvem esse problema através da *Foveação (Foveation)*, termo proveniente de Fóvea (área de alta resolução da retina de muitos animais). Foveação é a forma como células gangliais estão organizadas para coletar informações da retina. Ela objetiva promover uma compactação da informação a ser transmitida para o córtex. Em tais animais, a alta acuidade visual somente existe em uma pequena região angular no centro do campo visual. Apesar da compactação obtida por esse método, é preservado grande parte do campo visual e a alta resolução da periferia (BERNARDINO, 2004).

Para a modelagem computacional da operação de foveação é necessário a definição de um modelo de retina, sendo este dependente dos campos receptivos que modelam suas células gangliais e da forma como eles são organizados. A próxima seção apresenta a definição dos campos receptivos utilizados neste trabalho, enquanto a posterior, a forma como eles são organizados.

3.1 **Campo Receptivo da Retina**

Segundo Bernardino (2004), o termo *Campo Receptivo (RF)* surgiu depois que experimentos de neurofisiologia acharam células visuais sensíveis a estímulos somente em uma determinada área do campo visual. Descobriu-se também, que nem todas as partes do RF contribuem igualmente para a resposta das células, tendo portanto, um perfil descrevendo seu comportamento.

Para modelá-los introduziu-se o conceito de *Funções de Perfil (FP)*. Elas têm uma

limitada área espacial, onde o RF produz um estímulo; e um centro bem definido, onde o RF produz um estímulo máximo. Um melhor esclarecimento desses conceitos, pode ser encontrado em (EDELMAN, 1995; VALOIS; VALOIS, 1990).

Computacionalmente falando, a resposta de células visuais com um RF definido pode ser reproduzido aplicando-se uma convolução entre a FP do RF e um ponto (x, y) de uma imagem I . O RF funciona como um filtro que tem como objetivo extrair informações de uma determinada região da imagem (VALOIS; VALOIS, 1990).

3.1.0.1 Tipos de RFs

De acordo com a literatura, os RFs, computacionalmente modelados, podem assumir diferentes formas, como por exemplo retangular (BOLDUC; LEVINE, 1998) e circular (BOLDUC; LEVINE, 1998; BERNARDINO, 2004). Assim como, seguir diferentes FP's (BOLDUC; LEVINE, 1997; VALOIS; VALOIS, 1990): Uniforme, Gaussiana, Diferença-de-Gaussianas, Gabor, etc, podendo portanto, cada um deles responder de forma diferente para uma mesma entrada. Porém, em virtude da analogia biológica e dessa seção tratar dos RFs presentes no nível da retina, será descrito apenas como modelar RFs circulares seguindo a FP *Diferença-de-Gaussianas*.

3.1.0.2 RFs Circulares

Eles são ditos circulares, porque somente os pixels contidos em um determinado círculo contribuem para a resposta do RF. Os limites circulares são definidos por:

$$\phi(x, y) = \begin{cases} \phi(x, y) & \text{se } (x^2 + y^2) \leq r^2 \\ 0 & \text{se } (x^2 + y^2) > r^2 \end{cases} \quad (3.1)$$

onde, r será o raio a partir do qual os pixels contribuem com peso zero, independente da FP ϕ escolhida.

3.1.1 Função de Perfil Diferença-de-Gaussianas

A motivação para o uso da Diferença-de-Gaussianas (DOG) como FP, é modelar as *Células Gangliais* existentes no sistema visual humano, como pode ser visto em (VALOIS; VALOIS, 1990; BERNARDINO, 2004; BOLDUC; LEVINE, 1997; BALASURIYA; SIEBERT, 2003; HANSEN; BARATOFF; NEUMANN, 2000; RULLEN; THORPE, 2001).

Em (BERNARDINO, 2004), é dada uma descrição dessas células. Elas monitoram RFs

e constituem a saída da retina para o cérebro através do nervo ótico. Seus RFs têm um formato circular e uma estrutura *centro-periferia* (*center-surround*). A estrutura recebe essa denominação por possuir uma região central que ativa/inibe a saída e uma região envoltória com papel oposto ao do centro. Células que são ativadas com luz no centro e sombra fora do centro são chamadas de *Célula ON* e as células que respondem na situação oposta são chamadas de *Célula OFF*, ver Figura 13.

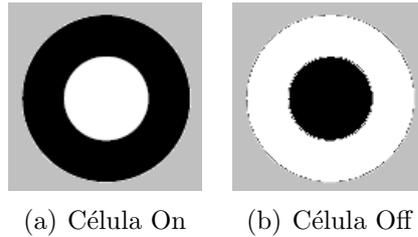


Figura 13: Centro-periferia (Center-Sorround).

Esse formato permite que elas computem o contraste entre o centro e a periferia com robustez quanto à variação de luminosidade.

A Figura 14 ilustra como a DOG modela uma célula ganglial.

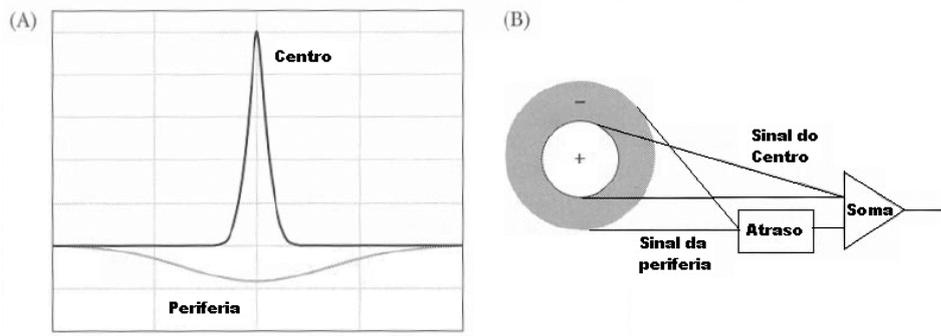


Figura 14: Modelagem Centro-periferia (Center-Sorround). (A) A resposta de uma célula ganglial linear no domínio do espaço (radialmente simétrica) (B) domínio temporal. Figura extraída de (BERNARDINO, 2004).

Uma DOG pode ser descrita por:

$$\phi(x, y) = A_1 e^{-\frac{(x^2+y^2)}{\sigma_1^2}} - A_2 e^{-\frac{(x^2+y^2)}{\sigma_2^2}}. \quad (3.2)$$

onde, as constantes A_1 , A_2 , σ_1 e σ_2 são escolhidas, como sugerido em (BOLDUC; LEVINE, 1997), de forma a garantir que exista um ponto $r' = \sqrt{(x'^2 + y'^2)}$ em que a função é zero (Figura 15). Além disso, elas devem garantir que r' seja aproximadamente metade do tamanho do RF ($\frac{r_i}{2}$) e que o somatório da parte positiva e da parte negativa seja unitário.

Essa segunda condição elimina a componente DC do RF, impedindo que ele responda de forma diferente para regiões de mesmo contraste e com quantidades de luz diferente.

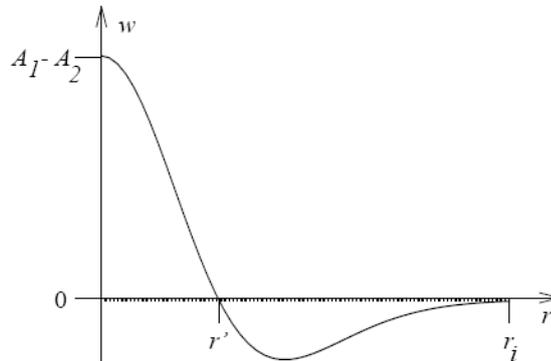


Figura 15: Diferença de Gaussianas - DOG. O raio do RF é dado por r_i e a função cruza o zero em r' . Figura extraída de (BOLDUC; LEVINE, 1997).

A definição dessas constantes foi feita como mostrado abaixo, dado raio r_i do RF:

- $\sigma_1 = \frac{r_i}{3}$, esse valor foi escolhido de forma a suavizar a transição nas bordas do RF, isto é, os valores na extremidades estejam próximos de zero;
- $\sigma_2 = \frac{\sigma_1}{1,5}$;
- Então é feito um pós processamento para obter as constantes A_1 e A_2 de acordo com as condições descritas acima.

A Figura 16 ilustra o exemplo de alguns RFs do tipo DOG com seus respectivos parâmetros de criação.

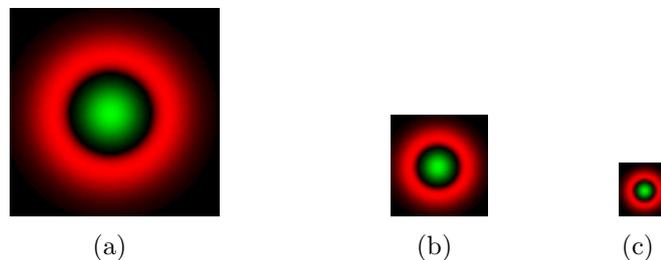


Figura 16: Exemplo de alguns RFs do tipo DOG. (a) $r_i = 40$ pixels. (b) $r_i = 20$ pixels. (c) $r_i = 10$ pixels. Quanto mais verde mais positivo. Quanto mais vermelho mais negativo.

3.2 Tipos de Distribuição

O *tipo de foveação* é que determina a forma como os campos receptivos serão espalhados ao longo da retina. Uma boa síntese das restrições quanto ao tamanho e distribuição dos RFs conhecidos pela psicofísica, neuro-anatomia e eletro-fisiologia podem ser achadas em (KOENDERINK; DOORN, 1978) apud (BERNARDINO, 2004) e é descrita abaixo:

- O diâmetro do menor RF é proporcional à excentricidade;
- Em qualquer excentricidade, todos os diâmetros maiores que a menor unidade correspondente estarão presentes;
- O tamanho médio do RF cresce linearmente com a excentricidade;
- O mapeamento do campo visual para o córtex é logarítmico e o córtex visual parece ser homogêneo;
- Em qualquer local da retina, RFs de muitos tamanhos estão presentes, porém os menores RFs estão localizados mais centralmente;
- A sobreposição relativa dos RFs é independente da excentricidade.

O tipo de organização dos RFs é que determina a quantidade de compactação a ser atingida pela foveação. Um exemplo disso pode ser encontrado em (BOLDUC; LEVINE, 1998, 1997), onde utiliza-se uma variável de resolução, ou *espaço variante* para diminuir a resolução seguindo alguma função (como $\text{Log}(z)$, $\text{Log}(z + c)$, $f(z)$), à medida que se afastam da fóvea, promovendo assim, uma redução dos dados a serem armazenados.

O nome *espaço variante* vem dessa variação proporcional à distância que ocorre na imagem. A Figura 17 ilustra esse conceito baseado na função de resolução $\log(z)$. A região de alta resolução, ou fóvea, pode ser vista na Figura 17(c). A região de baixa resolução na periferia, utiliza o sistema log-polar de coordenadas e pode ser visto na Figura 17(d). Pode-se notar por esse exemplo, uma clara redução no número de pixels da imagem da imagem original Figura 17(a), que caiu de 234.256 para 11.760, em relação a imagem reduzida Figura 17(b), também chamada de imagem da retina.



(a) imagem de entrada (484x484 pixels).



(b) imagem reduzida (ou imagem da retina), 20 vezes menor do que a imagem original (a).



(c) fóvea aumentada (50 pixels de raio).



(d) periferia aumentada (31x126 pixels).

Figura 17: Exemplo da redução de dados espaço variante. Esse exemplo é baseado no sistema de coordenadas log-polar. Figura extraída de (BOLDUC; LEVINE, 1998).

Encontram-se na literatura vários métodos para fazer-se a foveação de uma imagem, como por exemplo: Pulse-Coupled Neural Network (KINSER, 1999), Hermite Foveation (KRYLOV; KORTCHAGINE, 2004), Wavelet Foveation (CHANG; MALLAT; YAP, 1999; CHANG; YAP, 1997), Retina Auto-Organizada (BALASURIYA; SIEBERT, 2003, 2005) e Transformação Log-Polar (WALLACE et al., 1994; BERNARDINO, 2004; BOLDUC; LEVINE, 1997, 1998). Contudo, dado à motivação biológica e à quantidade de estudos encontrados na literatura, apenas o último será utilizado nesse trabalho.

3.2.1 Transformação Log-Polar

A *Transformação Log-Polar*, também conhecida como mapeamento logpolar, ou mapeamento $\log(z)$, é um método de foveação que pode possuir certas variações, porém sempre seguindo seus conceitos básicos. De acordo com (BERNARDINO, 2004), algumas de suas propriedades principais são:

- **Plausibilidade Biológica:** é baseado na distribuição dos RFs e no mapeamento retinocortical do sistema visual de mamíferos;
- **Mapeamento da Imagem:** os coeficientes de transformação foveal são arranjados

em uma imagem 2D (chamada de *imagem logpolar*) preservando a relação com os vizinhos, quase completamente, em relação a imagem original;

- **Invariante quanto a rotação e escala:** quando a imagem original é rotacionada ou escalada em relação ao centro (Fóvea), os padrões na imagem logpolar sofrerão apenas translações, preservando portanto o formato original.

Seundo Schwartz (1977), a região externa da retina de um primata é formada por anéis com aproximadamente o mesmo número de RFs. E a distância deles até o centro da retina pode ser expressa em termos de uma função exponencial. Analogamente, o mapeamento computacional pode ser obtido da seguinte forma:

$$w = \ln(z) \quad (3.3)$$

onde, w representa um ponto no plano cortical, e z representa um ponto no plano visual. Eles também podem ser representados como:

$$z = \rho e^{j\theta} \quad (3.4)$$

A Equação 3.3 pode ser escrita com variáveis reais:

$$\begin{aligned} x &= \ln(\rho) \\ y &= \theta \end{aligned} \quad (3.5)$$

onde, o ponto (x, y) está no plano cortical e o ponto (ρ, θ) está no plano polar visual.

Como apresentado em (BOLDUC; LEVINE, 1998) concordando com (SCHWARTZ, 1977), a Figura 18(a) representa o template do mapeamento baseado na função $\log(z)$ dividindo a imagem de entrada em região da fóvea e região da periferia.

A região da periferia consiste de um grupo de círculos concêntricos e com incrementos radiais exponenciais. Os círculos são divididos em setores equivalentes a um RF. Como pode ser visto na Figura 18(c), os RFs da periferia são arranjados em um grid retangular de acordo com suas coordenadas logpolar: um eixo representando o ângulo e o outro o logaritmo da componente radial.

A região da fóvea vista na Figura 18(b), só existe por razões práticas. Uma delas é evitar a singularidade da função $\log(z)$ em $z = 0$, a outra é evitar RFs de tamanho menor do que um pixel, pois não teria como representá-lo computacionalmente. Dentro dessa região os pixels permanecem da mesma forma que na imagem original, ou seja, o mapeamento é feito um para um.

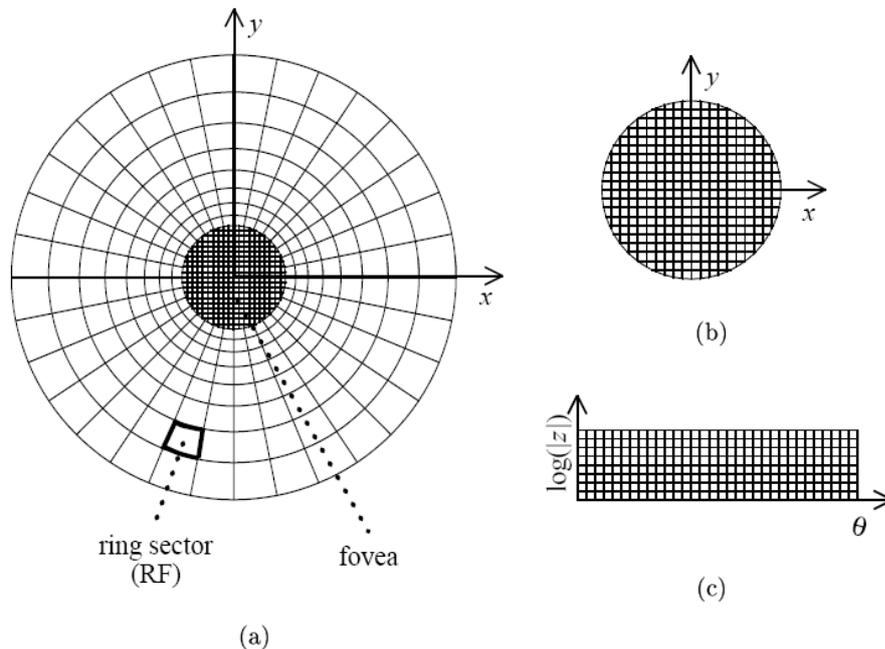


Figura 18: Modelo $\log(z)$. (a) template do mapeamento do domínio de entrada, (b) grid de saída da fóvea, (c) grid de saída da periferia. Cada pixel da periferia pertence a um setor denominado RF. Figura retirada de (BOLDUC; LEVINE, 1998)

Contudo, quando a foveação é feita como apresentado anteriormente, ela se distancia do modelo biológico ao considerar RFs de formato retangular e sem sobreposição. Pois como pode ser visto em (BOLDUC; LEVINE, 1998), uma mesma região do campo visual contribui para mais de um RF indicando haver um certo grau de sobreposição entre eles. Visando diminuir essa distância optou-se por uma de suas variações denominada de Transformação Logpolar Suave (*Smooth Logpolar Transform*), apresentada em (BERNARDINO, 2004).

3.2.2 Transformação Log-Polar Suave

A Transformação Log-Polar Suave utiliza RFs circulares e com um certo grau de sobreposição, assim como alguns modelos apresentados em (BOLDUC; LEVINE, 1998). Ela promove uma transição suave da região da fóvea para periferia, isto é, a distância e o tamanho dos RFs que compõem esse modelo de retina, são calculados de forma a manterem uma sobreposição constante. Tanto na direção radial, quanto na direção angular.

Para atingir esse objetivo, seguiu-se a abordagem adotada em (BERNARDINO, 2004), dividindo-se sua construção em três partes: distribuição radial suave, distribuição angular balanceada e tamanho dos RFs. As duas primeiras são independentes do RF utilizado e

apresentam o modo de organização dos RFs ao longo da retina. Já a terceira, apresenta uma forma de calcular o tamanho do RF de acordo com a organização proposta pelas duas primeiras.

1. **Distribuição radial suave, (BERNARDINO, 2004):** A definição de distribuição radial levará em conta duas variáveis de design, o raio da retina r_{retina} e o raio da fóvea r_{fovea} . Assumiu-se densidade uniforme na fóvea com um RF por pixel, e uma distribuição log-polar na periferia. Dado que ρ_i é a coordenada radial do i_{esimo} anel do centro para a periferia, então os valores de ρ_i podem ser definidos recursivamente por:

$$\rho_i = \begin{cases} \rho_{i-1} + 1 & \text{se } i \leq r_{fovea} \\ k\rho_{i-1} & \text{se } i > r_{fovea} \end{cases} \quad (3.6)$$

onde, $\rho_0 = 0$ e k é uma constante a definir.

Para forçar uma transição suave da fóvea para a periferia, fará-se a primeira coordenada radial da periferia ser igual a última coordenada radial da fóvea, como descrito abaixo:

$$\rho_{r_{fovea}+1} = \rho_{r_{fovea}} + 1 \quad (3.7)$$

$$\rho_{r_{fovea}+1} = k\rho_{r_{fovea}} \quad (3.8)$$

que resulta em:

$$k = \frac{r_{fovea} + 1}{r_{fovea}} \quad (3.9)$$

Reescrevendo a expressão de ρ_i :

$$\rho_i = \begin{cases} i & \text{se } i = 0, \dots, r_{fovea} \\ r_{fovea}k^{i-r_{fovea}} & \text{se } i > r_{fovea} \end{cases} \quad (3.10)$$

A representação gráfica para $f = 9$ é mostrada na Figura 19. Nota-se a transição suave da região da fóvea para a periferia.

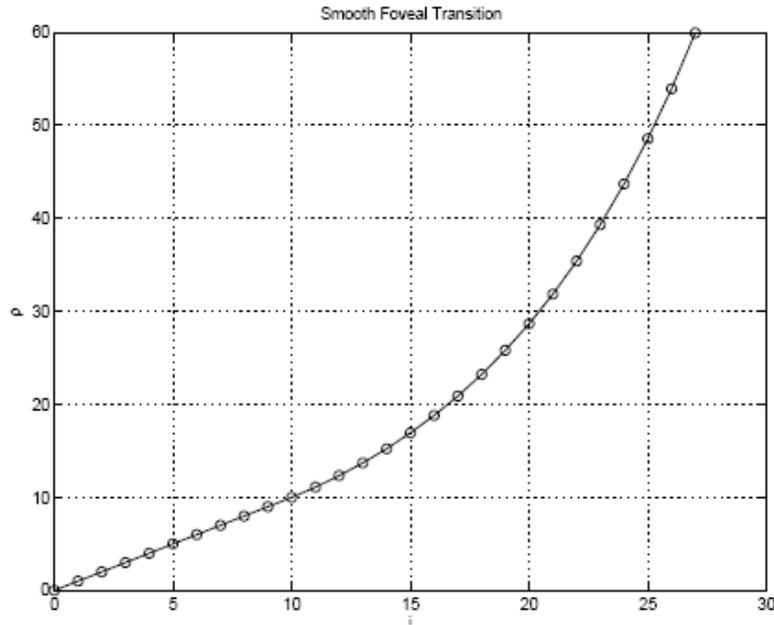


Figura 19: Distribuição radial com transição suave para $f = 9$. Figura retirada de (BERNARDINO, 2004).

2. **Distribuição angular balanceada, (BERNARDINO, 2004):** A distribuição angular é definida de forma a ter a mesma distância de seu vizinho angular que tem de seu vizinho radial. Esta definição só é relevante na região da periferia. De acordo com a Equação 3.10, a distância radial média entre RFs vizinhos da excentricidade i é dada por:

$$\Delta_r = \frac{\rho_{i+1} - \rho_{i-1}}{2} = \rho_i \frac{k - k^{-1}}{2} \quad (3.11)$$

Na direção angular a distância entre dois RFs vizinhos em uma mesma excentricidade i é dada por:

$$\Delta_a = 2\rho_i \sin \frac{\Delta_\theta}{2} \quad (3.12)$$

onde, Δ_θ é o passo angular.

Igualando as duas distâncias anteriores, Equações 3.11 e 3.12, obten-se:

$$\Delta_\theta = 2 \arcsin \frac{k - k^{-1}}{4} \quad (3.13)$$

A Figura 20(Esquerda) mostra o resultado da distribuição balanceada de RFs. Em casos práticos é útil ter os RFs localizados em pixels de valores inteiros sendo portanto necessário convertê-los. Na Figura 20(Direita) também é possível ver o resultado dessa conversão.

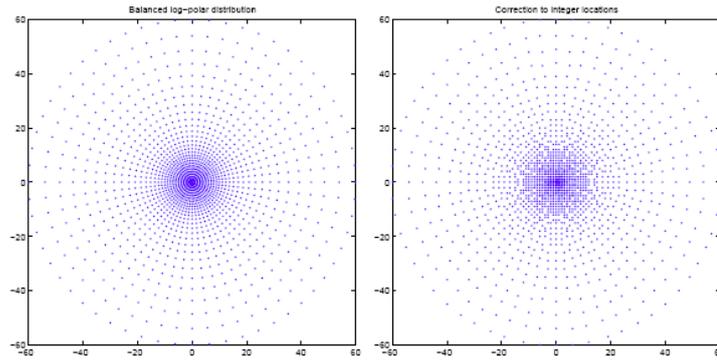


Figura 20: Distribuição de RFs pela retina com 128 pixels de diâmetro e fóvea com 8 pixels de raio. (Esquerda) Apresenta as posições em valores reais. (Direita) Apresenta as posições em valores inteiros. Figura retirada de (BERNARDINO, 2004).

3. **Tamanho dos RFs, (BERNARDINO, 2004):** Os tamanhos dos RFs são definidos a partir do tamanho dos RFs presentes na fóvea. Eles são calculados de forma a manter uma mesma porcentagem de sobreposição independente de sua excentricidade, seguindo a equação abaixo:

$$r_i = \begin{cases} r_0 & \text{se } i = 0, \dots, r_{fovea} \\ r_0 \rho_i^{\frac{k-k^{-1}}{2}} & \text{se } i > r_{fovea} \end{cases} \quad (3.14)$$

onde, r_i é o raio dos kernels da excentricidade i , e r_0 é o raio dos kernels presentes na fóvea.

3.3 Definição da Retina

Após a escolha do tipo de RF (DOG), feito na seção 3.1.1, e do tipo de distribuição (Transformação Log-Polar Suave), feito na seção 3.2.2, é possível definir uma retina RET a partir de três parâmetros de entrada: o raio da retina r_{retina} , o raio da fóvea r_{fovea} e o tamanho dos RFs presentes na fóvea r_0 .

Escolhidos os parâmetros, basta aplicar as regras definidas para a geração dos RFs e para geração do tipo de distribuição, a fim de se obter uma estrutura como à abaixo.

$$RET(r_{retina}, r_{fovea}, r_0) = \{(x_i, y_i), RF_i\} \quad \text{para } i = 1, \dots, N \quad (3.15)$$

onde, (x_i, y_i) é a posição cartesiana relativa dos RFs sendo $(x_1, y_1) = (0, 0)$ e localizada no centro da retina, RF_i é o RF referente a posição (x_i, y_i) , cujo o tamanho varia de acordo com a excentricidade, ou seja, distância do ponto $i = 1$, e N é o número de RFs presentes na retina.

3.3.1 Aplicando a Foveação

A convolução da imagem I no ponto (x_0, y_0) com o RF RF_i é definida como:

$$c(I(x_0, y_0), RF_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) RF_i(x - x_0, y - y_0) dx dy \quad (3.16)$$

Para aplicar-se a foveação com uma retina RET e o foco em um ponto (x_f, y_f) da imagem I , basta convoluir a imagem I com cada RF presente na retina em sua respectiva posição, como definido abaixo:

$$FOV(RET, I(x_f, y_f)) = \{(x_i, y_i), RF_i, v_i = c(I(x_f + x_i, y_f + y_i), RF_i)\} \text{ para } i = 1, \dots, N \quad (3.17)$$

A Figura 21 ilustra um exemplo de uma retina com foco de atenção no nariz. Com ela é possível ter uma representação visual da aplicação da foveação, pois a equação 3.17 representa a convolução de cada RF centrado em sua respectiva posição representadas pelos pontos verdes na imagem.



Figura 21: Exemplo dos pontos da retina com o foco de atenção no nariz. Os pontos verdes representam os centros de cada RF de uma retina com $r_{retina} = 26$, $r_{fovea} = 5$ e $r_0 = 5$.

3.3.2 Região de Interesse

Geralmente, a foveação é aplicada a toda a imagem a ser convertida tendo sua atenção voltada para a região apontada pela fóvea. Com isso, ao voltar-se a atenção da imagem para o nariz de uma pessoa, todo o ambiente em volta seria transformado juntamente com ele e, assim que o foco de atenção mudasse, seria necessário transformar toda a imagem

novamente.

Dado que nesse trabalho nem toda a imagem será utilizada para a geração da informação necessária para a identificação do indivíduo/localização de características, surge a necessidade da definição de uma região de interesse. Essa é apenas uma porção de interesse na imagem a ser transformada como, por exemplo, uma região envolvida por um raio que engloba todo o nariz.

A partir dessa definição, é possível reduzir o tamanho do raio da retina r_{retina} de forma a representar somente as regiões de interesse e não toda a imagem.

4 Representação da Informação

Ignorando a profundidade, o mundo visual é bi-dimensional, assim como as imagens da retina. Imagens bi-dimensionais podem ser analisadas utilizando-se a transformada de Fourier bi-dimensional, que permite decompô-las em componentes senoidais de frequência espacial, amplitude, fase e orientação (VALOIS; VALOIS, 1990). As Equações 4.1 e 4.2 representam a transformada de Fourier e sua inversa, respectivamente:

$$F(\alpha, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi(\alpha x + \beta y)} dx dy \quad (4.1)$$

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\alpha, \beta) e^{j2\pi(\alpha x + \beta y)} d\alpha d\beta \quad (4.2)$$

onde, x e y são as coordenadas, e α e β são as frequências horizontais e verticais.

A Equação 4.1 pode ser expandida para separar a parte real da imaginária resultando em:

$$F(\alpha, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cos(2\pi(\alpha x + \beta y)) dx dy - j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \sin(2\pi(\alpha x + \beta y)) dx dy \quad (4.3)$$

A partir da Equação 4.3 é possível representar uma frequência em particular na imagem através de uma amplitude e uma fase. Porém, como descrito em (VALOIS; VALOIS, 1990), apesar de um filtro de onda senoidal ter uma alta precisão em relação à informação de frequência, ele quase não tem informação a respeito da localização espacial. Ou seja, um padrão de frequência poderia estar localizado em várias posições e ainda produzir a mesma saída no filtro. Esse problema foi examinado por Gabor em 1946. Ele mostrou que um filtro temporal com uma queda gaussiana minimizaria o produto da incerteza de tempo e frequência, desde que a gaussiana modelasse uma transformada de Fourier, daí a função de Gabor.

4.1 Gabor

A função *Gabor* permite descrever uma porção da imagem através de um vetor de respostas em diferentes frequências, orientações e fases. A equação abaixo representa uma função de Gabor em duas dimensões:

$$g(x, y) = e^{-\pi\sigma^2(x^2 + \frac{y^2}{\gamma})} e^{j(2\pi\lambda(x \cos(\theta) + y \sin(\theta)) + \varphi)} \quad (4.4)$$

onde, x e y definem a posição espacial, σ é a abertura da gaussiana, λ é o comprimento de onda (seno e cosseno), γ é a forma da gaussiana, φ é a fase da senoide, e θ é a orientação da função.

As Figuras 22, 23, 24, 25 e 26 ilustram a influência dos parâmetros sobre a função Gabor, assim como descrito em (BOLME, 2003).

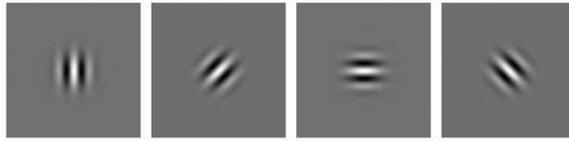


Figura 22: Exemplos de diferentes orientações da função Gabor. O parâmetro θ rotaciona a função em torno do seu centro. Valores de π a 2π são redundantes devido à simetria da função, para a parte real (cosseno) são idênticos e para a parte imaginária (seno) possuem o sinal invertido.

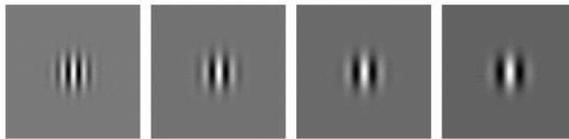


Figura 23: Exemplos de diferentes comprimentos de onda da função Gabor. O parâmetro λ modifica o comprimento de onda da senoide e por conseguinte a frequência. Valores menores respondem melhor à arestas finas e valores maiores respondem melhor à arestas grossas.

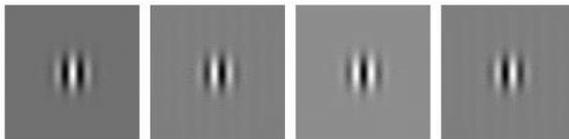


Figura 24: Exemplos de diferentes fases da função Gabor. O parâmetro φ modifica a fase da senoide. Como a função é complexa, ela é composta por uma parte real (representada pelo cosseno) e por uma parte imaginária (representada pelo seno). Portanto sua convolução gera dois valores defasados de $\frac{\pi}{2}$.

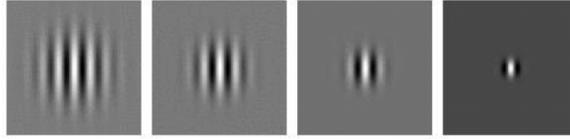


Figura 25: Exemplos de diferentes raios da gaussiana da função Gabor. O parâmetro σ modifica a abertura da gaussiana. Ele é responsável pela porção do kernel que afetará a imagem. Quanto maior a abertura da gaussiana, maior a influencia do kernel sobre a imagem. Ele geralmente é proporcional ao tamanho de λ .

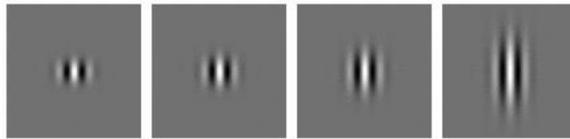


Figura 26: Exemplos de diferentes formas da gaussiana da função Gabor. O parâmetro γ modifica a forma bi-dimensional da gaussiana. Para o valor 1 a gaussiana fica com formato circular, quando vista de cima. Quanto mais o valor distancia-se de 1, mais ela toma a forma elíptica.

4.2 Gabor Wavelet

A tendência ao uso dessas funções no campo de visão computacional aumentou quando resultados de estudos aproximaram as respostas de células corticais às respostas das Gabors. Em (VALOIS; VALOIS, 1990), são apresentados dois modelos de córtex, um proveniente do estudo com gatos e outro do estudo com macacos, em que ambos armazenam informações provenientes da função Gabor em diferentes frequências e orientações.

Gabor Wavelets são um conjunto de kernels (ou filtros) gerados a partir de uma mesma função (Gabor) com a variação de alguns parâmetros. Elas têm como objetivo, extrair informações de frequência, orientação e fase de determinados locais da imagem, simulando assim, as informações armazenadas no córtex de alguns mamíferos, como dito acima. Existem na literatura muitos estudos, baseados nessa técnica (BOLME, 2003; WISKOTT et al., 1997; KEPENEKCI, 2001; WIEGHARDT; WÜRTZ; MALSBERG, 2002), aplicados à visão computacional, o que a consolida cada vez mais.

Como pode ser visto em (BOLME, 2003), é possível fazer uma analogia com Fourier e aplicar *Gabor Wavelets* em uma imagem a partir de convoluções. Para isso, basta:

$$c(I(x_0, y_0), G(\sigma, \lambda, \theta, \gamma)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y)g(x - x_0, y - y_0)dx dy \quad (4.5)$$

onde, $I(x_0, y_0)$ é a imagem a ser decomposta a partir do ponto (x_0, y_0) , e, $G(\sigma, \lambda, \theta, \gamma)$ é a *Gabor Wavelet* formada pela Equação 4.4 definida com os parâmetros: σ , λ , θ , γ e $\varphi = 0$.

Expandindo obtém-se:

$$c(I(x_0, y_0), G(\sigma, \lambda, \theta, \gamma)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) e^{-\pi\sigma^2((x-x_0)^2+(y-y_0)^2)} e^{2\pi j\lambda((x-x_0)\cos(\theta)+(y-y_0)\sin(\theta))} dx dy \quad (4.6)$$

$$\begin{aligned} c(I(x_0, y_0), G(\sigma, \lambda, \theta, \gamma)) = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) e^{-\pi\sigma^2((x-x_0)^2+(y-y_0)^2)} \\ & \cos(2\pi\lambda((x-x_0)\cos(\theta)+(y-y_0)\sin(\theta))) dx dy + \\ & j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) e^{-\pi\sigma^2((x-x_0)^2+(y-y_0)^2)} \\ & \sin(2\pi\lambda((x-x_0)\cos(\theta)+(y-y_0)\sin(\theta))) dx dy \end{aligned} \quad (4.7)$$

Assim como em Fourier, sua forma expandida também tem uma parte real a_{real} e outra imaginária a_{imag} :

$$c(I(x_0, y_0), G(\sigma, \lambda, \theta, \gamma)) = a_{real} + j a_{imag} \quad (4.8)$$

Essas podem ser representadas em coordenadas polares, por uma amplitude:

$$a = \sqrt{a_{real}^2 + a_{imag}^2} \quad (4.9)$$

E uma fase:

$$\phi = \begin{cases} \arctan\left(\frac{a_{imag}}{a_{real}}\right) & \text{se } a_{real} > 0 \\ \pi + \arctan\left(\frac{a_{imag}}{a_{real}}\right) & \text{se } a_{real} < 0 \\ \frac{\pi}{2} & \text{se } a_{real} = 0 \text{ e } a_{imag} \geq 0 \\ -\frac{\pi}{2} & \text{se } a_{real} = 0 \text{ e } a_{imag} < 0 \end{cases} \quad (4.10)$$

A Figura 27 ilustra algumas Wavelets formadas por Gabors com seus respectivos parâmetros para construção.

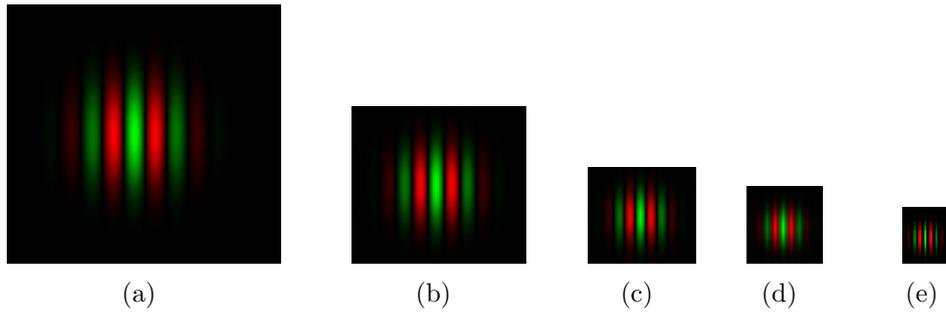


Figura 27: Exemplo de algumas Wavelets formadas por Gabor. (a) $\lambda = 16$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (b) $\lambda = 12$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (c) $\lambda = 8$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (d) $\lambda = 6$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (e) $\lambda = 4$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. Quanto mais verde mais positivo. Quanto mais vermelho mais negativo.

4.3 Componente DC

Em se tratando de imagens, a função 4.4 apresenta um problema. Em virtude dela carregar uma forte componente DC, não é possível obter-se uma invariância em relação à quantidade de luz, (MOVELLAN, 2002). Ou seja, frequências iguais, como por exemplo um plano preto e um plano branco, respondem de forma diferente. Para minimizar esse problema, ainda em (MOVELLAN, 2002) é apresentada uma forma de introduzir-se uma constante a fim de compensar essa componente DC. Sua nova forma é mostrada abaixo:

$$g(x, y) = e^{-\pi\sigma^2(x^2+y^2)}(e^{2\pi j\lambda(x \cos(\theta)+y \sin(\theta))} - e^{-\frac{\pi\lambda^2}{\sigma^2}}) \quad (4.11)$$

onde, x e y são a posição espacial, σ é o raio da gaussiana, λ é o comprimento de onda (seno e cosseno), θ é a orientação da função, e o segundo termo entre parênteses representa a correção da componente DC. Vale lembrar que foi atribuído 0 à fase φ e 1 à forma da gaussiana γ , já que esses parâmetros não serão modificados ao decorrer do trabalho.

4.4 Wavelets com Foveação

Como pode ser visto em (VALOIS; VALOIS, 1990), a retina envia sinais para outras partes do cérebro através do nervo ótico. Essa informação chega até o *Núcleo Geniculado Lateral* (LGN), que as projeta até o *Córtex Estriado* (Striate Cortex), onde ela é processada a fim de serem armazenadas em forma de orientações e frequências, como descrito anteriormente.

Inspirado nesse fluxo dos sinais biológicos, esse trabalho unificou as duas técnicas

(Foveação e Wavelets) com o intuito de obter a compactação da informação promovida pela foveação juntamente com a representatividade das Wavelets.

Entretanto, notou-se que a introdução da operação de foveação pode encarecer computacionalmente a extração dessas informações. Pois dependendo do tipo de retina escolhida, da quantidade de sobreposição dos RFs e do tamanho da região de interesse, o custo das convoluções que gerarão a saída da retina pode aumentar. Todavia, a quantidade de informação advinda da retina a ser processada posteriormente torna-se bem menor acelerando sua manipulação, pois as convoluções para a geração das informações do córtex seriam realizadas com um número de informação (pixels no caso sem foveação e saída da retina no caso com foveação) muito maior caso não houvesse a compactação obtida pela foveação.

Sendo assim, quanto maior a quantidade de pós-processamento necessário, isto é, quanto maior o número de wavelets a serem aplicadas, menor torna-se o custo percentual da operação de foveação. Esse processamento é aumentado à medida que se deseja extrair mais informação a respeito de uma região da imagem. Como por exemplo informação de rotação, na qual a informação teria que ser extraída em várias direções, para um mesmo ponto na imagem.

Na próxima seção, será descrito como esse processamento a posteriori é feito para gerar informações próximas às obtidas pelas Gabors Wavelets.

4.4.1 Processamento com Gabors

Após aplicar a foveação em um ponto (x, y) da imagem, serão gerados uma série de valores contidos em FOV (definidos na Equação 3.17) referentes a cada ponto da retina. Para aplicar uma Gabor Wavelet à saída da retina, basta multiplicar os valores de saída da foveação pelo seu peso correspondente na função Gabor, como abaixo:

$$GF(FOV(RET, I(x_f, y_f)), G(\sigma, \lambda, \theta, \gamma)) = \sum_{i=1}^N v_i g(x_i, y_i) \quad (4.12)$$

Expandindo:

$$GF(FOV(RET, I(x_f, y_f)), G(\sigma, \lambda, \theta, \gamma)) = \sum_{i=1}^N v_i e^{-\pi\sigma^2(x_i^2+y_i^2)} \cos(2\pi\lambda(x_i \cos(\theta) + y_i \sin(\theta))) + j \sum_{i=1}^N v_i e^{-\pi\sigma^2(x_i^2+y_i^2)} \sin(2\pi\lambda(x_i \cos(\theta) + y_i \sin(\theta))) \quad (4.13)$$

onde, $FOV(RET, I(x_f, y_f))$ é uma foveação aplicada com o foco em um ponto (x_f, y_f) da imagem I , $G(\sigma, \lambda, \theta, \gamma)$ é a *Gabor Wavelet* formada pela Equação 4.4 definida com os parâmetros: σ , λ , θ , γ e $\varphi = 0$, e v_i é o valor referente à posição (x_i, y_i) da retina.

Que pode ser escrita na forma:

$$GF(FOV(RET, I(x_f, y_f)), G(\sigma, \lambda, \theta, \gamma)) = a_{real} + ja_{imag} \quad (4.14)$$

onde, a_{real} representa a parte real e a_{imag} representa a parte imaginária.

Essas também podem ser representadas em coordenadas polares, por uma amplitude e uma fase, calculadas a partir das Equações 4.9 e 4.10.

Vale lembrar que a compensação da componente DC já foi feita na geração dos RFs da retina não sendo necessário fazê-la novamente nesse processamento.

A Figura 28 ilustra um exemplo dos valores a serem multiplicados com informação extraída pela retina, como definido na Equação 4.12. A partir dela é possível visualizar a redução na quantidade de pontos a serem processados para a geração das Gabor Wavelets.

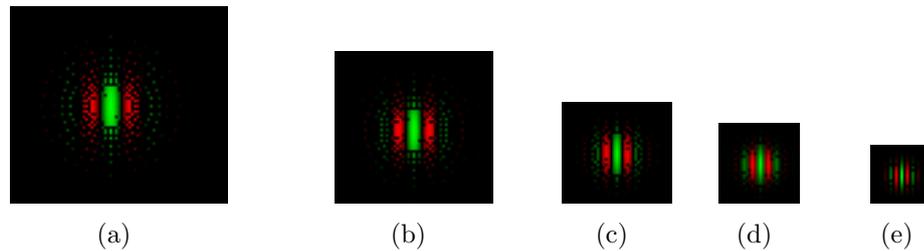


Figura 28: Exemplo do pós-processamento realizado na informação da retina. A retina utilizada foi gerada com 26 pixels de raio, uma fóvea de raio 5 pixels e com os kernels presentes na fóvea tendo 5 pixels de raio. (a) $\lambda = 16$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (b) $\lambda = 12$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (c) $\lambda = 8$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (d) $\lambda = 6$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (e) $\lambda = 4$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. Quanto mais verde mais positivo. Quanto mais vermelho mais negativo.

A Figura 29 ilustra um exemplo do resultado final obtido pela Equação 4.12.

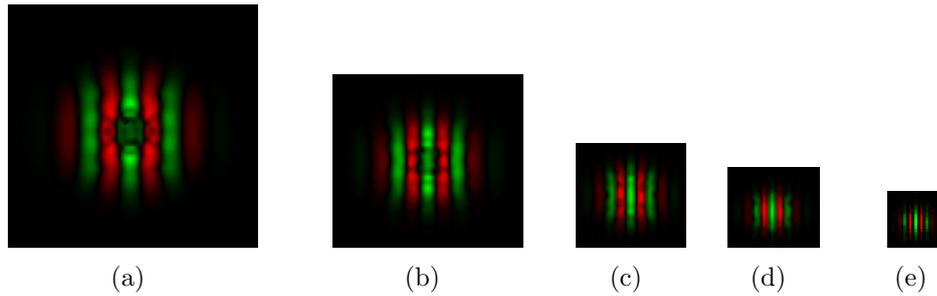


Figura 29: Exemplo de algumas Wavelets formadas por Gabor com Foveação. A retina utilizada foi gerada com 26 pixels de raio, uma fóvea de raio 5 pixels e com os kernels presentes na fóvea tendo 5 pixels de raio. (a) $\lambda = 16$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (b) $\lambda = 12$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (c) $\lambda = 8$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (d) $\lambda = 6$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. (e) $\lambda = 4$ pixels, $\theta = 0^\circ$, $\sigma = \lambda$ e $\varphi = 0$. Quanto mais verde mais positivo. Quanto mais vermelho mais negativo.

É possível perceber uma distorção na região da fóvea das Wavelets geradas a partir da foveação em relação às geradas diretamente por Gabor (Figura 27). Essa distorção é amplificada à medida que λ aumenta, isto é, a frequência diminui. Após a realização dos testes de reconhecimento será possível avaliar a influência dela no sistema.

4.5 Jets

O conceito de *Jet*, apresentado por Wiskott et al. (1997), é definido como sendo o conjunto de respostas de um grupo de Gabor Wavelets aplicado a um ponto de uma imagem. Para montá-lo, convolui-se cada Wavelet com a imagem a partir de um ponto (x, y) . Cada par (real, imaginário) obtido pelas convoluções gerará uma amplitude a_j e uma fase ϕ_j , como visto na seção 4.2. Esses novos pares, (a_j, ϕ_j) com $j = 1, \dots, N$, descreverão as informações de frequência em torno do ponto (x, y) . A equação abaixo descreve a forma de um Jet:

$$J = \{(x, y), (a_j, \phi_j)\} \quad \text{para } j = 1, \dots, N \quad (4.15)$$

onde, (x, y) é o ponto a partir do qual as convoluções foram aplicadas, (a_j, ϕ_j) é o par amplitude e fase obtidos através da J_{esima} Wavelet, e N é o número de Wavelets definidas para o Jet J .

É importante citar que o conceito de Jet é independente da Gabor ser gerada a partir de Foveação ou não, dado que as Wavelets geram o mesmo tipo de saída.

4.6 Similaridade entre Dois Jets

Medir a similaridade entre dois Jets é fundamental para o processo de comparação de partes de duas imagens. Com uma medida de similaridade definida, é possível dizer o quanto uma região em torno do ponto (x_m, y_m) de uma imagem I_i se parece com uma região em torno do ponto (x_n, y_n) de uma imagem I_j , segundo esta medida.

De acordo com Wiskott et al. (1997), Bolme (2003), essa comparação pode ser feita de várias formas: considerando somente a magnitude; considerando a magnitude e a fase; ou considerando a magnitude, a fase e um deslocamento estimado através da maximização da função de similaridade (esse requer o modelo proposto de Grafo). Porém, na abordagem adotada aqui, a similaridade é dependente da amplitude e da fase, de acordo com a equação abaixo:

$$S(J_1, J_2) = \frac{\sum_{j=1}^N a_{1j} a_{2j} \cos(\phi_{1j} - \phi_{2j})}{\sqrt{\sum_{j=1}^N a_{1j}^2 \sum_{j=1}^N a_{2j}^2}} \quad (4.16)$$

onde, J_1 e J_2 são os Jets a serem comparados. N é o número de Wavelets que compõem cada um dos Jets, a_{1j} e a_{2j} são as amplitudes referentes à J_{esima} Wavelet de J_1 e J_2 , respectivamente, e ϕ_{1j} e ϕ_{2j} são as fases referentes à J_{esima} Wavelet de J_1 e J_2 , respectivamente. Essa função computa um valor de similaridade entre -1 e 1 baseado na semelhança entre as amplitudes de cada Wavelet e ponderada pela semelhança entre as fases.

Portanto, quanto mais próximos as amplitudes e as fases comparadas, mais próxima de 1 será a saída.

5 *Identificação do Indivíduo*

A identificação facial é a parte do sistema responsável por atribuir uma identidade a uma face, isto é, ela deve receber uma face e associá-la, ou não, a alguma outra conhecida. Como foi visto na seção 2, existem muitas maneiras diferentes de fazer isso. O sistema aqui apresentado segue uma abordagem parecida com a de Wiskott et al. (1997) utilizando Jets (seção 4.5) para a descrição de certas regiões da face.

O processo de identificação é dividido em duas partes: uma de treinamento do sistema e outra de identificação da pessoa contida na imagem apresentada. A primeira, consiste na representação interna de cada indivíduo conhecido pelo sistema através da extração de informações da face de cada um deles. Já a segunda, consiste em representar a face contida na imagem a ser reconhecida através do mesmo tipo de informação obtida dos indivíduos conhecidos, para que essa possa ser comparada com cada uma das criadas durante o treinamento. A comparação é feita através de uma função que mede a similaridade entre elas.

5.1 **Representação Facial de um Indivíduo**

Assim como em (WISKOTT et al., 1997), cada face será representada através de Jets (Equação 4.15) localizados em pontos específicos da face. Porém, aqui os pontos não são encontrados a partir de grafos, mas sim de três listas relativas a uma das seguintes características: nariz, olho esquerdo e boca.

A escolha dessas três características foi feita seguindo os resultados obtidos por Souza (2005) que verificou serem bons pontos de partida para a descrição de uma face. A opção de utilizar um olho ao invés dos dois foi baseada na similaridade existente na face humana. Já a decisão entre o esquerdo e o direito foi aleatória em virtude da simetria entre eles.

Cada lista é composta por um grupo de pontos relativos a uma das características citadas, isto é, cada um deles tem sua posição calculada de acordo com a posição do nariz, olho esquerdo ou boca. Eles foram manualmente localizados utilizando uma imagem,

randomicamente selecionada na base de imagens, para ter uma idéia da posição do ponto em cada face. A seleção foi feita de forma que eles coincidisse com regiões da face que carregam informações relevantes, como por exemplo pontos abrangendo sobrancelha, proximidades do olho, em torno do nariz e da boca. Evitou-se a escolha de pontos próximos ao fundo da imagem a fim de que o método não fosse prejudicado ao variar o plano de fundo. A Figura 30 ilustra os selecionados para cada grupo.



Figura 30: Localização dos pontos faciais. Os pontos (+) vermelhos representam os pontos relativos ao nariz (*). Os pontos (+) verdes representam os pontos relativos ao olho esquerdo (*). Os pontos (+) azuis representam os pontos relativos à boca (*). Nos testes realizados verificou-se que com 17 pontos selecionados para cada característica, o que dá um total de 51 pontos, é suficiente para uma taxa de reconhecimento definida como boa (maior do que 80%).

A Figura 31 ilustra pontos faciais em algumas imagens, cujo a configuração foi definida a partir de outra imagem. As apresentadas abaixo utilizaram a configuração definida na Figura 30.



Figura 31: Exemplo da distribuição dos pontos faciais em outras imagens. Ilustração da distribuição dos pontos faciais, definidos a partir da Figura 30, em outras imagens da base. A Figura 30 foi selecionada randomicamente para a marcação dos pontos faciais. Os pontos (+) vermelhos representam os pontos relativos ao nariz (*). Os pontos (+) verdes representam os pontos relativos ao olho esquerdo (*). Os pontos (+) azuis representam os pontos relativos à boca (*).

Após a geração desses grupos de pontos, a representação de uma face pode ser obtida

extraindo-se um Jet (Equação 4.15) da posição real (posição do ponto na imagem em questão após a localização da característica referente à sua lista) de cada ponto contido em uma das três lista. A posição real é calculada somando-se a posição relativa do ponto à posição de sua característica representante na imagem em questão. Para exemplificar essa situação, imagine um nariz tendo a localização $(x_n, y_n) = (100, 150)$ em uma imagem I , e um ponto contido na lista relativa ao nariz tendo a localização $(x_2, y_2) = (10, -5)$. Isso resultará em uma posição real de $(x_{2r}, y_{2r}) = (110, 145)$ para o ponto em questão. A equação abaixo descreve a representação de uma face.

$$FACE = \{J_p\} \quad \text{para } p = 1, \dots, P \quad (5.1)$$

onde, J_p é o Jet (Equação 4.15) representante do ponto p , calculado a partir de uma das três listas, aplicado na imagem I , e P é o número de pontos contidos nas três listas.

A dependência dessas características para o cálculo da posição real dos pontos, torna também dependente, a identificação facial da localização desses três atributos. Sendo assim, além de receber a face como entrada, também será necessária a passagem das coordenadas (x, y) do *olho esquerdo*, do *nariz* e da *boca*. Portanto, seja manual ou automático como o apresentado na próxima seção (seção 6), algum método deverá ser utilizado para localizá-los.

Em virtude dessa dependência, qualquer variação na posição das características causa uma variação de igual tamanho na posição real dos pontos relativos. Isso, torna a representação interna de um indivíduo muito sensível a um possível erro na localização das características, ou seja, caso uma característica fosse localizada próxima, mas não na posição correta todos os pontos relativos a ela também estariam deslocados. Visando diminuir essa sensibilidade, ao invés de amostrar apenas o ponto relativo em sua posição real, amostrou-se um conjunto de pontos ao seu redor.

Para isso, criou-se uma lista de amostragem contendo pontos a serem somados no cálculo da posição real dos pontos relativos. Ela foi criada de forma empírica e contém um total de 25 pontos, indo de $(x, y) = (-2, -2)$ a $(x, y) = (2, 2)$, como pode ser visto na Figura 32.

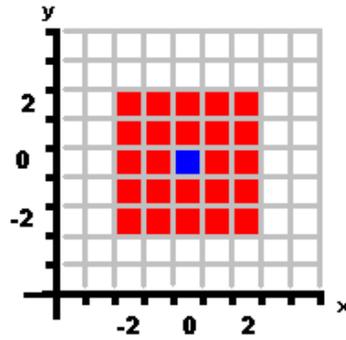


Figura 32: Exemplo da amostragem em cada ponto para identificação. O ponto em azul representa o ponto real na posição $(0, 0)$. E os pontos vermelhos são os pontos amostrados em torno do ponto $(0, 0)$ indo de $(-2, -2)$ a $(2, 2)$.

No sistema implementado, cada pessoa memorizada pode ter mais de uma foto. Considerando-se a amostragem apresentada acima, a representação interna da face de um indivíduo a ser armazenada pelo sistema difere um pouco da representação obtida de uma face a ser identificada (Equação 5.1). Pois, enquanto um ponto de uma face a ser identificada é representada por apenas um Jet (Equação 4.15), um ponto de uma face armazenada no sistema é representado por um conjunto de Jets sendo um para cada amostragem (Figura 32) e cada foto do indivíduo. Essa diferença nas representações permitem que os filtros sejam aplicados menos vezes durante a fase de identificação em relação a quantidade aplicada na etapa de treinamento. Isso, sem restringir o aumento da representatividade da região em torno de um ponto obtida pela amostragem. A equação abaixo descreve a representação interna de um indivíduo.

$$FACEIND = \{J_{paf}\} \quad \text{para } p = 1, \dots, P; a = 1, \dots, A; f = 1, \dots, F \quad (5.2)$$

onde, J_{paf} é o Jet (Equação 4.15) representante do um ponto p , calculado a partir de uma das três listas, e da amostra a aplicado a uma imagem I_f , I_f é imagem do indivíduo referente à foto f , na qual se conhece a posição das três características, P é o número de pontos contidos nas três listas, A é o número de amostras do ponto p , e F é o número de fotos do indivíduo em questão.

Sendo assim, cada uma dessas representações é utilizada em um momento diferente. A Equação 5.2 é requerida no processo de treinamento do sistema. Enquanto a Equação 5.1 é requerida quando uma imagem a ser reconhecida é apresentada para o sistema.

5.1.1 Definição dos Parâmetros

Nessa seção serão definidos todos os parâmetros utilizados na parte de identificação.

Especificação da Retina: Os parâmetros para a construção da retina utilizada na parte de identificação foram escolhidos de forma empírica e são definidos abaixo:

- $r_{retina} = 26$ pixels definindo a região de interesse, ver seção 3.3.2;
- $r_{fovea} = 5$ pixels;
- $r_0 = 4$ pixels;

Com isso, gerou-se uma retina com um passo angular de 10° e um total de 417 pontos.

Especificação das Wavelets: Tendo como base a abordagem adotada em (WISKOTT et al., 1997), os parâmetros que definem cada Wavelet a ser utilizada na parte de identificação, são apresentados abaixo:

- θ será definido para 6 diferentes orientações de 0 a π , isto é, $\theta \in \{0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \frac{4\pi}{6}, \frac{5\pi}{6}\}$. Testes realizados mostraram não ser necessário utilizar as 8 orientações como em (WISKOTT et al., 1997), pois essa diminuição não diminuiu a taxa de reconhecimento;
- λ será definido para 5 diferentes comprimentos de onda partindo de 4 pixels e incrementando de meia oitava, isto é, $\lambda \in \{4, 4\sqrt{2} \cong 6, 8, 8\sqrt{2} \cong 12, 16\}$.
- φ será definido para 2 diferentes fases, $\varphi \in \{0, \frac{\pi}{2}\}$, dado que 0 corresponde à parte real (função cosseno) e $\frac{\pi}{2}$ à parte imaginária (função seno).
- σ será definido para o mesmo tamanho do comprimento de onda, isto é, $\sigma = \lambda$.
- γ será definido para $\gamma = 1$. Isso não causará deformação na gaussiana dando-a um formato circular.

Essas definições, 6 orientações e 5 frequências, geram um total de 30 diferentes Wavelets, que possuem 2 fases cada uma. Ou seja, 60 kernels diferentes a serem convoluídos com a imagem nos pontos a serem analisados.

5.2 Identificação

Uma outra diferença em relação aos métodos apresentados por Wiskott et al. (1997), Bolme (2003) é a função de similaridade entre as faces. Essa função é a responsável por dizer o quanto uma face se parece com outra, para que, a partir daí, a mais similar

seja escolhida. Em (WISKOTT et al., 1997; BOLME, 2003), a similaridade foi calculada simplesmente tirando-se uma média entre a similaridade entre os Jets. Já no método aqui apresentado, ela é um pouco mais complexa e será descrita a seguir.

A função responsável por medir a similaridade entre uma dada face FA e um indivíduo armazenado no sistema FA_i , é definida pelas equações abaixo:

$$(j, s) = \max(\{S(J_p, J_{jpa f})\}) \quad \text{para } a = 1, \dots, A; f = 1, \dots, F; j = 1, \dots, NI \quad (5.3)$$

$$S(FA, FA_i) = \frac{1}{P} \sum_{p=1}^P \begin{cases} s & \text{se } s > \Omega_{lim} \text{ e } i = j \\ 0 & \text{se } s \leq \Omega_{lim} \end{cases} \quad (5.4)$$

onde, \max é a função que retorna o par (j, s) que possui a maior similaridade s , $S(J_p, J_{jpa f})$ é a similaridade entre dois Jets, Equação 4.16, J_p é o Jet (Equação 4.15) referente ao ponto p da face F , $J_{jpa f}$ é o Jet (Equação 4.15) referente ao ponto p , a amostra a e a foto f do indivíduo j , A é o número de amostras do ponto p , F é o número de fotos do indivíduo j , NI é o número de indivíduos conhecidos pelo sistema, FA é uma face da forma definida pela Equação 5.1, e FA_i é a representação interna da face do indivíduo i da forma definida na Equação 5.2, P é o número de pontos contidos nas três listas, e Ω_{lim} é um valor de corte para Jets que não atingiram um valor confiável de similaridade.

O motivo da diferença na função de similaridade (Equação 5.4) proposta nesse trabalho em relação à média utilizada em (WISKOTT et al., 1997; BOLME, 2003), foi evitar falsas contribuições introduzindo o conceito de Jet (Equação 4.15) vencedor definido pela Equação 5.3. Pois um sujeito que tenha um Jet (Equação 4.15) bem parecido, mas não o mais parecido, contribuiria tão bem quanto o mais parecido. Outra maneira de evitar falsas contribuições foi introduzir um limiar de similaridade. Isso, evitou que regiões que definitivamente não se parecem com o alvo, contribuíssem com o aumento da similaridade. Por exemplo, uma região com oclusão por óculos de sol definitivamente não deve contribuir para a similaridade da região dos olhos.

Tendo as similaridades entre as faces, é possível achar a sua identidade. Para isso, basta verificar qual indivíduo apresentou o maior valor. Deve-se considerar também o fato de o sistema ter que rejeitar, ou identificar como impostor (consultar seção 7 para maiores detalhes), os indivíduos que não atingirem uma certa confiabilidade. As equações abaixo definem como encontrar a identidade de uma face $FACE$.

$$(i, s) = \max(\{s = S(FACE, FACEIND_i)\}) \quad \text{para } i = 1, \dots, NI \quad (5.5)$$

$$ID(FACE) = \begin{cases} i & \text{se } s > \Omega_{rej} \\ IMPOSTOR & \text{se } s \leq \Omega_{rej} \end{cases} \quad (5.6)$$

onde, $S(FACE, FACEIND_i)$ é a função que mede a similaridade entre duas faces $FACE$ e $FACEIND_i$, como definido na Equação 5.4, $FACE$ é a face a ser identificada, $FACEIND_i$ é a representação interna do indivíduo i , max é a função que retorna o par (i, s) que possui a maior similaridade s , NI é o total de indivíduos representados pelo sistema, $IMPOSTOR$ são os indivíduos identificados como não pertencentes ao sistema, ou que não atingiram um valor de similaridade confiável, isto é, acima de Ω_{rej} .

6 Localização de Características

A localização de características é parte do sistema responsável por encontrar atributos da face (olhos, nariz, boca, sobrancelha, etc) que possam ser úteis na fase de identificação do indivíduo. Ela deve receber uma imagem e retornar a posição (x, y) de cada um dos pontos a serem localizados.

Como foi visto na seção 2, existem na literatura várias técnicas para realizar esse processo. Porém o sistema aqui desenvolvido continua utilizando os Jets, definidos na seção 4.5 e gerados a partir de Gabors com Foveação, para representação da informação de cada característica.

O processo de localização também é dividido em duas partes: uma de treinamento do sistema e outra de localização das características em questão. A primeira, consiste na criação de uma representação interna a partir de alguns indivíduos com a posição de suas características previamente conhecidas. Já a segunda, consiste na busca de cada característica a ser localizada. Essa será inspirada no modelo de sacada apresentado por Smeraldi, Capdevielle e Bigun (1999), Smeraldi, Carmona e Bigun (2000).

6.1 Representação das Características

Apesar do modelo de Smeraldi, Capdevielle e Bigun (1999), Smeraldi, Carmona e Bigun (2000) não utilizar o conceito de Jet apresentado por Wiskott et al. (1997), os dois são similares e utilizam vetores de respostas de Gabors para representar a informação em torno de um ponto da imagem.

Como visto na seção 5, a parte de identificação facial necessita da localização de três pontos específicos: nariz, olho esquerdo e boca. Sendo assim, é preciso ter uma representação interna de cada uma delas. Tais representações são criadas a partir de Jets (Equação 4.15) extraídos de cada um desses pontos, previamente conhecidos, de alguns indivíduos.

Contudo, a amostragem de apenas um ponto não foi suficiente para representar a

informação sobre uma característica. Então, criou-se uma lista para amostrar alguns pontos ao redor da posição real. Para calcular os pontos a serem amostrados basta somar a posição da característica com cada ponto contido na lista de amostragem. A Figura 33 ilustra os pontos a serem amostrados e a 34 ilustra um exemplo de amostragem no nariz.

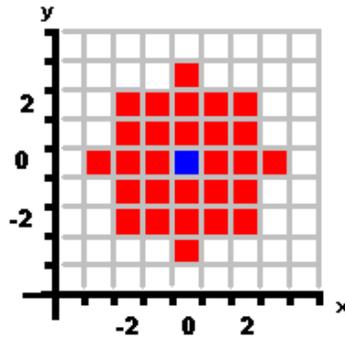


Figura 33: Exemplo da amostragem em cada ponto para a localização. O ponto em azul representa a localização da característica (0, 0). E os pontos vermelhos são os pontos amostrado circularmente em torno do ponto (0, 0). Ao todo a lista contém 29 pontos.



Figura 34: Exemplo da amostragem realizada no nariz. Os pontos vermelhos são pontos amostrado circularmente em torno da posição definida como nariz.

Com isso, uma característica passa a ser representada por Jets (Equação 4.15) extraídos ao redor de um ponto definido manualmente em cada foto utilizada para treinamento seguindo a amostragem definida nas Figuras 33 e 34. Essas fotos podem ser de indivíduos diferentes ou de um mesmo indivíduo em poses diferentes. A equação abaixo define a representação interna de uma característica c .

$$FEATURE_c = \{J_{ifa}\} \text{ para } i = 1, \dots, I; f = 1, \dots, F; a = 1, \dots, A \quad (6.1)$$

onde, J_{ifa} é o Jet (Equação 4.15) representante do ponto referente à característica c na amostragem a da foto f do indivíduo i , I é o número de indivíduos utilizados para representar a característica c , F é o número de fotos de cada indivíduo, e A é o tamanho da lista de amostragem.

É importante frisar que o grupo de indivíduos representantes de uma característica não é necessariamente o mesmo de uma outra.

6.2 Similaridade

A similaridade é a função responsável por dizer o quanto um Jet (Equação 4.15) representando um determinado ponto na imagem é similar a uma característica c .

A equação abaixo descreve a similaridade entre um Jet J e uma característica c .

$$S(J, FEATURE_c) = \max(\{S(J, J_{ifa})\}) \quad \text{para } i = 1, \dots, I; f = 1, \dots, F; a = 1, \dots, A \quad (6.2)$$

onde, \max é a função que retorna o maior valor de um conjunto, $S(J, J_{ifa})$ é a similaridade entre dois Jets, Equação 4.16, J é o Jet (Equação 4.15) referente ao ponto a ser comparado, e J_{ifa} é o Jet (Equação 4.15) referente à característica c na amostra a da foto f do indivíduo i .

6.3 Rotação

De acordo com os parâmetros escolhidos, a retina criada pode ter diferentes graus de flexibilidade quanto a rotação. Por exemplo, a Figura 35 ilustra uma retina com 36 graus de rotação e com um passo angular de 10° .

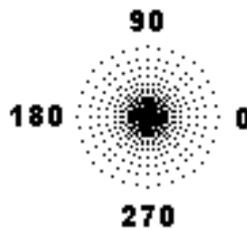


Figura 35: Exemplo dos pontos da retina. A retina apresenta um passo angular de 10° tendo informação suficiente para identificar 36 direções diferentes. Os ângulos 0, 90, 180 e 270 são mostrados na Figura.

O formato circular dos RFs que constituem a retina é um dos responsáveis por introduzir essa invariância. Pois, se a retina for aplicada em um ponto (x, y) de uma imagem I e ao seu ponto correspondente (x_r, y_r) da imagem I rotacionada de 90° em torno de (x, y) . Então, elas teriam os mesmos valores (ou bem próximos) com uma defasagem de

90°. Isso significa que se considerar-se a retina da Figura 35, então os valores dos pontos contidos no raio de 0° seriam os mesmos dos contidos no raio de 90°.

Tal invariância torna capaz a determinação de uma possível rotação da imagem em relação a uma característica procurada. Pois a partir do momento que o sistema foi treinado para uma característica na posição 0°, basta escolher a maior similaridade dos diferentes graus de liberdade de rotação da retina, como descrito pela equação abaixo.

$$\theta = \max(\{S(J_g, FEATURE_c)\}) \quad \text{para } g = 1, \dots, G \quad (6.3)$$

onde, \max é a função que retorna o grau de rotação g que apresentou a maior similaridade. $S(J_g, FEATURE_c)$ é a função de similaridade definida pela Equação 6.2. J_g é o Jet (Equação 4.15) referente ao grau de similaridade g . E, $FEATURE_c$ é a característica procurada.

Para gerar o Jet (Equação 4.15) referente a um determinado grau de rotação, basta adicionar o valor do ângulo relativo ao grau g á orientação de cada Wavelet geradora do Jet antes de realizar o processamento descrito pela Equação 4.12.

6.4 Busca

A *Busca* é a forma como o sistema varre a imagem procurando por uma determinada característica. O pensamento mais natural é utilizar uma busca exaustiva em toda imagem, em que, para cada ponto aplica-se a função de similaridade 6.2. Contudo, embora simples, essa abordagem demanda um tempo de processamento muito alto. Como alternativa, encontrou-se o modelo de sacada apresentado por Smeraldi, Capdevielle e Bigun (1999), Smeraldi, Carmona e Bigun (2000).

Segundo Smeraldi, Capdevielle e Bigun (1999), Smeraldi, Carmona e Bigun (2000), quando os seres humanos exploram uma cena natural, não rastreiam toda a imagem de forma exaustiva. Ao invés disso, realizam saltos rápidos - denominados *Sacada* - entre regiões de interesse, nas quais eles se fixam em média por 0,3 segundos. Toda a informação relevante é adquirida durante a fixação, embora grande parte do tempo seja gasto decidindo o próximo ponto a ser atingido.

Fazendo uma analogia com esse comportamento humano eles desenvolveram um sistema de busca visual baseado em sacadas. Ele é capaz de reduzir o tempo computacional durante a procura, pois não analisa todos os pontos da imagem, somente regiões de interesse. Em virtude da redução promovida por esse método, ele foi o escolhido para a busca

das características.

Vale a pena frisar que apesar do sistema proposto por eles ser motivado biologicamente, ele conflita em conceito, com o apresentado aqui. Pois nesse, a retina é construída a partir de DOG's que fornecem informações para a aplicação das Gabors, já naquele o processamento na imagem é feito diretamente a partir de Gabors. Esse conflito torna a região de interesse definida nessa seção diferente da definida em 3.3.2, porque embora tenha fundamento biológico, elas são utilizadas em momentos diferentes no sistema. Sendo assim, ao invés de chamá-la de região de interesse, ela será chamada de região de amostragem. Essa mudança de nomes almeja distanciar o processo de busca por sacada do modelo biológico a fim de diminuir o conflito conceitual.

Mesmo com esse conflito, o método foi fundamental na redução do tempo computacional e realização de testes mais representativos. Para melhor descrevê-lo, o processo de busca de uma característica c será dividido em três partes: análise da região de amostragem, decisão da próxima região de amostragem e decisão de parada. A Figura 36 apresenta o fluxograma geral da busca por sacada seguida por uma explicação mais detalhada de cada processo.

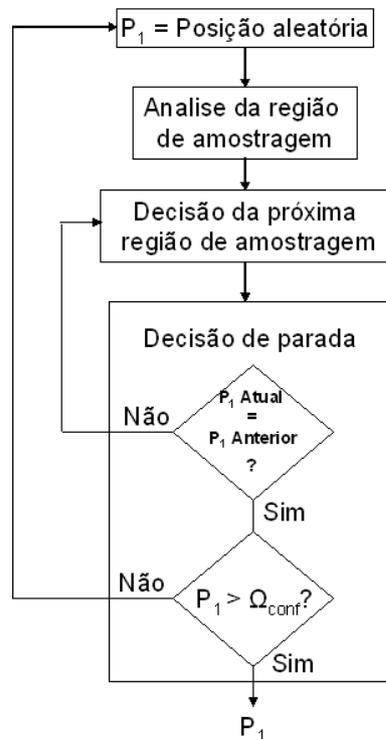


Figura 36: Fluxograma da busca por sacada.

1. **Análise da região de amostragem:** Nessa parte, mede-se a similaridade entre os pontos a serem amostrados a partir de um ponto inicial p_1 , e a característica c .

Para isso, cria-se uma lista de amostragem contendo a posição relativa dos pontos a serem amostrados. Essa lista segue uma amostragem log-polar, tendo o passo radial definido por:

$$\rho_i = B^i \quad \text{para } i = 1, 2, \dots \quad (6.4)$$

onde, B é um parâmetro que define a taxa de espaçamento radial entre os pontos a serem amostrados. E o passo angular é definido por:

$$\theta_j = A * j \quad \text{para } j = 1, 2, \dots \quad (6.5)$$

onde, A é um parâmetro que define o espaçamento angular entre os pontos a serem amostrados.

Portanto, pode-se dizer que para a definição de uma região de amostragem são necessários três parâmetros: a taxa de espaçamento radial B , o grau de espaçamento angular A e um raio máximo $RMAX$, em que $\rho_i \leq RMAX$. A Figura 37 ilustra um exemplo dos pontos amostrados durante uma sacada.

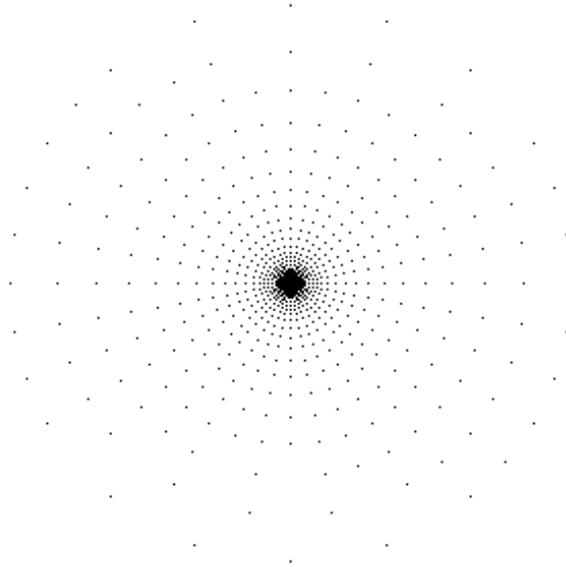


Figura 37: Exemplo de uma região de amostragem para a sacada. O ponto central encontra-se na posição $(0,0)$. Os parâmetros para a geração da lista foram $B = 1.2$, $A = 10$ e $RMAX = 150$.

Após a geração da lista de amostragem, extrai-se um Jet (Equação 4.15) referente a cada ponto real, que é calculado somando-se os pontos da lista com o ponto p_1 . Mede-se a similaridade de cada um deles com a característica c , gerando assim, a lista de análise da região de amostragem ARA , definida abaixo:

$$ARA(p_1, B, A, RMAX, FEATURE_c) = \{S(J_p, FEATURE_c)\} \quad \text{para } p = 1, \dots, P \quad (6.6)$$

onde, $S(J_p, FEATURE_c)$ é a função 6.2, J_p é o Jet (Equação 4.15) referente ao ponto real p obtido a partir do ponto p_1 e da lista de amostragem definida por B , A e $RMAX$, $FEATURE_c$ é a representação interna da característica c , e P é o número de pontos contidos na região de amostragem.

2. **Decisão da próxima região de amostragem:** Nessa parte, o sistema decide qual é o próximo destino da busca, ou seja, o próximo p_1 . Para isso, ele utiliza uma função de sacada, definida abaixo:

$$SACADA(ARA) = (x_{p_1}, y_{p_1}) = \max(ARA) \quad (6.7)$$

onde, ARA é lista de análise da região de amostragem. (x_{p_1}, y_{p_1}) é o próximo ponto p_1 . E, \max é a função que retorna o ponto (x, y) que obteve a maior similaridade em ARA .

A Equação 6.7, da forma como foi definida, direciona as sacadas para as regiões de maior similaridade com a característica procurada, isto é, para máximos locais.

3. **Decisão de parada:** Nessa parte, o sistema decide se a busca deve continuar ou terminar. Ela pode terminar ao encontrar a característica procurada ou após um número limite de sacadas.

Para verificar se a característica c foi encontrada, primeiro deve-se saber se a sacada convergiu para algum ponto. Ou seja, se o próximo ponto p_1 , obtido pela Equação 6.7, é igual ao p_1 anterior fazendo com que não haja uma nova região de amostragem. Caso não seja, faz-se uma sacada para a nova posição de p_1 reiniciando o processo. Caso seja, pode-se confirmar se c foi encontrada de acordo com um limiar de confiabilidade Ω_{conf} . Se a similaridade do ponto p_1 for maior do que o limiar Ω_{conf} , então ela foi encontrada e a procura termina retornando a posição de p_1 . Caso contrário, p_1 recebe posição aleatória e faz-se uma sacada para sua nova posição reiniciando o processo.

Se a busca não encontrar c após um determinado número de sacadas, então ela é encerrada sem uma posição definida.

A Figura 38 ilustra um exemplo dos passos descritos acima.

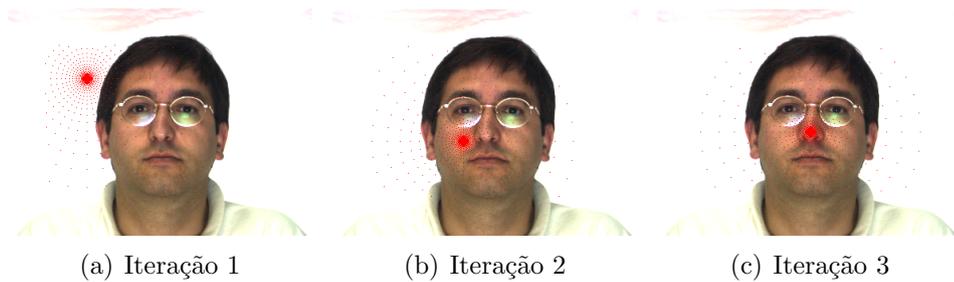


Figura 38: Exemplo da busca por sacada. (a) amostragem realizada a partir de uma posição aleatória. (b) amostragem realizada a partir do ponto com maior similaridade no passo anterior, (a). (c) amostragem realizada a partir do ponto com maior similaridade no passo anterior, (b). Nesse caso, o ponto de maior similaridade foi o central levando à convergência da busca. Como o ponto de convergência ultrapassou o limiar de confiabilidade, Ω_{conf} , a característica foi dada como encontrada. Caso não tivesse atingido tal limiar, a busca recomençaria a partir de uma posição aleatória.

Esse processo é o mesmo para todas as características procuradas. Porém, para a busca da primeira, c_1 , p_1 é iniciado com um valor aleatório. Já para as demais, c_i , p_1 é iniciado com o valor relativo à posição de c_i no indivíduo em que c_1 foi encontrado. Isto é, se c_1 teve a maior similaridade com indivíduo I e a distância de c_i relativo à c_1 no indivíduo I é (x_d, y_d) , então p_1 para a busca de c_i deve ser iniciado na posição de c_1 mais (x_d, y_d) .

Com o valor de p_1 indicando o local mais provável das características c_i , é possível fazer apenas uma busca local em torno da região apontada, ao invés de ter que procurar em toda a imagem. Portanto, pode-se diminuir o raio máximo da região de amostragem, o que diminui bastante a lista de amostragem.

A característica definida como c_1 foi o nariz. Ele foi escolhido pois é o que sofre a menor variação nas imagens. Os parâmetros da lista de amostragem utilizados para achar c_1 foram $B = 1.2$, $A = 10$ e $RMAX = 150$. Já os da localização das demais, c_i , foram $B = 1.3$, $A = 10$ e $RMAX = 30$.

6.4.1 Definição dos Parâmetros

Nessa seção serão definidos todos os parâmetros utilizados na parte de localização de características.

Especificação da Retina: Os parâmetros para a construção da retina utilizada na parte de localização de características foram escolhidos de forma empírica e são definidos

abaixo:

- $r_{retina} = 26$ pixels definindo a região de interesse, ver 3.3.2;
- $r_{fovea} = 5$ pixels;
- $r_0 = 5$ pixels;

Com isso, gerou-se uma retina com um passo angular de 10° , com 36 graus de liberdade de rotação e um total de 417 pontos.

Especificação das Wavelets: Os parâmetros foram definidos de forma empírica tendo como base o modelo de Wiskott et al. (1997). O nariz respondeu melhor para kernels definidos com σ inferior ao comprimento de onda λ . Já o olho esquerdo e a boca, responderam melhor com σ superior ao comprimento de onda λ . O restante dos parâmetros foram iguais para as três características e são definidos abaixo:

- θ será definido para 6 diferentes orientações de 0 a π , isto é, $\theta \in \{0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \frac{4\pi}{6}, \frac{5\pi}{6}\}$;
- λ será definido para 6 diferentes comprimentos de onda partindo de 4 pixels e incrementando de meia oitava, isto é, $\lambda \in \{4, 4\sqrt{2} \cong 6, 8, 8\sqrt{2} \cong 12, 16, 16\sqrt{2} \cong 22\}$;
- φ será definido para 2 diferentes fases, $\varphi \in \{0, \frac{\pi}{2}\}$, dado que 0 corresponde à parte real (função cosseno) e $\frac{\pi}{2}$ à parte imaginária (função seno).
- γ será definido para $\gamma = 1$. Isso não causará deformação na gaussiana dando-a um formato circular.

Essas definições, 6 orientações e 6 frequências, geram um total de 36 diferentes Wavelets, que possuem 2 fases cada uma. Ou seja, 72 kernels diferentes a serem convoluídos com a imagem nos pontos a serem analisados.

7 Experimentos

Um sistema de reconhecimento facial identifica um determinado indivíduo através de uma representação interna obtida previamente a partir de uma ou mais imagens dele. E para que se possa testá-lo eficientemente é necessário representar cada sujeito com imagens diferentes das que forem utilizadas durante os testes. Entende-se por imagem diferente como sendo, uma imagem obtida com um determinado intervalo temporal, com uma variação na expressão, com uma variação do ambiente, ou algo seguindo essa linha. Desse modo, para os casos de testes a serem descritos, as imagens serão separadas em dois grupos:

- **Imagens de Treino - (I_T):** Grupo de imagens para gerar a representação interna do sistema.
- **Imagens de Sondagem - (I_S):** Grupo de imagens, composto por imagens diferentes das contidas em I_T , para testar a capacidade de identificação do sistema.

O banco de imagens utilizado para a realização dos testes foi o *AR Face Database*. No apêndice A, é possível ver uma descrição sobre o banco de imagens e é definida uma nomenclatura para referenciar o conjunto de fotos (AR1, AR2, AR3, ..., AR26) durante os testes. As coordenadas (x, y) do nariz, do olho esquerdo e da boca, utilizadas durante os testes, foram manualmente localizadas por Souza (2005) em todas as imagens do banco.

7.1 Testes de Identificação

Segundo Yan e Osadciw (2005), os testes de identificação podem ser de dois tipos: *Conjunto Fechado*, quando todos os indivíduos contidos em I_S possuem pelo menos uma representação em I_T ; e *Conjunto Aberto*, quando pelo menos um indivíduo contido em I_S não possui uma representação em I_T , sendo que, esse(s) sem representação deve(m) ser identificado(s) como impostor(es).

Um teste do tipo *Conjunto Fechado* só pode gerar dois tipos de resposta: *Correto*,

quando o sujeito dado como entrada é igual ao encontrado pelo sistema; ou *Errado*, quando o sujeito dado como entrada é diferente do encontrado pelo sistema. Já um teste com uma imagem do tipo *Conjunto Aberto* pode gerar quatro tipos de resposta: *Identificação Correta* (IC), quando o sujeito dado como entrada é igual ao encontrado pelo sistema; *Rejeição Correta* (RC) quando o sujeito dado como entrada é identificado como impostor e não possui representação em I_T ; *Identificação Errada* (IE), quando o sujeito dado como entrada é diferente do encontrado pelo sistema. Uma IE pode ser dividida em dois casos: *Identificação Errada Trocada* (IET), quando identifica um sujeito como sendo um outro também presente no sistema e *Identificação Errada Impostor* (IEI), quando identifica um impostor como sendo alguém presente no sistema; ou *Rejeição Errada* (RE), quando o sujeito dado como entrada é identificado como impostor e possui representação em I_T .

Os testes de identificação foram elaborados de forma a avaliar o algoritmo em diferentes condições. Cada um deles será executado para os dois métodos, Gabor Wavelets com foveação (GWF) e sem foveação (GW). Todos eles seguiram a parametrização definida na seção 5 e tiveram as mesmas condições para os dois métodos. A seguir serão apresentados alguns testes realizados.

7.1.1 Teste de Expressões

Esse teste é do tipo *Conjunto Fechado* e teve como objetivo avaliar como o algoritmo de identificação responde à diferentes expressões faciais. Para isso, construiu-se vários grupos para I_T , sendo esses, formado por combinações diferentes de expressões da primeira sessão de fotos. Os conjuntos de fotos com expressões diferentes das presentes em I_T , tanto da primeira quanto da segunda sessão, constituíram o I_S respectivo.

Os resultados obtidos nos testes realizados com 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database*, são apresentados nas Tabelas 1 (primeira sessão de fotos) e 2 (segunda sessão de fotos). As Figuras 39 e 40 apresentam exemplos das expressões utilizadas.

Tabela 1: Testes de expressão da primeira sessão. Percentual de acerto para os diferentes I_S tendo como I_T os conjuntos apresentados na primeira coluna. A segunda coluna, M, determina o método, como definido em 7.1. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

I_T	I_S				
	M	AR1	AR2	AR3	AR4
AR1	GWF	—	100,0	99,15	81,20
	GW	—	98,29	98,29	79,49
AR1, AR2 e AR3	GWF	—	—	—	95,73
	GW	—	—	—	91,45
AR1, AR2 e AR4	GWF	—	—	100,0	—
	GW	—	—	100,0	—
AR1, AR3 e AR4	GWF	—	100,0	—	—
	GW	—	100,0	—	—
AR2, AR3 e AR4	GWF	100,0	—	—	—
	GW	100,0	—	—	—



Figura 39: Exemplo de fotos dos conjuntos AR1, AR2, AR3 e AR4, tiradas na primeira sessão. (a) Expressão neutra, (b) Sorrindo, (c) Zangado, (d) Gritando. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

Tabela 2: Testes de expressão da segunda sessão. Percentual de acerto para os diferentes I_S tendo como I_T os conjuntos apresentados na primeira coluna. A segunda coluna, M, determina o método, como definido em 7.1. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

I_T	I_S				
	M	AR14	AR15	AR16	AR17
AR1	GWF	100,0	93,16	95,73	66,67
	GW	100,0	92,31	96,58	62,39
AR1, AR2 e AR3	GWF	—	—	—	84,66
	GW	—	—	—	79,49
AR1, AR2 e AR4	GWF	—	—	99,15	—
	GW	—	—	98,29	—
AR1, AR3 e AR4	GWF	—	97,44	—	—
	GW	—	94,87	—	—
AR2, AR3 e AR4	GWF	99,15	—	—	—
	GW	98,29	—	—	—



Figura 40: Exemplo de fotos dos conjuntos AR14, AR15, AR16 e AR17, tiradas na segunda sessão. (a) Expressão neutra, (b) Sorrindo, (c) Zangado, (d) Gritando. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

A Tabela 3 apresenta os resultados obtidos com diferentes métodos encontrados na literatura utilizando a mesma base, (MARTINEZ; BENAVENTE, 1998), dos testes acima. Eles utilizaram o conjunto AR1 como I_T e AR2, AR3 e AR4 como I_S . Os resultados foram extraídos de (GUILLAMET; VITRIA,).

Tabela 3: Testes de expressão da primeira sessão encontrados na literatura. Percentual de acerto para os diferentes métodos apresentados na primeira coluna. AR1 foi utilizado como I_T . Foram utilizados 116 indivíduos (63 homens e 53 mulheres) para a realização desses testes. Resultados extraídos de (GUILLAMET; VITRIA,).

Método	I_S		
	AR2	AR3	AR4
FaceIt	96	93	78
Bayesian	72	67	41
PCA-50+L2 Norm	67	82	18
NMF-50+L2 Norm	61	78	14
NMF-50+L1 Norm	72	80	19
NMF-50+Correlation	73	77	18
PCA-100+L2 Norm	80	88	24
NMF-100+L2 Norm	62	85	09
NMF-100+L1 Norm	85	91	29
NMF-100+Correlation	89	90	28
PCA-150+L2 Norm	83	90	29
NMF-150+L2 Norm	66	87	09
NMF-150+L1 Norm	88	92	30
NMF-150+Correlation	93	95	36

A Tabela 4 apresenta os resultados obtidos com o método Análise Discriminante Linear (LDA) apresentado em (MARTINEZ; ZHANG,).

Tabela 4: Testes de expressão encontrados na literatura. Percentual de acerto do método LDA para os diferentes I_S tendo como I_T os conjuntos apresentados na primeira coluna. Foram utilizados 100 indivíduos para a realização desses testes. Resultados extraídos de (MARTINEZ; ZHANG,).

I_T	I_S							
	AR1	AR2	AR3	AR4	AR14	AR15	AR16	AR17
AR1, AR2 e AR3	—	—	—	83	—	—	—	62
AR1, AR2 e AR4	—	—	90	—	—	—	75	—
AR1, AR3 e AR4	—	97	—	—	—	77	—	—
AR2, AR3 e AR4	96	—	—	—	74	—	—	—

Conclusões

Após uma análise dos resultados obtidos para os testes de expressões, foi possível chegar a algumas conclusões apresentadas nos itens abaixo:

- **Diferença de expressão:** Quanto mais diferente é a expressão, menor o percentual de acerto. Por exemplo, a expressão neutra tem muito mais acertos do que a expressão gritando. Isso, porque ao gritar o sujeito abre a boca, enruga o nariz e alguns até fecham os olhos, o que causa grandes mudanças na face diminuindo a capacidade de identificação. Apesar dessas diferenças influenciarem negativamente nos resultados, elas não foram cruciais para invalidar os métodos. Pois, como pode ser visto na Tabela 1, eles mantiveram uma boa taxa de reconhecimento com o pior caso próximo aos 80%.
- **Diferença temporal:** A diferença temporal afetou um pouco a capacidade de reconhecimento do sistema. Isto é, as fotos da segunda sessão (Tabela 2) tiveram um percentual de identificação menor do que as da primeira (Tabela 1). Um dos motivos para que isso ocorresse foi a mudança de visual feita por alguns indivíduos de uma sessão para outra. Por exemplo, algumas pessoas usaram óculos durante a primeira sessão, mas não na segunda. Essas mudanças não influenciaram quando a expressão de teste foi a mesma de treino (como AR1 e AR14) mantendo-se em 100% de acerto, porém proporcionaram uma pequena queda quando elas se somaram às diferenças da expressão (como AR1 e AR15).
- **Aumento do conjunto de I_T :** O aumento do número de fotos representando um indivíduo (testes referentes às 4 últimas linhas de I_T nas Tabelas 1 e 2) mostrou-se uma boa opção de aumentar a taxa de identificação. Pois, partes da face que não foram identificadas em uma foto puderam ser identificadas em outra fazendo

com que o sistema tivesse uma maior representatividade de um mesmo indivíduo e acertasse mais.

- **(GWF e GW) x Outros:** Os resultados mostraram que tanto o GWF, quanto o GW, apresentaram melhores taxas de identificação do que os métodos mostrados nas Tabelas 3 e 4. Embora essa comparação não sirva para julgar se eles são bons ou ruins, ela serve para situar os métodos na literatura.
- **GWF x GW:** Em geral, o método GWF apresentou um percentual de identificação maior (em média 3%) do que o GW (Tabelas 1 e 2). Entretanto, para os testes de expressões, não houve nenhuma grande disparidade entre os dois.

Tendo em vista, as conclusões acima e o fato de que na maior parte dos testes de expressões os dois métodos apresentaram resultados próximos aos 100% de identificação, pode-se dizer que eles lidam bem com diferentes expressões.

7.1.2 Teste de Iluminação

Esse teste é do tipo *Conjunto Fechado* e teve como objetivo avaliar como o algoritmo de identificação responde à variações na iluminação. Para isso, utilizou-se o conjunto de fotos AR1 como I_T e AR5, AR6, AR7, AR18, AR19 e AR20 como I_S .

A Tabela 5 apresenta os resultados obtidos nos testes realizados com 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database*. A Figura 41 apresenta exemplos de fotos utilizadas nos testes de iluminação.

Tabela 5: Testes de iluminação. Percentual de acerto para os diferentes I_S tendo como I_T os conjuntos apresentados na primeira coluna. A segunda coluna, M, determina o método, como definido em 7.1. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

I_T	I_S						
	M	AR5	AR6	AR7	AR18	AR19	AR20
AR1	GWF	100,0	100,0	95,73	100,0	96,58	82,91
	GW	100,0	100,0	94,02	99,15	95,73	80,34

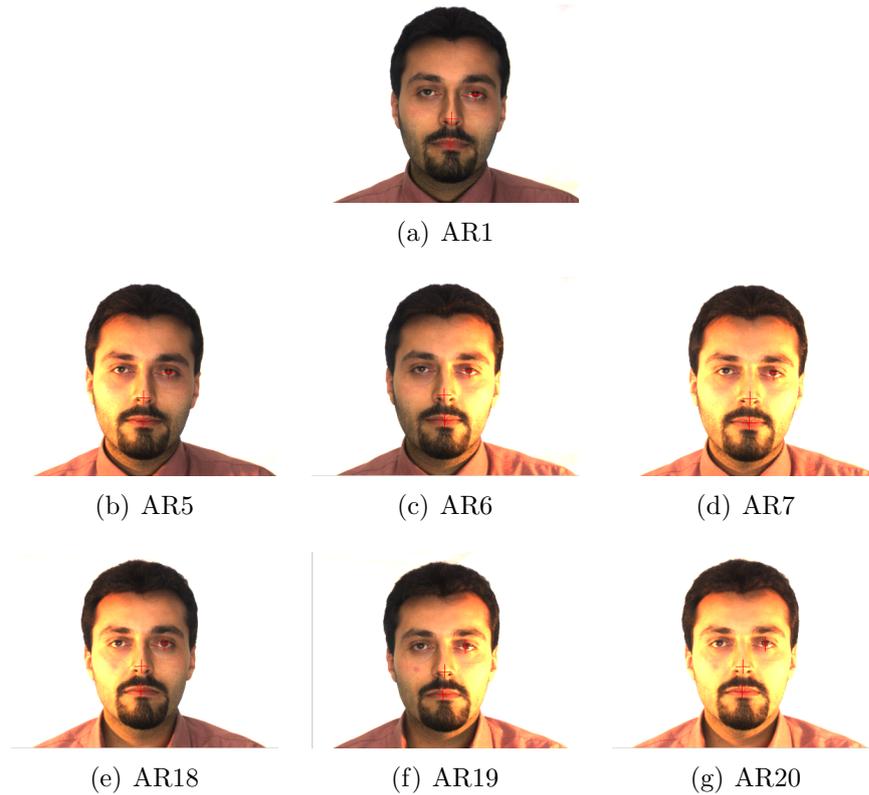


Figura 41: Exemplo de fotos dos conjuntos AR1, AR5, AR6 e AR7, tiradas na primeira sessão, e AR18, AR19 e AR20, tiradas na segunda sessão. (a) Expressão neutra, (b) Luz a esquerda, (c) Luz a direita, (d) Luz por todos os lados. (e) Luz a esquerda, (f) Luz a direita, (g) Luz por todos os lados. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

A Tabela 6 apresenta os resultados obtidos com diferentes métodos encontrados na literatura utilizando a mesma base, (MARTINEZ; BENAVENTE, 1998), dos testes acima. Eles utilizaram o conjunto AR1 como I_T e AR5, AR6 e AR7 como I_S . Os resultados foram extraídos de (GUILLAMET; VITRIA,).

Tabela 6: Testes de iluminação encontrados na literatura. Percentual de acerto para os diferentes métodos apresentados na primeira coluna. AR1 foi utilizado como I_T . Foram utilizados 116 indivíduos (63 homens e 53 mulheres) para a realização desses testes. Resultados extraídos de (GUILLAMET; VITRIA,).

Método	I_S		
	AR5	AR6	AR7
FaceIt	95	93	86
Bayesian	77	74	72
PCA-50+L2 Norm	77	76	57
NMF-50+L2 Norm	91	84	67
NMF-50+L1 Norm	93	87	69
NMF-50+Correlation	94	89	76
PCA-100+L2 Norm	86	86	69
NMF-100+L2 Norm	94	85	67
NMF-100+L1 Norm	97	94	87
NMF-100+Correlation	99	94	88
PCA-150+L2 Norm	85	87	71
NMF-150+L2 Norm	93	84	64
NMF-150+L1 Norm	98	97	92
NMF-150+Correlation	99	96	91

Conclusões

Após uma análise dos resultados obtidos para os testes de iluminação, foi possível chegar a algumas conclusões apresentadas nos itens abaixo:

- **Diferença de iluminação:** O aumento da quantidade de luz na cena diminuiu um pouco a capacidade de identificação. O maior efeito foi causado quando houve incidência bilateral (conjunto AR7, Tabela 5). Isso porque, o aumento da luz acabou causando uma saturação na região mais clara da imagem (como a pele dos indivíduos). Essa saturação deteriorou a informação de frequência a ser extraída pelas Wavelets introduzindo um certo ruído no processo de identificação. Apesar disso, elas não foram cruciais para invalidar os métodos. Pois eles mantiveram uma boa taxa de reconhecimento tendo o pior caso próximo aos 95%.
- **Diferença temporal:** A diferença temporal afetou um pouco a capacidade de reconhecimento do sistema (Tabela 5). Isto é, as fotos da segunda sessão (AR18, AR19 e AR20) tiveram um percentual de identificação menor do que as da primeira (AR5, AR6 e AR6). Para explicação do motivo visar os testes de expressões na seção 7.1.1.
- **(GWF e GW) x Outros:** Os resultados (Tabela 5) mostraram que tanto o GWF,

quanto o GW, apresentaram melhores taxas de identificação do que os métodos mostrados na Tabela 6.

- **GWF x GW:** Em geral, o método GWF apresentou um percentual de identificação maior (em média 2%) do que o GW (Tabela 5). Entretanto, para os testes de iluminação, não houve nenhuma grande disparidade entre os dois.

Tendo em vista, as conclusões acima e o fato de que na maior parte dos testes de iluminação os dois métodos apresentaram resultados próximos aos 100% de identificação, pode-se dizer que eles lidam bem com variações na quantidade de luz.

7.1.3 Teste de Oclusões Naturais

Esse teste é do tipo *Conjunto Fechado* e teve como objetivo avaliar como o algoritmo de identificação responde à oclusões naturais, como: óculos de sol escondendo os olhos e cachecol escondendo a boca. Para isso, utilizou-se o conjunto de fotos AR1 como I_T e AR8, AR9, AR10, AR21, AR22, AR23, AR11, AR12, AR13, AR24, AR25 e AR26 como I_S .

A Tabela 7 apresenta os resultados obtidos nos testes, relativos às oclusões causadas por óculos de sol, realizados com 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database*. A Figura 42 apresenta exemplos de fotos utilizadas nos testes de oclusão dos olhos.

Tabela 7: Testes de oclusão dos olhos. Percentual de acerto para os diferentes I_S tendo como I_T os conjuntos apresentados na primeira coluna. A segunda coluna, M, determina o método, como definido em 7.1. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

I_T	I_S						
	M	AR8	AR9	AR10	AR21	AR22	AR23
AR1	GWF	98,29	95,73	95,73	96,58	88,03	81,20
	GW	96,58	94,02	94,02	93,16	86,32	77,78

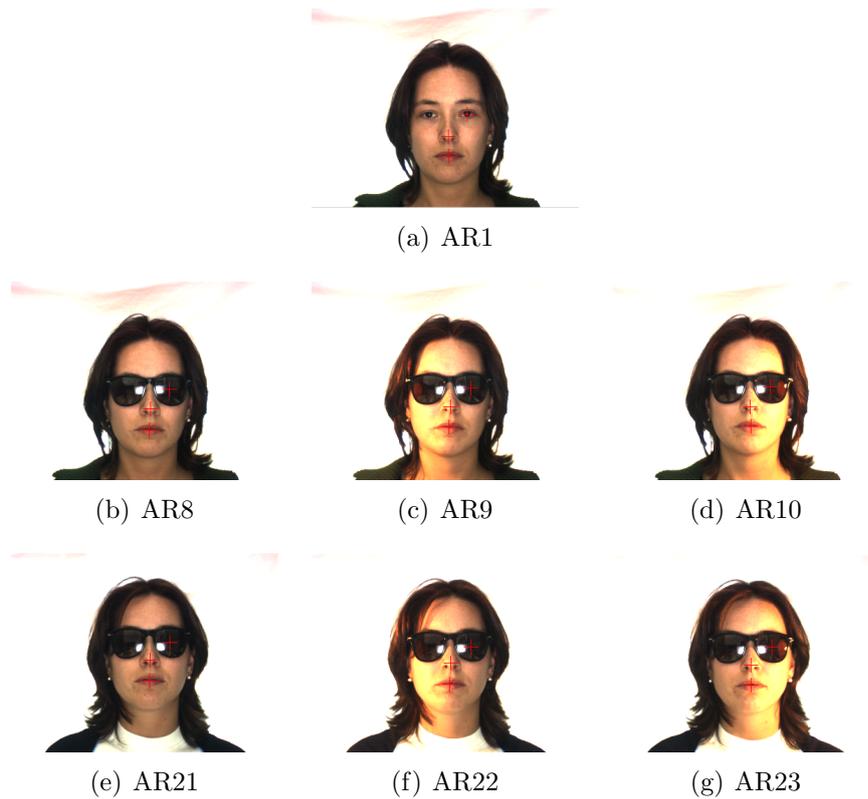


Figura 42: Exemplo de fotos dos conjuntos AR1, AR8, AR9 e AR10, tiradas na primeira sessão, e AR21, AR22 e AR23, tiradas na segunda sessão. (a) Expressão neutra, (b) Óculos de sol, (c) Óculos de sol e luz a esquerda, (d) Óculos de sol e luz a direita, (e) Óculos de sol, (f) Óculos de sol e luz a esquerda, (g) Óculos de sol e luz a direita. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

A Tabela 8 apresenta os resultados obtidos nos testes, relativos às oclusões causadas por cachecol, realizados com 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database*. A Figura 43 apresenta exemplos de fotos utilizadas nos testes de oclusão da boca.

Tabela 8: Testes de oclusão da boca. Percentual de acerto para os diferentes I_S tendo como I_T os conjuntos apresentados na primeira coluna. A segunda coluna, M, determina o método, como definido em 7.1. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

I_T	I_S						
	M	AR11	AR12	AR13	AR24	AR25	AR26
AR1	GW	95,73	94,02	84,62	96,58	87,18	70,94
	GW	94,02	94,02	80,34	92,31	82,91	65,81

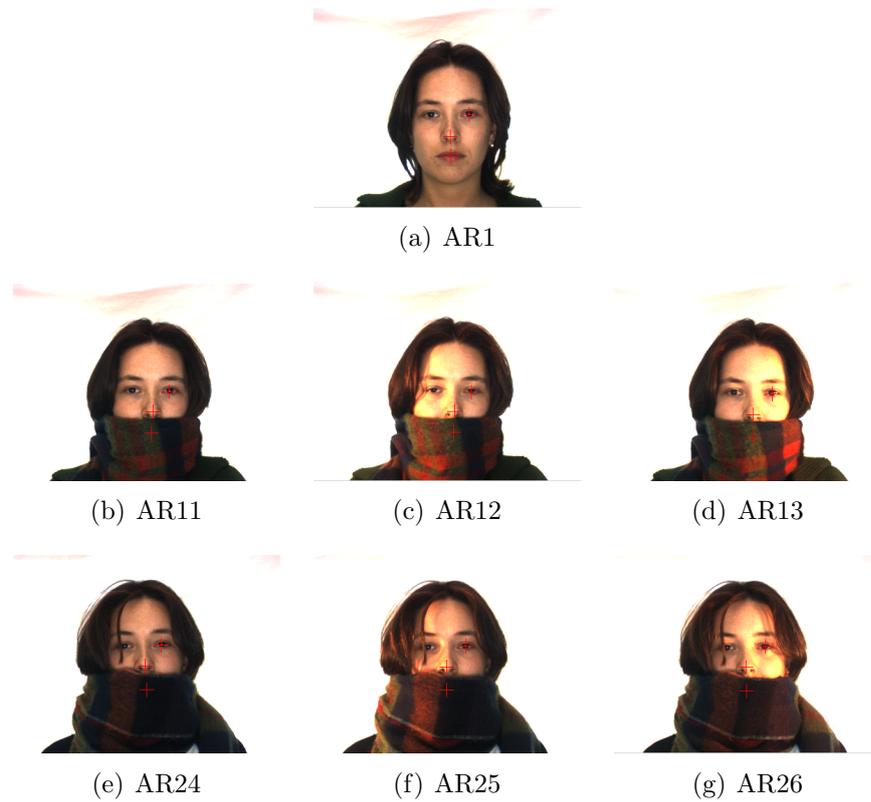


Figura 43: Exemplo de fotos dos conjuntos AR11, AR12 e AR13, tiradas na primeira sessão, e AR24, AR25 e AR26, tiradas na segunda sessão. (a) Expressão neutra, (b) Cachecol, (c) Cachecol e luz a esquerda, (d) Cachecol e luz a direita, (e) Cachecol, (f) Cachecol e luz a esquerda, (g) Cachecol e luz a direita. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

A Tabela 9 apresenta os resultados obtidos com diferentes métodos encontrados na literatura utilizando a mesma base, (MARTINEZ; BENAVENTE, 1998), dos testes acima. Eles utilizaram o conjunto AR1 como I_T e AR8, AR9 e AR10 como I_S . Os resultados foram extraídos de (GUILLAMET; VITRIA,).

Tabela 9: Testes de oclusão dos olhos encontrados na literatura. Percentual de acerto para os diferentes métodos apresentados na primeira coluna. AR1 foi utilizado como I_T . Foram utilizados 116 indivíduos (63 homens e 53 mulheres) para a realização desses testes. Resultados extraídos de (GUILLAMET; VITRIA,).

Método	I_S		
	AR8	AR9	AR10
FaceIt	10	08	06
Bayesian	34	35	28
PCA-50+L2 Norm	16	12	18
NMF-50+L2 Norm	16	10	12
NMF-50+L1 Norm	19	10	20
NMF-50+Correlation	23	12	17
PCA-100+L2 Norm	23	15	22
NMF-100+L2 Norm	14	11	12
NMF-100+L1 Norm	24	15	21
NMF-100+Correlation	32	19	24
PCA-150+L2 Norm	26	16	24
NMF-150+L2 Norm	17	12	09
NMF-150+L1 Norm	31	21	23
NMF-150+Correlation	38	21	23

A Tabela 10 apresenta os resultados obtidos com diferentes métodos encontrados na literatura utilizando a mesma base, (MARTINEZ; BENAVENTE, 1998), dos testes acima. Eles utilizaram o conjunto AR1 como I_T e AR11, AR12 e AR13 como I_S . Os resultados foram extraídos de (GUILLAMET; VITRIA,).

Tabela 10: Testes de oclusão da boca encontrados na literatura. Percentual de acerto para os diferentes métodos apresentados na primeira coluna. AR1 foi utilizado como I_T . Foram utilizados 116 indivíduos (63 homens e 53 mulheres) para a realização desses testes. Resultados extraídos de (GUILLAMET; VITRIA,).

Método	I_S		
	AR11	AR12	AR13
FaceIt	81	73	71
Bayesian	46	43	40
PCA-50+L2 Norm	44	38	37
NMF-50+L2 Norm	47	35	28
NMF-50+L1 Norm	59	35	32
NMF-50+Correlation	61	45	35
PCA-100+L2 Norm	59	50	47
NMF-100+L2 Norm	47	36	25
NMF-100+L1 Norm	66	55	46
NMF-100+Correlation	76	62	59
PCA-150+L2 Norm	62	57	48
NMF-150+L2 Norm	53	31	24
NMF-150+L1 Norm	73	57	48
NMF-150+Correlation	75	62	56

Conclusões

Após uma análise dos resultados obtidos para os testes de oclusão natural, foi possível chegar a algumas conclusões apresentadas nos itens abaixo:

- **Oclusões:** As oclusões reduziram um pouco a taxa de identificação (AR8, AR9, AR10, AR11, AR12 e AR13 presentes nas Tabelas 7 e 8). Isso ocorreu, porque a quantidade de regiões a serem analisadas diminuiu, causando assim, uma diminuição na certeza de escolha de cada indivíduo. Apesar disso, elas não foram cruciais para invalidar os métodos. Pois eles mantiveram uma boa taxa de reconhecimento, em média 95%, tendo o pior caso próximo aos 80%.
- **Variações:** Pelos resultados, nota-se que quanto mais variações (oclusão, iluminação e diferença temporal) são acrescentadas nas imagens, menor é a taxa de identificação. Porém, quando aplicadas uma a uma (AR8 e AR11 nas Tabelas 7 e 8), elas não causam tanto prejuízo quanto o acúmulo dessas variações, como por exemplo temporal, oclusão e iluminação (AR22, AR23, AR25 e AR26 nas Tabelas 7 e 8).
- **(GWF e GW) x Outros:** Os resultados (Tabelas 7 e 8) mostraram que tanto o GWF, quanto o GW, apresentaram taxas de identificação muito melhores do que os

métodos mostrados nas Tabelas 9 e 10.

- **GWF x GW:** Em geral, o método GWF apresentou um percentual de identificação maior (em média 4%) do que o GW (Tabelas 7 e 8). Entretanto, para os testes de oclusões naturais, não houve nenhuma grande disparidade entre os dois.

Tendo em vista, as conclusões acima e o fato de que na maior parte dos testes de oclusões naturais os dois métodos apresentaram resultados próximos ou acima de 95% de identificação, pode-se dizer que eles lidam bem com a falta de informação sobre algumas regiões.

7.1.4 Teste de Rejeição

Esse teste é do tipo *Conjunto Aberto* e teve como objetivo avaliar como o algoritmo de identificação responde ao ter que rejeitar indivíduos não pertencentes ao conjunto de imagens de treinamento (I_T).

Para isso, dividiu-se 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database* em dois grupos: os conhecidos e os desconhecidos pelo sistema. O primeiro grupo foi formado por 100 indivíduos selecionados randomicamente e o segundo pelos 17 restantes. Esse procedimento foi repetido três vezes gerando três conjuntos para testes C1, C2 e C3, com cada um deles possuindo uma combinação diferente de conhecidos e de desconhecidos. C1 contém apenas mulheres no grupo de desconhecidos. C2 contém apenas homens no grupo de desconhecidos. E, C3 contém mulheres e homens no grupo de desconhecidos. A Figura 44 apresenta exemplos de fotos utilizadas nos testes de rejeição.

A Tabela 11 apresenta os resultados obtidos nos testes para Equação 5.5 definida com $\Omega_{rej} = 0,1765$. E a Tabela 12 apresenta os resultados obtidos nos testes para Equação 5.5 definida com $\Omega_{rej} = 0,2941$.

Tabela 11: Testes de rejeição $\Omega_{rej} = 0,1765$. Percentual de acerto para os diferentes I_S tendo como I_T os conjuntos apresentados na primeira coluna. A segunda coluna, M, determina o método, como definido em 7.1. A terceira coluna TR, descreve os tipos de resposta, como visto em 7.1. As colunas restantes apresentam cada resultado de acordo com o I_S utilizado. Eles foram obtidos fazendo-se uma média entre os diferentes conjuntos de teste C1, C2 e C3. $\Omega_{rej} = 0,1765$ na Equação 5.5.

I_T	I_S								
	M	TR	AR2	AR3	AR4	AR14	AR15	AR16	AR17
AR1	GWF	IC	95,00	97,33	38,00	100,0	69,67	89,67	21,00
		IET	0,0	0,0	0,0	0,0	0,0	0,0	0,0
		RE	5,00	2,67	62,00	0,0	30,33	10,33	79,00
		RC	98,04	98,04	100,0	100,0	100,0	92,16	100,0
		IEI	1,96	1,96	0,0	0,0	0,0	7,84	0,0
	GW	IC	92,67	97,67	38,33	100,0	71,33	90,67	24,33
		IET	0,0	0,0	0,0	0,0	0,0	0,0	0,67
		RE	7,33	2,33	61,67	0,0	28,67	9,33	75,00
		RC	96,08	96,08	100,0	92,16	98,04	92,16	100,0
		IEI	3,92	3,92	0,0	7,84	1,96	7,84	0,0
AR1, AR2 e AR3	GWF	IC	—	—	72,67	—	—	—	37,33
		IET	—	—	0,0	—	—	—	0,0
		RE	—	—	27,33	—	—	—	62,67
		RC	—	—	98,04	—	—	—	100,0
		IEI	—	—	1,96	—	—	—	0,0
	GW	IC	—	—	70,67	—	—	—	38,33
		IET	—	—	0,0	—	—	—	0,0
		RE	—	—	29,33	—	—	—	61,67
		RC	—	—	98,04	—	—	—	98,04
		IEI	—	—	1,96	—	—	—	1,96

Tabela 12: Testes de rejeição $\Omega_{rej} = 0, 2941$. Percentual de acerto para os diferentes I_S tendo como I_T os conjuntos apresentados na primeira coluna. A segunda coluna, M, determina o método, como definido em 7.1. A terceira coluna TR, descreve os tipos de resposta, como visto em 7.1. As colunas restantes apresentam cada resultado de acordo com o I_S utilizado. Eles foram obtidos fazendo-se uma média entre os diferentes conjuntos de teste C1, C2 e C3. $\Omega_{rej} = 0, 2941$ na Equação 5.5.

I_T	I_S								
	M	TR	AR2	AR3	AR4	AR14	AR15	AR16	AR17
AR1	GWF	IC	60,67	91,00	10,00	100,0	33,67	76,00	6,67
		IET	0,0	0,0	0,0	0,0	0,0	0,0	0,0
		RE	39,33	9,00	90,00	0,0	66,33	24,00	93,33
		RC	100,0	100,0	100,0	100,0	100,0	100,0	100,0
		IEI	0,0	0,0	0,0	0,0	0,0	0,0	0,0
	GW	IC	63,00	92,00	9,33	99,67	36,33	77,67	6,0
		IET	0,0	0,0	0,0	0,0	0,0	0,0	0,0
		RE	37,00	8,00	90,67	0,33	63,67	22,33	94,00
		RC	100,0	100,0	100,0	100,0	100,0	100,0	100,0
		IEI	0,0	0,0	0,0	0,0	0,0	0,0	0,0
AR1, AR2 e AR3	GWF	IC	—	—	29,00	—	—	—	10,67
		IET	—	—	0,0	—	—	—	0,0
		RE	—	—	71,00	—	—	—	89,33
		RC	—	—	100,0	—	—	—	100,0
		IEI	—	—	0,0	—	—	—	0,0
	GW	IC	—	—	26,67	—	—	—	10,67
		IET	—	—	0,0	—	—	—	0,0
		RE	—	—	73,33	—	—	—	89,33
		RC	—	—	100,0	—	—	—	100,0
		IEI	—	—	0,0	—	—	—	0,0

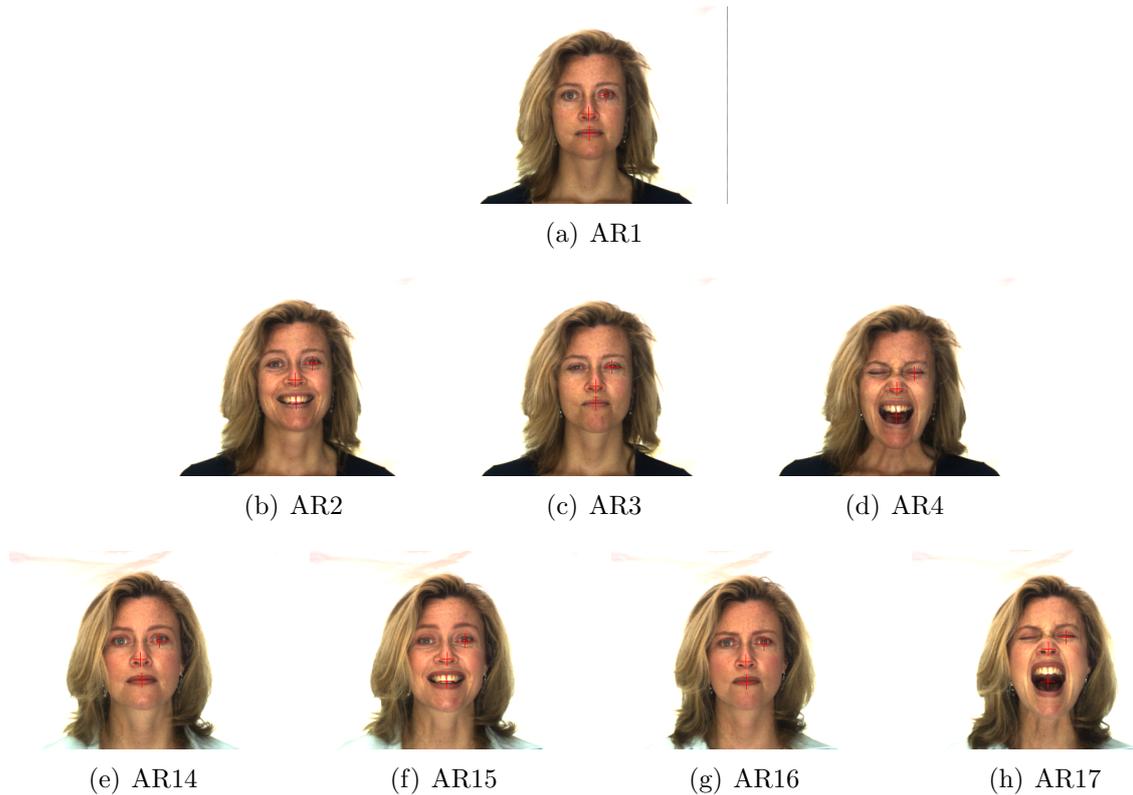


Figura 44: Exemplo de fotos dos conjuntos AR1, AR2, AR3 e AR4, tiradas na primeira sessão, e AR14, AR15, AR16 e AR17, tiradas na segunda sessão. (a) Expressão neutra, (b) Sorrindo, (c) Zangado, (d) Gritando, (e) Expressão neutra, (f) Sorrindo, (g) Zangado, (h) Gritando. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

Conclusões

Após uma análise dos resultados obtidos para os testes de rejeição, foi possível chegar a algumas conclusões apresentadas nos itens abaixo:

- **Parâmetro de confiabilidade:** Esse parâmetro possibilitou que o sistema desse respostas mais confiáveis. Por exemplo, nenhum teste realizado identificou um indivíduo com a identidade de outro (IET). Pois nesse caso, eles apresentaram baixa confiabilidade e foram descartados. Contudo, pagou-se o preço de recusar indivíduos que seriam corretamente identificados, mas com uma baixa confiabilidade. Para ver isso, basta comparar a quantidade de indivíduos corretamente identificados (IC nas Tabelas 11 e 12) com o percentual de acerto dos teste de expressões (Tabelas 1 e 2). Outro ponto analisado foi que, para atingir um grau de rejeição igualmente confiável, isto é, não permitir que intrusos fossem identificados como alguém do sistema (RC), foi necessário aumentar o parâmetro de confiabilidade. Isso implicou em uma diminuição na taxa de identificação do sistema (IC), pois ele se tornou menos flexível

à variações (como por exemplo mudança de expressão).

- **Variações:** As variações, temporal e de expressão, dificultaram o reconhecimento, pois elas diminuíram a confiabilidade do indivíduo sendo avaliado. O resultado tende a piorar quando há o acúmulo de variações (AR15, AR16 e AR17 nas Tabelas 11 e 12).
- **Aumento do conjunto de I_T :** O aumento do número de fotos representando um indivíduo mostrou-se uma boa opção para aumentar a taxa de identificação, mantendo-se a confiabilidade diante de variações (última linha de I_T nas Tabelas 11 e 12).
- **GWF x GW:** Os dois mostram-se igualmente bons em não trocar identidades (IET), porém o GWF apresentou-se melhor do que o GW em relação a rejeição de intrusos (RC), visar Tabelas 11 e 12.

Em geral, os dois métodos mostraram-se confiáveis com resultados próximos aos 100% de identificação e de rejeição. Entretanto, para mantê-los assim, foi necessário abrir mão da flexibilidade quanto às variações.

7.1.5 Teste de Erro na Localização

Esse teste é do tipo *Conjunto fechado* e teve como objetivo avaliar como o algoritmo de identificação responde a erros na localização das características da face (olho esquerdo, nariz e boca). Para isso, utilizou-se o conjunto de fotos AR1 como I_T e AR2, AR3 e AR4 como I_S fazendo as coordenadas das características desses variar, em relação às manualmente localizadas, de um número D de pixels em quatro direções: $y = 0$ e x crescente; $y = 0$ e x decrescente; $x = 0$ e y crescente; e, $x = 0$ e y decrescente.

A Tabela 13 apresenta os resultados obtidos nos testes realizados para diferentes I_S com 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database*. A Figura 45 apresenta exemplos de fotos utilizadas nos testes de erros de localização.

Tabela 13: Testes de erros de localização. Percentual de acerto para identificação realizada com as coordenadas das características deslocadas de diferentes quantidades em pixels D , definidas na primeira coluna, em quatro direções. A segunda coluna, M , determina o método, como definido em 7.1. Para a realização desses testes foram utilizados 117 indivíduos (64 homens e 53 mulheres) tendo $AR1$ como I_T . O I_S respectivo a cada teste é apresentado na segunda linha.

D	M	$y = 0 \text{ e } x = D$			$y = 0 \text{ e } x = -D$			$x = 0 \text{ e } y = D$			$x = 0 \text{ e } y = -D$		
		AR2	AR3	AR4	AR2	AR3	AR4	AR2	AR3	AR4	AR2	AR3	AR4
0	GWF	100,0	99,15	81,20	100,0	99,15	81,20	100,0	99,15	81,20	100,0	99,15	81,20
	GW	98,29	98,29	79,49	98,29	98,29	79,49	98,29	98,29	79,49	98,29	98,29	79,49
2	GWF	99,15	97,44	73,50	98,29	97,44	78,63	94,87	98,29	76,07	94,87	99,15	79,49
	GW	97,44	98,29	68,38	96,58	97,44	74,36	91,45	96,58	68,38	92,31	99,15	73,51
4	GWF	70,09	80,34	38,46	70,09	91,45	49,57	62,39	85,47	47,86	73,50	88,03	54,70
	GW	65,81	81,19	35,04	71,79	90,60	43,59	58,97	84,62	39,32	72,65	82,91	52,14
6	GWF	29,06	35,04	5,13	25,64	42,74	20,51	20,51	37,61	15,38	35,90	56,41	35,90
	GW	29,91	31,62	76,92	26,50	44,44	17,09	18,80	32,48	16,24	33,33	51,28	33,33
8	GWF	8,55	10,26	5,13	8,55	13,68	7,69	5,13	8,55	8,55	7,69	32,49	15,38
	GW	7,69	8,55	0,85	5,13	5,13	7,69	6,84	8,55	8,55	10,26	27,35	14,53

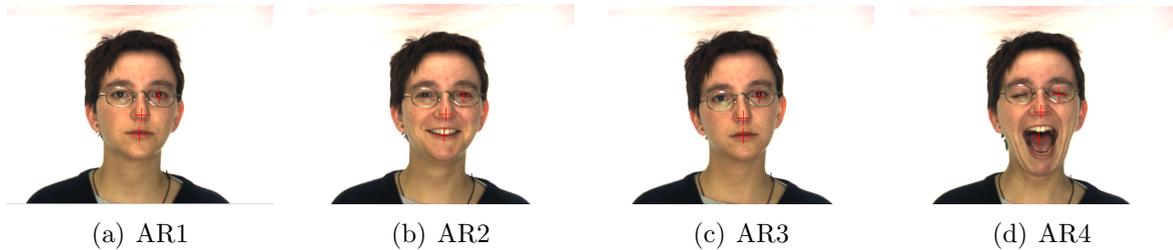


Figura 45: Exemplo de fotos dos conjuntos AR1, AR2, AR3 e AR4, tiradas na primeira sessão. (a) Expressão neutra, (b) Sorrindo, (c) Zangado, (d) Gritando. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

Conclusões

Após uma análise dos resultados obtidos para os testes de erros de localização (Tabela 13), foi possível perceber que, quanto maior a distância da posição real das características, menor é a taxa de identificação.

Os dois métodos apresentaram uma robustez, de aproximadamente quatro pixels, quanto ao erro de localização. Entretanto, quanto maior a variação na expressão (AR4 tem maior variação na expressão do que AR2 e AR3), mais rápido a taxa de acerto diminuiu com a distância.

Vale lembrar que esses testes avaliaram o pior caso, se houvesse erro de localização em todas as características.

7.1.6 Teste de Tamanho da Base de Imagens

Esse teste é do tipo *Conjunto fechado* e teve como objetivo avaliar como o algoritmo de identificação responde à base de imagens com diferentes tamanhos. Para isso, utilizou-se o conjunto de fotos AR1 como I_T e AR4 como I_S variando-se seus tamanhos. Primeiro realizou-se o teste com a base completa (117 indivíduos). Depois, diminuiu-se o tamanho de I_T e I_S . A diminuição foi feita de forma a manter as imagens classificadas erroneamente com a base completa, tanto nos testes com foveação, quanto nos sem. Pois assim, foi possível perceber o quanto o número de imagens afetou a capacidade de identificação. O conjunto AR4 foi escolhido como I_S , por ter apresentado um número maior de erros com a base completa.

A Tabela 14 apresenta os resultados obtidos nos testes realizados com 117, 80 e 40 indivíduos do banco de imagens *AR Face Database*. A Figura 46 apresenta exemplos de fotos utilizadas nos testes de tamanho da base de dados.

Tabela 14: Testes de tamanho da base de dados. Número de erros para identificação realizada com a base de imagens em diferentes tamanhos (117, 80 e 40 indivíduos). A primeira coluna, M, determina o método, como definido em 7.1. Para a realização desses testes utilizou-se AR1 como I_T e AR4 como I_S .

TW	117	80	40
GWF	22	20	17
GW	24	25	19



Figura 46: Exemplo de fotos dos conjuntos AR1 e AR4, tiradas na primeira sessão. (a) Expressão neutra, (b) Expressão neutra, (c) Gritando, (d) Gritando. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

Conclusões

Analisando-se os resultados do teste de tamanho da base de imagens (Tabela 14) percebe-se, que a diminuição dela diminuiu um pouco o número de erros cometidos pelo sistema. Porém ela não foi suficientemente significativa para dizer que os indivíduos retirados do teste foram responsáveis pela quantidade de erros cometidos. Isso permite inferir que o aumento do número de indivíduos não introduzirá uma sobreposição entre as classes capaz de invalidar a capacidade de reconhecimento dos dois métodos.

7.2 Testes de Localização de Características

Os testes de localização foram elaborados de forma a avaliar o algoritmo GWF em diferentes condições. Todos eles seguiram a parametrização definida na seção 6. A seguir serão apresentados alguns testes realizados.

7.2.1 Teste de Localização do Nariz

Esse teste teve como objetivo avaliar como o algoritmo de localização responde em diferentes situações para achar o nariz. Para isso, selecionou-se fotos do conjunto AR1 que fossem capazes de representar os narizes contidos na base de imagens. Um total de 7 indivíduos foi dado como suficiente para constituir um I_T bem representativo. Eles foram escolhidos de forma a abranger uma boa variação de narizes. O ponto inicial de busca, p_1 , foi definido como sendo o canto superior esquerdo da imagem em questão. Para geração dos resultados estipulou-se uma área em torno da posição, manualmente localizada, do nariz. Essa área foi definida a partir de um raio mínimo e um raio máximo. A Figura 47 ilustra alguns dos indivíduos utilizados para o treinamento, I_T .

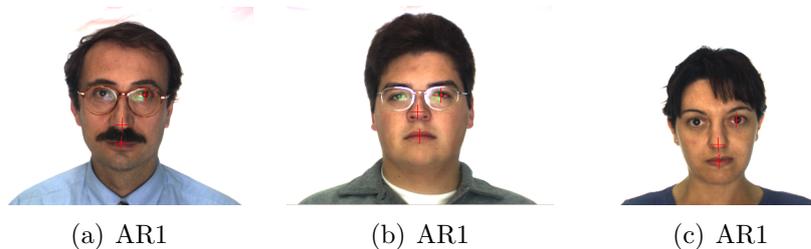


Figura 47: Exemplo de algumas fotos do conjunto AR1 utilizadas para o treinamento da localização do nariz. (a) Expressão neutra, (b) Expressão neutra, (c) Expressão neutra. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

As Tabelas 15 e 16 apresentam os resultados obtidos nos testes realizados para diferentes I_S com 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database*. As Figuras 48 e 49 apresentam alguns exemplos de narizes que não foram localizados corretamente.

Tabela 15: Testes de localização do nariz 1. Percentual de pontos contidos em uma determinada área para os diferentes I_S tendo como I_T fotos do conjunto AR1 de 7 indivíduos. A primeira coluna D , define em pixels a área (raio mínimo R_{min} e raio máximo R_{max}), em que os pontos estão contidos. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

$R_{min} < D \leq R_{max}$	AR2	AR3	AR4	AR8	AR11
$0 < D \leq 2$	30,77	35,04	29,06	28,21	34,19
$2 < D \leq 4$	40,17	34,19	42,74	28,21	35,90
$4 < D \leq 6$	22,22	21,37	17,09	26,50	15,38
$6 < D \leq 8$	5,13	4,27	5,13	5,13	3,42
$8 < D \leq 10$	0,00	0,00	0,00	0,00	0,00
$10 < D \leq \infty$	1,71	5,13	5,98	11,97	11,11

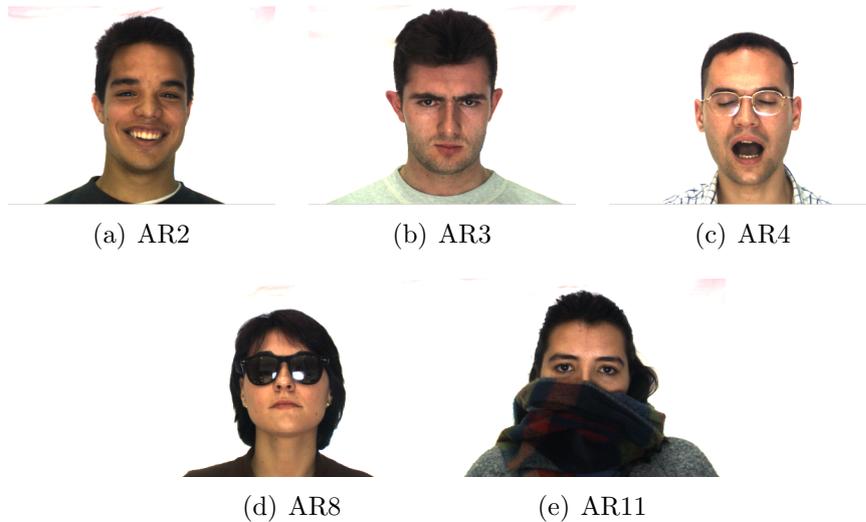


Figura 48: Exemplo de narizes localizados erroneamente, $D > 10$ pixels.

Tabela 16: Testes de localização do nariz 2. Percentual de pontos contidos em uma determinada área para os diferentes I_S tendo como I_T fotos do conjunto AR1 de 7 indivíduos. A primeira coluna D , define em pixels a área (raio mínimo R_{min} e raio máximo R_{max}), em que os pontos estão contidos. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

$R_{min} < D \leq R_{max}$	AR14	AR15	AR16	AR17	AR21	AR24
$0 < D \leq 2$	29,91	34,19	26,50	27,35	22,22	24,79
$2 < D \leq 4$	43,59	41,88	40,17	41,03	41,03	46,15
$4 < D \leq 6$	21,37	14,53	23,08	18,80	23,08	17,09
$6 < D \leq 8$	25,64	5,98	5,13	5,13	4,27	5,13
$8 < D \leq 10$	0,00	0,00	0,00	1,71	2,56	0,85
$10 < D \leq \infty$	2,56	3,42	5,13	5,98	6,84	5,98

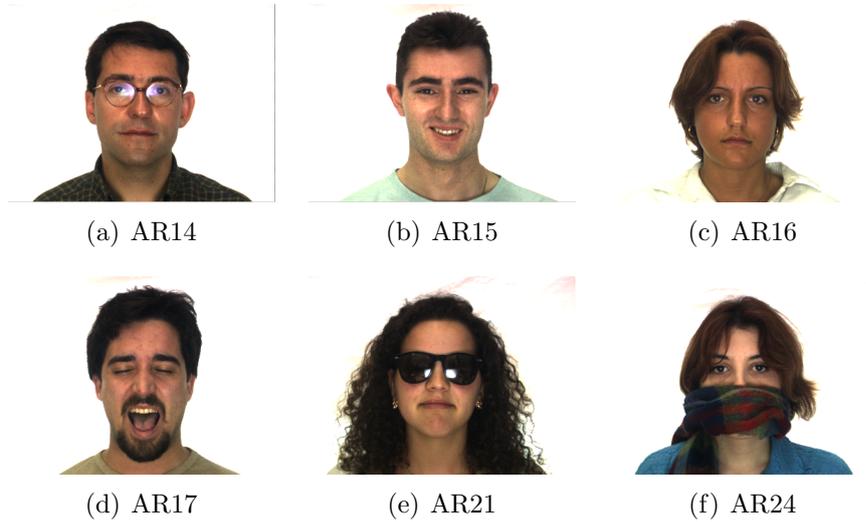
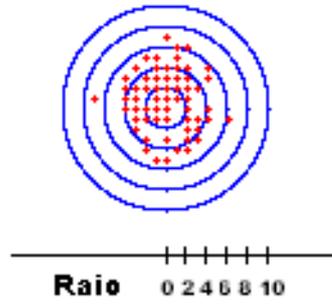
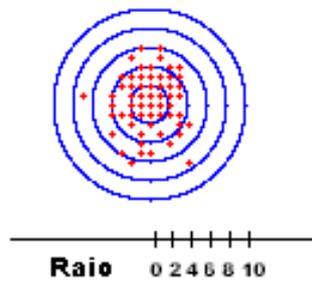


Figura 49: Exemplo de narizes localizados erroneamente, $D > 10$ pixels.

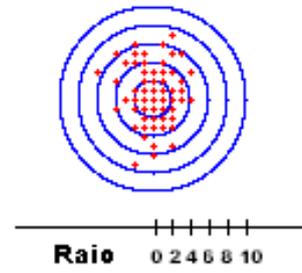
As Figuras 50 e 51 ilustram o espalhamento dos pontos localizados em cada ISON-DAGEM.



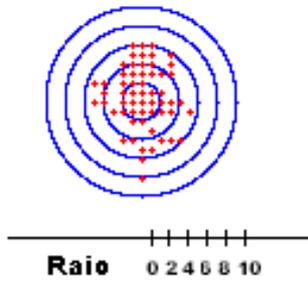
(a) AR2



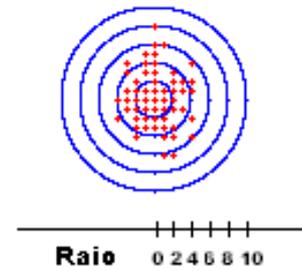
(b) AR3



(c) AR4



(d) AR8



(e) AR11

Figura 50: 1 Espalhamento dos pontos do nariz localizados em cada I_S . Os círculos azuis representam as áreas formadas pelos raios. Os pontos vermelhos representam os pontos localizados. O centro representa a posição do nariz manualmente localizado de cada imagem. A escala é dada em pixels.

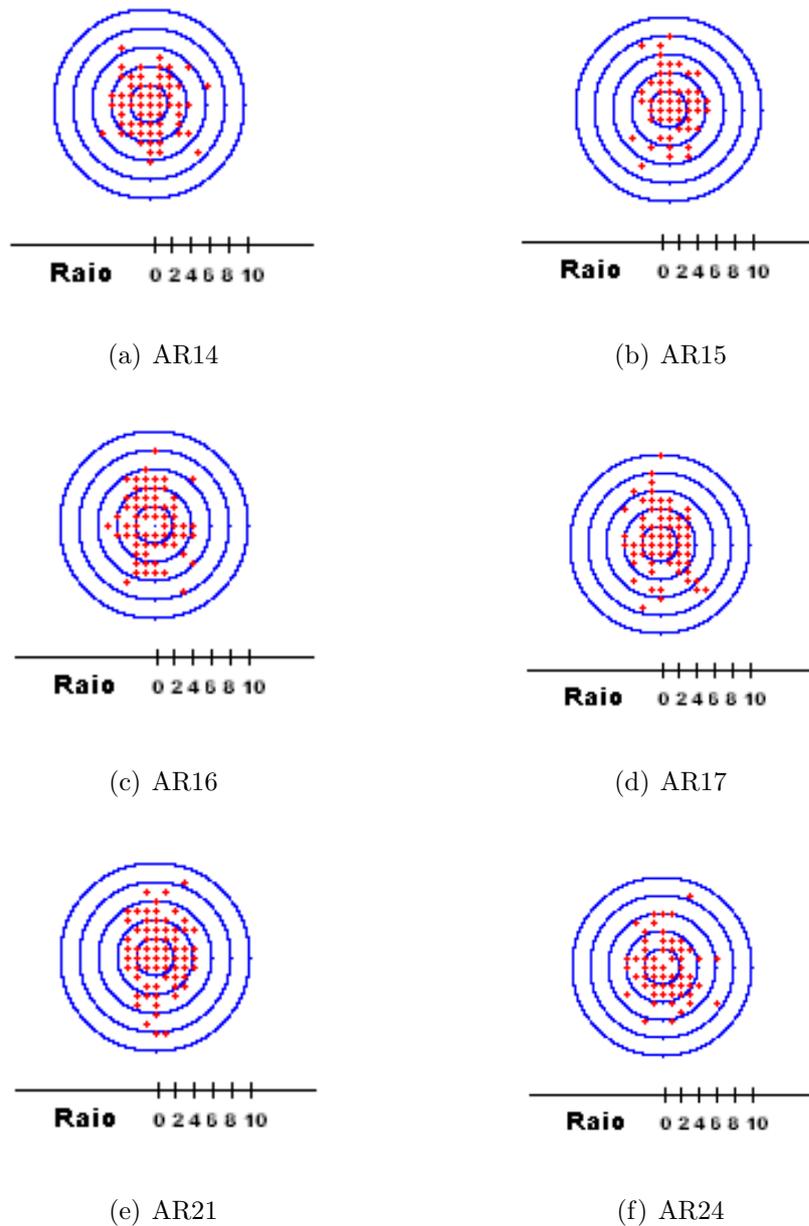


Figura 51: 2 Espalhamento dos pontos do nariz localizados em cada I_S . Os círculos azuis representam as áreas formadas pelos raios. Os pontos vermelhos representam os pontos localizados. O centro representa a posição do nariz manualmente localizado de cada imagem. A escala é dada em pixels.

Conclusões

Após uma análise dos resultados obtidos para os testes de localização do nariz, foi possível chegar a algumas conclusões apresentadas nos itens abaixo:

- **Diferentes expressões:** A diferença de expressão (AR2, AR3 e AR4 na Tabela 15) não afetou muito a localização. Isso, porque o nariz não sofreu grandes alterações

nas expressões avaliadas.

- **Iluminação:** Não realizou-se testes com diferenças na iluminação, porque a base de imagens utilizada contém fotos com a quantidade de luz muito alta. Isso causou a saturação de regiões claras implicando na perda de informações de frequência. O método não foi capaz de localizar nariz com alta luminosidade. Porém, alguns testes realizados durante a implementação mostraram que ele lida bem com a variação de luz quando não há muita perda de informação de frequência, uma das conseqüências da saturação.
- **Oclusões:** As oclusões (AR8 e AR11 na Tabela tab:testeLocalizacaoNariz1) afetaram um pouco a localização do nariz. Pois em algumas imagens, ela escondeu parte da região que representava o nariz.
- **Diferença temporal:** A diferença temporal da primeira para a segunda sessão (AR14, AR15, AR16, AR17, AR21 e AR24 na Tabela 16) não afetou a localização do nariz. Em alguns casos a localização foi até melhor.

Apesar do método não lidar bem com a iluminação, pode-se dizer que ele é eficiente na localização do nariz. Pois para condições normais a maior parte dos pontos foi localizada com uma distância menor do que 6 pixels da posição manualmente localizada.

7.2.2 Teste de Localização do Olho Esquerdo

Esse teste teve como objetivo avaliar como o algoritmo de localização responde em diferentes situações para achar o olho esquerdo. Para isso, selecionou-se fotos do conjunto AR1 que fossem capazes de representar os olhos contidos na base de imagens. Um total de 17 indivíduos foi dado como suficiente para constituir um I_T bem representativo. Eles foram escolhidos de forma a abranger uma boa variação de olhos. O ponto inicial de busca, p_1 , foi definido como sendo a posição do nariz do indivíduo em teste mais a posição do olho, em relação ao nariz, de algum dos indivíduos de I_T . Para geração dos resultados estipulou-se uma área em torno da posição, manualmente localizada, do olho. Essa área foi definida a partir de um raio mínimo e um raio máximo. A Figura 52 ilustra alguns dos indivíduos utilizados para o treinamento, I_T .

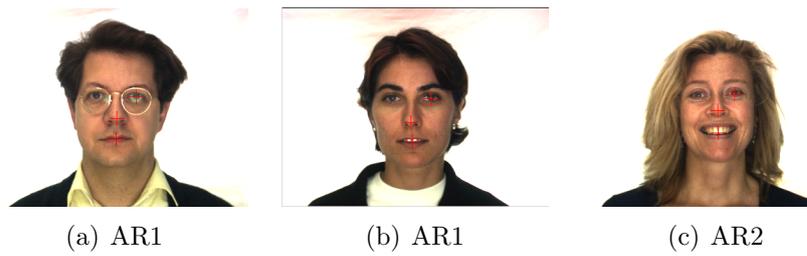


Figura 52: Exemplo de algumas fotos dos conjuntos AR1 e AR2 utilizadas para o treinamento da localização do olho esquerdo. (a) Expressão neutra, (b) Expressão neutra, (c) Expressão sorrindo. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

As Tabelas 17 e 18 apresentam os resultados obtidos nos testes realizados para diferentes I_S com 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database*. As Figuras 53 e 54 apresentam alguns exemplos de olhos que não foram localizados corretamente.

Tabela 17: Testes de localização do olho esquerdo 1. Percentual de pontos contidos em uma determinada área para os diferentes I_S tendo como I_T fotos do conjunto AR1 de 17 indivíduos. A primeira coluna D, define em pixels a área (raio mínimo R_{min} e raio máximo R_{max}), em que os pontos estão contidos. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

$R_{min} < D \leq R_{max}$	AR2	AR3	AR5	AR6	AR7	AR11	AR12	AR13
$0 < D \leq 2$	13,68	13,68	23,93	15,38	15,38	14,53	19,66	10,26
$2 < D \leq 4$	46,15	44,44	52,14	33,33	34,19	52,99	47,01	31,62
$4 < D \leq 6$	27,35	30,77	18,80	23,93	23,93	27,35	26,50	27,35
$6 < D \leq 8$	7,692	2,56	0,85	8,55	7,69	1,71	1,71	6,84
$8 < D \leq 10$	1,71	0,85	0,85	2,56	1,71	0,00	1,71	3,42
$10 < D \leq \infty$	3,42	7,69	3,42	16,24	17,09	3,42	3,42	20,51



Figura 53: Exemplo de olhos localizados erroneamente, $D > 10$ pixels.

Tabela 18: Testes de localização do olho esquerdo 2. Percentual de pontos contidos em uma determinada área para os diferentes I_S tendo como I_T fotos do conjunto AR1 de 17 indivíduos. A primeira coluna D , define em pixels a área (raio mínimo R_{min} e raio máximo R_{max}), em que os pontos estão contidos. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

$R_{min} < D \leq R_{max}$	AR14	AR15	AR16	AR18	AR19	AR20	AR24	AR25	AR26
$0 < D \leq 2$	22,22	16,24	14,53	19,66	6,84	9,40	22,22	19,66	11,11
$2 < D \leq 4$	54,70	52,99	39,32	54,70	39,32	38,46	51,28	50,43	29,91
$4 < D \leq 6$	17,95	21,37	26,50	17,95	26,50	20,51	18,80	22,22	23,08
$6 < D \leq 8$	2,56	3,42	6,84	3,42	10,26	10,26	2,56	1,71	7,69
$8 < D \leq 10$	0,85	0,00	1,71	0,00	1,71	1,71	0,85	0,00	0,85
$10 < D \leq \infty$	1,71	5,98	11,11	4,27	15,38	19,66	4,27	5,98	27,35

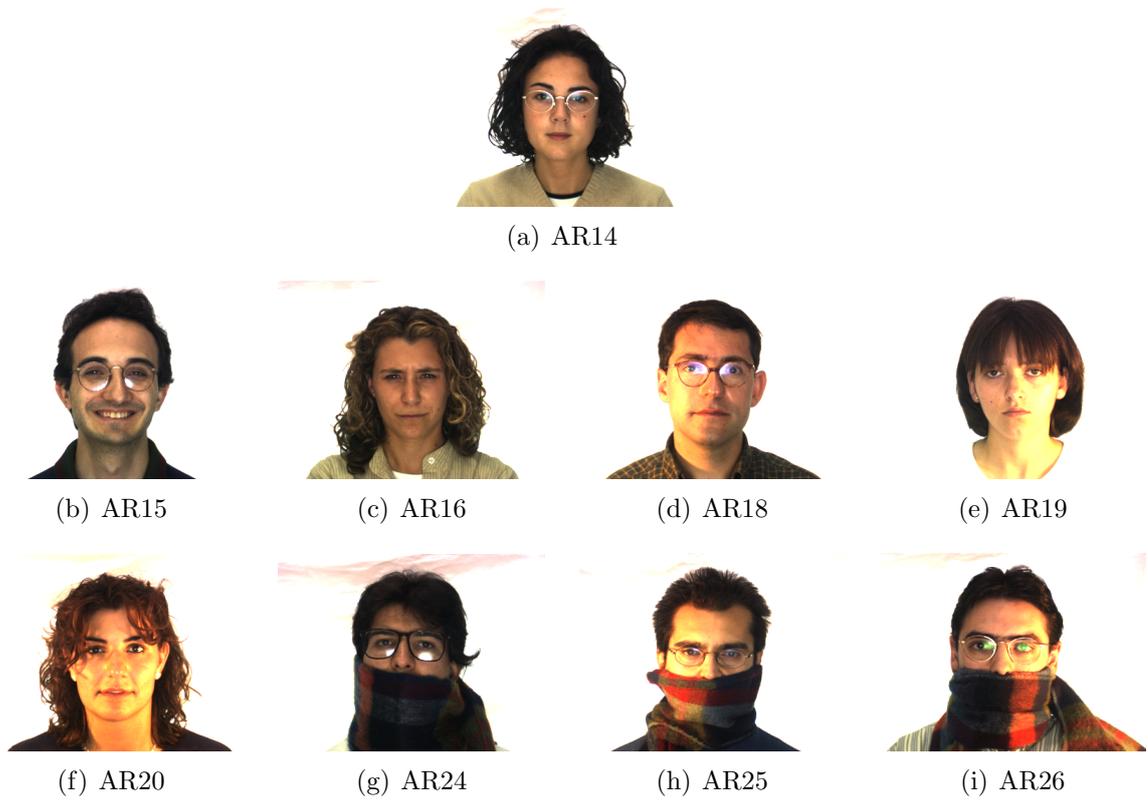
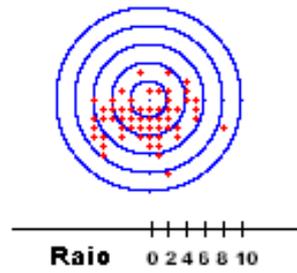
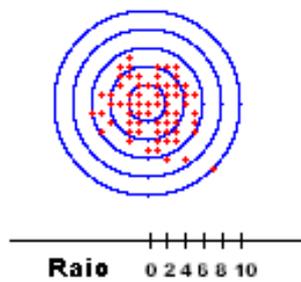


Figura 54: Exemplo de olhos localizados erroneamente, $D > 10$ pixels.

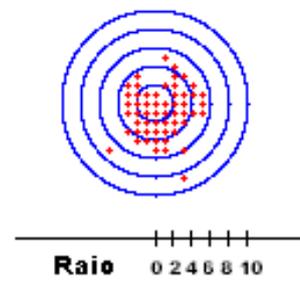
As Figuras 55, 56 e 57 ilustram o espalhamento dos pontos localizados em cada ISON-DAGEM.



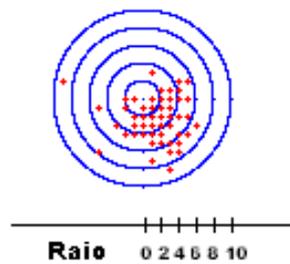
(a) AR2



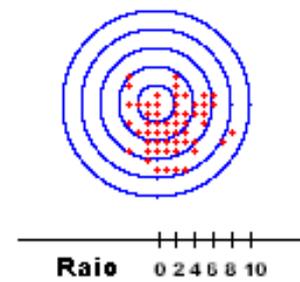
(b) AR3



(c) AR5

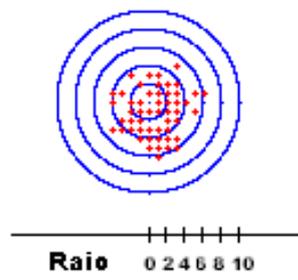


(d) AR6

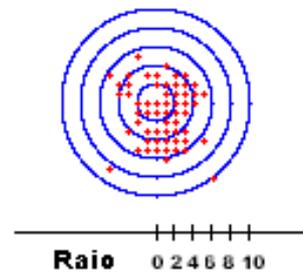


(e) AR7

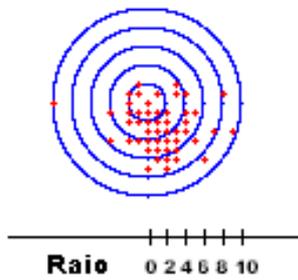
Figura 55: 1 Espalhamento dos pontos do olho esquerdo localizados em cada I_S . Os círculos azuis representam as áreas formadas pelos raios. Os pontos vermelhos representam os pontos localizados. O centro representa a posição do olho esquerdo manualmente localizado de cada imagem. A escala é dada em pixels.



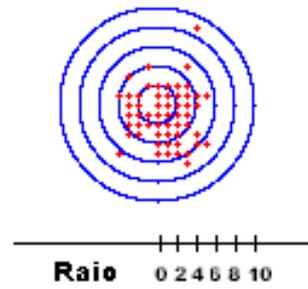
(a) AR11



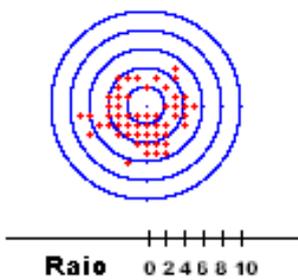
(b) AR12



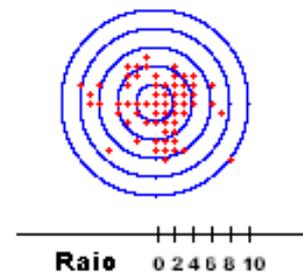
(c) AR13



(d) AR14



(e) AR15



(f) AR16

Figura 56: 2 Espalhamento dos pontos do olho esquerdo localizados em cada I_S . Os círculos azuis representam as áreas formadas pelos raios. Os pontos vermelhos representam os pontos localizados. O centro representa a posição do olho esquerdo manualmente localizado de cada imagem. A escala é dada em pixels.

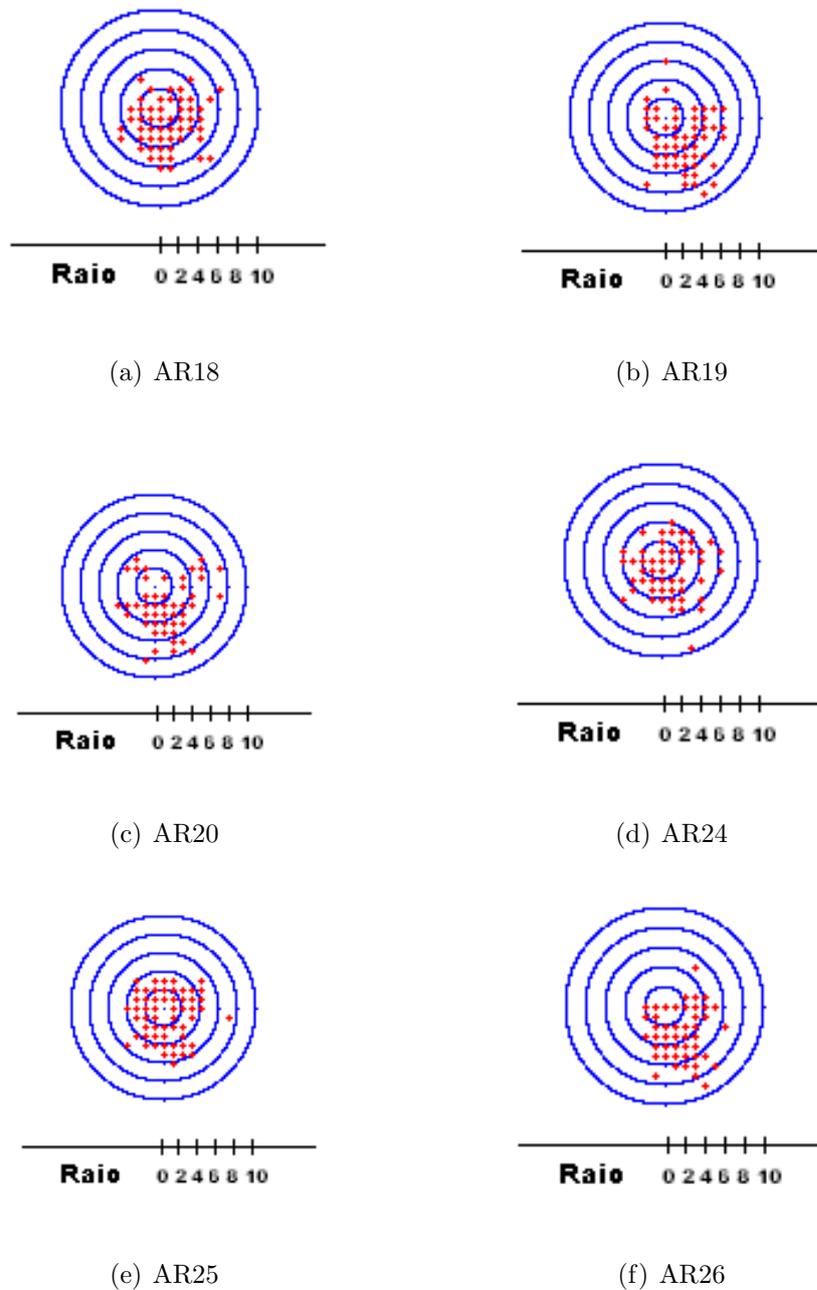


Figura 57: 3 Espalhamento dos pontos do olho esquerdo localizados em cada I_S . Os círculos azuis representam as áreas formadas pelos raios. Os pontos vermelhos representam os pontos localizados. O centro representa a posição do olho esquerdo manualmente localizado de cada imagem. A escala é dada em pixels.

Conclusões

Após uma análise dos resultados obtidos para os testes de localização do olho esquerdo, foi possível chegar a algumas conclusões apresentadas nos itens abaixo:

- **Diferentes expressões:** A diferença de expressão foi significativa para a localização

do olho esquerdo. O método lidou bem com os olhos abertos (AR2 e AR3 na Tabela 17), entretanto, para olhos fechados, como em algumas fotos de AR4 e AR17, a localização não foi boa. A grande diferença entre olhos abertos e fechados introduz ruídos para a localização dos olhos abertos, quando ambas são utilizadas no conjunto de treino. Isso faz com que sobrancelhas parecidas com cílios de olhos fechados sejam localizadas no lugar de alguns olhos abertos. Para evitar esse ruído, a expressão gritando (AR4 e AR17) não participou dos testes.

- **Iluminação:** Como o olho contém grandes partes escuras, a informação de frequência não foi completamente destruída. Isso possibilitou a localização de grande parte deles, mesmo assim, as imagens de alta luminosidade (AR6, AR7, AR13, AR19, AR20 e AR26 nas Tabelas 17 e 18) foram as com índice de erro mais alto.
- **Oclusões:** As oclusões de óculos de sol impediram a localização dos olhos sendo portanto eliminadas do teste.
- **Diferença temporal:** A diferença temporal da primeira para a segunda sessão (AR14, AR15, AR16, AR17, AR21 e AR24) não mostrou grande diferença nas taxas de localização do olho esquerdo.

Apesar do método não lidar bem com certas variações, pode-se dizer que ele é eficiente na localização do olho esquerdo. Pois para condições normais a maior parte dos pontos foi localizada com uma distância menor do que 6 pixels da posição manualmente localizada.

7.2.3 Teste de Localização da Boca

Esse teste teve como objetivo avaliar como o algoritmo de localização responde em diferentes situações para achar a boca. Para isso, selecionou-se fotos do conjunto AR1 e AR2 que fossem capazes de representar as bocas contidas na base de imagens. Um total de 16 indivíduos de AR1 e 5 de AR2 foi dado como suficiente para constituir um I_T bem representativo. Eles foram escolhidos de forma a abranger uma boa variação da boca. O ponto inicial de busca, p_1 , foi definido como sendo a posição do nariz do indivíduo em teste mais a posição da boca, em relação ao nariz, de algum dos indivíduos de I_T . Para geração dos resultados estipulou-se uma área em torno da posição, manualmente localizada, da boca. Essa área foi definida a partir de um raio mínimo e um raio máximo. A Figura 58 ilustra alguns dos indivíduos utilizados para o treinamento, I_T .

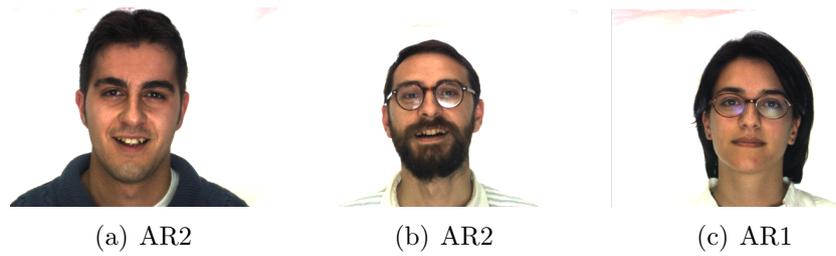


Figura 58: Exemplo de algumas fotos dos conjuntos AR1 e AR2 utilizadas para o treinamento da localização da boca. (a) Expressão sorrindo, (b) Expressão sorrindo, (c) Expressão neutra. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.

A Tabela 19 apresenta os resultados obtidos nos testes realizados para diferentes I_S com 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database*. A Figura 59 apresenta alguns exemplos de bocas que não foram localizadas corretamente.

Tabela 19: Testes de localização da boca. Percentual de pontos contidos em uma determinada área para os diferentes I_S tendo como I_T 16 fotos do conjunto AR1 e 5 do conjunto AR2. A primeira coluna D , define em pixels a área (raio mínimo R_{min} e raio máximo R_{max}), em que os pontos estão contidos. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

$R_{min} < D \leq R_{max}$	AR2	AR3	AR8	AR14	AR15	AR16	AR21
$0 < D \leq 2$	10,26	20,51	20,51	21,37	15,38	19,66	15,38
$2 < D \leq 4$	41,88	41,88	44,44	41,03	40,17	34,19	33,33
$4 < D \leq 6$	24,79	20,51	14,53	24,79	27,35	25,64	35,90
$6 < D \leq 8$	14,53	6,838	10,26	5,98	8,55	4,27	5,13
$8 < D \leq 10$	3,42	4,27	5,98	2,56	3,42	7,69	3,42
$10 < D \leq \infty$	5,13	5,98	4,27	4,27	5,13	8,55	6,84

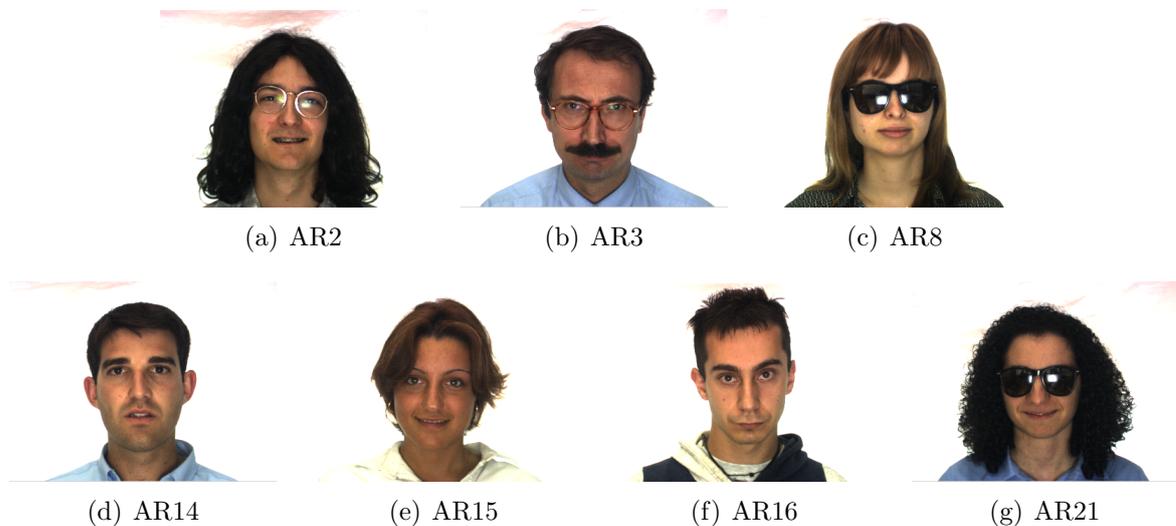
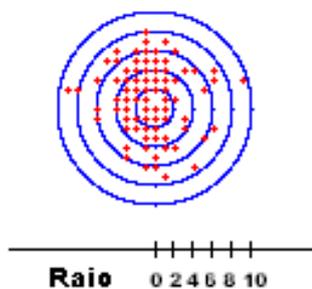
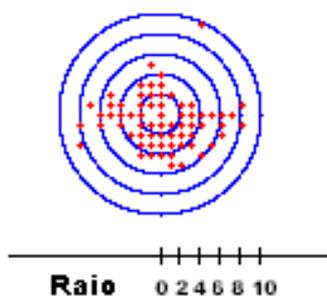


Figura 59: Exemplo de bocas localizadas erroneamente, $D > 10$ pixels.

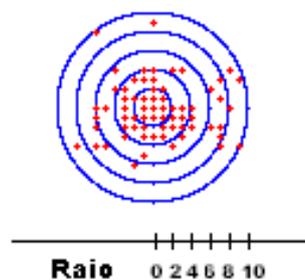
As Figuras 60, 61 e 57 ilustram o espalhamento dos pontos localizados em cada ISON-DAGEM.



(a) AR2



(b) AR3



(c) AR8

Figura 60: 1 Espalhamento dos pontos da boca localizados em cada I_S . Os círculos azuis representam as áreas formadas pelos raios. Os pontos vermelhos representam os pontos localizados. O centro representa a posição da boca manualmente localizada de cada imagem. A escala é dada em pixels.

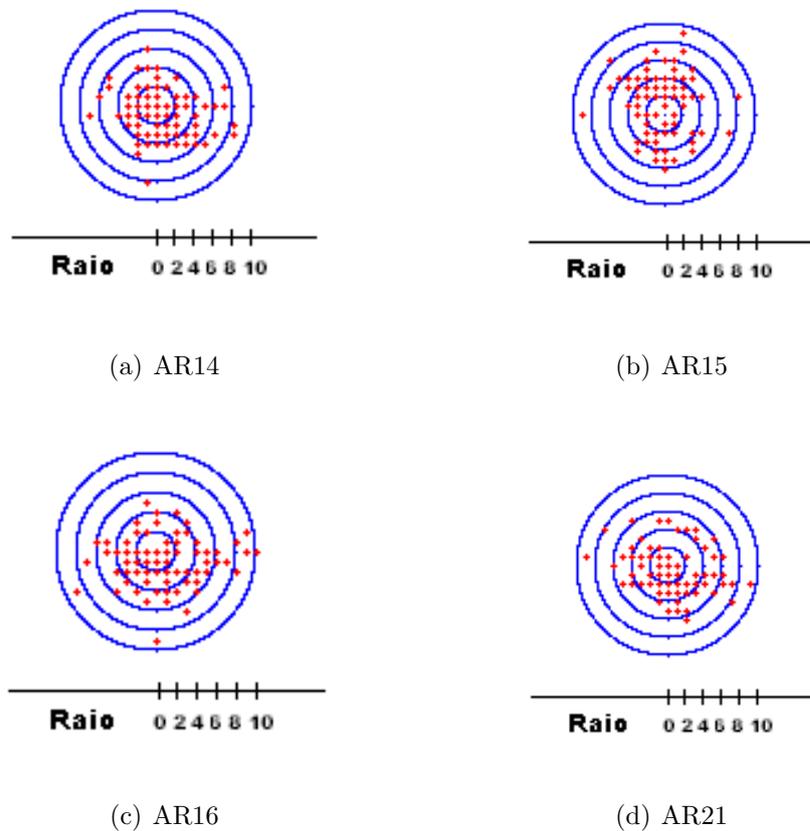


Figura 61: 2 Espalhamento dos pontos da boca localizados em cada I_S . Os círculos azuis representam as áreas formadas pelos raios. Os pontos vermelhos representam os pontos localizados. O centro representa a posição da boca manualmente localizada de cada imagem. A escala é dada em pixels.

Conclusões

Após uma análise dos resultados obtidos para os testes de localização da boca, foi possível chegar a algumas conclusões apresentadas nos itens abaixo:

- **Diferentes expressões:** A diferença de expressão foi significativa para a localização da boca. O método lidou bem com a boca fechada ou sorrindo (AR2 e AR3 na Tabela 19), entretanto, para boca aberta gritando (AR4 e AR17) o percentual de localização foi baixo. A grande diferença entre boca fechada e aberta introduz ruídos para a localização da boca, quando ambas são utilizadas no conjunto de treino. Para evitar esse ruído na localização, a expressão gritando (AR4 e AR17) não participou dos testes.
- **Iluminação:** Assim como o nariz, a localização da boca não foi possível na presença de alta luminosidade. Portanto, fotos com incidência de luz não participaram dos testes.

- **Oclusões:** As oclusões de cachecol impediram a localização da boca sendo portanto eliminadas do teste.
- **Diferença temporal:** A diferença temporal da primeira para a segunda sessão (AR14, AR15, AR16 e AR21 na Tabela 19) não mostrou grandes variações nas taxas de localização da boca.

Apesar do método não lidar bem com certas variações, pode-se dizer que ele é eficiente na localização da boca em condições normais, pois a maior parte dos pontos foi localizada com uma distância menor do que 6 pixels da posição manualmente localizada. Contudo, uma melhor abordagem deve ser utilizada a fim de evitar que variações como boca aberta e fechada influenciem na localização, uma vez que o sistema não lidou bem com elas. Ao invés de utilizar o centro da boca, como feito nesse teste, pode-se utilizar o canto da boca ou a parte do lábio inferior.

7.2.4 Teste de Rotação

O teste de rotação visa introduzir a possibilidade de descobrir a rotação de uma determinada característica da face. Escolheu-se o nariz como teste, por ser a primeira das três características a ser procurada. Para isso, selecionou-se 7 fotos do conjunto AR1 (as mesmas utilizadas na localização do nariz) para constituírem I_T . O ponto avaliado foi a posição do nariz manualmente localizada. Todas as imagens contidas em I_S foram rotacionadas de 90° em relação a posição manualmente localizada. As respostas do teste foram divididas em três grupos: correta (C), com $\theta = 90^\circ$; aproximadamente correta (AC), com $\theta = 80^\circ$ ou $\theta = 100^\circ$; errada (E), qualquer outro valor. A Figura 62 ilustra alguns dos indivíduos utilizados para o treinamento, I_T , e a Figura 63 ilustra a rotação de 90° aplicada nas fotos de sondagem, I_S .

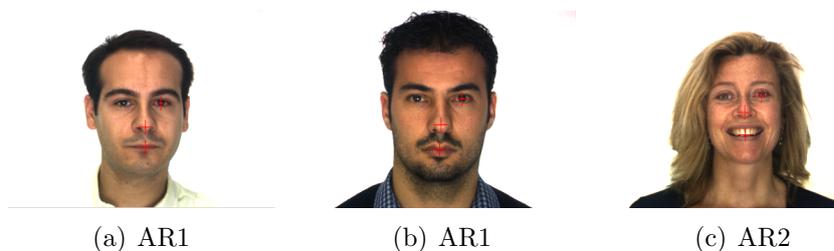


Figura 62: Exemplo de algumas fotos dos conjuntos AR1 e AR2 utilizadas para o treinamento do teste de rotação através do nariz. (a) Expressão neutra, (b) Expressão neutra, (c) Expressão sorrindo. Os pontos manualmente localizados estão evidenciados com uma cruz (+) vermelha.



Figura 63: Exemplo da rotação de 90° aplicada nas fotos utilizadas durante a sondagem, I_S .

A Tabela 20 apresenta os resultados obtidos nos testes realizados para diferentes I_S com 117 indivíduos (64 homens e 53 mulheres) do banco de imagens *AR Face Database*. A Figura 64 apresenta alguns exemplos de rotações que não foram identificadas corretamente.

Tabela 20: Testes de rotação. Percentual de fotos com um determinado tipo de resposta para os diferentes I_S tendo como I_T fotos do conjunto AR1 de 7 indivíduos. A segunda coluna TR, apresenta o tipo de resposta. Foram utilizados 117 indivíduos (64 homens e 53 mulheres) para a realização desses testes.

TR	AR2	AR3	AR4	AR8	AR11
C	67,52	73,50	66,67	51,28	69,23
AC	29,91	25,64	29,91	47,01	25,64
E	2,56	0,85	3,42	1,71	5,13

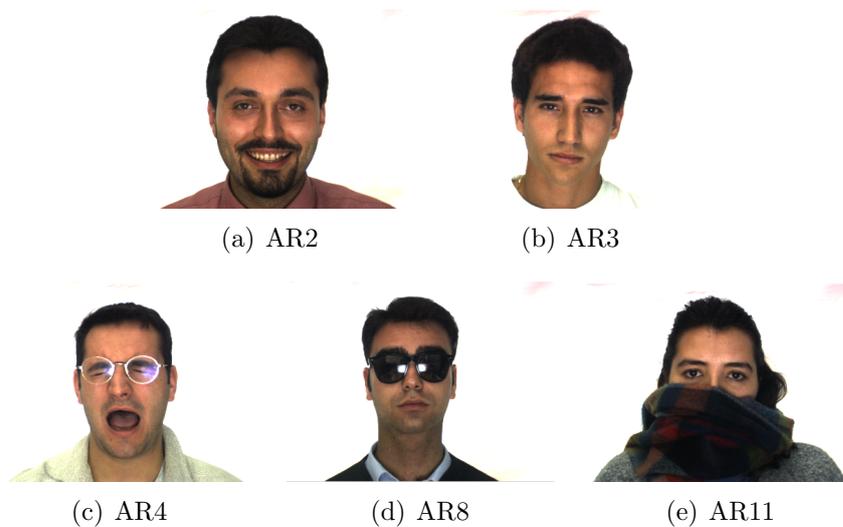


Figura 64: Exemplo de rotações identificadas erroneamente (E), ângulo maior do que 100° ou menor do que 80° .

A Figura 65 apresenta uma face com a cabeça levemente rotacionada.



Figura 65: Exemplo de uma Figura com inclinação na cabeça mesmo com a imagem em 0° .

Conclusões

Após uma análise dos resultados obtidos para os testes de rotação (Tabela 20), foi possível perceber que a maior parte das respostas identificou corretamente o ângulo de rotação do nariz. Algumas delas chegaram bem próximas errando apenas por um passo (o passo definido nessa configuração de foi de 10°). Um dos motivos para isso, é uma pequena inclinação existente na face de alguns indivíduos, como ilustrado na Figura 65. Apenas uma pequena parcela das fotos não foi corretamente classificada, o que permite confirmar a eficiência do método em identificar rotações.

8 *Conclusão e Trabalhos Futuros*

Após a análise dos vários aspectos estudados nesse trabalho, foi possível chegar à algumas conclusões a serem descritas a seguir.

Os algoritmos de identificação apresentados (GWF e GW) desempenharam otimamente o seu papel. Um destaque especial é dado para a função de similaridade entre as faces que ponderou bem as variações (temporal, iluminação, expressões e oclusões). Ela permitiu que, de maneira geral, os resultados tivessem altos índices de identificação. Mais altos inclusive do que os apresentados em alguns trabalhos na literatura. Além disso, ela permitiu um controle da confiabilidade do sistema possibilitando sua aplicação em situações, na quais haja a presença de intrusos. Porém, nesse caso deve-se abrir mão da flexibilidade dele quanto às variações.

Quando comparados, os dois métodos apresentaram resultados bem próximos (seção 7.1). Sendo que na maioria dos testes o GWF comportou-se ligeiramente melhor (de 2% a 4%) do que o GW. Isso mostrou que a compactação obtida na informação através da foveação, ao invés de atrapalhar a representatividade das regiões pelas Gabors Wavelets acabou melhorando-a na maioria dos casos. Dado que a representatividade do GWF melhorou em relação à do GW durante o processo de indentificação de indivíduos, foi possível utilizar o GWF ao invés do GW durante a localização de características. Isso tornou possível a redução do esforço computacional devido ao grande numero de pós-processamento (processamento realizado após a aplicação dos filtros DOG) requerido na localização, como por exemplo quantidade de Wavelets utilizadas e número de graus de rotação. Pois ao invés de aplicar novamente as Wavelets de Gabor na imagem (algoritmo GW), elas foram regeradas a partir de informações retornadas pelas DOGs (algoritmo GWF), que são em menor quantidade (número de pontos a serem processados na Figura 27 é bem maior do que os da Figura 28).

A sacada foi fundamental na redução do esforço computacional, durante a busca, para a realização dos testes de localização de características. Ela evitou que pontos contidos em regiões de baixa similaridade fossem amostrados desnecessariamente direcionando a atenção do sistema para as regiões de interesse. A utilização de uma região de amostragem

do tipo logpolar permitiu que as características fossem localizadas com precisão, dado que o centro da região de amostragem é bem denso; e em diferentes partes da imagem, dado que a periferia da região de amostragem é esparsa e ampla. Com isso, a atenção pode ser direcionada para qualquer região da imagem a fim de fazer um refinamento na localização. Como o tempo computacional para realização dos testes foi reduzido em mais de cinco vezes, pode-se concluir que o método de localização apresentado aqui é dependente de uma busca otimizada e que a sacada cumpriu bem esse papel.

A localização de características (seção 7.2) se mostrou muito bem quando não houve grandes mudanças entre o conjunto de treino e o de teste. Entretanto, não obteve resultados tão flexíveis quanto a parte de identificação. Pois, não foi capaz de lidar com variações (expressão, luz e oclusões) muito fortes. Isso torna essas variações dependentes do tipo de modelagem do problema. Pois em uma modelagem mais simplória pode-se definir como característica olho, nariz e boca. Contudo, um olho aberto em nada se parece com um olho fechado, o que permite expandi-los em duas novas classes de características olho aberto e olho fechado. O mesmo ocorre com a boca, que pode ser dividida em boca fechada, boca aberta, boca sorrindo e etc. Essa sub-divisão de classes com a participação do usuário do sistema na escolha do tipo de foto, por hora, é necessária, porque, diferentemente da parte de identificação que possui outras regiões a serem ponderadas, a parte de localização só conta com a região da própria característica.

Os testes de rotação (seção 7.2.4) mostraram ser possível identificar o ângulo de rotação de uma característica na imagem. Pois na maioria dos casos o ângulo foi corretamente identificado ou ficou muito próximo do valor real. A utilização do GWF nesses testes permitiu uma redução do esforço computacional, pois para cada região analisada as DOGs só foram aplicadas uma vez sendo o processamento posterior, isto é, geração das Wavelets em diferentes rotações, feito a partir da saída gerada por elas. Os resultados obtidos possibilitam que futuramente ela seja incorporada no processo de localização a fim de conseguir identificar faces com diferentes ângulos de rotação.

Embora os métodos aqui apresentados tenham desempenhado bem o papel para o qual foram propostos, ainda existem muitos pontos a serem investigados a fim de torná-los cada vez mais robustos às diversidades encontradas no dia a dia do mundo visual. Abaixo são listados alguns itens a serem estudados em trabalhos futuros:

- **Retina:** Aproveitar mais as regiões periféricas da retina para uma melhor representação de um determinado ponto na imagem. Pode-se também, utilizar outros modelos de retina para a realização da parte de foveação;

- **Rotação:** Investigar outras formas de processar a informação proveniente da retina para a identificação da rotação e unificá-la com a localização das características;
- **Escalamento:** Investigar a possibilidade de obter-se informações de escalamento da imagem, assim como foi feito para a rotação;
- **Sacada:** Aproveitar a informação da própria retina para realizar o movimento de sacada excluindo assim o conflito conceitual com o modelo proposto por Smeraldi, Capdevielle e Bigun (1999), Smeraldi, Carmona e Bigun (2000).
- **Características:** Testar outros pontos como características de partida para a identificação. Ao invés de centro da boca, utilizar um dos cantos da boca. Ao invés de centro do olho, utilizar canto da sobrancelha. Em fim, testar pontos menos sujeitos a variações.

Referências

- BALASURIYA, L. S.; SIEBERT, J. P. An artificial retina with a self-organised retinal receptive field tessellation. In: UNIVERSITY OF GLASGOW: DEPARTMENT OF COMPUTER SCIENCE. **Proceedings of the Biologically-inspired Machine Vision**. Aberystwyth: N/A, 2003. p. 34–42.
- BALASURIYA, L. S.; SIEBERT, J. P. A biologically inspired computational vision front-end based on a self-organised pseudo-randomly tessellated artificial retina. In: UNIVERSITY OF GLASGOW: DEPARTMENT OF COMPUTER SCIENCE. **Proceedings of the International Joint Conference on Neural Networks**. Montréal, 2005.
- BERNARDINO, A. J. M. **Binocular head control with foveal vision: methods and applications**. Tese (Doutorado) — Universidade Técnica de Lisboa, Apr 2004.
- BOLDUC, M.; LEVINE, M. D. A real-time foveated sensor with overlapping receptive fields. **Real-Time Imaging**, Academic Press Ltd., London, UK, UK, v. 3, n. 3, p. 195–212, 1997.
- BOLDUC, M.; LEVINE, M. D. A review of biologically motivated space-variant data reduction models for robotic vision. **Computer Vision and Image Understanding: CVIU**, v. 69, n. 2, p. 170–184, 1998. Disponível em: <citeseer.ist.psu.edu/bolduc96review.html>.
- BOLME, D. S. **Elastic bunch graph matching**. Dissertação (Mestrado) — Colorado State University, Fort Collins, Colorado, May 2003.
- BRUNELLI, R.; POGGIO, T. Face recognition: features versus templates. In: **IEEE Trans. Pattern Analysis and Machine Intelligence**. [S.l.: s.n.], 1993. v. 15, n. 10, p. 1042–1052), OPTannotate = .
- CHANG, E.; MALLAT, S.; YAP, C. **Wavelet foveation**. 1999. Disponível em: <citeseer.ist.psu.edu/article/chang99wavelet.html>.
- CHANG, E.-C.; YAP, C. K. A wavelet approach to foveating images. In: **SCG '97: Proceedings of the thirteenth annual symposium on Computational geometry**. New York, NY, USA: ACM Press, 1997. p. 397–399.
- EDELMAN, S. **Receptive fields for vision: from hyper-acuity to object recognition**. [S.l.], 1, 1995. Disponível em: <citeseer.ist.psu.edu/article/edelman95receptive.html>.
- GONZALEZ, R. C. **Digital image processing**. 2. ed. [S.l.]: Tom Robbins, 2002.
- GUILLAMET, D.; VITRIA, J. Classifying faces with non-negative matrix factorization.

- HANSEN, T.; BARATOFF, G.; NEUMANN, H. A simple cell model with dominating opponent inhibition for robust contrast detection. **Kognitionswissenschaft**, v. 9, n. 2, p. 93–100, 2000. Disponível em: <citeseer.ist.psu.edu/hansen00simple.html>.
- KEAT, T. P. L. Design of image-based face recognition system.
- KEPENEKCI, B. **Face recognition using gabor wavelet transform**. Dissertação (Mestrado) — The Middle East Technical University, Sep 2001.
- KINSER, J. M. Foveation by a pulse-coupled neural network. **IEEE-NN**, v. 10, n. 3, p. 621, May 1999. Disponível em: <citeseer.ist.psu.edu/221611.html>.
- KOENDERINK, J. J.; DOORN, A. J. V. Visual detection of spatial contrast; influence of location in the visual field, target extent and illuminance level. In: **Biological Cybernetics**. [S.l.: s.n.], 1978. v. 30, p. 157–167.
- KRYLOV, A.; KORTCHAGINE, D. Hermite foveation. In: UNIVERSITY OF GLASGOW: DEPARTMENT OF COMPUTER SCIENCE. **Graphicon-2004**. Moscow, Russia, 2004.
- MARTINEZ, A. M.; BENAVENTE, R. The ar face database. In: **CVC Technical Report**. [S.l.: s.n.], 1998. v. 24.
- MARTINEZ, A. M.; ZHANG, Y. Subset modelling of face localization error, occlusion, and expression.
- MOVELLAN, J. R. Tutorial on gabor filters. 2002.
- NES, A. **Hybrid system for face recognition**. Dissertação (Mestrado) — Norwegian University of Science and Technology, Jun 2003.
- ROWLEY, H. A. **Neural network-based face detection**. Tese (Doutorado) — Carnegie Mellon University, Pittsburgh, PA 15213, May 1999.
- RULLEN, R. V.; THORPE, S. J. Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. **Neural Computation**, v. 13, n. 6, p. 1255–1283, 2001. Disponível em: <citeseer.ist.psu.edu/vanrullen01rate.html>.
- SAVVY, G. **GIMP SAVVY**. <http://gimp-savvy.com/BOOK/index.html?node54.html>, Acessado em: abr. 2006.
- SCHWARTZ, E. L. Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. **Biological Cybernetics**, v. 25, n. 4, p. 181–194, 1977.
- SCHWARTZ, E. L. Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. **Vision Research**, v. 20, n. 8, p. 645–669, 1980.
- SCHWARTZ, E. L. On the mathematical structure of the visuotopic mapping of macaque striate cortex. **Science**, New Series, v. 227, n. 4690, p. 1065–1066, mar. 1985.
- SCHWARTZ, E. L. Computational studies of the spatial architecture of primate visual cortex: Columns, maps, and protomaps. In: PETERS, A.; ROCKLAND, K. (Ed.). **Primary Visual Cortex in Primates**. [S.l.]: Plenum Press, 1994, (Cerebral Cortex, v. 10).

- SCHWARTZ, E. L. Topographic mapping in primate visual cortex: Anatomical and computational approaches. In: KELLY, D. (Ed.). **Visual Science and Engineering: Models and Applications**. [S.l.]: New York : Marcel Dekker, 1994, (Optical Engineering, v. 43).
- SCHWARTZ, E. L. et al. Shape recognition and inferior temporal neurons. **Proceedings of the National Academy of Sciences of the United States of America**, v. 80, n. 18, p. 5776–5778, set. 1983.
- SENIOR, A. et al. Face detection in color images. **IEEE Trans. Pattern Anal. Mach. Intell.**, IEEE Computer Society, Washington, DC, USA, v. 24, n. 5, p. 696–706, 2002.
- SMERALDI, F.; CAPDEVIELLE, N.; BIGUN, J. Facial features detection by saccadic exploration of the Gabor decomposition and Support Vector Machines. In: **Proceedings of the 11th Scandinavian Conference on Image Analysis - SCIA 99, Kangerlussuaq, Greenland**. [s.n.], 1999. I, p. 39–44. Disponível em: <citeseer.ist.psu.edu/smeraldi99facial.html>.
- SMERALDI, F.; CARMONA, O.; BIGUN, J. Saccadic search with gabor features applied to eye detection and real-time head tracking. **Image and Vision Computing**, 2000. Disponível em: <citeseer.ist.psu.edu/smeraldi98saccadic.html>.
- SOUZA, A. F. Reconhecimento facial. 2005.
- TURK, M. A.; PENTLAND, A. P. Face recognition using eigenfaces. In: **In Proceedings of IEEE Computer Vision and Pattern Recognition**. Maui, Hawaii: [s.n.], 1991. p. 586–590), OPTannotate = .
- VALOIS, R. L. de; VALOIS, K. K. de. **Spatial Vision**. [S.l.]: Oxford university press, 1990. (Oxford Psychology Series 14).
- WALLACE, R. S. et al. Space variant image processing. **Int. J. Comput. Vision**, Kluwer Academic Publishers, Hingham, MA, USA, v. 13, n. 1, p. 71–90, 1994.
- WIEGHARDT, J.; WÜRTZ, R. P.; MALSBERG, C. von der. Gabor-based feature point tracking with automatically learned constraints. In: **Proceedings ECCV 2002, Copenhagen**. [s.n.], 2002. Disponível em: <citeseer.ist.psu.edu/596771.html>.
- WILSON, S. W. On the retino-cortical mapping. **International Journal of Man-Machine Studies**, v. 18, n. 4, p. 361–389, 1983.
- WISKOTT, L. et al. Face recognition by elastic bunch graph matching. In: SOMMER, G.; DANILIDIS, K.; PAULI, J. (Ed.). **Proc. 7th Intern. Conf. on Computer Analysis of Images and Patterns, CAIP'97, Kiel**. Heidelberg: Springer-Verlag, 1997. p. 456–463. Disponível em: <citeseer.ist.psu.edu/wiskott99face.html>.
- YAN, Y.; OSADCIW, L. Fusion in modular face recognition. In: **IEEE Transactions Pattern Analysis and Machine Intelligence**. [S.l.: s.n.], 2005.
- ZHAO, W. et al. **Face recognition: A literature survey**. 2000. Disponível em: <citeseer.ist.psu.edu/article/zhao00face.html>.

APÊNDICE A – AR Face Database

O banco de imagens *AR Face Database* foi criado por *Aleix Martinez* e *Robert Benavente* no *Centro de Computação Visual - (CVC)* em *U.A.B.*. Ele contém imagens (768 * 576 pixels com 24-bit de resolução colorida) de 126 faces, nas quais 70 são homens e 56 são mulheres, em 26 expressões ou momentos, dando um total de 3276 fotos. As imagens retratam diferenças nas expressões faciais, condições de iluminação, e oclusões (óculos de sol ou cachecol). As fotos foram tiradas sob condições estritamente controladas. Não houve restrições quanto às vestimentas (óculos, roupas, etc.), maquiagem ou corte de cabelo. Cada pessoa participou de duas sessões de fotos, tomadas com duas semanas de diferença (14 dias). O mesmo grupo de fotos foi tirado em cada sessão (MARTINEZ; BENAVENTE, 1998).

Foram necessárias algumas alterações na base de imagens original, sendo elas:

- Por motivos de performance, reduziu-se pela metade o tamanho das imagens passando-as para 384 * 288 pixels.
- Como os algoritmos desse trabalho são baseados em tons de cinza (ver (GONZALEZ, 2002; SAVVY, <http://gimp-savvy.com/BOOK/index.html?node54.html>, Acessado em: abr. 2006) para maiores detalhes), foi feita uma conversão de *RGB* para *Gray Scale* utilizando a seguinte fórmula:

$$R * 0.3 + G * 0.59 + B * 0.11 \tag{A.1}$$

- Devido à necessidade das coordenadas (x, y) de três características da face (Olho esquerdo, nariz e boca) tanto para os testes quanto para o funcionamento do sistema, foi feita uma localização manual de cada uma delas em todas as imagens utilizadas nos casos de teste. Essa localização foi realizada por Souza (2005).

A Figura 66 apresenta um exemplo do conjunto de fotos retiradas durante a primeira sessão e a Figura 67 apresenta um exemplo do conjunto de fotos retiradas durante a

segunda sessão, que ocorreu 14 dias após a primeira. Cada tipo de foto dos indivíduo foi nomeada e numerada de AR1 à AR26 para futuras referências.

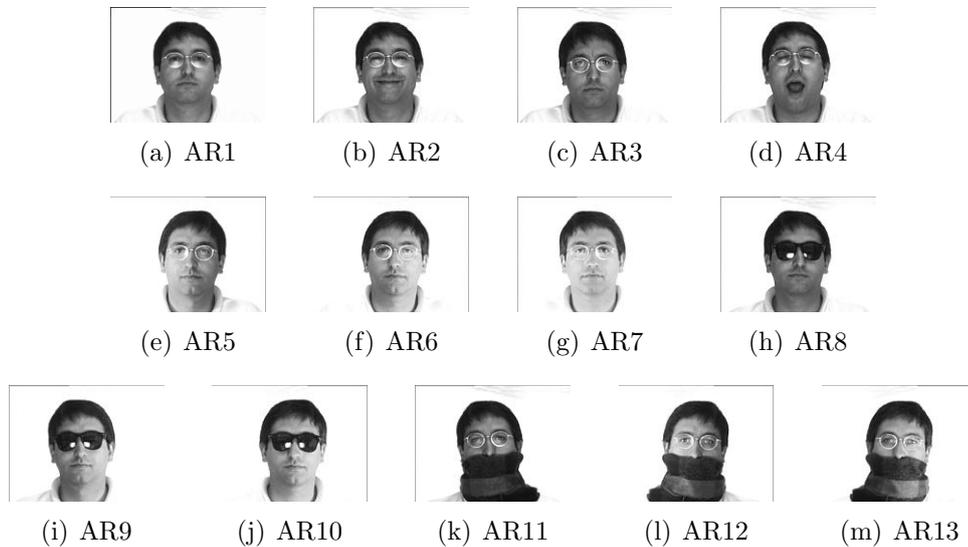


Figura 66: Exemplo das fotos relativas à primeira sessão. (a) Expressão neutra, (b) Sorrindo, (c) Zangado, (d) Gritando, (e) Luz à esquerda, (f) Luz à direita, (g) Luz por todos os lados, (h) Óculos de sol, (i) Óculos de sol e luz à esquerda, (j) Óculos de sol e luz à direita, (k) Cachecol, (l) Cachecol e luz à esquerda, (m) Cachecol e lua à direita.

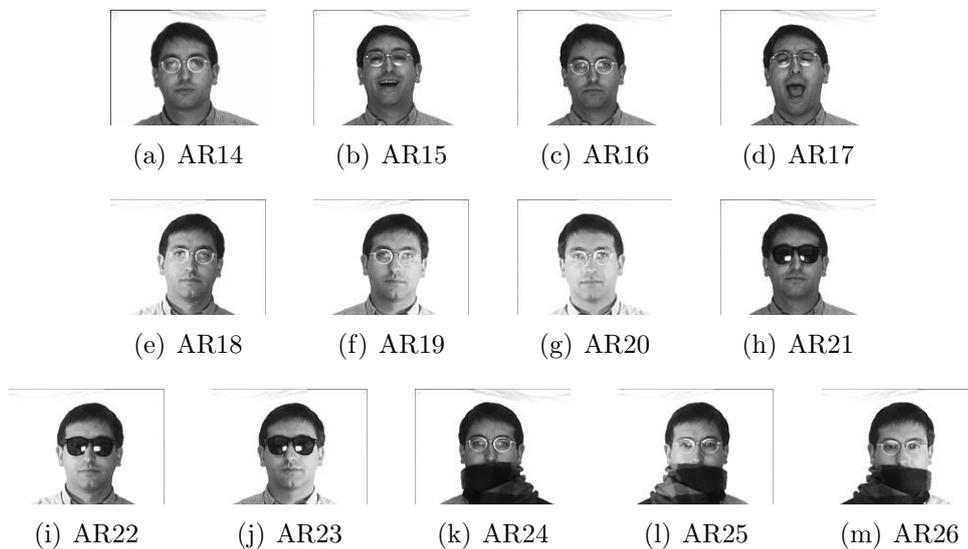


Figura 67: Exemplo das fotos relativas à segunda sessão. (a) Expressão neutra, (b) Sorrindo, (c) Zangado, (d) Gritando, (e) Luz à esquerda, (f) Luz à direita, (g) Luz por todos os lados, (h) Óculos de sol, (i) Óculos de sol e luz à esquerda, (j) Óculos de sol e luz à direita, (k) Cachecol, (l) Cachecol e luz à esquerda, (m) Cachecol e lua à direita.

A.1 Motivação para usar o AR Face Database

A diversidade apresentada nas fotos de cada indivíduo (expressões distintas, diferentes condições de luz, oclusões naturais e diferença temporal de 14 dias) permite a elaboração de testes capazes de dar uma idéia bem clara de como um sistema de reconhecimento facial responde a variadas situações.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)