

**Rastreamento eficiente de faces
em um subespaço wavelet**

Rogério Schmidt Feris

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE MESTRE
EM
CIÊNCIA DA COMPUTAÇÃO

Área de Concentração : Ciência da Computação

Orientador : Prof. Dr. Roberto Marcondes Cesar Junior

- São Paulo, Maio de 2001 -

Rastreamento eficiente de faces em um subespaço wavelet

Este exemplar corresponde à redação final
da dissertação devidamente corrigida e
apresentada por Rogério Schmidt Feris e
aprovada pela Comissão Julgadora.

São Paulo, 11 de maio de 2001

Banca Examinadora :

- Prof. Dr. Roberto Marcondes Cesar Junior (orientador) - IME-USP
- Prof. Dr. Carlos Hitoshi Morimoto - IME-USP
- Prof. Dr. Luciano da Fontoura Costa - IFSC-USP

Ao meu amigo de sempre, tio Antônio

Agradecimentos

Ao meu amigo e orientador Prof. Dr. Roberto Marcondes Cesar Junior, por tudo o que com ele aprendi, em ambiente propício à pesquisa de excelência, aliado ao permanente apoio afetivo/emocional.

Ao pesquisador Dr. Volker Krüger (*University of Maryland*), idealizador da técnica utilizada neste trabalho, pela sua amizade e por ter acompanhado com muito interesse minha pesquisa, através de contatos freqüentes via e-mail.

Ao Prof. Dr. Carlos Hitoshi Morimoto, em especial, por ter acompanhado meu trabalho, e pelas frutíferas conversações técnicas (e não técnicas) que tivemos.

Ao Teo, companheiro de trabalho, e amigo em todas as horas.

A todo pessoal do laboratório de processamento de imagens do IME-USP, pela acolhida, abertura, e disponibilidade em me ajudar.

Ao Sérgio, que colaborou com a formação do banco de seqüências de imagens utilizado neste trabalho, e com a realização de testes pertinentes ao módulo de rastreamento de faces.

À FAPESP (99/01487-1), pelo fundamental suporte financeiro dispensado à execução desta pesquisa.

À minha família, especialmente a meus pais e a minha irmã, pelo apoio constante, e por vibrarem sempre com as minhas conquistas pessoais e profissionais.

À minha namorada Evelyn, com especial carinho, pelo apoio, paciência e compreensão, neste período de árduo trabalho.

A Deus, sentido da minha vida, em quem encontro força e motivação para alcançar meus objetivos.

Resumo

O reconhecimento computacional de faces humanas a partir de seqüências de vídeo emerge nos dias atuais como um instigante e desafiador campo de pesquisa. Os resultados dos estudos realizados neste tema encontram diversas aplicações em sistemas de segurança, autenticação de usuários, interação homem-máquina e outras.

Em geral, três etapas são discriminadas em um sistema de reconhecimento de faces a partir de vídeo: (1) detecção e rastreamento da face do usuário; (2) segmentação e normalização da face em cada quadro da seqüência; e (3) extração de características e classificação. O objetivo do presente trabalho consiste no estudo e desenvolvimento de algoritmos de visão computacional relativos à primeira etapa, com ênfase no problema de rastreamento de faces.

A detecção da presença da face do usuário, em uma seqüência de imagens, foi realizada utilizando-se um procedimento baseado em um modelo estatístico de cor da pele. Em contrapartida, para efetuar o rastreamento da face, realizamos experimentos com uma técnica mais sofisticada, recentemente proposta na literatura, denominada *Gabor wavelet networks* (GWNs). O problema de localização e rastreamento de características faciais também foi tratado, visto que o módulo de reconhecimento, atualmente sendo implementado por outro aluno de mestrado [Campos, 2001], utiliza essas características para efetuar a classificação da face.

Como principal contribuição deste estudo, salientamos a proposição de um método eficiente de rastreamento de faces, realizado em um subespaço wavelet de baixa dimensionalidade. Mais especificamente, esse método constitui um aperfeiçoamento, em termos de eficiência, da técnica de rastreamento baseada em GWNs. A comparação de ambas abordagens, bem como a avaliação de performance da técnica introduzida, considerando aspectos como precisão, robustez e eficiência, serão discutidas ao longo deste trabalho.

Abstract

Computational human face recognition from video sequences emerges nowadays as an instigating and challenging research field. The results obtained from the study conducted over this topic encompass several applications in security systems, user authentication, human-machine interaction and so forth.

In general, three steps are distinguished in a face recognition system from video sequences: (1) user's face detection and tracking; (2) face segmentation and normalization in each frame; and (3) feature extraction and classification. The objective of the present work consists in studying and developing computer vision algorithms relative to the first step, with emphasis on the face tracking problem.

The user's face detection in an image sequence was performed by using a method based on a statistical skin-color model. On the other hand, to accomplish face tracking, we have carried out experiments with a more sophisticated technique, proposed recently in the literature, which is called Gabor wavelet networks (GWNs). The problem of locating and tracking facial features was also considered, since the recognition module, which is being developed by another master's student [Campos, 2001], uses these features to perform face classification.

As the main contribution of this thesis, we highlight the proposal of an efficient method for face tracking, carried out in a low-dimensional wavelet subspace. More specifically, this method poses a considerable enhancement over the tracking based on GWNs. The comparison between both approaches, as well as the performance assessment of the introduced technique, considering aspects such as precision, robustness and efficiency, are discussed along this work.

Sumário

1	Introdução	1
1.1	Objetivos	2
1.2	Contribuições	4
1.3	Organização da Dissertação	5
2	Revisão Bibliográfica	6
2.1	Detecção de Faces	6
2.1.1	Imagens Estáticas	7
2.1.2	Imagens Dinâmicas	9
2.2	Rastreamento de Faces	10
2.2.1	Rastreamento 2D	11
2.2.2	Rastreamento 3D	15
3	Introdução à Teoria de Wavelets	16
3.1	Transformada Contínua em Wavelets	18
3.2	Transformada Discreta em Wavelets	20
3.3	Wavelet Networks	21

4	Representação de Faces Utilizando GWNs	23
4.1	Representação Baseada em GWNs	24
4.2	Cálculo Direto dos Pesos	27
4.3	Reparametrização de uma GWN	31
4.4	Características da Representação	33
5	Detecção e Rastreamento de Faces	35
5.1	Detecção de Faces	35
5.1.1	Modelo Estatístico de Cor da Pele	35
5.1.2	Segmentação da Face	37
5.2	Rastreamento de Faces Utilizando GWNs	39
5.3	Localização e Rastreamento de Características Faciais	40
6	Rastreamento de Faces em um Subespaço Wavelet	44
6.1	Metodologia	46
6.2	Distância entre Pesos	47
6.3	Algoritmo	48
7	Resultados Experimentais	50
7.1	Avaliação da Técnica de Rastreamento de Faces	50
7.1.1	Análise de Precisão	50
7.1.2	Análise de Robustez	52
7.1.3	Análise de Eficiência	56
7.2	Aplicação: Reconhecimento de Faces a partir de Vídeo	58
7.2.1	Formação do Banco de Dados	58
7.2.2	Detecção da Face	59

7.2.3	Posicionamento das Características Faciais	59
7.2.4	Rastreamento da Face e das Características Faciais	60
7.2.5	Seleção e Normalização dos Melhores Quadros	61
8	Conclusões	62
A	Lista de Símbolos	64
B	Lista de Abreviaturas	67

Lista de Figuras

1.1	Extração e normalização das características faciais utilizadas para o reconhecimento do usuário.	4
2.1	Abordagens utilizadas no problema de detecção de faces.	7
2.2	Abordagens utilizadas no problema de rastreamento de faces.	11
3.1	(a) Exemplos de funções $g_{u,b}(t)$, com diferentes deslocamentos e frequências. (b) Exemplos de funções $\psi_{a,b}(t)$, com diferentes deslocamentos e escalas. . .	19
4.1	Função de Gabor 2D considerada (parte imaginária).	25
4.2	(a) Imagem de uma face. (b) Representação obtida pela GWN com 52 wavelets. (c) Posição das 16 wavelets de maior peso.	27
4.3	(a) Imagem de uma face. (b) Representações obtidas utilizando GWNs com 32, 52, 100 e 320 wavelets.	28
4.4	(a) Imagem de uma face. (b) Reconstrução com base na equação (4.6). (c) Reconstrução com pesos ótimos.	31
4.5	(a) Imagem da face em uma pose diferente. (b) Representação wavelet reparametrizada. (c) Posição das 16 wavelets de maior peso.	32
5.1	Amostra de cor da pele.	36
5.2	(a) Agrupamento de cores no espaço RGB. (b) Agrupamento de cores no espaço cromático.	36

5.3	(a) Imagem de entrada. (b) Imagem obtida a partir do modelo de cor da pele.	37
5.4	Exemplo de modelos de cor de pele e não-pele no caso unidimensional. Note que a distribuição de probabilidade de cor não-pele é uma constante, dada por um valor de limiar.	38
5.5	(a) Imagem resultante da limiarização. (b) Aplicação do operador de fechamento e filtro da mediana.	38
5.6	Face média e sua representação wavelet.	42
5.7	Localização de características faciais em diferentes indivíduos.	42
5.8	Localização de características faciais sob diferentes expressões faciais e mudanças de iluminação.	43
5.9	Rastreamento de características faciais utilizando GWNs.	43
6.1	Mapeamentos entre o espaço de imagens $L^2(\mathbb{R}^2)$ e o subespaço wavelet \mathbb{R}^N	45
7.1	Rastreamento de faces no subespaço wavelet, utilizando-se uma GWN com 116 wavelets.	51
7.2	Representações obtidas com 116, 51, 22 e 9 wavelets.	51
7.3	Estimação do parâmetro de posição c_x em cada quadro da seqüência.	53
7.4	Estimação do parâmetro de posição c_y em cada quadro da seqüência.	54
7.5	Estimação do parâmetro referente ao ângulo da face em cada quadro da seqüência.	55
7.6	Deslocamento da face em cada quadro da seqüência.	56
7.7	Comparação dos tempos exigidos por ciclo entre o rastreamento no espaço de imagens e o método proposto.	57
7.8	Detecção da face e localização das características faciais.	60
7.9	Rastreamento de olhos, nariz e boca. Note que o método é robusto a piscar de olhos, mudança homogênea de iluminação e diferentes expressões faciais.	61

Lista de Tabelas

2.1	Comparação de diferentes técnicas de rastreamento de faces.	12
2.2	Notação utilizada.	12

Capítulo 1

Introdução

Recentemente, com os avanços das técnicas de visão computacional, diversas pesquisas têm sido realizadas para implementar sistemas computacionais capazes de detectar a presença de pessoas em imagens, acompanhá-las, reconhecê-las e interpretar seus gestos. Exemplos concretos desse fato são evidenciados através da promoção de diversos congressos internacionais como o *IEEE Face and Gesture Recognition* e o *Perceptual User Interfaces*, além do considerável aumento do número de artigos publicados sobre esse assunto em anais de congressos e nas revistas especializadas em visão computacional.

Os resultados obtidos a partir dessas pesquisas são passíveis de aplicabilidade em diferentes áreas. Sistemas de reconhecimento de faces, por exemplo, podem ser utilizados na identificação de criminosos, busca de crianças desaparecidas, ou mesmo para permitir acesso a pessoas autorizadas a indústrias ou a bancos. Em geral, tais sistemas exigem, como pré-requisito, um módulo de detecção e rastreamento de faces, quando seqüências de vídeo são utilizadas. Mesmo quando somente imagens estáticas são tratadas, faz-se necessário um procedimento de localização da face para que o reconhecimento seja efetuado.

A detecção e o rastreamento de outras partes do corpo humano, além da face, é comumente requerida em sistemas de reconhecimento de gestos, os quais podem ser aplicados no ensino de movimentos, como dança ou atividades esportivas, e até mesmo para auxiliar pessoas portadoras de dificuldades especiais, traduzindo sinais para a linguagem falada. Nesse contexto, vale salientar, também, os sistemas de reconhecimento de expressões faci-

ais, que possibilitam às máquinas reagir apropriadamente ao estado emocional do usuário.

A utilização desses sistemas, de forma integrada, certamente proporcionará novas formas de interação homem-máquina, mais naturais e eficazes. Atualmente, a comunicação homem-computador é caracterizada pela utilização de dispositivos como teclado, mouse e por interfaces gráficas (*Graphical User Interfaces* - GUIs), baseadas em janelas, ícones e menus. Tendo em vista a atual possibilidade de aquisição de computadores pessoais eficientes, com entradas de áudio e vídeo a um custo relativamente baixo, surge o desafio de desenvolver novas interfaces mais naturais, em que as máquinas sejam capazes de “escutar” e “enxergar” o usuário, com o auxílio de técnicas de processamento da fala, visão computacional, aprendizado de máquina e outras.

Esse novo paradigma de interação homem-máquina é comumente referido como “interfaces de percepção do usuário” (*Perceptual User Interfaces* - PUIs)[Turk, 1998]. Nesse sentido, as interfaces baseadas em visão computacional cumprem um papel fundamental, concedendo às máquinas a capacidade de perceber visualmente o usuário, permitindo sua identificação e o reconhecimento de suas ações. Em geral, tais interfaces concentram-se em prover respostas para as seguintes questões[Cipolla and Pentland, 1998]: *Onde* está o usuário? *Quem* é o usuário? *O que* ele deseja, ou está fazendo, e *por quê*?

Deve-se mencionar que esse tema, além de constituir tópico atual de pesquisa em diferentes centros acadêmicos, tem atraído a atenção de vários líderes empresariais. Como exemplo, temos a criação do grupo de “Interfaces Baseadas em Visão” (<http://www.research.microsoft.com/vision/VisionBasedInterfaces>) no centro de pesquisa da Microsoft em Redmond, o projeto “BlueEyes” (<http://www.almaden.ibm.com/cs/blueeyes>) da IBM em Almaden e muitos outros.

1.1 Objetivos

O principal objetivo do presente trabalho consiste no estudo e desenvolvimento de algoritmos de visão computacional para a detecção e rastreamento de faces em seqüências de vídeo. Além disso, este trabalho está inserido em um projeto mais amplo, em desenvolvimento no departamento de Ciência da Computação do IME-USP, o qual visa a reconhecer pessoas a partir de seqüências de imagens.

Assim, assumimos que o usuário esteja localizado na frente de um computador, sendo filmado por uma câmera, em princípio estática, que adquire imagens coloridas. Supomos, também, um ambiente fechado, razoavelmente controlado, em que o usuário movimenta-se naturalmente diante da câmera.

Para efetuar-se o reconhecimento a partir da seqüência de imagens, três etapas são discriminadas: (1) detecção e rastreamento da face do usuário; (2) segmentação e normalização da face em cada quadro da seqüência; e (3) extração de características¹ e classificação. A primeira etapa consiste em, primeiramente, verificar se o usuário está presente diante da câmera, mediante a detecção da sua face. Uma vez detectada, a face é precisamente localizada e rastreada ao longo da seqüência de imagens. Logo após, em cada quadro da seqüência, a imagem da face é extraída e normalizada, caracterizando a segunda etapa. A normalização deve considerar aspectos geométricos (e.g. deformações rígidas da face) e radiométricos (e.g. variação nas condições de iluminação). Finalmente, a terceira etapa consiste no reconhecimento do usuário, através de um processo de extração de características e classificação da face.

Este trabalho concentra-se na primeira etapa, com maior ênfase no módulo de rastreamento de faces. O problema de localização e rastreamento de características faciais também será tratado, visto que módulo de reconhecimento, atualmente sendo implementado por outro aluno de mestrado do nosso grupo[Campos, 2001], utiliza essas características para efetuar a classificação da face.

Uma descrição detalhada do nosso projeto de reconhecimento de faces a partir de vídeo pode ser consultada no artigo[Feris et al., 2000b], incluído em anexo. Neste momento, é importante apenas salientar que o reconhecimento é realizado com base nas regiões dos olhos, nariz e boca do usuário. A figura 1.1 ilustra a extração e a normalização dessas características em um quadro de uma seqüência de imagens.

O tema focalizado neste trabalho, isto é, a tarefa de detectar a presença de uma face (e de suas características faciais) e acompanhá-la ao longo de uma seqüência de vídeo, constitui um tópico de pesquisa complexo e desafiador. Dificuldades aparecem com a

¹O termo “características” está sendo utilizado no sentido de “*features*” de reconhecimento estatístico de padrões, sendo mais amplo que o termo “características faciais”, o qual denota regiões como os olhos, nariz e boca do usuário.

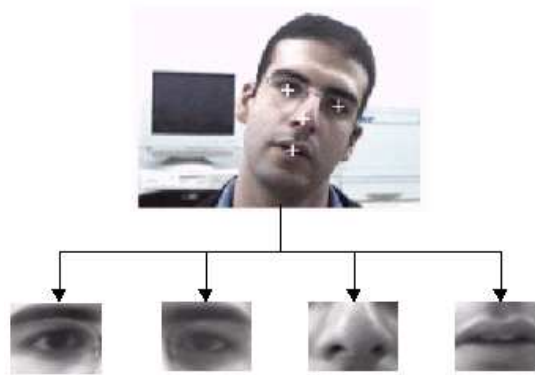


Figura 1.1: Extração e normalização das características faciais utilizadas para o reconhecimento do usuário.

presença de ruído nas imagens, mudanças nas condições de iluminação, e deformações rígidas e não rígidas da face, como variação de pose, escala e expressão facial.

A abordagem que escolhemos para tratar esses problemas baseia-se em uma técnica recentemente proposta por [Kruger and Sommer, 2000], denominada *Gabor wavelet networks* (GWNs). Tal técnica possibilita decompor a imagem de uma face em uma combinação linear de Gabor wavelets, em que os pesos e parâmetros das wavelets são determinados mediante um processo de otimização, visando preservar o máximo de informação da imagem. Esta representação apresenta propriedades de invariância com relação a distorções afins da face e variação homogênea das condições de iluminação.

1.2 Contribuições

As principais contribuições deste trabalho estão discriminadas abaixo:

- Amplo levantamento bibliográfico, com o estudo dos principais métodos que compõem o estado da arte no que se refere à detecção e rastreamento de faces em seqüências de imagens.
- Experimentos com detecção de faces, utilizando-se um modelo estatístico de cor de pele para localizar regiões de interesse e um procedimento de correlação para verificar a presença de uma face nessas regiões[Feris et al., 2000a].

- Aplicação da técnica baseada em *Gabor wavelet networks* para a localização e rastreamento de características faciais[Feris and Cesar, 2001].
- Proposição de um esquema para reconhecimento de faces a partir de seqüências de vídeo[Feris et al., 2000b], juntamente com o aluno de mestrado responsável pela implementação do módulo de reconhecimento.
- Proposição e avaliação de um método eficiente de rastreamento de faces, realizado em um subespaço wavelet de baixa dimensionalidade[Feris et al., 2001]. Essa é a principal contribuição original deste trabalho, e constitui um aperfeiçoamento da técnica de rastreamento de faces baseada em GWNs.

1.3 Organização da Dissertação

Esta dissertação está organizada da seguinte forma: inicialmente, no capítulo 2, será apresentada uma revisão bibliográfica sobre detecção e rastreamento de faces em seqüências de vídeo. O capítulo 3 apresenta uma breve introdução à teoria de wavelets, importante para o entendimento da representação de faces baseada em GWNs, a qual é detalhada no capítulo 4. O capítulo 5 trata dos problemas de detecção e rastreamento de faces, descrevendo, também, um procedimento para localização e rastreamento de características faciais. No capítulo 6, o tópico de rastreamento de faces é aprofundado, mediante a proposição de uma técnica eficiente baseada em um subespaço wavelet. A avaliação de performance da técnica introduzida, bem como os resultados obtidos no contexto da aplicação considerada, estão descritos no capítulo 7. Por fim, o capítulo 8 traça algumas conclusões sobre o tema, mencionando o desenvolvimento de trabalhos futuros. Em anexo, constam os principais artigos publicados neste período.

Capítulo 2

Revisão Bibliográfica

Diversos trabalhos já foram propostos para realizar a detecção e o rastreamento de faces em seqüências de imagens. Neste capítulo, iremos apresentar um panorama geral das principais técnicas utilizadas neste tópico de pesquisa, comentando suas vantagens e desvantagens.

2.1 Detecção de Faces

A detecção automática de faces humanas em imagens conduz a questões de extrema complexidade na área de visão computacional. De fato, dificuldades advém de fatores extrínsecos, como ruídos na imagem, variação das condições de iluminação, oclusão, bem como de fatores intrínsecos, como deformações e mudanças na aparência da face, devido, por exemplo, a diferentes expressões faciais. Assim, a maioria dos sistemas propostos na literatura impõem algumas restrições na aplicação considerada (e.g. processando imagens bem comportadas ou assumindo a face frontal na imagem).

A seguir, apresentaremos, primeiramente, os principais métodos de detecção de faces utilizados em imagens estáticas, para somente após discutir técnicas relacionadas com seqüências de vídeo. A figura 2.1 ilustra as diferentes abordagens que discutiremos ao longo desta seção.

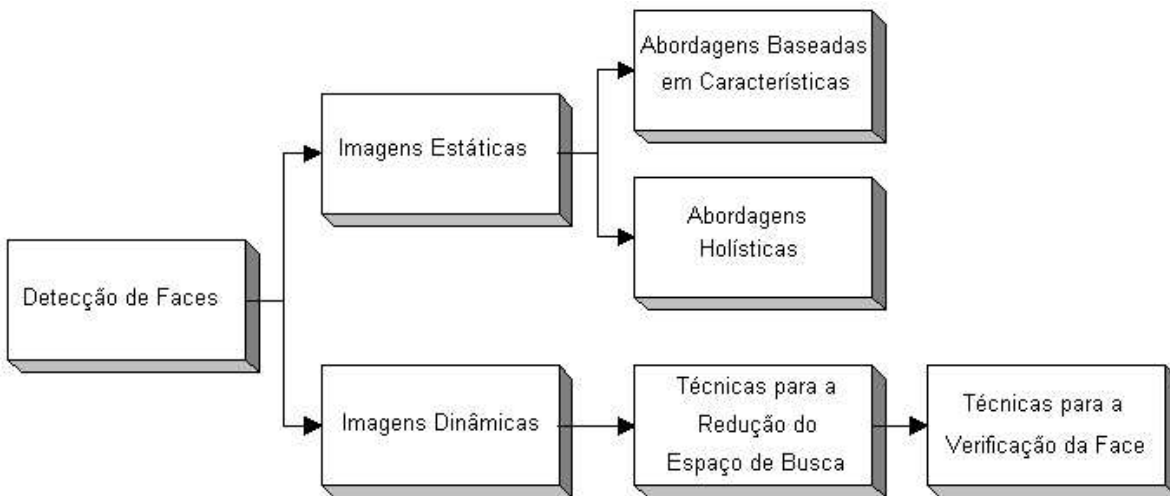


Figura 2.1: Abordagens utilizadas no problema de detecção de faces.

2.1.1 Imagens Estáticas

Basicamente, as técnicas propostas para a detecção de faces em imagens estáticas podem ser classificadas em abordagens baseadas em características e abordagens holísticas. As abordagens baseadas em características realizam o processo de detecção da face através da busca por características faciais (e.g. olhos, nariz e boca), levando em consideração, por exemplo, a relação geométrica entre essas características. Em contrapartida, as abordagens holísticas utilizam a informação da face como um todo para efetuar a detecção.

Abordagens Baseadas em Características

As primeiras técnicas propostas para a detecção de faces em imagens estáticas basearam-se na procura por características faciais mediante análise de formas[Sakai et al., 1969, Govindaraju et al., 1989]. Basicamente, essas técnicas processam o mapa de bordas da imagem considerada, sendo sensíveis a ruído.

A utilização de técnicas baseadas em casamento de padrões (*template matching*), em geral, não é adequada para detectar características faciais, visto que tais características não são rígidas. Com base nesse aspecto, o trabalho de [Yuille et al., 1989] propõe a utilização de padrões deformáveis para detectar elementos faciais. Essa abordagem, no

entanto, depende de uma correta inicialização de parâmetros e demanda elevado custo computacional.

Métodos similares utilizam modelos de contornos ativos [Blake and Isard, 1998], os quais constituem curvas que podem ser deformadas ou atraídas a um contorno da imagem, em resposta a um sistema de forças. Entretanto, esses métodos também dependem de uma boa inicialização para detectar corretamente uma característica facial.

Outras técnicas, como a utilização de modelos estatísticos, que exploram a relação geométrica das características faciais [Coelho et al., 1999], ou mesmo técnicas baseadas em correlação [Zelinsky and Heinzmann, 1996], já foram propostas na literatura. No entanto, no que se refere à detecção de faces em imagens estáticas, as abordagens holísticas são utilizadas, em geral, com maior eficácia.

Abordagens Holísticas

No processo de detecção de múltiplas faces em imagens estáticas, a maioria das abordagens holísticas utilizam-se de uma janela, a qual é deslocada na imagem, em várias escalas, extraindo regiões que são processadas por um método específico.

O trabalho de [Moghaddam and Pentland, 1997] utiliza PCA (*Principal Component Analysis*) e modelos probabilísticos para efetuar a detecção de faces nessas regiões. A referida técnica baseia-se na decomposição de uma imagem em uma combinação linear de *eigenfaces* [Turk and Pentland, 1991], que definem um espaço vetorial de dimensionalidade reduzida, apropriado para representação de faces. Os testes foram conduzidos apenas com imagens bem comportadas. Os principais problemas inerentes à representação baseada em PCA incluem sua sensibilidade em relação à variação de iluminação e pose da face, bem como a necessidade de alinhamento das imagens em um sistema de coordenadas comum.

Abordagens baseadas em redes neurais foram utilizadas com sucesso para detectar faces em imagens complexas. O trabalho de [Sung and Poggio, 1998] utiliza uma técnica de agrupamentos (*clustering*) para modelar a distribuição de faces, a qual é limitada por padrões não-face que costumam resultar em falsos alarmes. Com base em distâncias definidas em relação a este modelo, um perceptron multicamada é treinado para determinar a presença ou ausência de faces em regiões da imagem. Dentro desta abordagem, também

destaca-se o trabalho de [Rowley et al., 1998a], que utiliza um conjunto de redes neurais, juntamente com algumas heurísticas, para efetuar a detecção frontal de faces. Mediante o re-treinamento de padrões incorretamente classificados como faces, o sistema alcançou bons resultados em imagens complexas. Recentemente, uma extensão deste trabalho foi proposta para tratar a detecção de faces rotacionadas [Rowley et al., 1998b], utilizando-se uma rede neural específica para se determinar a orientação da face.

Outras abordagens, as quais consideram a detecção de faces em imagens estáticas coloridas, utilizam informação de cor [Wu et al., 1999] para selecionar regiões de interesse, ao invés de deslocar uma janela na imagem, como nas técnicas supracitadas. Tais abordagens são eficientes e apropriadas para o processamento de seqüências de imagens, sendo comentadas a seguir.

2.1.2 Imagens Dinâmicas

A despeito das técnicas supracitadas considerarem apenas imagens estáticas, requerendo elevado custo computacional, nos últimos anos, vários trabalhos concernentes à análise de seqüências de vídeo foram desenvolvidos. Neste caso, em geral, as imagens são de menor qualidade, mas os métodos podem dispor de mais informação, como movimento, por exemplo.

Basicamente, duas etapas podem ser discriminadas nos métodos de detecção de faces em seqüências de vídeo: redução do espaço de busca e verificação da face. A primeira etapa consiste na seleção eficiente de regiões candidatas na imagem. Na segunda etapa, a presença ou ausência de uma face é verificada nas regiões selecionadas, configurando o processo de detecção.

Embora abordagens diferentes já tenham sido propostas, as técnicas que compõem o estado da arte na literatura utilizam essas duas etapas para processar seqüências de vídeo com maior eficiência.

Redução do Espaço de Busca

As principais técnicas utilizadas para reduzir o espaço de busca por faces são baseadas em modelos de cor da pele humana [Yang et al., 1997, Kjeldsen and Kender, 1996]. Essas

abordagens são simples e demandam um esforço computacional mínimo, sendo adequadas para o processamento em tempo-real. Além disso, a utilização de modelos de cor possibilita a localização de faces de diferentes raças[Yang et al., 1997], sob diferentes orientações. A principal limitação da referida abordagem é evidenciada na presença de forte variação das condições de iluminação.

Deve-se ressaltar que, além de faces, outros objetos com cor da pele são selecionados através da utilização dessas técnicas. De fato, a segunda etapa faz-se necessária para verificar a presença de uma face nas regiões (*blobs*) de cor da pele selecionadas.

Outras abordagens para redução do espaço de busca por faces em imagens incluem subtração de *background*[Turk, 1996] e informação de movimento[McKenna et al., 1996]. Tais técnicas, no entanto, somente são adequadas para câmeras estáticas e cenas comportadas, com poucos objetos em movimento.

Vale salientar, ainda, a utilização da informação de profundidade[Feyrer and Zell, 1999] e de câmeras infra-vermelho (para localização das pupilas[Morimoto et al., 1998]), as quais também possibilitam a seleção de regiões candidatas.

Verificação da Face

Após a redução do espaço de busca, a segunda etapa é realizada com a verificação da presença ou ausência da face em cada região de interesse selecionada.

As principais técnicas utilizadas nesta etapa são baseadas em informações como textura [Dai and Nakano, 1996], piscar de olhos[Bala et al., 1997], simetria[Sun et al., 1998], forma [Feyrer and Zell, 1999], correlação com padrões[Darrel et al., 1998] e outras.

Em geral, quanto maior a integração de informações, maior é a robustez da detecção, porém, também maior é o custo computacional exigido.

2.2 Rastreamento de Faces

O rastreamento de faces consiste no acompanhamento da face detectada ao longo da seqüência de vídeo. Em outras palavras, a face deve ser constantemente localizada em cada

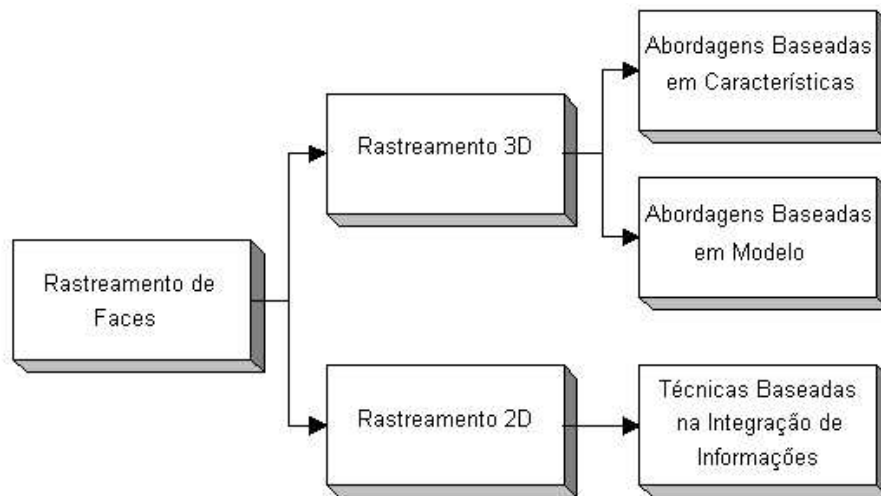


Figura 2.2: Abordagens utilizadas no problema de rastreamento de faces.

quadro da seqüência. Essa tarefa pode ser realizada no plano, configurando o rastreamento 2D, ou no espaço tridimensional, em que a posição e a orientação 3D da face são estimadas.

A figura 2.2 ilustra as abordagens que discutiremos ao longo dessa seção. Uma comparação dos trabalhos ora comentados é mostrada na tabela 2.1, a qual foi adaptada de [Toyama, 1998b]. As colunas dessa tabela denotam, da esquerda para direita, a referência ao trabalho, o objeto sendo rastreado (qualquer face, face particular, etc.), o método utilizado, o tempo de processamento e o *hardware* considerado. A notação utilizada está descrita na tabela 2.2.

2.2.1 Rastreamento 2D

A maior dificuldade inerente à pesquisa em rastreamento de faces consiste em efetuar esse processo com robustez e eficiência, permitindo o processamento em tempo-real. A seguir, apresentaremos as principais técnicas propostas para o rastreamento de faces 2D, discutindo aspectos como robustez e eficiência, com base na integração de informações.

Tabela 2.1: Comparação de diferentes técnicas de rastreamento de faces.

Rastreamento 2D de faces				
Trabalho	Objeto	Método	Tempo	Hardware
[Birchfield, 1998]	BCPE	C B	30Hz	Pm 200
[Crowley and Schwerdt, 1999]	QF	C F	15Hz	SGI
[McKenna et al., 1996]	QF	M T	5Hz	DC
[Darrel et al., 1998]	QF	C P T	12Hz	2 SGI
[Turk, 1996]	FO	B	33Hz	Pm 166
[Kruger and Sommer, 2000]	FP	T	25Hz	Pm450
[Hager and Belhumeur, 1998]	FP	T	30Hz	Pm 120
[Morimoto et al., 1998]	QF	F	30Hz	Pm II 333
[Maurer and Malsburg, 1996]	FP	F	NTR	Sun 20
Rastreamento 3D de faces				
[Cascia et al., 2000]	FP	T	5Hz	SGI O2
[Stiefelhagen and J.Yang, 1997]	QF	F	30Hz	HP-9000

Tabela 2.2: Notação utilizada.

Métodos de rastreamento de Faces 2D			
C	Cor	B	Borda (contorno)
M	Movimento	F	Característica (<i>Feature</i>)
P	Profundidade	T	Modelo Holístico (<i>Template</i>)
Objeto		Tempo/Hardware	
BCPE	<i>Blob</i> de cor da pele elíptico	NTR	Acima de 1Hz
QF	Qualquer face	Pm X	Pentium XMhz
FO	Contorno da face e dos ombros	SGI	Silicon Graphics
FP	Face particular	DC	Datacube MaxVideo 250
		Sun X	Sun Sparc X
		HP	Hewlett-Packard

Integração de Informações

Basicamente, os métodos propostos para o rastreamento de faces 2D utilizam uma ou mais das seguintes informações [Toyama, 1998b] (listadas, aproximadamente, em ordem crescente de tempo de processamento): *cor*, *movimento*, *profundidade*, *bordas*, *características da face* e *modelo holístico da face*.

As técnicas de rastreamento de faces baseadas em cor consistem, essencialmente, em acompanhar um *blob* de cor da pele ao longo da seqüência de imagens, com a possível utilização de filtros de Kalman para estimar a posição da face em cada quadro da seqüência [Olivier et al., 1997]. Entre outros, citamos os trabalhos de [Yang and Waibel, 1996] e [Raja et al., 1998] que apresentam um modelo estatístico de cor robusto a pessoas de diferentes raças e adaptável a diferentes condições de iluminação. Bons resultados também são apresentados no sistema proposto por [Birchfield, 1998], que utiliza restrições geométricas em adição à informação de cor para rastrear faces em tempo-real. O trabalho de [Crowley and Schwerdt, 1999] apresenta um sistema robusto de rastreamento de faces para compressão de vídeo, baseado em cor e correlação. Ainda neste contexto, é válido mencionar o trabalho de [Zarit et al., 1999] que mostra uma comparação entre dois métodos de segmentação de cor da pele em cinco diferentes espaços de cor.

Além de simplicidade e eficiência, as técnicas baseadas em cor permitem uma maior robustez com relação a deformações rígidas e não rígidas da face. No entanto, essas técnicas falham na presença de forte variação das condições de iluminação e podem ser afetadas por outros objetos com cor da pele na imagem.

O rastreamento de faces baseado na informação de movimento [McKenna et al., 1996] é comumente empregado na presença de ruídos na imagem e fundos (*background*) complexos. Porém, como já mencionado, essas técnicas, em geral, assumem a utilização de câmeras estáticas. Outra informação importante na segmentação de faces em imagens com fundos complexos é a profundidade, utilizada no trabalho de [Darrel et al., 1998].

O trabalho de [Turk, 1996] utiliza informação de borda para localizar e rastrear a face e os ombros do usuário em tempo-real. O mapa de bordas é computado a partir da imagem resultante da subtração de *background*, sendo que apenas um usuário deve estar presente diante da câmera. Outro trabalho baseado em bordas foi proposto por

[Silva et al., 1995] para a localização e o rastreamento de características faciais. Essa técnica falha em diversas situações, como presença de óculos e cabelos cobrindo a testa.

As abordagens baseadas em modelos holísticos da face utilizam um conjunto desses modelos, correspondendo às diferentes poses da face, ou então um único modelo deformável, que se ajusta à face em cada quadro da seqüência de imagens. Por exemplo, o recente trabalho de [Kruger and Sommer, 2000], o qual detalharemos ao longo desta dissertação, utiliza um modelo wavelet da face que pode ser deformado eficientemente em uma imagem.

A técnica proposta por [Hager and Belhumeur, 1998] considera modelos paramétricos de geometria e iluminação para efetuar o rastreamento de faces em tempo-real. Embora a face tenha sido modelada como um objeto rígido sob movimentos restritos, o método mostrou-se robusto a mudanças de expressão facial e pequenas rotações fora do plano. Neste contexto, um *framework* genérico é proposto para realizar o rastreamento eficiente de objetos que trata problemas como variação de pose, mudanças na iluminação e oclusão. Como limitação dessa técnica, salientamos os modelos de iluminação que são dependentes do usuário.

As abordagens baseadas em características efetuam o rastreamento de características (e.g. olhos, nariz, boca) ou pontos faciais específicos. Por exemplo, o sistema proposto por [Morimoto et al., 1998] realiza o rastreamento da face mediante a localização de pupilas em cada quadro da seqüência. Esse sistema, o qual utiliza fontes de luz infra-vermelho, é altamente confiável e adequado para aplicações em interfaces homem-máquina. Como limitação, citamos sua sensibilidade a movimentos bruscos da face do usuário.

O trabalho de [Maurer and Malsburg, 1996] descreve um sistema para rastrear pontos faciais utilizando a técnica *elastic graph matching*. Neste método, filtros de Gabor, com diferentes freqüências e orientações, são aplicados em algumas posições do interior da face, formando vetores de características, denominados *jets*. A face é então representada como um grafo, em que os nós correspondem aos *jets* e as arestas codificam informação geométrica. O rastreamento da face é realizado mediante um procedimento de casamento de grafos (*graph matching*) em cada quadro. A principal desvantagem dessa abordagem é o alto custo computacional requerido, o que conduz a uma taxa de processamento inadequada para aplicações em tempo-real.

2.2.2 Rastreamento 3D

Embora o rastreamento 3D de faces não pertença ao escopo deste trabalho, é válido mencionar, brevemente, algumas técnicas relativas a esse assunto. Basicamente, essas técnicas podem ser divididas em duas abordagens: baseadas em modelo e baseadas em características.

Abordagens Baseadas em Modelo

As abordagens baseadas em modelo [Cascia et al., 2000, Edwards et al., 1998] utilizam um modelo tridimensional da face para determinar a posição e a orientação 3D da face em cada quadro da seqüência. Por exemplo, o trabalho de [Cascia et al., 2000] utiliza um modelo cilíndrico no qual é projetada a textura da face. A determinação da posição e orientação do cilindro em cada quadro configura o processo de rastreamento. Em geral, as abordagens baseadas em modelo são robustas a diferentes deformações da face, mas demandam elevado custo computacional, não sendo, na maioria dos casos, passíveis de execução em tempo-real.

Abordagens Baseadas em Características

As abordagens baseadas em características determinam a orientação 3D da face a partir da posição de determinadas características faciais na imagem [Jebara and Pentland, 1997, Toyama, 1998a]. Por exemplo, o trabalho de [Stiefelhagen and J.Yang, 1997] determina a posição das pupilas, narinas e cantos da boca em cada quadro da seqüência, obtendo a orientação 3D da face a partir dessa informação. Geralmente, as abordagens baseadas em características são eficientes e permitem o processamento em tempo-real, mas são sensíveis a falhas quando os pontos característicos não são localizados (e.g. em virtude de oclusão ou variação de iluminação).

Capítulo 3

Introdução à Teoria de Wavelets

Apresentaremos, neste capítulo, uma breve introdução à teoria de wavelets, sob a ótica da análise de sinais. A formalização dessa teoria foi concretizada na década de 80[Grossmann and Morlet, 1984, Mallat, 1989] e no início da década de 90 [Daubechies, 1990], com base na generalização de conceitos já conhecidos, oriundos de diversos campos de pesquisa, como geofísica, análise e compressão de sinais, física e matemática. Desde então, a teoria de wavelets tem atraído a atenção de diversos pesquisadores, encontrando aplicações em diferentes áreas, particularmente em processamento de imagens e visão computacional[Costa and Cesar, 2001, Cesar and Costa, 1998]. Para uma revisão aprofundada sobre o tema, recomenda-se ao leitor o livro de Ingrid Daubechies [Daubechies, 1992], o qual apresenta uma análise matemática detalhada desta teoria.

O termo “wavelets” foi originariamente introduzido por J. Morlet, sendo a base matemática de suas idéias formalizada pelo físico teórico A. Grossmann [Grossmann and Morlet, 1984]. Os dados sísmicos estudados por Morlet exibiam conteúdos de frequência que mudavam rapidamente ao longo do tempo, para os quais a transformada de Fourier não era adequada como ferramenta de análise.

De fato, a transformada de Fourier não permite uma análise local do conteúdo de frequência do sinal. Eventos que venham a ocorrer em intervalos de tempos distintos, e mesmo bastante remotos, contribuem de maneira global para a transformada, afetando a representação como um todo. A equação (3.1) ilustra a transformada de Fourier de um sinal contínuo $f(t) : \mathbb{R} \rightarrow \mathbb{R}$. É importante notar que a transformada está baseada na

integração de todo o sinal para o cálculo de cada frequência.

$$F(u) = \int f(t)e^{-i2\pi ut} dt \quad (3.1)$$

A transformada janelada de Fourier, por outro lado, permite uma análise da frequência do sinal localmente no tempo. Neste caso, uma janela de observação é deslocada no domínio do tempo, e a transformada de Fourier da porção “visível” do sinal é calculada para cada posição da janela. Considerando $j(t)$ a função que cumpre o papel de janela, podemos definir a transformada janelada de Fourier de um sinal contínuo $f(t)$:

$$F(u, b) = \int j(t - b)f(t)e^{-i2\pi ut} dt \quad (3.2)$$

Quando a função janela $j(t)$ é definida como sendo uma função Gaussiana, a transformada janelada de Fourier é denominada transformada de Gabor. Sugere-se ao leitor a consulta do artigo [Gabor, 1946] para uma visão mais aprofundada dessa transformada.

Considerando-se $g_{u,b}(t) = j(t - b)e^{-i2\pi ut}$, a equação (3.2) pode ser escrita como:

$$F(u, b) = \int f(t)g_{u,b}(t)dt = \langle f, g_{u,b} \rangle \quad (3.3)$$

Assim, a transformada consiste no cálculo de produtos internos entre o sinal $f(t)$ e a função $g_{u,b}(t)$, a qual determina o deslocamento da janela de observação (parâmetro b) e a frequência da exponencial complexa (parâmetro u).

A transformada janelada de Fourier, porém, pode não ser adequada para a análise de determinados sinais, tendo em vista que o tamanho da janela de observação permanece constante para todas as frequências. Morlet introduziu uma nova transformada, denominada transformada em wavelets, na qual o tamanho da janela varia com a frequência, possibilitando que eventos de alta frequência possam ser localizados com uma maior resolução temporal, ao passo que componentes de baixa frequência possam ser analisados com maior resolução no domínio de Fourier. Intuitivamente, um evento de baixa frequência precisa de uma janela maior no domínio do tempo para ser observado.

3.1 Transformada Contínua em Wavelets

Definiremos, inicialmente, a transformada em wavelets considerando um sinal contínuo $f(t) : \mathbb{R} \rightarrow \mathbb{R}$. A generalização para sinais 2D $f(x, y)$ ou imagens, objeto de nosso estudo, pode ser obtida, tendo sido definida no trabalho de [Antoine et al., 1993]. A transformada contínua em wavelets de $f(t)$ é dada pela equação:

$$F(a, b) = \int f(t)\psi_{a,b}(t)dt \quad (3.4)$$

Nesta equação, os parâmetros $a > 0$ e b variam continuamente, sendo que as funções $\psi_{a,b}(t)$ são denominadas wavelets e definidas da seguinte forma:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \quad (3.5)$$

Claramente, tais funções são derivadas a partir de dilatações e translações de uma wavelet mãe ψ , a qual deve satisfazer o critério de admissibilidade:

$$C_\psi = 2\pi \int \frac{|\bar{\psi}(u)|^2}{|u|} du < \infty \quad (3.6)$$

onde $\bar{\psi}(u)$ é a transformada de Fourier de $\psi(t)$. Se $\bar{\psi}$ é uma função contínua, então C_ψ pode ser finito somente se $\bar{\psi}(0) = 0$, isto é, $\int \psi(x)dx = 0$.

Observando a equação (3.4), percebemos que a transformada em wavelets depende de dois parâmetros a e b , que correspondem, respectivamente, às informações de escala e tempo. É comum associar-se o inverso da variável de escala com a frequência. Assim, o sinal 1D $f(t)$ é desdobrado em uma transformada 2D $F(a, b)$, permitindo uma análise tempo-frequência do sinal, de forma semelhante à transformada janelada de Fourier. O semi-plano bidimensional (a, b) é comumente denominado *espaço de configurações* ou *plano tempo-escala* na literatura.

A equação (3.4) pode ser reescrita como o produto interno de $f(t)$ com uma função wavelet $\psi_{a,b}(t)$, isto é, $F(a, b) = \langle f, \psi_{a,b} \rangle$. Basicamente, este produto interno irá “medir” o conteúdo do espaço de configurações do sinal $f(t)$ em torno do ponto (a, b) nesse espaço.

É importante salientarmos a principal diferença entre a transformada em wavelets e a transformada janelada de Fourier, a qual reside na forma das funções $\psi_{a,b}(t)$ e $g_{u,b}(t)$, como ilustra a figura 3.1. Todas as funções $g_{u,b}(t)$ apresentam a mesma largura, a despeito

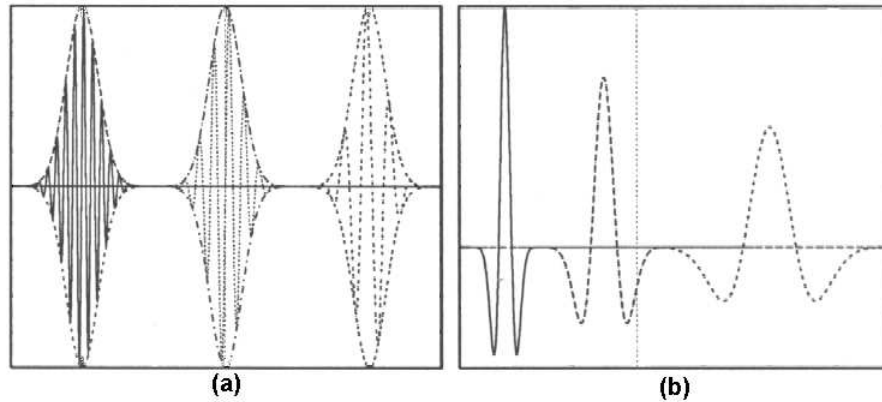


Figura 3.1: (a) Exemplos de funções $g_{u,b}(t)$, com diferentes deslocamentos e frequências. (b) Exemplos de funções $\psi_{a,b}(t)$, com diferentes deslocamentos e escalas.

do parâmetro de frequência u . Em contraste, a largura (no domínio do tempo) das funções $\psi_{a,b}(t)$ varia com a frequência. Mais especificamente, eventos de altas frequências conduzem a funções $\psi_{a,b}(t)$ mais finas (maior resolução temporal), enquanto que eventos de baixas frequências conduzem a funções $\psi_{a,b}(t)$ mais largas no domínio do tempo (maior resolução frequencial). Essa propriedade possibilita, por exemplo, que singularidades (eventos curtos de alta frequência) sejam analisadas com uma maior resolução temporal que outros eventos de mais longa duração.

A transformada em wavelets pode ser vista como a representação de um objeto, o qual pode ser uma função, imagem, sinal de som, etc. A equação (3.4), de fato, está decompondo $f(t)$ em uma superposição de wavelets $\psi_{a,b}(t)$, sendo que os coeficientes nessa superposição são exatamente obtidos pela transformada em wavelets de $f(t)$. Neste caso, é possível reconstruir $f(t)$ perfeitamente a partir de sua representação, como mostra a equação abaixo:

$$f(t) = \frac{1}{C_\psi} \int \int \langle f, \psi_{a,b} \rangle \psi_{a,b}(t) \frac{da db}{a^2} \quad (3.7)$$

A reconstrução, como pode ser notado, é obtida mediante uma integração sobre todo o espaço de configurações.

3.2 Transformada Discreta em Wavelets

Claramente, a representação baseada na transformada em wavelets contínua é altamente redundante. Entretanto, essa redundância pode ser reduzida ou eliminada mediante a restrição da escolha da wavelet ψ e pela discretização do espaço de configurações. Este processo origina a transformada discreta em wavelets, em que o sinal $f(t)$ é expressado como uma combinação linear de wavelets $\psi_{a,b}(t)$. Neste caso, os parâmetros a e b assumem valores discretos.

Vamos considerar, deste modo, uma discretização do espaço de configurações através do conjunto $S \subset \mathbb{R}^+ \times \mathbb{R}$, assumindo que $a \in \mathbb{R}^+$ e $b \in \mathbb{R}$ assumem valores discretos. Logo, podemos definir $B_\psi = \{\psi_{a,b} | (a,b) \in S\}$ como uma família discreta de wavelets.

É importante notar que o espaço de configurações não deve ser discretizado uniformemente, como na transformada de Fourier janelada discreta [Daubechies, 1990]. A discretização do parâmetro de translação depende da escala da wavelet. Assim, para altas frequências, as wavelets são transladadas por passos pequenos, enquanto que, para frequências baixas, a translação é feita com passos maiores.

Devido à discretização do espaço de configurações, a transformada em wavelets discreta inversa não está diretamente disponível [Daubechies, 1990]. Assim, surgem questões de como o sinal $f(t)$ pode ser reconstruído e se existem algoritmos numericamente estáveis para computar a transformada.

As respostas a essas questões dependem da escolha de ψ e da maneira como o espaço de configurações é discretizado. Basicamente, uma escolha adequada (que possibilita a reconstrução de $f(t)$) conduz a duas configurações distintas: sistemas discretos redundantes (frames) [Daubechies, 1990] e bases de wavelets, as quais podem ser ortogonais ou não ortogonais [Vetterli and Kovacevic, 1995].

No caso em que a família de wavelets B_ψ constitui uma base ortogonal, o sinal $f(t)$ pode ser expresso como:

$$f(t) = \sum_{(a,b) \in S} \langle \psi_{a,b}, f \rangle \psi_{a,b}(t) \quad (3.8)$$

Por exemplo, escolhendo-se como wavelet mãe a conhecida função de Haar, e mediante uma discretização apropriada do espaço de configurações, é possível gerar uma base

ortornormal de wavelets para o espaço $\mathbb{L}^2(\mathbb{R})^1$ [Daubechies, 1992].

No caso de bases não ortogonais ou sistemas discretos redundantes, o sinal $f(t)$ é reconstruído utilizando-se uma família de wavelets duais $\tilde{B}_\psi = \{\tilde{\psi}_{a,b} | (a,b) \in S\}$:

$$\begin{aligned} f(t) &= \sum_{(a,b) \in S} \langle \tilde{\psi}_{a,b}, f \rangle \psi_{a,b}(t) \\ &= \sum_{(a,b) \in S} \langle \psi_{a,b}, f \rangle \tilde{\psi}_{a,b}(t) \end{aligned} \quad (3.9)$$

Duas famílias de funções B_ψ e \tilde{B}_ψ são duais se para cada $\psi_i \in B_\psi$ e $\tilde{\psi}_j \in \tilde{B}_\psi$ é verificada a condição:

$$\langle \psi_i, \tilde{\psi}_j \rangle = \delta_{i,j} = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases} \quad (3.10)$$

A seção 4.2 apresentará esses conceitos com mais detalhes.

Finalmente, é válido mencionar o clássico artigo de [Mallat, 1989], que apresenta a transformada discreta em wavelets em um framework denominado “análise multiresolução”, propondo, concomitantemente, esquemas eficientes para o cálculo dessa transformada com base na teoria de filtragem.

3.3 Wavelet Networks

Em nosso trabalho, iremos utilizar a decomposição em wavelets de uma forma diferente das abordagens supracitadas. Nossa representação será baseada em uma técnica proposta por [Zhang and Benviste, 1992] denominada *wavelet network*. Nesta técnica, um sinal $f(t)$ é representado por um número fixo de wavelets, que correspondem aos nós de uma rede neural. Os coeficientes das wavelets são representados pelos pesos sinápticos desta rede. Mediante um procedimento de otimização, os pesos e parâmetros das wavelets são determinados de forma a melhor aproximar o sinal considerado. Sob a ótica da análise de sinais, $f(t)$ é aproximado por uma combinação linear de wavelets²:

$$\hat{f}(t) = \sum_{i=1}^M w_i \psi_{a_i, b_i}(t) \quad (3.11)$$

¹Uma função $f(t) : \mathbb{R} \rightarrow \mathbb{R} \in \mathbb{L}^2(\mathbb{R})$ se $\int |f(t)|^2 dt < \infty$

²Neste trabalho, não estamos interessados na reconstrução perfeita do sinal.

É importante salientar que os parâmetros a e b das wavelets são obtidos a partir do espaço contínuo de configurações, através do processo de otimização. A precisão da aproximação é dada pela escolha do número M de wavelets.

A transformada em wavelets pode ser facilmente estendida para o caso bidimensional [Antoine et al., 1993], com vistas à análise de imagens, por exemplo. Neste caso, porém, além dos parâmetros de dilatação e translação, um parâmetro adicional de rotação é utilizado na transformada. No próximo capítulo, iremos detalhar uma técnica baseada em *Gabor Wavelet Networks* para representação da imagem de uma face.

Capítulo 4

Representação de Faces Utilizando GWNs

A escolha da representação computacional de objetos em imagens cumpre um papel fundamental no desempenho da aplicação considerada, determinando aspectos como eficiência e robustez.

No que se refere à representação 2D de faces, duas abordagens são particularmente discriminadas [Brunelli and Poggio, 1993]: representações baseadas em características (*feature-based*) e representações baseadas em padrões (*template-based*). As representações baseadas em características [Cox et al., 1996, Yow and Cipolla, 1997] descrevem o objeto como uma coleção de suas características, as quais podem constituir medidas geométricas [Cox et al., 1996], ou até mesmo um conjunto de filtragens aplicadas localmente na imagem [Wiskott et al., 1997]. Em contraste, as abordagens baseadas em padrões [Rowley et al., 1998a, Turk and Pentland, 1991], caracterizam-se por representações holísticas, utilizando, por exemplo, todo o conjunto de *pixels* do objeto.

Neste capítulo, apresentaremos, com base no trabalho de [Kruger and Sommer, 2000], uma representação 2D de objetos baseada em *Gabor wavelet networks* (GWNs). Tal representação apresenta propriedades de ambas abordagens, baseada em características e baseada em padrões.

Basicamente, as GWNs possibilitam que a imagem discreta de uma face seja aproxi-

mada por uma combinação linear de Gabor wavelets, em que os pesos e parâmetros das wavelets são determinados mediante um processo de otimização, visando a preservar o máximo de informação da imagem. A representação obtida é caracterizada pela abordagem baseada em características, tendo em vista que a face é representada por uma coleção de Gabor wavelets específicas. De fato, somente os pesos e parâmetros otimizados de tais funções são armazenados na representação. Por outro lado, as GWNs permitem a reconstrução da face a partir do conjunto de Gabor wavelets, de modo que a representação possa ser tratada de forma holística, caracterizando a abordagem baseada em padrões. É interessante notar que o número de características utilizadas na GWN permite controlar a precisão da reconstrução da face.

Entre outras vantagens, a representação baseada em GWNs apresenta propriedades de invariância com relação a distorções afins da face e variação homogênea das condições de iluminação. Essas propriedades são evidenciadas no módulo de rastreamento de faces, o qual detalharemos adiante. A seguir, apresentaremos a técnica baseada em GWNs para representação de faces, elucidando suas principais características.

4.1 Representação Baseada em GWNs

Para definir uma *Gabor wavelet network*, consideramos primeiramente uma função wavelet mãe $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$, definida pela parte imaginária de uma função de Gabor 2D, caracterizada pela equação abaixo. A figura 4.1 ilustra graficamente essa função.

$$\psi(x, y) = \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \sin(x) \quad (4.1)$$

Seja $\Psi = \{\psi_{\mathbf{n}_1}, \dots, \psi_{\mathbf{n}_M}\}$ uma família de M Gabor wavelets, derivadas a partir de dilatações, translações e rotações de ψ , com a forma:

$$\begin{aligned} \psi_{\mathbf{n}_i}(x, y) &= \exp\left(-\frac{1}{2}[s_{x_i}((x - c_{x_i})\cos\theta_i - (y - c_{y_i})\sin\theta_i)]^2\right. \\ &\quad \left.+ [s_{y_i}((x - c_{x_i})\sin\theta_i + (y - c_{y_i})\cos\theta_i)]^2\right) \\ &\quad \times \sin(s_{x_i}((x - c_{x_i})\cos\theta_i - (y - c_{y_i})\sin\theta_i)) \end{aligned} \quad (4.2)$$

com $x, y \in \mathbb{R}$ e o vetor de parâmetros $\mathbf{n}_i = (c_{x_i}, c_{y_i}, \theta_i, s_{x_i}, s_{y_i}) \in \mathbb{R}^5$, onde c_{x_i}, c_{y_i} denotam a translação da Gabor wavelet, s_{x_i}, s_{y_i} denotam a dilatação e θ_i denota a orientação.

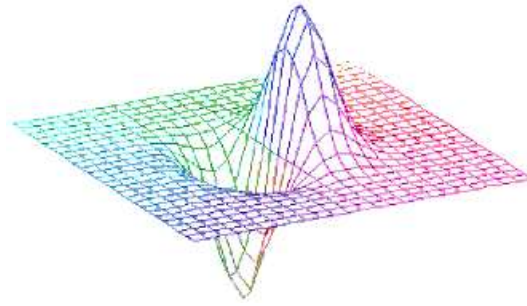


Figura 4.1: Função de Gabor 2D considerada (parte imaginária).

De acordo com [Zhang and Benviste, 1992], qualquer função $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, pertencente ao espaço das funções quadraticamente integráveis, isto é, $f \in \mathbb{L}^2(\mathbb{R}^2)$, pode ser representada por uma *wavelet network*.

Seja f a imagem contínua de uma face, a qual pode ser tomada como sendo uma função de $\mathbb{L}^2(\mathbb{R}^2)$. Vamos supor, sem perda de generalidade, que f possui média zero. A representação wavelet de f é obtida mediante a minimização da função energia abaixo:

$$E = \min_{\mathbf{n}_i, w_i \forall i} \left\| f - \left(\sum_{i=1}^M w_i \psi_{\mathbf{n}_i} \right) \right\|^2 \quad (4.3)$$

com relação aos pesos $w_i \in \mathbb{R}$ e aos parâmetros $\mathbf{n}_i \in \mathbb{R}^5$ das Gabor wavelets.

Assim, o conjunto de wavelets:

$$\begin{aligned} \Psi &= \{ \psi_{\mathbf{n}_1}, \dots, \psi_{\mathbf{n}_M} \} \text{ e o vetor} \\ \mathbf{w} &= (w_1, \dots, w_M) \end{aligned}$$

definem uma *Gabor wavelet network* (Ψ, \mathbf{w}) para uma função específica f .

Obtemos, então, a representação wavelet de f , a qual corresponde à aproximação \hat{f} da imagem original:

$$\hat{f} = \sum_{i=1}^M w_i \psi_{\mathbf{n}_i} \quad (4.4)$$

A equação (4.3) também pode ser aplicada considerando-se uma imagem discreta I^1

¹Nós utilizamos a notação f, g, \dots para referenciar funções contínuas, ao passo que a notação I, J, \dots refere-se explicitamente a imagens discretas em níveis de cinza.

(e.g. f amostrada). Assim, uma *Gabor wavelet network* é definida através de um vetor M -dimensional de pesos w_i e um vetor M -dimensional de Gabor wavelets $\psi_{\mathbf{n}_i}$, em que os pesos w_i e os vetores de parâmetros \mathbf{n}_i são escolhidos tal que a soma ponderada das Gabor wavelets $\psi_{\mathbf{n}_i}$ aproxima de forma ótima a imagem discreta I .

Deve-se mencionar que o trabalho de [Daugman, 1988] utiliza a equação (4.3) para determinar os pesos ótimos de um conjunto fixo de wavelets não ortogonais. Com a técnica baseada em GWNs, não somente os pesos, mas também os parâmetros ótimos de cada wavelet são determinados a partir do espaço contínuo de configurações.

Assim, uma única wavelet, com pesos e parâmetros otimizados, aproxima precisamente a estrutura local da imagem. Essa é a principal vantagem sobre o espaço discreto de configurações, em que uma combinação de wavelets deve ser utilizada para efetuar tal aproximação [Wiskott et al., 1997].

Dada uma imagem discreta I de uma face, a otimização da GWN, isto é, a minimização da função energia na equação (4.3), é realizada pelo método de otimização não linear Levenberg-Marquardt [Press et al., 1986]. Esse método possibilita determinar os parâmetros de um modelo contínuo² de modo a interpolar um conjunto de dados discretos. Em nosso caso, o modelo contínuo refere-se à combinação linear de Gabor wavelets. Os dados discretos correspondem ao conjunto de *pixels* da imagem I . Com o algoritmo de Levenberg-Marquardt, é possível determinar os pesos e parâmetros das wavelets (isto é, os parâmetros do modelo) de forma a interpolar os dados discretos da imagem da face. Mais especificamente, dado um conjunto de parâmetros iniciais do modelo, o algoritmo gradualmente (em vários ciclos) ajusta tais parâmetros até que o mínimo da função energia na equação (4.3) seja encontrado. É importante mencionar que esse método pode resultar na determinação de mínimos locais.

Como exemplo, consideremos a imagem da face ilustrada na figura 4.2(a). Utilizando uma GWN com somente 52 Gabor wavelets, inicializadas de forma homogênea na região interna da face, é possível gerar uma representação wavelet da face, a qual está ilustrada na figura 4.2(b). É importante mencionar que esta imagem reconstruída, também denominada *Gabor wavelet template* (GWT), é uma versão discreta do modelo wavelet contínuo da face. A figura 4.2(c) ilustra a posição das 16 wavelets de maior peso, após o

²O modelo depende dos parâmetros de forma não linear

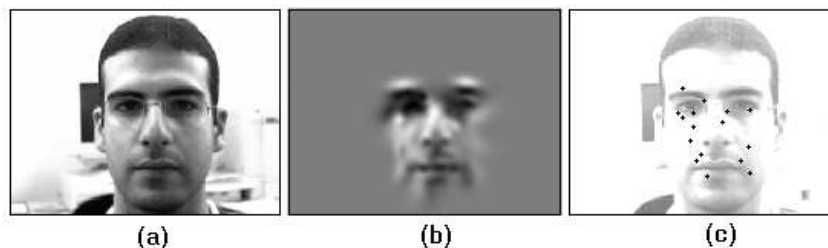


Figura 4.2: (a) Imagem de uma face. (b) Representação obtida pela GWN com 52 wavelets. (c) Posição das 16 wavelets de maior peso.

procedimento de otimização.

A precisão da representação da face é determinada pelo número M de wavelets da GWN. A escolha de M deve ser feita com base na aplicação desejada. Por exemplo, para problemas que requerem processamento em tempo real, um número pequeno de wavelets pode ser escolhido, gerando-se uma representação compacta e eficiente. No próximo capítulo, iremos mostrar que tal representação é adequada para o problema de rastreamento de faces, pois permite eficiência e adaptação a diferentes indivíduos. Por outro lado, se a aplicação exigir o processamento de informação mais precisa da face, então um valor mais alto para M deve ser escolhido.

Esta variabilidade da GWN na representação da informação da face é denominada atenção progressiva [Tsotsos, 1990]. A figura 4.3 ilustra este conceito, mostrando representações de uma face com diferentes níveis de precisão. É interessante notar que a GWN constitui uma *RBF network* (*Radial Basis Functions*) [Kosko, 1992], permitindo generalização dos dados de treinamento quando um pequeno número de funções base são utilizadas.

4.2 Cálculo Direto dos Pesos

Como já mencionado, no trabalho de [Daugman, 1988], a equação (4.3) é utilizada para o cálculo dos pesos de um conjunto de Gabor wavelets fixas. Visto que a otimização, em geral, caracteriza-se por um processo lento, iremos apresentar uma forma eficiente para o cálculo direto dos pesos de uma GWN. Esta seção servirá de base para a proposição de

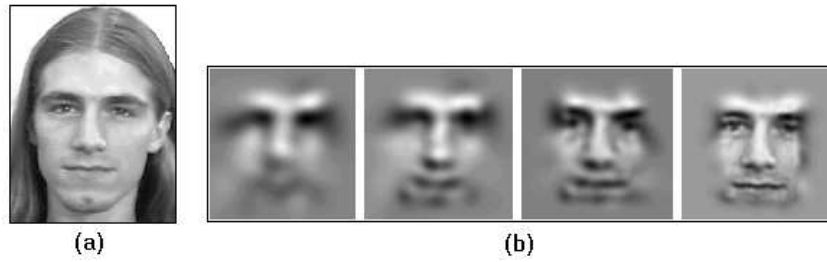


Figura 4.3: (a) Imagem de uma face. (b) Representações obtidas utilizando GWNs com 32, 52, 100 e 320 wavelets.

um novo método de rastreamento de faces, o qual será apresentado no capítulo 6, como principal contribuição do presente trabalho.

Inicialmente, é importante detalhar alguns conceitos apresentados na seção 3.2. Nesse sentido, vamos considerar uma família de wavelets $\Phi = \{\phi_1, \phi_2, \dots, \phi_N\}$ ³ que constituem uma base para $\mathbb{L}^2(\mathbb{R}^2)$, e uma função $f \in \mathbb{L}^2(\mathbb{R}^2)$, isto é, $f \in \langle \Phi \rangle$. Então existe um conjunto de pesos $\{w_i\}$ tal que:

$$f = \sum_{i=1}^N w_i \phi_i \quad (4.5)$$

No caso das wavelets $\{\phi_i\}$ formarem uma base ortogonal, os pesos $\{w_i\}$ são facilmente calculados pelo produto interno de f com cada wavelet ϕ_i em $\mathbb{L}^2(\mathbb{R}^2)$:

$$w_i = \langle f, \phi_i \rangle \quad (4.6)$$

Portanto, neste caso, a reconstrução de f é dada por:

$$f = \sum_{i=1}^N \langle f, \phi_i \rangle \phi_i \quad (4.7)$$

No entanto, quando a família de wavelets Φ não constitui uma base ortogonal, o cálculo dos pesos $\{w_i\}$ deve ser realizado de forma diferente. Uma maneira eficiente consiste em utilizar uma base dual de wavelets $\tilde{\Phi} = \{\tilde{\phi}_1, \tilde{\phi}_2, \dots, \tilde{\phi}_N\}$ [Vetterli and Kovacevic, 1995,

³A notação Φ será utilizada para designar wavelets genéricas, ao passo que a notação Ψ denotará a utilização de Gabor wavelets.

Daubechies, 1990], escolhida de forma a satisfazer a condição de biortogonalidade:

$$\langle \phi_i, \tilde{\phi}_k \rangle = \delta_{i,k} = \begin{cases} 1, & \text{se } i = k \\ 0, & \text{se } i \neq k \end{cases} \quad (4.8)$$

Claramente, quando a família de wavelets Φ constitui uma base ortogonal, então $\phi_i = \tilde{\phi}_i, \forall i$. No caso de uma base não ortogonal, os pesos $\{w_i\}$ são calculados eficientemente pelo produto interno de f com cada wavelet dual $\tilde{\phi}_i$ em $\mathbb{L}^2(\mathbb{R}^2)$:

$$w_i = \langle f, \tilde{\phi}_i \rangle \quad (4.9)$$

Assim, neste caso, a reconstrução de f é dada por:

$$f = \sum_{i=1}^N \langle f, \tilde{\phi}_i \rangle \phi_i \quad (4.10)$$

É interessante notar que a utilização de uma base dual também permite reconstruir f através da equação:

$$f = \sum_{i=1}^N \langle f, \phi_i \rangle \tilde{\phi}_i \quad (4.11)$$

Quando $f \notin \langle \Phi \rangle$, as equações acima podem ser utilizadas da mesma forma, mas a reconstrução de f não é perfeita. No entanto, o cálculo de pesos $\{w_i\}$, acima descrito, configura uma projeção ortogonal de f em $\langle \Phi \rangle$, caracterizando uma aproximação ótima da função no sentido do erro quadrado médio.

Utilizando os conceitos discutidos até o momento, podemos, agora, concentrar em nosso objetivo: determinar eficientemente os pesos $\{w_i\}$ para uma família de Gabor wavelets $\Psi = \{\psi_{\mathbf{n}_1}, \dots, \psi_{\mathbf{n}_M}\}$ de uma GWN.

Tendo em vista que as Gabor wavelets $\{\psi_{\mathbf{n}_i}\}$ não são ortogonais, os pesos $\{w_i\}$ serão calculados com a utilização de uma base dual $\tilde{\Psi} = \{\tilde{\psi}_{\mathbf{n}_1}, \dots, \tilde{\psi}_{\mathbf{n}_M}\}$, em que $\langle \psi_{\mathbf{n}_i}, \tilde{\psi}_{\mathbf{n}_k} \rangle = \delta_{i,k}$.

Considerando M o número de wavelets de uma GWN, vamos definir a matriz $\mathbf{A}_{M \times M}$, cujo termo geral é dado por $A_{ij} = \langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle$. Essa matriz é denominada matriz de

interferência das wavelets. As wavelets duais $\tilde{\psi}_{\mathbf{n}_i}$ podem ser então definidas, conforme a equação:

$$\tilde{\psi}_{\mathbf{n}_i} = \sum_{j=1}^M (A^{-1})_{ij} \psi_{\mathbf{n}_j} \quad (4.12)$$

De fato, tal definição satisfaz a condição de biortogonalidade $\langle \psi_{\mathbf{n}_i}, \tilde{\psi}_{\mathbf{n}_k} \rangle = \delta_{i,k}$, conforme mostrado a seguir:

$$\begin{aligned} \langle \psi_{\mathbf{n}_i}, \sum_{j=1}^M (A^{-1})_{kj} \psi_{\mathbf{n}_j} \rangle &= \int_{\mathbb{R}^2} \psi_{\mathbf{n}_i}(\mathbf{x}) \left[\sum_{j=1}^M (A^{-1})_{kj} \psi_{\mathbf{n}_j}(\mathbf{x}) \right] d\mathbf{x} \\ &= \sum_{j=1}^M (A^{-1})_{kj} \left[\int_{\mathbb{R}^2} \psi_{\mathbf{n}_i}(\mathbf{x}) \psi_{\mathbf{n}_j}(\mathbf{x}) d\mathbf{x} \right] \\ &= \sum_{j=1}^M (A^{-1})_{kj} \langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle \\ &= \sum_{j=1}^M (A^{-1})_{kj} A_{ji} \\ &= \delta_{i,k} \end{aligned} \quad (4.13)$$

Finalmente, podemos computar os pesos ótimos da GWN para aproximar uma função $f \in \mathbb{L}^2(\mathbb{R}^2)$. O cálculo dos pesos $\{w_i\}$ e a aproximação \hat{f} são dados pelas equações:

$$w_i = \langle f, \tilde{\psi}_{\mathbf{n}_i} \rangle \quad (4.14)$$

$$\hat{f} = \sum_{i=1}^M w_i \psi_{\mathbf{n}_i} \quad (4.15)$$

A figura 4.4(a) ilustra a imagem de uma face e a figura 4.4(c) mostra sua reconstrução com base nos pesos ótimos, considerando uma GWN com 320 wavelets. Se os pesos fossem calculados com a equação (4.6) teríamos o resultado da reconstrução ilustrado na figura 4.4(b).

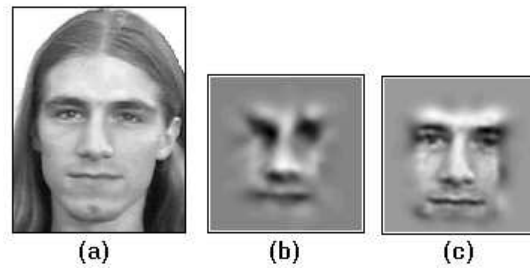


Figura 4.4: (a) Imagem de uma face. (b) Reconstrução com base na equação (4.6). (c) Reconstrução com pesos ótimos.

4.3 Reparametrização de uma GWN

Mostramos, anteriormente, como uma imagem discreta de uma face pode ser representada por uma combinação linear de Gabor wavelets, com parâmetros otimizados. Agora, iremos mostrar que essa representação pode ser distorcida de forma a *casar* com uma nova imagem da face, de modo que as wavelets estejam posicionadas nas mesmas características faciais da imagem original. Esse procedimento, denominado *reparametrização de uma GWN*, é realizado através da aplicação de uma transformação afim adequada na representação wavelet da face. O rastreamento de faces, a ser descrito no próximo capítulo, está baseado nesse conceito.

Como exemplo, considere a imagem da face ilustrada na figura 4.2(a). Seja G a GWN otimizada no interior da imagem dessa face. Considere, agora, a imagem da mesma face em uma pose diferente, como ilustra a figura 4.5(a). No processo de reparametrização, o conjunto de wavelets de G é registrado corretamente nas mesmas características faciais na imagem distorcida. É importante salientar que a reparametrização de uma GWN consiste na determinação dos parâmetros (translação, rotação, escala e cisalhamento) de qualquer deformação afim aplicada à imagem original. A figura 4.5(b) ilustra a representação discreta (GWT) reparametrizada, a qual foi obtida com 52 Gabor wavelets. A figura 4.5(c) mostra a posição das 16 wavelets de G com maior peso na imagem.

A reparametrização de uma GWN em uma nova imagem, isto é, a determinação dos parâmetros afins corretos, é feita com a utilização de uma superwavelet[Szu et al., 1992]. Seja $\Psi = (\psi_{\mathbf{n}_1}, \dots, \psi_{\mathbf{n}_M})$, $\mathbf{w} = (w_1, \dots, w_M)$ uma GWN otimizada em uma face. Uma



Figura 4.5: (a) Imagem da face em uma pose diferente. (b) Representação wavelet reparametrizada. (c) Posição das 16 wavelets de maior peso.

Gabor superwavelet (GSW) $\Psi_{\mathbf{n}} : \mathbb{R}^2 \rightarrow \mathbb{R}$ é definida como uma combinação linear de wavelets $\psi_{\mathbf{n}_i}$ tal que:

$$\Psi_{\mathbf{n}}(\mathbf{x}) = \sum_{i=1}^M w_i \psi_{\mathbf{n}_i}(\mathbf{S}\mathbf{R}(\mathbf{x} - \mathbf{c})) \quad (4.16)$$

em que os parâmetros do vetor \mathbf{n} da GSW Ψ definem a matriz de dilatação \mathbf{S} , a matriz de rotação \mathbf{R} e o vetor de translação \mathbf{c} sendo que:

$$\mathbf{S} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix}, \mathbf{R} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \mathbf{c} = (c_x, c_y)^T$$

Assim, a *Gabor superwavelet* $\Psi_{\mathbf{n}}$ é também uma wavelet que possui os parâmetros típicos de dilatação s_x, s_y , translação c_x, c_y e rotação θ . Dada uma nova imagem J , é possível deformar arbitrariamente a superwavelet através da otimização de seu vetor de parâmetros, de forma que a representação wavelet *case* com a face na imagem J . Este procedimento é realizado pela minimização da função energia abaixo, utilizando-se o algoritmo de Levenberg-Marquardt:

$$E = \min_{\mathbf{n}} \|J - \Psi_{\mathbf{n}}\|^2 \quad (4.17)$$

É importante notar que os parâmetros de uma wavelet incluem somente translação, dilatação e rotação. No entanto, o parâmetro de cisalhamento pode também ser incluído, permitindo qualquer deformação afim da GSW $\Psi_{\mathbf{n}}$. Assim, o parâmetro s_{xy} é adicionado ao vetor \mathbf{n} e a matriz de escala é modificada para:

$$\mathbf{S} = \begin{pmatrix} s_x & s_{xy} \\ 0 & s_y \end{pmatrix}$$

Considerando-se uma face com ≈ 60 pixels de largura, a inicialização dos parâmetros da superwavelet pode diferenciar dos parâmetros corretos por aproximadamente ± 10 px. nas direções x e y , 20% em escala e $\pm 10^\circ$ em rotação [Kruger and Sommer, 2000].

A reparametrização de uma GWN pode ser realizada em diferentes indivíduos, tendo em vista sua propriedade de generalização. De acordo com nossos experimentos, a reparametrização de uma GWN em uma imagem do mesmo indivíduo em que a mesma foi otimizada é sempre possível, mesmo considerando diferentes expressões faciais. No entanto, não podemos garantir que uma GWN otimizada em uma face específica possa ser reparametrizada em qualquer outra pessoa. Este problema será tratado na próxima seção, no item referente à localização de características faciais.

4.4 Características da Representação

A representação de faces baseada em GWNs apresenta diversas peculiaridades. A escolha da função de Gabor como wavelet mãe justifica-se pelo fato dos filtros de Gabor constituírem bons detectores de características em imagens [Manjunath and Chellapa, 1993], além de possibilitarem o melhor compromisso entre resolução de frequência e espacial (princípio de Heisenberg). Outro aspecto importante a ser ressaltado é que a utilização de filtros de Gabor em análise de imagens é biologicamente motivada, pois experimentos mostram que esses filtros modelam campos receptivos de células presentes no córtex visual humano [Daugman, 1985].

As principais características da representação de faces baseada em GWNs estão discriminadas abaixo:

- A representação apresenta propriedades de invariância com relação a deformações afins da face e variação homogênea das condições de iluminação. Essa segunda propriedade decorre do fato das Gabor wavelets possuírem média zero.
- O grau de precisão da representação depende do número de filtros da GWN, permitindo generalização dos dados de treinamento ou codificação de informação mais precisa da face.

- Os parâmetros das wavelets são determinados mediante um processo de otimização, de modo a representar o máximo de informação da imagem, considerando um determinado número de filtros.
- Os pesos ótimos de uma GWN estão diretamente relacionados com as respostas dos filtros de Gabor, refletindo a estrutura local da imagem.
- Os filtros de Gabor são bons detectores de características em imagens. Durante o processo de otimização, alguns filtros são atraídos por características locais da imagem, codificando informação geométrica. Os parâmetros de tais filtros tendem a assumir a mesma posição, orientação e escala das bordas da imagem.
- A representação baseada em GWNs é extremamente compacta. Por exemplo, o espaço ocupado pela representação ilustrada na figura 4.2 consiste em apenas 1040 bytes.

Outras vantagens e limitações desta representação serão apresentadas ao longo deste trabalho, particularmente nos experimentos realizados com detecção e rastreamento de faces e de características faciais.

As *Gabor wavelet networks* permitem representar qualquer objeto em uma imagem, embora o enfoque deste estudo seja restringido à representação de faces. No trabalho de [Kruger and Sommer, 2000], esta técnica foi eficazmente utilizada para rastreamento de faces, reconhecimento independente de expressão facial e estimação de pose. No próximo capítulo, iremos tratar o problema de detecção e rastreamento de faces e de características faciais em seqüências de imagens.

Capítulo 5

Detecção e Rastreamento de Faces

Conforme já mencionado, nosso objetivo consiste em, primeiramente, detectar a presença de uma face em uma seqüência de vídeo específica. Mostraremos, neste capítulo, que um modelo estatístico de cor de pele pode ser utilizado com eficácia para realizar esta tarefa. Após, iremos descrever o módulo de rastreamento de faces, bem como um procedimento para localizar e rastrear características faciais do usuário, utilizando GWNs.

5.1 Detecção de Faces

Tendo em vista os bons resultados obtidos pelos sistemas de detecção de faces baseados em cor, optamos pela escolha de tal abordagem. Basicamente, nosso método pode ser dividido em duas etapas: geração de um modelo estatístico de cor da pele e segmentação da face com base no modelo gerado.

5.1.1 Modelo Estatístico de Cor da Pele

A primeira etapa do método consiste em um treinamento supervisionado a partir de amostras de cor da pele humana. Um banco de 40 imagens de faces coloridas foi utilizado para realizar o treinamento. Tais imagens foram obtidas de pessoas de diferentes raças, idades e sexo, com variação das condições de iluminação. A figura 5.1 ilustra o processo de treinamento, em que uma região de cor da pele é selecionada, e sua representação



Figura 5.1: Amostra de cor da pele.

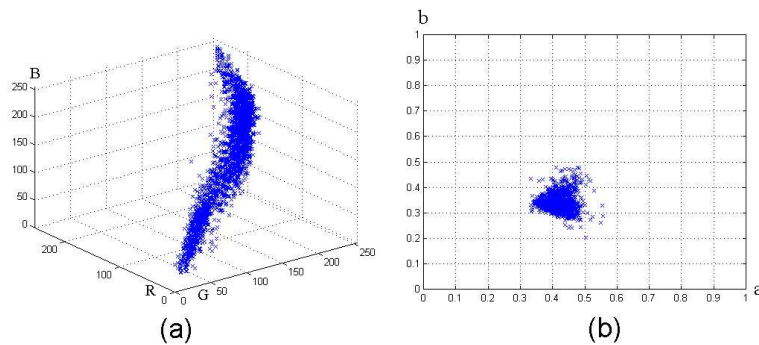


Figura 5.2: (a) Agrupamento de cores no espaço RGB. (b) Agrupamento de cores no espaço cromático.

RGB armazenada. A partir dos dados de treinamento, verificamos que as cores da pele humana formam um agrupamento (*cluster*) no espaço de cores, conforme ilustrado na figura 5.2(a).

É comum pensarmos que estas cores diferem significativamente de pessoa para pessoa, tendo em vista a diversidade de raças existentes. No entanto, o que realmente ocorre é uma maior diferença de brilho/intensidade e não de cor [Yang et al., 1997]. Assim, é possível reduzir a variância do agrupamento de cores da pele através da normalização de intensidade: $a = R/(R + G + B)$, $b = G/(R + G + B)$. As cores (a, b) são conhecidas como cores cromáticas, ou cores “puras”. A figura 5.2(b) ilustra o agrupamento de cores no espaço cromático.

De acordo com [Yang et al., 1997], a distribuição de cores da pele no espaço cromático pode ser eficazmente aproximada por um modelo Gaussiano. Assim, a partir dos dados

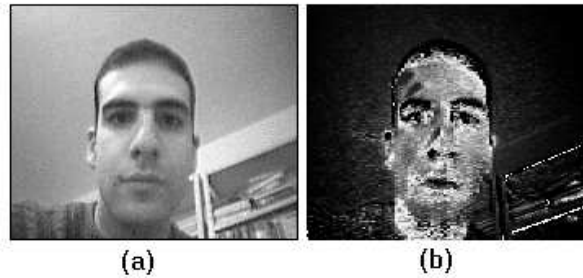


Figura 5.3: (a) Imagem de entrada. (b) Imagem obtida a partir do modelo de cor da pele.

de treinamento, podemos determinar os parâmetros -média e matriz de covariância - de uma Gaussiana bivariada, configurando o modelo de cor da pele.

5.1.2 Segmentação da Face

O procedimento de segmentação consiste em identificar as regiões (*blobs*) de cor da pele em uma imagem de entrada, com base no modelo gerado. Inicialmente, todos os pixels da imagem são convertidos para o espaço cromático. Mediante comparação com o modelo Gaussiano, uma nova imagem, em níveis de cinza, é gerada, sendo que a intensidade de cada pixel nesta nova imagem representa sua probabilidade de possuir cor da pele. A figura 5.3 ilustra este processo.

Para identificar os *blobs* de cor da pele, a imagem em níveis de cinza é convertida em uma imagem binária, com o ajuste de um valor de limiar (*threshold*) determinado experimentalmente.

Deve-se mencionar que essa técnica constitui um caso particular da teoria de decisão Bayesiana. De acordo com a regra de Bayes, um *pixel* é classificado com cor de pele se:

$$\begin{aligned}
 P(C_1)f(h|C_1) &\geq P(C_2)f(h|C_2) \\
 \text{ou} \\
 f(h|C_1) &\geq \frac{P(C_2)f(h|C_2)}{P(C_1)}
 \end{aligned} \tag{5.1}$$

onde C_1 é a classe “pele”, C_2 é a classe “não pele”, $P(C_1)$ e $P(C_2)$ são as probabilidades *a priori* dessas classes e $f(h|C_1)$ e $f(h|C_2)$ são as funções de densidade de probabilidade

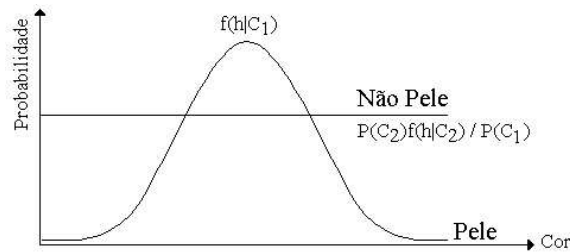


Figura 5.4: Exemplo de modelos de cor de pele e não-pele no caso unidimensional. Note que a distribuição de probabilidade de cor não-pele é uma constante, dada por um valor de limiar.

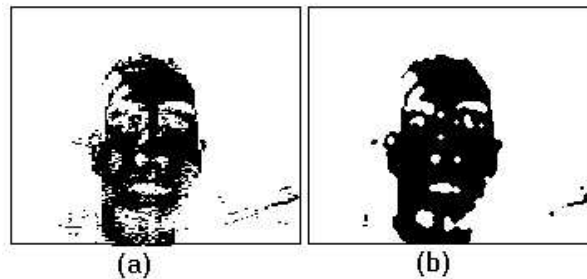


Figura 5.5: (a) Imagem resultante da limiarização. (b) Aplicação do operador de fechamento e filtro da mediana.

condicional da variável aleatória h , dado que o *pixel* pertence à classe C_1 e C_2 , respectivamente.

Em nosso caso, na equação (5.1), $f(h|C_1)$ corresponde ao modelo Gaussiano de cor da pele, enquanto que $P(C_2)f(h|C_2)/P(C_1)$ corresponde ao valor de limiar determinado experimentalmente. A figura 5.4 ilustra esse conceito.

Em geral, devido a ruídos e distorções na imagem de entrada, o resultado da limiarização pode gerar regiões de cor da pele particionadas (componentes não-conexas), além de permitir a existência de *pixels* de cor da pele isolados na imagem binária. Para solucionar esses problemas, utilizamos uma operação morfológica de fechamento seguida da aplicação do filtro da mediana. A figura 5.5 ilustra este processo.

A identificação dos limites dos *blobs* de cor da pele é então realizada através do algoritmo de crescimento de regiões. Regiões menores que 1% do tamanho da imagem são

descartadas. Além disso, são levados em consideração aspectos estruturais da face, eliminando regiões que não apresentam a estrutura de uma face. Isto é feito de uma forma simples, considerando a relação entre a altura e a largura do retângulo que define a região de cor da pele.

É importante salientar que as regiões determinadas podem corresponder a objetos com cor da pele diferentes de uma face. Logo, a verificação da presença ou ausência de uma face em tais regiões faz-se necessária. Isto é feito através de um procedimento simples de correlação que visa a localizar olhos nas regiões de cor da pele [Feris et al., 2000a]. Caso existam várias faces na imagem, aquela que estiver mais próxima da câmera será segmentada.

Outros experimentos com relação à verificação da presença ou ausência de uma face em regiões de cor da pele foram realizados utilizando descritores de Fourier (vide artigo em anexo) [Campos et al., 2000b]. No entanto, para obtermos um sistema realmente robusto, a integração de várias outras informações (movimento, simetria, etc.) deve ser realizada.

5.2 Rastreamento de Faces Utilizando GWNs

Após o procedimento de detecção, a face segmentada é acompanhada nas imagens subsequentes, ao longo da seqüência de vídeo. Este processo de rastreamento é realizado com a técnica baseada em GWNs, sendo robusto a transformações afins da face e mudanças homogêneas da iluminação.

Inicialmente, a representação wavelet é obtida para o *blob* de cor da pele determinado. A reparametrização da GWN pode ser então aplicada à seqüência de imagens, caracterizando o rastreamento da face. Assim, para cada quadro J_t referente ao instante t , os parâmetros \mathbf{n} da superwavelet $\Psi_{\mathbf{n}_t}$ são otimizados de acordo com a função energia abaixo:

$$E = \min_{\mathbf{n}_t} \|J_t - \Psi_{\mathbf{n}_t}\|^2 \quad (5.2)$$

O vetor de parâmetros \mathbf{n}_{t-1} é utilizado como valor inicial para a otimização no quadro J_t . Como as mudanças na imagem são pequenas de quadro para quadro, o processo de otimização converge rapidamente. Os valores iniciais para \mathbf{n}_0 no primeiro quadro são derivados a partir do *blob* de cor da pele.

O modelo adotado considera a face como um objeto planar, observado sob projeção ortográfica. Em consonância com o trabalho de [Hager and Belhumeur, 1998], verificamos que embora a face tenha sido modelada como um objeto rígido, com movimentos limitados, o algoritmo é capaz de rastrear o usuário ainda que este mude sua expressão ou realize rotações leves fora do plano.

5.3 Localização e Rastreamento de Características Faciais

A localização e rastreamento de características faciais constitui um tópico de pesquisa clássico em visão computacional. Nosso interesse neste problema reside no fato de que o reconhecimento de uma face pode ser realizado eficientemente utilizando apenas suas características faciais, como por exemplo, olhos, nariz e boca. Existem evidências biológicas [Moghaddam and Pentland, 1994] de que os olhos cumprem o papel mais importante no reconhecimento de uma face. O trabalho de [Campos et al., 2000a] demonstra a boa performance de um sistema de reconhecimento baseado apenas na região dos olhos, ao ser comparado com outro sistema que considera a face inteira. Além disso, trabalhos como [Moghaddam and Pentland, 1994, Brunelli and Poggio, 1993] mostram que o reconhecimento baseado em características faciais pode alcançar melhores resultados, tendo em vista que esta abordagem torna-se invariante a cortes de cabelo, penteados, maquiagem, etc.

Neste contexto, iremos apresentar um método baseado na reparametrização de uma GWN para localização e rastreamento de características faciais. Nossos experimentos referentes ao procedimento de localização foram conduzidos no banco de faces Yale, o qual consiste de 15 diferentes indivíduos, com 11 imagens por pessoa, apresentando diferentes expressões faciais (<http://cvc.yale.edu/projects/yalefaces/yalefaces.html>). Como este banco de dados contém apenas imagens em níveis de cinza, a região da face foi segmentada manualmente. Experimentos com imagens coloridas e detecção automática da face serão apresentados no próximo capítulo.

No capítulo anterior, foi mencionado que uma GWN de pouca precisão, isto é, com

um pequeno número de funções base, pode ser reparametrizada em diferentes indivíduos. No entanto, esta reparametrização não pode ser garantida em qualquer pessoa, tendo em vista que a mesma depende da similaridade entre a face de teste e a face em que a GWN foi otimizada, bem como do número de wavelets consideradas.

A solução que propomos para este problema consiste em otimizar a GWN em uma face média. Com este objetivo, foi selecionado um conjunto de 15 faces do banco de faces Yale, cada qual correspondendo a um indivíduo específico. As faces deste conjunto foram então registradas manualmente e a média calculada. A figura 5.6 ilustra a face média obtida, juntamente com sua representação wavelet, considerando apenas 52 funções base.

Utilizando a GWN otimizada nessa face média, a reparametrização funcionou perfeitamente em todas as imagens dos indivíduos presentes no banco de faces Yale. No entanto, é provável que esta técnica falhe ao considerarmos bancos de faces com um número bem maior de indivíduos.

Com vistas a realizar a localização de características faciais, pontos característicos são inicialmente localizados na face média. Os pontos que consideramos correspondem às pupilas, centro do nariz e centro da boca. Após, a GWN otimizada na face média é reparametrizada na imagem da face de teste. Nesta imagem, os pontos característicos são determinados mediante a aplicação de uma transformação afim adequada aos pontos originalmente marcados na face média. Os parâmetros corretos desta transformação são obtidos a partir do vetor de parâmetros da superwavelet $(s_x, s_y, s_{xy}, c_x, c_y, \theta)$, o qual é determinado pela equação (5.2) no processo de reparametrização.

A figura 5.7 ilustra a localização de características faciais em três indivíduos do banco de faces Yale. É interessante notar que o método é robusto à presença de óculos e barba na imagem da face. A figura 5.8 ilustra os resultados considerando imagens do mesmo indivíduo sob diferentes expressões faciais e mudanças de iluminação.

O rastreamento das características faciais é realizado mediante a aplicação de uma transformação afim aos pontos localizados, em cada quadro da seqüência¹. Claramente,

¹É importante salientar que não processamos diretamente os olhos, nariz e a boca do usuário; apenas obtemos uma estimativa da posição dessas características. Ainda assim, optamos pelo uso do termo “rastreamento de características faciais” para denotar o acompanhamento dos quatro pontos faciais de interesse no modelo wavelet.



Figura 5.6: Face média e sua representação wavelet.



Figura 5.7: Localização de características faciais em diferentes indivíduos.

os parâmetros corretos de cada transformação são derivados do vetor de parâmetros da superwavelet, em cada quadro. A figura 5.9 ilustra o rastreamento de características faciais em alguns quadros de uma seqüência de imagens.

Esse procedimento considera a geometria global da face, sendo robusto a deformações das características faciais, tais como piscar de olhos e sorriso, as quais constituem, em geral, uma situação crítica para a maioria das abordagens locais. Neste sentido, o método não exige a condição de alta correlação entre quadros adjacentes, como é exigido em técnicas baseadas em casamento de padrões (*template matching*), por exemplo.

É importante notar que o método pode ser facilmente generalizado para rastrear pontos característicos adicionais ou até mesmo regiões, como polígonos arbitrários ao redor dos olhos, nariz e boca. No capítulo 7, iremos apresentar alguns experimentos relacionados com esta abordagem.

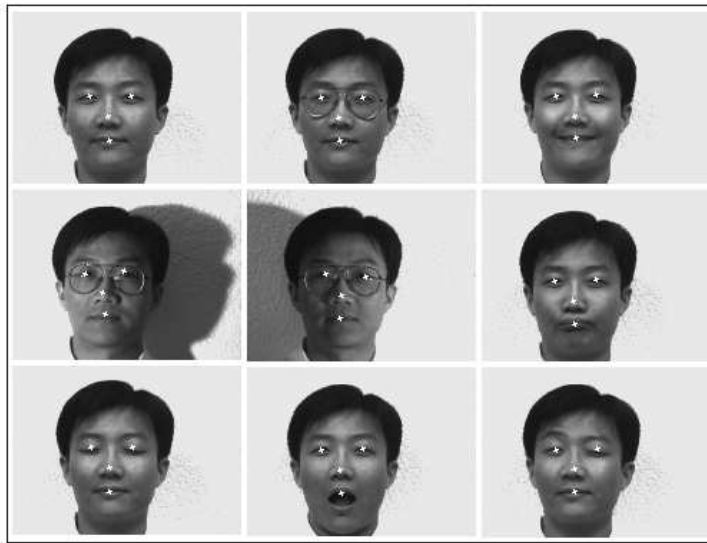


Figura 5.8: Localização de características faciais sob diferentes expressões faciais e mudanças de iluminação.



Figura 5.9: Rastreamento de características faciais utilizando GWNs.

Capítulo 6

Rastreamento de Faces em um Subespaço Wavelet

Neste capítulo, iremos apresentar, como principal contribuição científica deste trabalho, um método eficiente para rastreamento de faces baseado em GWNs. Basicamente, nosso objetivo consiste em realizar o rastreamento da face em um subespaço wavelet de dimensionalidade reduzida, visando a uma maior eficiência do processo.

Inicialmente, vamos considerar uma GWN (Ψ, \mathbf{w}) com $\Psi = \{\psi_{\mathbf{n}_1}, \dots, \psi_{\mathbf{n}_N}\}$ e $\mathbf{w} = (w_1, \dots, w_N)$. Conforme mencionado anteriormente, essa representação pode ser interpretada como uma superwavelet $\Psi_{\mathbf{n}}(\mathbf{x}) = \sum_{i=1}^N w_i \psi_{\mathbf{n}_i}(\mathbf{S}\mathbf{R}(\mathbf{x} - \mathbf{c}))$, onde N é o número de wavelets utilizadas na representação e $\mathbf{n} = (c_x, c_y, \theta, s_x, s_y, s_{xy})$ é o vetor de parâmetros da superwavelet, o qual define a matriz de dilatação \mathbf{S} , a matriz de rotação \mathbf{R} e o vetor de translação \mathbf{c} . Como visto na seção 5.2, o rastreamento da face pode ser obtido mediante a deformação afim da superwavelet $\Psi_{\mathbf{n}}$ em cada quadro da seqüência, conforme mostra a equação (5.2).

Claramente, esse método realiza um típico casamento de padrões (*template matching*) no espaço de imagens, em que o padrão, dado pela imagem obtida a partir da superwavelet, é distorcido gradualmente até *casar* com a face no quadro atual da seqüência. É interessante notar que os pesos w_i das wavelets em $\Psi_{\mathbf{n}}$ são constantes nesse processo. Assim, a deformação afim é capturada apenas pela deformação das wavelets, ao passo que o vetor de pesos permanece invariante.

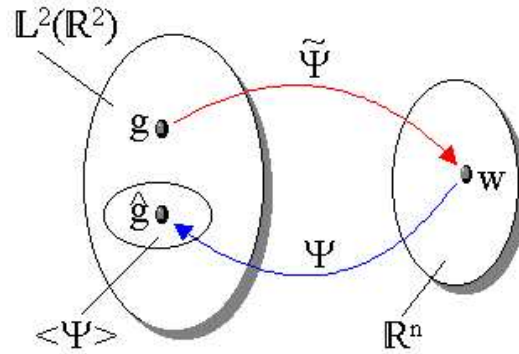


Figura 6.1: Mapeamentos entre o espaço de imagens $\mathbb{L}^2(\mathbb{R}^2)$ e o subespaço wavelet \mathbb{R}^N .

Com essa propriedade, concluímos que o rastreamento no espaço de imagens, acima descrito, pode também ser realizado com base apenas nos pesos das wavelets, isto é, em um espaço de dimensionalidade reduzida \mathbb{R}^N , o qual denominaremos *subespaço wavelet*.

De acordo com a seção 4.2, sabemos que uma imagem $g \in \mathbb{L}^2(\mathbb{R}^2)$ pode ser mapeada em um vetor $\mathbf{w} \in \mathbb{R}^N$ através do conjunto de wavelets duais $\tilde{\Psi} = \{\tilde{\psi}_{\mathbf{n}_1}, \dots, \tilde{\psi}_{\mathbf{n}_N}\}$. Tal mapeamento corresponde à projeção ortogonal de $g \in \mathbb{L}^2(\mathbb{R}^2)$ no subespaço $\langle \Psi \rangle \subset \mathbb{L}^2(\mathbb{R}^2)$. Analogamente, a reconstrução $\hat{g} \in \mathbb{L}^2(\mathbb{R}^2)$, isto é, o mapeamento de \mathbf{w} em $\mathbb{L}^2(\mathbb{R}^2)$, é obtido através do conjunto de wavelets Ψ . A figura 6.1 ilustra com clareza esses mapeamentos. Os espaços vetoriais envolvidos estão discriminados abaixo:

- Espaço de imagens $\mathbb{L}^2(\mathbb{R}^2)$, no caso contínuo. Uma imagem $g \in \mathbb{L}^2(\mathbb{R}^2)$ define um ponto nesse espaço.
- Espaço $\langle \Psi \rangle$, cuja base é definida pelas wavelets da representação. A reconstrução $\hat{g} = \sum_{i=1}^N w_i \psi_{\mathbf{n}_i}$ define um ponto nesse espaço. Claramente, $\langle \Psi \rangle$ constitui um subespaço do espaço de imagens.
- Subespaço wavelet \mathbb{R}^N . O vetor de pesos \mathbf{w} , obtido pela projeção ortogonal de g em $\langle \Psi \rangle$, define um ponto nesse espaço.

Esta discussão é perfeitamente válida para o caso de imagens discretas, as quais serão tratadas a seguir.

6.1 Metodologia

Para descrever o rastreamento no espaço \mathbb{R}^N , vamos considerar, inicialmente, uma GWN (Ψ, \mathbf{v}) otimizada na imagem de uma face, a qual desejamos rastrear em uma seqüência de vídeo. Como já mencionado, o vetor de pesos \mathbf{v} é obtido através da projeção ortogonal da imagem da face no subespaço $\langle \Psi \rangle$. Deste modo, temos o padrão da face mapeado no conjunto de pesos $\mathbf{v} \in \mathbb{R}^N$, os quais denominaremos *pesos de referência*.

Basicamente, o rastreamento no subespaço wavelet é realizado com a deformação afim do subespaço $\langle \Psi \rangle$, sendo aplicada até que o vetor de pesos $\mathbf{w} \in \mathbb{R}^N$, obtido pelo mapeamento ortogonal do quadro atual nesse subespaço, seja o mais próximo do vetor de pesos de referência \mathbf{v} . Esse procedimento assemelha-se com o método de rastreamento no espaço de imagens, com a diferença de que o processamento é realizado eficientemente no subespaço \mathbb{R}^N .

O mapeamento de imagens no \mathbb{R}^N é feito com um baixo custo computacional, através de um pequeno número de filtragens locais com as wavelets. De acordo com a seção 4.2, considerando uma imagem I , temos que $w_i = \langle I, \tilde{\psi}_{\mathbf{n}_i} \rangle$, em que $\tilde{\psi}_{\mathbf{n}_i} = \sum_j (A^{-1})_{ij} \psi_{\mathbf{n}_j}$. Simples transformações conduzem à equação abaixo:

$$w_i = \sum_j (A^{-1})_{ij} \langle I, \psi_{\mathbf{n}_j} \rangle \quad (6.1)$$

Assim, os pesos ótimos w_i são obtidos a partir de uma combinação linear de filtragens com as wavelets, onde os coeficientes são dados pela inversa da matriz $A_{ij} = \langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle$. É interessante notar que a matriz A_{ij} é, exceto por um fator de escala, invariante com relação a transformações afins da GWN¹. Isto permite que tal matriz seja calculada previamente, otimizando o cálculo dos pesos w_i , os quais são calculados eficientemente com a equação (6.1), através de uma aplicação local $\langle I, \psi_{\mathbf{n}_i} \rangle$ de cada uma das N Gabor wavelets $\psi_{\mathbf{n}_i}$, seguida por uma multiplicação com a inversa da matrix A_{ij} .

Seja $\mathbf{n} = (c_x, c_y, \theta, s_x, s_y, s_{xy})$ um vetor de parâmetros afim que configura uma parametrização para o subespaço $\langle \Psi \rangle$. Conforme descrito acima, o rastreamento no subespaço wavelet é obtido pela mudança gradual desses parâmetros até que o mapeamento do qua-

¹Os produtos internos $\langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle$ são alterados apenas por um fator de escala quando \mathbf{n}_i e \mathbf{n}_j são modificados pela mesma transformação afim.

dro atual J em $\mathbf{w} \in \mathbb{R}^N$ seja o mais próximo dos pesos de referência \mathbf{v} . Em outras palavras, devemos otimizar os parâmetros \mathbf{n} com relação à função energia:

$$E = \min_{\mathbf{n}} \|\mathbf{v} - \mathbf{w}\|_{\Psi} \quad (6.2)$$

$$\text{com } w_i = \sum_j \frac{1}{s_x \cdot s_y} (A^{-1})_{ij} \langle J, \psi_{\mathbf{n}_j}(\mathbf{S}\mathbf{R}(\mathbf{x} - \mathbf{c})) \rangle \quad (6.3)$$

onde \mathbf{S} é a matriz de dilatação, \mathbf{R} é a matriz de rotação e \mathbf{c} é o vetor de translação, todos definidos através do vetor $(c_x, c_y, \theta, s_x, s_y, s_{xy})$. Durante o rastreamento, essa otimização é feita para cada quadro da seqüência. Como existe pouca diferença entre dois quadros adjacentes, a otimização é rápida. Novamente, para minimizar a função energia (6.2), o algoritmo Levenberg-Marquardt foi utilizado.

6.2 Distância entre Pesos

Na equação (6.2), utilizamos a notação $\|\mathbf{v} - \mathbf{w}\|_{\Psi}$ para denotar a distância entre os vetores \mathbf{v} e \mathbf{w} no subespaço \mathbb{R}^N . No entanto, ainda não é claro como essa medida de distância deve ser calculada. De acordo com o trabalho de [Wiskott et al., 1997], a distância Euclidiana poderia ser utilizada entre esses dois vetores, mas tal medida trata igualmente as wavelets da representação, ainda que essas funções apresentem diferentes escalas.

Nós então propomos uma nova medida de distância, a qual é baseada na distância Euclidiana entre vetores do espaço de imagens. Assim, definimos a diferença $\|\mathbf{v} - \mathbf{w}\|_{\Psi}$ como a distância Euclidiana entre as duas respectivas imagens reconstruídas:

$$\|\mathbf{v} - \mathbf{w}\|_{\Psi} = \left\| \sum_{i=1}^N v_i \psi_{\mathbf{n}_i} - \sum_{j=1}^N w_j \psi_{\mathbf{n}_j} \right\|_2 \quad (6.4)$$

Várias transformações são então realizadas:

$$\begin{aligned} \|\mathbf{v} - \mathbf{w}\|_{\Psi} &= \left\| \sum_{i=1}^N v_i \psi_{\mathbf{n}_i} - \sum_{j=1}^N w_j \psi_{\mathbf{n}_j} \right\|_2 \\ &= \left[\int \left(\sum_{i=1}^N v_i \psi_{\mathbf{n}_i}(\mathbf{x}) - \sum_{j=1}^N w_j \psi_{\mathbf{n}_j}(\mathbf{x}) \right)^2 d\mathbf{x} \right]^{\frac{1}{2}} \end{aligned} \quad (6.5)$$

$$\begin{aligned}
 &= \left[\int \left(\sum_{i=1}^N \delta_i \psi_{\mathbf{n}_i}(\mathbf{x}) \right)^2 d\mathbf{x} \right]^{\frac{1}{2}} \quad \text{com } \delta_i = (v_i - w_i) \\
 &= \left[\int \left(\sum_{i=1}^N \sum_{j=1}^N \delta_i \delta_j \psi_{\mathbf{n}_i}(\mathbf{x}) \psi_{\mathbf{n}_j}(\mathbf{x}) \right) d\mathbf{x} \right]^{\frac{1}{2}} \\
 &= \left[\sum_{i=1}^N \sum_{j=1}^N \delta_i \delta_j \int \psi_{\mathbf{n}_i}(\mathbf{x}) \psi_{\mathbf{n}_j}(\mathbf{x}) d\mathbf{x} \right]^{\frac{1}{2}} \\
 &= \left[\sum_{i,j} \delta_i \delta_j \langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle \right]^{\frac{1}{2}} \\
 &= [\delta^T A_{ij} \delta]^{\frac{1}{2}} \quad \text{com } \delta = (\delta_1 \cdots \delta_N)^T
 \end{aligned} \tag{6.6}$$

Claramente, a medida de distância é caracterizada no espaço \mathbb{R}^N , sendo calculada eficientemente. A matriz de produtos escalares $A_{ij} = \langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle$ é a mesma matriz presente na eq. (6.1) e na seção 4.2. É interessante notar que, no caso das wavelets $\{\psi_{\mathbf{n}_i}\}$ serem ortonormais, A_{ij} é a matriz unitária e a equação (6.5) descreve a mesma medida de distância proposta em [Wiskott et al., 1997].

6.3 Algoritmo

Com o intuito de tornar mais clara a seqüência de etapas envolvidas no método de rastreamento de faces em um subespaço wavelet, apresentamos, a seguir, um algoritmo em alto nível do referido método. Por questões de simplicidade, o algoritmo considera a presença de uma face no primeiro quadro de uma seqüência de imagens, supondo-se conhecida a sua posição, orientação e escala.

Dada uma seqüência de imagens $\{J_1, J_2, \dots, J_N\}$:

→ Otimize a GWN (Ψ, \mathbf{v}) na face da imagem J_1 , obtendo, desta forma, os pesos de referência \mathbf{v} .

→ Inicialize o vetor de parâmetros $\mathbf{n}_1 = (c_x, c_y, \theta, s_x, s_y, s_{xy})$ com a posição, escala e orientação da face na imagem J_1 . Isto configura uma parametrização inicial do subespaço $\langle \Psi \rangle$.

Para i de 2 até N faça

→ Calcule \mathbf{w} , projetando J_i em $\langle \Psi \rangle$ com a equação (6.3).

Enquanto $\|\mathbf{v} - \mathbf{w}\|_{\Psi}$ não é mínimo faça

→ Ajuste o vetor de parâmetros \mathbf{n}_i (deformação do subespaço $\langle \Psi \rangle$) com o método Levenberg-Marquardt.

→ Calcule \mathbf{w} , projetando J_i em $\langle \Psi \rangle$ com a equação (6.3).

Fim Enquanto

→ Imprima o vetor \mathbf{n}_i : posição, escala e orientação da face no quadro J_i .

Fim Para

A técnica proposta apresenta um considerável aumento de performance quando comparada ao rastreamento no espaço de imagens. Essa vantagem, em termos de eficiência, deve-se ao fato de que o rastreamento é realizado em um subespaço wavelet de dimensionalidade reduzida, promovendo uma alta redução de dados. Além disso, deve-se notar que a reconstrução do padrão da face, bem como o cálculo da diferença *pixel a pixel* entre imagens, são fatores críticos em termos de tempo de processamento, dispensados no rastreamento baseado nos pesos das wavelets. O próximo capítulo apresentará uma avaliação quantitativa dessa técnica.

Capítulo 7

Resultados Experimentais

Neste capítulo, iremos descrever os experimentos realizados com as técnicas introduzidas neste trabalho, bem como discutir os resultados obtidos. Inicialmente, apresentaremos uma avaliação quantitativa do método de rastreamento de faces proposto no capítulo 6, considerando aspectos como atenção progressiva, robustez e eficiência. Logo após, os conceitos apresentados nessa dissertação serão reunidos e aplicados no problema de reconhecimento de faces a partir de seqüências de vídeo.

7.1 Avaliação da Técnica de Rastreamento de Faces

O método proposto para rastreamento de faces em um subespaço wavelet foi testado com sucesso em várias seqüências de imagens. Todas as seqüências de teste apresentavam um indivíduo em movimento, mais ou menos frontal à câmera, de forma que suas características faciais estivessem sempre visíveis. Os experimentos foram conduzidos visando a analisar a performance do rastreamento com relação a fatores como precisão, robustez e eficiência.

7.1.1 Análise de Precisão

A precisão do método de rastreamento foi avaliada considerando o conceito de atenção progressiva, isto é, variando o número de wavelets na representação. Inicialmente, grava-

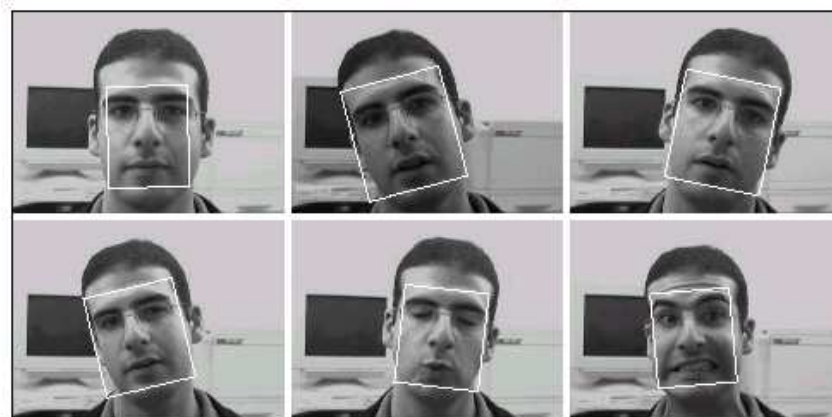


Figura 7.1: Rastreamento de faces no subespaço wavelet, utilizando-se uma GWN com 116 wavelets.

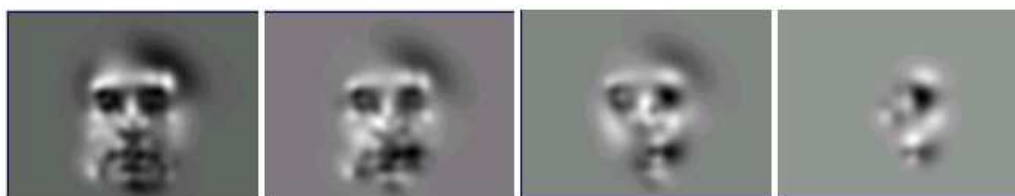


Figura 7.2: Representações obtidas com 116, 51, 22 e 9 wavelets.

mos uma seqüência de imagens de uma pessoa se movendo diante da câmera com diferentes expressões faciais. No interior da face desse indivíduo, uma GWN com 116 Gabor wavelets foi otimizada e utilizada para estimar os parâmetros afins de referência, em cada quadro da seqüência. A figura 7.1 ilustra alguns quadros do resultado da aplicação do novo método de rastreamento de faces nessa seqüência. É interessante notar a robustez do método com relação a deformações afins da face e a variações de expressão facial.

Com o objetivo de investigar o princípio da atenção progressiva no rastreamento da face, foram extraídas GWNs de menor precisão, a partir da representação de 116 wavelets, contendo somente as 51, 22 e 9 wavelets de maior peso, normalizadas com relação à escala. A figura 7.2 ilustra tais representações. Basicamente, queremos utilizar essas GWNs para analisar o decréscimo de precisão ao diminuir o número de wavelets na representação.

Assim, o rastreamento da face no subespaço wavelet foi realizado com cada uma das

GWNs citadas acima. As figuras 7.3, 7.4 e 7.5 ilustram a estimação dos parâmetros de posição c_x , c_y e o ângulo θ da face, respectivamente. Em cada gráfico, os parâmetros de referência são denotados pela linha pontilhada, ao passo que a linha contínua refere-se aos parâmetros estimados com 51, 22 e 9 wavelets. Os resultados obtidos com o rastreamento baseado no espaço de imagens são similares.

Mediante a análise visual desses gráficos, concluímos que a precisão do rastreamento depende diretamente do número de Gabor wavelets utilizadas na representação. De fato, a medida que diminuimos o número de wavelets, temos um decréscimo na precisão da estimação dos parâmetros em cada quadro da sequência. Por outro lado, temos um correspondente aumento de eficiência no rastreamento, como veremos adiante. Logo, a escolha do número de filtros depende da aplicação considerada, podendo ser ajustada de acordo com os recursos computacionais disponíveis.

7.1.2 Análise de Robustez

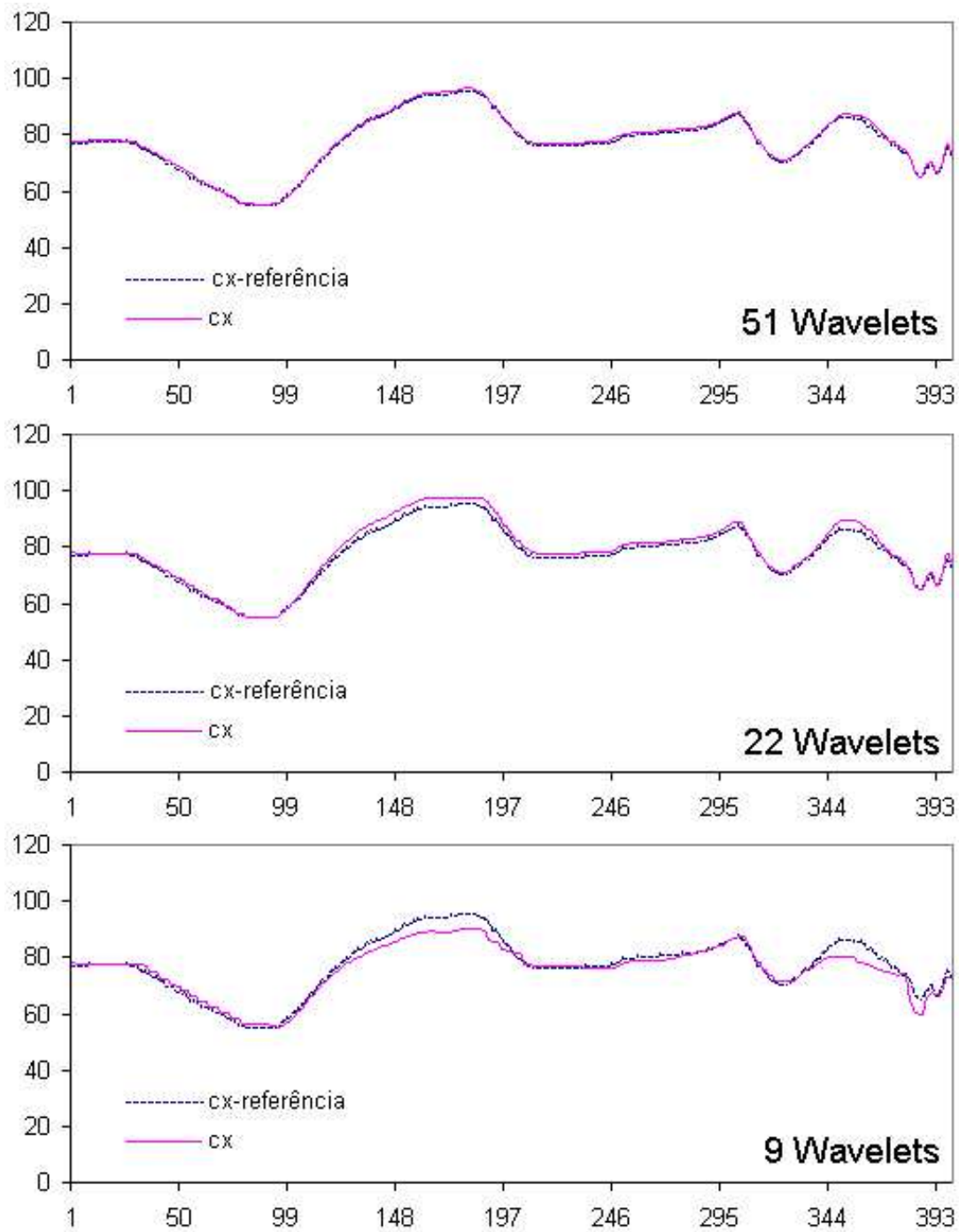
De acordo com nossos experimentos, o rastreamento no subespaço wavelet apresenta a mesma robustez do rastreamento baseado no espaço de imagens. Ambas abordagens são robustas à variação homogênea das condições de iluminação, bem como a deformações afins da imagem da face. É interessante notar que variações de expressão facial e rotações leves da face fora do plano são, em geral, corretamente aproximadas pelo modelo afim baseado em wavelets.

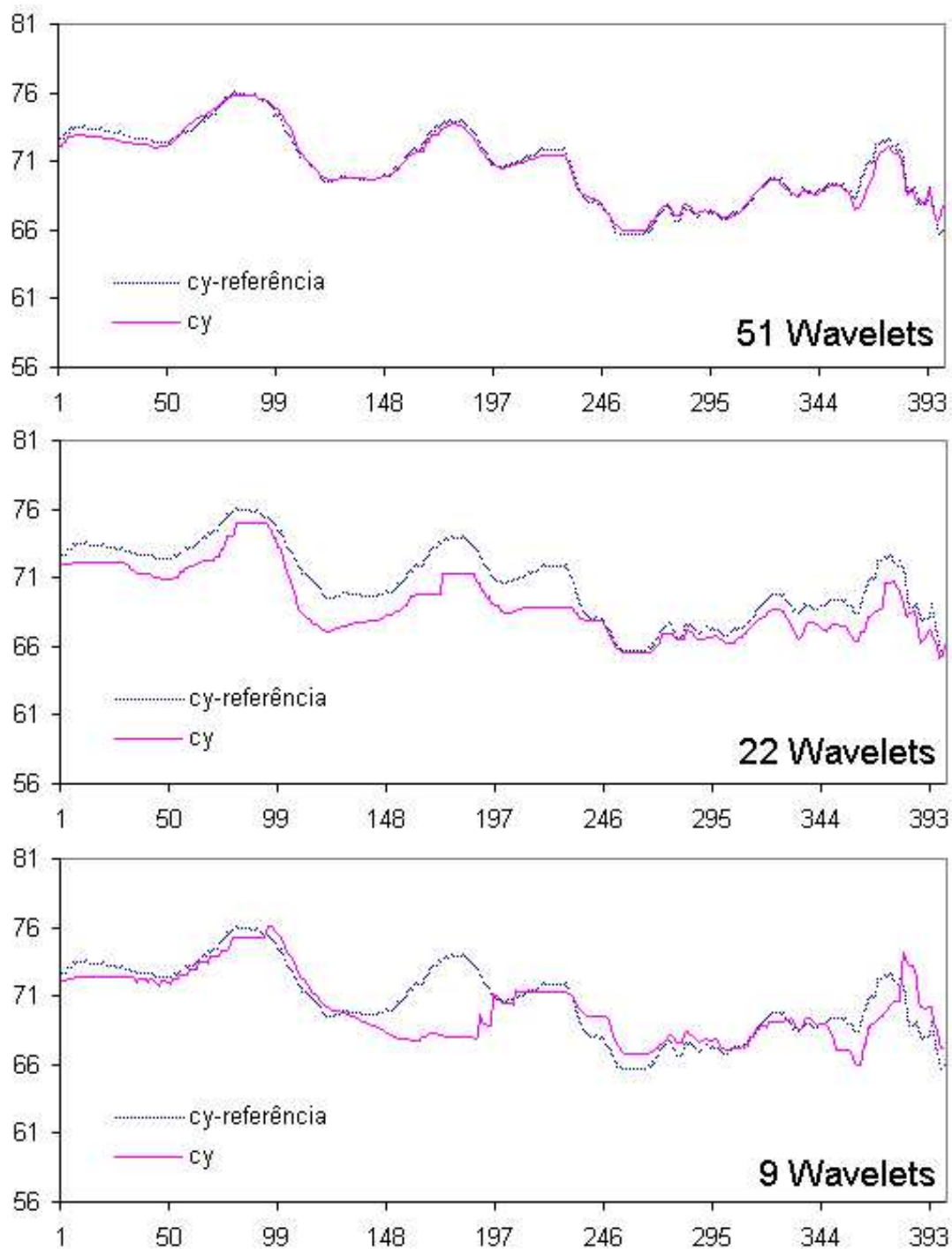
Os testes realizados mostraram, também, a robustez do método com relação à velocidade da face. Em outras palavras, o rastreamento é realizado mesmo na presença de considerável diferença entre dois quadros adjacentes.

O gráfico ilustrado na figura 7.6 mostra os deslocamentos (em *pixels*) da posição da face na sequência considerada. Tais deslocamentos foram calculados, em cada quadro, através dos parâmetros de referência c_x e c_y . Basicamente, considerando (c_x, c_y) a posição da face no quadro atual e (c_{xold}, c_{yold}) a posição da face no quadro anterior, o deslocamento d da face para o quadro atual foi calculado conforme a equação:

$$d = \sqrt{(c_x - c_{xold})^2 + (c_y - c_{yold})^2} \quad (7.1)$$

Observando o gráfico, percebemos que o deslocamento da face é relativamente alto em

Figura 7.3: Estimação do parâmetro de posição c_x em cada quadro da seqüência.

Figura 7.4: Estimação do parâmetro de posição c_y em cada quadro da seqüência.

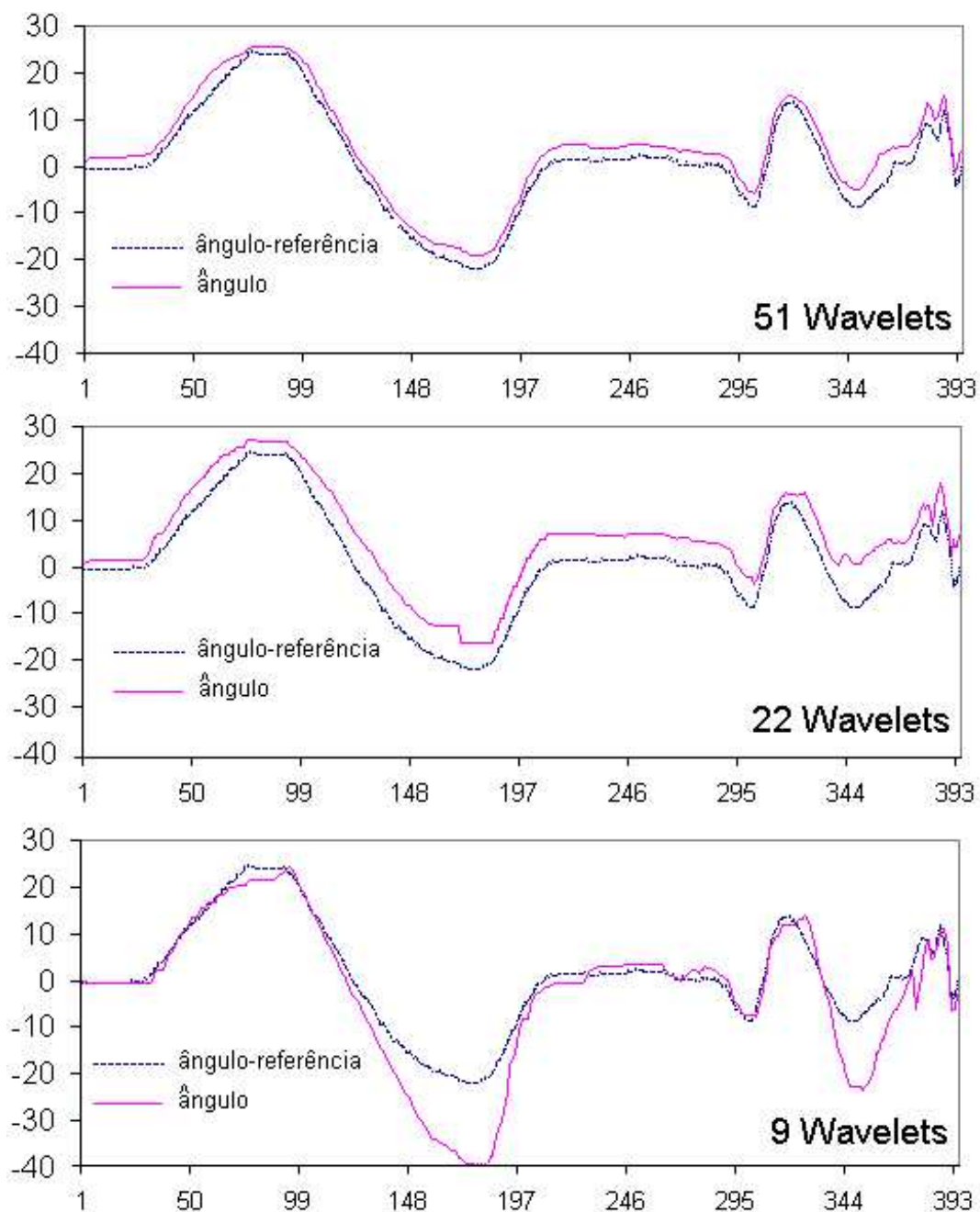


Figura 7.5: Estimação do parâmetro referente ao ângulo da face em cada quadro da seqüência.



Figura 7.6: Deslocamento da face em cada quadro da seqüência.

determinados quadros. Quando isso ocorre, o rastreamento é realizado com robustez, mas o tempo requerido pelo procedimento de reparametrização é aumentado, devido ao maior número de ciclos no processo de otimização.

7.1.3 Análise de Eficiência

A resolução das imagens mostradas na figura 7.1 é 160×120 pixels e o tamanho da face na qual a GWN foi otimizada é 50×65 pixels. Utilizando-se somente 9 wavelets, o tempo de computação para cada ciclo do algoritmo Levenberg-Marquardt é de 15ms, em um Athlon 1GHz rodando Linux. Uma performance ainda mais alta pode ser alcançada se considerarmos imagens da face de menor resolução, ou mesmo, considerando menos parâmetros (e.g. somente translação e rotação).

É importante salientar que o número de ciclos para cada quadro depende da distância entre os parâmetros iniciais e o mínimo local, sendo diretamente relacionado com a velocidade da face. O aumento do número de wavelets na representação promove um rastreamento mais preciso, mas também mais lento. Por exemplo, utilizando-se 51 wavelets, foi exigido um tempo de computação de 85ms por ciclo, em cada quadro da seqüência considerada.

Em comparação com o rastreamento baseado no espaço de imagens, observamos, na

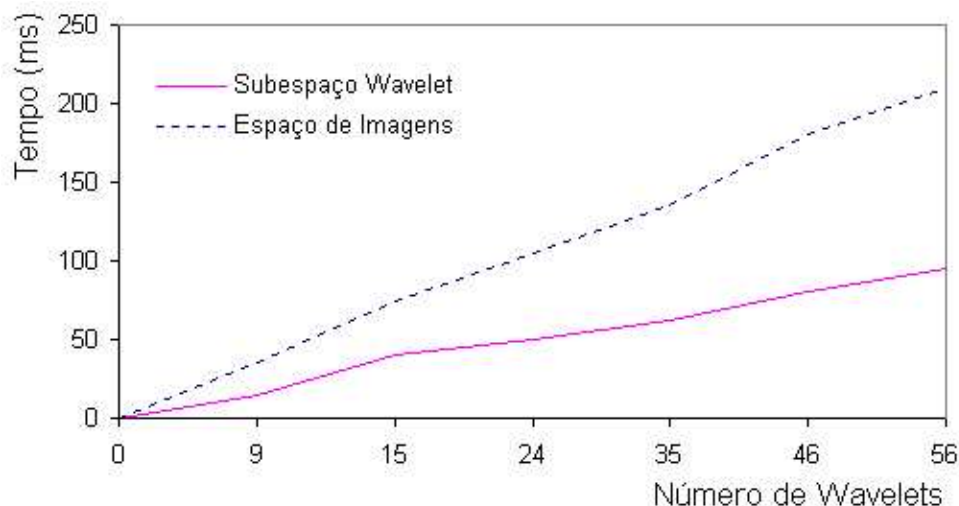


Figura 7.7: Comparação dos tempos exigidos por ciclo entre o rastreamento no espaço de imagens e o método proposto.

técnica proposta, uma economia de aproximadamente 50% do custo computacional exigido em cada ciclo do método Levenberg-Marquardt, com um pequeno aumento no número de ciclos requeridos. Acreditamos que a utilização de outros algoritmos, como os métodos Condensation[Isard and Blake, 1998] ou o *Sequential Importance Sampling* (SIS)[Li, 2000] possivelmente guiarão para um aumento ainda maior de eficiência. O gráfico ilustrado na figura 7.7 mostra a comparação dos tempos exigidos por ciclo entre o rastreamento no espaço de imagens e o método proposto.

Deve-se mencionar que o cálculo dos tempos de computação dependem sobremaneira do *hardware* e do sistema operacional utilizado. Além disso, os resultados foram apresentados considerando uma seqüência de vídeo específica. Apesar da obtenção de resultados similares em outras seqüências de teste, torna-se difícil traçar uma conclusão genérica.

7.2 Aplicação: Reconhecimento de Faces a partir de Vídeo

Os conceitos apresentados neste trabalho foram reunidos visando uma aplicação específica: reconhecimento de faces a partir de seqüências de vídeo. Basicamente, um sistema foi configurado com as seguintes etapas:

- Detecção da face;
- Posicionamento das características faciais;
- Rastreamento da face e das características faciais;
- Seleção e normalização dos melhores quadros.

As características faciais normalizadas são então enviadas ao módulo de reconhecimento, incumbido da identificação do usuário. O esquema desse método proposto para reconhecer pessoas a partir de vídeo está descrito no artigo [Feris et al., 2000b] em anexo.

7.2.1 Formação do Banco de Dados

Para realizar os testes do nosso sistema, integrado com o módulo de reconhecimento, um banco de seqüências de vídeo foi gerado, solicitando-se a colaboração de 10 pessoas do departamento. As imagens, bem como os experimentos realizados, serão armazenados em um CD, com a devida autorização das pessoas envolvidas.

Para cada pessoa, duas seqüências, de treinamento e de teste, foram gravadas em tempos distintos, com um intervalo de uma semana. É interessante notar as mudanças na aparência do indivíduo, após esse intervalo, como diferentes penteados, presença ou ausência de barba, óculos, etc.

Inicialmente, solicitamos que cada pessoa se posicionasse na frente de um computador, com o intuito de ser filmada por uma câmera colorida. Assumimos que a imagem da face do usuário seja frontal, para facilitar o processo de detecção. Em seguida, pedimos a cada pessoa para mover-se naturalmente diante da câmera, configurando um processo

de interação com a máquina. Nesse processo, as características faciais - olhos, nariz e boca - devem estar presentes em todos os quadros da seqüência.

Ao final, um total de 20 seqüências de imagens foram gravadas. Deve-se mencionar que este banco de dados está sendo expandido, de forma que uma avaliação mais completa e precisa será apresentada em trabalho futuro.

A seguir, discutiremos os resultados obtidos com cada módulo do sistema, no contexto da aplicação considerada.

7.2.2 Detecção da Face

Com relação à detecção de faces, o método mostrou-se robusto ao identificar regiões de cor da pele de pessoas de diferentes raças, idades e sexo. Entretanto, cores de pele escura, quando comparadas com cores de pele clara, corresponderam a pixels de menor intensidade na imagem em níveis de cinza. Isto se deve ao fato de que, no treinamento supervisionado, utilizou-se um maior número de cores de pele clara. Ainda assim, as cores de pele escura são classificadas corretamente quando da identificação de regiões de cor da pele. Em geral, o método falha sob alta variação das condições de iluminação ou quando a face está muito próxima de outros objetos com cor da pele.

A verificação da presença ou ausência da face na região de cor de pele foi realizada, como já comentado, utilizando-se uma técnica simples de correlação. Visando a aumentar a robustez desse módulo, pretendemos ainda utilizar uma combinação de outras informações, como movimento, simetria, profundidade, forma, e outras.

7.2.3 Posicionamento das Características Faciais

Uma vez que a face tenha sido detectada, sua informação de escala é obtida e a região de cor da pele é convertida para uma imagem em níveis de cinza. As características faciais são então localizadas através do processo de reparametrização da GWN, otimizada em uma face média. A posição e a escala do *blob* da face são utilizadas como parâmetros iniciais na reparametrização. A figura 7.8 ilustra a detecção de uma face no primeiro quadro de uma seqüência de imagens, bem como a localização de suas características faciais.



Figura 7.8: Detecção da face e localização das características faciais.

Consideramos que o método utilizado para o posicionamento das características faciais apresenta duas limitações. A primeira está relacionada com a reparametrização da GWN, visto que esse procedimento é adequado apenas quando bancos de dados com poucos indivíduos são tratados. A segunda limitação é derivada do fato de que a distância entre características faciais varia de pessoa para pessoa, ocasionando, em alguns casos, resultados imprecisos. Assim, seria interessante dispormos de um procedimento incumbido de ajustar a posição dos pontos característicos após a reparametrização da GWN. Como exemplo, nosso método poderia ser utilizado para a inicialização de contornos ativos[Blake and Isard, 1998], objetivando uma segmentação precisa das características faciais. Atualmente, uma aluna de doutorado do nosso grupo de pesquisa está investigando novos métodos para solucionar esse problema.

7.2.4 Rastreamento da Face e das Características Faciais

O rastreamento da face e das características faciais foi realizado corretamente nas seqüências de imagens consideradas, mostrando-se robusto a deformações afins da face, diferentes expressões faciais e mudanças homogêneas das condições de iluminação. A figura 7.9 ilustra o rastreamento de olhos, nariz e boca de uma seqüência específica, apresentando os quadros 60, 98 e 221.

De acordo com nossos experimentos, o rastreamento é realizado com maior precisão quando a GWN é otimizada na face do indivíduo considerado. A GWN otimizada na face média também possibilita o rastreamento, mas com menor precisão. Assim, após o posicionamento das características faciais, a representação wavelet da face é obtida, e o módulo de rastreamento acionado.

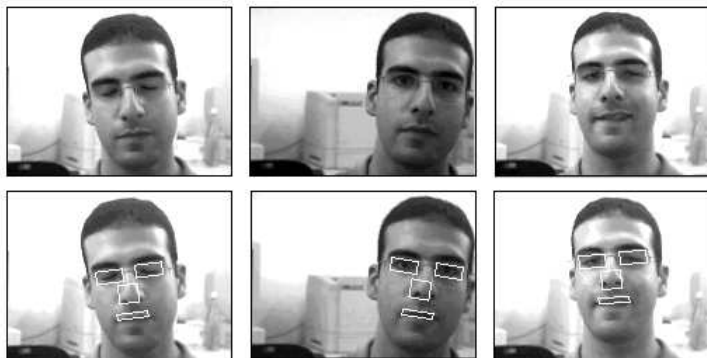


Figura 7.9: Rastreamento de olhos, nariz e boca. Note que o método é robusto a piscar de olhos, mudança homogênea de iluminação e diferentes expressões faciais.

7.2.5 Seleção e Normalização dos Melhores Quadros

Durante o processo de rastreamento, é possível que os pontos característicos não sejam precisamente localizados em determinados quadros. Isso ocorre quando o usuário realiza rotações leves fora do plano, modifica sua expressão facial, ou mesmo quando ocorrem mudanças nas condições de iluminação.

Assumindo que as características faciais são corretamente posicionadas inicialmente, é possível selecionar os melhores quadros e eliminar aqueles em que a localização dos pontos característicos não foi efetuada precisamente. Basicamente, se a reparametrização da GWN não foi realizada com precisão, isto é, se a função energia na equação (6.2) assumir um valor maior que um limiar, então o quadro pode ser eliminado.

Ao final, as características faciais dos quadros selecionados são extraídas, normalizadas, e enviadas ao módulo de reconhecimento. A normalização foi realizada aplicando-se uma transformação afim adequada nas regiões selecionadas, de acordo com os correspondentes parâmetros da superwavelet. Outros aspectos como normalização de iluminação precisam ser tratados posteriormente.

Capítulo 8

Conclusões

Neste trabalho, foram apresentados algoritmos de visão computacional para a detecção e rastreamento de faces humanas em seqüências de imagens. Mais especificamente, esse estudo concentrou-se na etapa de rastreamento de faces, sendo proposta, como principal contribuição, uma técnica eficiente para esse problema, baseada em um subespaço wavelet. Tal técnica foi avaliada considerando aspectos como precisão, robustez e eficiência. O problema de localização e rastreamento de características faciais, bem como alguns aspectos relativos à normalização de imagens, também foram comentados, visando a uma aplicação específica: reconhecimento de faces a partir de seqüências de vídeo.

Diversos experimentos foram realizados com a técnica baseada em *Gabor Wavelet Networks*, recentemente proposta na literatura. Entre outras vantagens, verificamos a robustez dessa técnica com relação a deformações rígidas e não rígidas da face no plano, bem como à variação homogênea das condições de iluminação. O compromisso entre precisão e eficiência é determinado pela escolha do número de wavelets na representação. Logo, essa escolha deve ser feita com base nas condições do problema e nos recursos computacionais disponíveis.

Os conceitos apresentados no presente estudo foram reunidos, ao final, objetivando a aplicação considerada, isto é, o reconhecimento de faces a partir de seqüências de vídeo. A pesquisa que subsidiou este trabalho possibilitou resposta a algumas questões, porém suscitou inúmeras outras, inerentes a toda investigação científica. A seguir, apresentamos tópicos a serem tratados em trabalho futuro:

- A verificação da presença ou ausência de uma face em um *blob* de cor de pele deve ser aperfeiçoada com a integração de novas informações como movimento, simetria, forma e outras.
- Pesquisas devem ser realizadas para se obter uma representação genérica de uma face baseada em GWNs. Atualmente, a representação utilizando-se uma face média é adequada apenas para banco de dados com poucos elementos.
- Durante o rastreamento, fatores como rotações leves fora do plano, variação nas condições de iluminação e mudanças de expressão facial são, em geral, bem aproximados pelo modelo wavelet afim. No entanto, é possível que, nessas condições, exista imprecisão no procedimento de posicionamento das características faciais. Esse problema também ocorre na localização dessas características utilizando-se uma GWN otimizada na face média, tendo em vista que a distância entre características faciais varia de pessoa para pessoa. Assim, seria interessante dispormos de um procedimento incumbido de ajustar a posição dos pontos característicos após a reparametrização da GWN, possibilitando uma segmentação precisa das características faciais. Esse tópico está sendo tratado atualmente em nosso grupo de pesquisa.
- Quando a face do usuário realiza rotações consideráveis fora do plano, o método de rastreamento falha. A solução para esse problema seria estender o atual modelo afim da face para uma representação wavelet 3D.
- O rastreamento de faces ainda pode ser aperfeiçoado em termos de eficiência, com a utilização de métodos tais como o Condensation [Isard and Blake, 1998] ou o *Sequential Importance Sampling* (SIS) [Li, 2000], ao invés do algoritmo de Levenberg-Marquardt.
- É interessante comparar os resultados obtidos neste trabalho com outras técnicas que compõem o estado da arte na literatura.

Apêndice A

Lista de Símbolos

$\langle f, g \rangle$	Produto interno em $L^2(\mathbb{R}^n)$: $\langle f, g \rangle = \int f(t)g(t)dt$
$L^2(\mathbb{R})$	Espaço das funções quadraticamente integráveis $f : \mathbb{R} \mapsto \mathbb{R}$
$L^2(\mathbb{R}^2)$	Espaço das funções quadraticamente integráveis $f : \mathbb{R}^2 \mapsto \mathbb{R}$
$g_{u,b}$	Função janela deslocada no tempo (parâmetro b) e modulada por uma exponencial complexa (parâmetro u)
ψ	Wavelet mãe
$\psi_{a,b}$	Wavelet ψ deslocada no tempo (parâmetro b), com uma escala específica (parâmetro a)
C_ψ	Condição de admissibilidade das wavelets
\mathbf{c}	Vetor de translação
\mathbf{R}	Matriz de rotação

\mathbf{S}	Matriz de escala
\mathbf{n}	Vetor de parâmetros da wavelet; $\mathbf{n} = (c_x, c_y, \theta, s_x, s_y)$ Vetor de parâmetros da superwavelet $\mathbf{n} = (c_x, c_y, \theta, s_x, s_y, s_{xy})$
$\psi_{\mathbf{n}_i}$	Gabor wavelet com vetor de parâmetros \mathbf{n}_i
\mathbf{w}, \mathbf{v}	Vetores de pesos $\mathbf{w} = (w_1, \dots, w_N)$, $\mathbf{v} = (v_1, \dots, v_N)$
f, g	Funções contínuas
\hat{f}	Aproximação de f ; Representação wavelet
I, J	Imagens discretas em níveis de cinza
$\Psi, \tilde{\Psi}$	Base biortogonal de Gabor wavelets $\Psi = \{\psi_{\mathbf{n}_1}, \dots, \psi_{\mathbf{n}_N}\}$ e $\tilde{\Psi} = \{\tilde{\psi}_{\mathbf{n}_1}, \dots, \tilde{\psi}_{\mathbf{n}_N}\}$
$\Phi, \tilde{\Phi}$	Base biortogonal de wavelets $\Phi = \{\phi_1, \dots, \phi_N\}$ e $\tilde{\Phi} = \{\tilde{\phi}_{\mathbf{n}_1}, \dots, \tilde{\phi}_{\mathbf{n}_N}\}$
$\langle \Psi \rangle$	Espaço vetorial definido pelas Gabor wavelets
E	Função energia
(Ψ, \mathbf{w})	<i>Gabor wavelet Network</i>
M, N	Número de wavelets
$\delta_{i,k}$	Função de Dirac $\delta_{i,k} = \begin{cases} 1, & \text{se } i = k \\ 0, & \text{se } i \neq k \end{cases}$

A_{ij}	Matriz de interferência de wavelets
$\Psi_{\mathbf{n}}$	Superwavelet
δ	Diferença entre dois vetores de pesos $\delta = (\mathbf{v} - \mathbf{w})$
\mathbb{R}^N	Subespaço wavelet

Apêndice B

Lista de Abreviaturas

GWN	<i>Gabor wavelet network</i>
GWT	<i>Gabor wavelet template</i>
GSW	<i>Gabor superwavelet</i>
PCA	<i>Principal component analysis</i>
RBF	<i>Radial basis functions</i>

Referências Bibliográficas

- [Antoine et al., 1993] Antoine, J., Carrete, P., Murenzi, R., and Piette, B. (1993). Image analysis with two-dimensional continuous wavelet transform. *Signal Processing*, 31:241–272.
- [Bala et al., 1997] Bala, L., Talmi, K., and Liu, J. (1997). Automatic detection and tracking of faces and facial features in video sequences. In *Picture Coding Symposium*, Berlin, Germany.
- [Birchfield, 1998] Birchfield, S. (1998). Elliptical head tracking using intensity gradients and color histograms. In *Proc. of IEEE CVPR*, pages 232–237, Santa Barbara, California.
- [Blake and Isard, 1998] Blake, A. and Isard, M. (1998). *Active Contours*. Springer.
- [Brunelli and Poggio, 1993] Brunelli, R. and Poggio, T. (1993). Face recognition: Features versus templates. *IEEE Transactions on PAMI*, 15(10):1042–1052.
- [Campos, 2001] Campos, T. (2001). Técnicas de seleção de características com aplicações em reconhecimento de faces. Master’s thesis, Universidade de São Paulo, São Paulo, Brasil.
- [Campos et al., 2000a] Campos, T., Feris, R., and Cesar, R. (2000a). Eigenfaces versus eigeneyes: First steps towards performance assessment of representations for face recognition. In *Lecture Notes in Artificial Intelligence*, volume 1793, pages 193–201, Acapulco, Mexico.
- [Campos et al., 2000b] Campos, T., Feris, R., and Cesar, R. (2000b). Improved face x non-face discrimination using fourier descriptors through feature selection. In *Brazilian*

- Symposium on Computer Graphics and Image Processing*, pages 28–35, Gramado-RS, Brazil.
- [Cascia et al., 2000] Cascia, M., Sclaroff, S., and Athitsos, V. (2000). Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *IEEE Transactions on PAMI*, 22(4):322–336.
- [Cesar and Costa, 1998] Cesar, R. and Costa, L. (1998). Neural cell classification by wavelets and multiscale curvature. *Biological Cybernetics*, 79(4):347–360.
- [Cipolla and Pentland, 1998] Cipolla, R. and Pentland, A. (1998). *Computer Vision for Human-Machine Interaction*. Cambridge University Press.
- [Coelho et al., 1999] Coelho, P., Esperança, C., and Oliveira, A. (1999). An image processing and believe network approach to face detection. In *Brazilian Symposium on Computer Graphics and Image Processing*, pages 177–186, Campinas-SP, Brazil.
- [Costa and Cesar, 2001] Costa, L. and Cesar, R. (2001). *Shape Analysis and Classification: Theory and Practice*. CRC Press Book Series on Image Processing.
- [Cox et al., 1996] Cox, I., Ghosn, J., and Yianilos, P. (1996). Feature-based face recognition using mixture-distances. In *Proc. of IEEE CVPR*, pages 209–216, Seattle, WA.
- [Crowley and Schwerdt, 1999] Crowley, J. and Schwerdt, K. (1999). Robust tracking and compression for video communication. In *ICCV'99 Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Corfu, Greece.
- [Dai and Nakano, 1996] Dai, Y. and Nakano, Y. (1996). Face-texture model-based on sgld and its application in face detection in a color scene. *Pattern Recognition*, 29(6):1007–1017.
- [Darrel et al., 1998] Darrel, T., Gordon, G., Woodfill, J., and Harville, M. (1998). A virtual mirror interface using real-time robust face tracking. In *Proc. of IEEE Face and Gesture Recognition*, Nara, Japan.
- [Daubechies, 1990] Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005.

- [Daubechies, 1992] Daubechies, I. (1992). *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia, Pennsylvania.
- [Daugman, 1985] Daugman, J. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized two-dimensional visual cortical filters. *Optical Society of America*, 2(7):1160–1168.
- [Daugman, 1988] Daugman, J. (1988). Complete discrete 2d gabor transform by neural networks for image analysis and compression. *IEEE Transactions on ASSP*, 36(7):1169–1179.
- [Edwards et al., 1998] Edwards, G., Taylor, C., and Cootes, T. (1998). Learning to identify and track faces in image sequences. In *Proc. of IEEE Face and Gesture Recognition*, pages 260–265, Nara, Japan.
- [Feris et al., 2000a] Feris, R., Campos, T., and Cesar, R. (2000a). Detection and tracking of facial features in video sequences. In *Lecture Notes in Artificial Intelligence*, volume 1793, pages 127–135, Acapulco, Mexico.
- [Feris et al., 2000b] Feris, R., Campos, T., and Cesar, R. (2000b). A project for face recognition from video sequences using gwn and eigenfeature selection. In *Workshop on Artificial Intelligence and Computer Vision*, pages 141–145, Atibaia - SP, Brazil.
- [Feris and Cesar, 2001] Feris, R. and Cesar, R. (2001). Detection and tracking of facial landmarks using gabor wavelet networks. In *Lecture Notes in Computer Science*, Rio de Janeiro, Brasil. Springer-Verlag Press.
- [Feris et al., 2001] Feris, R., Krüger, V., and Cesar, R. (2001). Efficient real-time face tracking in wavelet subspace. In *ICCV'2001 Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Vancouver, Canada.
- [Feyrer and Zell, 1999] Feyrer, S. and Zell, A. (1999). Tracking and pursuing persons with a mobile robot. In *ICCV'99 Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Corfu, Greece.
- [Gabor, 1946] Gabor, D. (1946). Theory of communication. *Institute of Electrical Engineers*, 93(26):429–457.

- [Govindaraju et al., 1989] Govindaraju, V., Sher, D., Srihari, R., and Srhari, S. (1989). Locating human faces in newspaper photographs. In *Proc. of IEEE CVPR*, pages 549–554.
- [Grossmann and Morlet, 1984] Grossmann, A. and Morlet, J. (1984). Decomposition of hardy functions into squared integrable wavelets of constant shape. *SIAM J. Math. Analysis*, 15:723–736.
- [Hager and Belhumeur, 1998] Hager, G. and Belhumeur, P. (1998). Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on PAMI*, 20(10):1025–1039.
- [Isard and Blake, 1998] Isard, M. and Blake, A. (1998). Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28.
- [Jebara and Pentland, 1997] Jebara, T. and Pentland, A. (1997). Parametrized structure from motion for 3d adaptative feedback tracking of faces. In *Proc. of IEEE CVPR*.
- [Kjeldsen and Kender, 1996] Kjeldsen, R. and Kender, J. (1996). Finding skin in color images. In *Proc. of IEEE Face and Gesture Recognition*, pages 312–318, Killington, VT.
- [Kosko, 1992] Kosko, B. (1992). *Neural Networks and Fuzzy Systems*. Prentice Hall.
- [Kruger and Sommer, 2000] Kruger, V. and Sommer, G. (2000). Gabor wavelet networks for object representation. Technical Report 2002, University Of Kiel, Germany.
- [Li, 2000] Li, B. (2000). Human and object tracking and verification in video. Technical Report CS-TR-4140, Center for Automation Research, University of Maryland.
- [Mallat, 1989] Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on PAMI*, 11(7):674–693.
- [Manjunath and Chellapa, 1993] Manjunath, B. and Chellapa, R. (1993). A unified approach to boundary perception: edges, textures, and illusory contours. *IEEE Transactions on Neural Networks*, 4(1):96–107.

- [Maurer and Malsburg, 1996] Maurer, T. and Malsburg, C. (1996). Tracking and learning graphs and pose on image sequence of faces. In *Proc. International Conference on Artificial Neural Networks*, Bochum.
- [McKenna et al., 1996] McKenna, S., Gong, S., and Collins, J. (1996). Face tracking and pose representation. In *British Machine Vision Conference*, Edinburgh.
- [Moghaddam and Pentland, 1994] Moghaddam, B. and Pentland, A. (1994). Face recognition using view-based and modular eigenspaces. In *Automatic Systems for the Identification and Inspection of Humans*, volume 2277. SPIE.
- [Moghaddam and Pentland, 1997] Moghaddam, B. and Pentland, A. (1997). Probabilistic visual learning for object representation. *IEEE Transactions on PAMI*, 19(7):696–710.
- [Morimoto et al., 1998] Morimoto, C., Koons, D., Amir, A., and Flickner, M. (1998). Real-time detection of eyes and faces. In *Workshop on Perceptual User Interfaces*, pages 117–120, San Francisco, California.
- [Olivier et al., 1997] Olivier, N., Pentland, A., and Berard, F. (1997). Lafter: Lips and face real-time tracker. In *Proc. of IEEE CVPR*.
- [Press et al., 1986] Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1986). *Numerical Recipes, The Art of Scientific Computing*. Cambridge University Press.
- [Raja et al., 1998] Raja, Y., McKenna, S., and S.Gong (1998). Tracking and segmenting people in varying lighting conditions using colour. In *Proc. of IEEE Face and Gesture Recognition*, Nara, Japan.
- [Rowley et al., 1998a] Rowley, H., Baluja, S., and Kanade, T. (1998a). Neural network-based face detection. *IEEE Transactions on PAMI*, 20(1):23–38.
- [Rowley et al., 1998b] Rowley, H., Baluja, S., and Kanade, T. (1998b). Rotation invariant neural network-based face detection. In *Proc. of IEEE CVPR*, pages 38–44, Santa Barbara, California.
- [Sakai et al., 1969] Sakai, T., Nagao, M., and Fujibayashi, S. (1969). Line extraction and pattern recognition in a photograph. *Pattern Recognition*, 1:233–248.

- [Silva et al., 1995] Silva, L., Aizawa, K., and Hatori, M. (1995). Detection and tracking of facial features. In *Proc. of SPIE Visual Communications and Image Processing*, Taiwan.
- [Stiefelhagen and J.Yang, 1997] Stiefelhagen, R. and J.Yang (1997). Gaze tracking for multimodal human computer interaction. In *Proc. of ICASSP*, Munich, Germany.
- [Sun et al., 1998] Sun, Q., Huang, W., and Wu, J. (1998). Face detection based on color and local symmetry information. In *Proc. of IEEE Face and Gesture Recognition*, Nara, Japan.
- [Sung and Poggio, 1998] Sung, K. and Poggio, T. (1998). Example-based learning for view-based human face detection. *IEEE Transactions on PAMI*, 20(1):39–50.
- [Szu et al., 1992] Szu, H., Telfer, B., and Kadambe, S. (1992). Neural network adaptive wavelets for signal representation and adaptation. *Optical Engineering*, 31(9):1907–1961.
- [Toyama, 1998a] Toyama, K. (1998a). "look, ma - no hands!" hand-free control with real-time 3d face tracking. In *Workshop on Perceptual User Interfaces*, pages 117–120, San Francisco, California.
- [Toyama, 1998b] Toyama, K. (1998b). Prolegomena for robust face tracking. Technical Report MSR-TR-98-65, Microsoft Corporation, Redmond, WA.
- [Tsotsos, 1990] Tsotsos, J. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sci.*, 13:423–469.
- [Turk, 1996] Turk, M. (1996). Visual interaction with lifelike characters. In *Proc. of IEEE Face and Gesture Recognition*, Killington, VT.
- [Turk, 1998] Turk, M. (1998). Moving from guis to puis. Technical Report MSR-TR-98-69, Microsoft Corporation, Redmond, WA.
- [Turk and Pentland, 1991] Turk, M. and Pentland, A. (1991). Face recognition using eigenfaces. In *Proc. of IEEE ICPR*, pages 586–591.

- [Vetterli and Kovacevic, 1995] Vetterli, M. and Kovacevic, J. (1995). *Wavelets and Sub-band Coding*. Prentice Hall.
- [Wiskott et al., 1997] Wiskott, L., Fellous, J., Krüger, N., and Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on PAMI*, 19(7):775–779.
- [Wu et al., 1999] Wu, H., Chen, Q., and Yachida, M. (1999). Face detection from color images using a fuzzy pattern matching method. *IEEE Transactions on PAMI*, 21(6):557–563.
- [Yang et al., 1997] Yang, J., Lu, W., and Waibel, A. (1997). Skin-color modeling and adaptation. Technical report, CMU-CS-97-146.
- [Yang and Waibel, 1996] Yang, J. and Waibel, A. (1996). A real-time face tracker. In *Proc. of IEEE Workshop on Applications of Computer Vision*, pages 142–147, Sarasota, Florida.
- [Yow and Cipolla, 1997] Yow, K. and Cipolla, R. (1997). Feature-based human face detection. *Image and Vision Computing*, 15(9):713–735.
- [Yuille et al., 1989] Yuille, A., Cohen, D., and Hallinan, P. (1989). Feature extraction from faces using deformable templates. In *Proc. of IEEE CVPR*, pages 104–109.
- [Zarit et al., 1999] Zarit, B., Super, B., and Quek, F. (1999). Comparison of five color models in skin pixel classification. In *ICCV'99 Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Corfu, Greece.
- [Zelinsky and Heinzmann, 1996] Zelinsky, A. and Heinzmann, J. (1996). Real-time visual recognition of facial gestures for human-computer interaction. In *Proc. of IEEE Face and Gesture Recognition*, pages 351–356, Killington, VT.
- [Zhang and Benviste, 1992] Zhang, Q. and Benviste, A. (1992). Wavelet networks. *IEEE Transactions on Neural Networks*, 3(6):889–898.