

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA ELÉTRICA E
COMPUTAÇÃO
DEPARTAMENTO DE COMUNICAÇÕES



Tese de Doutorado

**Um Método de Classificação de Texturas com Rotação
Baseado na Modelagem HMM de Características AM-FM**

Evandro Ottoni Teatini Salles
Orientador: Prof. Dr. Lee Luan Ling

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como requisito parcial para a obtenção do Título de Doutor em Engenharia Elétrica.

Banca Examinadora:

Prof. Dr. Adilson Gonzaga
Prof. Dr. Nelson Delfino d'Ávila Mascarenhas
Prof. Dr. Clésio Luis Tozzi
Prof. Dr. Fábio Violaro
Prof. Dr. João B. T. Yabu-uti

EESC – USP – São Carlos – SP
DC – UFSCar – São Carlos – SP
DCA - UNICAMP
DECOM – UNICAMP
DECOM – UNICAMP

Campinas – SP – Brasil
21 de Fevereiro de 2.001

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

Sa32m Salles, Evandro Ottoni Teatini
Um método de classificação de texturas com rotação baseado na modelagem HMM de características AM-FM / Evandro Ottoni Teatini Salles. --Campinas, SP: [s.n.], 2001.

Orientador: Lee Luan Ling.
Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Processamento de imagens. 2. Reconhecimento de padrões. 3. Wavelets (Matemática). 4. Processo estocástico. 5. Markov, Processos de. I. Lee, Luan Ling II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

RESUMO

A análise de texturas é uma importante ramificação da área de processamento de imagens devido a sua aplicação em diversos problemas industriais. Por exemplo, é possível aplicar as técnicas de análise de texturas para avaliar o processo de controle de qualidade na indústria têxtil, de couro, mármore e granito entre outras. Neste trabalho foi estudado o problema da classificação de texturas com e sem rotação. O banco de imagens utilizado foi construído baseado no álbum de texturas de Phill Brodatz, inicialmente proposto para artistas e *designers* mas que, com o passar do tempo foi aceito, pela comunidade de processamento de imagens, como um padrão na área para classificação e segmentação de texturas. No nosso modelo supomos que uma determinada textura é representada pela evolução de características instantâneas AM-FM pela extração do valor médio sobre o resultado das etapas de demodulação não linear. Para a detecção das características empregou-se o banco de filtros denominado Wavelets de Gabor, também referenciada como Transformada de Wavelets Contínuas. Após a demodulação AM-FM, pode-se interpretar as medidas AM-FM como uma representação robusta de uma textura específica. Para o problema da rotação de imagens, sobre o vetor de características, é aplicada uma técnica para correção da orientação. Segundo a idéia original, selecionou-se a abordagem por HMM para modelar a evolução das características. Os bons resultados alcançados provam o desempenho da técnica e são comparados com outras abordagens na literatura.

Palavras-Chave: Classificação de Texturas, características AM-FM, HMM, Wavelets de Gabor.

ABSTRACT

Texture analysis is an important ramification of the image processing research area because of its application in many industrial problems. For example, it is possible to employ the texture techniques to evaluate the quality control processes related to skin leather, textile and marble industry. In this work, the classification problem of non-rotated and rotated textures was studied. The image data bank used was based on the Phil Brodatz Album, which was initially proposed for artists and designers but, later, accepted by the image processing community as a standard pattern for texture analysis and segmentation. In our model, we suppose that textures can be characterized by the evolution of AM-FM instantaneous features, by means of expected values over the non-linear demodulation procedure result. To detect these features, it was employed a Gabor-Wavelet filter bank, also known as Continuous Wavelet Transform. After demodulation, one can interpret the AM-FM measures as a robust description of a specific texture. For the problem of texture rotation, over the feature vector, it was applied a technique to correct the inherent orientation. For feature vector classification the HMM approach to model these evolution features is used. The experimental results prove the proposed texture analysis is powerful when it is compared to other methods proposed in the literature.

Key Words: Texture Classification, AM-FM characteristics, HMM, Gabor Wavelets.

DEDICATÓRIA

Pela memória de meus pais André e Therezinha, por tudo que proporcionaram a mim;

Pelo amor, dedicação e compreensão de minha esposa Glézia.

AGRADECIMENTOS

Durante os quatro anos que passei em Campinas muitas amizades foram feitas e não poderia aqui deixar de destacar as pessoas que, de uma forma ou de outra, contribuíram para a realização desta tese. São meus sinceros agradecimentos:

ao meu orientador prof. Lee pela orientação e atenção dedicada a todo instante, além da confiança depositada em minha pessoa;

aos colegas do LRPRC e do DECOM com quem mantive grande amizade, aprendendo assim um pouco com cada um deles: Miguel, Nata, Lucila, Hugo, Mirian, Néstor, Tarciano, Ernesto, Flávio, André e Zimmer;

ao colega José Raimundo, pela amizade e pelas valiosas discussões sobre estatística;

a Jussara e Gonzaga, por me acolherem tão bem em Campinas e pelas dicas sobre a cidade e a UNICAMP;

aos meus irmãos Andréa, Angela, Alexandre e Cristiano, pela força nos momentos difíceis;

a meus tios Carmita e Norton, pelo incentivo e apoio incondicionais;

à família Dantas: Expedito, Cremilda, Glimalde, Rallyane e, porque não, Felipinho pelo apoio, suporte e confiança;

aos colegas do Depto. de Engenharia Elétrica – UFES pelo apoio e confiança na realização deste treinamento de doutorado.

Índice Analítico

Índice Analítico	vi
Índice de Figuras	viii
Índice de Tabelas	ix
Lista de Abreviaturas e Símbolos	x
Capítulo 1	11
Introdução	11
1.1) Texturas	11
1.2) Algumas Aplicações para Texturas	12
1.3) A Motivação para a Metodologia Empregada.....	13
1.4) O Objetivo desta Tese.....	15
1.5) Comentários Adicionais	15
1.6) A divisão dos capítulos nesta tese.....	16
Capítulo 2	17
Revisão Bibliográfica	17
2.1) Técnicas de segmentação de Texturas.....	17
2.2) Representação × Classificação.....	18
2.3) Técnicas baseadas em banco de filtros.....	19
2.4) Wavelets	36
2.5) Modelagem por Campos Aleatórios de Markov.....	40
Capítulo 3	41
A Modelagem das Texturas	41
3.1) O Modelo para as Texturas	41
3.2) Análise Tempo-Frequência	42
3.3) A Caracterização AM-FM	45
3.4) Sinais Multicomponentes	47
3.5) A Transformada de Wavelets Bi-dimensional	50
3.6) A Modelagem Utilizada.....	57
3.7) O Problema da Rotação de Texturas	62
Capítulo 4	65
A Modelagem Estatística do Problema	65
4.1) Formulação estatística do problema de classificação:.....	65
4.2) A Abordagem Markoviana.....	67
4.3) Os modelos ocultos de Markov	69
4.4) Considerações a cerca do modelo empregado e a estrutura de treinamento.....	75
Capítulo 5	79
Testes e Resultados	79
5.1) O Ambiente Desenvolvido.....	79
5.2) O Banco de Imagens de Texturas.....	80
5.3) Banco de Filtros de Gabor.....	82

5.3.1) Os parâmetros do HMM.....	82
5.4) Resultados Obtidos.....	83
5.4.1) Texturas em presença de rotação.....	88
Capítulo 6.....	93
Conclusões e Projetos Futuros.....	93
Contribuições da tese.....	93
Referências Bibliográficas	97
Apêndice A3.1	105
O Problema da Representação Única para Frequência Instantânea	105
Apêndice A3.2	108
A Questão da Validade da Modelagem AM-FM	108
Apêndice A3.3	111
Relação entre transformadas e banco de filtros	111
Apêndice A4.1	113
Algoritmo de Decodificação de Viterbi.....	113
Apêndice A4.2	115
Modelo de Misturas	115
Apêndice A4.3	116
Algoritmos EM e Reestimação de Baum-Welch.....	116
Apêndice A5.1	121
Relação dos tipos de Texturas tratadas	121

Índice de Figuras

- Figura 2.1: Representação \times Classificação **Erro! Indicador não definido.**
- Figura 2.2: Diagrama de blocos de um sistema genérico de análise de imagens por banco de filtros..... **Erro! Indicador não definido.**
- Figura 2.3: Texturas modeladas por textons. Os deslocamentos verticais e horizontais, Δx e Δy respectivamente, podem ser fixos ou variáveis, assim como a forma e a orientação de cada $t(x,y)$ **Erro! Indicador não definido.**
- Figura 2.4: Sistema não supervisionado proposto em [Unser&Eden, 89]. O elo de realimentação representa o processo de suavização iterativa..... **Erro! Indicador não definido.**
- Figura 2.5: Árvores de Decomposição e particionamento do plano tempo (t) \times frequência (ω) para um sinal 1D. Em a) Wavelets com 2 níveis de decomposição. b) Wavelet Packets com árvore, de 2 níveis, completa c) Wavelet Packet, de 3 níveis, com poda. **Erro! Indicador não definido.**
- Figura 3.1: Sinal contínuo avaliado por uma janela estreita. Observe que apenas alguns exemplares da janela foram mostrados. **Erro! Indicador não definido.**
- Figura 3.2: (a) Sinal com duas componentes a saber: incremento linear e decremento quadrático na frequência ao longo do tempo (em segundos). Em (b) observa-se seu espectrograma – frequência em hz. **Erro! Indicador não definido.**
- Figura 3.3: Textura e seus respectivos espectros de potência. As áreas mais claras de cada espectro representam pontos de potência elevada. **Erro! Indicador não definido.**
- Figura 3.4: Resposta (real e imaginária) de um filtro de Gabor assumidos em uma grade de 33×33 pontos com origem em (17, 17). **Erro! Indicador não definido.**
- Figura 3.5: Em a) determinação B_0 e B_r para o filtro no domínio da frequência (onde a elipse representa a curva de meia magnitude). Em b) parâmetros de orientação ρ e frequência radial $\omega = u_0$ **Erro! Indicador não definido.**
- Figura 3.6: Disposição espectral dos filtros de Gabor empregados nesta tese. Os eixos vertical e horizontal representam as frequências. Observe que devido à redundância espectral apenas dois quadrantes são ocupados pelos filtros. **Erro! Indicador não definido.**
- Figura 4.1: Dois exemplos de topologias para HMM's. O modelo b) foi utilizado nesta tese. **Erro! Indicador não definido.**
- Figura 5.1: Banco 1. Da esquerda para direita, de cima para baixo: D9, D12, D15, D17, D19, D24, D29, D38, D68, D84, D92, D94 e D112, amostras de 128×128 pixels. **Erro! Indicador não definido.**
- Figura A3.1.1: Espectro de potência: a) de um sinal real, b) de um sinal analítico. **Erro! Indicador não definido.**
- Figura A4.2.1: A figura mostra, na parte superior, um modelo HMM composto de dois estados e suas respectivas misturas são apresentadas nos gráficos da parte inferior. **Erro! Indicador não definido.**
- Figura A5.1: O banco total das 112 texturas, de D1 a D112 varridas da esquerda para direita, de cima para baixo. Cada amostra possui 128×128 pixels. **Erro! Indicador não definido.**

Índice de Tabelas

<i>Tabela 5.1: Teste inicial para o grupo das 13 texturas.</i>	Erro! Indicador não definido.
<i>Tabela 5.2: Configurações e taxa de acertos para classificar as 112 texturas.</i>	Erro! Indicador não definido.
<i>Tabela 5.3: Taxa de acertos para cada classe de textura, em %.</i>	Erro! Indicador não definido.
<i>Tabela 5.4: Matriz de confusão, com taxas em %, para a classificação do grupo das 13 texturas em presença de rotação.</i>	Erro! Indicador não definido.
<i>Tabela 5.5: Experimento empregando-se a técnica invariância à rotação sobre o banco 3 – as 112 texturas.</i>	Erro! Indicador não definido.
<i>Tabela 5.6: Tabela comparativa de alguns experimentos obtidos em relação a outros trabalhos.</i>	Erro! Indicador não definido.

Lista de Abreviaturas e Símbolos

Para facilitar a leitura, muitas abreviaturas foram mantidas no idioma inglês. Vetores estão em Negrito.

\mathfrak{R}^N	Conjunto dos números reais de dimensão $N \geq 1$
AM	Amplitude Modulada
AR	Auto Regressivo
\mathbf{C}^N	Conjunto dos números complexos de dimensão $N \geq 1$
DCT	Transformada de Cosseno Discreta
DFT	Transformada Discreta de Fourier
DHT	Transformada de Hadamard Discreta
dpi	pontos por polegada
DST	Transformada de Seno Discreta
EQM	Erro Quadrático Médio
FIR	Filtro de resposta impulsiva finita
FM	Frequência Modulada
HMM	Cadeia Oculta de Markov
KLT	Transformada de Karhunen-Loève
LLT	Transformações Lineares Localizadas
MAP	Estimador Máximo a Posteriori
nD	Espaço n-dimensional, $n=1, 2, \dots$
PCA	análise da componente principal
pdf	função de densidade de probabilidades (emprego da sigla no idioma inglês)
STFT	Transformada de Fourier de Tempo Curto
TR	Operador traço de uma matriz,
Tr	Transposição de matriz
WP	Wavelets Packets
\mathbf{Z}^+	Conjunto dos números inteiros positivos
\mathbf{Z}^N	Conjunto dos números inteiros de dimensão $N \geq 1$

Capítulo 1

Introdução

1.1) Texturas

Definir texturas não é uma tarefa fácil e ainda hoje não existe consenso de como fazê-lo [Mallat, 98]. Sua interpretação costuma envolver critérios subjetivos, o que dificulta o desejo de expressar formalmente o conhecimento que se tem acerca de um de seus exemplares. Dependendo do tipo de interpretação desejada, pode-se caracterizá-las de acordo com a percepção de conceitos tais como regularidade, granulosidade, repetitividade, homogeneidade, etc. Tais atributos são, em geral, de natureza complexa, dificultando qualquer procedimento de quantificação.

Contudo, sob o aspecto computacional, é necessário realizar medições de tal forma a representar suas diferentes classes. Desta forma, vários modelos foram introduzidos, analisando-se sob o ponto de vista estatístico, estocástico, determinístico, fractal, frequencial, etc., sendo esses, muitas vezes, baseados em estudos psico-visuais. Para se ter uma idéia, o trabalho pioneiro na área deve-se a B. Julesz, nos idos de 1962, que buscava uma análise psicológica sobre a percepção estocástica de texturas, interpretando-as como

campos aleatórios de Markov¹. Interessante notar que o tema das texturas é quase tão antigo quanto a própria área de processamento de imagens.

Desde então, muito se discutiu a respeito do tema e muitos modelos foram propostos. Alguns, devido ao pioneirismo, equivocados (como aconteceu justamente com a proposta de Julesz ao assumir que o ser humano só consegue diferir texturas distintas modeladas por até campos aleatórios de segunda ordem [Pratt, 91]). O que se apreende disto tudo é que não existe um tratamento fechado, definitivo, ainda mais para um tema tão fascinante como o é o estudo das texturas. A julgar pelo volume de documentos contabilizados apenas pelo autor desta tese (passando de 200 unidades, compreendidos entre artigos, teses, relatórios técnicos, etc. publicados desde 1962 até a data de dezembro de 2.000 [Mojsilovic *et. al.*, 00], o último até então), o tema está longe de se esgotar. Sua natureza um pouco fora de controle confere uma riqueza de informações sem par na área de processamento de imagens, acabando por justificar assim as diferentes abordagens ao tema.

1.2) Algumas Aplicações para Texturas

A relevância do tema não é puramente acadêmica, encontrando aplicações no âmbito industrial. Como exposto anteriormente, texturas são estruturas ricas em informações e estão presentes em áreas como se segue:

Controle do tamanho de *pelets*² numa usina de pelletização: neste caso, é muito importante a avaliação do volume das pelotas de minério na fabricação do aço. Uma vez na correia transportadora de uma usina siderúrgica, a aglomeração dos *pelets* produz uma imagem texturizada que pode ser utilizada para inferir o tamanho de cada pelota.

Sistemas de controle de qualidade na indústria têxtil e de processamento de couros: análise de texturas pode ser empregada para avaliar a qualidade de produtos têxteis, detectando um determinado tipo de defeito e inclusive classificando-o. O mesmo acontece para insumos como é o caso de produtos de couro, que possuem um alto comércio na Europa e Estados Unidos.

¹ Uma bom resumo da discussão sob estes aspectos pode ser encontrado em [Pratt, 91] [Salles, 94].

² Pelotas de minérios, e outros compostos, usadas na produção de aço.

Indústria de móveis de madeira e Indústria de Mármore e Granito: uma das dificuldades nestas duas áreas é manter as padronagens dos produtos acabados ou seja, uma vez terminado o móvel ou cortado os ladrilhos, deve-se garantir que todas as partes são de uma mesma padronagem. Este tipo de situação poderia ser avaliada por algum processo automático de avaliação das texturas envolvidas [Conners *et. al.*, 83].

1.3) A Motivação para a Metodologia Empregada

Representação de imagens por meio de banco de filtros gera mapas de características 2D e a partir destes são extraídos momentos estatísticos de diversas ordens. Este procedimento é largamente empregado em texturas, utilizando-se desde as máscaras de Laws até técnicas de Wavelets. Assim, assume-se um vetor multivariado de características, onde cada uma representa a medição proporcionada pelos momentos. Supondo que se esteja trabalhando com um banco de N filtros, o vetor obtido terá dimensão N e a ele está associado uma função de distribuição F . Uma dificuldade neste tipo de análise refere-se à determinação de F uma vez que não se conhece a priori sua forma e seus parâmetros.

Como hipótese simplificadora, usa-se assumir distribuição gaussiana mas isto ainda não diminui a complexidade que aparece ao se estimar seus parâmetros: vetor de médias e matriz de variância/covariância. Tudo depende da quantidade de observações disponíveis para treinamento e da dimensão N do vetor de características. Especificamente para a determinação de matrizes de variância/covariância não singulares, é necessário a estimação de $N \times (N+1)/2$ elementos (graus de liberdade da matriz), exigindo-se pelo menos $N+1$ exemplos de treinamento. Entretanto, em muitos casos, a quantidade de vetores empregada para a estimação fica bem abaixo do mínimo teórico necessário, sendo inclusive impraticável obtê-los.

Algumas soluções possíveis incluem a redução de dimensionalidade ou assumir independência entre as características fazendo a matriz de variância/covariância diagonal, exigindo-se assim a estimação de apenas N de seus elementos. Também é possível reduzir a dimensionalidade dos vetores de características por meio de uma análise das componentes principais (técnica essa muito empregada em classificação não supervisionada) ou através

do uso do discriminante de Fisher, quando se trabalha com análise supervisionada. Entretanto, a redução da dimensão não é garantia que se consiga estimar os parâmetros da distribuição. Assumir matriz diagonal implica em não levar em conta as correlações entre as variáveis do vetor de características, o que contribui para degradar o desempenho do classificador final. Ambas as técnicas, entretanto, se fazem necessárias já que em muitos casos não é possível a estimação adequada do modelo.

Em contrapartida, há muito já se sabe que uma das características mais importantes de um sinal refere-se à informação contida em sua componente de fase [Oppenheim *et. al.*, 97]. Para imagens, a situação não é diferente [Limt, 90]. Contudo, em boa parte dos estudos sobre texturas pouco se explorou este tipo de informação. É possível obter a componente frequencial instantânea (mais adequada para interpretação que a fase), a partir da derivação da fase instantânea [Cohen, 95] e sobre essa realizar as devidas análises. Identificar um sinal pela análise frequencial é uma tarefa comumente empregada por seres humanos, que o fazem na maioria das vezes, sem esforço. Acredita-se inclusive que os sistemas visuais humanos sejam dotados de estruturas especializadas na demodulação de tais características [Havlicek&Bovik, 95]. Assim sendo, a análise aqui presente toma como informação válida as componentes advindas tanto da magnitude do sinal como de sua componente frequencial (componentes AM-FM que serão discutidas no capítulo 3).

O emprego de filtros de Gabor ao estilo Wavelets confere uma boa resolução no plano tempo \times frequência e garante uma maneira eficiente de obter tais características. Uma etapa de demodulação é então necessária para a extração das componentes AM-FM, postas a seguir na forma de vetores de características. Entretanto, não se assume aqui que tais vetores sejam independentes, mesmo porque as etapas de extração não permitem que assim o faça. Ao contrário, uma textura é dita representada pela sequência dos vetores obtidos e uma forma eficiente de modelagem é obtida pelo uso de cadeias ocultas de Markov [Chen&Kundu, 94].

1.4) O Objetivo desta Tese

O objetivo desta tese é a proposição e desenvolvimento de um método capaz de aproveitar, de maneira eficiente, as informações de frequência e amplitude instantâneas localizadas, a fim de representar convenientemente uma classe de textura. Para fazer frente ao problema de estimação estatística apresentado, será utilizada uma aproximação por meio das cadeias ocultas de Markov (HMM) que, fica a meio termo entre supor vetores independentes e a estimação completa do modelo estatístico. Consigo, o uso do HMM traz uma interpretação interessante para a questão de classificação de texturas.

Serão avaliadas situações para texturas em presença ou não de rotação. O estudo da invariância à rotação aparece em diversas áreas do processamento de imagens e seu domínio é muito importante em tarefas de aplicações reais. Os testes serão realizados utilizando-se o álbum fotográfico de Phill Brodatz [Brodatz, 66], considerado um padrão para a área.

1.5) Comentários Adicionais

Alguns termos utilizados na literatura tais como Wavelets, Wavelets Packets, Algoritmos forward e backward, frame, etc. foram mantidos nesta tese no idioma inglês. Optou-se por fazê-lo desta forma pois, como tais termos são largamente empregados, o entendimento do assunto fica mais favorável. Em contrapartida, o termo *tiling*, muito empregado na literatura de análise em múltiplas resoluções, foi traduzido para ladrilhamento, em uma alusão direta do efeito da análise localizada sobre o plano de fase (ou plano tempo \times frequência). Ademais, muitas abreviaturas tais como HMM, EM, MAP, DCT, STFT, etc. foram deixadas também no idioma inglês (apresentadas inicialmente em Abreviaturas) para facilitar a identificação por parte de quem lê este trabalho. Também, agregou-se aos capítulos 3, 4 e 5 apêndices onde são discutidos tópicos relativos a seus desenvolvimentos. O objetivo é tornar a leitura menos pesada, podendo o leitor consultá-los ao final sem interromper a sequência normal da leitura.

1.6) A divisão dos capítulos nesta tese

Segue abaixo, a organização da tese em capítulos e alguns comentários relativos ao desenvolvimento de cada um deles.

Capítulo 2: A Revisão Bibliográfica

Neste capítulo é realizada uma revisão bibliográfica de trabalhos tanto relativos a classificação como segmentação de texturas. Foi dado maior ênfase às abordagens baseadas em banco de filtros, especialmente no que se refere à forma de seleção de seus parâmetros e como são extraídas as características.

Capítulo 3: A Modelagem das Texturas

Neste ponto da tese é introduzida a modelagem AM-FM assumida. São discutidos os temas referentes à extração de características, o algoritmo de demodulação empregado, além de questões relativas à validade do modelo. Outro assunto importante refere-se à forma de isolar ditas características, obtidas através de filtros de Gabor com resolução ao estilo transformada de Wavelets.

Capítulo 4) A Modelagem Estatística do Problema

Uma vez que se dispõe da sequência de vetores de características, este capítulo trata de como os mesmos são tratados a fim de representar uma classe de texturas. Portanto, são discutidos os aspectos relativos à cadeia oculta de Markov (técnica empregada), os algoritmos de implementação e os ajustes necessários para a adequação ao modelo de características.

Capítulo 5) Testes e Resultados

Foi desenvolvido uma bancada de testes com o objetivo de comparar os resultados obtidos com os apresentados na literatura. Desta forma, buscou-se manter a mesma forma de realização dos experimentos a fim de poder realizar as comparações com mais propriedade.

Capítulo 6) Conclusões e Projetos Futuros

Neste capítulo são apresentadas as conclusões e contribuições deste trabalho e propostos novos desenvolvimentos.

Referências Bibliográficas

Apêndices

Capítulo 2

Revisão Bibliográfica

2.1) Técnicas de segmentação de Texturas

Em classificação/segmentação de texturas existem dois paradigmas principais [Randen, 97]: extração de características seguida de agrupamento e modelos de textura voltados para segmentação. No primeiro caso aplica-se um conjunto de operações lineares e não lineares para que determinadas características sejam detectadas enquanto que no segundo busca-se ajustar determinados parâmetros para capturar características relevantes. Em geral, ambos podem se valer de algumas premissas da imagem através de uma abordagem bayesiana.

O segundo paradigma necessita de simplificações para se tornar aplicável. Por simplificações entende-se modelagem estacionária. Quanto mais restrita é a modelagem (invariância à rotação, translação, etc.), mais rígidas são as hipóteses assumidas a fim de se manter a estacionaridade. Mesmo para o primeiro caso, mais voltado para detecção e agrupamento de características, é comum assumir que o processo seja estacionário em sentido amplo, o que nem sempre é verdadeiro. O objetivo deste capítulo é apresentar os trabalhos mais relevantes na área e, especialmente, os que nortearam o desenvolvimento desta tese, e do qual vários outros são derivados. Neste capítulo serão tratados inclusive as

abordagens de segmentação (apesar de não ser o objetivo desta tese) que contribuíram em muito com a evolução do tema, e donde se apreendeu conhecimento para classificação de texturas.

2.2) Representação × Classificação

"One of the key tasks in any signal processing problem is to find an appropriate basis in which to represent the data" Dr. Gene Golub, Stanford University

A frase acima também pode ser aplicada para o caso de reconhecimento de texturas. O desejável é obter um conjunto de componentes (que gerem uma base ou um *frame*) onde as projeções do sinal em cada uma dessas proporcionem características importantes. O problema da representação é realizar uma transformação para que uma vez decomposto o sinal, sua reconstrução seja a mais fiel possível mesmo que determinados coeficientes da transformação sejam descartados. Já para a classificação de padrões, ou mais genericamente reconhecimento de padrões, nem sempre a projeção em ditas componentes é suficiente e nem sempre coeficientes importantes para representação também o são para uma correta classificação. A figura 2.1 abaixo mostra um exemplo onde utilizou-se a técnica de análise de componentes principais³, comumente empregada em problemas de representação, para decompor um sinal contendo 2 classes.

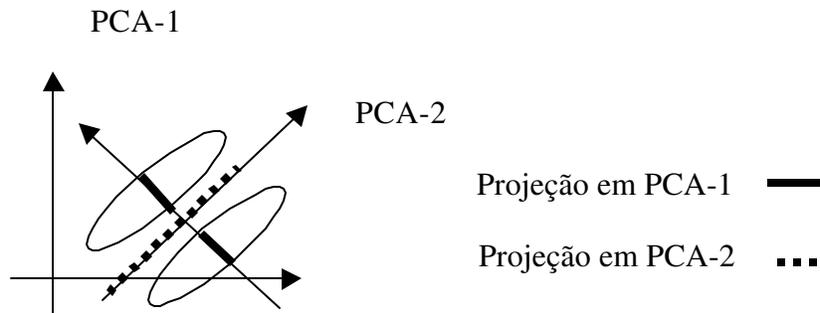


Figura 2.1: Representação × Classificação

Os eixos PCA-1 e PCA-2 representam a projeção do sinal obtida pela técnica. A projeção das classes em PCA-1 proporciona a melhor separabilidade entre classes entretanto, a projeção sobre PCA-2 dá o maior autovalor e conseqüentemente a maior variância/energia

³ A sigla para esta técnica no idioma inglês é PCA: Pincipal Component Analysis.

associada. Assim, para uma etapa de representação o eixo PCA-2 é o mais indicado (maior autovalor associado) entretanto proporcionaria resultados desastrosos para classificação, uma vez que não separa as classes.

Nas abordagens que serão discutidas a seguir, o ponto de partida é a seleção de componentes de decomposição, sendo que em alguns casos opta-se por selecioná-las através da otimização de um índice que mede a discriminação entre classes. Uma vez decomposto o sinal, aplica-se uma sequência de operadores ou assume-se determinados modelos para se efetivar a segmentação. As seções a seguir discutem as diversas técnicas existentes.

2.3) Técnicas baseadas em banco de filtros

A abordagem de segmentação/classificação de texturas segundo análise por bancos de filtros (filtros em multicanais) é muito utilizada no tratamento de texturas [Aach *et. al.*, 95]. Em [Randen&Husøy, 99a] é apresentado um ótimo estudo comparativo de diversas abordagens, com enfoque especial para o uso de banco de filtros no tratamento de texturas. Em especial, neste aspecto, os autores observaram que a estrutura básica de análise está baseada no diagrama abaixo:

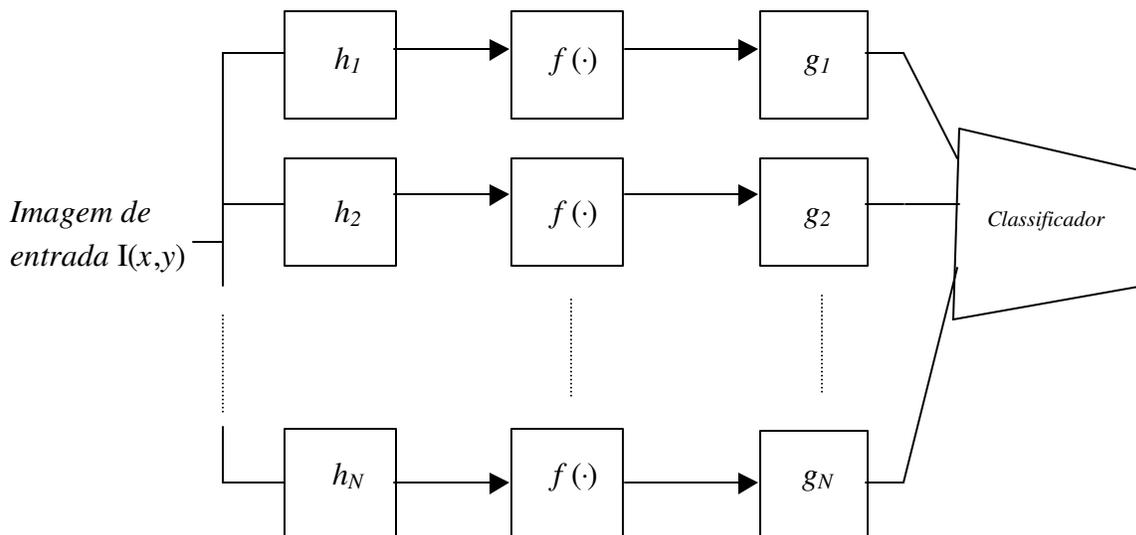


Figura 2.2: Diagrama de blocos de um sistema genérico de análise de imagens por banco de filtros.

De forma genérica, os filtros h_i de entrada são considerados como detectores de características espaciais tais como linhas retas orientadas, pontos, etc, ou como detectores de características espectrais da imagem. É importante observar que é impossível que um determinado filtro seja capaz de apresentar uma resolução arbitrária tanto no domínio espacial como no domínio espaço-frequencial correspondente, devido à dualidade espacial \times frequência espacial, pois não se pode desenvolver uma função que seja bem localizada em ambos os domínios [Randen&Husøy, 99a]. Qualquer escolha de filtros vai implicar em um compromisso, mesmo para os filtros de Gabor, filtros que possuem o menor compromisso tempo \times frequência. A etapa de processamento não linear, representada por $f(\cdot)$, é empregada tanto para ressaltar alguma característica estatística [Unser, 86],[Unser&Eden, 89],[Unser&Eden, 90],[Unser, 93] como para auxiliar na detecção de micro-características presentes na textura. Essas duas maneiras de ver o problema determinam o projeto dos filtros e de como são analisados os dados após a aplicação de $f(\cdot)$.

Apesar das diferentes interpretações de h_i e $f(\cdot)$, a última etapa é consenso entre as diversas abordagens (sobretudo em se tratando de segmentação): não se pode classificar um pixel $p(x,y)$ sem se observar seus vizinhos. Os operadores g_i , denominados de operadores para computação da energia local, propostos inicialmente por Laws [Laws, 80], são utilizados para a avaliação da energia em uma determinada janela centrada em um pixel $p(x,y)$. Isto é importante pois segmentar significa classificar pixel a pixel uma imagem, associando $p(x,y)$ a uma classe C (rotulando $p(x,y)$ com um valor numérico correspondente a C). Este procedimento equivale a decidir pelo valor do pixel de forma contextual, minimizando os erros de classificação. Na saída dos canais obtém-se um conjunto de imagens conhecidas como mapa de características por apresentar, cada uma delas, detalhes específicos e relevantes das texturas analisadas. Outros trabalhos [Unser&Eden, 90], [Unser, 93], sugerem etapas mais elaboradas tanto para os operadores $f(\cdot)$ como para g_i , entretanto, a estrutura básica é aquela apresentada na figura 2.2. Para os filtros h_i pode-se ter as seguintes opções de bancos de filtros:

- filtros fixos detectores de micro-características,
- filtros de Gabor,

filtros otimizados (em geral para o caso de 2 texturas),

Transformada de Wavelets, Wavelets de Gabor.

Já para os operadores não lineares é comum encontrar o emprego de:

operador módulo, $|\cdot|^\alpha$,

operador tangente hiperbólica $\tanh(\cdot)$.

E para a etapa de pós-filtragem g_i utiliza-se uma janela isotrópica do tipo gaussiana.

2.3.1) Filtros Fixos Detectores de Micro-características

Laws [Laws, 80] empregou um conjunto de máscaras, aplicadas sob a forma de banco de filtros, a fim de medir a energia em cada um dos canais de filtragem. As máscaras propostas são filtros FIR 2D especializados na detecção de determinadas características denominadas microtexturas tais como linhas retas orientadas, pontos, etc.. Assim, na saída de cada um destes filtros, ditas características encontram-se realçadas. Com o objetivo de utilizar este tipo de informação, Laws sugeriu uma metodologia de medição de energia, baseada no desvio padrão, que se comporta como um operador não linear. Portanto, ao final, dado N canais distintos, obtém-se um mapeamento não linear de uma imagem de dimensões $M \times M$ em N imagens $M \times M$.

Em [Unser, 86], Unser introduziu e desenvolveu a técnica a qual denominou Transformações Lineares Localizadas (do inglês LLT – *Local Linear Transforms*). A abordagem consiste em aplicar transformações lineares sobre regiões da textura, seguidas de operações não lineares⁴. Entretanto, ao contrário do proposto por Laws, a motivação para sua utilização baseou-se na análise estatística das texturas e não em conjecturas heurísticas. A transformação é efetivada por meio de um banco de filtros FIR 2D.

Posteriormente, em [Unser&Eden, 89] e [Unser&Eden, 90] o tema do pré-tratamento das texturas foi aprofundado com a avaliação de vários operadores não lineares a fim de garantir uma boa etapa de segmentação. Entretanto, de forma genérica, a abordagem segue a estrutura apresentada na figura 2.2 que, a partir deste ponto, será

⁴Na verdade, o ponto de partida para a proposta de Unser foi a técnica de Laws.

assumida como uma representação do trabalho apresentado pelos autores. Na figura 2.2 pode-se observar o sistema proposto onde, a partir de uma sub-imagem, gera-se um conjunto de canais pela sequência de operadores lineares/não lineares. Na saída de cada canal obtém-se um mapa de características.

O ponto chave do trabalho em questão é poder gerar um vetor de características para cada pixel $p(x,y)$ pertencente à imagem e associá-lo a uma dada textura, segundo uma abordagem supervisionada. O vetor pode ser composto de momentos estatísticos tais como média, variância, assimetria e curtosis, obtidos a partir da pdf do processo. Entretanto, a estimação da pdf não é viável por razões práticas e extrair ditos momentos do histograma da sub-imagem não tem muito sentido, além de ser pouco eficaz⁵. Assim, a abordagem proposta foi de aproximar a pdf de ordem N por N pdf's de ordem 1 (histogramas) através da aplicação de uma determinada transformação. Supõe-se que cada sub-imagem é a realização de um campo aleatório homogêneo e ergódico. Nesta situação, uma transformação do tipo Karhunen-Loève (KLT) diagonaliza a matriz de variância/covariância correspondente, descorrelacionando as variáveis do processo. Caso também se assuma que dito processo é gaussiano, o efeito obtido pela aplicação da KLT pode ser escrito segundo a equação abaixo:

$$f_x(x_1, x_2, \dots, x_N) = \prod_{i=1}^N f_x(x_i), \quad (2.1)$$

ou seja, a transformação torna independentes as variáveis da função de densidade de probabilidade $f_x(x_1, x_2, \dots, x_N)$. A partir de cada pdf marginal, pode-se extrair momentos estatísticos e montar um vetor de características para, posteriormente, classificar $p(x,y)$.

Aplicar a transformada de Karhunen-Loève exige a extração de autovalores e autovetores da matriz de variância/covariância do processo. Porém, na estimação desta matriz, a quantidade de vetores de observação necessários cresce com o aumento das dimensões dos mesmos. Para manter uma relação aceitável de custo/benefício, usa-se trabalhar com regiões de pequenas dimensões, o que limita este procedimento à detecção de

⁵ Extrai-se estatísticas que dizem respeito apenas ao nível de cinza do pixel em questão sem levar em conta a interrelação com seus vizinhos.

microtexturas. Cabe lembrar que a suposição de uma sub-imagem, de dimensões $M \times M$ é descrita a priori por uma pdf de ordem $N = M \times M$. Devido à estacionaridade assumida, em [Unser, 86] preferiu-se utilizar transformações ortogonais tais como a DFT, DCT, DST e DHT, referenciadas como “sub-ótimas”. Ditas transformações possuem implementação com grau de complexidade aceitável e não são dependentes dos dados/texturas, como é o caso da KLT. Em especial a DCT apresenta resultados muito próximos a aqueles obtidos pela aplicação da KLT quando se supõe um processo autoregressivo de ordem 1 com um índice de correlação ρ próximo de 1 ($\rho \rightarrow 1$).

Tais transformações sobre x , cada uma representada por um banco de filtros (vide Apêndice A3.3), geram uma nova variável gaussiana y . Observando-se a figura 2.2, antes da montagem do vetor de características é aplicado um operador não linear de tal forma que se tenha uma $u = |y|^\alpha$, para $\alpha=1$ e $\alpha=2$. O objetivo desta transformação é aumentar a separação entre classes, o que pode ser observado pela equação abaixo (supondo $E\{y\} = 0$ e $E\{y^2\} = \sigma^2$):

$$\mu_\alpha = E\{|z|^\alpha\} = \int_{-\infty}^{\infty} |z|^\alpha p(z) dz, \text{ para } z = \frac{y}{\sigma}, \quad (2.2)$$

De acordo com (2.2) e pela propriedade de $E\{|y|^\alpha\}$ [Papoulis, 84], pode-se escrever:

$$E\{u\} = E\{|y|^\alpha\} = \mu_\alpha \sigma^\alpha \quad e \quad (2.3.a)$$

$$E\{u^2\} = \gamma(\alpha) E\{u\} \quad . \quad (2.3.b)$$

O fato de que diferenças na dispersão de uma determinada classe são mapeadas em seu valor médio contribui para o emprego de métodos de agrupamentos, caso se deseje atacar o problema utilizando uma abordagem não supervisionada.

2.3.2 Filtros de Gabor

O uso dos filtros de Gabor em texturas foi proposto devido a sua semelhança com algumas estruturas encontradas no sistema visual humano [Malik&Perona, 85]. Em [Daugman, 88], Daugman estudou as similaridades entre as respostas impulsivas de um filtro de Gabor e de

estruturas corticais denominadas campos receptivos. Com base nesse estudo, foi introduzida uma rede neural supervisionada, apresentando camadas de pesos fixos (funções de Gabor) e outras de coeficientes ajustáveis. Sua preocupação era determinar os parâmetros que representassem uma dada imagem como uma combinação linear de funções elementares de Gabor, segundo se observa na equação (2.4):

$$I'(x, y) = \sum_{i=1}^N \alpha_i h_i(x, y), \quad (2.4)$$

onde:

$I'(x,y)$ é a aproximação de uma imagem $I(x,y)$,

α_i são os coeficientes da combinação linear,

$h_i(x,y)$ são as funções de Gabor.

Os filtros de Gabor são obtidos fazendo-se um envelope gaussiano modular uma exponencial complexa (como será visto no capítulo 3):

$$h(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \times \exp[j2\pi(Ux + Vy)], \quad (2.4.a)$$

para:

$$j = \sqrt{-1},$$

(x,y) representa o par das coordenadas da imagem no domínio espacial,

$(u=U, v=V)$ representa as frequências ao qual o filtro está centrado,

(σ_x, σ_y) determina o espalhamento espacial do envelope gaussiano.

A função de custo escolhida a ser minimizada foi o erro quadrático médio (EQM). Este tipo de representação garante que o conjunto de funções de Gabor seja completa mas não ortogonal (redundante), o que aumenta a carga computacional. Baseando-se em teorias propostas sobre o sistema visual humano, análise em canais múltiplos, apresentando resoluções distintas, Jain *et. al.* [Jain&Farrokhnia, 91] desenvolveram um sistema de segmentação de texturas não supervisionado empregando filtros de Gabor. A estrutura proposta segue o diagrama apresentado na figura 2.2, onde h_i representa tais filtros, $f(\cdot)$ representa um operador não linear e g_i é uma função gaussiana, para $i \in \mathbf{Z}^+$. h_i é parametrizado de tal forma que o conjunto de todos os filtros proporciona um

particionamento no plano de fase (plano tempo \times frequência) muito semelhante ao obtido pela transformada de Wavelets. Assim sendo, h_i é uma família de filtros orientados e dispostos de maneira que suas frequências centrais evoluam em oitavas no domínio das frequências espaciais. Neste caso, convencionou-se denominar tal estrutura como Wavelet de Gabor. No caso das Wavelets ortogonais diz-se que os coeficientes dos filtros geram uma base ortogonal para o espaço do sinal de entrada. Já para as Wavelets de Gabor diz-se que seus coeficientes geram um *frame* para o espaço de sinal. Com o objetivo de avaliar características na imagem que estejam em distintas rotações, Jain propôs que os filtros, em uma frequência radial específica, estejam orientados em ângulos de 0° , 45° , 90° e 135° .

A presença de operadores não lineares tem por finalidade detectar as estruturas básicas que constituem uma textura. Acredita-se que essas estruturas, definidas por Julesz *et. al.* [Julesz&Bergen, 83] como *blobs*, sejam similares àquelas detectadas pelo sistema visual humano ao avaliar uma imagem. Porém, não foi proposto e ainda não existe consenso em como detectá-las. O trabalho de Jain tornou-se relevante por apresentar uma abordagem propondo tais operadores, especificamente a tangente hiperbólica, como detector de *blobs*. É interessante notar que Unser, em [Unser, 86], também utilizou funções não lineares em cada um dos canais de detecção mas com o objetivo de extrair características estatísticas, mapeando a variância de um canal em um valor médio.

Para reduzir a carga computacional envolvida, em [Jain&Farrokhnia, 91] empregou-se uma metodologia de seleção dos filtros baseando-se no erro quadrático médio, avaliando-se a resposta de cada um e selecionando-se aqueles que proporcionem o EQM desejado. Ou seja, a escolha dos filtros foi obtida segundo um critério de representação das texturas e não discriminação entre as mesmas, critério este justificável devido ao caráter não supervisionado do problema. Após a obtenção dos mapas de características, utilizou-se uma técnica de análise de agrupamentos denominada CLUSTER para a associação de um pixel $p(x,y)$ a um agrupamento A_C . Novamente, $p(x,y)$ deve assumir como nível de cinza um valor numérico correspondente a A_C .

A presente abordagem tenta tornar o sistema robusto à rotação empregando discretização na orientação dos filtros. Entretanto, o problema ainda persiste já que não se consegue varrer todas as direções possíveis. O custo computacional é elevado pois a quantidade de filtros é elevada e não é possível aproveitar simetrias de construção dado que os mesmos são não ortogonais. Finalmente, a restrição de não supervisão dificulta a seleção de filtros que melhor discriminem as texturas, sendo necessário optar por um solução onde a representação seja a mais esparsa possível, o que nem sempre é alcançado.

Em [Dunn *et. al.*, 94], Dunn *et. al.* propuseram um modelo de texturas, voltado para segmentação, também empregando os filtros de Gabor. O modelo sugerido segue a teoria dos *textons* desenvolvida em [Julesz&Bergen, 83]. Assim, criou-se uma metodologia para estudar, de forma analítica, o problema de segmentação de duas texturas distintas. Assume-se que *textons* são os elementos básicos que compõem uma determinada textura. Cada um deles se caracteriza por apresentar tamanho, forma e orientação e são descritos por uma função $t(x,y)$, sendo possível trabalhar tanto com texturas determinísticas (uniformes) como não determinísticas (não uniformes). A técnica de segmentação empregada se caracteriza por ser supervisionada. A motivação principal para a modelagem apresentada está no desenvolvimento de filtros que produzam descontinuidades na região de separação entre duas texturas. O objetivo é realçar ao máximo a diferenciação entre as classes, de forma a obter um perfil de resposta que se aproxime ao caso degrau unitário. Entretanto, para casos mais realísticos, é possível observar quatro tipos de descontinuidades: degrau – sigmoidal, vale, crista e mudança da média local de um patamar. A figura 2.3 apresenta uma imagem I composta de duas texturas uniformes distintas.

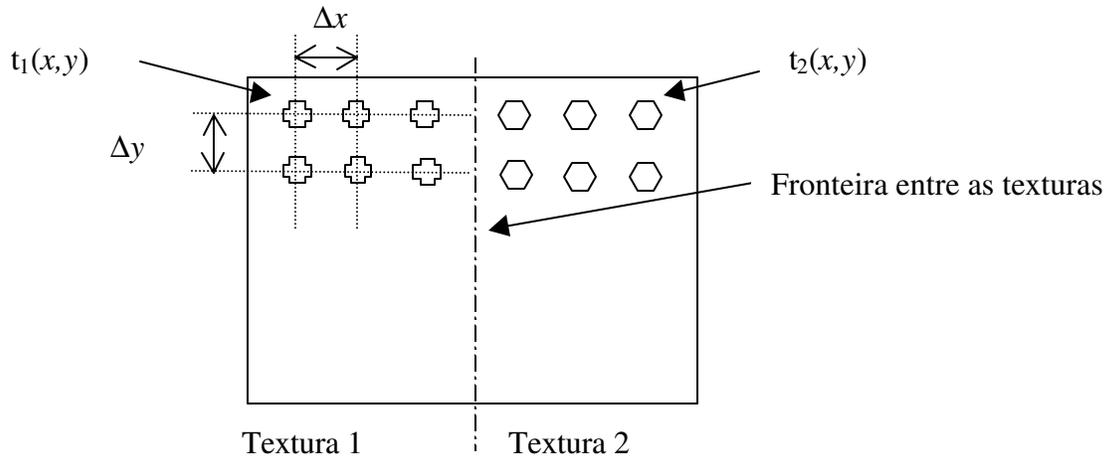


Figura 2.3: Texturas modeladas por *textons*. Os deslocamentos verticais e horizontais, Δx e Δy respectivamente, podem ser fixos ou variáveis, assim como a forma e a orientação de cada $t(x,y)$.

No caso uma textura uniforme T , de dimensões dadas por um conjunto suporte S_T , é possível caracterizá-la por uma função real e determinística $t(x,y)$ que se repete regularmente em S_T . Distribuição regular, neste caso, significa que os espaçamentos horizontal e vertical (Δx e Δy respectivamente) entre os $t(x,y)$ são fixos, conforme se observa na figura 2.3. Assume-se que $t(x,y)$ possui transformada de Fourier $T(u,v)$.

Assim, uma determinada imagem I , composta de duas texturas dispostas de forma adjacente, pode ser segmentada desde que o filtro selecionado esteja sintonizado em uma delas. Esta suposição é possível pois a modelagem homogênea proporciona espectros de potência com picos estreitos e centrados nas frequências (U,V) de repetição dos *textons*. Dunn empregou uma metodologia para determinar o par $(u=U,v=V)$ de T1 de forma que em uma imagem contendo T1 e T2, a resposta para T1 seja máxima em comparação com a resposta de T2. Uma vez determinado (U,V) , o filtro de Gabor é desenhado. As respostas do tipo degrau-sigmóide, vale e crista são típicos de textura homogênea. Uma resposta do tipo degrau indica que o filtro foi bem projetado, apresentando uma boa distinção entre classes e que as texturas são bem comportadas. Já a resposta em vale/crista ocorre quando duas texturas do mesmo tipo mas com fases distintas são dispostas lado a lado. Isto implica em uma resposta tipo vale/crista, para a modelagem apresentada, na região entre classes.

Para texturas não uniformes, o modelo acima deixa de valer. Neste caso, as respostas se aproximam do tipo degrau – sigmóide, apresentando certas variações de amplitude, sob a forma de ondulações, nos patamares da função degrau. Isto se deve à variação aleatória da textura em sua extensão. Ou seja, os *textons* deixam de ter tamanho, localização e orientação descritos de forma determinística. Para minimizar os problemas decorrentes desta situação empregou-se pós-filtragem gaussiana g_i onde a variação local em cada região é traduzida como um valor médio a fim de se ter uma resposta apropriada para segmentação. Entretanto, a pós-filtragem apresenta como efeito colateral uma degradação das respostas nas fronteiras entre texturas, tornando-as difusas e lentas.

O estudo do modelo sobre regiões não determinísticas produziu resultados interessantes. Em [Dunn&Higgins, 95][Weldon&Higgins, 96] avalia-se de forma mais rigorosa o problema de texturas não homogêneas, sendo proposto um novo modelo onde os *textons* são descritos como uma perturbação sobre a forma básica determinística $t(x,y)$. Esta perturbação foi aplicada aos atributos tamanho, localização e rotação, e $t(x,y)$ em uma outra, porém de caráter estocástico. Ou seja, os *textons* não são mais presos a uma rígida lei de formação. A imagem passa a ser definida como um conjunto de regiões disjuntas, cada uma contendo um *texton*, e descritas segundo uma variável aleatória. Como resultado desta abordagem, na saída do operador não linear $|\cdot|$, o processo é descrito pela pdf de Rice.

Este fato tem implicações muito interessantes em segmentação. Caso o filtro fosse aplicado a uma textura que não aquela utilizada no seu desenvolvimento, a pdf de saída obtida passa a ser dada pela equação de Rayleigh (e não a de Rice)! É interessante notar que a pdf de Rice descreve situações onde existe uma componente de forte energia somada a outras tantas, aleatórias, de menor energia. Já no caso de Rayleigh, as diversas componentes aleatórias contribuem na pdf final mas sem que haja uma componente significativa frente às demais. Então, dado o problema inicial de duas texturas e dado um filtro sintonizado para uma delas, a imagem final obtida terá duas regiões a saber: uma descrita pela pdf de Rayleigh e outra descrita pela pdf de Rice⁶. A classificação de cada pixel da imagem final é realizada a partir da comparação de um valor de limiar. Dado que

se dispõe das pdf's das regiões, encontrar o limiar se restringe a estimar os parâmetros de cada pdf, seguido da aplicação de um classificador bayesiano. Entretanto, os resultados obtidos dependem de um bom casamento entre o modelo proposto e as texturas a serem tratadas. Os autores apresentaram situações onde o modelo acima não vale, o que implicou na deterioração dos resultados de segmentação.

Em [Haley&Manjunath, 98] é abordado o problema de texturas em distintas orientações através de características extraídas por filtros de Gabor modificados, evitando-se o uso de pré-processamento para a correção das variações de brilho e contraste de distintas amostras de texturas. O uso de filtros analíticos possibilitou a extração da amplitude, frequência e fase associadas ao sinal de banda estreita. De cada uma das componentes acima, extrai-se características de envelope e direcionalidade. Extraiu-se seis mapas 2D de características, cada um deles corrigidos para tratar texturas em presença de rotação. A esta primeira etapa denominou-se de micromodelo. A abordagem segue obtendo-se momentos sobre tais mapas e a partir deste ponto o tratamento, denominado de macromodelo, supõe um modelo gaussiano multivariado não markoviano para cada classe de textura. No capítulo 5 serão discutidos os resultados desta técnica em relação à metodologia proposta nesta tese.

2.3.3) Filtro Ótimos

Uma das possíveis aplicações de segmentação de texturas refere-se à detecção de defeitos. Neste caso, assume-se dois tipos distintos de classes: não defeituosa e defeituosa. Para solucionar este tipo de problema é possível propor uma abordagem onde a extração de características é baseada em filtros ótimos. Estes filtros são desenhados de tal maneira que dado duas texturas dispostas como na figura 2.3, compondo a imagem de entrada $I(x,y)$, o objetivo é projetar um filtro de tal forma que as classes da imagem processada $O(x,y)$ sejam representadas por conjuntos disjuntos, formando uma partição do espaço de saída. Em termos da imagem $O(x,y)$, deseja-se que duas classes distintas sejam caracterizadas por

⁶ Estes dois tipos de distribuição, Rice e Rayleigh, são utilizadas para descrever efeitos de desvanecimento em sistemas de telecomunicações.

níveis de cinza distintos. Para tanto, usa-se um índice⁷ J que maximize a discriminação entre as classes, sendo J função dos parâmetros do filtro em questão. Como o objetivo é separação de classes/agrupamentos, J expressa o resultado em função de medidas estatísticas tais como média e variância, seguindo a abordagem do discriminante linear de Fisher.

Unser [Unser, 86] também tratou do problema de duas classes, propondo um índice que maximizasse a separação entre classes levando-se em conta as variâncias de cada uma delas. A equação do índice proposto é dada abaixo:

$$J_U(h) = \sum_{i=1}^N \frac{(\sigma_{2i}^2 - \sigma_{1i}^2)^2}{\sigma_{2i}^2 \sigma_{1i}^2}, \quad (2.5)$$

onde σ_{2i} e σ_{1i} representam as variâncias das classes 2 e 1 respectivamente, medidas logo após a aplicação do filtro h e o sub-índice i indexa o canal correspondente. Observe que mesmo para o problema de duas texturas, pode-se dispor de $N \geq 2$ canais de filtragem. Após uma transformação não linear é necessário a substituição das variâncias de cada canal pelas médias de cada canal. Em realidade, o que se tem é um índice que separa as classes pela escolha de componentes de decomposição no espaço do sinal onde a projeção da diferença entre as médias esteja maximizada. Assim, são justificáveis operações não lineares e filtragens iterativas, como propostas em [Unser&Eden, 89] [Unser&Eden, 90], para reduzir a variância dentro de cada classe. Pelos mesmos motivos expressos anteriormente, os autores optaram pelo uso de transformações sub-ótimas como a DCT, DST, etc..

Mahalanobis *et. al.* [Mahalanobis&Singh, 94] também propuseram uma função de custo para segmentar duas texturas. O ponto de partida é supor a existência de um filtro h que, ao ser correlacionado com a imagem de entrada $I(x,y)$, apresente na sua saída uma imagem contrastada. A expressão da função, parametrizada pelo filtro, é dada por:

$$J_{MS}(h) = \frac{E_I}{E_O}, \quad (2.6)$$

onde E_I e E_O representam respectivamente as energias da imagem de entrada de $I(x,y)$ e da imagem de saída $O(x,y)$. Como em [Unser, 86], este tipo de abordagem possui solução

⁷ A notação referente aos índices foi absorvida de [Etemad&Chellappa, 98].

analítica, obtida pela maximização do índice. O filtro é desenvolvido de tal modo que a energia em sua saída seja máxima para apenas uma determinada textura.

Em [Randen&Rusøy, 99b] [Randen, 97] foi proposto a utilização do discriminante de Fisher para a separação de duas classes. O índice empregado pode ser visto em (2.7).

$$J_F(h) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (2.7)$$

onde:

(μ_1, μ_2) representam as médias de cada classe;

(σ_1, σ_2) representam as variâncias de cada classe.

Na maximização de (2.7), foi necessário introduzir determinadas simplificações a fim de se obter uma solução analítica fechada. Para tanto, os autores assumiram que as texturas pudessem ser separáveis e caracterizadas por processos AR de ordem 1, o que é uma aproximação pobre.

Aqui também caberia como filtragem ótima de texturas a abordagem exposta em [Dunn&Higgins, 95]. O filtro de Gabor é sintonizado na frequência que descreve a repetição dos *textons*, além de ser proposto uma metodologia de se selecionar o limiar de decisão entre duas texturas, baseando-se na estimação da pdf de saída do valor absoluto da imagem filtrada.

Nas abordagens acima comentadas, filtragem ótima se refere à determinação de parâmetros baseando-se na modelagem estacionária da textura, o que nem sempre ocorre em realidade. Todas as técnicas possuem uma solução analítica para o caso de duas classes mas a introdução de mais classes dificulta a obtenção dos filtros através de equações fechadas. Para os índices J_U , J_{MS} e J_F o processo de maximização, obtido por derivação em relação aos parâmetros do filtro h exige a estimação de matrizes de correlação o que restringe o uso da abordagem a padrões do tipo microtexturas. Isto decorre da necessidade de se manter tais matrizes com dimensões não muito grandes tanto para se obter matrizes não singulares (problema de estimação) como para evitar carga computacional elevada. A

questão de matrizes não singulares é importante pois a equação final dos parâmetros de h , para os três índices citados, é dada pela solução de um problema de autovalores.

Em [Unser&Eden, 89], cada canal foi modificado para a disposição, apresentada pela figura 2.4 abaixo:

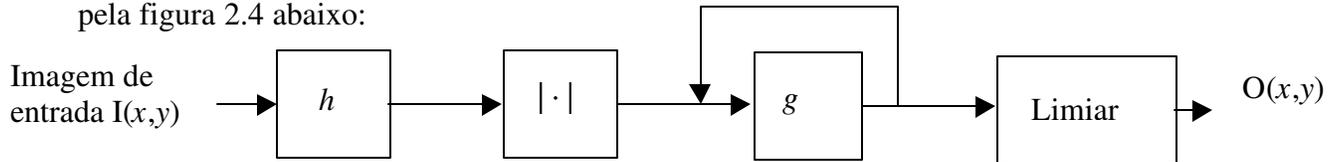


Figura 2.4: Sistema não supervisionado proposto em [Unser&Eden, 89]. O elo de realimentação representa o processo de suavização iterativa.

A introdução do filtro passa baixa de suavização iterativa, representado graficamente pela realimentação, tem por objetivo diminuir a variância dentro da classe. O filtro é dado por uma janela gaussiana que, a cada passo de iteração, realiza uma progressão de meia escala. Isto garante uma redução da variância por um fator de dois, passo a passo. A consequente redução da variância contribui para melhorar os resultados de classificação e, em especial para o problema de duas texturas, permite o desenvolvimento de um filtro ótimo mesmo para uma técnica não supervisionada. Para tanto, duas hipóteses foram testadas:

- a matriz de espalhamento entre classes é praticamente inalterada durante distintas resoluções, dado que esta é bastante afetada pela transformação não linear;
- grau de discriminação entre classes não varia de forma significativa de um passo de suavização para outro.

Estas duas suposições, confirmadas através de um teorema apresentado em [Unser&Eden, 89], permitiram que com apenas as matrizes de espalhamento total, obtidas em resoluções distintas, fosse possível encontrar uma transformação T , implementada por meio do banco de filtros. Finalmente, para realizar a segmentação da imagem, foi proposta uma técnica de seleção de limiar de decisão entre duas texturas de tal modo a maximizar a separação entre classes. Entretanto, um dos problemas que aparece com a pós-filtragem passa-baixas iterativa é a suavização em demasia das fronteiras entre texturas, o que dificulta justamente um processo de segmentação.

Em [Unser&Eden, 90], são avaliados outros operadores não lineares e suas consequências estatísticas sobre o problema da segmentação de texturas. Entretanto, a análise é realizada sobre o problema restrito de duas texturas, onde é possível supor determinadas condições que são difíceis de impor para o caso de múltiplas texturas.

2.3.4) A Modelagem AM–FM

Em uma sequência de artigos [Bovik *et. al.*, 90][Clark&Bovik, 89][Bovik, 91], Bovik *et. al.* sugerem que uma determinada textura $t(x,y)$ seja caracterizada por um sinal bidimensional limitado em banda, descrita pela equação (2.8):

$$t(x, y) = a(x, y) \cos[2\pi(Ux + Vy) + p(x, y)], \quad (2.8)$$

onde:

U e V são as frequências em torno das quais se concentra a banda,
 $a(x,y)$ representa o envelope de amplitude do sinal,
 $p(x,y)$ correspondente à informação contida na componente de fase.

Na verdade, a premissa básica deste modelo é supor que as informações de baixa frequência estão codificadas em $a(x,y)$ e as correspondentes de alta frequência apareçam na fase. Em [Bovik *et. al.*, 90], utilizou-se um banco de filtros de Gabor em distintas orientações e frequências radiais para a extração de características. Assumir (2.8) garante a extração das informações de amplitude, frequência e fase. Para a obtenção da frequência, utilizou-se o cálculo do gradiente da fase a fim de se obter uma indicação da direção de maior variação de frequência. Entretanto, (2.8) possui inconvenientes a serem discutidos no capítulo 3.

O modelo empregado segue a abordagem dada pela figura 2.2. Cada um dos filtros foi selecionado de tal forma a detectar picos no espectro de uma imagem exemplo $t(x,y)$. A seleção é feita através de um algoritmo de detecção de picos do espectro de potências $|T(u,v)|^2$. Supondo-se que existam N tipos distintos de $t(x,y)$, uma etapa de segmentação de texturas é levada a cabo através da regra de decisão:

$$(x,y) \in R_n \quad \text{se } n = \arg\{\max\{m_i(x,y)\}\}, \quad i \in \{1, \dots, N\}, \quad (2.9)$$

onde R_n representa a n -ésima região de uma imagem contendo diversas texturas e $m_i(x,y)$ é a amplitude resultante na saída do i -ésimo filtro $h_i(x,y)$ centrado nas frequências (U_i, V_i) e parâmetros σ_x, σ_y , conforme equação (2.4.a).

Entretanto, algumas classes de texturas não podem ser representadas unicamente por um sinal de banda estreita. Neste caso, o envelope obtido apresenta uma variação considerável em sua magnitude, o que inviabiliza o uso de (2.9). Para solucionar este problema, uma etapa de pós-filtragem g é introduzida a fim de se “suavizar” as ondulações, tornando as regiões mais homogêneas. Ou seja, a abordagem segue o esquema da figura 2.2. A regra de decisão é reescrita por:

$$(x,y) \in R_n \text{ se } n = \arg\{\max[g(x,y) * m_i(x,y)]\}, \quad i \in \{1, \dots, N\}, \quad (2.10)$$

onde $*$ representa a operação de convolução bidimensional. Em [Clark&Bovik, 89], estudos indicam que a demodulação da fase em imagens contendo distintas texturas mostrou-se útil para a determinação de fronteiras de segmentação. Em [Bovik, 91], o modelo (2.8) é acrescido de imperfeições e avalia-se sua validade.

Dando sequência à abordagem pela modelagem local, nos artigos [Bovik *et. al.*, 92][Havlicek&Bovik, 95][Havlicek *et. al.*, 00] são apresentados os conceitos de frequência emergente, coerência localizada e demodulação de características denominadas AM-FM a partir da aproximação por quase autofunções. A hipótese de que uma textura é descrita por processos de banda estreita continua valendo, sendo necessário o uso de uma transformação que possa detectar amplitude e frequência instantâneas. Neste caso, utiliza-se a transformação de Wavelet de Gabor, por sua propriedade de localização no domínio espacial \times frequência espacial. Como será visto no capítulo 3, tal abordagem é robusta, possuindo inclusive um algoritmo proposto. Nesta tese, algumas modificações serão propostas a fim de adaptar tal técnica às necessidades de classificação de texturas em presença de rotação.

2.3.5) A Utilização de Cadeias Ocultas de Markov

Em [Chen&Kundu, 95], Chen *et. al.* propuseram um sistema não supervisionado de segmentação de texturas, apresentando robustez à rotação. O procedimento de extração de características é baseado nas máscaras de Laws – micromáscaras (banco de filtros) como sabido, especializadas em primitivas da textura tais como linhas, pontos e curvas. Uma vez obtido mapas de características, mede-se a energia associada aos pixels através de operadores direcionais - macromáscaras, perfazendo um total de 16 (dezesseis) direções distintas. A partir do mapa final de características, teve lugar o processo de segmentação propriamente dito, sendo dividido em duas etapas. A primeira, denominada de segmentação grosseira, consiste em dividir a imagem em blocos de 16×16 pixels, extraindo-se características de cada um deles para um posterior treinamento de uma cadeia oculta de Markov. Esta etapa realiza uma sobre-segmentação gerando uma quantidade de classes acima daquelas originalmente presentes. Também utilizou-se um critério de medida de similaridade entre HMM para reduzir o efeito da segmentação excessiva, unindo-se⁸ os blocos cujos HMM sejam similares.

A segunda etapa, denominada de segmentação fina, foi proposta para corrigir os problemas da etapa anterior levando-se em conta a informação contextual de cada pixel. Para tanto, aplicou-se o método do quadrante na suavização da verossimilhança do espaço de características que, originalmente, foi introduzido para eliminar ruído em imagens sem porém degradar informação de fronteiras e bordas [Hisao&Sawchuk, 89a] [Hisao&Sawchuk, 89b]. Este problema não é tão simples já que tanto ruído como fronteiras caracterizam-se espectralmente por apresentarem componentes de alta frequência. Assim, um processo iterativo de treinamento dos HMM's foi realizado, unindo-se os modelos semelhantes até que não fosse mais possível fazê-lo.

Para o treinamento das cadeias ocultas, optou-se pelo procedimento do K-means Segmental e uma gaussiana por estado. No que tange à robustez à rotação os autores argumentam que ao empregar HMM não treinados previamente garante por si só robustez à rotação. Entretanto, os testes não são conclusivos já que não se compara a segmentação de

⁸ Na verdade, usava-se o modelo que apresentasse a menor distância entre os possíveis para um rótulo de pixel.

texturas idênticas apresentadas porém em rotações distintas. Ademais, como os HMM's não são pré-treinados, podem, a princípio, aprender qualquer textura rotacionada de qualquer ângulo.

Em [Chen&Kundu, 94], os autores trabalham com decomposição em sub-bandas, através da transformada de Wavelets, e empregando cadeias ocultas de Markov. É realizada uma decomposição ao estilo Wavelets Packets e para cada canal de saída são extraídos momentos estatísticos. O conjunto desses momentos define um vetor de características que evolui segundo se toma, sequencialmente, a saída de cada canal. Tal sequência é utilizada para treinar um HMM a partir de texturas giradas e não giradas. Não se verifica certa robustez à rotação uma vez que o treinamento foi realizado com diversas texturas giradas de 10° , de -170° a 180° e imagens com 512×512 . Ademais, o uso de filtros ortogonais introduz dificuldades por não haver muito controle de sua direção.

2.4) Wavelets

Em contrapartida à decomposição tradicional de Wavelets, Chang *et. al.* [Chang&Kuo, 93] estudaram de forma criteriosa a questão das frequências dominantes de uma textura. Um dos pontos mais importantes de seu trabalho foi concluir que as mais significativas informações espectrais estão distribuídas por canais de média frequência. Desta forma, empregou-se a expansão da imagem por meio de árvores de decomposição de Wavelets, que se diferencia da decomposição tradicional pela geração de tais canais. O algoritmo ramifica a árvore a partir do nó pai, comparando sua energia com a obtida pelos nós filho. A ramificação cessa quando a energia do nó filho é consideravelmente menor que aquela obtida pelo nó pai correspondente ou quando se atinge a profundidade máxima de ramificação. Com cada classe possuindo um mapa de energias dado por sua árvore, para a classificação de uma textura desconhecida realiza-se sua decomposição nos moldes acima e finalmente mede-se a distância do mapa obtido aos mapas previamente armazenados. A classificação é dada pela classe do mapa armazenado que apresentou a menor distância. Observe que a poda, na maneira como foi implementada, não garante que a árvore final seja uma base para a imagem decomposta ou seja, a árvore final pode ser não admissível.

Outras variações deste algoritmo também foram propostas mas observa-se que a abordagem é essencialmente voltada para representação de sinais. Selecionar os canais de maior energia para uma determinada classe não garante que os mesmos contribuam com as características mais discriminativas. Para um grande número de classes a serem tratadas, duas ou mais texturas podem compartilhar dos mesmos canais, já que a energia das subbandas de médias frequências é significativa, o que não contribui na discriminação entre elas. Entretanto, ao contrário das abordagens das máscaras de Laws e das transformadas lineares locais, o método proporciona uma decomposição que se adapta aos tipos de texturas tratados.

2.4.1) Wavelets Packets

A decomposição de um sinal em Wavelet Packets se distingue da decomposição tradicional em Wavelets pelo particionamento realizado no plano de fase [Mallat, 98]. No segundo caso apenas as bandas de baixa frequência são decompostas, assumindo-se que uma boa quantidade de informação do sinal está distribuída em tais regiões espectrais. De maneira diferente, as Wavelet Packets permitem qualquer decomposição diádica, fornecendo uma árvore de decomposição completa ou mesmo que seja aplicado algum algoritmo de poda, como será discutido a seguir. A figura abaixo mostra exemplos de árvores de decomposição para Wavelets e Wavelet Packets além do particionamento no plano tempo \times frequência.

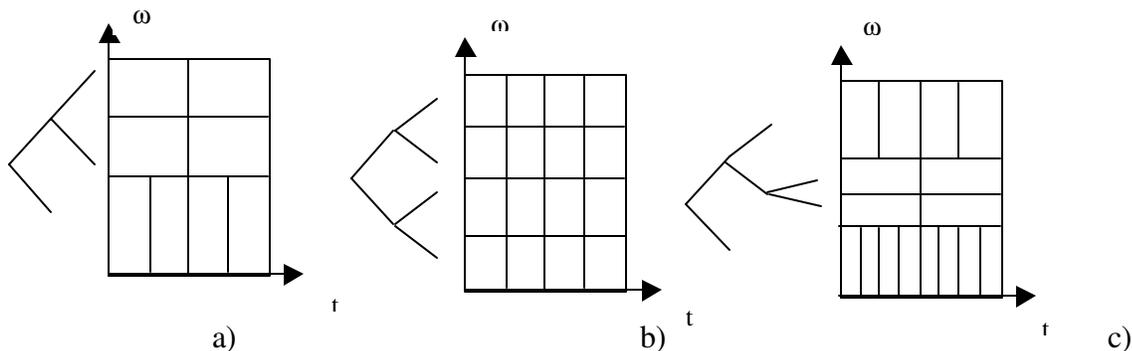


Figura 2.5: Árvores de Decomposição e particionamento do plano tempo (t) \times frequência (ω) para um sinal 1D. Em a) Wavelets com 2 níveis de decomposição. b) Wavelet Packets com árvore, de 2 níveis, completa c) Wavelet Packet, de 3 níveis, com poda.

Em [Etemad&Chellappa, 98] Etemad *et. al.* sugerem um sistema de classificação genérico através da transformada de Wavelet Packets. O ponto principal do trabalho é a geração de uma base de decomposição que melhor represente um determinado sinal (que pode ser de mais de uma dimensão). A obtenção da base é feita de forma adaptativa, a partir da estrutura de múltiplas escalas proporcionada pela geração de uma árvore tipo Wavelets Packets.

A abordagem original para determinação das melhores bases a partir de WP (que já foi aplicada a problemas de representação, classificação, codificação e extração de ruídos envolvendo diversos tipos de sinais) foi proposta por Coifman *et. al.* [Coifman&Wickerhauser, 92]. Coifman sugeriu um algoritmo de poda de nós, aplicável a uma árvore binária (no caso de imagens trabalha-se com a estrutura *quadtree*). Um índice de custo é escolhido de modo a avaliar a associação de um sinal a uma dada classe. O objetivo passa a ser de podar a árvore de tal maneira a minimizar o erro final de classificação. A poda é realizada comparando-se o custo de um determinado nó (nó-pai) com a soma dos custos de seus nós-filho. Estes últimos são mantidos se a soma de seus custos for *melhor* que o apresentado pelo nó pai, caso contrário os nós-filho são retirados da árvore. Para tanto, a árvore é varrida a partir de suas folhas em direção ao nó raiz. Uma vez podada a árvore, o resultado final é um conjunto de nós, cada nó representando uma componente da base que mais se adapta ao problema inicial. Desta forma, tem-se uma estrutura adaptativa que decompõe um sinal, dado um índice de custo e o resultado da poda é base para a imagem decomposta.

Uma das premissas básicas para que a complexidade do algoritmo não seja elevada é justamente assumir uma função de custo aditiva. Ademais, evita-se que seja necessário comparar o custo de um nó contra todos os outros existentes, o que geraria o problema de explosão combinatória. Entretanto, o resultado obtido segundo esta restrição (custo aditivo) diminui as chances de se obter uma solução otimizada. Assim sendo, nem sempre é possível encontrar medidas aditivas que conduzam, da melhor maneira possível, a obtenção da base ótima. Esta foi a principal modificação proposta em [Etemad&Chellappa, 98], onde utilizou-se algoritmos de busca e poda de árvore tais como o ramifica e poda ou eliminação

por retropropagação. Apesar de computacionalmente serem mais custosos, tais algoritmos apresentam melhores resultados quando operam sobre árvores que possuem dependência entre nós. O mesmo acontece para a escolha de custos não aditivos, que fornecem melhor desempenho para nós correlacionados em relação a medidas aditivas (correlação entre nós de diferentes escalas ocorre comumente em decomposição de Wavelets).

Dado que o objetivo é de classificar sinais/texturas, em [Unser&Eden, 90] optou-se por um índice parametrizado pela medida de espalhamento dentro e entre classes, oriundo do índice de Fisher [Duda&Hart, 72], cuja equação pode ser vista em (2.11):

$$J = \text{TR}(S_w^{-1}S_b), \quad (2.11)$$

onde:

S_w é a matriz de espalhamento dentro de cada uma das classes,

S_b é a matriz de espalhamento entre classes,

TR é o operador traço de uma matriz.

Uma vez obtida a árvore podada, são extraídas características a partir de momentos estatísticos, seguido de uma redução de dimensionalidade, e por fim a classificação por uma rede neural artificial – perceptron multicamadas com algoritmo de retropropagação. Para a geração da árvore de WP é preciso, antes de mais nada, selecionar um determinado filtro do tipo reconstrução perfeita. O filtro escolhido é, em geral, fruto de tentativa e erro pois ainda não existe uma avaliação qualitativa que o selecione dado uma classe de sinais [Strang&Nguyen, 96]. Entretanto, a escolha pode se basear em pistas heurísticas, como apresentado em [Coifman&Wickerhauser, 92].

Neste trabalho, os autores não abordaram o problema da variância ao deslocamento, efeito este que acontece quando se decompõe sinais através de bancos de filtros seguidos de sub-amostragem. Uma outra questão importante aparece quando se assume que as imagens a serem classificadas estejam giradas, diferentemente das usadas para treinamento. Em geral, decomposições segundo Wavelets separáveis para imagens empregam o produto tensorial entre filtros FIR-1D para obter-se filtros FIR-2D. Como consequência desta simplificação, os filtros obtidos não respondem de forma isotrópica, tendendo a favorecer

determinadas direções. Caso a textura possua alguma componente discriminatória em uma orientação específica, a mesma não será detectada de maneira satisfatória, o que empobrece os resultados de classificação sobretudo para texturas rotacionadas. Verificou-se uma diminuição no índice de acertos justamente quando são apresentadas texturas giradas de 90° e 180°, elevando a taxa de erros de 1% para 14%. Vale ressaltar que a classificação foi obtida para uma baixa quantidade de texturas, 10 no total, algumas delas não pertencentes ao álbum de Brodatz, o que inviabiliza a comparação de resultados.

2.5) Modelagem por Campos Aleatórios de Markov

Como exposto anteriormente, não é possível tratar pixels pertencentes a uma textura observando-se o histograma de primeira ordem da imagem. Esta foi uma das conclusões de Julesz que, na década de 60, realizando experimentos sobre a capacidade de discriminação do sistema visual humano concluiu (ainda que com algumas restrições) que texturas distintas podem ser percebidas se apresentarem diferenças em seus histogramas de segunda ordem. Este resultado fez surgir diversos estudos sobre o uso de matrizes de co-ocorrência e ditos histogramas no processamento de texturas – análise, classificação e segmentação [Salles, 94].

Uma suposição mais natural é assumir que em uma determinada região, cada um de seus pixels possua uma dependência com seus vizinhos. Este tipo de abordagem é geralmente posta sob a forma de um processo de Markov ou Campo Aleatório de Markov⁹. Uma das dificuldades deste método refere-se à estimação do modelo pois, quanto maior for sua ordem mais complexa se torna tal tarefa. A complexidade se refere tanto ao esforço computacional envolvido como na obtenção de exemplos suficientes a fim de garantir a estimação de seus parâmetros. Já existem trabalhos na literatura mesclando Wavelets ortogonais com os Campos aleatórios de Markov [Zhang *et. al.*, 98] sobretudo no que se refere à geração artificial de texturas e os resultados obtidos são bastante promissores.

⁹ Processos Estocásticos em duas dimensões são comumente referidos na literatura como Campos Aleatórios.

Capítulo 3

A Modelagem das Texturas

3.1) O Modelo para as Texturas

Imagens de texturas, em geral, se caracterizam por apresentar informações não estacionárias o que dificulta em muito sua análise, não sendo possível utilizar de técnicas como transformada de Fourier para obter alguma informação do processo. É necessário o uso de uma análise localizada, que seja capaz de avaliar as variações apresentadas ao longo de seu conjunto suporte. Esta abordagem pressupõe que, localmente, o processo dominante é estacionário de banda estreita e como tal possui associado a ele ou uma ou um conjunto de frequências locais, denominadas em [Bovik *et. al.*, 92] como frequências emergentes. Se estas frequências, quando avaliadas localmente, não variam demasiadamente diz-se que a região de análise em questão possui coerência localizada (ver apêndice A3.2).

Este capítulo trata da modelagem empregada na extração de características das texturas, trabalhando-se sobre o domínio espaço \times frequência espacial. Para um bom entendimento do modelo empregado, modelos multicomponentes AM-FM, é necessária a introdução de conceitos relativos à análise no espaço de fase, espaço tempo \times frequência, além dos conceitos relativos à transformação localizada em si.

3.2) Análise Tempo-Frequência

Como exemplos de sinais complexos é possível citar: sinais de eco de radar e sonar, sinais de fala, tipos variados de imagens, etc.. Segundo o exposto anteriormente, estes sinais não podem ser descritos, de forma inequívoca, através do comportamento global de seus espectros de potência, sendo necessário uma análise localizada. Gabor [Gabor, 46] foi um dos primeiros a tratar tal fato e como resultado propôs uma modificação na transformada de Fourier, adaptando-a para realizar a análise de forma localizada. Este tipo de transformação possui diversas denominações na literatura tais como Transformada de Fourier Janelada, Transformada de Fourier Dependente do Tempo ou Transformada de Fourier de Tempo Curto (do idioma inglês: STFT).

O efeito da localização é obtido com a ajuda de uma função janela que se translada sobre o sinal em estudo de forma a observar apenas uma parte de sua extensão total. Para fins de comparação, supondo a existência de um sinal unidimensional (1D) contínuo $s(t):\mathfrak{R}\rightarrow\mathfrak{R}$, a transformada de Fourier e a transformada de Fourier janelada são expressas segundo as equações (3.1) e (3.2) respectivamente:

$$S(\omega) = \int_{-\infty}^{\infty} s(t)e^{-j\omega t} dt, \quad (3.1)$$

para a frequência $\omega = 2\pi f$, e $j = \sqrt{-1}$;

$$S(\omega, b) = \int_{-\infty}^{\infty} s(t)g^*(t-b)e^{-j\omega t} dt, \quad (3.2)$$

onde

$g(t)$ = janela localizada, sendo $g^*(t)$ seu complexo conjugado,

b = deslocamento temporal dado na janela,

$g(t)e^{-j\omega t}$ = produto conhecido como função atômica, como será visto posteriormente.

Observe que a aplicação de (3.1) realiza uma transformação 1D para 1D ao passo que a equação (3.2) transforma de 1D para 2D. Este aumento na dimensão do sinal transformado é fruto da análise localizada e depende tanto da janela $g(t)$ como do deslocamento b empregado. A vantagem do emprego de (3.2) sobre (3.1) está na

capacidade da primeira equação em isolar uma determinada porção do sinal, porção esta “vista” através de $g(\cdot)$. Este ponto é bastante interessante pois o resultado obtido leva em conta não somente os detalhes locais do sinal como também mostra o quanto a janela afeta no cômputo final da equação (3.2). Ou seja, empregando análise localizada por meio de janelas, o ato de observar um fenômeno modifica-o¹⁰. Observar no presente caso implica em utilizar uma janela com boas propriedades de localização.

A capacidade de localizar uma determinada característica temporal tem o seu preço. Supondo que $s(t)$ apresente uma característica temporal não estacionária muito localizada, o desejável seria aplicar uma janela $g(t)$ tão estreita quanto fosse necessário para poder capturar dita variação. A figura 3.1 abaixo ilustra a situação:

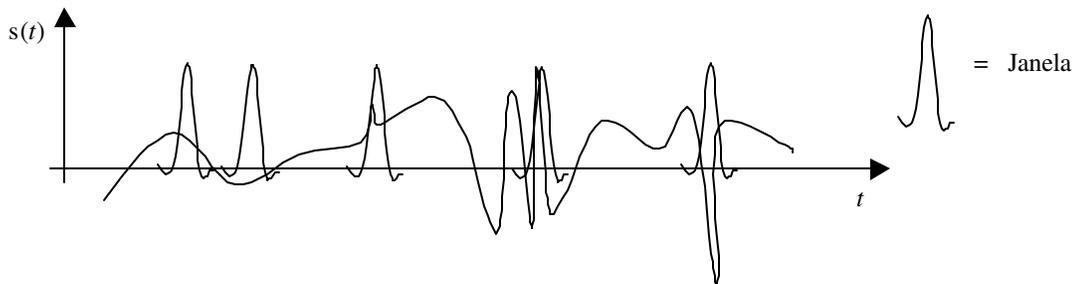


Figura 3.1: Sinal contínuo avaliado por uma janela estreita. Observe que apenas alguns exemplares da janela foram mostrados.

Como se observa, a janela possui as dimensões adequadas para capturar transições rápidas, porém o sinal também possui informação representada sob a forma de uma variação mais lenta, o que pede janelas de avaliação de maior duração temporal. Sinais apresentando variações lentas caracterizam-se por espectros de potência de banda estreita e, em contrapartida, sinais apresentando variações rápidas possuem espectros de banda larga. Assim sendo, não é possível a uma janela de dimensões fixas manter uma boa resolução tanto no domínio do tempo como no domínio da frequência simultaneamente. Esta dualidade entre os domínios temporal e frequencial é denominada de princípio da incerteza e algumas vezes referida na literatura de processamento de sinais como princípio da incerteza de Heisenberg, em homenagem ao pesquisador que primeiramente o enunciou no campo da Física Quântica.

¹⁰ Este efeito também acontece em determinadas análises da Física Quântica.

A incerteza é mensurada através do produto entre as variâncias temporal e frequencial, σ_t^2 e σ_f^2 respectivamente, em torno das respectivas médias de $g(t)$ e sua correspondente transformada de Fourier $G(f)$, conforme as equações abaixo [Mallat, 98]:

$$\sigma_t^2 \sigma_f^2 \geq \frac{1}{4}, \quad (3.3)$$

para

$$\sigma_t^2 = \frac{1}{2\pi \|g\|^2} \int_{-\infty}^{\infty} (t - \mu)^2 |g(t)|^2 dt, \text{ e} \quad (3.4)$$

$$\sigma_f^2 = \frac{1}{2\pi \|G\|^2} \int_{-\infty}^{\infty} (f - \xi)^2 |G(f)|^2 df, \quad (3.5)$$

onde μ é o valor médio de $g(t)$ e ξ é o valor médio de $G(f)$. A inequação em (3.3) só se torna uma igualdade para as janelas gaussianas, apesar dessas possuírem conjunto suporte infinito. O uso deste tipo de funções em transformadas de Fourier dependentes do tempo também é conhecida como Transformada de Gabor. Daugman [Daugman, 88] estendeu o conceito acima para sinais em duas dimensões e demonstrou que, novamente, a transformada janelada baseada em funções gaussianas apresenta a menor incerteza.

Como uma última observação, quando o objetivo é a representação tempo \times frequência, entre as escolhas possíveis, pode-se optar tanto pela Transformada de Fourier Janelada quanto pela distribuição de Wigner Ville. A primeira baseia-se na correlação do sinal com uma família de funções tempo \times frequência (não necessariamente gaussianas modulando exponenciais complexas) e a segunda correlaciona o sinal consigo mesmo, depois de transladado tanto no tempo quanto na frequência. Entretanto, o uso da distribuição de Wigner Ville, distribuição quadrática de energia no espaço de fase, fica limitado pelo aparecimento do termo cruzado [Cohen, 95]. O termo cruzado apresenta-se como informação espúria por não refletir uma condição do sinal, podendo inclusive tomar valores duas vezes maior que o próprio, além de haver a possibilidade de assumir valores negativos [Kwok&Jones, 00]. Assim sendo, devido aos inconvenientes apresentados pela distribuição quadrática acima, optou-se pela análise localizada, baseada em filtragem multi-canal pelos seguintes motivos (e outros fatores a serem abordados posteriormente):

Janelas gaussianas, quando empregadas nos filtros, proporcionam o menor produto de incerteza;
é possível modificar tais filtros, de tal forma a copiar as estruturas diádicas ao estilo Wavelets e assim capturar de maneira mais eficiente a evolução frequencial;
o uso dos filtros de Gabor permite extrair informações importantes como amplitude e frequência instantâneas, características utilizadas nesta tese;
filtros de Gabor possuem uma correspondência quase que direta com estruturas corticais, extremamente eficientes, denominadas de campos receptivos, que são responsáveis por boa parte do pré-processamento da visão.

3.3) A Caracterização AM–FM

Muitos processos físicos naturais manifestam-se sob a forma de amplitude modulada e frequência modulada, AM e FM respectivamente. A modificação da frequência instantânea ao longo do tempo pode caracterizar um sinal e não é sem motivo que assim o faz. Como exemplo, pode-se citar o som de um instrumento de cordas onde o instrumentista varia o comprimento das mesmas e conseqüentemente a frequência do sinal emitido [Cohen, 95]. A luz refletida de uma cena é outro exemplo clássico de percepção da frequência instantânea; ao percorrer os olhos por toda sua extensão experimenta-se a sensação de cores distintas ao longo do espaço. Em ambos os casos, a informação da amplitude está presente e é diretamente relacionada com a intensidade do sinal emitido pontualmente.

Este tipo de análise mostrou-se muito interessante na avaliação de texturas. A trama de uma textura, com um padrão que se repete exatamente, pode indicar múltiplas componentes frequenciais, sendo cada uma delas de valor fixo. Entretanto, texturas naturais não se manifestam desta forma. A pele de um réptil ou a grama de um jardim, à primeira vista, apresenta uma certa taxa de repetição de unidades básicas (as primitivas da textura) ainda que ao ser humano não lhe seja possível definir formalmente tais unidades e qual seria esta frequência de repetição.

Uma das maneiras de representar um determinado sinal utilizando-se tais características é conhecida como modelagem AM–FM. Sob o ponto de vista matemático, e trabalhando-se com sinais $s(t)$ já na forma analítica, é possível escrever ($j = \sqrt{-1}$):

$$s(t) = a(t)e^{j\varphi(t)}, \text{ para } s(t) : \mathfrak{R} \rightarrow \mathbf{C}, \quad (3.6)$$

onde:

$a(t)$ é a componente relativa à modulação de amplitude,

$\varphi(t)$ é a fase do sinal, cuja frequência instantânea é dada por $\omega(t) = \nabla\varphi(t)$.

Convém ressaltar que em geral $s(t) \in \mathfrak{R}$ e portanto seria naturalmente descrita por $s(t) = a(t)\cos(\varphi(t))$. Entretanto, como discutido no Apêndice A3.1, a representação real introduz certos inconvenientes e, por isto, o emprego da forma analítica.

Um dos pontos cruciais na detecção de características AM–FM é o uso de alguma transformação de Fourier janelada no intuito de avaliar localmente a amplitude e a frequência instantânea do sinal. O emprego das transformadas de Gabor, como apresentado anteriormente, encaixa-se muito bem para os objetivos atuais. E a disposição espectral dos filtros de Gabor é inspirada na transformada de Wavelets (conhecida como Wavelets de Gabor), possibilitando um bom compromisso na resolução de componentes não homogêneas. A particularidade do filtro de Gabor em possuir o menor produto de incerteza tempo \times frequência o capacita como etapa básica na avaliação das características AM–FM.

A opção pelo uso de Gabor é apresentada em [Bovik, 91] de maneira formal, onde expressa-se uma função objetivo dependente de um filtro que apresente a melhor propriedade de localização para o problema de segmentação de texturas. Uma vez definido este critério de otimalidade e partindo-se das expressões de variância espacial e espaço/frequencial escritas segundo um banco de filtros inicial hipotético, chega-se à conclusão que a transformação de Gabor é a que melhor atende aos requisitos impostos.

3.4) Sinais Multicomponentes

Sinais multicomponentes são aqueles que se caracterizam por apresentar regiões bem delineadas no espaço de fase [Cohen, 95]. Uma região bem delineada possui valores numéricos (magnitude) mais elevados que o resto dos pontos do plano tempo \times frequência. A figura 3.2.a mostra um sinal multicomponente discreto e determinístico com amplitude unitária e modulação em frequência. Observe que na figura 3.2.b as regiões delineadas possuem a forma de uma montanha e seu aspecto possibilita avaliar a largura de banda instantânea de cada componente do sinal. A região correspondente ao cume das montanhas define um conjunto de frequências instantâneas equivalentes ao centro da respectiva banda instantânea. Tais frequências nada mais são que máximos locais associados ao espectrograma do sinal. O espalhamento em torno deste conjunto de máximos, espalhamento este medido em relação ao eixo das frequências, define a largura de banda instantânea. Para sinais 2D, a visualização não é possível uma vez que o plano de fase gerado possui dimensão 4.

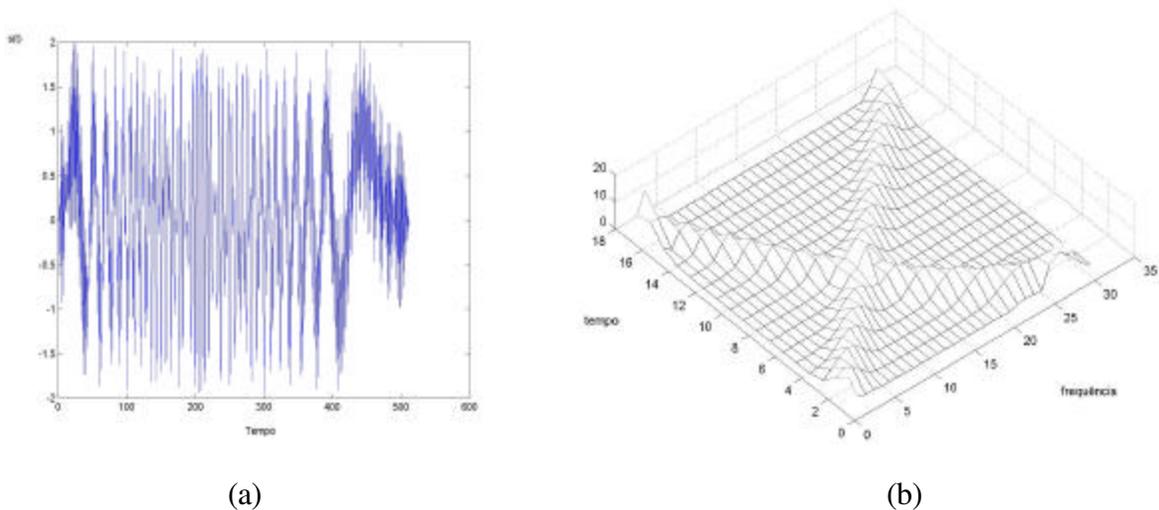


Figura 3.2: (a) Sinal com duas componentes a saber: incremento linear e decremento quadrático na frequência ao longo do tempo (em segundos). Em (b) observa-se seu espectrograma – frequência em hz.

Desta forma, o sinal passa a ser caracterizado pelos pontos de máximo do espectrograma e suas correspondentes frequências instantâneas. Entretanto, dado um sinal

não estacionário, nem todas as regiões do plano de fase são relevantes para sua representação. Muitas delas são relativas às deficiências de digitalização, contaminação por ruído ou não possuem mesmo componentes importantes, mascarando assim as verdadeiras características tempo \times frequência. A solução apresentada na literatura [Mallat, 98] [Clark&Bovik, 89] [Havlicek&Bovik, 95] [Delprat *et. al.*, 92] [Gonnet&Torresani, 94], com algumas diferenças entre si, é baseada na detecção dos máximos locais do espectrograma, segundo o exposto anteriormente. Este tipo de análise é denominada em [Havlicek *et. al.*, 00] como análise da componente dominante e em [Gonnet&Torresani, 94] como detecção do “cume da montanha” no plano de fase, sendo indicados para seguir a evolução não estacionária de um sinal ao longo de seu conjunto suporte.

Porém, quando se avalia uma textura, nem sempre é desejável atribuir a uma região apenas uma componente AM–FM. Muitas vezes, para capturar toda a sua essência, é importante avaliar a informação contida em outros canais que não necessariamente caracterizam-se por apresentar uma frequência instantânea dominante. A figura 3.3 abaixo relaciona uma região de 128×128 pixels de duas texturas e seus respectivos espectro de potências.

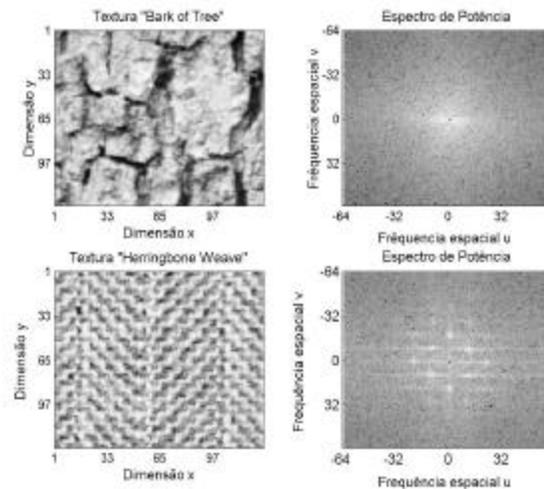


Figura 3.3: Textura e seus respectivos espectros de potência. As áreas mais claras de cada espectro representam pontos de potência elevada.

Como se observa, ambos os espectros são bastante ricos em informação e foram obtidos a partir das imagens com média zero. Devido à existência de um espalhamento da informação espectral e sob o ponto de vista de classificação de texturas, a seleção da componente de maior valor no correspondente espectrograma 4D pode não ser capaz de conferir à etapa de classificação a capacidade necessária de decisão entre classes. Fang Liu *et. al.* [Liu&Picard, 99], analisando de forma estocástica e não localizada, aproximaram texturas homogêneas por processos estacionários passíveis de serem descritos pela decomposição de Wold¹¹. Outros autores [Haley&Manjunath, 98], tratando o problema da rotação de texturas, empregaram filtros de Gabor modificados em multiresolução e advogam que deve-se tomar todas as informações possíveis, e de todos os canais, de forma simultânea.

A abordagem proposta nesta tese lida com o problema de maneira distinta. Para a questão da classificação de texturas assume-se que as informações necessárias são obtidas de uma análise cuidadosa dos espectros de potência localizados resultantes de cada canal e que uma determinada textura pode ser caracterizada pelas amplitude e frequência instantâneas locais. O modelo utilizado para uma classe é dado pela equação multicomponente (assumindo representação analítica):

$$t(\mathbf{x}) = \sum_{k=1}^K a_k(\mathbf{x}) e^{j\varphi_k(\mathbf{x})}, \text{ para } t(\mathbf{x}) : \mathfrak{R} \rightarrow \mathbf{C}, \mathbf{x} \in \mathbf{Z}^2, \quad (3.7)$$

onde:

$$j = \sqrt{-1},$$

$t(\mathbf{x})$ é a textura a ser modelada,

$a_k(\mathbf{x})$ é a k-ésima componente relativa à modulação de amplitude,

$\varphi_k(\mathbf{x})$ é a k-ésima fase do sinal, cuja k-ésima frequência instantânea é dada por $\omega_k(\mathbf{x}) = \nabla \varphi_k(\mathbf{x})$.

Ao assumir que uma determinada imagem é descrita por (3.7), não se sabe de antemão quantas são as componentes existentes em cada uma delas e se ao menos através

¹¹ Decomposição que expressa um processo estacionário segundo a soma de um processo predito e outro regular [Hayes, 96].

da filtragem por banco de filtros será possível detectá-las. Desta forma, é necessário arbitrar uma certa quantidade de componentes fixando-se o número de filtros locais empregados. Por conseguinte, cada filtro do banco de filtros de Gabor isola uma porção do espectro que por sua vez é caracterizada por uma componente de (3.7).

A abordagem da componente dominante pode ser indicada para duas situações distintas. Em um primeiro caso assume-se que uma textura varia suas características ao longo de seu domínio espacial. Assim, ao se dividir uma imagem de, por exemplo, 512×512 pixels em áreas de 16×16 pixels, empregando-se a abordagem das frequências instantâneas relativas aos máximos locais do espectrograma, é possível associar a cada sub-região 16×16 pixels uma (01) frequência e uma (01) amplitude instantânea. Este expediente é utilizado em [Havlicek&Bovik, 95] e [Havlicek *et. al.*, 00] para o problema de segmentação de texturas que, como se sabe, exige uma análise muito delicada nas regiões de fronteira entre classes distintas. E justamente neste ponto surge o problema do compromisso da sub-região de análise. Segmentar exige trabalhar o mais localmente possível, o que não é muito indicado para texturas de grão grosso, texturas essas que pedem sub-regiões de maiores dimensões, segundo se observa na figura (3.3). Como o presente trabalho trata do problema de classificação de texturas (a segunda situação), sejam elas de grão fino ou não, as dimensões assumidas são de no mínimo 64×64 pixels, obtendo-se melhores resultados para 128×128 pixels. Portanto, o espectro da imagem assumida como exemplo de treinamento contém as características espectrais instantâneas para a correta representação da classe em questão. E não se faz restrição ao tipo de textura a ser classificada desde que se escolha de maneira conveniente as dimensões da área de análise. Observe que texturas de caráter fractal não são tratadas corretamente pelo modelo aqui proposto.

3.5) A Transformada de Wavelets Bi-dimensional

Existe um interesse particular na estrutura Wavelet pois, de acordo com o comentado na seção 3.2, o emprego da STFT com janela de dimensões fixas não consegue seguir de forma eficiente todas as variações sofridas pela frequência do sinal sob análise. A transformada Wavelet decompõe um sinal a partir de componentes de resoluções distintas,

permitindo que determinadas características sejam melhor “capturadas”. Um exemplo clássico são os sinais de eco de radar, sinais que apresentam grande variação na frequência instantânea. Em especial para as Wavelets de Gabor, alia-se o desempenho da localização tempo \times frequência dos filtros de Gabor, em distintas frequências radiais, com a propriedade de invariância ao deslocamento espacial. Porém, tal arranjo apresenta a desvantagem de ser uma transformação não ortogonal, o que eleva sua complexidade computacional.

Para apresentar tal transformação assume-se, primeiramente, que a transformada de Fourier $F(u, v)$ de uma imagem $f(x, y)$ bi-dimensional é dada pela seguinte equação:

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp^{-j2\pi(ux+vy)} dx dy, \quad (3.8)$$

onde x, y, u e $v \in \Re$ e $j = \sqrt{-1}$ e

(x, y) representa o par das coordenadas da imagem no domínio espacial,

(u, v) representa as frequências do espectro da imagem.

De maneira mais conveniente, é possível expressar (3.8) na notação vetorial, onde define-se $\mathbf{w} = [2\pi u \ 2\pi v]^T$, $\mathbf{x} = [x \ y]^T$, para Tr = transposto. Assim sendo, a equação (3.8) pode ser reescrita como:

$$F(\mathbf{w}) = \int_{\mathbf{x} \in \Re^2} f(\mathbf{x}) \exp^{-j\mathbf{w}^T \mathbf{x}} d\mathbf{x}. \quad (3.9)$$

Agora, assumamos uma função $\psi(\mathbf{x})$ com boa localização espaço \times frequência espacial, de tal forma a atender os seguintes requisitos [Gonnet&Torresani, 94]:

$\psi(\mathbf{x})$ deve possuir conjunto suporte compacto em uma determinada vizinhança e deve assumir valor máximo em $\mathbf{x} = [0, 0]^T$,

a transformada de Fourier de $\psi(\mathbf{x})$, dada por $\mathfrak{F}\{\psi(\mathbf{x})\} = \Psi(\mathbf{w})$ toma valor máximo local para $\mathbf{w} = \mathbf{w}_0$ com um bom decaimento a partir destes valores de frequências (definidas por \mathbf{w}_0),

$\Psi(\mathbf{w})$ deve satisfazer ao critério de admissibilidade segundo a equação abaixo:

$$\int_{\mathbf{w} \in \mathfrak{R}^2} \frac{|\Psi(\mathbf{w})|^2}{\|\mathbf{w}\|} < \infty. \quad (3.10)$$

Da equação (3.10) conclui-se que devido a presença de \mathbf{w} no denominador do termo integrando, uma condição necessária para satisfazer a desigualdade é:

$$\Psi(\mathbf{0}) = 0, \therefore \mathfrak{I}^{-1}\{\Psi(\mathbf{w})\}_{\mathbf{w}=[0,0]^{Tr}} = \int_{\mathbf{x} \in \mathfrak{R}^2} \psi(\mathbf{x}) d\mathbf{x} = 0 \quad (3.11)$$

Ou seja, a função localizada $\psi(\mathbf{x})$ constitui um filtro passa-banda. Este ponto é muito interessante pois um conjunto de tais funções transladadas no espaço e modulando exponenciais complexas pode gerar o espaço \mathfrak{R}^2 (com o auxílio de uma função extra associada ao filtro passa-baixa correspondente. Outra interpretação comum é associar dito arranjo ao conceito de janelas empregadas em análise espectral [Hayes, 96]). Esta também é a estrutura básica da transformada de Wavelets, entretanto modifica-se a operação de modulação pela operação de mudança da escala da função. Em termos matemáticos, pode-se escrever [Strang&Nguyen, 96] ($j = \sqrt{-1}$):

Funções Janelas: $\psi(\mathbf{x})$ modula (uma exponencial complexa) e translada-se

$$\Psi_{m,\mathbf{b}}(\mathbf{x}) = e^{jm\mathbf{W}^{Tr}\mathbf{x}}\psi(\mathbf{x} - \mathbf{b}), \quad (3.12.a)$$

Funções Wavelets: $\psi(\mathbf{x})$ dilata-se e translada-se

$$\Psi_{s,\mathbf{b}}(\mathbf{x}) = s^{k/2}\psi(s^k\mathbf{x} - \mathbf{b}), \quad (3.12.b)$$

para:

m = fator de modulação,

k = fator de mudança de escala,

s = escala inicial,

\mathbf{b} = deslocamento espacial,

$\mathbf{W} = [2\pi U \ 2\pi V]^{Tr}$ representa um par de frequências específicas.

As funções acima são conhecidas como funções atômicas pois “ladrilham” o espaço de sinal. Existem certas condições que determinam se ambas as funções, janelas e Wavelets, geram o espaço bi-dimensional; tudo depende da escolha de seus parâmetros [Strang&Nguyen, 96]. Uma vez que o espaço possa ser gerado, tanto por (3.12.a) como por

(3.12.b), diz-se que tais funções, sob o aspecto de geração de espaços vetoriais, são denominadas de *frame*. No presente trabalho, como o interesse é obter características locais de uma textura, não sendo necessária a reconstrução do sinal, basta assumir que as funções de decomposição ψ atendam às três premissas básicas listadas anteriormente. Outras considerações, a respeito da disposição espectral dos filtros, serão comentadas posteriormente.

3.5.1) Os Filtros de Gabor

Os filtros de Gabor são obtidos fazendo-se um envelope gaussiano modular uma exponencial complexa, como se escreve abaixo:

$$h(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \times \exp[j2\pi(Ux + Vy + \phi)] \text{ para } j = \sqrt{-1}, \quad (3.13)$$

(x, y) representa o par das coordenadas no domínio espacial,

$(u=U, v=V)$ representa as frequências ao qual o filtro está centrado,

(σ_x, σ_y) determina o espalhamento espacial do envelope gaussiano,

ϕ é uma componente de fase da equação da exponencial complexa (em geral $\phi = 0$).

A figura abaixo mostra as partes real e imaginária de um filtro de Gabor.

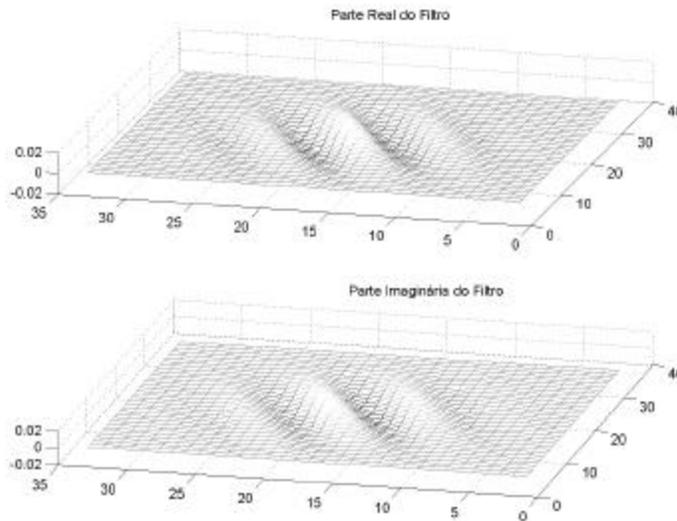


Figura 3.4: Resposta (real e imaginária) de um filtro de Gabor assumidos em uma grade de 33x33 pontos com origem em (17, 17).

Quando assume-se que $\sigma_x \neq \sigma_y$, a resposta espectral do filtro privilegia uma determinada orientação pois, naturalmente, cada canal está centrado em uma frequência radial ω_r e orientado segundo um ângulo ρ de acordo com as equações (3.14) abaixo:

$$\omega_r = \sqrt{\omega_U^2 + \omega_V^2}, \text{ para } \omega_U=2\pi U \text{ e } \omega_V=2\pi V \quad (3.14.a)$$

$$\rho = \arctan\left(\frac{\omega_V}{\omega_U}\right), \quad (3.14.b)$$

$$\sigma_u = \frac{1}{2\pi\sigma_x} \text{ e } \sigma_v = \frac{1}{2\pi\sigma_y}. \quad (3.14.c)$$

O uso dessa desigualdade ($\sigma_x \neq \sigma_y$) mostrou-se útil para o tratamento de texturas em presença de rotação, uma vez que para texturas anisotrópicas é importante avaliar seu comportamento nas diferentes direções.

É interessante assumir a transformação abaixo sobre o par (x,y) para (x',y') e $\phi=0$:

$$h(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2}\right)\right] \times \exp[-j2\pi(Ux + Vy)], \quad (3.15)$$

cujos parâmetros são apresentados como se segue:

θ determina a disposição angular do filtro no plano espectral.

$$\begin{cases} x' = x \cos \theta + y \sin \theta \\ y' = y \cos \theta - x \sin \theta \end{cases}, \text{ que reescrito como uma de matriz de rotação tem-se:}$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (3.16)$$

Usa-se fazer $\rho = \theta$ e para M orientações tem-se que $\theta = m\pi/M$, $m \in \{0, 1, \dots, M-1\}$. Desta forma, é possível posicionar um filtro em qualquer localização do plano espectral dado um par (U,V) e dada uma orientação θ . A resposta em frequência espacial para o filtro da expressão (3.15) é dada a seguir:

$$H(u, v) = \exp\{-2\pi^2[(u' - U')^2\sigma_u^2 + (v' - V')^2\sigma_v^2]\}, \quad (3.17)$$

onde u' , v' , U' e V' sofreram a transformação de rotação dada pela matriz de rotação da equação (3.16). A partir de (3.17) pode-se definir as larguras de banda de frequência radial B_r com $\omega_r=u_0$ e de orientação B_θ tal como se segue:

$$B_r = \log_2 \left(\frac{u_0 + (2 \ln 2)^{0.5} \sigma_u}{u_0 - (2 \ln 2)^{0.5} \sigma_u} \right), \quad (3.18)$$

$$B_\theta = 2 \tan^{-1} \left(\frac{(2 \ln 2)^{0.5} \sigma_u}{u_0} \right). \quad (3.19)$$

Ambas as equações foram obtidas supondo largura de banda medindo-se até as frequências de metade da magnitude em relação ao meio da faixa. A figura 3.5 mostra a relação de B_θ e B_r com a curva de meia magnitude de um filtro no domínio da frequência, assim como seus parâmetros de orientação ρ (já igual a θ) e de frequência radial u_0 .

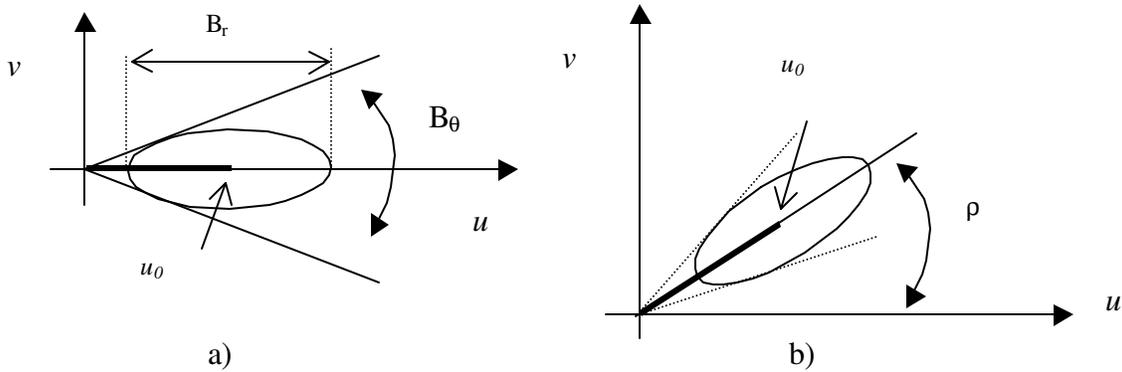


Figura 3.5: Em a) determinação B_θ e B_r para o filtro no domínio da frequência (onde a elipse representa a curva de meia magnitude). Em b) parâmetros de orientação ρ e frequência radial u_0 .

A expressão (3.15) pode ser reescrita na forma apresentada abaixo:

$$h(\mathbf{x}) = \frac{|\Gamma|}{2\pi} \exp(-0.5 \mathbf{x}^{Tr} \Gamma^2 \mathbf{x}) \exp(j \mathbf{W}^{Tr} \mathbf{R}_\theta \Gamma \mathbf{x}), \quad (3.20)$$

onde

$$\mathbf{R}_\theta = \begin{bmatrix} \cos \theta & \text{sen} \theta \\ -\text{sen} \theta & \cos \theta \end{bmatrix}, \quad \text{ou seja } \mathbf{R}_\theta^{Tr} = \mathbf{R}_\theta^{-1} \therefore \mathbf{R}_\theta^{Tr} \mathbf{R}_\theta = \mathbf{I}, \quad \mathbf{I} = \text{matriz identidade},$$

$$\Gamma = \begin{bmatrix} 1/\sigma_x & 0 \\ 0 & 1/\sigma_y \end{bmatrix}, \quad \Gamma^2 = \Gamma \Gamma = \Gamma^{Tr} \Gamma, \quad |\Gamma| = \det(\Gamma),$$

$$\mathbf{W} = [2\pi U \ 2\pi V]^{Tr}, \quad \mathbf{x} = [x \ y]^{Tr}, \quad \text{para } x, y, U \text{ e } V \in \Re \text{ e } j = \sqrt{-1}.$$

Observe que em contraste com as equações (3.12.a) e (3.12.b), a definição de funções ψ a partir dos filtros $h(\mathbf{x})$ envolve tanto o conceito de modulação como de multiresolução. A

matriz Γ controla a largura de banda efetiva de cada filtro e o vetor de frequências $[U, V]^T$ determina a localização do filtro no espectro de frequências espaciais. Já a orientação é controlada pela matriz R_θ . Ou seja, $h(\mathbf{x})$ emprega tanto a modulação complexa como o “ladrilhamento” ao estilo Wavelets. Filtros de Gabor 2D obtidos desta forma são conhecidos como Wavelets de Gabor. É interessante notar que alguns autores [Hsim&Li, 98][Hsim, 00] propuseram o uso das Wavelets moduladas e sob este aspecto as transformadas de Gabor podem ser definidas como tal (apesar de que a proposta possui cunho genérico).

Agora, de acordo com o exposto anteriormente, a transformação de $t(\mathbf{x})$ por Wavelets $\psi(\mathbf{x})$ gera os coeficientes $z(\mathbf{b})$ dados por (supondo aqui que $\mathbf{x}, \mathbf{b} \in \mathfrak{R}^2$):

$$z(\mathbf{b}) = \langle t(\mathbf{x}), \psi(\mathbf{x} - \mathbf{b}) \rangle = \int_{\mathbf{x} \in \mathfrak{R}^2} t(\mathbf{x}) \psi^*(\mathbf{x} - \mathbf{b}) d\mathbf{x}, \quad \therefore z(\mathbf{b}) = t(\mathbf{b}) * \psi^*(-\mathbf{b}), \quad (3.21)$$

(onde $*$ representa convolução espacial). Ou seja, a transformação pode ser realizada por um procedimento de convolução desde que o filtro $h(\mathbf{x}) = \psi^*(-\mathbf{x})$, em $\mathbf{x} = \mathbf{b}$. Para o filtro da expressão (3.15), de maneira genérica, $h(\mathbf{x}) = \psi_{\Gamma, \mathbf{w}, \mathbf{b}, R_\theta}^*(-\mathbf{x})$. O tratamento equivalente, para $\mathbf{x} \in \mathbf{Z}^2$, entre transformada e banco de filtros pode ser encontrado no apêndice A3.3.

3.5.2) A disposição dos filtros

Do exposto acima, observa-se que os parâmetros de seleção dos filtros de Gabor são a evolução das frequências centrais de cada um deles e suas respectivas larguras de banda. O controle destas variáveis pode conferir ao conjunto a característica de transformação ao estilo Wavelets. Assim, uma condição importante refere-se à cobertura do espectro de potências de forma a capturar todas as informações importantes e como compatibilizá-las com a etapa de classificação.

A localização dos filtros no espectro de potências é determinada pela frequência radial ω_r , descrita em (3.14.a), que descreve um círculo de raio de mesmo valor. M filtros de frequência central ω_r , com largura de banda de frequência radial e fase definidas por (3.18) e (3.19) respectivamente, podem ser dispostos de tal forma que os picos de

magnitude de meia potência tocam-se mutuamente. Para 4 valores distintos de ω_r e 8 orientações, o arranjo utilizado pode ser visualizado na figura 3.6 abaixo.

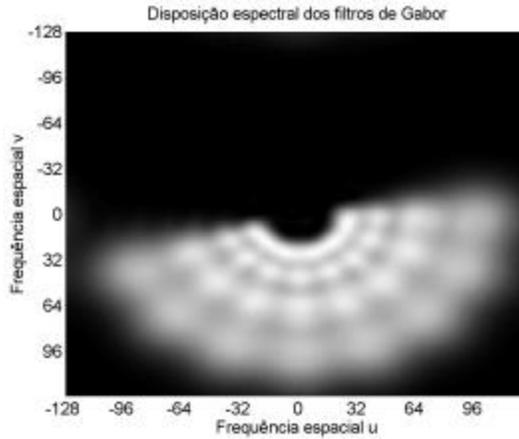


Figura 3.6: Disposição espectral dos filtros de Gabor empregados nesta tese. Os eixos vertical e horizontal representam as frequências. Observe que devido à redundância espectral apenas dois quadrantes são ocupados pelos filtros.

3.6) A Modelagem Utilizada

A presente modelagem propõe o uso de cadeias ocultas de Markov para o sequenciamento das características AM–FM de cada classe de textura. Este classificador tem como função objetivo a maximização da verossimilhança usada para a representação das texturas. A classificação se dá por representação e não discriminação entre classes e seu bom desempenho depende de como se dispõe os filtros no plano espectral. A seguir, descreve-se o modelo de extração de características.

Seja uma textura $f(\mathbf{x})$, $f(\mathbf{x}):\mathfrak{R}^2 \rightarrow \mathfrak{R}$ $\mathbf{x} \in \mathbf{Z}^2$ e uma transformação H que permite o mapeamento $f(\mathbf{x}) \rightarrow t(\mathbf{x})$, $t(\mathbf{x}):\mathfrak{R}^2 \rightarrow \mathbf{C}$ onde assume-se $t(\mathbf{x})$ multicomponente AM–FM dada

por $t(\mathbf{x}) = \sum_{i=1}^K a_i(\mathbf{x})e^{j\phi_i(\mathbf{x})}$ (ver Apêndice A3.1). Seja também um banco de filtros de Gabor

$h(\mathbf{x}) = \psi^*(-\mathbf{x})$, que atendem às condições de admissibilidade expostas anteriormente. Assim sendo, é possível escrever que (para uma componente i apenas $t_i(\mathbf{x}) = a_i(\mathbf{x})e^{j\phi_i(\mathbf{x})}$):

$$z(\Gamma, \mathbf{b}, \mathbf{w}, R_\theta) = \left\langle t(\mathbf{x}), \Psi_{\Gamma, \mathbf{w}, \mathbf{b}, R_\theta}(\mathbf{x}) \right\rangle. \quad (3.22)$$

Dado que o objetivo desta tese é a representação das texturas através do sequenciamento de características instantâneas, a dificuldade que aparece neste ponto diz respeito à definição do modelo dos coeficientes $z(\cdot)$ em (3.22). Tal modelo é importante pois as medidas dos valores instantâneos localizados AM-FM fazem sentido desde que seja possível assumir que $z(\cdot)$ também é representado por uma função analítica. Assim sendo, expressões analíticas do tipo $a(\mathbf{x})e^{j\phi(\mathbf{x})}$ regem tanto as componentes das texturas como também as funções resultantes da decomposição segundo os núcleos de Gabor. Na verdade, sua validade para $z(\cdot)$ depende de algumas considerações e parte delas foi discutida anteriormente.

Dentre as abordagens possíveis, optou-se pela aproximação por quase-autofunções [Havlicek&Bovik, 95] (discutida no apêndice A3.2), onde a resposta de cada filtro é expressa na equação abaixo:

$$z(\mathbf{x}) = t(\mathbf{x}) * h(\mathbf{x}). \quad (3.23)$$

Se a imagem $t(\mathbf{x})$ for monocromática (no sentido de haver apenas um único nível de cinza sobre toda sua extensão), para o qual as informações de amplitude a e frequência \mathbf{W} instantâneas são constantes, $t(\mathbf{x})$ é dada pela expressão $t(\mathbf{x}) = ae^{j\mathbf{W}^T \mathbf{x}}$, $\mathbf{W} = [2\pi U_0 \ 2\pi V_0]^T$. Desta forma, empregando os filtros de Gabor localizados pode-se afirmar, de forma exata, que:

$$z(\mathbf{x}) = t(\mathbf{x}) * h(\mathbf{x}) = t(\mathbf{x})H(\mathbf{W}), \quad (3.24.a)$$

onde $H(\cdot)$ é a resposta em frequências espaciais do filtro centrado em \mathbf{W} . Ou seja, $t(\mathbf{x})$ passa a ser uma autofunção tendo associada a ela autovalores dados por $H(\mathbf{W})$. Entretanto, as características de boa localização dos operadores de Gabor, em conjunto com a propriedade de coerência localizada (o que implica em funções de modulação apresentando variações suaves – apêndice A3.2) permitem assumir a seguinte aproximação:

$$z(\mathbf{x}) = t(\mathbf{x}) * h(\mathbf{x}) \approx t(\mathbf{x})H(\nabla\phi(\mathbf{x})), \quad (3.24.b)$$

onde agora não se restringe $t(\mathbf{x})$ a ser monocromática mas que atenda à propriedade de coerência localizada, sendo modelada pela equação $t(\mathbf{x}) = a(\mathbf{x})e^{j\phi(\mathbf{x})}$. Sob esta ótica é que

se introduz a aproximação por quase autofunções [Havlicek&Bovik, 95]. Portanto, a equação (3.24.b) é reescrita, a partir de (3.21) e (3.22) como se segue:

$$z(\Gamma, \mathbf{b}, \mathbf{W}, \mathbf{R}_\theta) \approx t(\mathbf{x}) * h_{\Gamma, \mathbf{W}=\nabla\varphi(\mathbf{x}), \mathbf{b}, \mathbf{R}_\theta}(\mathbf{x}). \quad (3.25)$$

O uso de tal modelo é atrativo pois permite que se desenvolva um algoritmo para a extração da frequência instantânea e da amplitude instantânea. Se $\mathbf{x} \in \mathfrak{R}^2$, a obtenção da frequência instantânea é possível, dada por uma expressão analítica não linear, a partir da aplicação do operador gradiente em (3.24.b) e isolando-se o termo relativo à frequência [Havlicek&Bovik, 95]. Contudo, para a forma discreta, $\mathbf{x} \in \mathbf{Z}^2$, é necessário a introdução de um operador linear invariante ao deslocamento equivalente a um gradiente discreto, descrito em [Havlicek *et. al.*, 00] por:

$$g_i(\mathbf{x}) = \delta(\mathbf{x}+n_1\mathbf{e}_i)+q\delta(\mathbf{x}+n_2\mathbf{e}_i), \quad (3.26)$$

para:

$\delta(\mathbf{x})$: Delta de Kronecker,

\mathbf{e}_i : vetor unitário indicando uma das duas direções possíveis: $i=1,2$

n_1, n_2 : deslocamento desejado em cada direção, com valores no conjunto $\{-1,1\}$,

q : constante que pode assumir valores no conjunto $\{-1,1\}$.

A obtenção da frequência instantânea é levada a cabo através da convolução:

$$y(\mathbf{x}) = z(\mathbf{x}) * g_i(\mathbf{x}). \quad (3.27)$$

onde $y(\mathbf{x})$ representa o mapa 2D de frequências instantâneas. Em (3.27) aplica-se a aproximação das quase autofunções onde, após manipulações algébricas, obtém-se o algoritmo de determinação da frequência instantânea:

$$\mathbf{e}_i^T \nabla\varphi(\mathbf{x}) = \arcsin \left[\frac{z(\mathbf{x} + \mathbf{e}_i) - z(\mathbf{x} - \mathbf{e}_i)}{2jz(\mathbf{x})} \right], \quad (3.28)$$

$j = \sqrt{-1}$ e assumiu-se $n_1 = q = 1$ e $n_2 = -1$. Também a partir da aproximação das quase autofunções, e uma vez obtida $\nabla\varphi(\mathbf{x})$ a determinação da amplitude instantânea $a(\mathbf{x})$ é, após manipulações algébricas, dada por:

$$a(\mathbf{x}) = \left| \frac{z(\mathbf{x})}{\Psi[\nabla\varphi(\mathbf{x})]} \right|. \quad (3.29)$$

As características instantâneas obtidas produzem como resultado um conjunto de sinais bidimensionais, ou mapas instantâneos localizados.

Neste ponto tem lugar o procedimento de redução de dimensionalidade, onde extrai-se o valor médio de cada um dos mapas acima, a partir da saída de cada canal. Ou seja, o modelo proposto assume que $f(\mathbf{x})$ pode ser representada por funcionais aplicados às componentes espectrais localizadas AM-FM. Localmente, extrai-se as referidas medidas, amplitude instantânea $a(\mathbf{x})$ e frequência instantânea $\mathbf{e}_i^{Tr} \nabla \varphi(\mathbf{x})$, $i=1,2$ e calcula-se:

$$c_1(\omega_r, \theta_j) = \frac{\sum_x \sum_y \mathbf{e}_{i=1}^{Tr} \nabla \varphi(x, y)}{XY}, \quad (3.30.a)$$

$$c_2(\omega_r, \theta_j) = \frac{\sum_x \sum_y \mathbf{e}_{i=2}^{Tr} \nabla \varphi(x, y)}{XY}, \quad (3.30.b)$$

$$c_3(\omega_r, \theta_j) = \frac{\sum_x \sum_y a(x, y)}{XY}, \quad (3.30.c)$$

$$c_4(\omega_r, \theta_j) = \frac{\sum_x \sum_y \sqrt{(\mathbf{e}_{i=2}^{Tr} \nabla \varphi(x, y))^2 + (\mathbf{e}_{i=1}^{Tr} \nabla \varphi(x, y))^2}}{XY}, \quad (3.30.d)$$

$$c_5(\omega_r, \theta_j) = \frac{\sum_x \sum_y \arctan\left(\frac{\mathbf{e}_{i=2}^{Tr} \nabla \varphi(x, y)}{\mathbf{e}_{i=1}^{Tr} \nabla \varphi(x, y)}\right)}{XY}, \quad (3.30.e)$$

para X e Y representando as dimensões do conjunto suporte da frequência instantânea e amplitude instantânea. Cada característica $c_k(\omega_r, \theta_j)$, $k \in \{1, \dots, 5\}$ é o resultado de um canal com frequência radial ω_r , $r \in \{1, \dots, R\}$, onde $\omega_{r=1} \neq 0$ e orientação θ_j , $j \in \{0, \dots, J-1\}$, perfazendo um total de $R \times J$ filtros. Nem todas as características apresentadas em (3.30) são utilizadas para classificação ao mesmo tempo.

Uma vez aplicado dito procedimento a todos os canais, monta-se a sequência de características desejada. O cálculo dos funcionais acima é feito para cada um dos canais de maneira a formar uma sequência de vetores. Nesta sequência fica representada como se comporta a textura sob o ponto de vista de evolução espectral de suas componentes de

frequência e amplitude instantâneas. Tal forma de representar o problema, através de uma sequência de vetores de características AM–FM, é capaz de capturar a essência de uma textura.

A modelagem, a partir deste ponto possui caráter estocástico. Assume-se que dita sequência de vetores pode ser descrita por uma cadeia oculta de Markov, HMM, a ser discutido no próximo capítulo. Como será visto, um HMM é composto de estados e a cada um deles é possível associar uma mistura de funções de densidade de probabilidade, mistura esta relativa aos vetores característicos. Como o número de características extraídas localmente é pequeno, consegue-se realizar uma estimativa de ditas pdf's, o que contribui com a taxa de acertos do modelo. Em contrapartida, algumas abordagens tradicionais [Haley&Manjunath, 98] extraem um grande número de características e terminam em assumir modelos gaussianos não markovianos para cada textura, além de não serem capazes de estimar a matriz de variância/covariância.

Cabe aqui ressaltar que com a cadeia oculta de Markov utilizada não se representa um modelo markoviano de n-ésima ordem, mesmo porque empregou-se nesta tese HMM's de ordem 1. A utilização de funcionais, aplicados à saída de filtros bem localizados no espaço, possibilita que a sequência de vetores seja bem representada por HMM's. E, muito importante é o fato da boa localização dos filtros (característica não explorada em [Chen&Kundu, 95][Chen&Kundu, 94]) pois, de forma aproximada, é possível supor uma dependência forte entre canais vizinhos, diminuindo-se à medida que se toma canais não contíguos.

No que tange ao algoritmo HMM, um passo inicial de agrupamento é necessário visto que não é possível atribuir sentido físico aos estados da cadeia, como será tratado no próximo capítulo. Em continuação, executa-se uma etapa de pré-treinamento segundo o procedimento dos estados otimizados. O processo de treinamento em si, baseado no algoritmo da Re-estimação, continua por ajustar os parâmetros do modelo até que se atenda a um critério de parada. Como toda a etapa de aprendizagem é proposta segundo a função objetivo de maximização da verossimilhança, a modelagem empregada classifica segundo o

paradigma da melhor representação. Uma vez que todos os HMM's estejam treinados, cada um deles relativo a uma textura, executa-se o algoritmo de Viterbi que devolve o HMM que modela a classe de textura procurada.

Desta forma, supondo um conjunto de $R \times J$ filtros, R frequências radiais e J orientações distintas conforme apresentado na figura (3.6), com P exemplos por classe, é possível escrever:

Procedimento de extração de características e treinamento dos HMM's:

para uma determinada frequência radial ω_r , são realizadas $K \times J$ medidas (K funcionais, como os apresentados na equação (3.30), para cada um dos J filtros orientados), para $k \in \{1, \dots, K\}$, $r \in \{1, \dots, R\}$ e $j \in \{0, \dots, J-1\}$, monta-se o vetor de evolução das características $\mathbf{v}_{r,j} = [c_1 \ c_2 \ \dots \ c_k]^{Tr}$ e $k \leq K$, constrói-se a sequência de observações $O_p = \{\mathbf{v}_{1,0} \ \mathbf{v}_{1,1} \ \dots \ \mathbf{v}_{1,J-1}, \mathbf{v}_{2,0} \ \dots \ \mathbf{v}_{R,J-1}\}$, para $p \in \{1, \dots, T\}$, de posse das diversas sequências O_p , treina-se um HMM para cada classe de textura.

Procedimento de classificação de texturas:

Para uma determinada sequência O :
executa-se os passos 1. 2. e 3. da etapa de treinamento,
aplica-se o algoritmo de Viterbi e seleciona-se a classe vencedora.

3.7) O Problema da Rotação de Texturas

Outro problema clássico em processamento de imagens é o reconhecimento de texturas em presença de rotação. Dito problema apresenta um grau de dificuldade mais elevado, que puramente a classificação de texturas, pela necessidade de se empregar características invariantes ao giro. Observe que não se consegue corrigir a rotação de uma textura, como é feito em determinadas abordagens de processamento de imagens, por não ser possível fixar uma referência. Desta forma, existem diversas técnicas para o tema [Haley&Manjunath, 98][Wouwer, 98][Tan, 98] que tratam a questão empregando momentos estatísticos invariantes à rotação extraídos a partir dos dados de saída de um banco de filtros. E, uma vez que a etapa de extração de características tenha sido finalizada,

assume-se vetores de distribuição normal não markoviana e desenha-se o classificador de Bayes correspondente.

Pode-se adaptar o modelo de classificação proposto para também tratar texturas que sofreram rotação. Para isto, suponha o conjunto de $R \times J$ filtros conforme apresentado anteriormente. Na presente abordagem, a redução de dimensionalidade é obtida pelo cálculo do valor esperado aplicado a cada mapa 2D de características. Assim, expressa-se cada textura por uma sequência de vetores de médias, de acordo com as expressões (3.30.a), (3.30.b), (3.30.c), (3.30.d) e (3.30.e), sendo essas funções do ângulo de orientação θ e da frequência instantânea ω_r . Tais características são variantes à rotação visto que, para uma mesma ω_r , $c_k(\omega_r, \theta_j) \neq c_k(\omega_r, \theta_m)$ com $j \neq m$, $j, m \in \{0, \dots, J-1\}$, k indexando uma das cinco equações em (3.30).

Entretanto, como observado em [Wouwer, 98] [Tan, 98], obter $c_k(\omega_r, \theta_j)$ a partir de uma textura e extrair $c_k(\omega_r, \theta_m)$ da mesma textura, agora girada de $\Delta\theta = \theta_k - \theta_m$, proporciona $c_k(\omega_r, \theta_j) = c_k(\omega_r, \theta_m)$ (no presente caso, as características que atendem a esta propriedade são dadas para $k=3,4,5$ em (3.30)). Fixando-se ω_r , e devido à propriedade do conjugado simétrico para imagens (assim, não se realiza a transformação analítica sobre a imagem) [Limt, 90], os filtros necessitam apenas de estarem dispostos para valores discretos de θ no intervalo $0 \leq \theta \leq \pi$. Isto faz com que $c_k(\omega_r, \theta_j)$ seja periódica em θ , fixando-se ω_r , podendo ser expandida em série de Fourier conforme a equação a seguir [Tan, 98]:

$$c_k(\omega_r, \theta_j) = \sum_{n=0}^{J-1} d_n e^{i2\pi jn/J}, \text{ para } d_n = \frac{1}{J} \sum_{j=0}^{J-1} c_k(\omega_r, \theta_j) e^{-i2\pi jn/J} \text{ e } i = \sqrt{-1}, \quad (3.31)$$

onde, na expressão acima $|d_n| = |d_{J-n}|$ e , proporcionando uma quantidade de $(J-1)/2$ ou $J/2$ características independentes (dependendo-se se J é ímpar ou par). Ademais, tomar o valor absoluto de d_n significa retirar a dependência angular o que pode levar a alguma degradação no resultado de classificação. Os procedimentos de treinamento e classificação são sumarizados abaixo:

Procedimento de extração de características e treinamento dos HMM's:

para uma determinada frequência radial ω_r são realizadas $K \times J$ medidas (K funcionais, como os apresentados na equação (3.30), para cada um dos J filtros orientados), para $k \in \{1, \dots, K\}$, $r \in \{1, \dots, R\}$ e $j \in \{0, \dots, J-1\}$, monta-se o vetor de evolução das características $\mathbf{v}_{r,j} = [c_1 \ c_2 \ \dots \ c_k]^T$, uma vez obtidos $\mathbf{v}_{r,j}$, toma-se todos os correspondentes α_k , para cada um dos J filtros, e corrige-se a rotação. Os novos vetores são $\mathbf{r}_{r,n}$, $n \in \{0, \dots, J/2-1\}$, constrói-se a sequência de observações $O_p = \{\mathbf{r}_{1,0}, \mathbf{r}_{1,1}, \dots, \mathbf{r}_{1,J/2-1}, \mathbf{r}_{1,0}, \dots, \mathbf{r}_{R,J/2-1}\}$, para $p = \{1, 2, \dots, T\}$, dada as diversas sequências de treinamento O_p , treina-se um HMM para cada classe de textura.

Procedimento de classificação de texturas:

Para uma determinada sequência O :
 executa-se os passos 1, 2, 3, e 4 da etapa de treinamento,
 aplica-se o algoritmo de Viterbi e seleciona-se a classe vencedora

Capítulo 4

A Modelagem Estatística do Problema

4.1) Formulação estatística do problema de classificação:

A abordagem de classificação empregada nesta tese é supervisionada e o problema se restringe a classificar M texturas distintas ω , $i \in \{1, \dots, M\}$, cada uma delas representada por uma sequência O composta de vetores de observações \mathbf{o}_t , $t \in \{1, \dots, T\}$. Ou seja, O obedece a uma distribuição de probabilidade $P(O|\omega)$ com as respectivas probabilidades a priori $P(\omega)$ para a classe ω . O objetivo da classificação é selecionar a classe ω tal que uma vez observado O , a probabilidade a posteriori $P(\omega|O)$ seja máxima, o que pode ser escrito como:

$$P(\omega_i|O) > P(\omega_k|O), \quad \forall i \neq k. \quad (4.1)$$

Empregando-se o teorema de Bayes, é possível escrever:

$$P(\omega_i | O) = \frac{P(O | \omega_i)P(\omega_i)}{P(O)}, \quad (4.2)$$

onde:

$P(\omega_i|O)$: probabilidade a posteriori da ocorrência da classe ω uma vez observado O ,

$P(O|\omega)$: distribuição condicionada da observação O dado a classe ω ,

$P(\omega)$ e $P(O)$: probabilidades marginais da classe ω e de O , respectivamente.

A equação (4.2) propõe uma abordagem de classificação a partir de probabilidades marginais e condicionais. $P(\omega)$ é a probabilidade a priori de ocorrência da classe ω onde, para o presente caso, assume-se distribuição uniforme (todas as M classes são equiprováveis) e $P(O)$ se comporta como uma constante de normalização pois diz respeito apenas aos vetores de observação. Resta, portanto, obter a probabilidade condicionada $P(O|\omega)$. Entretanto, esta é uma das grandes dificuldades na área de reconhecimento de padrões. Determinar a lei de probabilidade de cada classe não é uma tarefa simples pois, em geral, tudo de que se dispõe são medições realizadas sobre o processo que a representa. Assim sendo, por se tratar de uma abordagem supervisionada, a estimação dos termos presentes na equação (4.2) é realizada admitindo-se que os mesmos seguem um modelo de distribuição parametrizado por λ_i que os rege, permitindo que se reescreva (4.2) conforme abaixo:

$$P(\omega_i | O, \lambda_i) = \frac{P(O | \omega_i, \lambda_i) \cdot P(\omega_i | \lambda_i)}{P(O | \lambda_i)} \quad (4.3)$$

A equação (4.3) foi obtida após alguma manipulação algébrica e a suposição de que $P(\omega|O)$ passa a ser dada por $P(\omega|O, \lambda_i)$ [Morgan&Bourlard, 95]. λ_i representa o conjunto de parâmetros do modelo, parâmetros estes que serão ajustados para cada tipo de textura.

Isto posto, a classificação de uma determinada textura é obtida avaliando-se $P(\omega_i|O, \lambda_i)$ para $i \in \{1, \dots, M\}$, e decidindo-se pela classe ω que maximiza a equação (4.3). Portanto a classificação, tal como foi proposta, equivale à implementação do procedimento MAP¹²: maximização da função de probabilidade a posteriori. Uma vez estimado os M modelos, decide-se pela classe da textura, indexada por i^* , segundo a expressão:

$$i^* = \arg \max_i P(\omega_i | O, \lambda_i), \quad \forall i \in \{1, \dots, M\}. \quad (4.4)$$

Além de uma etapa de aprendizagem, onde os parâmetros do modelo são estimados a partir de um conjunto de vetores de observação de treinamento, deve-se dispor de uma metodologia que, dado um conjunto de vetores de observação de teste, seja possível avaliar o desempenho de tais modelos quanto a erros de classificação, a ser visto posteriormente.

¹² O que não impede da utilização da máxima verossimilhança para estimar $P(O|\omega_i)$ em (4.2).

4.2) A Abordagem Markoviana

A modelagem a ser seguida nesta tese baseia-se na evolução das características AM-FM de uma textura, conforme apresentado no capítulo 3. Utilizando a transformada não ortogonal das Wavelets de Gabor, as medidas locais apresentam, porém, certa dependência mesmo levando-se em conta a localização dos filtros. Portanto, o processo O resultante é composto de vetores aleatórios estatisticamente dependentes que, sob o ponto de vista da teoria das comunicações, pode ser representado por uma fonte com memória, ou fonte de Markov [Wells, 99]. Posto desta maneira, O possui uma determinada função de distribuição $P(O|\omega, \lambda_i)$ expressa, na forma fatorada, de acordo com a equação abaixo [Papoulis, 84]:

$$P(O|\omega, \lambda_i) = P(\mathbf{o}_{t=1}|\omega, \lambda_i) \prod_{t=2}^T P(\mathbf{o}_t|\mathbf{o}_{t-1}, \dots, \mathbf{o}_t, \dots, \mathbf{o}_1, \omega, \lambda_i). \quad (4.5)$$

Entretanto, trabalhar com os vetores aleatórios levando-se em conta todas as dependências possíveis pode tornar a modelagem impraticável¹³. Uma opção é reduzir a “memória” do modelo, restringindo-se o grau de dependência. Como consequência, é possível escrever (4.5) como se observa a seguir:

$$P(O|\omega, \lambda_i) = P(\mathbf{o}_{t=1}|\omega, \lambda_i) \prod_{t=2}^T P(\mathbf{o}_t|\mathbf{o}_{t-1}, \dots, \mathbf{o}_t, \dots, \mathbf{o}_{t-k}, \omega, \lambda_i), \quad (4.6)$$

para $1 \leq k < T$,

onde, na expressão (4.6), k impõe a ordem do modelo de Markov. A questão agora é como determinar k e como avaliar (4.6). Apesar da redução da memória, o problema de estimação pode ainda não ter solução satisfatória tanto em relação aos resultados finais da classificação, devido ao problema da insuficiência de dados para a obtenção de λ_i , como em relação ao custo computacional envolvido. Na seção 4.3 será discutido uma alternativa para esta situação, que se constitui nos modelos ocultos de Markov (do idioma inglês: HMM).

Antes porém, algumas considerações importantes devem ser feitas sobre as etapas de treinamento e classificação, expressas pelas equações (4.7.a) e (4.7.b) a seguir. No

¹³Cabe ressaltar que na presente tese, a modelagem é realizada sobre o espaço de características AM-FM e não diretamente sobre a imagem.

primeiro caso, deseja-se obter o modelo λ_i dado uma sequência de treinamento O_i associada à classe ω e no segundo caso deseja-se obter a classe ω dado uma sequência de observações O e todos os modelos treinados, ou seja:

$$\arg \max_{\lambda_i} P(\omega_i | O_i, \lambda_i), \quad (4.7.a)$$

$$i^* = \arg \max_i P(\omega_i | O, \lambda_i), \quad \forall i \in \{1, \dots, M\}, \quad (4.7.b)$$

onde i^* é o índice da classe vencedora.

Em ambos os procedimentos é necessário maximizar a função de probabilidade da classe condicionada à sequência de observações e ao modelo sob questão. Como a obtenção de $P(\omega_i | O_i, \lambda_i)$ e $P(\omega_i | O, \lambda_i)$ é uma tarefa complexa aplica-se, da mesma forma que em (4.2), o teorema de Bayes conforme se observa abaixo (treinamento e classificação respectivamente):

$$P(\omega_i | O_i, \lambda_i) = \frac{P(O_i | \omega_i, \lambda_i) P(\omega_i | \lambda_i)}{P(O_i | \lambda_i)}, \quad (4.8.a)$$

$$P(\omega_i | O, \lambda_i) = \frac{P(O | \omega_i, \lambda_i) P(\omega_i | \lambda_i)}{P(O | \lambda_i)}. \quad (4.8.b)$$

Analisando-se primeiramente a etapa de treinamento e substituindo-se (4.8.a) em (4.7.a) obtém-se:

$$\arg \max_{\lambda_i} \frac{P(O_i | \omega_i, \lambda_i) P(\omega_i | \lambda_i)}{P(O_i | \lambda_i)}. \quad (4.9)$$

A equação (4.9) requer a avaliação de três termos: a probabilidade de uma observação dado a classe e o modelo $P(O_i | \omega_i, \lambda_i)$, a probabilidade de uma classe dada o modelo $P(\omega_i | \lambda_i)$, e a probabilidade da observação dada o modelo $P(O_i | \lambda_i)$. Em relação a $P(\omega_i | \lambda_i)$, é possível assumir independência entre ω e λ_i pois, além de que cada modelo é treinado em separado, assume-se classes equiprováveis. O termo $P(O_i | \lambda_i)$ é frequentemente assumido como $P(O_i | \lambda_i) = P(O_i)$, para evitar complicações que seriam impostas ao HMM caso fosse calculado como tal [Morgan&Bourlard, 95]. Portanto, a expressão final que representa a etapa de treinamento pode ser simplificada segundo a equação abaixo:

$$\arg \max_{\lambda_i} P(O_i | \omega_i, \lambda_i). \quad (4.10)$$

Ao assumir (4.10) como uma função de verossimilhança, sua maximização pode ser implementada em computadores digitais através de algoritmos tais como Newton-Raphson ou o EM (contração das iniciais das palavras *Expectation* e *Maximization*)¹⁴. O primeiro método, apesar de fornecer bons resultados, exige a inversão de uma matriz tipo Hessiana, obtida pela derivada segunda da variável a ser maximizada. Assim sendo, por não haver garantias de que essa matriz seja não singular ou pelo menos bem condicionada, a melhor opção é a utilização do método do EM. Especialmente para o caso de treinamento de HMM's, a maximização da sequência de vetores de observação é obtida por uma variante do EM denominada algoritmo de Reestimação de Baum-Welch ou *forward – backward*.

No que tange à etapa de classificação, equação (4.8.b), as considerações apresentadas na etapa de treinamento, supondo que $P(\omega_i | \lambda_i) = P(\omega_i)$ e $P(\omega_i) = P(\omega_j)$, $\forall i, j$, continuam válidas. Especificamente para a classificação, assume-se $P(O | \lambda_i) = P(O)$ uma vez que os modelos já se encontram pré-treinados e os parâmetros relativos a λ_i são fixos. Classificar, portanto, significa reescrever (4.7.b) segundo a expressão abaixo:

$$i = \arg \max_i P(O | \omega_i, \lambda_i), \quad \forall i \in \{1, \dots, M\}. \quad (4.11)$$

Dado que uma textura é representada pela sequência de vetores de observações, a determinação de classe vencedora em (4.11) é realizada por meio do algoritmo de Viterbi, a ser apresentado mais adiante.

4.3) Os modelos ocultos de Markov

Como visto anteriormente, assumir para um modelo uma ordem elevada pode inviabilizar sua estimativa. Por outro lado, para o caso de uma fonte markoviana de ordem 1, cada estado possui uma relação um para um, determinística, com uma observação a ele associado. Assim, o uso de modelos de Markov de ordem 1 fica muito restrito e impede que se obtenha as variações naturais das observações. Entretanto, o problema pode ser visto de

¹⁴ Como não se conhece a forma das funções a priori não é possível empregar diretamente os resultados clássicos de estimadores de parâmetros pela máxima verossimilhança.

outra forma: supõe-se modelos de Markov de ordem 1 para os estados e permite-se que cada um deles “gere” os vetores de observações segundo uma função de distribuição de probabilidade. Desta forma, o processo relativo às observações fica atrelado a um segundo processo, regido pela propriedade de Markov. Observe que não se supõe que os vetores de observação, que representam justamente os dados a serem modelados – processo visível, possuam característica markoviana; apenas a sequência de estados, processo oculto, é modelado como tal (processo de Markov de ordem $k = 1$).

Com isto, dada uma observação \mathbf{o}_t , não é possível associá-la a um determinado estado, como o é para cadeias markovianas tradicionais. A relaxação da rígida relação entre estado/observação nas cadeias de Markov para estado/função de distribuição de observação confere a este novo modelo, denominado de modelo oculto de Markov ou HMM, a capacidade de capturar as diferentes nuances do padrão a ser representado.

Os modelos ocultos de Markov são comumente empregados na representação de sequências de dados, sendo bastante utilizados em processamento de voz [Rabiner, 89] e imagens [Aas *et. al.*, 99]. Formalizando-se o exposto acima, tais modelos se caracterizam, essencialmente, pela evolução de dois processos estocásticos a saber:

um processo não observável relativo à propriedade markoviana de evolução temporal de estados;

um processo observável relativo à sequência de dados em si, sequência esta baseada no processo não observável.

A combinação destes dois processos confere poder de análise aos HMM's. Por exemplo, em aplicações de processamento de voz [Rabiner, 89], uma determinada elocução, representada por uma sequência de vetores de características O , é extraída desse sinal onde cada observação \mathbf{o}_t refere-se a uma medição multivariada relativa ao instante de tempo t . Um procedimento semelhante é realizado quando se aplica dita técnica no reconhecimento de palavras manuscritas [Aas *et. al.*, 99]. Ou seja, em ambos os casos o sinal (1D para voz e 2D para palavras manuscritas) é suposto evoluir de forma estocástica, cabendo aos HMM's descrevê-los. Via de regra, esses sinais são particionados (no tempo ou no espaço) e técnicas de extração de características são aplicadas a fim de se reduzir a

dimensionalidade sem perder a capacidade de representação. Como nos exemplos acima os sinais evoluem naturalmente, as características extraídas (sob a forma de vetores de observação) são tratadas segundo uma sequência de dados e sobre essas é que se aplica a modelagem por meio das cadeias ocultas de Markov.

Algumas modificações devem ser feitas na modelagem estatística exposta anteriormente com o intuito de implementá-la via HMM. Para o modelo proposto de caracterização das texturas: sequência de vetores de observação extraídos segundo características locais AM–FM, as equações das probabilidades passam a ter um parâmetro a mais, dado pelos estados da cadeia de Markov. Entretanto, como foi apresentado nas equações (4.9) e (4.10) o termo chave das etapas de treinamento e classificação é dado por $P(O|\omega, \lambda_i)$.

Antes de mais nada, deve-se formalizar a apresentação das variáveis empregadas e a notação a ser utilizada:

Sequência O constituída de T vetores de observações \mathbf{o}_t (vetores – coluna)

$$O = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t, \dots, \mathbf{o}_T\}, t \in \{1, \dots, T\}. \quad (4.12)$$

Um conjunto de estados $q_j^t, j \in \{1, \dots, N\}$ e $t \in \{1, \dots, T\}$. A notação para os estados pode ser simplificada, conforme se discutirá a seguir.

O vetor de probabilidades iniciais dos estados

$$\boldsymbol{\pi} = [P(q_1|t=1), P(q_2|t=1), \dots, P(q_N|t=1)]^{Tr}, \text{ para } Tr = \text{transposto}. \quad (4.13)$$

Uma matriz de probabilidade de transição de estados A , de dimensões $N \times N$, onde cada um de seus elementos é representado por $a_{jl}, j, l \in \{1, \dots, N\}$. O modelo empregado nesta tese assume a propriedade da homogeneidade: dado dois estados, a probabilidade de transição entre eles independe sequência de evolução (amostras em diferentes frequências radiais) em que a mesma ocorre ou seja, supondo-se N estados q_j^t :

$$P(q_{j+1}^{t+1} | q_j^t) = P(q_{j+1} | q_j). \quad (4.14)$$

Tal propriedade assume condição estacionária para a cadeia de estados. É possível trabalhar com modelos não estacionários mas o custo envolvido eleva-se em demasia.

Um conjunto relativo às probabilidades das observações emitidas em cada estado: $B=\{P(\mathbf{o}_t | q_j)\}$, $t \in \{1, \dots, T\}$ e $j \in \{1, \dots, N\}$. Como o modelo apresentado é contínuo, $P(\mathbf{o}_t | q_j)$ é genericamente dada por um modelo de misturas. Na literatura é comum o emprego da notação $b_j(\mathbf{o}_t)$ em referência a modelos discretos ou seja, ao invés de trabalhar com funções de distribuição $P(\mathbf{o}_t | q_j)$, $b_j(\mathbf{o}_t)$ diz respeito à probabilidade de símbolos emitidos pelo modelo. Nesta tese, apesar de tratar exclusivamente de modelos contínuos, empregar-se-á $b_j(\mathbf{o}_t)$ como referência a $P(\mathbf{o}_t | q_j)$, a fim de simplificar a notação.

Para um determinado HMM, usa-se representar seu modelo pela tripla $\lambda=(A,B,\pi)$. E, além da seleção prévia do número N de estados, deve-se também determinar suas interconexões de acordo com o problema a ser tratado, ficando assim refletido nos elementos da matriz estocástica A a topologia escolhida (ou seja $a_{ij} \neq 0$). Por exemplo, em processamento de voz [Rabiner, 89] é comum o uso do modelo esquerda-direita, conforme se observa na figura 4.1.a, obtendo-se como consequência uma matriz A esparsa. No presente caso optou-se pelo uso da topologia ergódica, figura 4.1.b, onde permite-se transição entre todos os estados, obtendo-se uma matriz A cheia.

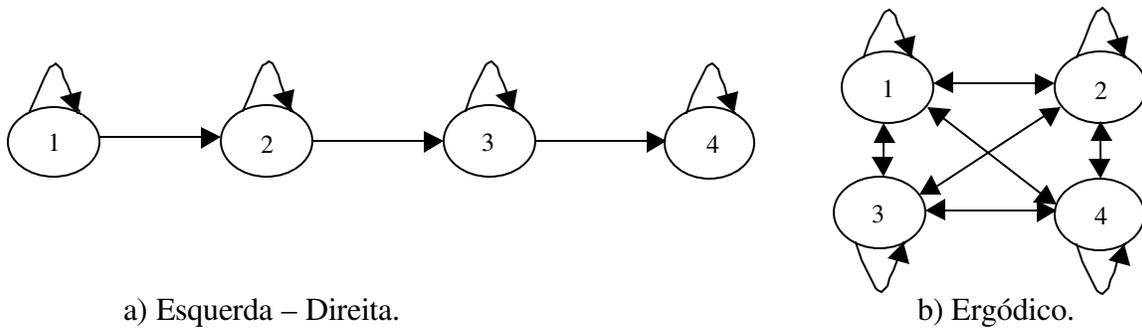


Figura 4.1: Dois exemplos de topologias para HMM's. O modelo b) foi utilizado nesta tese.

De acordo com o exposto sobre HMM e uma vez formalizada a notação, é importante explicitar as relações entre os vetores de observações e os estados do modelo. Isto é necessário visto que todos os parâmetros relativos aos estados, que constituem o processo oculto, serão extraídos a partir de O (processo visível). Assumindo-se que o modelo é de ordem 1, é possível escrever:

$$P(\mathbf{o}_t | q_j, \mathbf{o}_{t-1}, \mathbf{o}_{t-2}, \dots, \mathbf{o}_1) = P(\mathbf{o}_t | q_j), \quad (4.15.a)$$

$$P(q_{j+1} | q_j, \mathbf{o}_{t-1}, \mathbf{o}_{t-2}, \dots, \mathbf{o}_1) = P(q_{j+1} | q_j). \quad (4.15.b)$$

Ou seja, o passado de toda a cadeia fica resumido ao estado q_j , tanto no que se refere à emissão de probabilidade como no que se refere à transição para um novo estado. A partir de (4.15.a) e (4.15.b) e da independência assumida no modelo¹⁵, é possível chegar ao valor da probabilidade da cadeia em emitir uma determinada sequência O . Para isto, primeiramente, constrói-se a função de probabilidade conjunta entre os estados (processo não observável) e a sequência O (processo observável) conforme abaixo:

$$P(O, Q | \lambda) = P(q^1) \prod_{t=1}^{T-1} P(q^{t+1} | q^t) \prod_{t=1}^T P(\mathbf{o}_t | q^t), \quad (4.16)$$

onde Q se refere a uma sequência de estados. Observe que a emissão de probabilidade de \mathbf{o}_t em (4.16) depende apenas do estado atual, em contrapartida com o outro modelo possível onde a emissão de \mathbf{o}_t depende do estado atual e do anterior.

Para encontrar $P(O|\lambda)$ basta obter a função marginal de (4.16) aplicando-se:

$$P(O | \lambda) = \sum_{\{Q_s\}} P(O, Q_s | \lambda), \quad (4.17)$$

onde $\{Q_s\}$ representa o conjunto de todas as sequências de estados possíveis.

Entretanto, avaliar a expressão acima constitui exatamente o primeiro dos três problemas clássicos relativos à abordagem HMM: dada uma sequência de observações O e um modelo λ como calcular de forma eficiente $P(O|\lambda)$? A dificuldade imposta em (4.17) diz respeito à complexidade computacional envolvida. Como solução para este problema foram propostos dois algoritmos, de complexidade reduzida, denominados de *forward* e *backward*, sendo que ambas as soluções chegam ao mesmo resultado. A abordagem *forward* estima a probabilidade a partir do instante de tempo $t = 1$ propagando-a para frente (daí o nome *forward*) e a abordagem *backward* o faz a partir do instante $t = T$ propagando-a para trás (*backward*). Os detalhes de implementação podem ser encontrados no Apêndice A4.3.

A etapa de treinamento continua baseada no algoritmo (4.10), agora porém com a introdução do processo oculto relativo aos estados. Assim sendo, pode-se escrever:

$$\arg \max_{\lambda_i} P(O_i | \omega_i, \lambda_i), \text{ que equivale a } \arg \max_{\lambda_i} \sum_{\{Q_s\}} P(O_i, Q_s | \omega_i, \lambda_i), \quad (4.18)$$

onde $\{Q_s\}$ representa o conjunto de todas as sequências de estados possíveis.

O processo de treinamento constitui o terceiro problema clássico dos modelos HMM, além de ser considerado como o mais complexo [Rabiner, 89]. Não existe um procedimento analítico para a estimação dos parâmetros $\lambda=(A,B,\pi)$ a fim de se maximizar $P(O|\lambda)$. Como exposto anteriormente, os métodos disponíveis podem ser derivados ou do algoritmo EM ou de alguma técnica de gradiente. A solução mais empregada em cadeias ocultas de Markov foi proposta por Baum *et. al.* (referenciada em [Rabiner, 89]) e baseia-se no procedimento do EM, sendo batizada de algoritmo de re-estimação de Baum-Welch. Esta técnica também é conhecida como *forward-backward* em alusão ao uso dos dois algoritmos utilizados para a solução do primeiro problema clássico. O Apêndice A4.3 apresenta mais detalhes sobre o algoritmo.

Na etapa de classificação deve-se associar uma classe a uma dada sequência de observações O , segundo o exposto pela equação (4.11). Mas como agora introduziu-se a sequência de estados ao modelo, esta etapa é obtida através dos seguintes passos:

avalia-se para cada modelo λ_i a melhor sequência de estados que explica O , aqui assumida como Q_i^* ,

para cada Q_i^* (que representa um determinado modelo λ_i) obtém-se a probabilidade de emissão de O ,

classifica-se finalmente O associando-a ao modelo que apresentar a sequência de estados mais provável.

De acordo com o algoritmo de classificação expresso em (4.11), classificar equivale a:

$$i^* = \arg \max_i P(O | \omega_i, \lambda_i, Q_i^*), \quad \forall i \in \{1, \dots, M\}. \quad (4.19)$$

¹⁵ O que caracteriza um dos pontos fracos dos modelos HMM.

Ou seja, para classificar primeiramente deve-se avaliar a sequência de estado ótima para cada modelo e justamente esta avaliação constitui-se no segundo problema clássico dos HMM: dado uma sequência de observações O e um modelo λ , como inferir a sequência de estados ótima que gera O ?

4.4) Considerações a cerca do modelo empregado e a estrutura de treinamento

É possível atribuir sentido físico aos estados, associando a cada um deles (ou mesmo a um grupo deles) um significado particular, a partir do conhecimento adquirido sobre o problema. Por exemplo, para reconhecimento de voz um grupo de estados pode ser utilizado na representação de um determinado fonema ou, como em reconhecimento de palavras manuscritas, associa-se estados a primitivas da escrita como linhas retas, curvas, etc.. Dependendo da complexidade do problema, pode-se intercalar vários modelos de HMM na obtenção da solução. Para os exemplos acima, a interconexão de vários HMM modela de palavras a frases. Mas tudo depende da associação entre os estados e seus significados. Para a área de processamento de voz diz-se tratar de um procedimento de transcrição fonética e para a área de reconhecimento de palavras manuscritas realiza-se a segmentação em suas componentes primitivas. Devido ao grau de complexidade envolvido em ditas tarefas, ambos os procedimentos são realizados com o acompanhamento do homem a fim de se chegar nos resultados desejados. Em processamento de texturas, o equivalente às primitivas ainda não é consenso na área. Um dos primeiros trabalhos a se preocupar com o tema data de 1983 [Julesz&Bergen, 83] propondo os *textons* como elemento básico da textura. Entretanto, não existe procedimento padrão a cerca de como detectá-lo, apesar de algumas tentativas discutidas no capítulo 2 desta tese.

Portanto, a fim de atribuir significado físico aos estados, optou-se por um procedimento padrão de agrupamento das características AM-FM, empregando-se para isto a métrica euclidiana. Inicialmente, dado T vetores e N estados, atribui-se os T/N primeiros vetores ao estado 1 e assim sucessivamente. Uma vez terminado o procedimento de agrupamento, cada estado fica responsável por emitir vetores de observações segundo uma

determinada função de probabilidades e sua função de densidade de probabilidade é expressa por meio de um modelo de misturas a fim de capturar a diversidade do agrupamento das observações. Uma breve introdução ao tema de modelo de misturas pode ser vista no Apêndice A4.2.

O processo de treinamento exposto em (4.18) exige, conforme [Juang&Rabiner, 90], que todas as sequências de estados sejam consideradas no cálculo da verossimilhança, o que pode apresentar dificuldades numéricas. Uma outra forma de proceder o treinamento de HMM's é conhecida como "K-means Segmental". O procedimento continua baseado no algoritmo (4.10) modificando-se porém a forma de avaliação de $P(O_i|\omega, \lambda_i)$, aqui aproximada por $P^{\max}(O_i|\omega, \lambda_i)$, verossimilhança dos estados otimizados, e dada por:

$$P^{\max}(O_i|\omega, \lambda_i) = \max_{Q_s} P(O_i, Q_s|\omega, \lambda_i). \quad (4.20)$$

Ou seja, o cálculo da verossimilhança segundo (4.18) é modificado aplicando-se o operador "max" (máximo), ao contrário da somatória, sobre todas as sequências de estado possíveis. Desta forma, a mais provável sequência de estados – aquela que melhor "explica" a sequência O , é levada em conta [Morgan&Bourlard, 95]. O novo procedimento de treinamento passa a ser dado conforme o algoritmo abaixo:

$$\arg \max_{\lambda_i} P(O_i|\omega, \lambda_i), \text{ que equivale a } \arg \max_{\lambda_i} \{ \max_{Q_s} P(O_i, Q_s|\omega, \lambda_i) \}. \quad (4.21)$$

A introdução do operador não linear "max", implementado pelo método de Viterbi, dificulta a análise de convergência do método (sua prova encontra-se em [Juang&Rabiner, 90]). Outrossim, não existe um formalismo analítico que avalia (4.21) de maneira direta. Para isto, a operação se procede em passos de iteração que se alternam entre determinação da melhor sequência Q_i^* , seguida da determinação dos parâmetros da melhor sequência. O algoritmo final pode ser reescrito conforme abaixo:

$$\bar{\lambda}_i = \arg \max_{\lambda_i} \{ \max_{Q_s} P(O_i, Q_s|\omega, \lambda_i) \}. \quad (4.22)$$

O processo começa a partir de um modelo inicial λ_i e as iterações se sucedem com a substituição de λ_i por $\bar{\lambda}_i$ até que se atenda a uma condição de parada. O termo K-means

empregado na denominação do método diz respeito à avaliação do modelo de misturas. Durante a aplicação de (4.22), uma vez obtida a sequência ótima Q_i^* , resta determinar a função de densidade de probabilidade de cada um de seus estados. Como supõe-se o modelo dado por misturas de gaussianas, aplica-se o K-means sobre o conjunto de observações associadas a um determinado estado, calculando-se os pesos, os vetores média e as matrizes de variância/covariância. O algoritmo completo do K-means segmental é apresentado a seguir.

Algoritmo K-means Segmental:

Passo 1: Agrupamento dos vetores de observações.

Passo 2: Cálculo da probabilidade inicial π_i :

$$\pi_i = \frac{\text{Número de ocorrências } \{\mathbf{o}_t \in q_j\}}{\text{Número de amostras de treinamento}}, \text{ para } 1 \leq j \leq N. \quad (4.23)$$

Passo 3: Cálculo da transição de estados – determinação da matriz A:

$$a_{ij} = \frac{\text{Número de transições } q_j \rightarrow q_i}{\text{Número de transições a partir de } q_j}, \text{ para } 1 \leq j, l \leq N. \quad (4.24)$$

Passo 4: Cálculo dos parâmetros do modelo de misturas de cada estado:

Assume-se M_x misturas em cada estado, indexadas por m .

Estimação dos vetores média $\mathbf{v}_{j,m}$ (modelo de mistura m do estado q_j):

$$\mathbf{v}_{j,m} = \frac{1}{N_{j,m}} \sum_{\mathbf{o}_t \in \text{estado } j, \text{ modelo } m} \mathbf{o}_t, \quad (4.25)$$

para $N_{j,m}$ ocorrências de \mathbf{o}_t associado ao modelo m do estado q_j .

Estimação das $C_{j,m}$ matrizes de variância/covariância:

$$C_{j,m} = \frac{1}{N_{j,m}} \sum_{\mathbf{o}_t \in j,m} (\mathbf{o}_t - \mathbf{v}_{j,m})(\mathbf{o}_t - \mathbf{v}_{j,m})^{Tr}, \text{ onde } Tr \text{ indica transposição matricial.} \quad (4.26)$$

A estimação do peso $\mathbf{w}_{j,m}$ das misturas é baseado no número de observações alocadas para cada componente

$$\mathbf{w}_{j,m} = \frac{\text{número de ocorrências de } \mathbf{o}_t \text{ associado ao modelo } m \text{ do estado } q_j}{\text{número de ocorrências de } \mathbf{o}_t \text{ associado estado } q_j}. \quad (4.27)$$

Passo 5: Obter a sequência ótima Q_i^* atual através do algoritmo de Viterbi.

Passo 6: Se a sequência ótima Q_i^* da iteração anterior for igual a Q_i^* atual então finaliza-se o processo de aprendizagem. Caso contrário, atualiza-se a associação de estados e retorna-se ao Passo 2.

Capítulo 5

Testes e Resultados

) O Ambiente Desenvolvido

Todo o sistema foi desenvolvido sobre o ambiente MATLABTM, aproveitando-se as facilidades de implementação por ele proporcionadas. Este ambiente, criado na década de 70 nas universidades norte-americanas, permite diminuir o tempo de “prototipagem” de sistemas complexos. Hoje em dia, com a adição de diversas ferramentas, organizadas em caixas de trabalho denominadas “TOOLBOX”, é possível realizar inclusive aplicações profissionais.

No que tange à maneira de escrever um programa em MATLABTM deve-se estar atento para evitar funções de execução lenta. Como tal linguagem é de propósito geral, é comum basear-se no estilo de outras como FORTRAN, C ou PASCAL o que acarreta numa diminuição de desempenho. Melhores resultados são obtidos programando-se na forma matricial evitando-se o uso de laços sobretudo para tais operações. Outra de suas vantagens refere-se à visualização de dados em geral, possível através de um conjunto de recursos gráficos de fácil utilização. O MATLABTM também possui funções de leitura e escrita especiais para arquivos (com diversos filtros para tipos distintos de imagens), permitindo a

criação e a utilização de um banco de imagens. Por todas as facilidades enumeradas acima, optou-se pelo desenvolvimento do sistema em tal ambiente.

5.2) O Banco de Imagens de Texturas

Talvez, um dos mais famosos e utilizados banco de testes em processamento de imagens seja o álbum de Phill Brodatz [Brodatz, 66]. Muitas de suas fotos sobre texturas são empregadas em diversos artigos científicos, tornando-se praticamente um padrão internacional. Seu uso na literatura técnica permite comparar o resultado de diferentes abordagens para os temas de classificação e segmentação. O álbum é composto de 112 texturas, listadas no apêndice A5.1, algumas delas repetidas em diferentes escalas. Na figura A5.1 é apresentado um mosaico contendo todas as texturas. Pela diversidade das imagens (grão fino, grão grosso, homogêneas, não homogêneas, fractais, etc.), vários autores selecionam dentre as 112 as mais adequadas para o tipo de técnica empregada [Mojsilovic *et. al.*, 00][Unser, 93][Chang&Kuo, 93][Tan, 98][Manjunath&Ma, 96]. Neste trabalho, optou-se por avaliar também o resultado sobre todas as imagens.

Portanto, o banco completo de imagens implementado consta das 112 texturas, cada uma digitalizada em 512×512 pixels, 8 bits por pixel (em tons de cinza), por meio de um *scanner* de resolução de 200 dpi. Devido ao número limitado de exemplos por classe, as amostras de uma de textura possuem ou 64×64 ou 128×128 pixels, conforme o tipo de testes a ser realizado. Independente das dimensões das amostras, todas foram tomadas sem superposição, perfazendo por classe de textura um total de 64 exemplares para o caso de 64×64 pixels e 16 exemplares para o caso de 128×128 pixels.

A escolha destas dimensões diz respeito à forma de tratamento das características extraídas. Determinadas texturas de grão grosso exigem amostras de dimensões elevadas para que seja eficiente a representação de sua classe. Entretanto, um aumento no tamanho das imagens acarreta uma diminuição na quantidade de exemplos tanto para treinamento como para testes. Como consequência, uma menor quantidade de exemplos para treinamento pode inviabilizar a estimação da matriz de variância/covariância de cada estado do HMM, já que se trabalha com matriz cheia. A seleção dos valores 64×64 e 128×128 foi

portanto obtida através de realização de testes onde a dimensão mínima foi limitada inferiormente pela “grossura” da textura e o valor superior pelo número de exemplos restantes para testes. Por último, todas as imagens utilizadas nos experimentos possuem média zero.

Uma vez definida estas dimensões, admite-se que metade do número de amostras é empregada para treinamento e outra metade para a avaliação dos resultados (exceção a ser feita para o caso de rotação de texturas). Este tipo de procedimento foi escolhido por estar em concordância com aquele empregado na literatura e assim permitir a comparação, com mais propriedade, entre diferentes abordagens. A partir do banco de dados completo, foram criados novos bancos, compostos de sub-grupos das 112 texturas. Este procedimento visa comparar os resultados obtidos pela presente técnica com outros encontrados na literatura.

Os bancos utilizados seguem abaixo:

Banco 1: D9, D12, D15, D17, D19, D24, D29, D38, D68, D84, D92, D94 e D112;

Banco 2: D3, D4, D6, D9, D11, D16, D19, D21, D24, D29, D34, D36, D52, D53, D55, D57, D65, D68, D74, D77, D78, D79, D82, D83, D84, D92, D95, D102, D103 e D105;

Banco 3: as 112 texturas do álbum de Brodatz (o banco completo);

Banco 4: Todas as texturas do banco 1, aqui giradas no sentido anti-horário em ângulos de 30°, 45°, 60°, 90°, 120° e 150°;

Banco 5: D10, D12, D15, D17, D22, D24, D34, D37, D53, D56, D65, D68, D80, D82 e D84, giradas no sentido anti-horário em ângulos de 30°, 45°, 60°, 90°, 120° e 150°.

Novamente, as amostras extraídas a partir dos bancos, para compor o conjunto de imagens de treinamento e testes, foram feitas sem superposição. Também evitou-se girar as imagens artificialmente no computador devido ao “aspecto serrilhado”, efeito que acontece nesses casos. Optou-se por empregar no treinamento apenas imagens não giradas (0°) e utilizar as imagens giradas apenas para avaliar os resultados. Experimentos preliminares indicaram que, para o problema de rotação de texturas, uma boa proporção entre a quantidade de exemplos para treinamento e testes é de 40 amostras para treinamento e 32 amostras para testes.

5.3) Banco de Filtros de Gabor

Como discutiu-se no capítulo 3, a disposição espectral dos filtros de Gabor deve ser tal que permita a extração das características AM–FM. A detecção de tais características depende do quão boa é a localização dos filtros que, por serem não ortogonais, aumentam em muito a complexidade computacional. Assim sendo, foi necessário estabelecer uma relação de compromisso entre velocidade de processamento e o resultado de classificação. O objetivo proposto foi de obter uma boa classificação evitando-se um número muito grande de frequências radiais e orientações. Testes iniciais mostraram que o aumento excessivo desses parâmetros não implica em um aumento proporcional da taxa de classificação. Em contrapartida, um “ladrilhamento” excessivo do espaço de fase, devido à não ortogonalidade da transformação, incrementa consideravelmente a complexidade computacional. Após diversas experimentações optou-se por utilizar 4 valores distintos de frequência radial e 8 orientações, entre 0 e π , conforme apresentado na figura 3.6 do capítulo 3.

5.3.1) Os parâmetros do HMM

Foram avaliadas duas configurações: HMM’s com 8 estados e uma gaussiana por estado e HMM’s com 4 estados e 2 gaussianas por estado. No segundo caso, empregou-se o algoritmo do EM para a estimação do modelo de misturas. O uso de mais gaussianas por estado fica limitado à quantidade de exemplos por classe. Uma vez que cada função de densidade de probabilidade é descrita por um vetor de médias e uma matriz de variância/covariância, a estimação dessa última depende da quantidade de amostras disponíveis. Para vetores de dimensão N são necessários $N+1$ exemplos por classes na estimação. Consequentemente, para vetor de dimensão 3 são necessários 4 exemplos distintos. Testes com 3 gaussianas não se mostraram efetivos provavelmente devido à baixa quantidade de vetores para estimação da referida matriz (lembrando que em se tratando das imagens de 128×128 , cada tipo de textura dispõe de 16 exemplos: 8 para treinamento e 8 para classificação).

Vale mencionar que uma abordagem muito comum em reconhecimento de fala, e que poderia ser adaptada neste caso, é a aplicação de uma transformação ortogonal no

intuito de reduzir a magnitude dos termos relativos à covariância na matriz de variância/covariância de forma a assumí-la diagonal . Isto supõe um processo estacionário autoregressivo que é descorrelacionado por dita transformação. O interessante é que a transformação empregada, MEL Cepstral, cumpre dupla função: descorrelaciona os dados e permite a separação entre a fonte de excitação do filtro equivalente e o aparato vocal [Young *et. al.*, 97].

5.4) Resultados Obtidos

O primeiro experimento refere-se classificação das 13 texturas, banco 1, apresentadas na figura 5.1. A tabela 5.1 apresenta os resultados obtidos. Foram testadas diferentes configurações para o HMM, avaliando-se o emprego de:

K-Means Segmental (K-Means Seg.) apenas,

K-Means Segmental (K-Means Seg.) seguido de HMM.

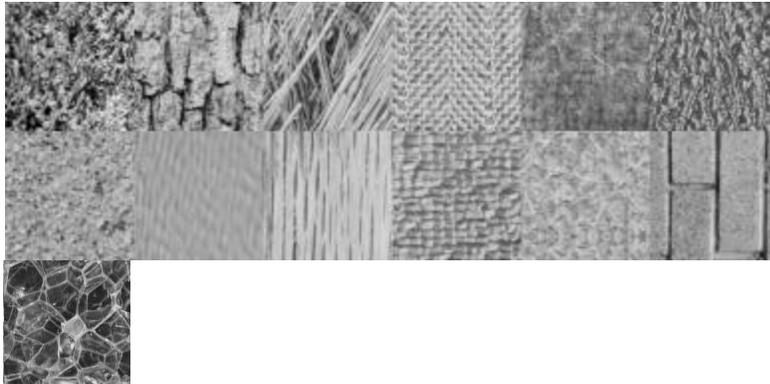


Figura 5.1: Banco 1. Da esquerda para direita, de cima para baixo: D9, D12, D15, D17, D19, D24, D29, D38, D68, D84, D92, D94 e D112, amostras de 128×128 pixels.

Para o uso de mais de uma gaussiana (gaus.), foi utilizado o algoritmo do EM na estimação de seus parâmetros. As características c_1 , c_2 e c_3 , listadas na tabela, são as apresentadas no capítulo 3, equações (3.30.a), (3.30.b) e (3.30.c) respectivamente.

Teste	K-Means Seg.	HMM	Características	Dimensões	Taxa de Acertos
5.1.1	4 estados/2 gaus.	Não	c_1, c_2, c_3	64×64	98,8%
5.1.2	4 estados/2 gaus.	Sim	c_1, c_2, c_3	128×128	100,0%

Tabela 5.1: Teste inicial para o grupo das 13 texturas.

No teste 5.1.1 (texturas de 64×64), 32 amostras foram utilizadas para treinamento e as restantes 32 para classificação. Já no teste 5.1.2 (texturas de 128×128) 8 amostras foram utilizadas para treinamento e 8 para testes. Para efeito de comparação, utilizando 12 das 13 texturas do banco 1 (excluindo-se D112) e imagens de dimensões de 256×256 , o que acentua texturas de grão fino, em [Unser, 93] Unser empregou decomposição por Wavelets biortogonais (extraindo momentos estatísticos de cada canal, e classificador de mínima distância) e obteve uma taxa de acerto de 98,83% para 64 amostras de treinamento e 64 para testes.

A importância deste primeiro banco deve-se ao ajuste inicial dos parâmetros do sistema. Tal procedimento foi necessário pois, como visto anteriormente, existem muitas variáveis tanto no que se refere ao HMM como também nos filtros de Gabor. Para o HMM foi necessário a definição do número de estados, número este igual à quantidade de agrupamentos realizados previamente. Primeiramente, empregou-se o algoritmo sugerido em [Wu&Wei, 96] para a implementação da função agrupamento, onde inicia-se com um número N de grupos até reduzir-se para os M grupos desejados ($M = \text{número de estados e } N > M$). Entretanto, na presente abordagem, esta forma de implementação não se mostrou muito eficiente. Acredita-se que este procedimento tira do HMM a capacidade de capturar a distribuição estatística dos dados. Na tabela 5.1 o melhor resultado foi obtido assumindo-se mais de uma função de densidade de probabilidade por estado e permitindo-se que o EM ajustasse os parâmetros da mistura de modelos (teste 5.1.2).

O segundo experimento emprega o banco 2, onde classificam-se 30 texturas, as mesmas utilizadas em [Chang&Kuo, 93]. Neste trabalho, os autores comparam diversas técnicas tais como Wavelets, Wavelets Packets, filtros de Gabor, DCT, DHT e filtros de Laws. O treinamento é realizado para amostras de 512×512 pixels. Na aplicação de cada técnica acima, extrai-se momentos estatísticos e utiliza-se uma função discriminante de Bayes, uma vez que se assumiu vetores de características multivariados e de distribuição Gaussiana. Isto posto, o melhor resultado obtido em [Chang&Kuo, 93] foi para estruturas do tipo Wavelet Packets seguido de algoritmo de poda, com taxa de acerto de 99,6% e, o pior resultado é obtido a partir dos filtros de Gabor cuja taxa de acertos foi de 93,5 %. Com

a presente abordagem, e sobre o banco 2, obteve-se taxa de acerto de 100% utilizando-se amostras de dimensões 128×128 . Acredita-se que a avaliação das características AM-FM é mais eficiente que a mera extração de momentos estatísticos diretamente aplicados sobre canais, mesmo que se empregue transformações em múltiplas resoluções como Wavelets. Outro ponto que contribuiu para a taxa de 100% é a utilização de um classificador tipo HMM que confere maior flexibilidade de adaptação ao real modelo dos dados.

Como último resultado sobre as amostras do banco 2, em [Mojsilovic *et. al.*, 00], empregando-se 23 das 30 possíveis, amostras de 128×128 pixels, e trabalhando-se com seleção da melhor base de Wavelets (que no caso foi a base biortogonal 3.3 com filtros duais) obteve-se uma taxa de acerto máxima de 99,35%. Vale ressaltar que o objetivo dos autores é avaliar de forma mais criteriosa a escolha do banco de filtros para classificação de texturas, sob o ponto de vista de transformada de Wavelets. O vetor de características empregado é composto pelas energias das sub-bandas de decomposição e a classificação foi feita pelo classificador da menor distância. Também estudou-se o desempenho da classificação tomando-se apenas aquelas sub-bandas de maior energia, com resultados que vão desde 61,70% a 97,39% (a taxa de 99,35% foi obtida medindo-se a energia sobre todas as sub-bandas).

O próximo experimento refere-se à classificação das 112 texturas do banco de Brodatz, banco 3, e os resultados são apresentados na tabela 5.2. Observa-se que o uso do algoritmo do EM elevou a taxa de acertos tanto para o caso de 64×64 pixels como também para 128×128 pixels. Em amostras de dimensão 64×64 pixels a melhor taxa obtida é dado pelo teste 4. Para efeito de comparação em [Manjunath&Ma, 96], trabalhando-se com 109 amostras (onde excluiu-se D31, D32 e D99), obteve-se uma taxa de 72,13%. Neste artigo, a abordagem utiliza filtros de Gabor na forma piramidal, dispostos em 4 frequências radiais e 6 orientações distintas (entre 0 e π), e aplica sobre os mapas 2D, oriundos de cada canal, momentos estatísticos como média e variância. A técnica segue assumindo-se pdf gaussiana e implementando-se a função discriminante equivalente.

Teste	K-Means Seg.	HMM	Características	Dimensões	Taxa de Acertos
5.2.1	8 estados/1 gaus.	Não	c1, c2, c3	64 × 64	75,84 %
5.2.2	8 estados/1 gaus.	Sim	c1, c2, c3	64 × 64	75,64%
5.2.3	4 estados/2 gaus.	Não	c1, c2, c3	64 × 64	75,25% (sem EM)
5.2.4	4 estados/2 gaus.	Sim	c1, c2, c3	64 × 64	76,87%
5.2.5	4 estados/2 gaus.	Não	c1, c2, c3	128 × 128	85,94%
5.2.6	4 estados/2 gaus.	Não	c1, c2, c3	128 × 128	82,59% (sem EM)

Tabela 5.2: Configurações e taxa de acertos para classificar as 112 texturas.

Na tabela 5.3, mostrada a seguir, é apresentada a taxa de acertos obtida para cada textura (em %), resultado este referente ao teste 5.2.5 onde se conseguiu uma taxa de classificação total igual a 85,94%.

Tabela 5.3: Taxa de acertos para cada classe de textura, em %.

D1 = 100.0	D29 = 100.0	D57 = 100.0	D85 = 100.0
D2 = 100.0	D30 = 37.5	D58 = 62.5	D86 = 100.0
D3 = 100.0	D31 = 75.0	D59 = 75.0	D87 = 100.0
D4 = 100.0	D32 = 87.5	D60 = 100.0	D88 = 100.0
D5 = 50.0	D33 = 87.5	D61 = 75.0	D89 = 62.5
D6 = 100.0	D34 = 100.0	D62 = 62.5	D90 = 100.0
D7 = 25.0	D35 = 100.0	D63 = 50.0	D91 = 87.5
D8 = 62.5	D36 = 87.5	D64 = 75.0	D92 = 100.0
D9 = 100.0	D37 = 62.5	D65 = 100.0	D93 = 100.0
D10 = 100.0	D38 = 100.0	D66 = 100.0	D94 = 87.5
D11 = 100.0	D39 = 75.0	D67 = 100.0	D95 = 100.0
D12 = 100.0	D40 = 100.0	D68 = 100.0	D96 = 100.0
D13 = 50.0	D41 = 100.0	D69 = 100.0	D97 = 100.0
D14 = 87.5	D42 = 100.0	D70 = 100.0	D98 = 87.5
D15 = 100.0	D43 = 12.5	D71 = 100.0	D99 = 75.0
D16 = 100.0	D44 = 37.5	D72 = 87.5	D100 = 50.0
D17 = 100.0	D45 = 62.5	D73 = 75.0	D101 = 87.5
D18 = 100.0	D46 = 100.0	D74 = 100.0	D102 = 75.0
D19 = 87.5	D47 = 100.0	D75 = 100.0	D103 = 25.0
D20 = 100.0	D48 = 75.0	D76 = 87.5	D104 = 50.0
D21 = 100.0	D49 = 100.0	D77 = 100.0	D105 = 100.0
D22 = 100.0	D50 = 75.0	D78 = 100.0	D106 = 12.5
D23 = 75.0	D51 = 100.0	D79 = 12.5	D107 = 100.0
D24 = 100.0	D52 = 100.0	D80 = 100.0	D108 = 75.0
D25 = 100.0	D53 = 100.0	D81 = 100.0	D109 = 50.0
D26 = 100.0	D54 = 37.5	D82 = 100.0	D110 = 87.5
D27 = 100.0	D55 = 100.0	D83 = 100.0	D111 = 100.0
D28 = 100.0	D56 = 100.0	D84 = 100.0	D112 = 100.0

As baixas taxas de acertos de algumas texturas possuem causas diversas. Para o caso de D5, D30, D31 e D37, texturas de grão grosso, a causa principal deve-se à inadequação das dimensões da janela com a estrutura da textura. Outras texturas também de grão grosso, como a D56, apresentaram uma taxa de 100,0% de acerto mas a regularidade de sua estrutura básica de repetição é muito bem definida, fato que não ocorre para as anteriores. Neste aspecto, é possível interpretar o resultado obtido por D25 e D48, parede de tijolos e papel perfurado respectivamente pois, em ambos os casos, as amostras de 128×128 não capturaram a essência da textura. A taxa de 100% obtida por D25 refere-se à microtextura presente no tijolo em si e não arranjo por ele proporcionando. O mesmo tipo de análise vale para D48.

Um outro problema diz respeito à não homogeneidade (o que, indiretamente, pode ser um problema de inadequação do tamanho da amostra). Como exemplos pode-se citar: D43, D44, D45, D58, D59, D61, D62, D63, D90, D100. Tais texturas são bastante irregulares, o que dificulta a extração de características. D43 e D44 são imagens com a predominância de 2 níveis de cinza e de difícil análise sob o ponto de vista de textura. Para a obtenção de boas taxas de acerto, a digitalização deveria ser conduzida tomando-se regiões específicas, o que não é de objetivo deste trabalho. Observe que conta muito a região utilizada no processo de digitalização, como é o caso de D91, foto de uma nuvem; em sua totalidade a imagem não é homogênea mas, durante seu particionamento em amostras de 128×128, alguns exemplos recaíram justamente sobre uma porção muito regular, o que proporcionou a taxa de 87,5% de acerto. D58, D59 e D61 são fotos de mármore europeu, que caracterizam texturas não homogêneas e de grão grosso. Para estes casos, a presente abordagem não permite que se obtenha uma taxa de acertos elevada.

Outro tipo de imagem que aparece e sua estrutura não é considerada na técnica são aquelas que possuem componente fractal. Como exemplo é possível citar imagens de pedras tais como a D23, D30, D31 e D54. Todas elas apresentam pedras semelhantes na forma mas diferentes no tamanho. Como não existe uma predominância de uma sobre a outra, as taxas de acerto obtidas não foram elevadas.

5.4.1) Texturas em presença de rotação

Com o objetivo de avaliar o desempenho do método para imagens em presença de rotação, foram propostos alguns testes e seus resultados comparados aos obtidos na literatura. Primeiramente, modificou-se as características empregadas, substituindo-se $c1$ e $c2$ por não atenderem ao novo modelo. De acordo com o exposto na seção 3.7 do capítulo 3, as medidas que se adaptam à metodologia de invariância à rotação são $c3$, $c4$ e $c5$, equações (3.30.c), (3.30.d) e (3.30.e) respectivamente. Como existe uma redução de informação, optou-se por trabalhar inicialmente com amostras de dimensões de 64×64 pixels, evitando-se problemas de estimação das matrizes de variância/covariância.

Isto posto, para um HMM de 4 estados 2 e gaussianas por estado e a configuração inicial dos filtros de Gabor (não havendo necessidade de modificação para nenhum dos testes desta tese), a taxa total de acertos foi de 90,38%, taxa obtida sem o emprego de imagens giradas para treinamento. A matriz de confusão obtida é apresentada sob a forma da tabela 5.4.

	D9	D12	D15	D17	D19	D24	D29	D38	D68	D84	D92	D94	D112
D9	81,25	0	0	0	0	18,75	0	0	0	0	0	0	0
D12	0	90,625	3,1125	0	0	6,25	0	0	0	0	0	0	0
D15	0	0	100	0	0	0	0	0	0	0	0	0	0
D17	0	0	0	100	0	0	0	0	0	0	0	0	0
D19	0	0	0	0	65,625	0	21,875	0	0	0	0	12,5	0
D24	0	0	0	0	0	100	0	0	0	0	0	0	0
D29	0	0	0	0	0	0	100	0	0	0	0	0	0
D38	3,125	0	0	0	0	3,125	0	93,75	0	0	0	0	0
D68	0	0	0	0	0	0	0	0	100	0	0	0	0
D84	0	0	0	0	0	0	0	0	6,25	93,75	0	0	0
D92	0	3,125	3,125	0	0	0	0	0	0	0	93,75	0	0
D94	0	0	0	0	0	0	3,125	0	0	0	0	96,875	0
D112	0	0	0	0	12,5	0	0	0	0	0	0	28,125	59,375

Tabela 5.4: Matriz de confusão, com taxas em %, para a classificação do grupo das 13 texturas em presença de rotação.

Em [Haley&Manjunath, 98], para o mesmo conjunto de amostras usado na tabela 5.4, mas trabalhando-se com dimensões de 128×128 e empregando-se texturas giradas tanto para treinamento como para testes, obteve-se uma taxa de 96,8%. Ou seja, sob o ponto de vista de robustez, a abordagem apresentada nesta tese foi capaz de transformar, de forma eficiente, características variantes à rotação para características invariantes à rotação.

No que se refere à obtenção dos mapas 2D, tanto a abordagem desta tese como a apresentada em [Haley&Manjunath, 98] são semelhantes pois ambas extraem informações de amplitude e frequência instantâneas, havendo diferenças no que se refere à estrutura dos filtros de Gabor utilizados, a disposição espectral dos mesmos e o modelo de extração das características instantâneas. Ademais, a forma de processar a rotação modifica-se uma vez que em [Haley&Manjunath, 98] opta-se por realizar a correção ainda sobre os mapas 2D de saída de cada canal e na presente técnica cria-se um vetor de características a partir do valor médio dos mapas, e sobre este é aplicado a correção.

Ainda sobre os resultados da tabela 5.4, observe que as 13 texturas selecionadas possuem características tanto isotrópicas (sem uma direção predominante) como anisotrópicas (com a predominância de mais de uma direção) e, como exposto em [Tan, 98], características extraídas de texturas isotrópicas são essencialmente invariantes à rotação. Contudo, esta divisão é muito severa uma vez que determinadas texturas possuem uma natureza ambígua neste sentido. Um exemplo é o que ocorre com D19: percebe-se uma trama suave anisotrópica mas a componente isotrópica relativa à distribuição dos níveis de cinza predomina. Portanto, uma abordagem mais de acordo com a situação é denominar as texturas por preferencialmente isotrópica ou preferencialmente anisotrópica.

Os baixos valores de taxa de acertos das texturas D9, D19 e D112 é atribuída à técnica de tratamento de invariância à rotação. Para estes casos, a correção da orientação representa perda de informação (seção 3.7 do capítulo 3) uma vez que não existe direção predominante. É importante lembrar que este mesmo banco de dados, sem a presença de rotação e sem o descarte de características, a abordagem desta tese proporcionou taxa de acerto de 100%.

Em [Tan, 98], propõe-se um novo sub-grupo de texturas (aqui equivalente ao banco 5), exclusivamente anisotrópicas, composta das seguintes classes: D10, D12, D15, D17, D22, D24, D34, D37, D53, D56, D65, D68, D80, D82 e D84. As 15 imagens foram digitalizadas em 512×512 pixels e as amostras geradas possuem dimensões de 128×128 pixels, formando um banco de dados de 600 imagens, 40 para cada classe. Empregando filtragem multicanal por filtros de Gabor e classificação por distância euclidiana, obteve-se uma taxa de acertos de 89,3%. Os autores também utilizaram exemplos de textura giradas para treinar o sistema.

Este mesmo sub-grupo de imagens foi avaliado pelo presente método (AM-FM + HMM), com a mesma configuração utilizada para o banco das 13 texturas giradas, e a taxa de acertos obtida foi de 88,33% (amostras de 64×64 pixels). E, como mencionado no começo do capítulo, não se empregou imagens giradas na etapa de treinamento. Os resultados obtidos são muito próximos, evidenciando a capacidade da técnica em tratar texturas giradas sem seu devido treinamento.

Também em [Haley&Manjunath, 98], empregando-se a estrutura desenvolvida para texturas giradas, testou-se seu desempenho sobre as 109 amostras propostas em [Manjunath&Ma, 96] (excluindo-se D31, D32 e D99), obtendo-se uma taxa de acertos de 80,6%. Observe que foi empregada a técnica para tratar texturas giradas na classificação de texturas não giradas. Neste sentido, dada a técnica desenvolvida nesta tese para rotação de texturas, elaborou-se um experimento a fim de medir seu desempenho quando aplicada às 112 texturas não giradas, armazenadas no banco 3. Os resultados são apresentados na tabela 5.5.

Teste	K-Means Seg.	HMM	Características	Dimensões	Taxa de Acertos
5.5.1	4 estados/2 gaus.	Não	c3, c4, c5	64×64	69,36%
5.5.2	4 estados/1 gaus.	Sim	c3, c4, c5	128×128	84%

Tabela 5.5: Experimento empregando-se a técnica invariância à rotação sobre o banco 3 – as 112 texturas.

É interessante notar que em 5.5.2, devido a problemas relativos ao número de amostras para treinamento, foi necessário reduzir o número de gaussianas para 01 por estado. Apesar da redução, a taxa de acertos obtida foi elevada, quase que se aproximando ao resultado obtido pelo teste 5.2.5. O bom desempenho obtido neste caso reforça a capacidade das características empregadas, assim como o desempenho da modelagem proporcionada pela etapa de classificação.

O mesmo tipo de experimento acima (tratamento de texturas giradas aplicado a texturas não giradas) também foi aplicado ao banco 2, obtendo-se uma taxa de acertos de 99,17%. Em [Haley&Manjunath, 98], obteve-se uma taxa de acertos de 100%. A tabela a seguir mostra um resumo dos resultados mais expressivos obtido com a técnica e os compara com outras abordagens.

Banco	Referência	Comentários	Acertos
109 Texturas	[Manjunath&Ma, 96]	Técnica específica para texturas não giradas	72,13%
Banco 3 (112 texturas)	[esta tese]	Técnica específica para texturas não giradas	85,94%
109 Texturas	[Haley&Manjunath, 98]	Técnica para texturas giradas	80,6%
Banco 3 (112 texturas)	[esta tese]	Técnica para texturas giradas	84%
Banco 2	[Haley&Manjunath, 98]	Técnica para texturas giradas	100%
	[esta tese]	Técnica para texturas giradas	99,17%
	[Chang&Kuo, 93]	Técnica específica para texturas não giradas	99,6%
	[esta tese]	Técnica específica para texturas não giradas	100%
23 texturas do Banco 2	[Mojsilovic <i>et. al.</i> , 00]	Técnica específica para texturas não giradas	99,35%
Banco 4	[Haley&Manjunath, 98]	Treinamento com amostras giradas	96,8%
	[esta tese]	Treinamento sem amostras giradas	90,38%
Banco 5	[Tan, 98]	128×128 pixels - Treinamento com amostras giradas	89,3%
	[esta tese]	Treinamento sem amostras giradas	88,33%

Tabela 5.6: Tabela comparativa de alguns experimentos obtidos em relação a outros trabalhos.

Capítulo 6

Conclusões e Projetos Futuros

Contribuições da tese

Quanto ao sequenciamento das características: A representação de uma textura é feita sob a forma de evolução de características instantâneas AM-FM. Alguns trabalhos reconhecem a importância dessas medidas mas não a usam de forma efetiva. Apesar de já se utilizar HMM's em texturas, até então não se levava em conta informações tão relevantes como a frequência instantânea e como seu comportamento ao longo da textura é capaz de caracterizá-la.

Quanto à correlação entre os vetores de características: O uso de HMM se adapta de forma eficiente ao tipo de vetores de características proporcionado pelas transformadas de Gabor. Por serem estas não ortogonais, não se deve assumir vetores independentes, fato este negligenciado na literatura. Os modelos ocultos de Markov não modelam completamente a estatística dos vetores de características mas capturam as dependências entre características de canais distintos, evitando-se o emprego de modelos de Markov explícitos que dificultam a estimação de seus parâmetros. Por fim, utilização do algoritmo EM na estimação das pdf's de cada estado permitiu melhores resultados.

Quanto à invariância à rotação: Foi possível adaptar um método de invariância à rotação para trabalhar com características instantâneas AM–FM. Sua utilização através de cadeias ocultas de Markov demonstrou melhores resultados que os obtidos na literatura uma vez que apenas empregando texturas não giradas no treinamento obteve-se uma boa taxa de acertos.

Quanto à extração de características relativas aos mapas AM–FM: Nesta tese, a correção de rotação foi realizada sobre o valor médio dos mapas de características e não diretamente sobre esses. Tal fato permitiu uma diminuição na carga computacional, além de possibilitar maior robustez na classificação. Como resultado, foi possível classificar texturas giradas sem que no processo de treinamento fosse necessário apresentar tais exemplos.

Os resultados obtidos, em muitos dos casos, demonstraram a robustez da técnica para classificação de texturas, provando ser possível adaptá-la para a questão de texturas giradas. No que se refere ao banco de dados empregado, os bons resultados obtidos demonstram que o sequenciamento de características AM–FM representa convenientemente a maior parte das amostras de texturas existentes. O uso do HMM se justifica tanto pela forma de representação proposta como para capturar, de maneira mais eficiente, as estatísticas de vetores de características obtidos a partir de transformações não ortogonais.

Temas a serem investigado

Assim como foi discutido no capítulo 2, pode ser interessante algum tipo de pré-processamento não linear para ressaltar determinadas características. Esta técnica é largamente empregada em segmentação de texturas mas pode ser extrapolada para o problema de classificação. Inclusive, o pré-processamento pode ser indicado para que seja possível ressaltar os textons e assim permitir, caso se opte pelo emprego de HMM's, que seja associado sentido físico aos estados.

Um próximo tema a ser abordado é a proposição de uma metodologia menos custosa para sinais não estacionários. Para isto, deseja-se investigar decomposição por

Wavelets Packets a fim de decidir pela melhor base sob o ponto de vista AM–FM. A partir de uma decomposição em uma árvore completa total, ortogonal e em multiresolução, aplica-se um algoritmo de poda iniciando-se pelos nós folha da árvore em direção aos nós raiz. A função de decisão de poda deve refletir a capacidade de detecção AM–FM.

Alguns pontos importantes devem ser avaliados com cuidado:

Qual a melhor função base?

De que forma o uso de bases ortogonais deteriora o produto tempo \times frequência?

Como definir a função de custo?

Bases ortogonais tendem a privilegiar determinadas direções. Quais são as implicações para o caso de características AM–FM?

O interessante desta abordagem é que, dado qualquer forma resultante da poda da árvore de Wavelets Packets, sempre será possível definir um conjunto de funções de reconstrução ortogonal do sinal em questão.

Referências Bibliográficas

- [Aach *et. al.*, 95] T. Aach, A. Kaup, R. Mester, “On Texture Analysis: Local Energy Transforms versus Quadrature Filters”, *Signal Processing*, 45 (1995) pp. 173-181.
- [Aas *et. al.*, 99] K. Aas, L. Eikvil, R. B. Huseby, “Applications of hidden Markov chains in image analysis”, *Pattern Recognition*, 32, (1999), pp. 703-713.
- [Bilmes, 98] J. A. Bilmes, “A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models”, International Computer Science Institute, Berkeley, CA, TR-97-021, April, 1998.
- [Bishop, 95] C. M. Bishop, “*Neural Networks for Pattern Recognition*”, Oxford: Oxford University Press, 1995.
- [Bovik *et. al.*, 90] A. C. Bovik, M. Clark, W. Geiler, “Multichannel Texture Analysis Using Localized Spatial Filters”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, january 1990, pp. 55-73.
- [Bovik *et. al.*, 92] A. C. Bovik, N. Gopal, T. Emmoth, A. Respreto (Palacios), “Localized Measurement of Emergent Image Frequencies by Gabor Wavelets”, *IEEE Trans. On Information Theory*, vol. 38, no.2 march 1992, pp. 644-664.
- [Bovik, 91] A. C. Bovik, “Analysis of Multichannel Narrow-Band Filters for Image Texture Segmentation”, *IEEE Trans. on Signal Processing*, vol. 39, no. 9, september 1991, pp. 2025-2043.
- [Brodatz, 66] P. Brodatz, “*Textures: A Photographic Album for Artists and Designers*”, Mineola, NY: Dover Publications, 1966.
- [Carlson, 86] A. B. Carlson, “*Communication Systems*”, McGraw-Hill, third edition, 1986.

- [Chang&Kuo, 93] T. Chang, C.-C. Jay Kuo, "Texture Analysis and Classification with Tree-Structured Wavelet Transform", *IEEE Trans. On Image Processing*, vol. 2, no. 4, october 1993, pp. 1549-1560.
- [Chen&Kundu, 94] J.-Lin Chen, A. Kundu, "Rotation and Gray Scale Transform Invariant Texture Identification Using Wavelet Decomposition and Hidden Markov Models", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, february 1994, pp. 208-214.
- [Chen&Kundu, 95] J.-Lin Chen, A. Kundu, "Unsupervised Texture Segmentation Using Multichannel Decomposition and Hidden Markov Models", *IEEE Trans. On Image Processing*, vol. 4, no. 5, may 1995, pp. 603-619.
- [Clark&Bovik, 89] M. Clark, A. C. Bovik, "Experiments in Segmenting Texton Patterns Using Localized Spatial Filters", *Pattern Recognition*, vol. 22, no. 6, pp. 707-717, 1989.
- [Cohen, 95] L. Cohen, "*Time-Frequency Analysis*", Englewood Cliffs, NJ: Prentice Hall PTR, 1995.
- [Coifman&Wickerhauser, 92] R. R. Coifman, M. V. Wickerhauser, "Entropy Based Algorithms for Best Basis Selection", *IEEE Transaction on Information Theory*, vol. 32, march 1992, pp. 712-728.
- [Connors *et. al.*, 83] R. W. Connors, C. W. McMillin, K. Lin, R. E. V.-Espinosa, "Identifying and Locating Surface Defects in Wood: Part of an Automated Lumber Processing System", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 5, No. 6, november, 1983, pp. 717-728.
- [Daugman, 88] J. G. Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression", *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 36, no. 7, 1988, pp. 1169-1179.
- [Delprat *et. al.*, 92] N. Delprat, B. Escudié, P. Guillemain, R. Kronland-Martinet, P. Tchamitchian, B. Torrèsani, "Asymptotic Wavelet and Gabor Analysis: Extraction of Instantaneous Frequencies", *IEEE Trans. On Information Theory*, vol. 38, no.2 march 1992, pp. 644-664.
- [Duda&Hart, 72] R. O. Duda, P. Hart, "*Pattern Classification and Scene Analysis*", New York: Wiley, 1973.

- [Dunn *et. al.*, 94] D. Dunn, W. E. Higgins, J. Wakeley, “Texture Segmentation Using 2-D Gabor Elementary Functions”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, No. 2, pp. 130-149, february 1994.
- [Dunn&Higgins, 95] D. Dunn, W. E. Higgins, “Optimal Gabor Filters for Texture Segmentation”, *IEEE Trans. on Image Processing*, vol. 4, No. 7, pp. 947-964, july 1995.
- [Etemad&Chellappa, 98] K. Etemad, R. Chellappa, “Separability-Based Multiscale Basis Selection and Feature Extraction for Signal and Image Classification”, *IEEE Trans. On Image Processing*, vol. 7, no. 10, october 1998.
- [Gabor, 46] D. Gabor, “Theory of Communication”, *Journal of IEE*, vol. 93, 1946, pp. 429-457.
- [Gonnet&Torresani, 94] C. Gonnet, B. Torresani, “Local Frequency Analysis with Two-Dimensional Wavelet Transform”, *Signal Processing*, 37 (1994), pp. 389-404.
- [Haley&Manjunath, 99] G. M. Haley, B. S. Manjunath, “Rotation-Invariant Texture Classification Using a Complete Space-Frequency Model”, *IEEE Trans. on Image Processing*, vol. 8, n. 2, february 1999, pp. 255-269.
- [Havlicek *et. al.*, 00] J. P. Havlicek, D. S. Harding, A. C. Bovik, “Multidimensional Quasi-Eigenfunction Approximations and Multicomponent AM–FM Models”, *IEEE Trans. On Image Processing*, vol. 9, no. 2, february 2000, pp. 227-241.
- [Havlicek&Bovik, 95] J. P. Havlicek, A. C. Bovik, “AM–FM Models, the Analytic Image, and Nonlinear Demodulations Techniques”, *Center for Vision and Images Sciences*, CVIS –TR-95-001, march, 1995.
- [Hayes, 96] M. H. Hayes, “*Statistical Digital Signal Processing and Modeling*”, John Wiley & Sons, Inc., 1996.
- [Hisao&Sawchuk, 89a] J. Y. Hsiao, A. A. Sawchuk, “Supervised Textured Image Segmentation Using Feature Smoothing and Probabilistic relaxation Techniques”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 12, december 1989, pp. 1279-1292.
- [Hisao&Sawchuk, 89b] J. Y. Hsiao, A. A. Sawchuk, “Unsupervised Textured Image Segmentation Using Feature Smoothing and Probabilistic Relaxation Techniques”, *Computer Vision, Graphics, and Image Processing*, vol. 48, october – 1989, pp. 1-21.

- [Hsim&Li, 98] H.-C. Hsim, C.-C. Li, “An Experiment on Texture Segmentation Using Modulated Wavelets”, *IEEE Trans. on Systems, Man, and Cybernetics – Part A: Systems and Humans*, vol. 28, no. 5, september 1998, pp. 720-725.
- [Hsim, 00] H.-C. Hsim, “Texture Segmentation Using Modulated Wavelets”, *IEEE Trans. on Image Processing*, vol. 9, no. 7, july 2000, pp. 1299-1302.
- [Jain&Farrokhnia, 91] A. K. Jain, F. Farrokhnia, “Unsupervised Texture Segmentation Using Gabor Filters”, *Pattern Recognition*, vol. 24, no. 12, pp. 1167-1186, 1991.
- [Juang&Rabiner, 90] B.-Hwang Juang, L. R. Rabiner, “The Segmental K-Means Algorithm for Estimating Parameters of Hidden Markov Models”, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 9, september 1990, pp. 1639-1641.
- [Julesz&Bergen, 83] B. Julesz, J. P. Bergen, “Textons, the Fundamental Elements in Preattentive Vision and Perception of Textures”, *Bell Syst. Tech. J.*, vol. 62, pp. 1619-1645, 1983.
- [Kwok&Jones, 00] H. K. Kwok, D. L. Jones, “Improved Instantaneous Frequency Estimation Using an Adaptive Short-Time Fourier Transform”, *IEEE Trans. on Signal Processing*, vol. 48, no. 10, october 2000, pp. 2964-2972.
- [Laws, 80] K. I. Laws, “Textured Image Segmentation”, Ph.D. Thesis, Department of Electrical Engineering, IPI, University of Southern California, Los Angeles, 1980.
- [Lim, 90] J. S. Lim, “Two-Dimensional Signal and Image Processing”, Englewood Cliffs, NJ: PTR Prentice Hall, 1990.
- [Liu&Picard, 99] F. Liu, R. W. Picard, “A Spectral 2-D Wold Decomposition Algorithm for Homogeneous Random Field”, *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Phoenix, Arizona, March 15-19, 1999.
- [Mahalanobis&Singh, 94] A. Mahalanobis, H. Singh, “Application of Correlation Filters for Texture Recognition”, *Applied Optics*, vol. 33, no. 11, april 1994, pp. 2173-2179.
- [Malik&Perona, 85] J. Malik, P. Perona, “Preattentive Texture Discrimination with Early Vision Mechanisms”, *Journal of Optical Society American*, 2, 1985, pp. 1160-1169.
- [Mallat, 98] S. Mallat, “A Wavelet Tour to Signal Processing”, *Academic Press*, 1998.
- [Malvar, 92] H. S. Malvar, “*Signal Processing with Lapped Transform*”, Norwood, MA: Artech House, 1992.

- [Manjunath&Ma, 92] B. S. Manjunath, W. Y. Ma, “Texture Features for Browsing and Retrieval of Image Data”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, August 1996, pp. 837-842.
- [Mojsilovic *et. al.*, 00] A. Mojsilovic, M. V. Ppopovic, D. M. Rackov, “On the Selection of an optimal Wavelet Basis for Texture Characterization”, *IEEE Trans. on Image Processing*, vol. 9, n. 12, december 2000, pp. 2043-2050.
- [Morgan&Bourlard, 95] N. Morgan, H. A. Bourlard, “Neural Networks for Statistical Recognition of Continuous Speech”, *Proceedings of IEEE*, vol. 83, no. 5, may 1995, pp. 743-769.
- [Oppenheim *et. al.*, 97] A. V. Oppenheim, A. S. Willsky, S. H. Nawab, “Signal & Systems”, Upper Saddle River, NJ: Prentice Hall Signal Processing Series, 1997.
- [Papoulis, 84] A. Papoulis, “*Probability, Random Variables, and Stochastic Processes*”, McGraw Hill International, 1984.
- [Pratt, 91] W. K. Pratt, “*Digital Image Processing*”, A Wiley Interscience Publication, second Edition, 1991.
- [Proakis&Manolakis, 96] J. G. Proakis, D. G. Manolakis, “*Digital Signal Processing*”, Upper Saddle River, NJ: Prentice Hall PTR, 1996.
- [Rabiner, 89] L. R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, *Proceedings of IEEE*, vol. 77, no. 2, february 1989, pp. 257-286.
- [Randen&Husøy, 99a] T. Randen, J. H. Husøy, “Filtering for Texture Classification: A Comparative Study”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291-310, april 1999.
- [Randen&Husøy, 99b] T. Randen, J. H. Husøy, “Texture Segmentation Using Filters with Optimized Energy Separation”, *IEEE Trans. on Image Processing*, vol. 8, no. 4, pp. 571-582, april 1999.
- [Randen, 97] T. Randen, “Filter and Filter Bank Design for Image Texture Recognition”, Doctoral Dissertation, Norwegian University of Science and Technology, Stavanger, Norway, 1997.

- [Salles, 94] E. O. T. Salles, “Segmentação de Texturas - Uma Aplicação para Detecção de Defeitos em Superfícies”, Tese de Mestrado. PPGEE, Universidade Federal do Espírito Santo – julho/1994.
- [Strang&Nguyen, 96] G. Strang, T. Nguyen, “*Wavelets and Filter Banks*”, Wellesley, MA: Wellesley-Cambridge Press, 1996.
- [Tan, 98] T. N. Tan, “Rotation-Invariant Texture Features and Their Use in Automatic Script Identification”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, july 1998, pp. 751-756.
- [Unser&Eden, 89] M. Unser and M. Eden, “Multiresolution Feature Extraction and Selection for Texture Segmentation”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, No. 7, july 1989, pp. 717-728.
- [Unser&Eden, 90] M. Unser and M. Eden, “Nonlinear Operartors for Improving Texture Segmentation Based on Features Extracted by Spatial Filtering”, *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 20, No. 4, july 1990, pp. 804-815.
- [Unser, 86] M. Unser, “Local Linear Tranforms for Texture Measurements”, *Signal Processing*, 11 (1986) pp. 61-79.
- [Unser, 93] M. Unser, “Texture Classification and Segmentation Using Wavelets Frames”, *IEEE Trans. on Image Processing*, vol. 4, No. 11, november 1993, pp. 1549-1560.
- [Weldon&Higgins, 96] T. P. Weldon, W. E. Higgins, “Design of Multiple Gabor Filters for texture Segmentation”, *Proc. ICASSP – 96*, May 1 – 10, Atlanta, GA, USA.
- [Wells, 99] R. B. Wells, “*Applied Coding and Information Theory for Enginners*”, Upper Saddle River NJ: Prentice Hall, 1999.
- [Wouwer, 98] G. V. De Wouwer, “*Wavelets for Multiscale Texture Analysis*”, Ph.D. Thesis, Universitaire Instelling Antwerpen, Antwerpen – 1998.
- [Wu&Wei, 96] W.-R. Wu, S-C. Wei, “Rotation and Gray-Scale Transform Invariant Texture Classification Using Spiral Resampling, Subband Decomposition, and Hidden Markov Model”, *IEEE Trans. On Image Processing*, vol. 5, no. 10, october 1996, pp. 1423-1434.
- [Youn *et. al.*, 97] S. Young, J. Odell, V. Valtchev, P. Woodland, “*The HTK Book for HTK Version 2.1*”, Cambridge University, 1997.

[Zhang *et. al.*, 98] J. Zhang, D. Wang, Q. N. Tran, “A Wavelet-Based Multiresolution Statistical Model for Texture”, *IEEE Trans. On Image Processing*, vol. 7, no. 11, november 1998, pp. 1621-1626.

APÊNDICE A3.1

O Problema da Representação Única para Frequência Instantânea

É comum a modelagem de um determinado sinal real $s(t)$ segundo a equação abaixo:

$$s(t) = \sum_{i=1}^K a_i(t) \cos(\varphi_i(t)), \text{ para } s(t) : \mathfrak{R} \rightarrow \mathfrak{R}. \quad (\text{A3.1.1})$$

Entretanto, a representação $s(t)$ por (A3.1.1) acarreta ambiguidades na determinação da frequência instantânea [Cohen, 95] pois existem diversas combinações possíveis para $a_i(t)$ e $\varphi_i(t)$ que satisfazem a (A3.1.1). Um outro problema acontece quando se deseja medir a frequência média de $s(t)$. Por exemplo, supondo $s(t)$ um sinal real e limitado em banda, seu espectro de potências é simétrico (propriedade da transformada de Fourier para sinais reais) e conseqüentemente a frequência média possui valor nulo, conforme se observa na figura A3.1.1.a abaixo.

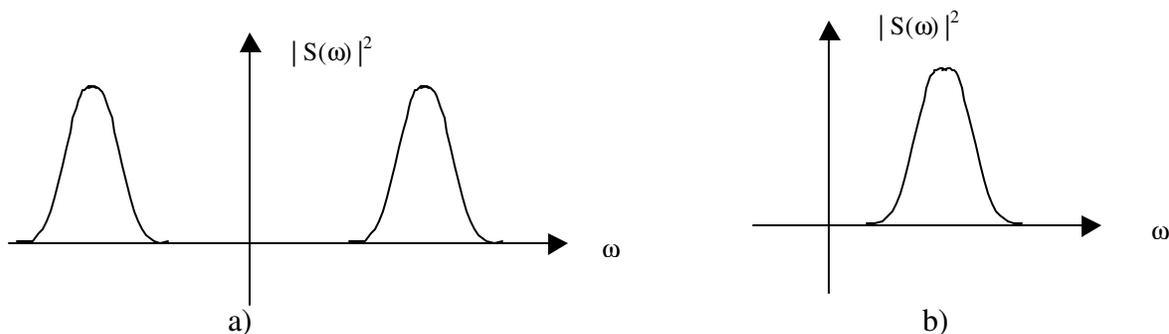


Figura A3.1.1: Espectro de potência: a) de um sinal real, b) de um sinal analítico.

Para evitar tais problemas emprega-se a representação de $s(t)$ na forma de sinal analítico, onde adiciona-se ao sinal real uma parte imaginária a fim de selecionar, de maneira única, um par de amplitude instantânea e de frequência instantânea. Assim, $s(t)$ sofre uma transformação, com a adição de uma componente imaginária $u(t)$, tal que é

possível escrever (tomando-se, para fins de simplificação, apenas uma componente de (A3.1.1)):

$$s_a(t) = s(t) + ju(t), \quad j = \sqrt{-1}. \quad (\text{A3.1.2})$$

A operação (A3.1.2) resulta em um novo sinal $s_a(t)$, agora analítico, desde que seja expressa por meio de:

$$s_a(t) = s(t) + \frac{j}{\pi} \int_{-\infty}^{\infty} \frac{s(t')}{t-t'} dt', \quad (\text{A3.1.3})$$

sendo que a integral acima é avaliada no valor principal de Cauchy [Proakis&Manolakis, 96]. A componente imaginária introduzida em (A3.1.3) refere-se à transformada de Hilbert \mathbf{H} do sinal real. Assim, pode-se escrever $u(t) = \mathbf{H}\{s(t)\}$ por:

$$\mathbf{H}\{s(t)\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(t')}{t-t'} dt', \quad (\text{A3.1.4})$$

que impõe uma rígida interdependência entre as partes real e imaginária do sinal transformado. Consequentemente, a magnitude e a fase de $s_a(t)$ não podem ser expressas de forma arbitrária e tal fato garante a representação única desejada. Na construção de (A3.1.3) a nova representação $s_a(t)$ implica que sua transformada $S_a(\omega)=0$ para $\omega < 0$. Assim sendo, um sinal analítico $s_a(t)$ é essencialmente complexo e completamente caracterizado por sua parte real $s(t)$.

No caso de uma imagem $f(\mathbf{x})$, assumindo-se amplitude real e conjunto suporte discreto, tem-se a equação (A3.1.5) conforme abaixo:

$$t(\mathbf{x}) = \sum_{i=1}^K a_i(\mathbf{x}) e^{j\phi_i(\mathbf{x})}, \quad \text{para } t(\mathbf{x}) : \mathfrak{R}^2 \rightarrow \mathbf{C}, \quad \mathbf{x} \in \mathbf{Z}^2, \quad (\text{A3.1.5})$$

$$\text{para } t(\mathbf{x}) = f(\mathbf{x}) + \mathbf{H}\{f(\mathbf{x})\}. \quad (\text{A3.1.6})$$

Neste caso, a transformação de Hilbert para o caso 2D pode ser obtida por (A3.1.7) [Havlicek&Bovik, 95]:

$$\mathbf{H}\{f(\mathbf{x})\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{t(\mathbf{x}' - p\mathbf{e}_i)}{p} dp, \quad (\text{A3.1.7})$$

onde \mathbf{e}_i representa um vetor em uma direção arbitrária da imagem.

A transformada de Hilbert é necessária no mapeamento do sinal real para a forma complexa que, atribuindo valores iguais a zero aos espectro de potências relativo às frequências negativas da transformada do sinal. Em especial, esta é uma forma eficiente de implementação utilizada para sinais discretos, sendo este o método implementado, selecionando-se uma das duas frequências possíveis e atribuindo-se valor zero para a porção do espectro relativa às frequências negativas.

APÊNDICE A3.2

A Questão da Validade da Modelagem AM-FM

O objetivo deste apêndice é introduzir os conceitos referentes à modelagem AM–FM sem ser demasiado extenso. A opção pela utilização de seus resultados deve-se não somente por ser um tema bastante rico, mas por ser especialmente indicado para imagens não homogêneas, o que costuma acontecer com texturas. Para uma discussão mais profunda, sugere-se as referências [Bovik *et. al.*, 92] [Havlicek&Bovik, 95] e [Havlicek *et. al.*, 00], ao longo do qual todo o desenvolvimento da abordagem é apresentado.

O uso da modelagem AM–FM aplicada a um sinal representado na sua forma analítica pressupõe determinadas condições. Primeiramente, não se deve confundir um sinal na forma em quadratura com a forma analítica, apesar da expressão matemática ser a mesma para ambos, dada por (A3.2.1) (onde analisa-se apenas uma componente de uma imagem):

$$t(\mathbf{x}) = a(\mathbf{x})e^{j\varphi(\mathbf{x})}. \quad (\text{A3.2.1})$$

A forma analítica impõe que o conteúdo espectral da amplitude seja menor que o conteúdo espectral do termo $e^{j\varphi(\mathbf{x})}$ [Cohen, 95], o que não é garantido para a forma em quadratura. Dizer que um é menor que o outro significa que se o espectro de $a(\mathbf{x})$ está contido em uma determinada região espectral A1 e o espectro de $e^{j\varphi(\mathbf{x})}$ está contido em outra região espectral A2, sendo ambas as regiões disjuntas e A1 localizada na porção de mais baixas frequências. Ou seja, o conteúdo de baixa frequência é dado pelo termo $a(\mathbf{x})$ e o de alta frequência dado pelo termo $e^{j\varphi(\mathbf{x})}$.

Quando o modelo AM–FM é apresentado, duas características importantes devem ser destacadas: frequências emergentes e coerência localizada. O conceito das frequências emergentes em processamento de imagens está ligado a situações onde imagens não

homogêneas apresentam comportamentos locais, havendo modificação de dito comportamento ao longo do conjunto suporte dos dados. Daí a necessidade de transformações com boa localização espacial – frequência espacial. Em se tratando de texturas, seria o equivalente a observar, para um mesmo tipo e sobre toda sua extensão, regiões que se assemelham localmente. Desta forma, é possível assumir certa homogeneidade local introduzindo-se, conseqüentemente, o conceito de coerência local. Assumir a propriedade da coerência garante que as funções componentes $a(\mathbf{x})$ e $\varphi(\mathbf{x})$ dos modelos (3.6) e (3.7) são suaves em todas as direções de avaliação local no conjunto suporte, permitindo a aplicação de algum procedimento de medição. Sob o ponto de vista de suavidade, a coerência das componentes fica quantificada pelo valor de suas normas no espaço de Sobolev¹⁶. Observe que garantir regiões de frequências emergentes unicamente não garante o sucesso do método. É preciso dar sustentação matemática para que procedimentos de medição torne atrativa a teoria. Isto é conseguido assumindo-se a coerência localizada.

Outro ponto importante diz respeito à aproximação por quase autofunções, apresentada na equação (3.25). Como se sabe, apenas para imagens monocromáticas essa suposição é válida. Contudo, devido à utilização dos filtros de Gabor, e de sua propriedade de localização, é intuitivo assumir tal aproximação. Se se analisa sob o aspecto de frequências emergentes e coerência localizada, é possível encontrar um filtro cuja frequência central e largura de banda se adaptem à informação localizada da imagem em questão. E é interessante notar que justamente neste aspecto se baseiam as abordagens de filtros ótimos apresentadas no Capítulo 2, pois tudo que se deseja nesta situação é encontrar um filtro dedicado para cada textura. A potencialidade da aproximação por autofunções aliada à características AM–FM reside em fazer suposições bastante plausíveis a respeito das imagens tratadas: localmente são emergentes e coerentes, sob este aspecto são homogêneas, o que não indica seu uso para imagens do tipo fractal. Em [Havlicek&Bovik, 95] são apresentadas condições a respeito do uso da aproximação por quase autofunções. Neste caso, avalia-se o erro introduzido ao se assumir tal aproximação e comprova-se que o

¹⁶ O espaço de Sobolev contém todas as funções $f(\mathbf{x})$, $\mathbf{x} \in \mathfrak{R}^d$, com s derivadas no espaço normado de p -ésima ordem L^p [Strang&Nguyen, 96].

mesmo é limitado, permanecendo baixo ao tratar imagens com características emergentes e coerência localizada.

O conceito de frequência local também foi proposto em [Delprat *et. al.*, 92] onde trata-se inclusive da modelagem multicomponente. Em [Gonnet&Torresani, 94] aplica-se tal abordagem a imagens bem comportadas (imagens obtidas por meio de funções matemáticas que implementam o modelo (3.7)). A grande vantagem do modelo AM-FM por quase autofunções é, além da apresentação de solução analítica para conjunto suporte contínuo, a proposição de um algoritmo robusto para imagens discretas (conjunto suporte discreto e níveis de cinza discretos).

APÊNDICE A3.3

Relação entre transformadas e banco de filtros

Neste trabalho utiliza-se os termos banco de filtros e transformada de maneira indistinta e o objetivo deste apêndice é mostrar que é possível assumí-lo. Para tanto, suponha um filtro FIR 2D, com conjunto suporte retangular e de dimensões $N \times M$, cuja resposta pulsada é dada por $h(n, m)$. A operação de filtragem linear de uma imagem $t(n, m)$ que gera uma imagem $y(n, m)$ na saída deste filtro é dada por:

$$y(n, m) = h(n, m) * t(n, m) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} h(i, j) t(n-i, m-j), \quad (\text{A3.3.1})$$

onde o uso de $*$ representa a convolução bidimensional. Para um determinado par (n, m) , a equação acima pode ser reescrita como um produto interno [Malvar, 92]. Primeiramente, ordena-se lexicograficamente as linhas de $h(n, m)$ e $t(n, m)$ como mostrado a seguir:

$$\mathbf{h} = \begin{bmatrix} h(0,0) \\ \vdots \\ h(0, M-1) \\ h(1,0) \\ \vdots \\ h(1, M-1) \\ \vdots \\ h(N-1, M-1) \end{bmatrix} \quad \text{e} \quad \mathbf{t}_{n,m} = \begin{bmatrix} t(n, m) \\ \vdots \\ t(n, m-M+1) \\ t(n-1, m) \\ \vdots \\ t(n-1, m-M+1) \\ \vdots \\ t(n-N, m-M) \end{bmatrix}, \quad (\text{A3.3.2})$$

ou seja, define-se duas variáveis vetoriais \mathbf{h} e $\mathbf{t}_{n,m}$, sendo $\mathbf{t}_{n,m}$ um bloco da imagem indexado pelo par (n, m) . Portanto, para cada posição (n, m) da imagem de saída é possível escrever:

$$y(n, m) = \mathbf{h}^T \mathbf{t}_{n,m}. \quad (\text{A3.3.3})$$

Suponha agora um conjunto de L filtros \mathbf{h}_i , $1 \leq i \leq L$, escritos na forma lexicográfica conforme (A3.3.2). Cada um destes L filtros projeta um valor distinto relativo a $\mathbf{t}_{n,m}$ segundo (A3.3.3). Assim sendo, é possível escrever:

$$[y_1(n, m), y_2(n, m), \dots, y_L(n, m)]^{\text{Tr}} = [\mathbf{h}_1 : \mathbf{h}_2 : \dots : \mathbf{h}_L]^{\text{Tr}} \mathbf{t}_{n,m}, \quad (\text{A3.3.4})$$

onde define-se a matriz de transformação $\mathbf{H}_T = [\mathbf{h}_1 : \mathbf{h}_2 : \dots : \mathbf{h}_L]^{\text{Tr}}$. Expresso segundo esta maneira, as colunas de \mathbf{H}_T geram o espaço do vetor $\mathbf{Y} = [y_1(n, m), y_2(n, m), \dots, y_L(n, m)]^{\text{Tr}}$. Portanto, um banco de filtros pode ser interpretado como uma operação de transformação e a base do novo espaço é gerada a partir dos vetores coluna (coeficientes do filtro escritos lexicograficamente) de \mathbf{H}_T .

APÊNDICE A4.1

Algoritmo de Decodificação de Viterbi

O algoritmo de Viterbi é uma técnica originalmente desenvolvida para a decodificação de códigos convolucionais através da maximização da verossimilhança da sequência de informação decodificada [Carlson, 86]. Entretanto, seu uso também se popularizou na determinação da máxima verossimilhança para modelos de Markov. Sua utilização é importante pois evita o sequenciamento de dois estados j e l com probabilidade de transição $a_{jl} = 0$, como pode ocorrer com o método ditreto [Rabiner, 89]. Na seleção de Q_i^* o algoritmo toma sequencialmente cada vetor de observações $\mathbf{o}_t \in O$ e determina o estado j que proporciona a maior probabilidade. Dita probabilidade é apresentada na expressão abaixo:

$$\delta_j(t) = \max_{q^1, q^2, \dots, q^{t-1}} P(q^1, q^2, \dots, q^{t-1} = j, \mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t \mid \lambda_t), t \leq T \text{ e } i \in \{1, \dots, M\}. \quad (\text{A4.1.1})$$

Sua implementação para tempos posteriores a t é obtida aplicando-se o procedimento da indução, conforme abaixo:

$$\delta_j(t+1) = [\max_j \delta_t(j) a_{jl}] \cdot P(\mathbf{o}_{t+1}, q_t \mid \lambda_t). \quad (\text{A4.1.2})$$

Uma vez varrida toda a sequência de observações basta recuperar o caminho de máxima verossimilhança. Para isto, adiciona-se ao algoritmo uma variável auxiliar $\psi_j(t)$ somente para fazer a “marcação” do melhor caminho. Na forma algorítmica, pode-se escrever o procedimento de Viterbi como se segue abaixo [Rabiner, 89]. Para simplificar a notação, empregou-se $b_j(\mathbf{o}_t)$ no lugar de $P(\mathbf{o}_t \mid q_j)$.

Inicialização:

$$\begin{aligned} \delta_j(t) &= \pi_j b_j(\mathbf{o}_1) \quad 1 \leq j \leq N, \\ \psi_j(1) &= 0. \end{aligned} \tag{A4.1.3}$$

Recursão:

$$\delta_l(t) = \max_{1 \leq j \leq N} [\delta_j(t-1) a_{jl}] \cdot b_l(\mathbf{o}_t); \quad 2 \leq t \leq T; \quad 1 \leq l \leq N, \tag{A4.1.4}$$

$$\psi_l(t) = \arg \max_{1 \leq j \leq N} [\delta_j(t-1) a_{jl}]; \quad 2 \leq t \leq T; \quad 1 \leq l \leq N. \tag{A4.1.5}$$

Finalização:

$$P(O | \lambda_i) = \max_{1 \leq j \leq N} [\delta_j(T)], \quad 1 \leq j \leq N; \quad 1 \leq i \leq M, \tag{A4.1.6}$$

$$q_T^* = \arg \max_{1 \leq j \leq N} [\delta_j(T)] \tag{A4.1.7}$$

Obtenção da melhor seqüência de estados Q^*

$$Q_i^* = \{\psi_{q^*(t+1)}(t+1)\}_{t=T-1}^{t=1} \tag{A4.1.8}$$

onde $q^*(t)$ é o melhor estado no instante t .

APÊNDICE A4.2

Modelo de Misturas

Modelo de misturas é a técnica que permite representar uma função de densidade de probabilidade $p(\mathbf{x})$ desconhecida através de uma combinação linear de funções base (ou núcleos), de acordo com a equação abaixo:

$$p(\mathbf{x}) = \sum_{j=1}^{M_x} p_j(\mathbf{x} | m) P_j(m), \quad \sum_{j=1}^{M_x} P_j(m) = 1, \quad 0 \leq P_j(m) \leq 1 \quad \text{e} \quad \int p_j(\mathbf{x} | m) d\mathbf{x} = 1, \quad (\text{A4.2.1})$$

onde $p_j(\mathbf{x} | m)$ é a função base ponderada por $P_j(m)$. Este tipo de abordagem, por aproveitar as vantagens da estimação paramétrica e não paramétrica é denominada de semi-paramétrica [Bishop, 95]. As funções base em geral são densidades de probabilidade do tipo log-côncava ou elípticamente simétricas [Rabiner, 89]. Uma escolha comum é supor gaussianas e a seleção dos parâmetros de cada uma das M_x possíveis é através de estimadores da máxima verossimilhança, como são as equações de estimação apresentadas no passo 4 do algoritmo do K-means Segmental. É possível demonstrar que a modelagem apresentada por (A4.2.1) pode aproximar, com precisão arbitrária, qualquer uma determinada pdf desde que se tenha um número M_x suficiente de funções base, um procedimento confiável para a seleção de parâmetros da função base e da ponderação $P_j(m)$. A figura abaixo mostra dois estados onde as pdf's correspondentes são dadas por modelo de misturas.

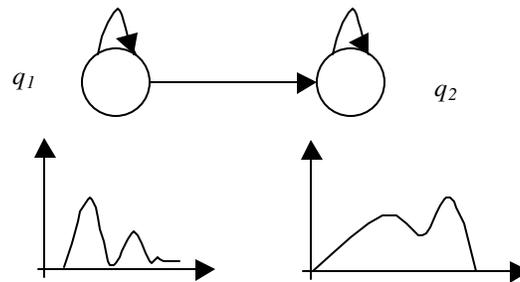


Figura A4.2.1: A figura mostra, na parte superior, um modelo HMM composto de dois estados e suas respectivas misturas são apresentadas nos gráficos da parte inferior.

APÊNDICE A4.3

Algoritmos EM e Reestimação de Baulm-Welch

Antes de apresentar o algoritmo de Reestimação de Baum-Welch é necessário introduzir os algoritmos *forward* e *backward*, que são básicos para o seu entendimento. Conforme exposto anteriormente, dada uma sequência de observações O e um modelo λ é necessário desenvolver uma metodologia que calcule, de forma eficiente, $P(O|\lambda)$. Ambos, *forward* e *backward*, o fazem de maneira satisfatória, a um custo computacional da ordem de N^2T , contra um custo para o método direto da ordem de $2TN^T$. O ponto chave é assumir uma estrutura em treliça de observações e aplicar um procedimento de indução. Para isto, duas novas variáveis são introduzidas aos métodos, conforme abaixo:

$$\text{Para } forward: \alpha_j(t) = P(\mathbf{o}_1, \dots, \mathbf{o}_t, q'_j | \lambda) \quad (A4.3.1)$$

$$\text{Para } backward: \beta_j(t) = P(\mathbf{o}_{t+1}, \dots, \mathbf{o}_T | q'_j, \lambda) \quad (A4.3.2)$$

Onde $\alpha_j(t)$ é a probabilidade conjunta da sequência parcial de observação dos t primeiros vetores, uma vez que o estado em t é q_j , dado um modelo λ . Já para o caso *backward*, $\beta_j(t)$ é a probabilidade da sequência parcial de observações do instante $t+1$ até T , dado um modelo λ e que o estado em t é q_j . Aplicando-se indução em ambos os casos obtém-se [Rabiner, 89]:

Algoritmo *forward*:

Inicialização:

$$\alpha_j(1) = \pi_j b_j(\mathbf{o}_1), \quad 1 \leq j \leq N \quad (A4.3.3)$$

Indução para frente:

$$\alpha_j(t+1) = \left[\sum_{j=1}^N \alpha_j(t) a_{jl} \right] b_j(\mathbf{o}_{t+1}) \quad (A4.3.4)$$

$$1 \leq t \leq T-1, \quad 1 \leq l \leq N \quad (A4.3.5)$$

Finalização:

$$P(O | \lambda) = \sum_{j=1}^N \alpha_j(T) \quad (A4.3.6)$$

Algoritmo *backward*:

Inicialização:

$$\beta_j(T) = 1, \quad 1 \leq j \leq N \quad (A4.3.7)$$

Indução para trás:

$$\beta_j(t) = \sum_{j=1}^N a_{jl} b_l(\mathbf{o}_{t+1}) \beta_l(t+1) \quad (A4.3.8)$$

$$t = T-1, \dots, 1 \quad 1 \leq j \leq N \quad (A4.3.9)$$

Finalização:

$$P(O | \lambda) = \sum_{j=1}^N \beta_j(t) \pi_j b_j(\mathbf{o}_1) \quad (A4.3.10)$$

O problema de obter o modelo $\lambda=(A,B,\boldsymbol{\pi})$ é um pouco mais complicado visto que dispõe-se apenas da sequência O , processo visível, tanto para a obtenção dos parâmetros relativos à sua distribuição, dada por B , como também para encontrar os parâmetros da matriz estocástica A que rege o processo oculto. Ou seja, aparentemente não existe suficiente quantidade de variáveis para estimar as leis estatísticas de ambos os processos. Felizmente, é possível estimar o modelo λ a partir desta informação incompleta e uma maneira de se fazê-lo é aplicando a técnica do EM; também conhecida como algoritmo de reestimação de Baum-Welch quando usada na obtenção de parâmetros de HMM's.

O problema estimação de parâmetros de modelos a partir de variáveis incompletas já foi muito estudado na estatística com diversos métodos propostos para resolvê-lo. Em comum entre eles está a suposição de uma variável intermediária, chamada de dado latente, que facilita a otimização da verossimilhança [Bilmes, 98]. De posse dos dados latentes e uma vez explicitada os parâmetros $\boldsymbol{\theta}$ do modelo, executa-se um processo iterativo que culmina com a determinação de $\boldsymbol{\theta}$.

Suponha a existência de um conjunto de observações X , dados completos, dependentes de parâmetros $\boldsymbol{\theta}$ e suponha também que só é possível observá-lo através de um conjunto de variáveis $Y = T(X)$, dados incompletos, onde T representa uma transformação não inversível (tipo muitos para um), ou seja $\mathbf{y} \in Y \subseteq \mathcal{R}^l$ e $\mathbf{x} \in X \subseteq \mathcal{R}^n$, para $n > l$. A idéia principal é explicitar a estimativa dos parâmetros $\boldsymbol{\theta}$ sob a forma $\boldsymbol{\theta}(k)$, onde k indexa iterações, de tal forma a garantir que:

$$l(\boldsymbol{\theta}^{(k+1)} | \mathbf{y}) - l(\boldsymbol{\theta}^{(k)} | \mathbf{y}) \geq 0, \quad (\text{A4.3.11})$$

onde $l(\cdot)$ representa uma função de verossimilhança.

O algoritmo EM para a obtenção dos parâmetros $\boldsymbol{\theta}$ é expresso segundo se observa a seguir (supondo-se uma estimativa inicial $\boldsymbol{\theta}(0)$):

Procedimento E: Obtenção do valor esperado da verossimilhança

No passo $(k+1)$, dado que $\Theta(k)$ já está disponível, calcula-se:

$$Q(\Theta; \Theta(k)) = E\{\ln(p_x(x, \Theta | X; \Theta(k)))\}, \quad (\text{A4.3.12})$$

Procedimento M: maximização de $\Theta(k)$

$$\Theta(k+1) = \frac{\partial Q(\Theta; \Theta(k))}{\partial \Theta} = 0. \quad (\text{A4.3.13})$$

O procedimento termina quando $\|\Theta(k+1) - \Theta(k)\| \leq \epsilon$. $Q(\cdot)$ representa a função de verossimilhança parametrizada por $\Theta(k)$. Supondo-se que a pdf de cada estado seja uma mistura de gaussianas, como em (A4.2.1), a etapa **M** maximiza os parâmetros de cada componente do modelo, a saber: a ponderação $P(m)$, o vetor de médias \mathbf{v}_m e a correspondente matriz de variância/covariância C_m .

A reestimação de Baum-Welch, como o título sugere, também se baseia na técnica de, a cada iteração, obter uma estimativa dos parâmetros sobre o valor esperado. Para relembrar, o problema é encontrar $\lambda = (A, B, \pi)$ que maximize $P(O|\lambda)$, sendo O uma sequência de treinamento referente a uma textura. A partir das probabilidades *forward* e *backward*, é possível determinar a probabilidade do HMM estar em um determinado estado j em um dado instante de tempo t a partir de:

$$\xi_{j,l}(t) = P(q(t) = j, q(t+1) = l | \lambda, O) = \frac{\alpha_j(t) a_{j,l} b_l(\mathbf{o}_{t+1}) \beta_l(t+1)}{P(O | \lambda)}. \quad (\text{A4.3.14})$$

De posse de $\xi_{j,l}(t)$ define-se uma nova variável $\gamma_j(t)$ segundo a expressão abaixo:

$$\gamma_j(t) = \sum_{l=1}^N \xi_{j,l}(t). \quad (\text{A4.3.15})$$

A equação (A4.3.15) informa a probabilidade do HMM estar no estado j , no instante de tempo t , dado a observação de O . Para a estimação dos parâmetros do modelo basta empregar as equações (A4.3.14) e (A4.3.15) em conjunto com o conceito de frequência estatística, como se segue:

$$\boldsymbol{\pi}_l = \gamma_l(1) \therefore \boldsymbol{\pi} = [\gamma_1(1), \gamma_2(1), \dots, \gamma_N(1)]^{Tr}, \text{ Tr} = \text{transposto}, \quad (\text{A4.3.16})$$

$$a_{j,l} = \frac{\sum_{t=1}^{T-1} \xi_{j,l}(t)}{\sum_{t=1}^{T-1} \gamma_j(t)}, \quad (\text{A4.3.17})$$

$$b_l(s) = \frac{\sum_{t=1}^{T-1} \gamma_l(t)}{\sum_{t=1}^{T-1} \gamma_j(t)}, \quad (\text{A4.3.18})$$

A equação (A4.3.18) refere-se a HMM's utilizando emissão de símbolos \mathbf{v}_s . Mas, para o caso onde a função de densidade de probabilidade de cada estado é descrita por misturas de modelo, conforme (A4.1.1), introduz-se mais um índice (além dos índices do tempo e dos estados), relativo à quantidade de modelos por estado. Assim sendo, modifica-se (A4.3.18) para a estimação dos parâmetros da mistura de modelos de cada estado, mantendo-se as equações referentes à estimação da matriz A e do vetor de probabilidades iniciais, conforme apresenta-se a seguir [Rabiner, 89]:

$$P_j(m) = \frac{\sum_{t=1}^T \gamma_{j,m}(t)}{\sum_{t=1}^T \sum_{m=1}^{Mx} \gamma_{j,m}(t)}, \quad (\text{A4.3.19})$$

$$\mathbf{v}_{j,m} = \frac{\sum_{t=1}^T \gamma_{j,m}(t) \cdot \mathbf{o}_t}{\sum_{t=1}^T \gamma_{j,m}(t)}, \quad (\text{A4.3.20})$$

$$\mathbf{C}_{j,m} = \frac{\sum_{t=1}^T \gamma_{j,m}(t) \cdot (\mathbf{o}_t - \mathbf{v}_{j,m})(\mathbf{o}_t - \mathbf{v}_{j,m})^{Tr}}{\sum_{t=1}^T \gamma_{j,m}(t)}, \text{ Tr} = \text{transposto}, \quad (\text{A4.3.21})$$

onde $\gamma_{j,m}(t)$ é dada por:

$$\gamma_{j,m}(t) = \frac{\alpha_j(t) \beta_j(t) \frac{P_j(m) G(\mathbf{o}_t, \mathbf{v}_{j,m}, \mathbf{C}_{j,m})}{\sum_{j=1}^N \alpha_j(t) \beta_j(t) \sum_{l=1}^{Mx} P_j(l) G(\mathbf{o}_t, \mathbf{v}_{j,l}, \mathbf{C}_{j,l})}}{\sum_{j=1}^N \alpha_j(t) \beta_j(t) \sum_{l=1}^{Mx} P_j(l) G(\mathbf{o}_t, \mathbf{v}_{j,l}, \mathbf{C}_{j,l})} \quad (\text{A4.3.22})$$

$G(\mathbf{o}_t, \mathbf{v}_{j,m}, C_{j,m})$ representa o núcleo gaussiano com vetor de medias $\mathbf{v}_{j,m}$ e matriz de variância/covariância $C_{j,m}$ estimada a partir dos vetores de observações \mathbf{o}_t .

APÊNDICE A5.1

Relação dos tipos de Texturas tratadas

D1	Woven aluminum wire	D28	Beach sand
D2	Fieldstone	D29	Beach sand
D3	Reptile skin	D30	Beach pebbles-translucent effect
D4	Pressed cork	D31	Beach pebbles with hard, dry appearance
D5	Expanded mica	D32	Pressed cork
D6	Woven aluminum wire	D33	Pressed cork
D7	Fieldstone	D34	Netting
D8	Abstract illusion of woven wire	D35	Lizard skin
D9	Grass lawn	D36	Lizard skin
D10	Crocodile skin	D37	Water
D11	Homespun woolen cloth	D38	Water
D12	Bark of tree	D39	Lace
D13	Bark of tree	D40	Lace
D14	Woven aluminum wire	D41	Lace
D15	Straw	D42	Lace
D16	Herringbone weave	D43	Varied swinging of light bulb on a length of wire
D17	Herringbone weave	D44	Swinging lights in a darkened room
D18	Raffia weave	D45	Abstract effect of swinging light
D19	Woolen cloth	D46	Woven brass mesh
D20	French canvas	D47	Woven brass mesh
D21	French canvas	D48	Perforated masonite panel
D22	Reptile skin	D49	Straw screening
D23	Beach pebbles	D50	Raffia woven with cotton threads
D24	Pressed calf leather	D51	Raffia woven with cotton threads
D25	Ceramic-coated brick wall	D52	Oriental straw cloth
D26	Ceramic-coated brick wall	D53	Oriental straw cloth
D27	Beach sand and pebbles translucent effect		

D54	Beach pebbles	D87	Sea fan, fossilized with a coral covering
D55	Straw matting	D88	Dried hop flowers
D56	Straw matting	D90	Clouds
D57	Handmade paper	D89	Dried hop flowers
D58	European marble	D91	Clouds
D59	European marble	D92	Pigskin
D60	European marble	D93	Fur. Hide of unborn calf
D61	European marble	D94	Brick wall
D62	European marble	D95	Brick wall
D63	European marble	D96	Brick wall
D64	Handwoven Oriental rattan	D97	Cross section of a weathered tree
D65	Handwoven Oriental rattan	D98	Crushed rose quartz
D66	Plastic pellets	D99	Crushed rose quartz
D67	Plastic pellets	D100	Ice crystals on an automobile
D68	Wood grain	D101	Cane
D69	Wood grain	D102	Cane
D70	Wood grain	D103	Loose burlap
D71	Wood grain	D104	Loose burlap
D72	Tree stump, used as a chopping block	D105	Cheesecloth
D73	Soap bubbles stump	D106	Cheesecloth
D74	Coffee beans	D107	Japanese rice paper
D75	Coffee beans	D108	Japanese rice paper
D76	Oriental grass fiber cloth	D109	Handmade paper, grassy fiber
D77	Cotton canvas	D110	Handmade paper, grassy fiber
D78	Oriental straw cloth	D111	Plastic bubbles
D79	Oriental grass fiber cloth	D112	Plastic bubbles
D80	Oriental straw cloth		
D81	Oriental straw cloth		
D82	Oriental straw cloth		
D83	Woven matting		
D84	Raffia looped to a high pile		
D85	Straw matting with bamboo effect		
D86	Ceiling tile		

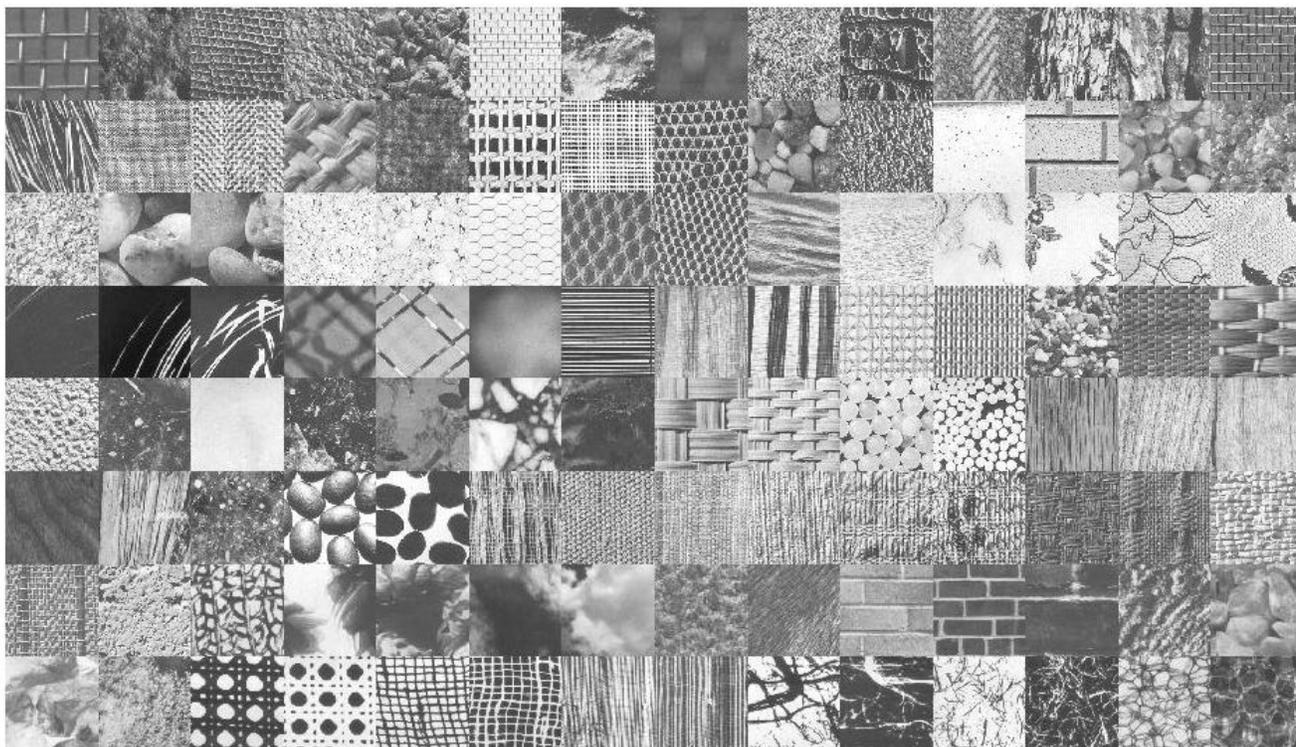


Figura A5.1: O banco total das 112 texturas, de D1 a D112 varridas da esquerda para direita, de cima para baixo. Cada amostra possui 128×128 pixels.

