



PALAVRAS CHAVES/KEY WORDS

AUTORES  
AUTHORS

MODELOS ESTATÍSTICOS CLASSIFICAÇÃO DE PADRÕES  
PROCESSAMENTO DE IMAGENS ARMA

AUTORIZADA POR/AUTHORIZED BY

*Ralf Gielow*  
Ralf Gielow  
Pres. Cons. Pós-Graduação

AUTOR RESPONSÁVEL  
RESPONSIBLE AUTHOR

*Luciano Vieira Dutra*  
Luciano Vieira Dutra

DISTRIBUIÇÃO/DISTRIBUTION

INTERNA / INTERNAL  
 EXTERNA / EXTERNAL  
 RESTRITA / RESTRICTED

REVISADA POR / REVISED BY

*Nelson Mascarenhas*  
Nelson D.A. Mascarenhas

CDU/UDC

621.396.5

DATA / DATE

ABRIL, 1990

TÍTULO/TITLE	PUBLICAÇÃO Nº PUBLICATION NO
	INPE-5067-TDL/406
CLASSIFICAÇÃO DE TEXTURAS USANDO MODELOS ARMA E DISTÂNCIAS DA FUNÇÃO DE AUTOCORRELAÇÃO	
AUTORES/AUTHORSHIP	Luciano Vieira Dutra

ORIGEM  
ORIGIN

PG/DPI

PROJETO  
PROJECT

CAP

Nº DE PAG.  
NO OF PAGES

162

ULTIMA PAG.  
LAST PAGE

A.4

VERSÃO  
VERSION

Nº DE MAPAS  
NO OF MAPS

RESUMO - NOTAS / ABSTRACT - NOTES

Com o aumento da resolução das imagens de satélite de recursos terrestres, a informação de textura torna-se cada vez mais importante como auxílio na detecção e medidas dos objetos presentes nessas imagens. Neste trabalho é proposto um método para descrever e classificar texturas baseado em modelos autorregressivos e de média móvel (ARMA) bidimensionais. Esses modelos são derivados de modelos unidimensionais estimados sobre séries unidimensionais obtidas pela concatenação de linhas ou colunas da imagem. A imagem é filtrada pelos filtros inversos derivados desses modelos, gerando tantos canais quantos forem as classes de texturas. Considera-se a distribuição conjunta das saídas dos filtros para cada classe. A imagem é classificada ponto a ponto usando um classificador de máxima verossimilhança, ou por amostras (conjunto conectado de pixels) usando os critérios de distância JM, de máxima verossimilhança amostral ou maioria. É proposto também um novo classificador de amostras baseado em distâncias da função de autocorrelação bidimensional. Os métodos foram testados usando imagens de radar de abertura sintética, do livro de Brodatz e imagens SPOT pancromáticas. Os modelos para as classes definidas foram apresentados e discutidos. Os resultados das classificações mostraram substancial melhoria no desempenho médio (uma estimativa para a probabilidade de erro), principalmente para os casos onde a diferença entre a média das classes é pequena. Foi observado também um bom desempenho quando se usa o classificador de amostras baseado em distância entre distribuições.

OBSERVAÇÕES/REMARKS

Tese de Doutorado em Computação Aplicada, aprovada em 14 de dezembro de 1989.

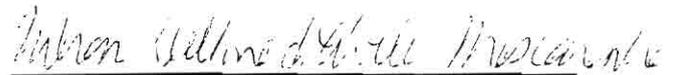
Aprovada pela Banca Examinadora  
em cumprimento a requisito exigido  
para a obtenção do Título de Doutor  
em Computação Aplicada

Dr. Ivan Jelinek Kantor



Presidente

Dr. Nelson Delfino d'Ávila Mascarenhas



Orientador

Dr. Gerald Jean Francis Banon



Orientador

Dr. Fernando Toshinori Sakane



Membro da Banca  
-convidado-

Dr. Jacques Szczupak



Membro da Banca  
-Convidado-

Dr. Osamu Saotome



Membro da Banca  
-Convidado-

Candidato: Luciano Vieira Dutra

São José dos Campos, 14 de dezembro de 1989

*Na verdade, na  
verdade vos digo que  
quem ouve a minha  
palavra, e crê na-  
quele que me enviou,  
tem a vida eterna, e  
não entrará em con-  
denação mas passou  
da morte para a  
vida.*

*(Jo 5, 24)*

Ao Lucas, Suzana, Isabel e Cristina.

Aos meus pais.

### AGRADECIMENTOS

O trabalho apresentado não seria possível sem a participação dos meus orientadores Dr. Nelson D.A. Mascarenhas e Dr. Gerald J.F. Banon, a quem agradeço sinceramente. Estendo meus agradecimentos à Sueli P. Castellari e Silvia S. Ii pelo desenvolvimento de programas utilizados nos experimentos, e também à Janete da Cunha pela datilografia de parte da tese; ao Prof. Pedro A. Morettin do IME da USP pela cessão de bibliografia fundamental no desenvolvimento do trabalho; aos membros da Banca pelas valiosas sugestões apresentadas e ao INPE pelo suporte material e financeiro. Finalmente quero agradecer a todos aqueles que me animaram, a Carlos A. Nobre e principalmente aos meus filhos e esposa, a paciência com que compreenderam a minha ausência nas inúmeras noites em que não pude estar com eles.

## ABSTRACT

*As the spatial resolution of remote sensing imagery increases, texture information becomes an important feature to help in the process of detection and measurement of objects present in these images. Several methods for describing texture have been proposed in the literature, although, up to the present time, no formal definition of texture has been stated. In this work a method for the description and classification of textures using two-dimensional autorregressive moving average (ARMA) models is proposed. These models are derived from uni-dimensional models estimated over uni-dimensional series obtained through the concatenation of rows or columns of the image. The image is filtered by inverse filters derived from the models, generating as many channels (residuals) as the number of textures classes. The joint distribution of the filters outputs is considered for each class. The image is classified on a pixel by pixel basis using a maximum likelihood classifier or on a sample basis (set of connected pixels) using the JM-distance, sample maximum likelihood or majority criteria. A new sample classifier, based on distances defined over two-dimensional auto-correlation functions, is also proposed. The methods were tested using synthetic aperture radar (SAR), Brodatz' book and SPOT panchromatic images. The obtained models for the defined classes were presented and discussed, relating them to the characteristics of the textures and the formation process of the image. The classification results showed substancial improvement of the average performance (an estimation of the correct classification probability) mainly for the cases in which the differences of classes means were negligible. Good performance was also verified using the sample classifier based on distances between two-dimensional auto-correlation functions.*

## SUMÁRIO

	<u>Pág.</u>
LISTA DE FIGURAS .....	<i>xv</i>
LISTA DE TABELAS.....	<i>xvii</i>
LISTA DE SÍMBOLOS.....	<i>xix</i>
LISTA DE SIGLAS E ABREVIATURAS.....	<i>xxi</i>
<u>CAPÍTULO 1 - INTRODUÇÃO</u> .....	1
<u>CAPÍTULO 2 - MODELOS AUTO-REGRESSIVOS E MÉDIA MÓVEL</u> .....	5
2.1 - Definições .....	5
2.2 - Modelos mistos ARMA .....	6
2.3 - Propriedades dos modelos ARMA .....	8
2.3.1 - Função de autocorrelação e autocovariância.....	8
2.3.2 - Equação de Yule-Walker .....	10
2.3.3 - Relação entre parâmetros ARMA e a função de autocorrelação.....	11
2.3.4 - Relação dos processos AR com predição linear.....	14
2.3.5 - Solução recursiva para YW e a função de autocorrelação parcial...	17
2.3.6 - Coeficientes de reflexão.....	18
2.3.7 - Condições de estabilidade e estacionariedade .....	19
2.4 - Estimativa dos parâmetros .....	20
2.4.1 - Estimativa dos parâmetros auto-regressivos .....	20
2.4.2 - Estimativa de parâmetros modelos MA .....	22
2.4.3 - Estimativa preliminar de parâmetros de modelos ARMA.....	23
2.4.4 - Estimativa de máxima verossimilhança de modelos ARMA .....	26
2.4.5 - Erro padrão dos parâmetros estimados .....	28
2.5 - Modelos não estacionários e sazonais.....	29
2.5.1 - Modelos não estacionários .....	29
2.5.2 - Modelos sazonais .....	30
2.6 - Identificação do modelo.....	33
2.7 - Diagnóstico e revisão do modelo.....	35
2.7.1 - Obtenção dos resíduos.....	35
2.7.2 - Uso dos resíduos para modificar o modelo .....	37
2.8 - Resumo da estratégia de modelamento .....	37

<u>CAPÍTULO 3 - MODELOS MULTIVARIADOS E BIDIMENSIONAIS</u> .....	39
3.1 - Modelos multivariados .....	39
3.1.1 - Definições .....	39
3.1.2 - Propriedades e estimativa do modelo MARMA.....	40
3.1.3 - Identificação e verificação .....	42
3.2 - Modelos bidimensionais .....	43
3.2.1 - Definições e regiões de suporte .....	43
3.2.2 - Função de autocorrelação(FAC) e autocovariância .....	47
3.2.3 - O modelo AR, a equação de Yule-Walker 2D e outras propriedades.	48
3.2.4 - Método para estimativa preliminar de parâmetros de modelos ARMA-2D .....	49
3.2.4.1 - As dificuldades para se obter os parâmetros do modelo ARMA-2D completo .....	49
3.2.4.2 - Utilização de modelos unidimensionais sazonais para gerar modelos bidimensionais .....	51
3.2.4.3 - Relacionamento entre FACs de séries bidimensionais e FACs de séries unidimensionais concatenadas .....	52
3.2.4.4 - O efeito da inclusão de zeros entre segmentos concatenados.....	55
3.2.4.5 - A determinação dos coeficientes preliminares de modelos AR 2D ..	56
3.2.5 - Identificação e estimativa de modelo ARMA 2D usando concatenação	59
3.2.6 - Condições de estabilidade e estacionariedade .....	62
 <u>CAPÍTULO 4 - CLASSIFICAÇÃO ESTATÍSTICA DE IMAGENS.</u> .....	 65
4.1 - Introdução.....	65
4.2 - O classificador pontual de máxima verossimilhança .....	65
4.3 - Classificação de amostras.....	66
4.4 - Proposta de um classificador de amostras de textura usando a função de autocorrelação 2D .....	68

<u>CAPÍTULO 5 - ATRIBUTOS DE TEXTURA POR FILTRAGEM INVERSA ..</u>	73
5.1 - Revisão da literatura.....	73
5.2 - Comentários sobre a bibliografia e proposta do método.....	76
5.3 - Classificação ponto a ponto usando filtros inversos .....	78
5.4 - A classificação de amostras.....	92
<u>CAPÍTULO 6 - AQUISIÇÃO E IDENTIFICAÇÃO: RESULTADOS.....</u>	95
6.1 - Metodologia.....	95
6.2 - Testes práticos e discussões .....	96
6.2.1 - Texturas de Brodatz .....	96
6.2.2 - Radar de abertura sintética (SAR580) .....	100
6.2.3 - Imagem SPOT - cidade de São Paulo .....	103
<u>CAPÍTULO 7 - CLASSIFICAÇÃO: RESULTADOS .....</u>	107
7.1 - Introdução.....	107
7.2 - Resultados .....	108
7.2.1 - Texturas de Brodatz .....	108
7.2.1.1 - Diagnóstico dos resíduos.....	117
7.2.2 - Radar de abertura sintética (SAR580) .....	120
7.2.2.1 - Classificação de amostras.....	122
7.2.2.2 - Verificação do branqueamento .....	123
7.2.3 - Imagem São Paulo - SPOT .....	126
7.3 - Conclusões e sugestões para estudos futuros.....	127
REFERÊNCIAS BIBLIOGRÁFICAS .....	131
APÊNDICE A - CÁLCULO DA DIVERGÊNCIA ENTRE DUAS CLASSES.	

## LISTA DE FIGURAS

	<u>Pág.</u>
2.1 - Representação de série temporal como saída de filtro linear excitado por ruído branco $a_t$ .....	6
3.1 - Função matriz de autocorrelação.....	43
3.2 - Região de suporte de meio plano não simétrico.....	44
3.3 - Suporte de quarto de plano.....	45
3.4 - Região de suporte de meio plano não simétrico truncado.....	45
3.5 - Diagrama de relacionamento espacial da série do exemplo da Equação 3.13.....	50
3.6 - Ajuste de um modelo de série temporal espacial de dimensão $m=1$ no plano, fazendo coordenada $t \equiv m$ . A região de suporte não é equivalente nem à NSHP nem à QP. ....	51
3.7 - Transformação de séries bidimensionais em unidimensionais por concatenação de linhas .....	52
3.8 - Relacionamento entre correlações calculadas sobre a série concatenada e o posicionamento no plano dos pontos envolvidos .....	53
3.9 - Indicação dos termos acumulados erroneamente em função da localização no suporte NSHP (números inferiores). Índices sobre a série concatenada $i_c$ (números superiores), $M=7$ .....	54
3.10 - Colocação de zeros entre segmentos concatenados.....	55
3.11 - Aspecto de linhas concatenadas e intercaladas com zero .....	56
3.12 - Exemplo de FAC-2D representada em matriz $7 \times 5$ ( $M = 3, N = 5$ )....	62
5.1 - Extração de atributos de textura na filtragem inversa .....	80
6.1 - Mosaico das texturas de Brodatz (1956): (a) grama,(b) tecido, (c) madeira.....	97
6.2 - Modelos ARMA para a classe grama: (a) modelo AR, parte MA é nula exceto $\theta_{0,0}$ - AIC = 3363, (b) modelo ARMA, AIC = 4472.....	98
6.3 - Filtros inversos das texturas: (a) grama - modelo do tipo AR, (b) tecido - ARMA, (c) madeira - MA.....	99
6.4 - Imagem SAR580: (a) floresta, (b) cultura anual, (c) restos de cultura	100
6.5 - Configuração de coeficientes para a classe Floresta, modelos (a) AR(21), (b) AR(21) LOG, (c) MA(21), (d) MA(21) LOG, (e) ARMA(2,10), (f) ARMA(2,10) LOG.....	102

6.6 - Modelos auto-regressivos para as classes Cultura Anual nos campos Normal (a), e Log (b), e Restos de Cultura nos campos Normal (c) e Log (d). .....	103
6.7 - Imagem Spot - São Paulo mostrando as classes cidade e montanha ...	104
6.8 - Filtros inversos para as classes (a) Floresta e (b) Cidade. ....	105
7.1 - Imagem filtrada pelo filtro inverso da grama.....	109
7.2 - Amostras de treinamento (áreas maiores) e teste para as classes grama, tecido e madeira. ....	109
7.3 - Áreas teste para levantamento da FAC-2D. ....	118
7.4 - Representações bi-dimensionais para a FAC-2D para as classes (a) grama, (b) tecido e (c) madeira .....	118
7.5 - FAC-2D para amostras das classes (a) grama, (b) tecido e (c) madeira no filtro inverso grama.....	119
7.6 - Amostras teste para classificação .....	122
7.7 - Amostras para aquisição da FAC-2D. ....	124
7.8 - FAC-2D das amostras das classes (a) floresta, (b) restos de cultura e (c) cultura anual para o canal radar original. ....	124
7.9 - FAC-2D das amostras das diversas classes obtidas sobre a imagem de radar filtrada pelo filtro inverso da floresta. ....	125

## LISTA DE TABELAS

	<u>Pág.</u>
6.1 - Variância de resíduos SPOT .....	105
7.1 - Matriz de classificação para os canais filtrados (%).....	111
7.2 - Matriz de classificação para o canal original (%) .....	111
7.3 - Matriz de classificação de amostras, canais residuais - critério Distância JM.....	112
7.4 - Matriz de classificação de amostras, canais residuais - critério Distância Mahalanobis Média. ....	112
7.5 - Matriz de classificação de amostras, canais residuais - critério Maioria.....	113
7.6 - Matriz de classificação de amostras, canal original - critério Distância JM.....	113
7.7 - Matriz de classificação de amostras, canal original - critério Distância Mahalanobis Média. ....	114
7.8 - Matriz de classificação de amostras, canal original - critério Maioria.....	114
7.9 - Matriz de classificação de amostras, critério distância entre FAC's (DFAC) - canal original.....	115
7.10 - Matriz de classificação de amostras, critério distância pond. entre FAC's (DPAC) - canal original.....	116
7.11 - Matriz de classificação de amostras, critério log da dife- rença quadrática (DLM) - canal original.....	116
7.12 - Matriz de classificação de amostras, critério distância entre log's dos momentos (DLDM) - canal original. ....	117
7.13 - Variância na saída dos filtros aplicados a cada área de treina- mento (filtros utilizados: auto-regressivos). ....	120
7.14 - Índices de desempenho da matriz de classificação, áreas de treina- mento SAR-580 - 3 classes. ....	121
7.15 - Índices de desempenho da matriz de classificação, canal original processado por log - áreas de treinamento .....	121
7.16 - Classificação de amostras para imagens de radar usando-se canal original ou canais filtrados .....	123
7.17 - Índices de desempenho da matriz de classificação, áreas de treina- mento São Paulo - SPOT.....	127

## LISTA DE SÍMBOLOS

- $a_t$  - Série temporal ruído branco ou "shock"
- $A_t$  - Série temporal vetorial ruído branco
- $A_i$  - Area de amostra ou região amostral em uma imagem
- $d$  - Ordem de diferenciação de uma série temporal cujo modelo seja um modelo ARMA integrado, modelo ARIMA(p,d,q)
- $e_t$  - Resíduo de predição
- $g_i(*)$  - Função discriminante para a classe  $H_i$
- $H_i$  - Rótulo da i-ésima classe conhecida
- $B$  - Operador atraso, "backward shift operator",  $Bz_t = z_{t-1}$
- $DJM_{k_i}$  - Distância de Jeffreys-Matusita entre classes  $H_k$  e  $H_i$
- $c_k$  - Estimativa para  $\gamma_k$ , a autocovariância
- $K$  - Número de classes em consideração
- $m$  - Ordem de um preditor linear unidimensional
- $M$  - Número de pixels no segmento a concatenar
- $N$  - Número de observações ou amostras de um processo estocástico
- $p$  - Ordem de um modelo AR, ou da parte AR de um modelo ARIMA
- $P$  - Ordem da parte AR sazonal de um modelo ARMA sazonal
- $q$  - Ordem de um modelo MA, ou da parte MA de um modelo ARIMA
- $Q$  - Ordem da parte MA sazonal de um modelo ARMA sazonal
- $r_k$  - Estimativa para  $\rho_k$ , o coeficiente de autocorrelação
- $S_x$  - Região de suporte bidimensional de rótulo  $X$
- $S_x^o$  - Região de suporte  $S_x$  não incluindo a origem
- $z_t$  - Série temporal, sequência de valores no tempo
- $z$  - Transformada  $Z$
- $Z_t$  - Série temporal vetorial
- $\alpha$  - Coeficiente de predição linear
- $\phi$  - Parâmetro autorregressivo de um modelo ARMA univariado
- $\Phi$  - Matriz de coeficientes autorregressivos de um modelo ARMA multivariado
- $\gamma_k$  - Função de covariância (FA) para um processo estocástico estacionário
- $\Gamma$  - Matriz de covariância para um processo multivariado

- $\lambda$  - Correlação cruzada entre a entrada e a saída de um filtro-modelo ARMA
- $\mu$  - Média de um processo estocástico estacionário
- $\omega_t$  - Amostra de um processo estocástico  $z_t$
- $\pi$  - Parâmetro de uma expansão AR infinita de um processo ARMA
- $\Psi$  - Coeficiente de uma expansão de um modelo ARMA em termos de um modelo MA de ordem infinita
- $\rho_x$  - Função coeficiente de autocorrelação (FAC) de um processo estocástico estacionário
- $\sigma_x$  - Desvio padrão da v.a.  $X$
- $\Sigma$  - Matriz de covariância
- $\theta$  - Parâmetro média-móvel de um modelo ARMA univariado
- $\Theta$  - Matriz de coeficientes média móvel de um modelo ARMA multivariado.

#### Índices Superiores

- T - Indica transposta
- + - Indica pseudo - inversa

#### Índices Inferiores

- t - Tempo
- k - Atraso ou "lag", intervalo entre dois instantes de tempo
- ij - Índices de coeficientes bidimensionais ou de componentes de matrizes, podem conforme contexto significar a ordem da recursão e o número do coeficiente. São separados por vírgula quando necessário para clareza dos campos dos índices.
- $i_c$  - Índice sobre a série concatenada

#### Símbolos especiais

- - Til sobre uma v.a. ou processo estocástico indica sua versão centralizada ou seja com a média subtraída
- ˆ - Acento circunflexo sobre v.a. indica estimativa daquela variável
- - Traço sob a variável indica vetor
- - Traço sobre a variável indica média
- $\nabla$  - Operador diferença,  $\nabla z_t = z_t - z_{t-1}$

## LISTA DE SIGLAS E ABREVIATURAS

- AR - Autorregressivo (modelo)
- ARMA - Autorregressivo e média móvel (do inglês "autorregressive moving average"), modelo
- DFAC - Distância entre funções de autocorrelação
- DPAC - Distância ponderada entre funções de autocorrelação
- DLM - Distância euclidiana entre logaritmos com sinal dos momentos da FAC-2D
- DLDM - Distância logaritmica da diferença quadrática dos momentos
- DJM - Distância de Jeffreys-Matusita
- FA - Função de autocovariancia
- FAC - Função de autocorrelação
- FAP - Função de autocorrelação parcial
- FPA - Filtro passa-altas
- FPB - Filtro passa-baixas
- MA - Média móvel ("moving average")
- v.a. - Variável aleatória
- YW - Yule-Walker, equação de
- YW-2D- Yule-Walker de duas dimensões, equação de

## CAPÍTULO 1

### INTRODUÇÃO

Na medida em que se aumenta a resolução dos sensores dos satélites de observação da Terra e conseqüente aumento de percepção de detalhes, aumenta também a importância do uso da informação de textura para auxiliar o processo de detectar e mensurar os objetos presentes nas imagens.

Mesmo em outros ambientes, como os industriais, a textura é fator primordial para distinção dos objetos da cena. Não existe ainda uma definição formal de textura, porém dois tipos de métodos aparecem na literatura (Haralick, 1979) tentando descrevê-la: os métodos estatísticos e os determinísticos, que se aplicam conforme o grau de regularidade que caracteriza a textura em estudo. Texturas de objetos feitos pelo homem são melhor descritos por métodos determinísticos enquanto que texturas naturais são normalmente estatísticas. A Seção 5.1 apresenta alguns métodos estatísticos existentes na literatura.

Esses métodos, desenvolvidos em nível teórico para exemplos bem específicos, são, porém, muito custosos e de eficiência discutível, principalmente quando se trata de classificar cenas obtidas de plataformas remotas, ou seja, distinguir diferentes objetos em imagens de sensoriamento remoto. Um dos principais problemas que aparece, se refere ao fato de que esses métodos requerem grande massa de dados (grandes parcelas de imagem com textura uniforme) para determinar os parâmetros do classificador, o que geralmente não é disponível, além do custo computacional elevado. Muitos desses processos de forma alguma se adapta ao problema de sensoriamento remoto.

Por isso, na prática tem sido utilizado quase que exclusivamente os chamados classificadores pontuais que, embora eficientes, são classificadores que assinalam classes aos pixels da imagem de forma individualizada, ou seja, não levam em consideração a informação dos pixels vizinhos (textura) para decidir qual a classe que deve ser assinalada ao ponto sendo classificado.

Numa tentativa de se manter a eficiência, aproveitando de certa forma o relacionamento local entre pixels, tem sido desenvolvido métodos heurísticos de extração de atributos espaciais tanto para classificação de regiões (Kettig e Landgrebe, 1976), quanto para classificação pontual (Dutra e Mascarenhas, 1984; Dondes e Rosenfeld, 1982; Andrade, 1989).

Mais recentemente Therrien (1983) e Gambotto (1980) apresentaram um método mais formalizado para descrição de texturas baseado na teoria de séries temporais bidimensionais (modelos AR-2D), utilizando esses modelos para classificação de imagens digitais.

O principal objetivo desse trabalho, estendendo o trabalho de Therrien (1983) e Gambotto (1980) em vários aspectos a serem discutidos posteriormente, é a utilização de modelos estatísticos do tipo ARMA (ver Capítulo 1) para caracterização formal de texturas irregulares e a utilização desses modelos para ajudar o processo de determinação dos diferentes objetos existentes em imagens de sensoriamento remoto. Além disso é proposto nesse trabalho um método de discriminação de texturas baseados em distâncias da função de autocorrelação da imagem (Seção 4.4).

O modelamento de texturas é efetuado dentro de uma perspectiva de *análise de sistemas*. Sistemas lineares são completamente caracterizados pelo tipo de suas entradas, saídas e a função de transferência que governa o processo da entrada para a saída. Na análise de sistemas pode-se identificar as seguintes instâncias (Bennet, 1979):

- a) *Identificação e estimação*: quando se conhece a entrada e a saída-função de transferência desconhecida.
- b) *Convolução e previsão* ("Forecast"): saída do sistema desconhecida
- c) *Deconvolução e antevisão* ("backforecast"): entrada desconhecida.
- d) *Identificação ARMA* : entrada do sistema e função de transferência desconhecida.
- e) *Simulação*: entrada e saída desconhecida

Considera-se o tratamento da imagem como um problema de identificação ARMA ("autorregressive moving-average"), pois apenas a saída,

que é a imagem, é observada. A identificação ARMA considera que o sistema é excitado por ruído branco que passa por um filtro (ou modelo) linear. Em casos multiclases o filtro pode ser considerado função determinística do espaço e seria característico de regiões com diferentes texturas na imagem.

A obtenção dos modelos de cada textura pode então ser conjugada com a classificação estatística de imagens, para ajudar no particionamento da imagem em estudo.

O Capítulo 2 apresenta os modelos ARMA unidimensionais e os processos de identificação e estimação correspondentes, consolidando algumas idéias encontradas nos campos da Engenharia, Economia e Estatística. A teoria aqui apresentada, é base para apresentação no Capítulo 3 dos modelos multivariados (ou multicanais ou múltiplos- Seção 3.1), que guardam uma relação com os modelos bidimensionais (2D - Seção 3.2), que são os aplicados às imagens.

A obtenção de modelos 2D é, no entanto, bem mais complicada que a identificação de modelos 1D, como discutido na Seção 3.2.4. Por isso é proposto (Seções 3.2.4 e 3.2.5) um método para identificação de modelos 2D, baseado na concatenação de linhas ou colunas da imagem e posterior análise unidimensional sobre a série concatenada. Nesse capítulo são discutidos ainda aspectos importantes de modelagem 2D, tais como: estabilidade de modelos 2D, regiões de suporte, e outras propriedades. Discutem-se também as funções de autocorrelação multivariadas e bidimensionais e suas estimativas. Propõe-se afinal uma 'máquina' para apresentação da FAC 2D para auxiliar sua visualização e seus intervalos de confiança.

No Capítulo 4 são apresentados os métodos clássicos de classificação estatística pontual e de amostras. É proposto também um novo método (em quatro versões), para classificação de amostras de texturas, baseado na FAC 2D e seus momentos (Seção 4.4).

O Capítulo 5 apresenta numa revisão da literatura, os métodos comumente referenciados de descrição, discriminação e síntese de texturas baseados em modelos estatísticos. Em seguida é apresentado um comentário

sobre esses métodos e é proposto um novo método para classificação pontual e de regiões baseado em modelos ARMA e filtragem inversa. É apresentado também na página 77 um resumo sobre as principais propostas da Tese. O método proposto é então desenvolvido com a derivação teórica dos parâmetros do classificador. O impacto do método sobre a probabilidade de erro é discutido num exemplo. Finalmente é discutido como o esquema de filtragem inversa casada pode ser aplicado para classificação de amostras.

O Capítulo 6 dá detalhes sobre a aquisição dos dados para modelamento, e apresenta como resultado os modelos obtidos para diversas classes definidas em três tipos de imagens digitais: uma imagem mosaico de texturas de Brodatz (1956), uma imagem de radar de abertura sintética e uma imagem do satélite francês SPOT, de 10m de resolução, sobre a cidade de São Paulo. No Capítulo 7 é apresentado um detalhamento do processo de classificação ponto a ponto e de regiões, e é apresentado o resultado de classificação das classes anteriormente definidas sob forma de tabelas de contigência, sobre áreas teste, tanto para o esquema de classificação ponto a ponto como para o processo de classificação de regiões. Ainda nesse capítulo são apresentadas as conclusões e sugestões para o futuro.

## CAPÍTULO 2

### MODELOS AUTO-REGRESSIVOS E MÉDIA MÓVEL

Discutem-se neste capítulo as chamadas séries temporais que são dados observados de um sistema a intervalos de tempo igualmente espaçados. A partir dessas observações pretende-se prever o comportamento futuro do sistema ("forecast"), ou mesmo controlá-lo através de esquemas adequados.

Por exemplo, no problema de previsão de vendas, as vendas  $z_t$  do corrente mês e as vendas  $z_{t-1}$ ,  $z_{t-2}$ , ... dos meses anteriores são disponíveis e serão usadas para prever as vendas dos meses futuros.

A idéia de se usar um modelo matemático para descrever o comportamento de um sistema é bem estabelecida e se as leis físicas que regem um certo fenômeno são conhecidas, cálculos exatos podem ser feitos para prever o comportamento do sistema: o modelo é determinístico. Provavelmente nenhum fenômeno é completamente determinístico porque fatores desconhecidos podem influenciar o sistema, e principalmente para fenômeno onde isto é mais evidente tal como na previsão de vendas, pode-se derivar um modelo que leve em conta a probabilidade de um certo valor futuro cair entre determinados limites. Esse fenômeno é chamado de um processo estocástico. Uma determinada série temporal é considerada amostra de uma população infinita de tais séries temporais.

Uma classe importante de modelos estocásticos são os chamados modelos estacionários, os quais assumem que o processo permanece em equilíbrio em torno de um certo nível médio constante.

#### 2.1 - DEFINIÇÕES

O operador  $B$ , definido como  $Bz_t = z_{t-1}$  é o chamado operador *atraso* ("backward shift operator") e também  $B^m z_t = z_{t-m}$ . A operação inversa é realizada pelo operador *avanço*  $F = B^{-1}$  ("forward shift operator"). O operador diferença é igual a:

$$\nabla z_t = z_t - z_{t-1} = (1-B)z_t \quad (2.1)$$

O inverso do operador  $\nabla$  é o operador  $S$  (somatória), que é dado por:

$$\nabla^{-1} z_t = S z_t = \nabla^{-1} z_t = (1 - B)^{-1} z_t = (1 + B + B^2 + \dots) z_t = \sum_{j=0}^{\infty} z_{t-j} \quad (2.2)$$

Os modelos aqui usados, são baseados na idéia de que as séries temporais, nas quais valores sucessivos são altamente dependentes, são considerados gerados por uma série de 'shocks' independentes e de média nula, que fazem parte de um processo chamado ruído branco de média zero e variância  $\sigma^2$ , que excita um filtro linear (Fig. 1).



Fig. 1 - Representação de série temporal como saída de filtro linear, excitado por ruído branco  $a_t$ .

A operação de filtragem toma uma soma ponderada das observações prévias tal que:

$$z_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \dots = \mu + \Psi(B)a_t \quad (2.3)$$

Em geral  $\mu$  é o parametro que determina a média do processo. Se os coeficientes  $\Psi_i$  forem em número finito ou infinito mas a série for convergente, o filtro é dito estável e o processo é estacionário.

## 2.2 - MODELOS MISTOS ARMA

Em sua forma mais geral o comportamento do filtro pode ser descrito pelo modelo auto-regressivo de média móvel (Moretin e Toloi, 1985):

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \dots + \phi_p \bar{z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.4)$$

ou  $\phi(B)\bar{z}_t = \theta(B)a_t$

onde  $\bar{z}_t = z_t - \mu$ ,  $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  e  $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$

O modelo deve ser escrito assim no sentido de obter o menor número de parâmetros estimados possível (essa é a condição chamada de *parcimônia*).

O modelo da Equação 2.3 em condições convenientes pode ser escrito como soma ponderada dos valores passados de  $\bar{z}_t$ :

$$\bar{z}_t = \pi_1 \bar{z}_{t-1} + \pi_2 \bar{z}_{t-2} + \dots + a_t \quad (2.5)$$

Tanto a Forma 2.3 e a Forma 2.5 contêm número infinito de elementos, o que dificulta o processo de modelagem. Os casos notáveis são quando a Forma 2.3 ou 2.5 contém número finito de elementos. No primeiro caso o processo é chamado de média móvel (MA - moving average) e corresponde à Forma 2.4 com  $\phi(B)=1$  e no segundo caso o modelo é chamado auto-regressivo (AR) e corresponde à 2.4 com  $\theta(B)=1$ ;  $p$  e  $q$  determinam a ordem do modelo. A Equação 2.4 pode ser reescrita como:

$$\bar{z}_t = \frac{\theta(B)}{\phi(B)} a_t = \frac{1 - \theta_1 B - \dots - \theta_q B^q}{1 - \phi_1 B - \dots - \phi_p B^p} a_t \quad (2.6)$$

Então um processo misto ARMA pode ser pensado como saída de um filtro linear cuja função de transferencia é a razão de dois polinômios e a entrada é ruído branco ou seja um filtro formado pela cascata de 2 outros, um só de polos outro só de zeros.

Tomando a transformada  $z$  de um processo que segue o modelo da Equação 2.4 obtém-se:

$$(1 - \phi_1 z^{-1} - \dots - \phi_p z^{-p}) Z(z) = (1 - \theta_1 z^{-1} - \dots - \theta_q z^{-q}) A(z)$$

Comparando com a Equação 2.6 observa-se que:

$$H(z) = \frac{Z(z)}{A(z)} = \frac{\theta(z^{-1})}{\phi(z^{-1})} \quad (2.6b)$$

### 2.3 - PROPRIEDADES DOS MODELOS ARMA

#### 2.3.1 - FUNÇÃO DE AUTOCORRELAÇÃO E AUTOCOVARIÂNCIA

Seja  $z_t$  um processo estocástico — A covariância entre  $z_t$  e sua amostra  $z_{t+k}$  separada por  $k$  intervalos de tempo é chamada autocovariância e é definida por:

$$\gamma_{t,t+k} = \text{cov}[z_t, z_{t+k}] = E[(z_t - \mu)(z_{t+k} - \mu)] \quad (2.7)$$

Nesta definição o valor da covariância depende do tempo  $t$  e do intervalo  $k$ . Um processo é dito estritamente estacionário se a distribuição conjunta associada com as observações  $z_{t1}, z_{t2}, \dots, z_{tm}$  é a mesma do que as associadas a  $z_{t1+k}, \dots, z_{tm+k}$  para qualquer  $t$ . Uma condição menos restritiva é a condição de estacionariedade fraca quando somente os momentos até certa ordem são considerados apenas função do atraso  $k$ . Neste trabalho consideraremos estacionariedade até segunda ordem. Assim (2.7) pode ser reescrita:

$$\gamma_k = \text{cov}[z_t, z_{t+k}] \quad (2.8)$$

A autocorrelação relativa a um atraso  $k$ ,  $\rho_k$  (também chamado de coeficiente de autocorrelação) pode ser definida como:

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad (2.9)$$

$\gamma_k$  e  $\rho_k$  considerados como função de  $k$ , são chamados função de autocovariância (FA) e função coeficiente de autocorrelação (FAC) respectivamente. Para a estimativa da função de autocorrelação quando se tem um conjunto de  $N$  observações, Box e Jenkins (1970) apresentam como mais satisfatória a seguinte expressão:

$$r_k = \frac{c_k}{c_0} \quad (2.10)$$

onde  $c_k = \frac{1}{N} \sum_1^{N-k} (z_t - \hat{z})(z_{t+k} - \hat{z})$ ,

$$c_{-k} = c_k,$$

e onde  $\hat{z} = \frac{1}{N} \sum_1^{N-k} z_t$

Essa estimativa é polarizada e é preferida à estimativa não polarizada onde a soma dos produtos é dividida por  $(N-k-1)$  porque em geral a variância desse estimador é menor Marple (1987, p. 148), e conduz a matrizes de autocorrelação sempre positivo-definidas. A estimativa polarizada quando usada nas equações de Yule-Walker (ver seção seguinte) corresponde ao método da *autocorrelação* (Therrien et al., 1986; Marple, 1987, p. 218) para resolução de problemas de predição linear (Seção 2.3.4) e é equivalente a estender o limite superior da somatória na Expressão 2.10 para  $N$ , e considerar os dados que faltam como zero. Por isso o método é também chamado *janelado*. À estimativa não polarizada corresponde o método da *covariância* ou *não janelado*. Essas denominações para os métodos são utilizadas por razões históricas e não têm relação com o significado padrão em estatística.

### 2.3.2 - EQUAÇÃO DE YULE-WALKER

Considerando-se o modelo AR(p), (p é a ordem do modelo) obtido a partir da Equação 2.4:

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \dots + \phi_p \bar{z}_{t-p} + a_t$$

e multiplicando ambos os lados da equação por  $\bar{z}_{t-k}$  e tomando o valor esperado podemos deduzir a equação

$$\gamma_k = \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} \quad k > 0$$

dividindo tudo por  $\gamma_0$ :

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (2.11)$$

Se substituirmos  $k=1,2, \dots, p$  na Equação 2.11, obtém-se um conjunto de equações lineares para  $\phi_1, \dots, \phi_p$  em termos de  $\rho_1, \dots, \rho_p$ :

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1 + \dots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \dots + \dots + \phi_p \end{aligned}$$

ou em forma matricial:

$$\underline{\phi} = \begin{bmatrix} \phi_1 \\ \cdot \\ \cdot \\ \phi_p \end{bmatrix} \quad \underline{\rho}_p = \begin{bmatrix} \rho_1 \\ \cdot \\ \cdot \\ \rho_p \end{bmatrix}$$

$$\underline{P}_p = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & \cdots & 1 \end{bmatrix}$$

$$\underline{\rho}_p = \underline{P}_p \underline{\phi} \tag{2.12}$$

A Equação 2.12 é chamada de equação de Yule-Walker, e provê uma maneira aproximada de estimar os parâmetros de uma série AR(p), que pode ser obtida substituindo os coeficientes de correlação teóricos por estimados e calculando a Equação 2.13.

$$\hat{\underline{\phi}} = \hat{\underline{P}}_p^{-1} \underline{r}_p \tag{2.13}$$

### 2.3.3 - RELAÇÃO ENTRE OS PARÂMETROS ARMA E A FA

Se a Equação 2.4 - Modelo ARMA completo - for multiplicada por  $\tilde{z}_{t-k}$  (daqui para frente dispensaremos o  $\tilde{\cdot}$ , ou seja, consideraremos que a média é sempre subtraída) e tomado o valor esperado, pode-se deduzir:

$$\gamma_k = \sum_{i=1}^p \phi_i \gamma_{k-i} - \sum_{i=0}^q \theta_i \lambda_{k-i} \tag{2.14}$$

onde  $\theta_0 = -1$  e  $\lambda_k = E(a_t z_{t+k})$

A correlação cruzada  $\lambda_k$  entre a entrada e a saída pode ser expressa em termos da expansão MA( $\infty$ ) (Equação 2.3) (Marple, 1987, p. 181) :

$$\lambda_k = \begin{cases} \sigma_a^2 & k = 0 \\ \sigma_a^2 \Psi & k < 0 \\ 0 & k > 0 \end{cases} \quad (2.15)$$

A relação final entre os parâmetros ARMA e a função de autocovariância (FA) do processo fica então, usando a Equação 2.15:

$$\gamma_k = \begin{cases} \sum_{i=1}^p \phi_i \gamma_{k-i} - \sigma_a^2 \sum_{i=k}^q \theta_i \Psi_{i-k} & 0 \leq k \leq q \end{cases} \quad (2.16a)$$

$$\gamma_k = \begin{cases} \sum_{i=1}^p \phi_i \gamma_{k-i} & k > q \end{cases} \quad (2.16b)$$

$$\gamma_{-k} \quad k < 0 \quad (2.16c)$$

Observando-se a Equação 2.16 verifica-se que os parâmetros auto-regressivos do modelo são relacionados por um conjunto de equações lineares com a FA. A Equação 2.16 pode ser avaliada para os  $p$  índices de atraso  $q < k \leq q+p$ , por exemplo, e as equações agrupadas na matriz:

$$\begin{bmatrix} \gamma_q & \gamma_{q-1} & \dots & \gamma_{q-p+1} \\ \gamma_{q+1} & \gamma_q & & \gamma_{q-p+2} \\ \vdots & & \ddots & \vdots \\ \gamma_{q+p-1} & \dots & & \gamma_q \end{bmatrix} \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \gamma_{q+1} \\ \vdots \\ \gamma_{q+p} \end{bmatrix} \quad (2.17)$$

Dados portanto a FA para os atrasos  $q-p+1$  a  $q+p-1$  os parâmetros AR podem ser achados separadamente dos parâmetros MA. Essa é a chamada equação de Yule-Walker modificada (Banon, 1971).

Infelizmente os parâmetros MA do modelo não podem ser achados simplesmente como solução de equações lineares, pois estão convoluídos com os coeficientes das expansão MA infinita, resultando em uma relação não linear com a FA.

Fazendo  $q=0$  obtemos as relações para o processo AR puro:

$$\gamma_k = \begin{cases} \sum_{i=1}^p \phi_i \gamma_{k-i} & k > 0 \\ \sum_{i=1}^p \phi_i \gamma_{k-i} + \sigma_a^2 & k = 0 \end{cases} \quad (2.18)$$

que pode ser agrupada na expressão:

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_p \\ \gamma_1 & \gamma_0 & & \gamma_{p-1} \\ \vdots & & \ddots & \vdots \\ \gamma_p & \gamma_{p-1} & & \gamma_0 \end{bmatrix} \begin{bmatrix} 1 \\ -\phi_1 \\ \vdots \\ -\phi_p \end{bmatrix} = \begin{bmatrix} \sigma_a^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (2.19a)$$

ou

$$\Gamma_p \begin{bmatrix} 1 \\ -\underline{\phi} \end{bmatrix} = \begin{bmatrix} \sigma_a^2 \\ \underline{0} \end{bmatrix} \quad (2.19b)$$

Essa é uma forma mais completa para a Equação 2.12, onde a variância do ruído aparece explicitamente.

A relação entre a FA e o modelo MA pode ser obtida fazendo  $p=0$  e notando que  $\Psi_k = -\theta_k$  para  $1 \leq k \leq q$  :

$$\gamma_k = \begin{cases} 0 & k > q \\ \sigma_a^2 \sum_{i=k}^q \theta_i \theta_{i-k} & 0 \leq k \leq q \end{cases} \quad (2.20)$$

Existe portanto uma relação não linear entre a FA e os parâmetros MA.

### 2.3.4 - RELAÇÃO DOS PROCESSOS AR COM PREDIÇÃO LINEAR

O problema da predição linear consiste em estimar uma amostra futura a partir de  $m$  amostras do passado mais recente, por meio de uma relação linear:

$$\hat{z}_t = \sum_{i=1}^m \alpha_i z_{t-i} \quad (2.21)$$

O erro ou resíduo de predição é definido por:

$$e_t = z_t - \hat{z}_t \quad (2.22)$$

Esse resíduo é estacionário e sua variância é dada por:

$$\sigma_e^2 = E(e_t^2) \quad (2.23)$$

Procuram-se os coeficientes  $\alpha_i$  que minimizam a Equação 2.23 e esses são encontrados como solução das equações normais (Giordano e Hsu, 1985, Cap. 3), estrutura portanto idêntica à expressão de estimativas do modelo AR.

Duas distinções no entanto devem ser feitas:

1. Nenhuma suposição foi feita sobre a caracterização estatística do erro de predição, mas os coeficientes do filtro preditor obedecem à relação:

$$\Gamma_m \begin{bmatrix} 1 \\ -\underline{\alpha} \end{bmatrix} = \begin{bmatrix} \sigma_e^2 \\ \underline{0} \end{bmatrix} \quad (2.24)$$

2. O erro é não correlacionado com os dados, e portanto com a predição linear (princípio da ortogonalidade), mas não é em geral ruído branco. Isso só acontece quando  $m = p$  ou seja a ordem do filtro de predição for igual ou maior que a ordem de um processo AR(p) gerando  $z_t$ . Nessas condições o filtro é chamado de "branqueador"

Os coeficientes do filtro, na prática, podem ser obtidos substituindo-se as correlações teóricas por amostrais, na equação de Yule-Walker. Utilizando os dados diretamente na forma de um sistema de equações sobredeterminado a solução clássica da pseudo-inversa pode ser obtida.

Seja  $\underline{\omega} = [ \omega_1, \dots, \omega_N ]^T$  uma sequência de amostras conhecidas do processo  $z_t$ , então para cada  $\omega_j$ ,  $m \leq j \leq N$ , de (2.22) :

$$\omega_j = \hat{\omega}_j + e_j \quad (2.25a)$$

Usando (2.21):

$$\omega_j = \sum_{i=1}^m \alpha_i \omega_{j-i} + e_j \quad (2.25b)$$

Em forma matricial :

$$\underline{\omega}_m = \Omega \underline{\alpha} + \underline{e}_m \quad (2.25c)$$

$$\text{onde } \underline{\omega}_m^T = [ \omega_{m+1}, \dots, \omega_n ], \quad (2.25d)$$

$$\underline{e}_m^T = [ e_{m+1}, \dots, e_n ], \quad (2.25e)$$

$$\Omega = \begin{bmatrix} \omega_m & \omega_{m-1} & \dots & \omega_1 \\ \omega_{m+1} & \omega_m & \dots & \omega_2 \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{N-1} & \dots & \dots & \omega_{N-m} \end{bmatrix}, \quad (2.25f)$$

$$\alpha^T = [ \alpha_1, \dots, \alpha_m ]. \quad (2.25e)$$

Quer-se achar o valor de  $\alpha$  tal que  $e_m^T e_m$  seja mínimo; então a solução de mínimos quadrados será (Mascarenhas e Velasco, 1989):

$$\hat{\alpha} = (\Omega^T \Omega)^{-1} \Omega^T \omega_m = \Omega^+ \omega_m \quad (2.26)$$

onde  $\Omega^+$  é a chamada pseudo-inversa de  $\Omega$

Se compararmos a Equação 2.26 com a Equação 2.13 ver-se-á que a matriz  $c_0 \hat{P}_p$  corresponde à  $\Omega^T \Omega$  e  $c_0 \Gamma_p$  à  $\Omega^T \omega_m$ , com a distinção de que as estimativas da FA usadas na Equação 2.13 não são as mesmas das calculadas implicitamente na Equação 2.26. A diagonal principal de  $\Omega^T \Omega$  não tem os valores exatamente iguais porque existem diferenças nos limites dos cálculos em cada posição da diagonal principal de  $\Omega^T \Omega$ .

Portanto, além das estimativas polarizadas e não polarizadas para as autocorrelações, existe também a estimativa calculada implicitamente pela Equação 2.26. Como nós podemos observar entretanto em Brockwell e Davis (1987),  $n^{-1}(\Omega^T \Omega) \xrightarrow{P} \gamma_0 P_p$  e  $n^{-1} \Omega^T \omega_n \xrightarrow{P} \gamma_0 \rho$ , quando  $n \rightarrow \infty$  ( $\xrightarrow{P}$  significa convergência em probabilidade, ou seja  $x_n$  de uma sequência  $x_t$  converge em probabilidade para  $x$  ( $x_n \xrightarrow{P} x$ )  $\Leftrightarrow P(|x_n - x| > \epsilon) \rightarrow 0$  quando  $n \rightarrow \infty$ ). Este fato significa portanto, como assegura Brockwell e Davis (1987, p. 257), que as estimativas da pseudo-inversa e as estimativas de Yule-Walker convergem em probabilidade para o mesmo valor.

No caso geral de predição linear quando os resíduos são correlacionados entre si, a estimativa para  $\alpha$  na Equação 2.26 pode ser

polarizada (Bennet, 1979) Quando se conhecer a matriz de covariância dos resíduos ( $\Sigma_e$ ), o melhor estimador é o estimador de mínimos quadrados generalizado (GLS). Para o Modelo 2.25c (sobredeterminado) o valor de  $\hat{\alpha}$  é dado por:

$$\hat{\alpha} = (\Omega^T \Sigma_e^{-1} \Omega)^{-1} \Omega^T \Sigma_e^{-1} \omega_m \quad (2.27)$$

Esse estimador é BLUE ("best linear unbiased estimator"), reduzindo portanto a variância dos resíduos. A matriz  $\Sigma_e$  normalmente não é conhecida, podendo no entanto ser estimada por processos iterativos, tomando como início a estimativa de mínimos quadrados simples.

O uso de GLS não é frequente em predição linear devido à dificuldade de se obter a matriz de covariância do resíduo.

### 2.3.5 - SOLUÇÃO RECURSIVA PARA A EQUAÇÃO DE YW E FUNÇÃO DE AUTOCORRELAÇÃO PARCIAL (FAP).

A matriz da Expressão 2.19 além de Toeplitz é simétrica pois a cada entrada  $a_{ij}$  corresponde a entrada  $a_{ji}$  de igual valor. Para solução de equações desse tipo existe uma solução recursiva simples, chamada algoritmo de Levinson. Com esse algoritmo pode-se determinar sequencialmente os coeficientes do modelo de predição linear em ordem crescente. Eventualmente, o m-ésimo coeficiente do modelo torna-se zero ou próximo de zero, se o processo de origem for gerado por um modelo AR(m-1). À sequência dos últimos coeficientes do modelo de predição linear em função da ordem do modelo dá-se o nome de função de autocorrelação parcial (FAP), e se constitui numa ferramenta de auxílio à identificação do modelo como se verá na Seção 2.5.

Algoritmo de Levinson:

$$\hat{\phi}_1 = \hat{\rho}_1 \quad (2.28a)$$

$$\hat{\phi}_{p+1,j} = \hat{\phi}_{p,j} - \hat{\phi}_{p+1,p+1} \hat{\phi}_{p,p-j+1} \quad j=1, \dots, p \quad (2.28b)$$

$$\hat{\phi}_{p+1,p+1} = \frac{\Gamma_{p+1} - \sum_{j=1}^p \hat{\phi}_{p,j} \Gamma_{p+1-j}}{1 - \sum_{j=1}^p \hat{\phi}_{p,j} \Gamma_j} \quad (\text{FAP}) \quad (2.28c)$$

Exemplo coeficientes de terceira ordem:

$$\hat{\phi}_{31} = \hat{\phi}_{21} - \hat{\phi}_{33} \hat{\phi}_{22}$$

$$\hat{\phi}_{32} = \hat{\phi}_{22} - \hat{\phi}_{33} \hat{\phi}_{21}$$

$$\hat{\phi}_{33} = \frac{\Gamma_3 - \hat{\phi}_{21} \Gamma_2 - \hat{\phi}_{22} \Gamma_1}{1 - \hat{\phi}_{21} \Gamma_1 - \hat{\phi}_{22} \Gamma_2} \quad (2.29)$$

### 2.3.6 - COEFICIENTES DE REFLEXÃO

Os coeficientes  $\hat{\phi}_{ii}$  (definidos pela Equação 2.28 são chamados de função de autocorrelação parcial (FAP(i)) na literatura estatística, porque representam uma correlação normalizada entre  $z_t$  e  $z_{t+i}$  com as correlações de  $z_{t+1}$  até  $z_{t+i-1}$  removidas. O negativo desses coeficientes,  $k_i = -\hat{\phi}_{ii}$ , são frequentemente chamados de coeficientes de reflexão devido a sua interpretação como parâmetros físicos de um modelo em tubo acústico ou em modelos sísmicos (Robinson e Treitel, 1980).

Pode-se definir um erro de predição inversa ("backwards"), onde os coeficientes para cada ordem do preditor são os mesmos do preditor direto, mas onde se usam as  $m$  amostras para adiante (portanto anti-causal) e derivar uma relação (Marple, 1987) envolvendo os erros de predição direto e inverso de um processo estocástico complexo.

A predição inversa é dada por:

$$\hat{z}_t^i = \sum_{j=1}^m \alpha_j^i z_{t+j} \quad (2.30)$$

O erro de predição inversa é

$$e_t^i = z_{t-m} - \hat{z}_{t-m}^i \quad (2.31)$$

definido dessa forma para que  $e_t$  e  $e_t^i$  sejam função de um mesmo conjunto de dados  $(z_{t-m}, \dots, z_t)$

A relação entre  $e_t$  e  $e_t^i$ , levando-se em consideração a ordem  $m$  do preditor linear e os coeficientes de reflexão é dado por (Marple, 1987):

$$e_{(m)t} = e_{(m-1)t} + k_m e_{(m-1)(t-1)}^i \quad (2.32a)$$

$$e_{(m)t}^i = e_{(m-1)(t-1)}^i + k_m^* e_{(m-1)t} \quad (2.32b)$$

onde \* = conjugado.

A Equação 2.32 provê uma interpretação alternativa do filtro de predição como uma estrutura em treliça. Os parâmetros em cada estágio do filtro treliça são os coeficientes de reflexão. Os resíduos de um filtro preditor de maior ordem são função dos resíduos diretos e inversos dos filtros de ordem imediatamente inferior. Esse tipo de filtro tem vantagens de implementação pois minimizam a propagação dos erros de arredondamento e tem menor sensibilidade ao valor dos coeficientes.

### 2.3.7 - CONDIÇÕES DE ESTABILIDADE E ESTACIONARIDADE

Na Seção 2.1 foi introduzido o conceito de estacionaridade, hipótese pela qual um processo em particular permanece estável em torno de uma média. Isto também significa (Box e Jenkins, 1970) que as raízes da equação característica  $\phi(B) = 0$ , da parte auto-regressiva do modelo ARMA gerador estão fora do círculo unitário no plano complexo.

Adicionalmente, a condição denominada de invertibilidade deve ser observada pelo modelo ARMA (Box e Jenkins, 1970). Essa condição pode ser ilustrada pela idéia de que, para que um processo MA(q) puro possa ser representado por um AR infinito de maneira convergente, o inverso de  $\theta(B)$  deve ser uma série convergente ou seja as raízes de  $\theta(B) = 0$  devem cair fora do círculo unitário.

As duas condições se resumem na chamada condição de fase mínima (Oppenheim e Schaffer, 1975; Dudgeon e Mersereau, 1984), que implica que o sinal e sua transformada Z apresentam as seguintes propriedades: é causal, os polos e zeros de  $H(z)$  estão dentro do círculo unitario e tem a maior parte da energia perto da origem (no espectro de potência) e além disso  $H$  e  $H^{-1}$  são somáveis absolutamente.

Claerbout (1976) aponta também que a energia de um sinal de fase mínima é maior ou igual que a energia de qualquer outra onda com o mesmo espectro.

## 2.4 - ESTIMATIVA DOS PARÂMETROS

Ver-se-á nesta seção como, uma vez fixada a ordem do modelo, obter o valor dos coeficientes do modelo ARMA.

### 2.4.1 - ESTIMATIVA DOS PARÂMETROS AUTO-REGRESSIVOS

Seja  $\omega^T = [\omega_1, \dots, \omega_N]$  uma sequência de amostras conhecidas do processo  $z_t$ ; deseja-se, a partir desses dados, estimar os parâmetros de um modelo do tipo AR(p), p dado, ajustado a esses dados. Existem diversos métodos de estimação de parâmetros entre os quais pode-se citar (Melsa e Cohn, 1978): mínimos quadrados (LS), máxima verossimilhança (ML) e máximo a posteriori (MAP), que diferem entre si na quantidade de conhecimento estatístico que se possa ter a respeito do parâmetro sendo estimado e do comportamento estatístico do ruído. O método MAP não será discutido aqui.

Pelo método dos mínimos quadrados, conforme visto na Seção 2.3.4, a solução é obtida pela substituição, nas equações de YW, das

correlações teóricas pelas estimadas, além disso, a solução também pode ser obtida pela pseudo inversa da matriz de dados. O que muda basicamente nessas soluções são as fórmulas usadas para o cálculo das estimativas das autocorrelações.

O método ML, dá uma estimativa de melhor qualidade do que LS, pois permite incorporar na solução o conhecimento a priori a respeito do ruído que afeta o modelo (Equação 2.25c). Ao se ajustar o modelo aos dados contidos em  $\omega$ , obtém-se o seguinte sistema na forma matricial:

$$\underline{\omega}_m = \Omega \underline{\phi} + \underline{e}_m \quad (2.33)$$

onde como na Equação 2.25  $\underline{e}_m^T = [e_{m+1}, \dots, e_N]$  e  $\underline{\phi}$  ocupa o lugar dos parâmetros  $\underline{\alpha}$  pois a ordem do preditor nesse caso é igual à ordem do modelo AR(p).

A estimativa ML para  $\underline{\phi}$ , quando o resíduo  $\underline{e}_m$  tem distribuição gaussiana, é dada por (Melsa e Cohn, 1978):

$$\hat{\underline{\phi}} = (\Omega^T \Sigma_e^{-1} \Omega)^{-1} \Omega \Sigma_e^{-1} \underline{\omega}_m \quad (2.34)$$

Observe que para esse caso a solução é idêntica à GLS (Equação 2.27). Essa estimativa ML porém não é "exata" pois na construção dos sistema de equações, parte dos dados disponíveis foi utilizada como condição inicial (implícito na formação do vetor  $\underline{\omega}_m$ ). Essa é a estimativa de máxima verossimilhança *condicional* àquelas condições iniciais.

A partir de uma melhor aproximação da verossimilhança "verdadeira" dos dados disponíveis Box e Jenkins (1970) apresentam a seguinte expressão para a estimativa ML de  $\underline{\phi}$ :

$$\hat{\underline{\phi}} = D_p^{-1} \underline{d} \quad (2.35)$$

onde

$$d = \begin{bmatrix} D_{12} \\ \cdot \\ \cdot \\ D_{1,p+1} \end{bmatrix} \quad D_p = \begin{bmatrix} D_{22} & D_{23} & \dots & D_{2,p+1} \\ D_{23} & D_{33} & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ D_{2,p+1} & \dots & \dots & D_{p+1,p+1} \end{bmatrix}$$

$$D_{ij} = D_{ji} = \tilde{\omega}_i \tilde{\omega}_j + \tilde{\omega}_{i+1} \tilde{\omega}_{j+1} + \dots + \tilde{\omega}_{n+1-j} \tilde{\omega}_{n+1-i} \quad (2.36)$$

obs:  $\tilde{\omega}_i$  versão centralizada de  $\omega_i$ .

Desenvolvendo 2.36 para alguns termos na diagonal principal de  $D_p$  temos:

$$D_{22} = \omega_{n-1}^2 + \dots + \omega_3^2 + \omega_2^2$$

$$D_{33} = \omega_{n-2}^2 + \dots + \omega_3^2$$

$D_{44} = \dots$  diminuindo o número de termos na somatória.

Como se pode observar as estimativas ML aproximadas conduzem a soluções semelhantes às da equação de YW com as correlações estimadas, com variação observada no cálculo dessas mesmas autocorrelações. Dos métodos vistos até agora, o último apresentado é o que apresenta maior aproximação com a chamada verossimilhança exata (Priestley, 1981).

#### 2.4.2 - ESTIMATIVAS DE PARÂMETROS DE MODELOS MA

Não existe uma maneira direta para estimar os parâmetros de modelos MA, do tipo da equação de YW, mas assim como se pode ajustar modelos AR de ordens crescentes aos dados  $\omega^T = [\omega_1, \dots, \omega_n]$  aplicando o algoritmo de Levinson às autocovariâncias amostrais, pode-se ajustar modelos de média móvel

$$\tilde{z}_t = a_t - \theta_{m1} a_{t-1} - \dots - \theta_{mm} a_{t-m} \quad (2.37)$$

de ordens crescentes  $q= 1, 2, \dots$ , por meio do algoritmo das inovações (Brockwell e Davis, 1987), onde  $\theta_{mn}$  significa o  $n$ -ésimo coeficiente de um modelo de ordem  $m$  e  $v_a = \sigma_a^2$  é a variância de  $a_t$ :

Algoritmo da inovações:

$$\hat{v}_{a0} = c_0 \text{ (variância estimada do sinal } z_t)$$

$$\hat{\theta}_{m,m-k} = -\hat{v}_{ak}^{-1} \left[ c_{m-k} - \sum_{j=0}^{k-1} \hat{\theta}_{m,m-j} \hat{\theta}_{k,k-j} \hat{v}_{aj} \right] \quad (2.38a)$$

$k = 0, \dots, m-1$ , ;  $c_k$  = é a autocovariância estimada para o atraso  $k$ .

$$\hat{v}_{am} = c_0 - \sum_{j=1}^{m-1} \hat{\theta}_{m,m-j}^2 v_j \quad (2.38b)$$

Marple (1987) apresenta um método alternativo para estimação de parâmetros MA, que usa a representação de um processo MA por um modelo AR de ordem elevada. Os parâmetros AR(M), onde  $M \gg q$  são possíveis de calcular usando por exemplo o algoritmo de Levinson. Como os parâmetros MA tem uma relação convolucional com os parâmetros AR visto que  $TZ(MA(q))/TZ(AR(\infty)) = 1$ , (onde TZ significa a transformada  $z$ ) constrói-se um sistema de equações que permite calcular os parâmetros MA já que conheço os parâmetros AR. Não seria possível usar as Equações 2.20 porque a variável incógnita aparece no produto  $\theta_i \theta_{i-k}$  dificultando a resolução do sistema.

#### 2.4.3 - ESTIMATIVA PRELIMINAR DE PARÂMETROS DO MODELOS ARMA

Seja o modelo ARMA completo, de média zero, estável

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.4)$$

A condição de estabilidade assegura que a expansão MA( $\infty$ ) converge.

$$z_t = \sum_{j=0}^{\infty} \Psi_j a_{t-j} = \Psi(B) a_t \quad (2.3)$$

Se for observada a interpretação do ponto de vista de transformada  $z$  (Equação 2.6b), deduz-se que  $\Psi(z^{-1}) = \frac{\theta(z^{-1})}{\phi(z^{-1})}$  ou :

$$\theta(z^{-1}) = \Psi(z^{-1})\phi(z^{-1}) \quad (2.40)$$

Tomando a anti-transformada de 2.40 observa-se que se obtém a convolução dos termos em  $\Psi$  e  $\phi$  resultando (Brockwell e Davis, 1987, p. 243):

$$\begin{aligned} \Psi_0 &= 1 \\ \theta_j &= \Psi_j + \sum_{i=1}^{\min(j,p)} \phi_i \Psi_{j-i} \quad j= 1, 2, 3, \dots \end{aligned} \quad (2.41)$$

onde por convenção  $\theta_j = 0$  para  $j > q$  e  $\phi_j = 0$  para  $j > p$ . Observa-se a partir da Equação 2.41 que é possível, utilizando-se dessas relações para  $j=q+1$  até  $q+p$ , extrair os  $p$  valores de  $\phi_i$  já que nesse intervalo  $\theta_j = 0$ , desde que se conheça  $\Psi_j$ . Para estimar os necessários  $\Psi_1, \dots, \Psi_{p+q}$  pode-se usar as estimativas geradas pelo algoritmo das inovações. Substituindo  $\Psi_j$  por  $-\hat{\theta}_{mj}$  em 2.41 e resolvendo as equações resultantes, para  $m$  suficientemente grande (cf. discutido adiante):

$$\hat{\theta}_{mj} = \theta_j + \sum_{i=0}^{\min(j,p)} \phi_i \hat{\theta}_{m,j-1}, \quad j= 1, 2, \dots, p+q \quad (2.42)$$

Da Equação 2.41 com  $j = q+1, \dots, q+p$ , observa-se que  $\hat{\phi}$  deve satisfazer a equação:

$$\hat{\theta}_m = \hat{\theta} \hat{\phi} \quad (2.43)$$

onde

$$\hat{\underline{\theta}}_m = \begin{bmatrix} \hat{\theta}_{m,q+1} \\ \hat{\theta}_{m,q+2} \\ \vdots \\ \hat{\theta}_{m,q+p} \end{bmatrix} \quad \hat{\underline{\phi}} = \begin{bmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{bmatrix}$$

$$\hat{\underline{\theta}} = \begin{bmatrix} \hat{\theta}_{m,q} & \hat{\theta}_{m,q-1} & \dots & \hat{\theta}_{m,q+1-p} \\ \hat{\theta}_{m,q+1} & \hat{\theta}_{m,q} & \dots & \hat{\theta}_{m,q+2-p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}_{m,q+p-1} & \dots & \dots & \hat{\theta}_{m,q} \end{bmatrix}$$

Tendo sido resolvido 2.43 para  $\hat{\underline{\phi}}$  então  $\hat{\underline{\theta}}$  é achado, usando-se 2.42, de:

$$\hat{\theta}_j = \hat{\theta}_{m,j} - \sum_{i=1}^{\min(j,p)} \hat{\phi}_i \hat{\theta}_{m,j-1}, \quad j=1,2, \dots, q \quad (2.44)$$

A variância do ruído branco é calculada a partir de 2.38b:

$$\sigma_a^2 = \hat{V}_{am} \quad (2.45)$$

Um critério para a escolha de m suficientemente grande para se considerar uma boa aproximação para um modelo MA( $\infty$ ) é escolher m de forma a serem ajustados modelos MA(m) de ordem crescente, de tal forma que se minimize o índice do critério AIC, que será discutido na Seção 2.6.2.

Marple (1987) apresenta um outro método para a estimativa preliminar dos parâmetros  $\hat{\theta}$  e  $\hat{\phi}$  em dois passos; primeiro são estimados os parâmetros AR utilizando-se de uma variante da equação de YW modificada, chamada de equação de YW *modificada de mínimos quadrados*. A série é filtrada para se obter resíduos do tipo MA. Um modelo MA(q) é ajustado a esses resíduos usando-se o método anteriormente descrito.

Vale observar que a Equação 2.43 é uma equação do tipo YW onde os coeficientes da expansão MA( $\infty$ ) substituem as correlações amostrais.

#### 2.4.4 - ESTIMATIVA DE MÁXIMA VEROSSIMILHANÇA PARA O MODELO ARMA(p,q)

Seja  $\omega^T = [ \omega_1, \omega_2, \dots, \omega_n ]$  um vetor de observações de um processo estocástico  $z_t$  de média zero. Conhece-se a ordem do modelo ARMA(p,q) que gerou o processo, mas não se conhece os parâmetros  $\phi_1, \dots, \phi_p; \theta_1, \dots, \theta_q$  e  $\sigma_a^2$  ( a média supõe-se zero ), ou seja as observações  $\omega_i$  foram extraídas de uma distribuição conjunta  $p(\omega | \phi, \theta, \sigma_a^2)$ . Tendo as observações em mãos, deseja-se saber quais são os valores para os parâmetros mais prováveis de terem dado origem a esse dados. Na estimativa de máxima verossimilhança os parâmetros são considerados determinísticos, e deseja-se então maximizar uma função  $L(\phi, \theta, \sigma_a^2 | \omega)$  que é da mesma forma que  $p(\omega | \phi, \theta, \sigma_a^2)$ .

Assume-se que as observações advém de distribuição normal. Para início da análise, considerar-se-á, no entanto, a distribuição do ruído de entrada ou "shocks", que supõem gaussianos e independentes entre si. Sendo a distribuição de 1 "shock" simples dado por:

$$p(a_t | \sigma_a^2) = \frac{(2\pi)^{-\frac{1}{2}}}{\sigma_a} \exp\left[-\frac{a_t^2}{2\sigma_a^2}\right] \quad (2.47)$$

A distribuição conjunta dos "shocks" de entrada será:

$$p(a_1, \dots, a_n) = (2\pi)^{-\frac{1}{2}n} \sigma_a^{-n} \exp\left[-\frac{1}{2\sigma_a^2} \sum_{t=1}^n a_t^2\right] \quad (2.48)$$

mas de 2.4:

$$a_t = \omega_t - \phi_1 \omega_{t-1} - \dots - \phi_p \omega_{t-p} + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (2.49)$$

Substituindo 2.49 em 2.48 obtém-se a densidade de probabilidade conjunta de  $\omega$  dado que o Jacobiano da transformação é unitário (Nelson, 1973; Priestley, 1981):

$$p(\omega | \underline{\phi}, \underline{\theta}, \sigma_a^2) = (2\pi)^{-\frac{n}{2}} \sigma_a^{-n} \exp\left[-\frac{1}{2\sigma_a^2} \sum_{t=1}^n (\omega_t - \phi_1 \omega_{t-1} - \dots - \phi_p \omega_{t-p} + \dots + \theta_q a_{t-q})^2\right] \quad (2.50)$$

A função de verossimilhança dos dados é portanto

$$L(\underline{\phi}, \underline{\theta}, \sigma_a^2 | \underline{\omega}) = (2\pi)^{-\frac{1}{2}n} \sigma_a^{-n} \exp\left[-\frac{1}{2\sigma_a^2} \sum_{t=1}^n a_t(\underline{\phi}, \underline{\theta})^2\right] \quad (2.51)$$

Pode-se observar que a Expressão 2.49 nem sempre pode ser calculada para todos os índices  $t$ , pois, conforme a ordem do modelo não se tem os valores de  $a_t$  ou  $\omega_t$  para  $t \leq 0$ . Para resolver o problema ou se ajustam os limites da somatória em 2.51 ou se assumem os valores não existentes como zero. Box e Jenkins (1970) sugerem também um método denominado "back-forecasting" para estimar os valores iniciais necessários para o cálculo de  $L(\bullet)$ . Assim sendo a função de verossimilhança  $L(\bullet)$  deixa de ser *exata* pois não usa todos os dados disponíveis ou então é necessário estabelecer hipóteses para as condições iniciais (c.i.). De qualquer forma para um grande número de dados de treinamento as c.i. não influenciariam muito o resultado. Refira-se a (Priestley, 1981) para maiores detalhes sobre a elaboração da verossimilhança exata.

O que se procura são os valores de  $\underline{\phi}$ ,  $\underline{\theta}$  e  $\sigma_a^2$  que maximizam a Equação 2.51. Para facilitar os cálculos e eliminando-se os valores constantes, basta maximizar a expressão:

$$\text{Log } L(\bullet) \propto l(\underline{\phi}, \underline{\theta}, \sigma_a^2) = -n \ln \sigma_a - \frac{S(\underline{\phi}, \underline{\theta})}{2\sigma_a^2} \quad (2.52)$$

onde

$$S(\underline{\phi}, \underline{\theta}) = \sum_{t=1}^n a_t(\underline{\phi}, \underline{\theta})^2 \quad (2.53)$$

Para maximizar 2.52 basta inicialmente minimizar  $S(\bullet)$ , e uma vez calculado esse valor e os parâmetros  $\hat{\underline{\phi}}$  e  $\hat{\underline{\theta}}$ , o valor de  $\sigma_a^2$  é dado por (derivando e igualando a zero):

$$\hat{\sigma}_a^2 = \frac{S(\hat{\underline{\phi}}, \hat{\underline{\theta}})}{n} \quad (2.54)$$

No cálculo dos parâmetros de modelos ARMA, em geral a expressão  $S(\bullet)$  será altamente não linear em relação aos parâmetros; a solução direta é portanto muito difícil na maioria dos casos. Para resolução aplicam-se então métodos de estimação não linear, por exemplo, um método baseado em aproximação linear do modelo, o chamado método de Gauss-Newton (GN) (Nelson, 1973), que é um método iterativo e que utiliza inicialmente as estimativas preliminares geradas pelos esquemas descritos nas seções anteriores. As diversas estimativas para os parâmetros são geradas até que  $S(\hat{\underline{\phi}}, \hat{\underline{\theta}})$  não mude mais dentro de certos limites e os valores alcançados para  $\hat{\underline{\phi}}$  e  $\hat{\underline{\theta}}$  na última iteração são as estimativas de máxima verossimilhança procurada.

Em princípio a função  $S(\underline{\phi}, \underline{\theta})$  pode ter vários mínimos locais, mas usando-se as estimativas preliminares como partida, normalmente se alcança o menor valor entre os mínimos (Nelson, 1973).

#### 2.4.5 - ERRO PADRÃO DOS PARÂMETROS ESTIMADOS

As estimativas de máxima verossimilhança dos parâmetros são v.a. desde que são função dos dados. Da teoria estatística (Nelson, 1973) sabe-se que em condições mais ou menos gerais para "grandes" amostras as estimativas são distribuídas normalmente com matriz de covariância dada por:

$$V(\underline{\beta}) = 2\sigma_a^2 \begin{bmatrix} \frac{\partial^2 S(\underline{\beta})}{\partial \beta_1^2} & \dots & \frac{\partial^2 S(\underline{\beta})}{\partial \beta_1 \partial \beta_{p+q}} \\ \vdots & \ddots & \vdots \\ \cdot & \dots & \frac{\partial^2 S(\underline{\beta})}{\partial \beta_{p+q}^2} \end{bmatrix}^{-1} \quad (2.55)$$

onde  $\underline{\beta} = [\underline{\phi}, \underline{\theta}]$

Para o cálculo de 2.55 utilizam-se os valores de  $\underline{\phi}$  e  $\underline{\theta}$  da última iteração para o cálculo de  $S(\underline{\beta})$  pelo procedimento de GN. Sob hipótese

nula o estimador de cada parâmetro tem variância  $\frac{\partial^2 S(\underline{\beta})}{\partial \beta_i^2} = \sigma_{\beta_i}^2$ . Portanto para

se saber determinar se determinado parâmetro é diferente de zero num nível de 5% basta verificar se  $|\beta_i| > 2\sigma_{\beta_i}$ . É possível estabelecer também, um intervalo de confiança para os parâmetros preliminares, como demonstrado em Brockwell e Davis (1987).

## 2.5 - MODELOS NÃO ESTACIONÁRIOS E SAZONAIS

### 2.5.1 - MODELOS NÃO ESTACIONÁRIOS

Quando uma série em estudo não é estacionária, o que ocorre frequentemente, procura-se através de transformações nos dados torná-la estacionária. Uma transformação muito utilizada é a função logarítmica ou a raiz quadrada por exemplo. Certos cuidados no entanto devem ser tomados quando se quiser fazer a previsão dos dados originais, pois nem sempre a previsão no campo original é dada pela inversa da transformação utilizada (Pereira, 1984, Seção 9.5).

O tipo de não-estacionariedade mais comum é a chamada não-estacionariedade homogênea quando sucessivas diferenças da série em estudo

são estacionárias, ou seja,  $z_t$  é dita uma série não-estacionária homogênea se a série :

$$w_t = \nabla^d z_t = (1-B)^d z_t \quad (2.57)$$

for estacionária;  $d$  é a ordem da diferenciação e  $(1-B)z_t = z_t - z_{t-1}$

Uma série temporal com não-estacionariedade homogênea é chamada um processo ARIMA (p,d,q) (Autoregressive Integrated Moving Average), onde  $d$  é o grau de diferenciação e  $w_t = \nabla^d z_t$  é um processo ARMA(p,q).

Como exemplo, podemos observar uma série cuja média seja função linear ascendente no tempo (também chamada de não-estacionariedade em *tendência*). Basta diferenciar-se uma vez e facilmente observa-se que tal série terá média constante. Muitas vezes, porém, será necessário conjugar uma transformação e uma diferenciação, para tornar a série estacionária.

Na Seção 2.6 ver-se-á como identificar uma possível não-estacionariedade a partir da análise da FAC e da FAP.

### 2.5.2 - MODELOS SAZONAIS

Sazonalidade é um fenômeno que ocorre com muita frequência em diversos processos naturais, particularmente os processos econômicos, onde praticamente tudo se refere a períodos anuais. Existem 2 métodos principais que procuram modelar a sazonalidade: um determinístico, onde o processo sazonal é considerado uma soma de um processo determinístico periódico com um processo estocástico; outro estatístico que permite certa aleatoriedade na componente sazonal. Aqui o processo estocástico é considerado uma multiplicação de componentes, uma não sazonal para modelar os aspectos intra-períodos e outras componentes quantas forem os aspectos periodicos determinados.

No primeiro método denominado de "decomposição clássica" o processo  $z_t$  é considerado como (Brockwell e Davis, 1987):

$$z_t = m_t + s_t + w_t \quad (2.58)$$

onde  $m_t$  é uma função que varia lentamente, chamada de tendência,  $s_t$  é a componente sazonal e  $w_t$  é o processo ARMA.

Observe que nesse caso  $w_t$  já é considerado estacionário pois o termo tendência é retirado da série, por exemplo, ajustando uma curva polinomial aos dados observados da série e retirando esses valores determinísticos em função de  $t$ , da série observada. A diferenciação sequencial discutida na seção anterior também poderia eventualmente ser aplicada aqui para se alcançar a estacionariedade. O termo  $s_t$  - sazonalidade-, aqui também uma componente determinística seria calculada, por exemplo, por considerar a média de valores correspondentes dentro de cada estação (por exemplo: a média do número de acidentes de carro em Janeiro num período de dez anos, depois de retirada a leve tendência do aumento do número de carros vendidos de ano para ano). Essas médias dos valores nas posições temporais dentro da estação, definem um período da função que se repete periodicamente durante todo o tempo a que se referem as observações. Das observações  $\omega_t$ , instâncias do processo  $z_t$ , extrai-se portanto a série  $w_t$  que é a série a ser modelada tendo sido retirada a tendência e a sazonalidade. Outros métodos podem ser vistos em (Brockwell e Davis, 1987; Granger e Newbold, 1977).

O modelo multiplicativo denominado SARIMA  $(p,d,q) \times (P,D,Q)_s$  (Seasonal Autoregressive Integrated Moving Average) é uma generalização do processo ARIMA (Nelson, 1973; Granger e Newbold, 1977), que leva em consideração a dependência de observações afastadas por períodos de tamanho  $s$ . Assim tem-se o modelo:

$$z_t = \kappa_1 z_{t-s} + \dots + \kappa_p z_{t-sp} + \eta_t - \delta_1 \eta_{t-s} - \dots - \delta_Q \eta_{t-sQ} \quad (2.59)$$

onde  $\eta_t$  é processo ruído branco e  $z_t$  tem média zero e é estacionário.

Denotando a diferença sazonal por  $w_t = z_t - z_{t-s}$  ou  $w_t = (1 - B^s)z_t$  para permitir séries estacionárias, chega-se ao modelo:

$$(1 - \kappa_1 B^S - \dots - \kappa_p B^{sP}) (1 - B^S)^D z_t = (1 - \delta_1 B^S - \dots - \delta_q B^{sQ}) \eta_t \quad (2.60)$$

onde D permite diferenças sazonais de maior ordem.

O modelo acima contém defeitos pois não contempla relações intra estações, ou seja as autocorrelações intermediárias às estações seriam zero, o que não é muito razoável na maioria dos casos.

Para modelar então essas interações entre atrasos unitários supõe-se que o ruído  $\eta_t$  siga um modelo ARMA.

Chega-se finalmente ao modelo geral multiplicativo :

$$(1 - \kappa_1 B^S - \dots - \kappa_p B^{sP}) (1 - \phi_1 B - \dots - \phi_p B^P) (1 - B)^d (1 - B^S)^D z_t = (1 - \delta_1 B^S - \dots - \delta_q B^{sQ}) (1 - \theta_1 B - \dots - \theta_p B^q) a_t \quad (2.61a)$$

Em uma notação mais compacta:

$$\kappa_p(B^S) \phi_p(B) \nabla^d \nabla^D = \delta_q(B^S) \theta_q(B) a_t + \mu \quad (2.61b)$$

O modelo é multiplicativo no sentido que a série final passa por 2 filtros ARMA, um simples e outro sazonal.

A estimação de parâmetros de um modelo SARIMA não é muito diferente (Brockwell e Davis, 1987, p. 314), de um processo ARIMA. A diferença  $w_t = (1 - B)^d (1 - B^S)^D z_t$  constitui um processo ARMA(p + sP, q + sQ) onde alguns dos coeficientes são zero. Expandindo-se as relações multiplicativas dos parâmetros do modelo para um dado conjunto p,P,q,Q,s e relacionando-se com os valores calculados para o modelo ARMA, acham-se os valores para  $\kappa(\bullet), \phi(\bullet), \delta(\bullet)$  e  $\theta(\bullet)$ .

## 2.6 - IDENTIFICAÇÃO DO MODELO

Nos desenvolvimentos efetuados até agora foi suposto que se conheciam as ordens do modelo ARMA ou (SARIMA se for o caso). A determinação da ordem entretanto não é trivial e se constitui na realidade na fase mais difícil do uso da metodologia de séries temporais, denominada fase de identificação do sistema.

A primeira ferramenta de análise da série temporal é o formato da própria série temporal (usualmente mostrada em um gráfico). Uma tendência polinomial ou mesmo exponencial (como é o caso da inflação brasileira) pode ser observada e essa tendência retirada. Transformações do tipo log ou raiz quadrada podem ser utilizadas, normalmente sugeridas por um conhecimento a priori sobre o processo. O objetivo é produzir uma série estacionária. Como alternativa para o caso de tendência polinomial pode-se optar por diferenças sucessivas da série, até que essa alcance um aspecto de média constante. Mesmo ainda nesse caso, ao se analisar a FAC, pode ser que se note que a mesma cai muito lentamente, sugerindo que o polinômio auto-regressivo modelo do processo tenha um zero muito próximo do círculo unitário. Uma amostra finita nunca terá uma FAC constante, mesmo que a série tenha sido gerada a partir de um modelo integrado de ordem  $d=1$ . Nesse caso será necessária ainda mais uma diferenciação em "lags" unitários. Uma vez eliminada a forte correlação unitária a sazonalidade poderá se manifestar. Reforço na FAC em atrasos periódicos sugere um modelo sazonal; se nesses atrasos a FAC cai lentamente pode ser necessária uma diferenciação sazonal.

Como se pode notar a fase de identificação necessita de uma certa experimentação. No caso de dúvida o analista retém algumas alternativas para escolha futura ao analisar o comportamento do resíduo, assunto que será discutido na seção sobre diagnóstico.

Depois de obtida a estacionariedade da série resta estabelecer as ordens  $p, q, P, Q$  das diferenças  $w_t = (1 - B)^d (1 - B^s)^D$ . Da teoria (Seção 2.3) sabe-se que a FAC se anula a partir do "lag"  $q$  para os modelos MA e a FAP também se anula a partir do atraso  $p$  para modelos AR; para o modelo ARMA completo nem a FAC nem a FAP se anulam. Observando-se portanto a FAC e a

FAP pode-se ter uma idéia do modelo subjacente a um processo. Na prática porém é preciso estabelecer um critério para saber quando as estimativas não são nulas, devido à variância dos estimadores das autocorrelações. Segundo Granger e Newbold (1977) que citam Anderson (1942) os estimadores da FAC e FAP são aproximadamente gaussianos para todos os "lags", quando o sinal é ruído branco, com variância  $n^{-1}$  (onde  $n$  é número de amostras), assim pode-se usar o limite de  $\pm 2n^{-1/2}$  (equivalente ao dobro do desvio padrão da distribuição do estimador) como um indicativo mais ou menos grosseiro para saber se as funções não são zero em certos atrasos, afastando-se assim da hipótese de ruído branco. Dessa maneira se obtém imediatamente dois valores provisórios para  $p$  e  $q$  ( $\bar{p}$  e  $\bar{q}$ ), correspondendo aos valores de "lag" para os quais a FAP e a FAC caem dentro do intervalo  $\pm 2n^{-1/2}$ , portanto consideradas nulas. Para se obter um valor para  $P$  e  $Q$  quando for o caso, analisa-se as  $FAC(s)$  e  $FAP(s)$  onde  $s$  é o período e  $l=0,1,2 \dots$

Para tornar o problema de identificação bem definido, foram introduzidos diversos índices, entre os quais, o mais utilizado é o chamado índice de AKAIKE (AIC) (Priestley, 1981), que é posto da seguinte forma:

$$AIC(p,q) = n \ln \hat{\sigma}_{p,q}^2 + 2(p+q) \quad (2.62)$$

onde  $\hat{\sigma}_{p,q}^2$  é igual à variância estimada para o ruído para um modelo ARMA(p,q).

A idéia geral para o uso desse critério de qualidade para o ajustamento é que, em geral ao se aumentar o número de parâmetros do modelo a variância do resíduo diminui e o resíduo fica mais "branco". Entretanto aumentar demasiadamente o número de parâmetros é indesejável, pois, além de se diminuir a precisão na estimativa dos parâmetros, esse fato contraria o chamado princípio da parcimônia, pelo qual se deve procurar modelos em sua forma equivalente a mais compacta possível.

De posse dos valores de  $\bar{p}$  e  $\bar{q}$  iniciais, escolhem-se algumas ordens  $p < \bar{p}$  e  $q < \bar{q}$  para estimar os parâmetros dos modelos ARMA(p,q). O modelo escolhido é aquele que minimiza o índice AIC.

Ainda assim verifica-se a necessidade de se ter que estimar os parâmetros de um número grande de modelos, principalmente se houver opções com relação ao grau de diferenciação e transformações utilizadas na série original. Para diminuir essas dificuldades pode-se se utilizar da informação sobre os desvios-padrões dos estimadores dos parâmetros para eliminar coeficientes pouco significativos. A literatura apresenta também outras funções auxiliares para a identificação do modelo tais como as autocorrelações inversas, autocorrelação parcial inversa, arranjos  $R^*$  e  $S^*$  e o método Corner; todos esses revisados em Pereira (1984). Esses métodos também facilitam a identificação de processos SARIMA pois muitas dessas funções auxiliares indicam quais coeficientes intermediários devem ser eliminados. No caso de identificação de modelos sazonais o índice AIC deve ser modificado para (Brockwell e Davis, 1987, p. 296):

$$AIC_s = 2 n \ln \sigma_s + 2m \quad (2.63)$$

onde  $m$  é o número de coeficientes estimados.

## 2.7 - DIAGNÓSTICO E REVISÃO DO MODELO

Depois de estimado um modelo tentativo para os dados existentes, é necessário verificar a adequabilidade do modelo à esses mesmos dados. Esse procedimento é denominado de *diagnóstico* e serve também para orientar a modificação do modelo quando necessário. O diagnóstico se baseia nos chamados *resíduos*.

### 2.7.1 - OBTENÇÃO DOS RESÍDUOS

Suponha que o modelo ARMA(p,q) obtido a partir de 2.61, onde  $w_t = \nabla^d \nabla^D z_t$

$$\phi(B)w_t = \theta(B)a_t \quad (2.64)$$

foi ajustado às observações  $\omega_n = [\omega_1, \dots, \omega_n]$  com estimativas  $\hat{\phi}$  e  $\hat{\theta}$ .

A quantidade  $\hat{a}_t$  em (2.65) é conhecida como *resíduo*:

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)\omega_t \quad (2.65)$$

Se este modelo estiver correto, então os resíduos  $\hat{a}_t$  constituem um processo ruído branco e como visto antes, as autocorrelações amostrais desses resíduos devem ser aproximadamente zero para "lags" maiores que zero. Na prática porém como os coeficientes de 2.65 são estimados, então a sequência  $\hat{a}_t$  não é sequência independente e identicamente distribuída Brockwell e Davies (1987, p. 299), e a distribuição das correlações amostrais de  $\hat{a}_t(\hat{\phi}, \hat{\theta})$  não é a mesma de  $\hat{a}_t(\phi, \theta)$ , embora se possa esperar que as correlações amostrais de  $\hat{a}_t$  possam dar informação valiosa a respeito da adequabilidade do modelo. O que sucede é que, como apontam Granger e Newbold (1977); Brockwell e Davis (1987) e outros, o desvio padrão de  $r_k(\hat{a}(\hat{\phi}, \hat{\theta}))$ , pode ser consideravelmente menor que o desvio padrão do estimador com o valor real dos parâmetros se esses fossem conhecidos, embora o limite de  $\pm 2n^{-1/2}$  para se considerar uma autocorrelação como zero ainda continue válido para grandes atrasos. Para atrasos pequenos (menor que 6 para modelos não muito complicados (Granger e Newbold, 1977), o intervalo de confiança pode cair até a metade com relação ao intervalo que seria se conhecesse os parâmetros reais (Box e Jenkins, 1970). Como a variação do intervalo de confiança é função complicada do modelo, Granger e Newbold (1977) sugerem que se use o intervalo de confiança padrão observando se, pelo menos para  $k < 6$ , o valor da autocorrelação se afasta dos limites do intervalo de confiança, ficando mais próximo de zero.

No campo de processamento de sinais a operação de cálculo dos resíduos é denominada de *filtragem inversa*, no sentido em que a filtragem direta se dá na configuração teórica do modelo.

No cálculo dos resíduos, aparece a questão das condições iniciais, já mencionada na seção de estimativa de máxima verossimilhança. Normalmente utiliza-se os mesmos critérios já mencionados, mesmo porque ao término da última iteração no cálculo da estimativa de max. ver. os resíduos já estão estabelecidos.

### 2.7.2 - USO DOS RESÍDUOS PARA MODIFICAR O MODELO

Suponha que a análise dos resíduos indicou a não adequabilidade do modelo

$$\phi(B)z_t = \theta(B)a_t \quad (2.67)$$

e que o modelo identificado e estimado para o resíduo foi:

$$\phi'(B)a_t = \theta'(B)a'_t \quad (2.68)$$

Substituindo-se o Modelo 2.68 em 2.67 temos o novo modelo tentativa:

$$\phi'(B)\phi(B)z_t = \theta'(B)\theta(B)a'_t \quad (2.69)$$

### 2.8 - RESUMO DA ESTRATÉGIA DE MODELAMENTO DE SÉRIES TEMPORAIS

- 1) A série é diferenciada em relação ao atraso unitário ou sazonal, para produzir estacionariedade. Transformações nos dados ou decomposição sazonal poderão ser necessárias.
- 2) Inspeccionando-se a FAC e a FAP da série transformada, ordens máximas provisórias para p e q são escolhidas.
- 3) Vários modelos ARMA(p,q)  $p \leq \bar{p}$  e  $q \leq \bar{q}$  são estimados, o AIC calculado. Os erros sobre os coeficientes podem orientar quais outros modelos tentar.
- 4) Escolhe-se o modelo com menor AIC.
- 5) Realiza-se um diagnóstico sobre os resíduos para observar se os mesmos aparentam ser provenientes de um processo ruído branco.
- 6) Modifica-se o modelo, caso existam correlações significativas nos resíduos.

## CAPÍTULO 3

### MODELOS MULTIVARIADOS E BIDIMENSIONAIS

#### 3.1 - MODELOS MULTIVARIADOS

Uma extensão direta das séries temporais univariadas são as chamadas séries temporais multivariadas, ou múltiplas, ou multicanais ou vetoriais, onde se coletam dados de diferentes fontes que tem relação entre si, em períodos de tempo igualmente espaçados. Muitos trabalhos em processamento de imagens utilizam uma formulação vetorial ora considerando uma linha como vetor e as colunas como índices, ora considerando uma coluna de um conjunto de linhas como um vetor e posição na linha como índice (por exemplo: Therrien, 1986).

Neste trabalho, não será usado diretamente a formulação vetorial, mas para esclarecer algumas deduções e facilitar algumas comparações, apresentar-se-á uma breve introdução ao assunto.

Utilizam-se eventualmente as siglas ARMAV ou MARMA para identificar esses modelos.

##### 3.1.1 - DEFINIÇÕES

Sejam  $Z_t^T = [ z_{1t}, \dots, z_{mt} ]$  um vetor de  $m$  séries temporais centralizadas (isto é subtraídas das médias) e  $A^T = [ a_{1t}, \dots, a_{mt} ]$  um vetor de  $m$  ruídos brancos e com matriz de covariância cruzada:

$$E(A_t A_{t+k}^T) = \begin{cases} \Sigma & k = 0 \\ 0 & k \neq 0 \end{cases} \quad (3.1)$$

Suponha que  $Z_t$  é estacionário (no sentido lato), eventualmente depois de transformações convenientes, isto é, cada  $z_t$  é estacionário e as covariâncias cruzadas  $cov(z_{it}, z_{jt+k})$ , só dependam de  $k$  ( $\forall i, j$ ).

DEFINIÇÃO 3.1 - (Série Temporais Multivariadas)  $Z_t$  é um processo MARMA(p,q) se satisfaz a equação diferença:

$$Z_t - \Phi_1 Z_{t-1} - \dots - \Phi_p Z_{t-p} = A_t - \Theta_1 A_{t-1} - \dots - \Theta_q A_{t-q} \quad (3.2)$$

onde  $\Phi_i$  e  $\Theta_i$  são matrizes (mxm) de parâmetros.

A Equação 3.2 pode ser escrita em forma mais compacta como:

$$\Phi(B)Z_t = \Theta(B)A_t \quad (3.3)$$

onde

$$\Phi(B) = \Phi_0 - \Phi_1 B - \dots - \Phi_p B^p$$

$$\Theta(B) = \Theta_0 - \Theta_1 B - \dots - \Theta_q B^q$$

$$\Phi_0 = \Theta_0 = I_m$$

### 3.1.2 - PROPRIEDADES E ESTIMATIVA DO MODELO MARMA.

PROPOSIÇÃO 3.1 - (Estabilidade do filtro-modelo) O modelo é dito estável (gera uma série estacionária) se os zeros do polinômio  $|\Phi(B)|$  estão fora do círculo unitário. Neste caso o modelo pode ser escrito na forma de média móvel de ordem infinita.

DEMONSTRAÇÃO: ver Hannan (1970) e Pereira (1984)

OBSERVAÇÃO 3.1 - Existe uma forma alternativa para o modelo vetorial onde  $\Phi_0 \neq I_m$  mas  $\Sigma$  é diagonal, sendo essa forma preferida por economistas (Pereira,1984). Gambotto(1980) define  $\Phi_0$  como triangular superior, com 1's na diagonal principal, para que se facilite a correspondência entre o modelo vetorial e bidimensional sobre suporte de meio plano não simétrico.

Para modelos vetoriais, no entanto, aparece uma dificuldade adicional, que é o chamado problema da *identificabilidade* (Priestley, 1981,

p. 801; Brockwell e Davies, 1987, p. 418) ou seja não basta exigir inversibilidade do modelo ( $|\Theta(B)|$  com zeros fora do círculo unitário) para que se garanta a unicidade entre os parâmetros do modelo e sua função matriz de autocovariância. Para o modelo na forma da Equação 3.3 e com as condições de estabilidade e inversibilidade resolvidas, uma condição necessária e suficiente para representação única é:

$$\text{Rank}(\Phi_p, \Theta_q) = m$$

É possível também, como no caso univariado, estabelecer uma relação entre os parâmetros MARMA e a função matriz de covariância (Kay, 1988; Pereira, 1984), aqui definida como :

$$\Gamma(k) = E(Z_t Z_{t+k}^T) \quad (3.4)$$

Para o modelo MAR(p) a relação entre os parâmetros e a função matriz de covariância é a equação de Yule-Walker que no formato matricial se torna:

$$\Gamma_p \Phi = \underline{\Gamma} \quad (3.5)$$

onde

$$\Gamma_p = \begin{bmatrix} \Gamma(0) & \Gamma(-1) & \dots & \Gamma(-p+1) \\ \Gamma(+1) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \Gamma(-1) \\ \Gamma(p-1) & \dots & \Gamma(+1) & \Gamma(0) \end{bmatrix}$$

e

$$\underline{\Phi} = \begin{bmatrix} \Phi_1 \\ \vdots \\ \vdots \\ \Phi_p \end{bmatrix} \quad \underline{\Gamma} = \begin{bmatrix} \Gamma(1) \\ \vdots \\ \vdots \\ \Gamma(p) \end{bmatrix}$$

$\Gamma_p$  é uma matriz simétrica já que  $\Gamma(-k) = \Gamma^T(k)$  ( $T$  significa transposta), e além disso é "block" Toeplitz o que permite usar o algoritmo de Levinson multicanal (Wiggins e Robinson, 1965), para obter os parâmetros dos modelos MAR(p) em ordem crescente. Usa-se para isso as estimativas, preferivelmente polarizadas, para garantir que  $\hat{\Gamma}_p$  seja positivo definida. (Kay, 1988, p. 463).

Brockwell e Davis (1987) apresenta uma versão do algoritmo das inovações para o caso MMA(q), que obtém recursivamente os parâmetros do modelo. Não há, aparentemente porém, um trabalho que apresente uma maneira sistemática de se obter as estimativas preliminares para um modelo MARMA(p,q) completo. Kay (1988) sugere a possibilidade de uso da equação de YW modificada para se obter os parâmetros AR e por filtragem inversa obter uma seqüência MA. Assinala porém que não se conhece uma maneira recursiva de se obter os parâmetros de modelo desse tipo. Brockwell e Davis (1987) sugere que se use como estimativas iniciais, as estimativas univariadas das componentes, para o procedimento de máxima verossimilhança, que pode ser condicional ou exata (Tiao e Box, 1981; Pereira, 1984).

### 3.1.3 - IDENTIFICAÇÃO E VERIFICAÇÃO

Como no caso univariado as duas principais ferramentas de identificação são a função matriz de autocorrelação (FMAC) e a função matriz de aut. parcial (FMAP). Para se facilitar a interpretação, no lugar de se apresentar uma seqüência de matrizes com valores reais em função do tempo, Tiao e Box (1981) sugerem utilizar-se o símbolo "+" para indicar que uma estimativa é maior que 2 desvios padrões do estimador, "-" para indicar uma estimativa menor que menos duas vezes o desvio padrão e "." quando está no intervalo. As correlações cruzadas são normalizadas pelas autocovariancias

do atraso zero, e as correlações parciais são obtidas utilizando-se do procedimento de Levinson multicanal.

A função matriz de autocorrelação teria o seguinte aspecto para um modelo do tipo MMA(2) (m=2) (Figura 3.1):

atraso	0	1	2	3	...
matrizes	$\begin{bmatrix} + & - \\ - & \cdot \end{bmatrix}$	$\begin{bmatrix} + & + \\ - & - \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	

Fig. 3.1 - Função matriz de autocorrelação.

Pereira (1984) apresenta em seu trabalho, outras funções que como a FMAP e FMAP ajudam no processo de determinação da ordem do modelo. Da mesma forma que no caso univariado, para tornar o problema de indentificação bem definido, usa-se um índice que é calculado para cada um dos modelos propostos. Seleciona-se aquele para o qual esse índice é mínimo. Um dos índices mais utilizados é o índice de Akaike para o caso multivariado como definido em Brockwell e Davies (1987).

### 3.2 - MODELOS BIDIMENSIONAIS

#### 3.2.1 - DEFINIÇÕES E REGIÕES DE SUPORTE

Uma imagem é função de duas variáveis independentes e assim a noção de causalidade, (muito útil em sistemas 1D, onde essa propriedade é uma restrição física razoável e está ligada a consecução de sistemas de processamento em tempo real) não é tão naturalmente estendida a sistemas 2D pois somente uma das variáveis independentes pode ser associada ao tempo.

Entretanto, uma propriedade de sistemas 1D causais, a *computabilidade recursiva* também é útil em sistema 2D. *Computabilidade recursiva* nos permite calcular a saída baseado num conjunto finito de saídas

passadas, assim como um conjunto finito de condições iniciais e entradas recentes.

DEFINIÇÃO 3.2 - (Computabilidade Recursiva): Um sistema é *recursivamente computável*, se existe um ordenamento para o processamento tal que, os valores que se necessitam para calcular o valor corrente são sempre disponíveis.

Modelos ARMA unidimensionais supõem a causalidade, assim para a extensão 2D dos modelos ARMA se supõe a computabilidade recursiva (Kay, 1988; Dudgeon e Mersereau, 1984). A região de suporte no plano mais geral e que possui a propriedade de computabilidade recursiva é a região definida como o suporte de meio plano não simétrico ("NSHP Suport") (Figura 3.2).

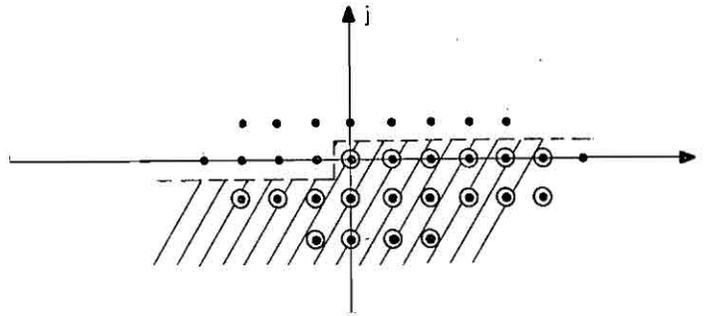


Fig. 3.2 - Região de suporte de meio plano não simétrico .

(região hachurada na figura)

Essa região de suporte é naturalmente definida pelo esquema de varredura para aquisição de imagens em câmaras de TV. Um caso especial ocorre quando o suporte é confinado a um quarto de plano (QP) como mostrado na Figura 3.3.

Uma versão finita do suporte de meio plano não simétrico pode ser definida pela Equação 3.7.

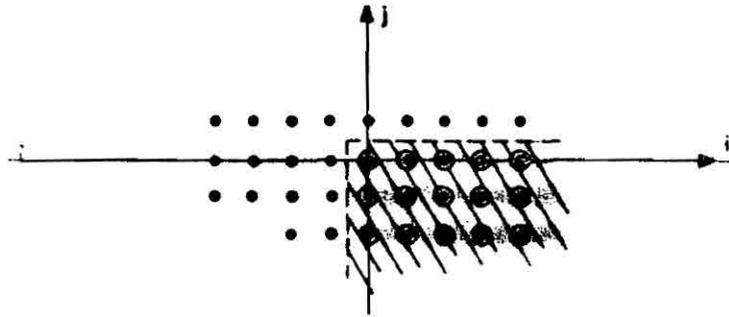


Fig. 3.3 - Suporte de quarto de plano.

DEFINIÇÃO 3.3 - (Suporte de meio plano não simétrico truncada): A região de suporte de meio plano não simétrico truncada (TNSHP), é a região do plano definida pela Equação 3.7 e corresponde a um subconjunto do suporte de meio plano não simétrico (Figura 3.4).

$$S_x(x_1, x_2, x_3) = \{(i, j) \mid i = -x_3, \dots, 0, \dots, x_1; j = -x_2, \dots, -1\} \cup \{(i, j) \mid j=0; i = 0, \dots, x_1\} \quad (3.7)$$

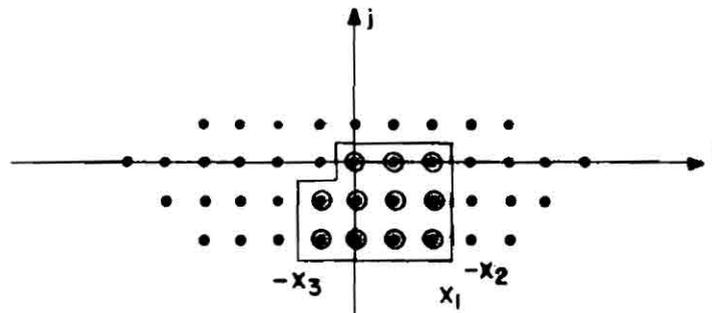


Fig. 3.4 - Região de suporte de meio plano não simétrico truncada.

A região NSHP é portanto igual à região TNSHP quando  $x_1, x_2, x_3 \rightarrow \infty$ . A região Q.P. é obtida fazendo  $x_3 = 0$ . Therrien et al. (1986) provam que a região NSHP é suficientemente geral para caracterizar qualquer processo aleatório com função covariância positivo definida.

$$H(z_m, z_n) = \frac{\sum_{(i,j) \in S_q} \theta_{i,j} z_m^{-i} z_n^{-j}}{\sum_{(i,j) \in S_p} \phi_{i,j} z_m^{-i} z_n^{-j}} = \frac{\Theta(z_m^{-1}, z_n^{-1})}{\Phi(z_m^{-1}, z_n^{-1})} = \frac{Z(z_m, z_n)}{A(z_m, z_n)} \quad (3.9)$$

### 3.2.2 - FUNÇÃO DE AUTOCORRELAÇÃO E AUTOCOVARIANÇA

Os coeficientes preliminares dos modelos são em geral obtidos da função de autocovariância do processo estocástico. No caso bidimensional, o campo aleatório é considerado estacionário (no sentido lato) se as estatísticas do processo até 2ª ordem não dependem da posição no plano, mas apenas do intervalo espacial.

A função ou sequência de autocovariância é definida por

$$\gamma_{k,l} = \text{cov}(z_{m,n}, z_{m+k, l+n}) = E\{(z_{m,n} - \mu_z)(z_{m+k, n+l} - \mu_z)\} \quad (3.10)$$

onde  $\mu_z = E(z_{m,n})$

A expressão para a estimativa polarizada da autocovariância para o campo aleatório, utilizando amostras em uma área retangular de  $MXN$  amostras é dada por (Marple, 1987):

$$c_{k,l} = \begin{cases} \frac{1}{MN} \sum_{m=0}^{M-1-k} \sum_{n=0}^{N-1-l} (z_{m+k, n+l} - \hat{\mu}_z)(z_{m,n} - \hat{\mu}_z); & k \geq 0, l \geq 0 \\ \frac{1}{MN} \sum_{m=0}^{M-1-k} \sum_{n=-1}^{N-1} (z_{m+k, n+l} - \hat{\mu}_z)(z_{m,n} - \hat{\mu}_z); & k \geq 0, l < 0 \\ c(-k, -l) & k \leq 0, \forall l \end{cases} \quad (3.11)$$

### 3.2.4 - MÉTODO PARA ESTIMATIVA PRELIMINAR DE PARÂMETROS DE MODELOS ARMA 2D.

#### 3.2.4.1 - AS DIFICULDADES DE SE OBTER OS PARÂMETROS DO MODELO ARMA COMPLETO 2D.

Como foi observado na seção anterior são maiores as dificuldades para a extensão da teoria 1D para o campo 2D. Como Marzetta (1980) assevera, apenas se pode estender as propriedades de teoria de predição linear quando se usa o tipo de suporte contínuo (meio plano) como definido em seu trabalho. Como na prática isso não é possível, dificuldades continuam. O autor inclusive sugere o estudo de suporte descontínuo 1D como base para o estudo e extensão para o caso 2D. Como aparentemente não existe uma maneira sistemática de ajustar modelos de média móvel puros à sequência bidimensional e por conseguinte os modelos ARMA completos, na prática tem sido utilizado basicamente modelos AR e de predição linear (Modelos MVPR) em suporte de quarto de plano, para se utilizar do algoritmo Levinson 2D.

Algumas referências (Tou, 1981; Tou et al., 1976) têm utilizado modelos ARMA completo com região de suporte de quarto de plano, propostos com base na inspeção visual das FAC's e FAP's unidimensionais nas direções vertical e horizontal. Tal método no entanto se ressentem da maior sistematização e serve apenas para modelos bem curtos propostos com base empíricas. Permanece inclusive grande dificuldade na estimativa preliminar dos parâmetros da parte MA.

A literatura de séries temporais espaciais apresenta em muitos pontos semelhanças com a teoria de séries bidimensionais. Nestas séries uma coordenada é sempre o tempo e as outras coordenadas independentes são espaciais (Aroian, 1985; Haugh, 1984). Na equação diferença que define esse tipo de modelo a variável de saída no tempo  $t$  e em certa localização espacial é resultado da combinação linear de amostras desse mesmo processo estocástico em tempos anteriores e localização espaciais contíguas, assim como de "shocks" anteriores no tempo em localizações contíguas e uma amostra de "shock" no tempo presente e na mesma localização espacial. Por exemplo, a Equação 3.13

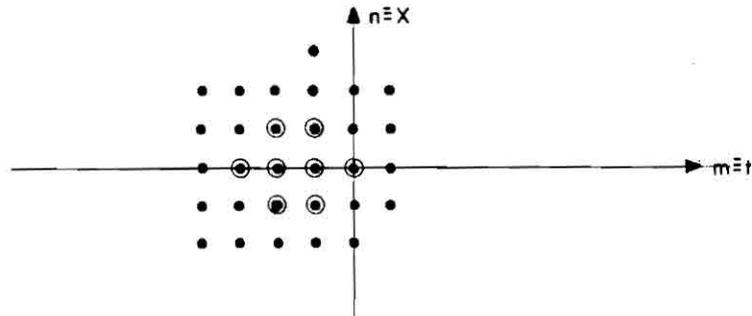


Fig. 3.6 - Ajuste de um modelo de série temporal espacial de dimensão  $m = 1$ , no plano, fazendo coordenada  $t \equiv m$ . A região de suporte não é equivalente nem à NSHP nem à Q.P.

#### 3.2.4.2 - A UTILIZAÇÃO DE MODELOS UNIDIMENSIONAIS SAZONAIS PARA GERAR MODELOS 2D.

Uma maneira alternativa é procurar adaptar o uso de modelos unidimensionais para o caso bidimensional (imagens).

MacCormick e Jayaramamurthy (1974) consideraram modelos sazonais unidimensionais obtidos a partir de linhas concatenadas da imagem para modelar texturas. Usaram os modelos para síntese com razoável sucesso; porém, texturas com estrias no sentido da concatenação não obtiveram bom desempenho. Os modelos utilizados foram selecionados de forma empírica a partir de indicações obtidas na plotagem das funções de autocorrelação e FAP. Uma dificuldade adicional reside no fato de que modelos sazonais dão origem apenas a modelos bidimensionais multiplicativos, o que nem sempre é satisfatório.

Jain (1989) apresenta o uso de modelos unidimensionais AR para caracterizar a imagem como sendo um sinal de varredura saída de uma câmara quando a relação entre linhas é desprezada, em problemas de compressão de imagens.

que correspondem a correlações entre linhas sobrepostas. Dos atrasos  $\frac{M}{2} + 1$  até  $M$  as correlações calculadas incluem um maior número de fatores que correspondem a pontos de linhas superiores e de linhas inferiores e à esquerda dos pontos superiores. De  $M$  até  $\frac{3M}{2}$  a relação se torna majoritária entre pontos superiores e inferiores à direita. As correlações calculadas em atrasos múltiplos de  $M$  correspondem a correlações bidimensionais entre pontos na vertical. O exemplo da Figura 3.8 ilustra o relacionamento existente entre as correlações calculadas sobre a série concatenada e a série bidimensional para o caso do cálculo da correlação do atraso  $M-1$ . No exemplo fica claro que a maior parte dos produtos cruzados se refere ao cálculo da correlação  $\gamma_{-1,-1}$ . Apenas um produto, que se ajusta dentro do tamanho do segmento corresponde à correlação  $\gamma_{0,M-1}$ .

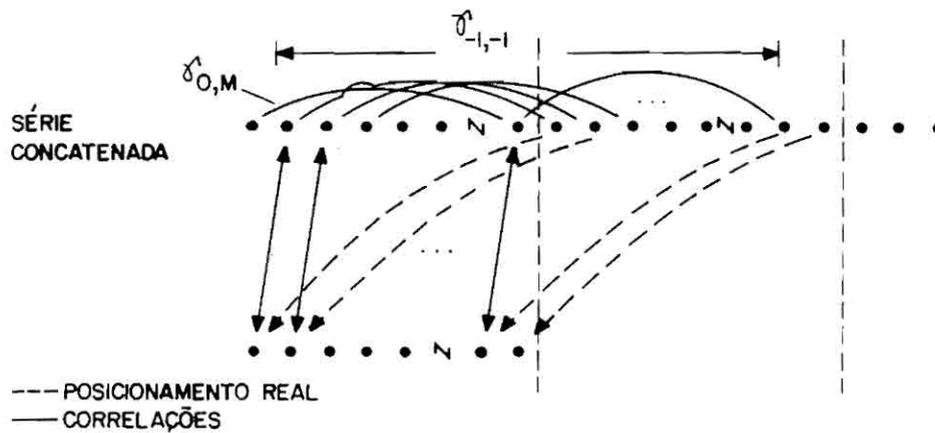


Fig. 3.8 -Relacionamento entre correlações calculadas sobre a série concatenada e o posicionamento no plano dos pontos envolvidos (para o atraso  $M-1$ ).

Do raciocínio exposto conclui-se que, se  $M$  for suficientemente grande, o cálculo da FAC sobre a série concatenada gera e distribui as correlações bidimensionais sobre um suporte de meio plano não simétrico.

Os valores das autocorrelações calculadas sobre a série unidimensional acumulam tantos erros quanto mais se afastarem os atrasos dos atrasos  $0, M, 2M$ , etc em direção a  $\frac{(2n+1)M}{2}$ . Passando para os atrasos  $\frac{(2n+1)M}{2} - 1$  os fatores acumulados passam a preponderar para os atrasos bidimensionais de uma linha abaixo. A Figura 3.9 indica, em função da posição do ponto no NSHP, o número de fatores acumulados erroneamente.

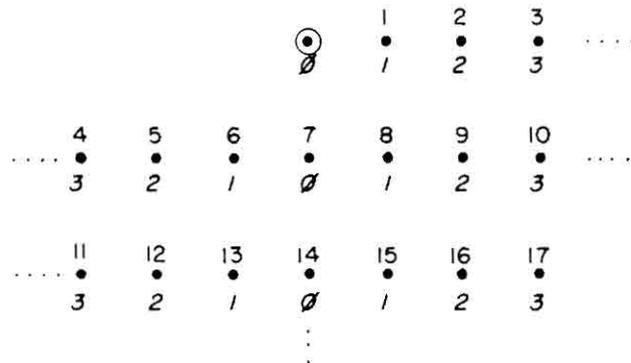


Fig. 3.9 - Indicação dos termos acumulados erroneamente em função da localização em suporte NSHP (número inferiores). Índices sobre a série concatenada  $i_c$  (números superiores),  $M=7$ .

Deve-se observar também que as correlações mais importantes que são justamente aquelas mais próximas do ponto de referência incluem menor número de fatores errôneos.

A transformação do índice  $i_c$  sobre a série concatenada para o índice  $(i,j)$  bidimensional é dada pelo seguinte algoritmo:

ALGORITMO 3.1 - (Transformação de coordenadas unidimensionais para bidimensionais)

$i_c = 0,1,2,3, \dots$                       coeficiente da série unidimensional concatenada

$t = \frac{i_c}{M}$                                       (divisão inteira)

se

(esquerda)       $(t-1)M + \frac{M}{2} < i_c < tM \Leftrightarrow (i,j) = \left[ -(M-i_c \text{ Mod } M), -t \right]$                       (3.14a)

(direita)         $(tM) \leq i_c \leq tM + \frac{M}{2} \Leftrightarrow (i,j) = \left[ i_c \text{ Mod } M, -t \right]$                       (3.14b)

Para evitar a inclusão de fatores que se acumulem erroneamente para um certo atraso equivalente no plano, seria necessário alterar a fórmula do cálculo da autocorrelação, alterando os limites da acumulação dos autoprodutos que seriam divididos em blocos, ou seja, usar-se-ia a Equação 3.11 numa versão linearizada. Outra possibilidade seria incluir zeros entre segmentos concatenados conforme se verá na seção seguinte.

#### 3.2.4.4 - O EFEITO DA INCLUSÃO DE ZEROS ENTRE SEGMENTOS CONCATENADOS

Uma alternativa para não se alterar a fórmula de cálculo da FAC unidimensional seria incluir M-1 zeros entre uma linha concatenada e outra. Esse procedimento separa os produtos antes acumulados erroneamente (Figura 3.10).

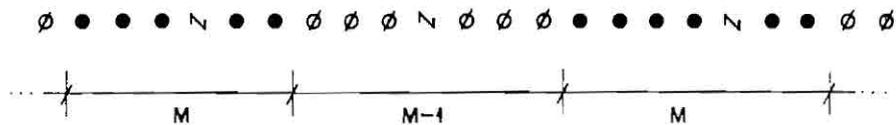


Fig. 3.10 - Colocação de zeros entre segmentos concatenados.

Esse esquema corresponderia ao chamado método da "correlação" ou janelado para predição linear (Therrien et al., 1986, p. 535 ) (ver também Seção 2.3.1 e refs.).

De qualquer maneira para M grande, mesmo não sendo feito o janelamento, os termos agregados incorretamente ao somatório tenderiam a dar contribuições globais nulas pois estão distantes entre si, logo independentes, e portanto não correlacionados.

Esse raciocínio leva a crer então que a estimativa da FAC sobre os dados linearizados reproduzem com razoável fidelidade as reais estimativas bidimensionais.

TABELA 3.1

CORRESPONDÊNCIA ENTRE A POSIÇÃO CONCATENADA E A POSIÇÃO NO PLANO: CASOS COM E SEM ZEROS ADICIONADOS.

SEM ZEROS					COM ZEROS				
$i_c$	(i,j)	OK	ZERO	ERRO	$i_c$	(i,j)	OK	ZERO	ERRO
1	(1,0)	6	0	1	1	(1,0)	6	7	0
2	(2,0)	5	0	2	2	(2,0)	5	8	0
3	(3,0)	4	0	3	3	(3,0)	4	9	0
4	(-3,-1)	4	0	3	4	(4,0)	3	10	0
5	(-2,-1)	5	0	2	5	(5,0)	2	11	0
6	(-1,-1)	6	0	1	6	(6,0)	1	12	0
7	(0,-1)	7	0	0	7	(-6,-1)	1	12	0
8	(1,-1)	6	0	1	8	(-5,-1)	2	11	0
9	(2,-1)	5	0	2	9	(-4,-1)	3	10	0
10	(3,-1)	4	0	3	10	(-3,-1)	4	9	0
11	(-3,-2)				11	(-2,-1)	5	3	0
12	...				12	(-1,-1)	6	7	0
13					13	(0,-1)	7	6	0
					14	(1,-1)	6	7	0
					15	(2,-1)	5	8	0
					16	(3,-1)	4	9	0
					17	...			

Uma ordem  $p=10$ , sobre a série linearizada, é suficiente para gerar coeficientes num suporte  $S(1,1,1)$ . (ver Figura 3.9 e Tabela 3.1).

Um modelo AR(10) dá origem à equação de YW unidimensional dada por (Equação 2.18)

$$\gamma_s = \sum_{i=1}^{10} \phi_i \gamma_{s-i} \quad s > 0 \quad (2.18a)$$

Ao se comparar a Equação 3.16 com a Equação 3.18 pode-se observar que pelo método da concatenação são gerados todos os coeficientes até os limites  $\frac{m}{2}$  no sentido da linha concatenada.

Porém, como as correlações estimadas pelo processo concatenado são praticamente as mesmas do que as correlações calculadas pelos estimadores 2D e considerando que, por hipótese, os coeficientes AR no caso 2D não existem fora de  $S(1,1,1)$  então se espera que em 3.16  $\phi_{20} = \phi_{30} = \phi_{-3,-1} = \phi_{-2,-1} = \phi_{2,-1} = \phi_{3,-1} = 0$ . Nestas condições a Equação 3.16 é equivalente a Equação 3.18. O mesmo ocorre com as outras equações do Sistema 2.18 que encontrem correspondência no suporte NSHP. Assim, observa-se que a equação YW-2D gera as mesmas estimativas que a equação de YW concatenada.

O fato de que os coeficientes auto-regressivos são nulos fora do suporte TNSHP, é usado na fase de identificação do modelo 2D observando-se os coeficientes gerados por um ajuste de 1 modelo AR de ordem conveniente, sobre a série unidimensional. Os coeficientes que estiverem dentro do intervalo de confiança são considerados nulos. Dessa maneira a região de suporte se define automaticamente.

O mesmo também acontece para modelos ARMA em geral onde os coeficientes da parte média móvel do modelo também se distribui no suporte NSHP. Dos coeficientes cujos valores caírem dentro do intervalo de confiança são considerados nulos e também a região de suporte de média móvel será determinada automaticamente.

### 3.2.5 - IDENTIFICAÇÃO E ESTIMATIVA DE MODELO ARMA 2D USANDO CONCATENAÇÃO

#### Estimativas preliminares

Conforme discutido na seção anterior, a estrutura de autocorrelação levantada sobre a série concatenada (se o tamanho do segmento a ser concatenado for suficientemente grande) é equivalente à estrutura de autocorrelação 2D, só que apresentada em formato concatenado ou sazonal e existe um mapeamento de índices concatenados para os índices sobre o plano.

### A Estimativa de Máxima Verossimilhança

No cálculo da estimativa de máxima verossimilhança os resíduos são calculados a cada iteração. Ocorre também aqui o fenômeno do cruzamento de fronteiras, mas aqui o efeito é menor do que no cálculo das autocorrelações pois apenas 1 entre M elementos vai ter sua previsão comparada com um valor de 1 ponto que não realmente seu vizinho. Como esse ponto não é correlacionado com seu vizinho o resíduo tenderia ser um pouco maior. O efeito global no entanto pode ser desprezado.

### O Diagnóstico Sobre os Resíduos

De qualquer forma ao se aplicar o teste diagnóstico sobre os resíduos pode-se confirmar a adequabilidade do modelo. No caso unidimensional uma simples inspeção no gráfico das autocorrelação (médias retiradas) pode-se rejeitar ou não a hipótese das autocorrelações sejam oriundas de processo ruído branco se algum valor ultrapassar o intervalo de confiança de  $\pm 1.96N^{-1/2}$ , traçado no gráfico como 2 retas paralelas.

No caso bidimensional, torna-se mais complicada a representação da FAC 2D, pois a impressão de números a serem comparados com o intervalo de confiança torna a visualização confusa.

Tiao e Box (1981) propuseram para o caso multivariado uma *máquina* para visualização da função matriz de autocorrelação (Seção 3.1.3). Baseado nessa idéia é proposta também uma *máquina* para visualizar a FAC 2D: imprime-se uma matriz cujos elementos são os símbolos: + se a autocorrelação superar positivamente o intervalo de confiança, - se superar negativamente e . se ficar no interior. Os símbolos são colocados numa matriz de dimensão  $(2M + 1) \times N$  onde M e N são valores máximos calculados para a FAC que é simétrica na parte superior. O valor 1 indica o ponto de referência (0,0).

$$A(z_m, z_n) \neq 0 \quad |z_m| = 1 \quad \text{e} \quad |z_n| \leq 1 \quad (3.20)$$

Gambotto (1980) sustenta que (3.20) vale para suporte TNSHP, só que nesse caso  $z_m$  e  $z_n$  não são mais intercambiáveis. Apesar da condição 3.20 ser mais simples ainda assim é necessário certos procedimentos, algo complicados para testar as condições (Dudgeon e Mersereau, 1984).

Um critério mais simplificado apresentado por Aroian (1985) estabelece que os polos  $z_m$  da equação característica  $A(z_m, z_n)$  devem cair dentro do círculo unitário, quando  $z_n = 1$ , e  $z_n$  também devem cair dentro do círculo unitário considerando-se  $A(1, z_n)$ . Como esse teste é muito mais simples que 3.20 conjetura-se que não é equivalente à estabilidade no sentido BIBO. Willsky (1979, p. 174) afirma que a solução da equação normal para o suporte TNSHP é estável.

No processo de identificação proposto a identificação e cálculo dos parâmetros é feita sobre o modelo linearizado, logo os cálculos das condições de estabilidade são feitos sobre o modelo linearizado sazonal, conforme regras estabelecidas na Seção 2.3.7. Se o modelo for estável também o será no plano pois é possível pensar num mapeamento que linearizaria a imagem, aplicar-se-ia o filtro sazonal (por causa da computabilidade recursiva) e deslinearizaria-se novamente. Ocorre que a rigor a série temporal linearizada não é estacionária por causa da transição de fronteiras, mas como esses pontos são minoria e não correlacionados entre si, pode-se conjeturar que a série linearizada é estacionária quando o campo aleatório é estacionário.

## CAPÍTULO 4

### CLASSIFICAÇÃO ESTATÍSTICA DE IMAGENS

#### 4.1 - INTRODUÇÃO

Neste capítulo será oferecida uma revisão da teoria de classificação estatística de imagens, tanto para o caso de classificação ponto a ponto, quanto para o caso de classificação de amostras. Será proposta também uma *distância* entre funções de autocorrelação bidimensional para utilização em esquemas de classificação de amostras.

Classificação Estatística de Imagens se refere ao conjunto de técnicas que visa particionar uma imagem em diversas classes de significado homogêneo, valendo-se de critérios e modelos estatísticos para o tratamento das grandezas envolvidas. A classificação pode ser feita ponto a ponto, quando a rotulação da imagem é feita um ponto por vez, ou por amostras quando a decisão se dá considerando vários pontos (normalmente conectados espacialmente). A rotulação é única e é efetuada a todos os pontos simultaneamente.

#### 4.2 - O CLASSIFICADOR PONTUAL DE MÁXIMA VEROSSIMILHANÇA

O classificador de máxima verossimilhança é um caso particular do classificador bayesiano quando os custos são zero e um (Mascarenhas e Velasco, 1989) e a distribuição dos atributos é gaussiana. Um ponto da imagem ou pixel, representado por um vetor de atributos  $\underline{x} = [x_1, \dots, x_n]$  é classificado como pertencente a uma certa classe  $H_i$  se :

$$g_i(\underline{x}) = -\frac{1}{2} \ln |\Sigma_i| - \frac{1}{2}(\underline{x} - \underline{\mu}_i)^T \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i) - \ln P(H_i) \geq g_j(\underline{x}) \quad (4.1)$$

onde:  $g_j(\underline{x})$  tem a mesma forma de  $g_i(\underline{x})$ ,  $\Sigma_i$  é a matriz de covariância espectral para a classe  $H_i$ ,  $\underline{\mu}_i$  é a média para a mesma classe  $H_i$  ( $i, j = 1, \dots, K$ ) e  $P(H_i)$  é a probabilidade a priori para essa classe.

O termo  $d_{T_i} = (\underline{x} - \underline{\mu}_i)^T \Sigma_i^{-1} (\underline{x} - \underline{\mu}_i)$  é a chamada distância de Mahalanobis entre  $\underline{x}$  e o centro da classe considerada.

Se as probabilidades a priori e os determinantes das matrizes de covariância são iguais então o ponto é classificado àquela classe cuja distância de Mahalanobis do ponto do centro da classe é a menor de todas.

Costuma-se também impor um limite máximo à distância de Mahalanobis (ou limite mínimo à função discriminante  $g(\underline{x})$ ) acima (abaixo) do qual o ponto é considerado não classificado.

No desenvolvimento desse classificador foi tomado como hipótese que não havia correlação espacial entre pixels adjacentes ou seja que toda a informação necessária para classificar um ponto estivesse contida em seus atributos espectrais, o que nem sempre é verdade (Foglein e Kittler, 1983). Tal classificador, porém, é muito usado por causa de sua simplicidade e relativa robustez.

Antes da classificação propriamente dita, é necessário realizar uma fase denominada de treinamento. Nessa fase são adquiridos os parâmetros (médias e matrizes de covariância) para definição do classificador a partir de áreas de classificação previamente conhecida.

#### 4.3 - CLASSIFICAÇÃO DE AMOSTRAS

Classificadores de amostras classificam todos os pontos pertencentes a uma região, ou amostra, a uma dentre várias classes conhecidas a priori. Os classificadores de amostras mais utilizados na literatura são os classificadores de máxima verossimilhança amostral e o classificador de amostras por mínima distância JM (Jeffreys-Matusita) (Souza et al., 1980). Digno de nota também é o método de classificação de amostras de texturas de Faugeras e Pratt (1980).

O método de classificação de amostras por máxima verossimilhança amostral considera que uma amostra ou região é formada por vetores aleatórios independentes e governada por distribuição gaussiana. As classes têm probabilidades a priori iguais. Nesse critério uma amostra  $A_i$  é associada à classe  $H_i$  se (Souza et al., 1980):

$$g_i(A_j) \geq g_j(A_j) \quad \text{para } j = 1, \dots, K \text{ classes} \quad (4.2)$$

onde

$$g_j(A_j) = - \sum_{\forall x_k \in A_j} \left[ (x_k - \underline{\mu}_j)^T \Sigma_j^{-1} (x_k - \underline{\mu}_j) + \ln |\Sigma_j| \right] \quad (4.3)$$

e  $\mu_j$  e  $\Sigma_j$  são as médias e matrizes de covariâncias para as classes  $j = 1, \dots, K$  conhecidas a priori.

Analisando a Expressão 4.3 observa-se que, a menos do termo  $\ln |\Sigma_j|$ , a classe escolhida é aquela que minimizar a somatória das distâncias de Mahalanobis dos pontos da amostra ao centro de cada classe.

Outro classificador de amostras é baseado no critério da distância estatística. *Distância estatística* é um índice utilizado para medir distância entre distribuições ou seja quanto maior a semelhança entre as distribuições (normalmente consideradas gaussianas) menor é a distância estatística. Da amostra desconhecida levantam-se os parâmetros da distribuição gaussiana que se supõe governa essa amostra. A amostra é associada à classe cuja distância estatística for menor. Para distribuições gaussianas a distância JM (de Jeffreys-Matusita) entre duas classes  $H_k$  e  $H_l$  (Souza et al., 1980) é dada por

$$J_{kl} = 2(1 - e^{-\alpha_{kl}}) \quad (4.4)$$

$$\alpha_{kl} = \frac{1}{8} (\underline{\mu}_k - \underline{\mu}_l)^T \Sigma^{-1} (\underline{\mu}_k - \underline{\mu}_l) - \frac{1}{2} \ln \left| \frac{\det \Sigma}{[\det \Sigma_k \times \det \Sigma_l]^{1/2}} \right| \quad (4.5)$$

onde  $\Sigma = \frac{1}{2} [\Sigma_k + \Sigma_l]$

Uma amostra  $A_i$  cujo rótulo é desconhecido, é classificado para uma classe  $H_i$ , se, levantados a média e matriz de covariância dessa amostra,

$$J_{ii} \leq J_{ik} \quad k = 1, \dots, K \quad (4.6)$$

Faugeras e Pratt (1980) classificam amostras de textura levantando a função de autocorrelação (FAC) de diversas subamostras dessa mesma textura. Das FAC's das subamostras são extraídos momentos que se supõe distribuídos gaussianamente no conjunto das subamostras. Além disso, um operador decorrelativo tal como um gradiente digital ou laplaciano digital é aplicado sobre a amostra. Atributos extraídos sobre os histogramas, calculados nas subamostras do campo decorrelacionado, são então usados conjuntamente com os momentos de FAC 2D para medir a separabilidade entre texturas utilizando-se da distância B (relacionada com a distância JM discutida anteriormente).

Um terceiro critério heurístico pode ser sugerido para classificar uma amostra, associando-a à classe que tenha maioria numérica dentro da amostra quando os pontos são classificados isoladamente.

#### 4.4 - PROPOSTA DE UM CLASSIFICADOR DE AMOSTRAS USANDO A FAC 2D

O método de Faugeras e Pratt (1980) discutido na seção passada não é prático para classificar amostras de textura pois necessita de um grande número de subamostras dentro da amostra para gerar uma quantidade suficiente de dados para uma estimativa satisfatória da média e matriz de covariância utilizadas no cálculo da distância B. Ocorre entretanto que a experiência mostra diferenças notáveis entre FAC's de texturas visualmente distintas. Para medir essa diferença e usar essa medida para classificar texturas, são sugeridas distâncias entre FAC's que não necessitam amostras de treinamento tão extensas quanto no caso de Faugeras e Pratt (1980).

O primeiro método (em duas variantes) mede a distância direta entre duas FAC's, que é definida como a integral da diferença quadrática entre as FAC's. O segundo método (também em duas variantes) baseia-se na extração de momentos da FAC (como foi feito por Faugeras e Pratt (1980)), tanto para a textura padrão como para a desconhecida. A textura desconhecida é classificada como sendo similar à textura cuja distância entre vetores de momentos for menor.

Primeiro método

DEFINIÇÃO 4.1 - *Distância entre funções de autocorrelação.* Sejam duas funções de autocorrelação  $\rho_1(k,l)$  e  $\rho_2(k,l)$  de duas séries bidimensionais estacionárias no sentido amplo, e definidas num intervalo de  $-K \leq -k \leq K$ ,  $-L \leq l \leq L$ . A distância entre essas duas funções de autocorrelação (DFAC) é dada pela fórmula:

$$DFAC(\rho_1, \rho_2) = \sum_{k=1}^K \sum_{l=-L}^L [\rho_1(k,l) - \rho_2(k,l)]^2 + \sum_{l=1}^L [\rho_1(0,l) - \rho_2(0,l)]^2 \quad (4.7)$$

OBSERVAÇÃO 4.1 - A distância é definida em meio plano não simétrico, pois a FAC 2D é simétrica em torno do ponto central.

OBSERVAÇÃO 4.2 - A distância definida em 4.7 não é sensível nem a escala nem à média. Ao substituir o coeficiente de autocorrelação pela autocovariância a medida passaria a ser sensível a escala. Com o uso da autocorrelação seria sensível também a média. A escolha de qual estatística de segunda ordem utilizar depende do uso necessário. Para classificar texturas padrões em ambientes com diferentes iluminações é melhor usar a formulação como foi proposta. Em casos de classificação de regiões em sensoriamento remoto, onde diferenças de escala e média interessam, é necessário utilizar-se da autocorrelação.

Uma versão alternativa para o cálculo de distância entre FAC's 2D é o uso da ponderação para atrasos "logs" maiores para compensar o efeito da maior variância dos estimadores para atrasos maiores.

DEFINIÇÃO 4.2 - *Distância Ponderada entre FAC's:* A distância ponderada entre as duas funções de autocorrelação (DPAC) é dada por

$$DPAC(\rho_1, \rho_2) = \sum_{k=1}^K \sum_{l=-L}^L \left[ \frac{\rho_1(k,l)}{p_1(k,l)} - \frac{\rho_2(k,l)}{p_2(k,l)} \right]^2 + \sum_{l=1}^L \left[ \frac{\rho_1(0,l)}{p_1(0,l)} - \frac{\rho_2(0,l)}{p_2(0,l)} \right]^2 \quad (4.8)$$

OBSERVAÇÃO 4.3 - Não existe na literatura formulação teórica para a variância do estimador da FAC 2D. Box e Jenkins (1970, pág. 34) apresentam uma expressão para a variância do estimador do FAC 1D quando o processo estocástico é do tipo AR(1). Fazendo uma extensão para o caso 2D e supondo separabilidade e igual coeficiente de correlação para as direções vertical e horizontal ter-se-ia a seguinte expressão para a variância do estimador  $\hat{\rho}(k,l)$  do coeficiente de autocorrelação

$$v(k,l) = \text{var} (\hat{\rho}(k,l)) = \frac{1}{N} \left[ \frac{(1 + \phi) (1 - \phi^{2(k+l)})}{1 - \phi} \right] - 2(k + l) \phi^{2(k+l)} \quad (4.9)$$

onde N é o número de amostras e  $\phi = \rho(1,0) = \rho(0,1)$  é o coeficiente de autocorrelação da imagem que é suposta igual para direção vertical e horizontal.

Os valores limites para a variância de  $\hat{\rho}$  são: para  $\hat{\rho}(0,0) = 1$ ,  $\text{var}(\hat{\rho}) = 0$ ; para

$$\rho(\infty,\infty) \Rightarrow \text{var}(\hat{\rho}) \rightarrow \frac{1}{N} \left[ \frac{1 + \phi^2}{1 - \phi^2} \right] \quad (4.10)$$

onde a variância se estabiliza rapidamente.

O peso a ser usado na Equação 4.8 é dado por

$$p_i(k,l) = \sqrt{N v_i(k,l)} \quad (4.11)$$

OBSERVAÇÃO 4.4 - O uso da ponderação no cálculo da distância pode, no entanto, violar um dos postulados de definição de distância pois, conforme os pesos usados, a distância pode se anular sem que  $\rho_1 = \rho_2$  para todo k,l.

#### Segundo método

Os momentos da função de autocorrelação são obtidos a partir da função geradora de momentos da FAC 2D (Jain, 1989):

$$S(u,v) = \sum_{k=0}^K \sum_{l=-L}^L (k - m_k)^u (l - m_l)^v \rho(k,l) \quad (4.12)$$

onde

$$m_k = \sum_{k=0}^K \sum_{l=-L}^L k \rho(k,l) \quad (4.13)$$

$$m_l = \sum_{k=0}^K \sum_{l=-L}^L l \rho(k,l) \quad (4.14)$$

e  $\rho(k,l)$  é a FAC de uma certa textura  $T$ . Os momentos mais utilizados (Jain, 1989) que definirão o vetor de atributos da FAC que descrevem uma determinada textura são os momentos  $S(0,2)$ ,  $S(2,0)$ ,  $S(1,1)$  e  $S(2,2)$ . Ocorre que esses momentos têm escalas muito diferentes. Para compensar o efeito da escala na distância, a distância log transformada é definida como:

$$\begin{aligned} DLM^2(T_1, T_2) = & (\log S_1(0,2) - \log S_2(0,2))^2 + \\ & (\log S_1(2,0) - \log S_2(2,0))^2 + \\ & (\log S_1(1,1) - \log S_2(1,1))^2 + \\ & (\log S_1(2,2) - \log S_2(2,2))^2 \end{aligned} \quad (4.15)$$

onde  $\log(S_i) = \text{sign}(S_i) \cdot \log |S_i|$  e

onde  $S_1(u,v)$  e  $S_2(u,v)$  são as funções geradoras para as texturas  $T_1$  e  $T_2$ .

Uma versão alternativa para essa distância é a distância log da diferença:

$$\begin{aligned} DLDM(T_1, T_2)^2 = & \log(S_1(0,2) - S_2(0,2))^2 + \\ & \log(S_1(2,0) - S_2(2,0))^2 + \\ & \log(S_1(1,1) - S_2(1,1))^2 + \\ & \log(S_1(2,2) - S_2(2,2))^2 \end{aligned} \quad (4.16)$$

Definidas as distâncias, é possível estabelecer então o seguinte critério para a classificação de amostras de textura:

Sejam uma amostra de textura A a ser classificada e um conjunto de texturas conhecidas  $T_i$  ( $i = 1 \dots K$ ) com suas FAC's também conhecidas.

A amostra A é classificada como sendo da classe de  $T_i$  se

$$D(A, T_i) \leq D(A, T_j) \quad i \neq j \quad i, j = 1 \dots K$$

onde K é o número de classes e D é uma das distâncias desenvolvidas DFAC, DPAC ou DLM, DLDM.

## CAPÍTULO 5

### EXTRAÇÃO DE ATRIBUTOS DE TEXTURA POR FILTRAGEM INVERSA

#### 5.1 - REVISÃO DA LITERATURA

Os métodos estatísticos de modelamento e discriminação de texturas encontrados na literatura (Jain, 1981; Jain, 1989; Haralick, 1979) podem ser subdivididos nas seguintes categorias:

- métodos baseados na função de autocorrelação que usam basicamente os momentos da FAC como atributos para discriminação; o método proposto na Seção 4.3 é um exemplo nessa categoria.
- métodos baseados em transformadas
- densidade de bordas
- atributos de histogramas
- métodos baseados em modelos estatísticos.

Dentre os métodos baseados em modelos estatísticos podemos destacar os que usam os modelos markovianos, os métodos de predição linear, também chamados de representação de predição de mínima variância (MVPR) e os métodos que utilizam versões unidimensionais e bidimensionais dos modelos ARMA.

O principal objetivo desse trabalho, como foi discutido na Introdução, é a utilização de modelos do tipo ARMA bidimensionais para caracterizar texturas aleatórias, comumente encontradas em imagens de sensoriamento remoto, e utilizar esses modelos para colaborar na discriminação de superfícies texturadas.

Modelos ARMA para processamento de imagens começaram a ser utilizados no início dos anos setenta inicialmente para síntese de imagens. MacCormick e Jayaramamurthy (1974) utilizaram modelos sazonais para modelar e posteriormente sintetizar texturas; como o modelo sazonal é multiplicativo esse tipo de modelo não se adequa a todos os tipos de textura estatística, onde em muitos casos o modelo 2D não é multiplicativo.

Tou (1981), Tou et al. (1986) utilizaram-se, com algumas limitações, de modelos bidimensionais para síntese de texturas (conf. Seção 3.2.4); usando suporte de QP e modelos bem curtos. Chen (1980) utiliza modelos ARMA-2D sobre suporte QP para redução de ruído. Heitz et al. (1987) utiliza modelos bidimensionais auto-regressivos e vetoriais para restauração de imagens.

Foglein e Kittler (1983) estudaram o efeito das correlações dos pixels na separabilidade entre classes. Os autores salientaram que a hipótese de independência espacial, quando falsa, leva a estimativas polarizadas dos parâmetros do classificador diminuindo assim a precisão do classificador. Usando modelos unidimensionais (para linha ou coluna), propostos por outros autores, chegaram a conclusão de que dependendo do filtro-modelo da imagem, o uso de dados decorrelacionados pelo filtro inverso melhora a precisão dos classificador (para modelos tipo passa-baixa, o uso de filtro-inverso do tipo passa-alta não ajuda).

Tubbs e Coberly (1978) ajustaram um modelo vetorial AR(1) aos dados Landsat e demonstraram que as estimativas comumente utilizadas para os parâmetros são polarizados. Sugerem um estimador não polarizado.

Therrien (1983) desenvolveu um método de classificação aplicando conceitos de estimação de parâmetros diretamente ao campo aleatório para calcular qual a melhor partição da imagem. O autor utilizou-se de modelos AR-2D, com média constante para toda a imagem, e derivou dois métodos de classificação, um sob critério de máxima verossimilhança (ML) e outro de máximo a posteriori (MAP). No primeiro critério chega-se a um método bem simples, onde o pixel é associado à classe cujo filtro inverso produz o menor resíduo quadrático normalizado.

Para o método MAP a solução é mais trabalhosa e recai num procedimento iterativo. Além disso, o autor demonstra que, mesmo que o modelo não seja AR e sim uma representação de predição de mínima variância, o procedimento utilizado em média classifica corretamente. Em Therrien et al. (1986) os autores apresentam uma expressão para o classificador de amostras por máxima verossimilhança baseado nos mesmos princípios descritos em

Therrien (1983). Aqui a amostra é associada à classe que obtiver menor somatória dos resíduos normalizados sobre a região a ser classificada.

Gambotto (1980) trabalhando independentemente apresenta um resumo de sua tese, em muitos pontos semelhante ao trabalho de Therrien, onde utiliza modelos auto-regressivos vetoriais e bidimensionais para síntese, classificação e segmentação de texturas. Apresenta também uma equivalência entre modelos MAR (AR vetorial) e AR-2D e um método de classificação de amostras baseado na mínima variância dos resíduos obtidos por filtros inversos casados às texturas (vale também para predição linear). Sugere, alternativamente, verificar as propriedades estatísticas dos resíduos obtidos filtrado-se as amostras pelos diversos filtros inversos. O filtro inverso que tornar textura da amostra um campo aleatório branco corresponde à classe correta. Esse critério, porém, não é unicamente definido, podendo haver casos em que dois filtros branqueiem uma determinada amostra dentro dos limites aceitos. O autor, assim como Therrien (1983) utiliza a hipótese de média constante para toda a imagem e usa apenas modelos AR.

Haralick (1979) sugere um esquema de classificação usando filtragem inversa que assinala ao pixel a classe que dê menor módulo do resíduo desde que esse resíduo seja menor que um certo limiar, caso contrário o ponto é borda.

O método de extração de atributos de textura por filtragem inversa pode ser considerado um aperfeiçoamento de métodos de extração de atributos espaciais (Dutra e Mascarenhas, 1984; Dondes e Rosenfeld, 1982; Andrade, 1989). Nestes métodos normalmente se utilizam filtros passa-baixos (FPB) em conjunção com filtros passa-altas (FPA). O FPB produz canal com redução da variabilidade dos atributos espectrais e o FPA produz canal que é *proporcional* à rugosidade local. A separação dessas duas informações provoca uma melhoria da precisão da classificação.

## 5.2 - COMENTÁRIOS SOBRE A BIBLIOGRAFIA E PROPOSTA DO MÉTODO.

A análise da literatura permite os seguintes comentários:

- 1) Ao se avaliar a evolução dos métodos de classificadores estatísticos, para uso cotidiano em sensoriamento remoto, observa-se que praticamente nada do que foi desenvolvido, foi utilizado operacionalmente além do classificador ML simples. Os métodos de extração de atributos espaciais se ressentem de uma estrutura teórica mais sólida. Os que se utilizam da teoria de séries temporais encontram sua dificuldade na área de identificação do modelo e na falta de algoritmos adequados para estimativa de modelos média-móvel 2D, além de dificuldades adicionais na medida da estabilidade do filtro modelo. Métodos de classificação que utilizam contexto normalmente conduzem a expressões complexas e muitas vezes intratáveis; mesmo com simplificações recaem em procedimentos iterativos que são dispendiosas e com dificuldades teóricas remanescentes.
- 2) Em geral, esses métodos aumentam a carga computacional consideravelmente, mas com o aumento da capacidade de processamento das estações de trabalho, e a utilização de processadores de sinal, a utilização de métodos baseados em texturas passa a ser factível, a custo acessível.
- 3) As principais dificuldades observadas nos métodos propostos por Tou (1981), Therrien (1983), Gambotto (1980) e outros, revistos na seção anterior, se concentram na fase de identificação, na fase da estimativa de parâmetros modelos MA (daí o uso quase que exclusivo de modelos AR em suporte de quarto de plano) e custo computacional elevado. Além disso, no ponto de vista de desempenho do classificador, observa-se que o esquema de classificação proposto por Therrien (1983), no caso um estimador de máxima verossimilhança para uma partição da imagem aplicado diretamente ao campo aleatório, conduz a uma expressão que não contempla a correlação existente entre as saídas dos filtros. O autor considera também, que a média é constante para toda a imagem, o que não acontece em geral.

- 4) Os algoritmos estatísticos de classificação (Capítulo 4) utilizam normalmente a hipótese de que os pontos de imagem são espacialmente independentes; imagens de sensoriamento remoto e mesmo imagens padrão, como as imagens de Brodatz (1956), no entanto, não seguem essa hipótese e são altamente correlacionadas espacialmente. Isso significa que as variações de pixel para pixel vizinhos não são altamente aleatórias, mas seguem um certo padrão de "determinação" que é justamente a textura. Se a correlação entre pixels fosse 1, a imagem seria uma função determinística (no caso um plano). Se a correlação fosse 1 em certos intervalos sazonais a imagem seria periódica com formato definido pelas condições iniciais.

Como foi observada na Introdução, pretende-se melhorar o desempenho dos classificadores estatísticos ponto a ponto e de amostra, utilizando-se de modelos estatísticos. Um dos usos dos classificadores de amostra supõe a utilização de segmentadores que dividem a imagem em regiões de textura uniforme, o que não será tratado aqui.

Considerando a literatura consultada e os comentários efetuados, decidiu-se então realizar um trabalho que desenvolva os seguintes aspectos:

- 1) Desenvolvimento e aplicação de um método sistemático para identificação de modelos ARMA-2D em geral a partir de séries concatenadas e considerações sobre o intervalo de confiança para os parâmetros (Seções 3.2.4 e 3.2.5).
- 2) Proposta de um esquema de classificação baseado na extração de atributos de textura por filtragem inversa por modelos casados. Os atributos são considerados conjuntamente para considerar a correlação entre filtros. Os pixels são classificados com base nos novos atributos usando classificação pontual (por máxima verossimilhança) e por amostra (Distância JM e método máxima verossimilhança discutidos na Seção 4.2). Não há hipóteses restritivas com relação à média da classe (a ser desenvolvido nas seções seguintes).

- 3) A classificação de amostras usando atributos gerados por filtragem inversa é comparada com a classificação de amostras usando os métodos clássicos aplicados diretamente sobre a imagem e com o método proposto na Seção 4.3, de distância entre funções de autocorrelações.
- 4) É verificada a viabilidade de se classificar as amostras testando o branqueamento das amostras pelos diversos filtros. Para isso foi utilizada a 'máquina' apresentadora de autocorrelação 2D definida e proposta na Seção 3.2.5. Conjuntamente é verificada a hipótese de classificação por mínima variância dos resíduos. A mesma máquina é utilizada como diagnóstico sobre resíduos sobre áreas de treinamento para verificar a adequabilidade do modelo.
- 5) Utilizando o método de identificação desenvolvido estuda-se o modelamento de imagens de radar de abertura sintética. Imagens de radar apresentam um filtro de ruído denominado "speckle" que se supõe multiplicativo. Estudos existentes sobre a estrutura de correlação espacial nesse tipo de imagens são cotejadas com modelos obtidos no campo homomórfico e normal (Seção 6.2.1).
- 6) Os métodos são também testados em imagens padrão de Brodatz (1956) e em uma imagem SPOT da região de São Paulo (Seções 6 e 7).

### 5.3 - A CLASSIFICAÇÃO PONTO A PONTO USANDO FILTROS INVERSOS

Ao contrário dos métodos clássicos onde se usa basicamente a informação espectral para a classificação ponto a ponto ou por amostras, nessa seção vai ser desenvolvido um método para utilização da informação de textura para ajudar no processo de classificação. A informação espectral nesse caso é menos importante.

Vamos supor que cada região da imagem  $I$ , além de pertencer a uma mesma classe  $H_c$ , possua um só tipo de textura e essa textura pode ser descrita por um modelo ARMA do tipo definido na Seção 3.2.1 (Equação 5.1). Cada pixel aqui será descrito por apenas um valor ou atributo: nível de cinza.

$$\sum_{(i,j) \in S_p} \phi_{i,j,H_c} \tilde{z}_{m-i, n-j, H_c} = \sum_{(i,j) \in S_q} \theta_{i,j,H_c} a_{m-i, n-j, H_c} \quad (5.1a)$$

$$\phi_{0,0,H_c} = -1 ; \quad \theta_{0,0,H_c} = -1$$

onde  $S_p$  e  $S_q$  são regiões de suporte para a parte AR e a parte MA do modelo respectivamente; conforme definido em 3.8 (supõe-se: computabilidade recursiva) e

$$\tilde{z}_{m,n,H_c} = z_{m,n,H_c} - \mu_{H_c} , \quad (5.1b)$$

$\mu_c$  é a média da classe  $H_c$  ,  $\tilde{z}_{m,n,H_c}$  é o nível de cinza centralizado para a classe  $H_c$  e  $\phi_{m,n,H_c}$  e  $\theta_{m,n,H_c}$  são os parâmetros do modelo da classe  $H_c$ .

Como as regiões de suporte permitem a computabilidade recursiva, é possível obter da imagem, dado o modelo e um conjunto conveniente de condições iniciais (c.i.), os "shocks" que a geraram a imagem.

Resolvendo 5.1 para  $a_{m,n}$  temos

$$a_{m,n,H_c} = \sum_{(i,j) \in S_p} \phi_{i,j,H_c} \tilde{z}_{m-i, n-j, H_c} + \sum_{(i,j) \in S'_q} \theta_{ij} a_{m-i, n-j, H_c} \quad (5.2a)$$

No caso típico os modelos verdadeiros das texturas não são conhecidos, sendo que foram estimados a partir dos dados, assim ao aplicar a Equação 5.1 sobre a imagem  $I$  e lembrando que  $\tilde{z}_{m,n,H_c} = z_{m,n,H_c} - \mu_c$  onde  $\mu_c$  a média da sinal na classe  $H_c$  , pode-se obter uma estimativa do "shocks" que deram origem a cada região da imagem.

$$\hat{a}_{m,n,H_c} = \sum_{(i,j) \in S_p} \hat{\phi}_{i,j,H_c} \tilde{z}_{m-i, n-j, H_c} + \sum_{(i,j) \in S'_q} \hat{\theta}_{ij,H_c} \hat{a}_{m-i, n-j, H_c} \quad (5.2b)$$

Essa estimativa para o "shock" é chamado o *resíduo* calculado pela filtragem inversa.

Therrien (1983) desenvolveu um método para classificação ponto a ponto onde o pixel (m,n) é assinalado à classe  $H_c$  tal que:

$$h_c = \frac{\hat{a}_{m,n,H_c}^2}{\sigma_c^2} + \ln \sigma_c^2 \quad (5.3)$$

é mínimo, onde  $\hat{a}_{m,n,H_c}$  é o resíduo calculado no ponto pelo filtro  $H_c$  e  $\sigma_c^2$  é a variância estimada para o resíduos.

Conforme já mencionado anteriormente tal regra de decisão não leva em conta correlação que existe entre as saídas dos filtros; ou seja não incorpora ao processo de decisão o estudo do comportamento do filtro quando este filtra texturas não correspondentes. A existência desta correlação entre filtros ficara mais evidente nos desenvolvimentos que se seguirão.

O método proposto então é o seguinte: Suponha-se que deseje classificar uma imagem em K classes, onde cada classe constitui-se de regiões de textura uniforme. O primeiro passo do processo é identificar e estimar os modelos de cada textura a partir de sub-regiões ou amostras conhecidas na imagem. Em seguida, num segundo passo, a imagem é filtrada por filtros inversos correspondentes a cada modelo (Figura 5.1).

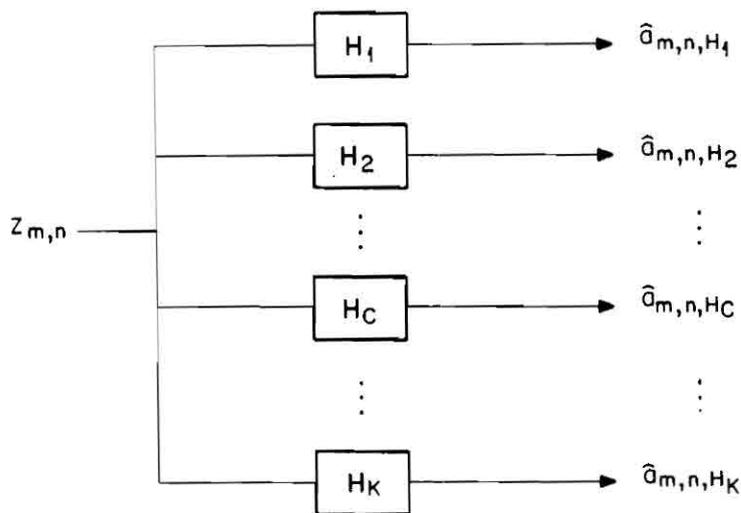


Fig. 5.1 - Extração de atributos de textura na filtragem inversa.

No procedimento de Therrien (1983) o ponto é classificado com base no menor resíduo normalizado, mas agora é necessário aprender como cada filtro casado a uma textura, "vê" as outras texturas.

Para isso cada ponto da imagem é associada a um vetor de atributos cujos componentes são os resíduos extraídos por filtragem inversa aplicando-se os K filtros à imagem toda. Esses atributos são gaussianos multivariados já que cada componente é gaussiana, além disso podem ser considerados independentes entre si, pois o resíduo é branco quando o filtro inverso é casado à textura, e os outros componentes, correspondentes aos filtros não casados, tem a correlação espacial muito diminuída devido ao caráter rejeita-faixa dos filtros inversos com maior reforço para as altas frequências. Como observa Faugeras e Pratt (1980), filtragem passa-alta tende a decorrelacionar o campo imagem.

Torna-se então conveniente a aplicação do classificador de máxima verossimilhança ponto a ponto, discutido na Seção 4. Aqui as condições são mais favoráveis, pois é atendida a suposição de pixels espacialmente independentes entre si, efetuada na dedução do classificador de máxima verossimilhança.

O *treinamento*, (ver Seções 4.2 e 7.1) que não existe no método de Therrien (1983), deve ser efetuado preferivelmente sobre as mesmas áreas onde foram obtidos os dados para confecção do modelo. Cada área de treinamento é filtrada por cada um dos filtros-modelo existentes, observando-se os vetores de atributos que informam como os filtros veem aquela área. O resultado do treinamento são vetores média e matriz de covariância dos efeitos de todos os filtros sobre a área de treinamento de determinada classe.

DISCUSSÃO 5.1 - *Qual a idéia básica do procedimento?* A idéia básica do método proposto é a seguinte: por força do processo de modelamento, a variância do resíduo sobre área filtrada com filtros casados é menor do que a variância dos resíduos obtidos por modelos não ajustados aquela área. Assim, os componentes dos vetores de atributos que correspondam às regiões usadas, tendem a dar valores absolutos baixos enquanto o restante dão maiores valores absolutos. Isso faz com que os vetores de atributos de regiões casadas

a um dos modelos treinados, tendam a agregar em torno dos outros vetores de um mesmo tipo de textura e a separarem de vetores de atributos sobre regiões de outra textura.

O mesmo não acontece por exemplo quando se usa extração de atributos espaciais de rugosidade, que tendem a dar valores altos para texturas díspares, que tenham rugosidade marcante e semelhante.

Esse princípio se aplica também a representação de predição de mínima variância (modelo MVPR), apenas que o branqueamento dos componentes fica de certa forma comprometida. A vantagem dessa representação é que poder-se-ia estipular regiões de suporte de tamanho fixo. Mínimos quadrados generalizados poderia ser utilizado para melhorar as estimativas de representação de predição de mínima variância pois diminuiria ainda mais a variância dos resíduos, além de um branqueamento maior da mesma. As estimativas generalizadas de mínimos quadrados, são, entretanto, mais difíceis de calcular e recaem normalmente em processos iterativos (Seção 3).

MÉTODO 5.1 - *Classificação de texturas usando resíduos*: O método proposto pode ser assim sintetizado: 1) para cada textura presente na imagem obtém-se um modelo ARMA, MVPR ou GLS. 2) Cada área de treinamento é filtrada para se obter resíduos referentes aos modelos usados. 3) A partir dos resíduos é feito o treinamento do classificador MAXVER. 4) A imagem inteira ou amostras são classificadas, usando-se os atributos extraídos por filtragem inversa e atribuindo o ponto à classe da menor distância de Mahalanobis ou o critério de classificação de amostras desejado.

OBSERVAÇÃO 5.1 - Outras representações de suporte não causais poderiam ser utilizadas, mas esses modelos não possuem a propriedade de inversibilidade, ou sejam não poderiam ser utilizadas para síntese, sendo portanto menos versáteis.

Para os modelos AR, MVPR e GLS, é possível derivar expressões teóricas para especificação do classificador de máxima verossimilhança sem precisar calcular os resíduos, bastando conhecer o modelo e a função de autocorrelação espacial.

Para isso é necessário redefinir as Equações 5.1b e 5.2b para representar as saídas dos filtros inversos de cada classe, quando filtram uma região de uma certa classe  $H_c$ , por exemplo.

Partindo da Equação 5.2b considere-se a v.a.  $e_{m,n,c,d}$  como sendo o resíduo do filtro inverso  $H_d$  quando filtra uma região de classe  $H_c$ , no ponto  $(m,n)$ :

$$e_{m,n,c,d} = \sum_{(i,j) \in S_p} \hat{\phi}_{i,j,d} \tilde{z}_{m-i, n-j,c,d} + \sum_{(i,j) \in S'_q} \hat{\theta}_{i,j,d} e_{m-i, n-j,c,d} \quad (5.4a)$$

onde

$$\tilde{z}_{m,n,c,d} = z_{m,n,c} - \mu_d \quad (5.4b)$$

e

$(\hat{\phi}_d, \hat{\theta}_d, \mu_d)$  são os parâmetros estimados para o filtro da classe  $H_d$  e  $z_{m,n,c}$  é um pixel que pertence à classe  $H_c$ .

Deseja-se classificar uma imagem (1 canal apenas) de maneira ponto a ponto, em  $K$  classes com texturas homogêneas. Supõe-se que os atributos das classes possuem distribuição gaussiana. Um procedimento de estimação e identificação realizado sobre dados extraídos de áreas de treinamento escolhidas para cada classe, determina um modelo estatístico para cada textura.

Da Equação 5.4a pode-se obter uma expressão para o filtro inverso auto-regressivo onde  $e_{m,n,c,d}$  representa o "shock" obtido no formato de filtragem inversa AR. Fazendo-se  $\theta_{i,j,d} = 0$  para  $(i,j)$  diferente de  $(0,0)$  temos

$$e_{m,n,c,d} = - \sum_{(i,j) \in S_p} \hat{\phi}_{i,j,d} (z_{m-i, n-j,c} - \mu_c) \quad (5.5a)$$

onde  $\phi_{0,0} = -1$

A Equação 5.4 pode ser representada pelo produto escalar

$$e_{c,d} = \underline{\phi}_d^T \tilde{z}_{c,d} = \underline{\phi}_d^T (z_c - \underline{\mu}_d) \quad (5.5b)$$

onde  $\tilde{z}_{c,d}$  é um conjunto convenientemente ordenado dos termos  $z_{m,n,c} - \mu_d$ ,  $\underline{\phi}_d$  é o vetor de coeficientes do filtro auto-regressivo para a classe  $H_d$ , e  $\underline{\mu}_d$  é o vetor de médias para a classe  $H_d$ .

Alternativamente 5.5b pode ser escrita como

$$e_{c,d} = \tilde{z}_{c,d}^T \underline{\phi}_d \quad (5.5c)$$

A imagem é filtrada por cada um dos filtros determinados e a cada ponto das áreas de treinamento é atribuído um vetor cujos componentes são os resíduos de cada filtro.

PROPOSIÇÃO 5.1 - (*Parâmetros do classificador de máxima verossimilhança*). Os parâmetros vetor média ( $\underline{m}_c$ ) e matriz de covariância ( $\Sigma_c$ ) de uma certa classe  $H_c$ , parâmetros de um classificador de máxima verossimilhança, quando se utilizam os atributos gerados por filtragem inversa são dados por:

$$\underline{m}_c = \begin{bmatrix} \underline{\phi}_1^T (\underline{\mu}_c - \underline{\mu}_1) \\ \vdots \\ \underline{\phi}_c^T (\underline{\mu}_c - \underline{\mu}_c) = 0 \\ \vdots \\ \underline{\phi}_k^T (\underline{\mu}_c - \underline{\mu}_k) \end{bmatrix} \quad (5.6a)$$

e também

$$\Sigma_c = \begin{bmatrix} \phi_1^T \\ \vdots \\ \phi_c^T \\ \vdots \\ \phi_k^T \end{bmatrix} R_c \begin{bmatrix} \phi_1 & \cdots & \phi_c & \cdots & \phi_k \end{bmatrix} \quad (5.11c)$$

onde  $R_c = E(\underline{z}_c \underline{z}_c^T) - \underline{\mu}_c \underline{\mu}_c^T$ , é a matriz de covariância espacial para a classe  $H_c$  e  $(\phi_i, \underline{\mu}_i)$  são os parâmetros do modelo AR, MVPR, ou GLS para a classe  $H_i$ .

DEMONSTRAÇÃO: - Para estabelecer as estatísticas necessárias ao classificador transformado é necessário então derivar o vetor média e a matriz de covariância de cada classe. O vetor  $\underline{m}_c$  média dos resíduos dos diversos filtros sobre uma área de treinamento  $T_c$  da classe  $H_c$  é dado por

$$\underline{m}_c = \begin{bmatrix} E(e_{c1}) \\ \vdots \\ E(e_{ci}) \\ \vdots \\ E(e_{ck}) \end{bmatrix} \quad (5.6b)$$

onde

$$E(e_{ci}) = E(\phi_i^T \bar{z}_{ci}) = E(\phi_i^T (z_c - \underline{\mu}_i)) = \phi_i^T E(z_c - \underline{\mu}_i) = \phi_i^T (\underline{\mu}_c - \underline{\mu}_i)$$

De (5.6.b) conclui-se que o vetor média dos resíduos sobre uma área de classe  $H_c$  vai se colocar no espaço, em posições proporcionais às diferenças entre médias das diversas classes.

A matriz de covariância para a classe  $H_c$  ou seja calculada sobre a área de treinamento  $T_c$  é dada por

$$\Sigma_c = [\text{cov}(\mathbf{e}_{ci} \mathbf{e}_{cj})]_{K \times K} \quad i, j = 1, \dots, K \quad (5.7)$$

Mas

$$\begin{aligned} \text{Cov}(\mathbf{e}_{ci} \mathbf{e}_{cj}) &= E\left[(\mathbf{e}_{ci} - E(\mathbf{e}_{ci})) (\mathbf{e}_{cj} - E(\mathbf{e}_{cj}))\right] = \\ &= E(\mathbf{e}_{ci} \mathbf{e}_{cj}) - 2 E(\mathbf{e}_{ci}) E(\mathbf{e}_{cj}) + E(\mathbf{e}_{ci}) E(\mathbf{e}_{cj}) = \\ &= E(\mathbf{e}_{ci} \mathbf{e}_{cj}) - E(\mathbf{e}_{ci}) E(\mathbf{e}_{cj}) \end{aligned} \quad (5.8)$$

Desenvolvendo o 1º termo de 5.8 e usando 5.5b e 5.5c

$$E(\mathbf{e}_{ci} \mathbf{e}_{cj}) = E\left[\phi_i^T \tilde{z}_{ci} \tilde{z}_{cj}^T \phi_j\right] = \phi_i^T E(\tilde{z}_{ci} \tilde{z}_{cj}^T) \phi_j \quad (5.9)$$

Mas

$$\begin{aligned} E\left[\tilde{z}_{ci} \tilde{z}_{cj}^T\right] &= E\left[(z_c - \underline{\mu}_i) (z_c^T - \underline{\mu}_j^T)\right] = \\ &= E(z_c z_c^T) - E(z_c) \underline{\mu}_j^T - \underline{\mu}_i E(z_c^T) + \underline{\mu}_i \underline{\mu}_j^T \\ &= E(z_c z_c^T) - \underline{\mu}_c \underline{\mu}_j^T - \underline{\mu}_i \underline{\mu}_c^T + \underline{\mu}_i \underline{\mu}_j^T \end{aligned} \quad (5.10)$$

Substituindo-se 5.10 em 5.9, depois substituindo em 5.8 usando 5.6b tem-se

$$\begin{aligned} \text{Cov}(\mathbf{e}_{ci} \mathbf{e}_{cj}) &= \phi_i^T E(z_c z_c^T) \phi_j - \phi_i^T \underline{\mu}_c \underline{\mu}_j^T \phi_j \\ &- \phi_i^T \underline{\mu}_i \underline{\mu}_c^T \phi_j + \phi_i^T \underline{\mu}_i \underline{\mu}_j^T \phi_j - \phi_i^T (\underline{\mu}_c - \underline{\mu}_i) \phi_j^T (\underline{\mu}_c - \underline{\mu}_j) \end{aligned}$$

usando o fato de que  $\phi^T z = z^T \phi$ , cancelam-se vários termos, assim =

$$\text{Cov}(\mathbf{e}_{ci} \mathbf{e}_{cj}) = \phi_i^T \left[ E(z_c z_c^T) - \underline{\mu}_c \underline{\mu}_c^T \right] \phi_j \quad (5.8a)$$

O termo  $E\left[ \underline{z}_c \underline{z}_c^T \right] - \underline{\mu}_c \underline{\mu}_c^T$  é a matriz de covariância espacial, que já foi calculada sobre a área de treinamento da classe  $H_c$  para o processo de modelagem, e é denotada por  $R_c$ , ou seja

$$R_c = E\left[ \underline{z}_c \underline{z}_c^T \right] - \underline{\mu}_c \underline{\mu}_c^T \quad (5.11a)$$

Assim, a matriz de covariância sobre os atributos de resíduos e' dada por (Equação 5.7):

$$\Sigma_c = \begin{bmatrix} \underline{\phi}_1^T R_c \underline{\phi}_1 & \underline{\phi}_1^T R_c \underline{\phi}_2 & \dots & \underline{\phi}_1^T R_c \underline{\phi}_k \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\phi}_k^T R_c \underline{\phi}_1 & \dots & \dots & \underline{\phi}_k^T R_c \underline{\phi}_k \end{bmatrix} = \quad (5.11b)$$

$$= \Sigma_c = \begin{bmatrix} \underline{\phi}_1^T \\ \vdots \\ \underline{\phi}_c^T \\ \vdots \\ \underline{\phi}_k^T \end{bmatrix} R_c \begin{bmatrix} \underline{\phi}_1 & \dots & \underline{\phi}_c & \dots & \underline{\phi}_k \end{bmatrix} \quad (5.11c)$$

OBSERVAÇÃO 5.2 - No caso de modelos do tipo MVPR ou GLS a região de suporte é fixa, e a matriz  $R_c$  tem a dimensão de  $M \times M$ , onde  $M$  é o número de pontos na região de suporte, por outro lado, como na modelagem AR a região não é fixa é necessário escolher uma região de suporte que englobe todas as estimadas e colocar coeficientes não estimados como zero.

OBSERVAÇÃO 5.3 - Não se consegue uma fórmula fechada para o caso ARMA em geral porque a relação entre os parâmetros do modelo e a FAC 2D não é

linear. O princípio da utilização dos resíduos da filtragem inversa, como atributos para classificação contínua, porém, o mesmo.

DISCUSSÃO 5.2 - *O método diminui a probabilidade de erro?* - O objetivo de se procurar novas metodologias para classificação de imagens é em geral a procura de métodos que diminuam a  $P(E)$ . Ocorre entretanto que, mesmo com hipóteses simplificadoras tais como distribuição gaussiana para as densidades envolvidas e modelos AR para as classes, o cálculo da  $P(E)$  não é imediato, principalmente para casos multivariados. Utilizam-se então critérios indiretos expressos por medidas de distância entre distribuições tais como a distância JM (visto no Capítulo 4) e a divergência (Mascarenhas e Velasco, 1989). Assim mesmo a utilização desses critérios para uma prova de cunho geral, tendo em vista o método aqui desenvolvido, não é factível.

Analizou-se entretanto um exemplo, com base na comparação da divergência das classes antes e depois da extração de atributos por filtros inversos, e verificou-se teoricamente um aumento nesse índice (e por conseguinte um aumento no poder de separabilidade das classes neste novo espaço), tanto maior quanto maior for a diferença entre as texturas.

EXEMPLO 5.1 - *A divergência entre 2 classes de textura utilizando-se de atributos obtidos por filtragem inversa.:*

Supondo-se que uma imagem é composta por duas texturas ( $H_1$  e  $H_2$ ) auto-regressivas de ordem 1 no sentido das linhas; deseja-se calcular a divergência entre as duas classes quando se utilizam atributos obtidos por filtragem inversa.

Os modelos para cada classe são:

$$\text{Classe 1} \quad \tilde{z}_{m,1,1} = \phi_{1,1} \tilde{z}_{m-1,1,1} + a_{m,1,1} \quad (5.12)$$

$$\text{Classe 2} \quad \tilde{z}_{m,2,2} = \phi_{1,2} \tilde{z}_{m-1,2,2} + a_{m,2,2} \quad (5.13)$$

onde  $\tilde{z}_{m,c,c} = z_{m,c} - \mu_c$ ,  $\mu_c$  é a média da classe  $H_c$ ,  $z_{m,c}$  é o valor do nível de cinza de um pixel localizado na  $m$ -ésima posição e que pertence à classe  $H_c$ ,

$\phi_{1,1}$  e  $\phi_{1,2}$  são os coeficientes do modelo auto-regressivo para as classes  $H_1$  e  $H_2$  (conforme Box e Jenkins, 1970,  $\phi_{1,1} = \rho_1$  e  $\phi_{1,2} = \rho_2$  onde  $\rho_1$  e  $\rho_2$  são as correlações horizontais em atrasos unitários para as classes  $H_1$  e  $H_2$ ). O coeficiente bidimensional (m,n) usado na definição da Equação 5.4 é simplificado aqui para índice m.

As Equações 5.12 e 5.13 podem ser colocadas no formato de filtragem inversa vetorial

$$a_{11} = \phi_1^T \tilde{z}_{11} \quad (5.14)$$

$$a_{22} = \phi_2^T \tilde{z}_{22} \quad (5.15)$$

onde

$$\phi_1 = \begin{bmatrix} 1 \\ -\rho_1 \end{bmatrix} \quad \tilde{z}_{cd} = \begin{bmatrix} \tilde{z}_{m,c,d} \\ \tilde{z}_{m-1,c,d} \end{bmatrix} = \begin{bmatrix} z_{m,c} \\ z_{m-1,c} \end{bmatrix} - \begin{bmatrix} \mu_d \\ \mu_d \end{bmatrix}$$

$$\phi_2 = \begin{bmatrix} 1 \\ -\rho_2 \end{bmatrix} \quad \text{ou } \tilde{z}_{cd} = z_c - \mu_d$$

Usando 5.6

$$\underline{m}_1 = \begin{bmatrix} E(a_{11}) \\ E(a_{12}) \end{bmatrix} = \begin{bmatrix} 0 \\ \phi_2^T (\underline{\mu}_1 - \underline{\mu}_2) \end{bmatrix} = \begin{bmatrix} 0 \\ \phi_2^T \underline{\delta} \end{bmatrix}$$

$$\underline{m}_2 = \begin{bmatrix} E(a_{21}) \\ E(a_{22}) \end{bmatrix} = \begin{bmatrix} \phi_1^T (\underline{\mu}_2 - \underline{\mu}_1) \\ 0 \end{bmatrix} = \begin{bmatrix} -\phi_1^T \underline{\delta} \\ 0 \end{bmatrix} \quad (5.16)$$

$$\text{onde } \underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2 = \begin{bmatrix} \delta \\ \delta \end{bmatrix}$$

De 5.11 temos a expressão para o cálculo das matrizes de covariâncias para as classes  $H_1$  e  $H_2$  calculados sobre os atributos (resíduos) gerados pelos filtros  $\phi_1$  e  $\phi_2$ .

$$\Sigma_1 = \begin{bmatrix} \phi_1^T R_1 \phi_1 & \phi_1^T R_1 \phi_2 \\ \phi_2^T R_1 \phi_1 & \phi_2^T R_1 \phi_2 \end{bmatrix} \quad (5.17)$$

$$\Sigma_2 = \begin{bmatrix} \phi_1^T R_2 \phi_1 & \phi_1^T R_2 \phi_2 \\ \phi_2^T R_2 \phi_1 & \phi_2^T R_2 \phi_2 \end{bmatrix} \quad (5.18)$$

onde

$$R_1 = E \left[ \begin{matrix} z_1 & z_1^T \end{matrix} \right] - \underline{\mu}_1 \underline{\mu}_1^T = \sigma_1^2 \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \quad (5.19)$$

$$R_2 = \sigma_2^2 \begin{bmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{bmatrix} \quad (5.20)$$

Sob hipótese de que  $\sigma_1 = \sigma_2 = \sigma$  (i. é, as texturas diferem apenas pela média e estrutura de correlação espacial) os termos das matrizes  $\Sigma_1$  e  $\Sigma_2$  são dadas por

$$\phi_1^T R_1 \phi_1 = \sigma^2 [1 \quad -\rho_1] \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -\rho_1 \end{bmatrix} = \sigma^2 [1 - \rho_1^2 \quad 0] \begin{bmatrix} 1 \\ -\rho_1 \end{bmatrix} = \sigma^2 (1 - \rho_1^2)$$

$$\phi_2^T R_1 \phi_2 = \sigma^2 [1 \quad -\rho_2] \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -\rho_2 \end{bmatrix} = \sigma^2 [1 - \rho_1 \rho_2 \quad \rho_1 \rho_2] \begin{bmatrix} 1 \\ -\rho_2 \end{bmatrix} =$$

$$\sigma^2 [1 - \rho_1 \rho_2 - \rho_1 \rho_2 + \rho_2^2] = \sigma^2 [1 - 2\rho_1 \rho_2 + \rho_2^2].$$

$$\phi_1^T R_1 \phi_2 = \phi_2^T R_1 \phi_1 = \sigma^2(1 - \rho_1^2)$$

Dai:

$$\Sigma_1 = \sigma^2 \begin{bmatrix} 1 - \rho_1^2 & 1 - \rho_1^2 \\ 1 - \rho_1^2 & 1 - 2\rho_1\rho_2 + \rho_2^2 \end{bmatrix} \quad (5.21)$$

Similarmente

$$\Sigma_2 = \sigma^2 \begin{bmatrix} 1 - 2\rho_1\rho_2 + \rho_1^2 & 1 - \rho_2^2 \\ 1 - \rho_2^2 & 1 - \rho_2^2 \end{bmatrix} \quad (5.22)$$

A divergência entre 2 classes com distribuição gaussiana é dada por (Mascarenhas e Velasco, 1989):

$$D_{ij} = \frac{1}{2} \text{tr} \left\{ \left( \Sigma_i - \Sigma_j \right) \left( \Sigma_j^{-1} - \Sigma_i^{-1} \right) \right\} + \frac{1}{2} \text{tr} \left\{ \left( \Sigma_i^{-1} + \Sigma_j^{-1} \right) \left( \underline{m}_i - \underline{m}_j \right) \left( \underline{m}_i - \underline{m}_j \right)^T \right\} \quad (5.24)$$

Usando-se os parâmetros calculados nas Equações 5.16, 5.21 e 5.22, pode-se calcular agora a divergência entre as duas classes do exemplo desenvolvido. O primeiro termo da divergência, que é devido à diferença entre as variâncias das classes, é dado por (Apêndice A):

$$\text{Divergência das Variâncias} = (\rho_1 - \rho_2) \left( \frac{\rho_1}{(1 - \rho_1^2)} - \frac{\rho_2}{(1 - \rho_2^2)} \right)$$

O segundo termo é devido à diferença entre as médias das classes no espaço transformado:

$$\text{Divergência das médias} = \frac{\delta^2}{2\sigma^2} \left( 2 + \frac{(1 - \rho_1)}{(1 + \rho_1)} + \frac{(1 - \rho_2)}{(1 + \rho_2)} \right)$$

A expressão completa da divergência compreende a soma dos dois termos acima, e pode-se salientar alguns valores limites desse índice nos seguintes casos:

1) Para  $\rho_1, \rho_2 \rightarrow 1 \Rightarrow D = \frac{\delta^2}{\sigma^2}$

2) Para  $\rho_1, \rho_2 \rightarrow 0 \Rightarrow D = \frac{2\delta^2}{\sigma^2}$

3) Para  $\rho_1, \rho_2 \rightarrow -1 \Rightarrow D = \infty$

Considerando a divergência dos canais originais ( $D = \frac{\delta^2}{\sigma^2}$ ) verifica-se que principalmente quando as correlações são negativas a divergência aumenta, e assim a P(E) diminui.

#### 5.4 - A CLASSIFICAÇÃO DE AMOSTRAS

O método de classificação de amostras proposto no Capítulo 4 pode ser aplicado diretamente aos campos filtrados conforme a metodologia desenvolvida nesse capítulo. A propriedade de branqueamento do filtros inversos torna a utilização desses classificadores mais conveniente, pois na dedução desses classificadores foi utilizada também a hipótese de independência espacial. Testes serão efetuados com imagens de sensoriamento remoto e os resultados comparados com a utilização desses classificadores sem filtragem inversa.

Uma outra maneira a ser testada, também como sugere Gambotto (1980), é verificar o branqueamento dos campos residuais. Para isso será aplicada a máquina sugerida no Capítulo 3 para apresentação da FAC 2D. Pode-se de antemão prever, no entanto, que esse critério pode não dar resposta única pois eventualmente pode ocorrer que determinada amostra pode ser considerada branqueada por mais de um filtro ou então não ser branqueado por nenhum filtro (nesse caso a amostra é rejeitada).

Finalmente, os métodos mencionados são também comparados com o método da classificação de amostras de texturas pela medida da distância entre FAC's 2D e distância entre momentos de FAC 2D sugeridos no Capítulo 4.

## CAPÍTULO 6

### AQUISIÇÃO E IDENTIFICAÇÃO: RESULTADOS

#### 6.1 - METODOLOGIA

A sistemática adotada para adquirir os dados das classes e identificar os modelos, conforme os princípios expostos na Seção 3.2.5, segue os seguintes passos:

- a) Adquirir os dados referentes às áreas de treinamento, concatenando-se segmentos de linha ou coluna. Segmentos de tamanho 20 a 30 são o tamanho ideal, mas restrições práticas podem reduzir o tamanho necessário.
- b) Retirar a média do sinal e plotar as funções FAC e FAP. Pela observação dessas funções determinar  $p$  e  $q$  tal que elas se anulem (ou sejam consideradas dentro do intervalo de confiança).
- c) Obter as estimativas preliminares para  $AR(p)$ ,  $MA(q)$  e  $ARMA(p,q)$ . Metodicamente eliminar as ordens de modelos que resultarem não causais ou não invertíveis.
- d) De posse do erro padrão para os coeficientes preliminares, zerar os não significativos (esses parâmetros não serão mais considerados para otimização).
- e) Submeter os três modelos à estimativa de máxima verossimilhança. Escolher o de menor índice de AKAIKE (AIC). Observar que a natureza originalmente bi-dimensional dos dados vai revelar rapidamente a região de suporte do filtro. Pode ser necessário reter mais de um filtro quando as diferenças de AIC não forem significativas.
- f) Fazer o teste diagnóstico do modelo para verificar se o resíduo é branco. Caso contrário testar os outros modelos ou alterá-los adicionando ou retirando parâmetros com valor próximo a 2 vezes o desvio padrão do estimador.

## 6.2 - TESTES PRÁTICOS E DISCUSSÕES

Para avaliar o desempenho do método estabelecido, foram efetuados testes práticos em três tipos de imagens: imagens padrões de texturas digitalizadas do livro de Brodatz (1956), imagem de radar de abertura sintética (imagens SAR) e imagens obtidas a partir de sensores óticos de satélites de recursos naturais.

Nesse capítulo apresentam-se os resultados da identificação dos modelos das texturas existentes nessas imagens, e no capítulo seguinte (Capítulo 7) são apresentados os resultados da aplicação desses modelos em problemas de classificação de padrões.

Foi utilizado o programa PEST de identificação e estimação de parâmetros de modelos ARMA para séries univariadas. Este programa acompanha o livro de Brockwell e Davis (1987). Foram obtidas áreas de treinamento, para cada classe, com 1000 pontos resultantes da concatenação de 100 linhas de 10 pontos cada. Para cada classe definida foram levantados os modelos AR(21), MA(21) e ARMA(2,10), usando a metodologia apresentada. Os modelos de ordem 21 compreendem duas linhas abaixo do ponto de referência e o de ordem (2,10) uma linha apenas.

### 6.2.1 - TEXTURAS DE BRODATZ

Foi construída uma imagem mosaico de três texturas, digitalizando-as do livro de Brodatz (1956), a saber: grama, tecido e madeira (Figura 6.1). Conforme a metodologia exposta na Seção 6.1 e com as considerações expostas na Seção 6.2 foram determinados os modelos das classes em estudo.

Para a classe grama foi possível determinar apenas o modelo AR-2D a partir de um modelo AR(21) concatenado e um modelo ARMA-2D a partir de um modelo ARMA(11,11) concatenado. Todas as outras opções média móvel resultaram não-causais.

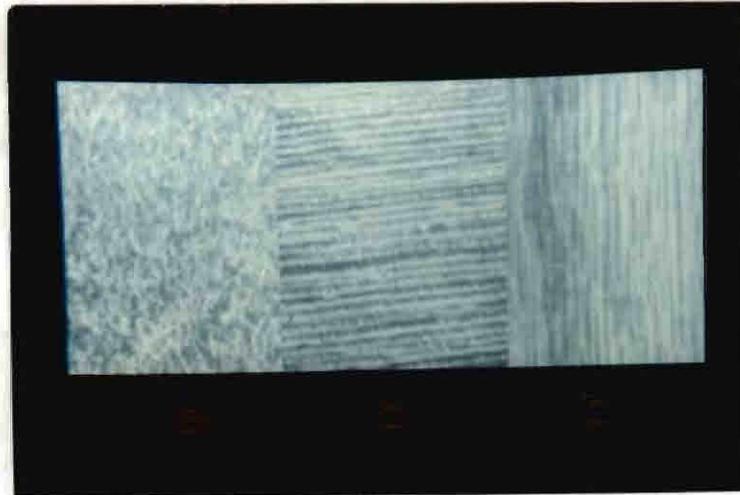


Fig. 6.1 - Mosaico de texturas de Brodatz( 1956): (a) grama, (b) tecido (c) madeira.

A Figura 6.2 apresenta os modelos da classe grama e respectivos índices de AKAIKE(AIC). Nesta Figura 6.2 estão representados para cada modelo, os coeficientes do filtro e sua distribuição espacial relativa. O símbolo  $\square$  representa a origem da região de suporte, ou seja, a coordenada (0,0). Significa também que o ponto nessa coordenada relativa é calculado em função dos pesos (coeficientes) nas regiões de suporte AR e MA. A Figura 6.2 pode ser considerada também a representação visual da Equação 6.1 que é obtida a partir da Equação 3.8. O modelo nessa forma pode ser utilizado para síntese, ou seja, utilizando-se o filtro-modelo sobre um campo aleatório branco. Sob certas condições iniciais, gera-se a textura, com a aparência visual da sua imagem de origem. Ao ponto processado pelo modelo deve ainda ser adicionada a média da região.

$$z_{ij} = \sum_{(i,j) \in S_q} \hat{\theta}_{ij} a_{m-i, n-j} - \sum_{(i,j) \in S'_p} \hat{\phi}_{ij} z_{m-i, n-j} \quad (6.1)$$

AR			MA		
	□	.612		1	0.0
.227	-.859	-.465	0.0	0.0	0.0
-.160	-.225	.121			

(a)

AR		MA	
□	.515		1
.645	-.274	.229	.415

(b)

Fig. 6.2 - Modelos ARMA para a classe grama (a) modelo AR, parte MA é nula, exceto  $\theta_{0,0}$  - AIC = 4463, (b) modelo ARMA AIC = 4472.

Nota-se entretanto não haver diferenças significativas entre os índices de AKAIKE dos modelos para a grama. Esse fenômeno foi observado também para as outras classes em estudo ou seja, entre os modelos que resultaram causais e invertíveis, poucas eram as diferenças entre os índices AIC. Uma verificação sobre a variância dos resíduos quando se usa um ou outro modelo também não mostrou grandes diferenças significando igual adequabilidade de cada modelo. Ainda assim optou-se pelo modelo de menor AIC.

Os modelos para serem utilizados em classificações de padrões, promovendo branqueamento, devem ser utilizados em sua forma denominada filtro inverso, pois o que se deseja é obter uma estimativa do campo aleatório branco (resíduos) que deu origem à imagem.

Da Equação 3.8 e considerando que  $\theta_{00} = -1$  obtém-se a Equação 6.2.

$$\hat{a}_{mn} = - \sum_{(i,j)} \sum_{\in S'_q} \hat{\theta}_{i,j} a_{m-i, n-j} + \sum_{(i,j)} \sum_{\in S_p} \hat{\phi}_{i,j} z_{m-i, n-j} \quad (6.2)$$

Dáí se pode observar que no formato filtro inverso o modelo tem seus coeficientes com sinal trocado. A Figura 6.3 apresenta os modelos escolhidos para as três classes (segundo o critério de menor AIC) em sua forma gráfica de filtro inverso.

O símbolo □ significa nesse caso que o resíduo é calculado nessa posição relativa. Deve-se observar também que se usou o suporte de meio plano superior para haver uma relação direta com os programas de filtragem que procuram normalmente de cima para baixo.

AR			MA		
-.121	+.225	+.160			
+.465	-.859	-.227		0.0	
-.612	1		0.0	□	0.0

(a)

AR		MA	
.535	- .675	0.0	-.221
-.817	1	0.0	□

(b)

AR			MA		
	0.0		-.340	-.610	-.181
0.0	0.0	0.0	-.574	-.954	-.305
0.0	1	0.0	-.452	□	

(c)

Fig. 6.3 - Filtros inversos das texturas: (a) grama - modelo do tipo AR, (b) tecido - ARMA, (c) madeira - MA.

Os modelos para as três classes do experimento - grama, tecido e madeira - exemplificam um caso em que diferentes tipos de modelos são os mais convenientes para representar texturas. A utilização exclusiva de modelos AR-2D, como é feita na maioria dos casos citados na literatura, não possibilita a representação adequada de certas texturas, ou mesmo impossibilita a representação por este tipo de modelo linear. Não existe, por exemplo, nas tentativas de identificação, um modelo AR-2D causal para a classe madeira.

#### 6.2.2 - RADAR DE ABERTURA SINTÉTICA (SAR580)

Testes também foram efetuados em uma imagem de radar de abertura sintética da missão SAR-580 (Figura 6.4) sobre uma área da Alemanha, sensor cujo uso se torna cada vez mais intenso devido à faculdade de se obter imagens em quaisquer condições de tempo. O radar apresenta entretanto desafios de processamento, pois apresenta o fenômeno da presença de um ruído abrangente denominado "speckle" que se supõe multiplicativo, e que mascara a textura natural do terreno.

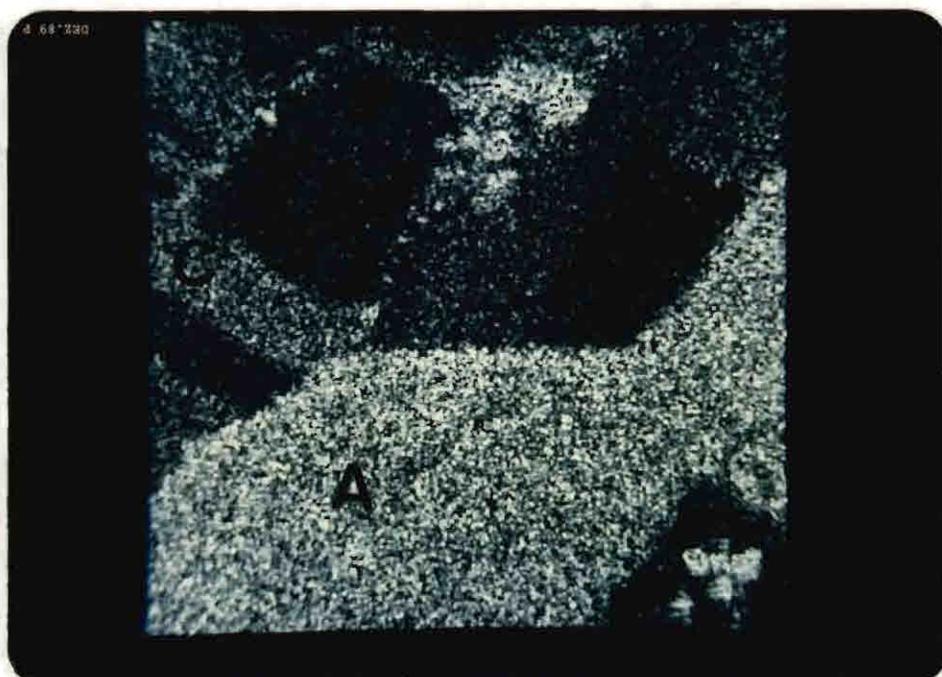


Fig. 6.4 - Imagem SAR580 (a) floresta, (b) cultura anual, (c) restos de cultura.

Li e Banon (1988), mencionam diversos tipos de filtros que procuram diminuir esse ruído sem afetar grandemente a textura natural. No presente trabalho, no entanto, não foi realizada nenhuma filtragem na imagem antes do processo de modelagem, com objetivo de avaliar o impacto do "speckle" sobre a determinação do modelo de textura em radar.

As áreas de treinamento foram obtidas concatenando-se pequenos segmentos de linha. O autor, em trabalho anterior (Dutra, 1988), discute os casos de concatenação de colunas e segmentos puramente lineares.

Os modelos foram levantados, considerando-se também o logarítmo dos dados, operação essa sugerida pelo caráter multiplicativo do ruído "speckle". A operação de logaritmo é muito utilizada em teoria de séries temporais para tornar a série estacionária ou obter modelos mais parcimoniosos, por melhor adequação de hipóteses com relação ao caráter do ruído. No campo homomórfico o ruído quando originalmente multiplicativo torna-se aditivo, compatível com as hipóteses dos modelos ARMA.

A Figura 6.5 apresenta a configuração dos modelos levantados para a classe Floresta. Observa-se que os coeficientes significativos se agrupam em torno do ponto de referência e ocupam o espaço de suporte de meio plano não simétrico, como explicado no Capítulo 3. Testes diagnósticos efetuados sobre os resíduos obtidos pelos filtros inversos correspondentes a esses modelos, usando-se ainda o programa PEST, demonstraram que estão bem adequados, pois só apresentam correlações significativas no atraso 0 (sobre a série concatenada).

Observa-se que o modelo ARMA(2,10) (na notação unidimensional), equivalente ao modelo ARMA( $S_p(0,1,1), S_q(0,2,0)$ ), é o modelo mais parcimonioso, pois possui menor AIC que os outros modelos (Dutra, 1988), e apenas dois coeficientes a considerar. Deve-se notar, no entanto que as diferenças entre os AIC's foram muito pequenas e não significativas, o que sugere a possibilidade de uso de quaisquer dos modelos. Ao contrário da teoria, é sempre possível na prática ajustar mais de um modelo, a um certo conjunto de dados. Isto ocorre porque, o conjunto de dados sendo finito, também o é a FAC e a FAP. Teste diagnóstico sobre o resíduos obtidos pelos diferentes modelos,

confirmam a possibilidade de todos poderem representar adequadamente o campo aleatório.

	□	.250			1	.124
.055	.604	-.111	.05	.079	.450	
	-.222				-.142	
	(a)				(b)	
	□	.268			1	.127
.076	.667	.200		.083	.456	.093
	.120					
	(c)				(d)	
	□	.265 (AR)			1	.130 (AR)
	.693 (MA)				.449 (MA)	
	(e)				(f)	

Fig. 6.5 -Configuração de coeficientes para a classe Floresta, modelos (a) AR(21), (b) AR(21) LOG, (c) MA(21), (d) MA(21) LOG, (e) ARMA(2,10), (f) ARMA(2,10) LOG.

A Figura 6.6 apresenta os modelos AR(21), nos campos *log* e *normal* para as classes Cultura Anual e Restos de Cultura. Observando-se esses modelos e os modelos da Figura 6.5, verifica-se que aqueles identificados no campo homomórfico parecem indicar menor correlação global pois as regiões de suporte são menores, assim como os próprios coeficientes, indicando assim maior independência do ruído nesse campo. Pelo princípio da parcimônia esses modelos são mais apropriados, confirmando de certa forma a hipótese de ruído multiplicativo. Outra conclusão possível, é que o "speckle" fortemente influencia a determinação do modelo, não havendo muita diferença entre eles.

	□	.500	-.145		1	.277
.05		.683	-.320		.075	.429
		-.290	.121			-.119
	(a)				(b)	
	□	.469	-.180	.063	1	.288
.051		.622	-.337	.102	.098	.380
		-.252	.098			-.115
	(c)				(d)	

Fig. 6.6 - Modelos Autorregressivos para as classes Cultura Anual nos campos Normal (a) e Log (b) e Restos de Cultura nos campos Normal(c) e Log(d).

### 6.2.3 - IMAGEM SPOT - CIDADE DE SÃO PAULO

Para testar o método em classificação de imagens óticas foi utilizada uma imagem SPOT (Figura 6.7) sobre a cidade de São Paulo com duas classes: cidade e montanha. Ao aplicar o método apresentado na Seção 6.1 foi observada uma característica que parece comum às imagens de sensoriamento em geral, que é a alta correlação espacial dessas imagens. Este fato foi observado em várias imagens que foram avaliadas anteriormente, com diversos tipos de classes, tais como classes agrícolas, urbanas e outras, quando se notou que a alta correlação espacial dificultava a estimação de modelos, pois os polos da função transferência dos modelos se posicionavam perto do círculo unitário. Além disso, em geral, as texturas não eram extensas e variavam bastante de região para região.

Escolheu-se a presente imagem por possuir texturas uniformes relativamente extensas. A alta correlação espacial se traduz em uma FAC sobre a série concatenada que cai muito lentamente sugerindo uma diferenciação sobre a série concatenada para torná-la estacionária. Ocorre que uma diferenciação sobre a uma série equivaleria a uma diferenciação no sentido da concatenação (no caso do presente experimento na direção das linhas) sobre a imagem, tornando o resultado não simétrico. Não seria possível também reconstituir a imagem a partir da série diferenciada visto que a média não é

estacionária sobre a imagem e depende da classe da região. Assim, optou-se por não realizar a diferenciação.

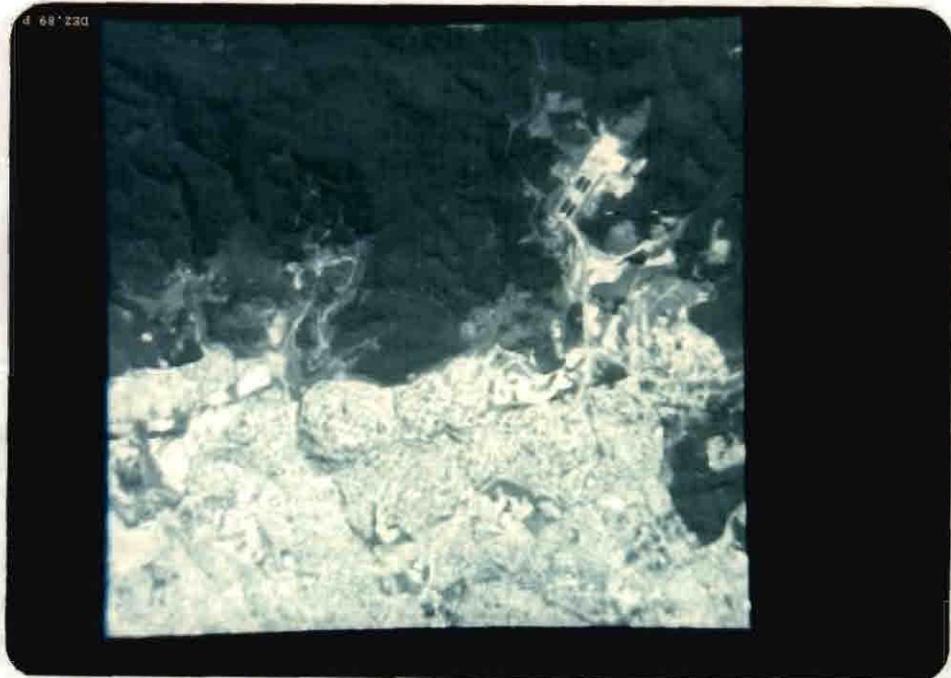


Fig. 6.7 - Imagem SPOT - São Paulo mostrando as classes cidade e montanha

Os modelos obtidos para as classes no formato de filtros inversos são dados na Figura 6.8. Outros tipos de modelos resultaram não causais ou não-invertíveis principalmente por causa da não-estacionariedade (ACF cai devagar) das classes da imagem.

Nota-se uma razoável diversidade nesses modelos o que se refletiu em uma grande diversidade da variância das saídas desses filtros inversos quando filtram a classe correta e a área da classe não correta (Tabela 6.1).

TABELA 6.1

VARIÂNCIA DE RESÍDUOS SPOT

ÁREA TESTE	FILTROS			
	ORIGINAL	TEÓRICO	MONTANHA	CIDADE
Montanha	2,20	1,11	1,02	2,19
Cidade	90,96	24,30	32,95	20,90

AR				MA		
	.177		-.253			
			-.213		0	
-.158	.232	-.616	1	0	□	0

(a)

AR				MA		
		-.426	-.109	-.373		-.479
+.093	-.142	0	1	-.682		□

(b)

Fig. 6.8 - Filtrros inversos para as classes (a) Floresta e (b) Cidade.

## CAPÍTULO 7

### CLASSIFICAÇÃO - RESULTADOS

#### 7.1 - INTRODUÇÃO

Uma vez determinado o modelo referente às classes de texturas presentes na imagem, o processo de classificação se desenvolve em três passos conforme a Seção 5.3.

- a) *Extração dos atributos* - fase em que o canal (ou canais) original é filtrado por cada um dos filtros inversos correspondentes a cada modelo de textura.
- b) *Treinamento* - obtenção das matrizes de covariância e média das classes sobre os canais filtrados. A matriz de covariância e média quando o filtro é AR podem ser calculadas teoricamente como indicado na Seção 5.3. De qualquer forma é possível filtrar as áreas de treinamento e dessas áreas obter os parâmetros necessários para o treinamento do classificador. Normalmente se utiliza a mesma área de treinamento para modelagem e para o classificador.
- c) *Classificação* - utilizam-se os processos desenvolvidos no Capítulo 4, ou seja, pode-se utilizar os classificadores ponto a ponto e o classificadores de amostra.

No Capítulo 4 foi proposto um método de classificação de amostras por distância entre FAC's. Esse método também vai ser cotejado com os métodos já desenvolvidos, sendo que os resultados são apresentados em forma de matriz de classificação.

Deve ser observado que poderão ocorrer problemas nas fronteiras devido ao tamanho não unitário da região de suporte do filtro. Esse problema, porém, é minimizado pelo método pois permite a consecução de modelos parcimoniosos, ou seja, com pequenas regiões de suporte.

## 7.2 - RESULTADOS

São apresentados aqui os resultados da classificação de texturas para os diversos tipos de imagens testadas.

Para o experimento da imagem de Brodatz (1956) (Seção 6.2.1) é apresentado um conjunto completo de testes que inclui a classificação de amostras de texturas por distâncias da FAC. Para o experimento da imagem de radar (Seção 6.2.2) são apresentados os resultados da classificação ponto a ponto e de amostras sobre os atributos residuais, e exemplos de aplicação da máquina de apresentação de FAC. Finalmente, na Seção 7.2.3, apresentam-se os resultados da classificação ponto a ponto e de amostras de 2 classes sobre a imagem SPOT em São Paulo.

Como afirmado na Seção 7.1, os parâmetros de classificação por método de máxima verossimilhança quando se usam modelos AR, MVPR ou GLS podem ser calculados teoricamente. Porém, como foi necessário filtrar todos os canais para posterior classificação e considerando que outros tipos de modelos não permitem cálculo teórico, optou-se por treinar o classificador diretamente sobre os resíduos. Testes efetuados com máscaras auto-regressivas curtas, mostravam no entanto que as matrizes calculadas teoricamente e as calculadas diretamente sobre os resíduos são muito semelhantes.

### 7.2.1 - TEXTURAS DE BRODATZ

O canal textura original foi filtrado pelos filtros inversos da Figura 6.3 dando origem a três canais. A Figura 7.1 apresenta o aspecto do canal original (ver Figura 6.1) quando filtrado pelo filtro inverso da classe grama.

Foi coletada uma amostra para treinamento por classe (áreas maiores mostradas na Figura 7.2) para o canal original e para o conjunto de três canais filtrados. Essas áreas de treinamento não coincidem com a área de extração do modelo.



Fig. 7.1 - Imagem filtrada pelo filtro inverso da grama.

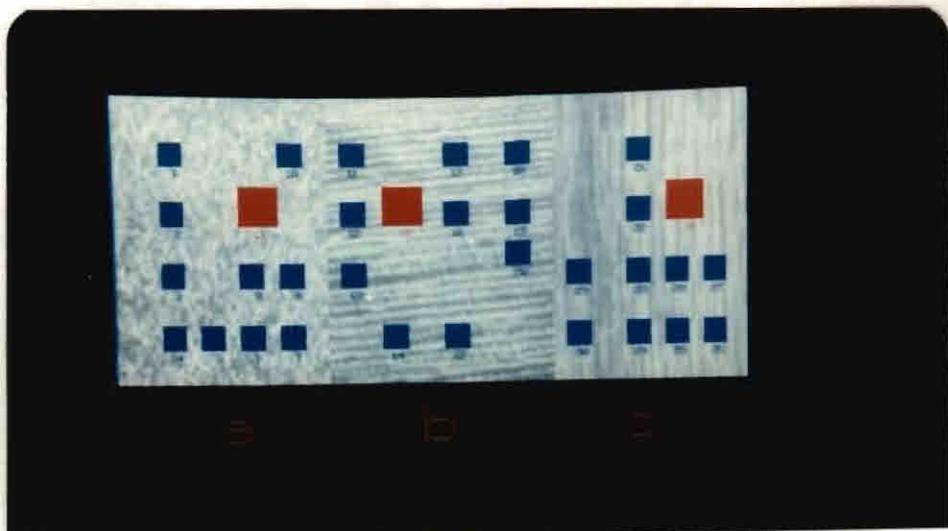


Fig. 7.2 - Amostra de treinamento (áreas maiores) e teste, para as classes grama, tecido e madeira.

A Tabela 7.1 apresenta a matriz de classificação para as áreas de treinamento quando se utilizam os canais filtrados. Matriz de classificação é uma forma de verificação da qualidade de definição do classificador, pois apresenta em forma de tabela como os pontos de áreas de classificação conhecidas a priori são classificados para as diversas classes definidas. Quando um classificador é bem definido espera-se que todos os pontos de uma área de classificação conhecida sejam realmente classificados para aquela classe. Pode-se então definir sobre os dados da matriz de classificação os seguintes índices: Desempenho médio ( $D_m$ ) que informa a porcentagem de pontos classificados corretamente sobre o número total de pixels das amostras; Abstenção média ( $A_m$ ) que significa a porcentagem do número de pontos não classificados (todos os pontos que superarem os limiares para as funções discriminantes de todas as classes; porcentagem por área de teste marcada na coluna N da matriz de classificação) sobre o número de pontos total. Finalmente a Confusão média ( $C_m$ ) que é a porcentagem de pontos classificados erroneamente. Trata-se de uma estimativa para a  $P(E)$  do classificador (embora nesse caso um estimador otimista) definida sobre a mesma área de treinamento do classificador.

A Tabela 7.2 apresenta os resultados da classificação para as mesmas áreas de treinamento definidos sobre o canal original, e podem ser comparados com os resultados da Tabela 7.1. Observa-se uma sensível melhoria nos índices pois quando se usa os canais filtrados, o  $D_m$  é maior e a  $C_m$  é menor.

Deve-se notar que os resultados só não são melhores pois as médias das diversas classes de textura estão muito próximas entre si, o que prejudicaria a separabilidade entre as classes.

A Figura 7.2 apresenta também a localização das 10 amostras (retângulos menores) por classe definidas para avaliar o esquema de classificação de regiões quando se usam os canais filtrados.

TABELA 7.1

MATRIZ DE CLASSIFICAÇÃO PARA OS CANAIS FILTRADOS(%)

ÁREA DE TREINAMENTO	N	CLASSES		
		GRAMA	TECIDO	MADEIRA
Gramma	3,2	29,1	23,7	43,9
Tecido	4,4	15,8	48,9	30,9
Madeira	0,1	11,6	11,1	77,1

$D_m = 51,7$

$A_m = 2,6$

$C_m = 45,7$

TABELA 7.2

MATRIZ DE CLASSIFICAÇÃO PARA O CANAL ORIGINAL(%)

ÁREA DE TREINAMENTO	N	CLASSES		
		GRAMA	TECIDO	MADEIRA
Gramma	0,0	18,3	26,5	55,3
Tecido	3,3	22,7	34,3	39,7
Madeira	0,2	22,9	12,1	64,8

$D_m = 39,1$

$A_m = 1,2$

$C_m = 59,7$

Os dados das classificações de regiões pelos critérios de mínima distância JM, mínima distância de Mahalanobis média e maioria de pontos classificados ponto a ponto, são mostrados nas Tabelas de 7.3 a 7.5 para os canais filtrados. Não há rejeição nesta classificação.

TABELA 7.3

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CANAIS RESIDUAIS - CRITÉRIO DISTÂNCIA JM

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Gramma	80	20	0
Tecido	0	100	0
Madeira	0	0	100

$$D_m = 93.3$$

$$C_m = 6.7$$

TABELA 7.4

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CANAIS RESIDUAIS - CRITÉRIO DISTÂNCIA MAHALANOBIS MÉDIA

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Gramma	80	20	0
Tecido	0	100	0
Madeira	20	0	80

$$D_m = 86.7$$

$$C_m = 13.3$$

TABELA 7.5

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CANAIS RESIDUAIS - CRITÉRIO MAIORIA

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Gramma	10	40	50
Tecido	0	90	10
Madeira	0	0	100

$$D_m = 66.7$$

$$C_m = 33.3$$

Observando-se as tabelas verifica-se o melhor aproveitamento para o critério de menor distância JM. Nas Tabelas 7.6, 7.7 e 7.8 são apresentadas as matrizes de classificação para os mesmos critérios antes mencionados, mas utilizando-se apenas do canal original.

TABELA 7.6

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CANAL ORIGINAL - CRITÉRIO DISTÂNCIA JM

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Gramma	20	60	20
Tecido	0	100	0
Madeira	0	60	40

$$D_m = 53.3$$

$$C_m = 46.7$$

TABELA 7.7

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CANAL ORIGINAL - CRITÉRIO: DIST. MALAHANOBIS MÉDIA

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Gramma	30	60	10
Tecido	0	100	0
Madeira	0	70	30

$$D_m = 53.3$$

$$C_m = 46.7$$

TABELA 7.8

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CANAL ORIGINAL - CRITÉRIO MAIORIA

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Gramma	0	60	40
Tecido	0	80	20
Madeira	0	10	90

$$D_m = 56.7$$

$$C_m = 43.3$$

Como se pode observar, houve real melhoria dos índices de classificação correta e diminuição da confusão média.

Para avaliar o desempenho das distâncias definidas para a FAC e para os momentos da FAC, foram estabelecidas as FAC-2D de referência e teste sobre as mesmas áreas utilizadas para treinamento e teste do classificador de máxima verossimilhança, utilizando-se amostras do canal original.

As Tabelas 7.9, 7.10, 7.11 e 7.12 apresentam as matrizes de classificação para os seguintes critérios respectivamente: DFAC, DPAC, DLM e DLDM (ver Seção 4.4).

TABELA 7.9

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CRITÉRIO DISTÂNCIA ENTRE FAC'S (DFAC) - CANAL ORIGINAL

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Gramma	100	0	0
Tecido	0	100	0
Madeira	10	0	90

$$D_m = 96.7$$

$$C_m = 3.3$$

Observa-se que o primeiro método (DFAC e DPAC) que envolve a diferença quadrática entre FAC's é o que apresenta melhores resultados, inclusive melhores do que a utilização de filtros inversos. As medidas que envolvem log's dos momentos não demonstraram bom resultado.

TABELA 7.10

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CRITÉRIO DISTÂNCIA POND. ENTRE FAC'S (DPAC) - CANAL ORIGINAL

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Grama	100	0	0
Tecido	0	100	0
Madeira	10	0	90

$$D_m = 96.7$$

$$C_m = 3.3$$

TABELA 7.11

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CRITÉRIO LOG DA DIFERENÇA QUADRÁTICA (DLM) - CANAL ORIGINAL

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Grama	50	40	10
Tecido	10	30	60
Madeira	10	10	80

$$D_m = 53.3$$

$$C_m = 46.7$$

TABELA 7.12

MATRIZ DE CLASSIFICAÇÃO DE AMOSTRAS  
CRITÉRIO DISTÂNCIA ENTRE LOG'S DOS MOMENTOS (DLDM) - CANAL ORIGINAL

AMOSTRAS	CLASSES		
	GRAMA	TECIDO	MADEIRA
Gramma	20	80	0
Tecido	0	40	60
Madeira	10	30	60

$$D_m = 40$$

$$C_m = 60$$

7.2.1.1 - DIAGNÓSTICO DOS RESÍDUOS

Para estudo da máquina de apresentação da função de autocorrelação como método para verificar à qual classe pertence uma amostra foi realizado o seguinte experimento: A Figura 7.3 apresenta sobre a imagem original duas áreas de aquisição da FAC-2D de 30x30 pixels por classe, sendo que uma delas coincide com a área de treinamento usado pelo classificador de máxima verossimilhança pontual, mas não coincide com a área de treinamento utilizada para a modelagem, pois para uma estimação adequada da FAC-2D por limites estabelecidos é necessário uma área maior do que a utilizada para concatenação.

A Figura 7.4 apresenta nessa forma a FAC-2D (limiar de significância = 0.067) das texturas originais estudados. Observam-se claramente as diferenças notáveis entre as FAC'S dessas texturas, sugerindo assim a razão do bom desempenho da medida de distância sugeridas. Por outro lado, quando se comparam visualmente duas FAC'S-2D de amostras distintas de mesma textura as diferenças são pequenas.

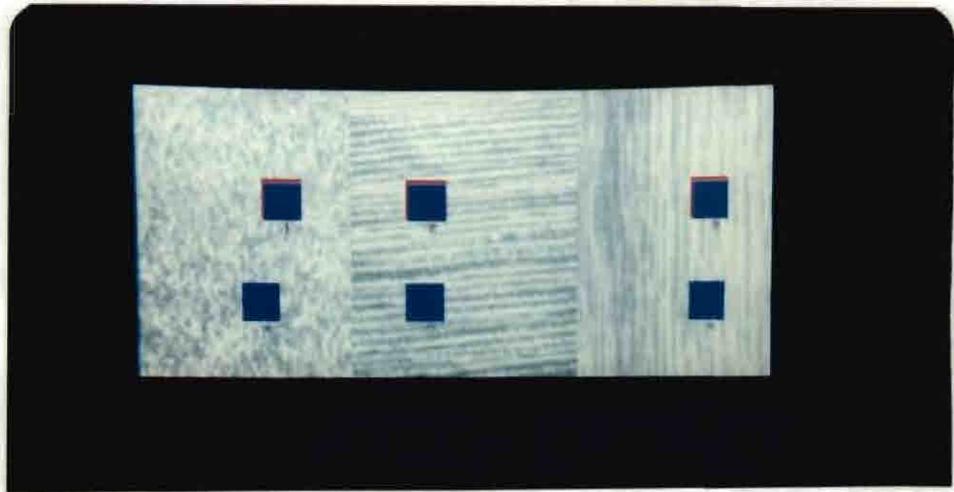
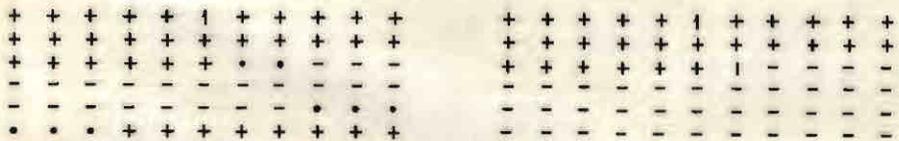


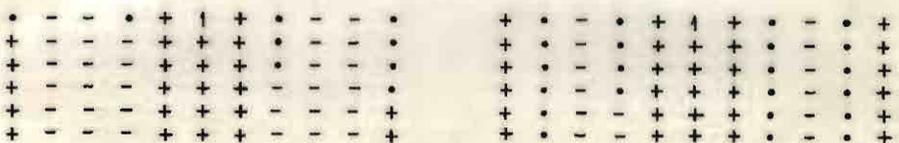
Fig. 7.3 - Áreas teste para levantamento da FAC-2D.



(a)



(b)



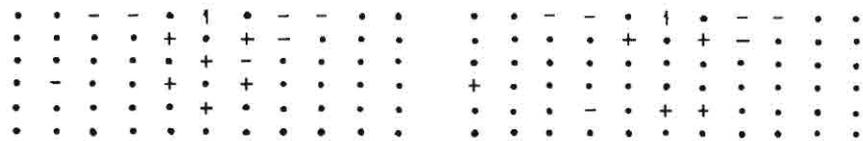
(c)

Fig. 7.4 - Representações bi-dimensionais para a FAC-2D para as classes (a) grama, (b) tecido e (c) madeira.

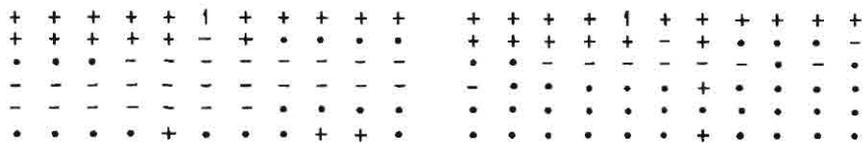
“+” significa correlação positiva significativa, “-” correlação negativa significativa, “•” correlação não significativa.

A verificação do branqueamento total de uma certa textura, quando filtrada por filtro inverso casado, poderia ser considerado como um método para classificar texturas; ocorre que esse método não é unicamente definido pois amostras de resíduos de diferentes texturas poderiam se mostrar ambos eventualmente brancas ou nenhuma delas, tudo isso dependente da amostra em particular e dos filtros envolvidos. O que se pode esperar no entanto é um certo branqueamento da textura casada ao filtro inverso. É o que se verifica ao examinar-se as FAC's-2D da imagem mosaico processada pelo filtro inverso da grama (Figura 7.5; limiar de significância = 0.067).

A FAC-2D da amostra referente à classe grama mostrou um número grande de valores considerado dentro do intervalo de confiança (demonstrando assim o branqueamento) enquanto que as amostras obtidas nas outras texturas dentro do canal inverso grama demonstraram ainda correlações significativas e em distribuições semelhantes ao caso não filtrado. Compare a Figura 7.5 com a Figura 7.4.



( a )



( b )



( c )

Fig. 7.5 -FAC-2D para amostras das classes: (a) grama, (b) tecido e (c) madeira com o filtro inverso grama.

7.2.2 - RADAR DE ABERTURA SINTÉTICA (SAR580)

A Tabela 7.13 apresenta a variância das áreas de treinamento das classes Floresta, Cultura Anual e Restos de Cultura. Cada área é filtrada pelo seu filtro casado e pelos filtros das outras classes. Como se espera a variância é menor na saída do filtro casado àquela área de treinamento, embora nesse caso as diferenças sejam bem menores do que no caso da texturas da imagem de São Paulo (Seção 6.2.3). A classe Floresta tem variância original de 2515, Cultura tem variância 1791 e Restos de Cultura tem variância de entrada de 480.

TABELA 7.13  
VARIÂNCIA NA SAÍDA DOS FILTROS APLICADOS A CADA ÁREA DE TREINAMENTO  
(FILTROS UTILIZADOS: AUTO-REGRESSIVOS)

ÁREA DE TREINAMENTO	<u>FILTRO</u> → VAR. ORIGINAL	FLORESTA	CULTURA ANUAL	RESTOS CULTURA
Floresta	2515	1561	1679	1595
Cultura Anual	1791	1105	745	776
Restos Cultura	480	271	256	247

A Tabela 7.14 apresenta os índices de desempenho da matriz de classificação (com limiar de classificação de 95%), para as áreas de treinamento, para o caso do SAR580 em 3 classes, onde se pode observar a melhoria no poder de classificação desde que a distância JM aumentou, indicando maior poder de separabilidade das classes no espaço transformado. Tal comportamento se refletiu em melhoria nos índices. Essa melhoria, entretanto, não foi muito grande, o que se supõe por causa da textura do "speckle" que mascara a textura natural do terreno.

TABELA 7.14

ÍNDICES DESEMPENHO DA MATRIZ DE CLASSIFICAÇÃO  
ÁREAS DE TREINAMENTO SAR-580 - 3 CLASSES

ÍNDICE	CANAIS	
	ORIGINAL	FILTRADOS
DJM	.5855	.718371
Dm	53.10	58.13
Am	.27	.87
Cm	46.63	41.00

A Tabela 7.15 apresenta os resultados de desempenho calculados pelas matrizes de classificação para os casos de canais processados por curva log. Do canal processado por logaritmo são obtidos três canais residuais usando-se os filtros inversos identificados para esse caso.

TABELA 7.15

ÍNDICES DE DESEMPENHO DA MATRIZ DE CLASSIFICAÇÃO  
CANAL ORIGINAL PROCESSADO POR LOG - ÁREAS DE TREINAMENTO

ÍNDICES	CANAIS	
	LOG DO ORIGINAL	FILTRADOS DE LOG
DJm	.4390	.5984
Dm	50.00	53.00
Am	2.3	5.50
Cm	47.7	41.50

Observa-se também que para esse caso houve ligeira melhoria nos índices do experimento ao se classificar as áreas de treinamento com canal log original para o caso de três canais filtrados; ocorre porém que o desempenho, quando se usa o logaritmo do canal original, é pior do que ao se classificar diretamente o canal original, o mesmo acontecendo com os canais filtrados com e sem log. O modelo que se supõe multiplicativo para imagens de radar não mostrou influência no resultado. Cabe também observar que a saturação imposta pelo processo de reescalonamento logaritmico pode ter ocasionado maior confusão entre os canais e assim diminuir o Dm.

#### 7.2.2.1 - CLASSIFICAÇÃO DE AMOSTRAS

Para avaliar a classificação de amostras em imagens de radar foram obtidas 24 amostras além das áreas de treinamento conforme mostrado na Figura 7.6.

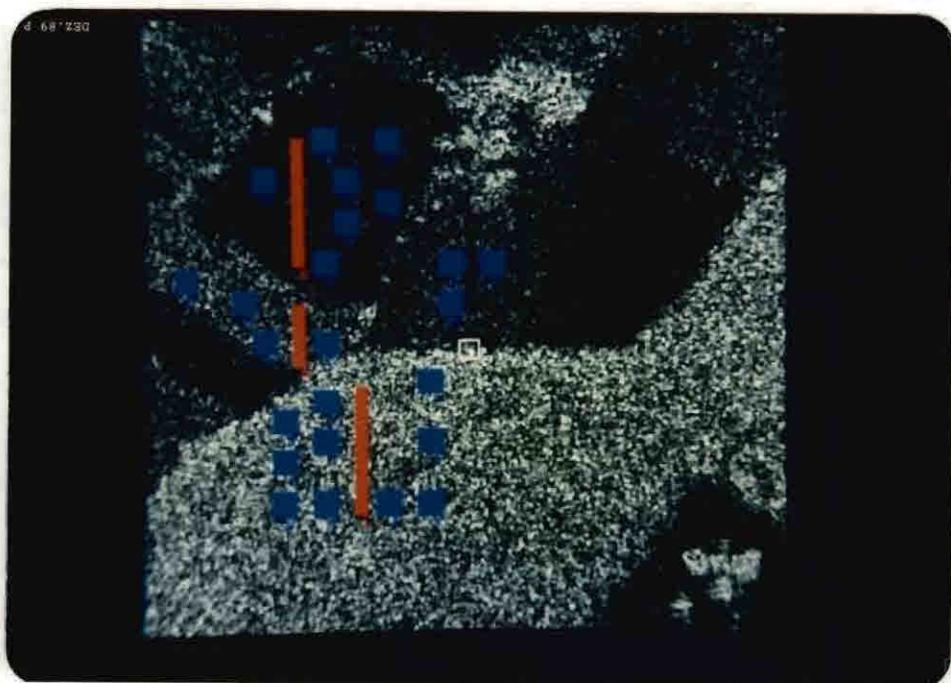


Fig. 7.6 - Amostras teste para classificação.

O índices de desempenho de classificação de amostras estão sumarizados na Tabela 2.16 para os casos de canal original e conjunto de canais filtrados, para os critérios de distância JM (DJM), distância de Mahalanobis (MAH) média e maioria de classificação ponto a ponto (MAI).

TABELA 7.16

CLASSIFICAÇÃO DE AMOSTRAS PARA IMAGENS DE RADAR  
USANDO-SE CANAL ORIGINAL OU CANAIS FILTRADOS

CANAIS	ORIGINAL			FILTRADOS			
	CRITÉRIOS	DJM	MAH	MAI	DJM	MAH	MAI
	ÍNDICES						
	Dm	100,0	91,6	83,3	100,0	100,0	91,6
	Cm	0,0	8,4	16,7	0,0	0,0	8,4

Os índices, embora apresentem uma certa melhoria quando se usam os canais filtrados, não permitem afirmar categoricamente a superioridade do desempenho, pois ocorre que ao se classificar amostras com o canal original já os índices são ou se aproximam de 100%. Ocorre que nesse caso as amostras já se separam pela média e a ajuda da discriminação pela textura é secundária, ao contrário do que ocorre com o exemplo da imagem mosaico de Brodatz, onde a média é similar entre as texturas. Nesse caso a discriminação da textura é fundamental para a separabilidade das amostras.

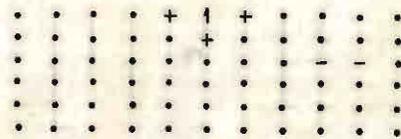
7.2.2.2 - VERIFICAÇÃO DE BRANQUEAMENTO

A Figura 7.7 apresenta amostras adquiridas para levantamento da FAC-2D sobre a imagem de radar. Algumas das áreas adquiridas coincidem em parte com as áreas de treinamento para a aquisição do modelo. Ocorre que as áreas para concatenação tem largura de 10 pixels e as áreas para levantamento da FAC têm tamanho de 20x20.

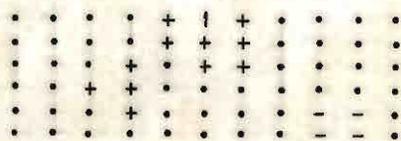
A Figura 7.8 apresenta a FAC-2D, com limiar de significância de 0.1, para as áreas de treinamento das três classes definidas. Nota-se logo que essas imagens apresentam uma FAC-2D que se anula muito mais rapidamente que as FACs das texturas de Brodatz. É a influência do ruído "speckle" que sobressai.



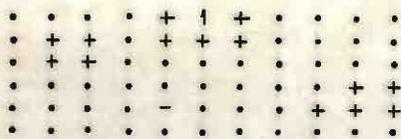
Fig. 7.7 Amostras para aquisição da FAC-2D.



( a )



( b )



( c )

Fig. 7.8 - FAC-2D das amostras das classes (a) floresta, (b) restos de cultura e (c) cultura anual para o canal radar original.

A Figura 7.9 apresenta as FAC-2D (limiar de significância de 0.1) do canal radar original, filtrado pelo filtro inverso da Floresta, para as três classes definidas. Nota-se o branqueamento da amostra relativa à classe floresta, principalmente para atrasos unitários, enquanto que as amostras de outras classes quando filtradas pelo filtro inverso de floresta ainda guardam correlações significativas residuais embora menos que no canal original.

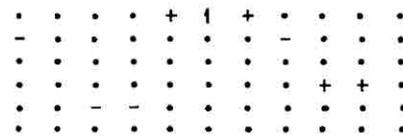
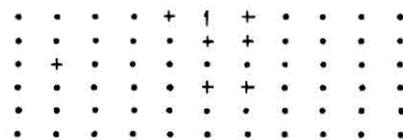
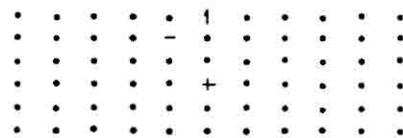


Fig. 7.9 - FAC-2D das amostras das diversas classes obtidas sobre a imagem de radar filtrada pelo Filtro inverso da floresta.

Fica claro observando-se esses resultados que a verificação do branqueamento não deve ser utilizado como critério de classificação de amostras mas apenas serve como auxílio qualitativo para verificação do ajuste de modelos pois conforme os modelos avaliados poderá haver dois modelos que branqueiem uma certa amostra ou como ocorre no exemplo da Figura 7.9(a) ainda resta uma ou outra correlação significativa. Ocorre que se utilizam em geral os limiares de significância fixa desenvolvidos para o caso 1D. Teoricamente porém, os limiares dependem do modelo subjacente ao processo e

não existem na literatura maiores estudos sobre modelos em geral e nem sobre particularidades de processos 2D.

A análise das amostras quando filtrados pelos outros filtros inversos mostrou resultados semelhantes ao verificado para o filtro floresta.

### 7.2.3 - IMAGEM SÃO PAULO - SPOT

Na Seção 6.2.3 verificou-se notável variação no resultados das variâncias sobre áreas de treinamento, maior do que a notada no caso do radar (Seção 6.2.2). Ocorre que essa variação não se refletiu positivamente sobre os resultados do cálculo dos índices de desempenho sobre as matrizes de classificação das áreas de treinamento (Tabela 7.17; limiar de rejeição = 99%) - casos A (canal original) e B (canais filtrados). Para explicar tal resultado negativo duas considerações são feitas:

- 1) O desempenho médio é muito alto já com o canal original, deixando pouco espaço para melhoria desses índices. 2) Além disso, como o filtro inverso da classe montanha é passa-alta, por causa das altas correlações espaciais dessa classe, quando esse filtro passa pela cidade provoca um grande aumento na rugosidade dessa região, com impacto negativo sobre a classificação pontual. Este efeito é semelhante ao efeito de diminuição da divergência (e por conseguinte da  $P(E)$ ) quanto  $\rho_1, \rho_2 \rightarrow 1$  no Exemplo 5.1.

Como sugerem Foglein e Kittler (1983), e confirmado no exemplo 5.1, não se deve usar os canais filtrados quando o filtro é passa-alta. Assim realizou-se a experiência de não incluir o canal inverso correspondente à classe montanha, e manteve-se o canal original e o canal correspondente ao resíduo do filtro inverso da cidade (caso C Tabela 7.17). Nota-se discreta melhoria nos índices desse último caso. Tendo-se em vista porém, que esses índices são estimadores para a  $P(E)$  com uma certa variância, não se pode absolutamente afirmar o melhor desempenho no caso C.

TABELA 7.17

ÍNDICES DESEMPENHO DA MATRIZ DE CLASSIFICAÇÃO  
ÁREAS DE TREINAMENTO SÃO PAULO - SPOT

ÍNDICE	CASOS		
	<i>A(Original)</i>	<i>B(Filtrados)</i>	<i>C(Misto)</i>
DJM	1.39	1.15	1.40
Dm	98.65	93.5	98.9
Am	1.3	0.35	0.85
Cm	0.05	6.15	0.25

Foi realizado também um teste de classificação de amostras com essa imagem SPOT São Paulo e os resultados foram semelhantes aos obtidos com classificação pontual. Foram obtidas 10 amostras por classe de 20 x 20 pontos cada. Para o critério de DJM as 20 amostras foram classificadas corretamente tanto no caso do canal original e com os 2 canais filtrados. Para critérios de distância de Mahalanobis média e maioria, houve uma amostra que foi classificada erroneamente com o uso do canal original. Tal amostra passou a ser classificada corretamente com o uso dos demais canais filtrados. Assim, devido ao fato de que já com o uso dos canais originais as amostras foram classificadas corretamente em sua quase totalidade a melhoria da classificação utilizando-se os canais filtrados não foi muito grande.

7.3 - CONCLUSÕES E SUGESTÕES PARA ESTUDOS FUTUROS

Foi apresentado um método para classificação pontual de texturas estatísticas, baseado em utilização de atributos extraídos por filtragem inversa. Para isso foi necessário desenvolver uma metodologia para estimar modelos ARMA-2D em geral, baseado na utilização de séries concatenadas. Foram deduzidas teoricamente as equações para o cálculo dos parâmetros do classificador de máxima verossimilhança (amostral e pontual) quando o modelo é auto-regressivo.

Avaliou-se também um caso onde foi demonstrado que a melhoria da classificação depende da diversidade dos modelos estimados, ou seja, de quão distantes (diferentes) são as texturas. Foi verificado também o desempenho dos classificadores amostrais de mínima distância JM, de mínima distância de Mahalanobis média e um classificador de amostras por maioria de classes classificados ponto a ponto, quando se usa os atributos extraídos por filtros inversos.

O método proposto se mostrou eficaz quando realmente é necessário utilizar-se a textura para distinção dos objetos da cena e quando esta textura é homogênea em regiões mais ou menos amplas. Quando a diferença de médias entre texturas for tal que por si só já produza um alto índice de desempenho médio ou quando a textura não for muito homogênea para a região em estudo o ganho é marginal.

Foi proposto também um método para classificação de textura baseado em medida de distâncias entre FAC'S e entre momentos dessa FAC-2D.

Em geral, foi verificado que as imagens de sensoriamento remoto apresentam alta correlação espacial em certos tipos de classes, como água e coberturas vegetais, pois a FAC cai muito lentamente, denunciando polos próximos do círculo unitário. Supõe-se que o efeito de filtro passa-baixas (fpb) imposto pelo sistema sensor (constante para todas às classes) aumente esta característica. A remoção desse polo diminuiria o efeito de fpb, além de provocar um realce nas imagens. Para o futuro pretende-se estudar esse efeito utilizando-se de imagem desconvoluídas a partir de modelos finais existentes (Fonseca e Mascarenhas, 1988) e desenvolver um método para estimar o modelo do sensor diretamente dos dados.

No teste do método com imagens de radar, verificou-se que os modelos não diferem muito entre si, o que refletiu em pequena diferença nos modelos das diversas texturas das imagens teste e pequeno ganho nas matrizes de classificação. Aqui conjetura-se que o ruído "speckle" de certa forma mascara as texturas naturais dificultando a determinação de seus modelos. Por outro lado pretende-se estudar a possibilidade de determinação de modelo ARMA para descrever esse ruído "speckle" e a utilização desse modelo para

restauração dessas imagens. Pode-se também estudar o efeito do uso dos filtros de extração do ruído "speckle" na separabilidade das classes.

No campo log, o desempenho dos classificadores utilizando-se ou não atributos ou filtragens inversas deram resultados piores do que no campo original, porém não muito distantes, inviabilizando qualquer juízo a respeito da hipótese de modelo multiplicativo a partir desse experimento. O efeito de saturação imposto pela transformação log pode sugerir também uma provável explicação para o pior desempenho nesse campo.

Os resultados da utilização do método em imagens padrão de Brodatz foram bons, ressaltando-se os resultados da classificação de amostras. O mosaico preparado é um exemplo típico de um caso onde apenas a textura é fator de discriminação, pois a diferença entre médias no caso é desprezível. Observou-se, no entanto, que os modelos não conseguiram branquear totalmente, principalmente as duas texturas semi-estatísticas tecido e madeira. Para esse tipo de textura sugere-se no futuro estudar um método misto onde uma componente determinística seria extraída das texturas e o resíduo seria modelado por séries temporais. Esse método é baseado nos modelos clássicos de decomposição de séries sazonais em componentes estatísticos e determinísticos.

Observou-se também que, em geral, os índices de desempenho para a classificação de amostras foram melhores quando se utilizou do critério de distância JM, e pior quando se utilizou o critério de maioria.

O resultado de classificação de amostras de textura por minimização da distância entre funções de autocorrelação foi considerado muito bom. Não houve entretanto diferenças ao se usar distâncias ponderadas ou não, assim sugere-se utilizar-se das distâncias simples, inclusive porque evita as dificuldades na definição da distância ponderada de garantir igualdade entre FAC'S quando a distância é zero.

Esse método é portanto bem superior ao método sugerido de medir a distância entre momentos da FAC e ao método de Faugeras e Pratt (1980) que necessita de grande número de áreas de treinamento.

Outras sugestões de trabalho futuro que se podem mencionar:

- Elaboração de um estimador de máxima verossimilhança verdadeiramente 2D, para determinação dos coeficientes dos modelos.
- Utilização de modelos para as classes que levem em conta contexto (para classificação MAP).
- Utilização de suporte fixo para identificação de modelos de predição linear e/ou mínimos quadrados generalizados.
- Melhoria nos métodos de determinação das ordens do modelo utilizando-se de outros métodos mencionados na literatura tais como matrizes  $R^*$  e  $S^*$  e funções de autocorrelação inversos.
- Utilização de modelos vetoriais num esquema de um vetor por linha ou conjunto de linhas que predizem um conjunto de pontos à frente.
- Estudo da influência da resolução das imagens nos modelos das texturas e uso desses modelos para interpolação.
- Uso dos modelos para síntese de imagens realistas.
- Elaboração de métodos de classificação de amostras por distâncias diretas entre modelos.
- Uso de métodos de séries temporais para determinação de bordas de texturas.

Pode-se prever também a necessidade, ao se colocar em utilização prática o método aqui proposto, de se construir um sistema especialista que haveria de decidir quando se deve ou não se utilizar dos canais filtrados e quando seria necessário decompor a textura em componentes determinísticas e estatísticas. Auxiliaria também no processo de aquisição das amostras (tamanho e quantidade) e no processo de identificação dos modelos.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ANDRADE, V.M.B. *Classificação de imagens de alta resolução usando atributos de textura*. (Dissertação de Mestrado em Computação Aplicada) - Instituto de Pesquisas Espaciais, São José dos Campos, fev. 1989. 102 p. (INPE-4888-TDL/379).
- AROIAN, L.A. Time series in M dimensions: past, present and future. In: ANDERSON, O.D.; ORD, J.K.; ROBINSON, E.A., ed. *Time series analysis: theory and Practice*, 6., Amsterdam, Elsevier, 1985. Proceedings of the International Conference held at Nottingham University, 1985. p. 241-261.
- BANON, G.J.F. *Etude d'algorithmes d'estimation des paramètres pour l'identification adaptative en temps reel des processus linéaires perturbés par un bruit corrélé*. (These de Docteur-Ingénieur). Université Paul Sabatier, Toulouse, Jan. 1971. 166 p. (N<sup>o</sup> ordre: 269).
- BENNET, R.J. *Spatial time series: analysis, forecasting and control*. London, Pion, 1979.
- BOX, G.E.P.; JENKINS, G.M. *Time series analysis: forecasting and control*. Holden Day, San Francisco, 1970.
- BROCKWELL, P.J.; DAVIS, R.A. *Time series: theory and methods*. Berlin, Springer-Verlag, 1987.
- BRODATZ, P. *Textures: a photograph album for artists and designers*. New York, NY, Dover, 1956.
- CHAPARRO, L.F.; BOUDAUD, L. Two-dimensional linear prediction covariance method and its recursive solution. *IEEE Transactions on Systems Man and Cybernetics*, SMC-17(4):617-621, July/Aug. 1987.
- CLAERBOUT, J.F. *Fundamentals of geophysical data processing*. New York, NY, McGraw-Hill, 1976.

- CHEN, C.H. On two-dimensional ARMA models for image analysis. In: INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION, 5., Miami Beach, Fl, Dec. 1-4, 1980. *Proceedings*. New York, NY, IEEE, 1980, p. 1128-1131.
- DONDES, P.A.; ROSENFELD, A. Pixel classification based on gray level and local 'busyness'. *IEEE Transactions Pattern Analysis and Machine Intelligence*, PAMI-4(1):79-83, Jan. 1982.
- DUDGEON, D.E.; MERSEREAU, R.M. *Multidimensional digital signal processing*, Englewood Cliffs, NJ, Prentice Hall, 1984.
- DUTRA, L.V. Identificação estatística preliminar de imagens de radar: modelos autorregressivos e média móvel (ARMA). In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 5., Natal, 11-15 out. 1988. *Anais*. São José dos Campos, INPE, 1988, v. 2, p. 517-524.
- DUTRA, L.V.; MASCARENHAS, N.D.A. Some experiments with spatial feature extration methods in multispectral classification. *International Journal of Remote Sensing*, 5(2):303-313, 1984.
- FAUGERAS, O.D.; PRATT, W.K. Decorrelation methods of texture feature extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2:323-332, July 1980.
- FOGLEIN, J.; KITTLER, J. The effect of pixels correlations on class separability. *Pattern Recognition Letters* 1:401-407, July 1983.
- FONSECA, L.M.G.; MASCARENHAS, N.D.A. Methods for combined image interpolation - restoration through a FIR filter design technique. In: INTERNATIONAL SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING. Congress 16., Kyoto, Japan, July, 1988. Commission III. *International Archives of Photogrammetry and Remote Sensing*, Kyoto, ISPRS, 1988, v. 27, Part B3, p. 96-205.

- GAMBOTTO, J.P. Two dimensional time series for textures. In: SIMON, J.C.; HARALICK, R.M., ed. *Digital image processing*, New York, NY, D. Reidel Publishing, 1980.
- GIORDANO, A.A.; HSU, F.M. *Least square estimation with applications to digital signal processing*. New York, NY, John Wiley, 1985.
- GRANGER, C.W.J.; NEWBOLD, P. *Forecasting economic time series*. New York, NY, Academic Press, 1977.
- HANNAN, E.J. *Multiple time series*. New York, NY, John Wiley, 1970.
- HARALICK, R.M. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786-804, May 1979.
- HAUGH, L.D. An introductory overview of some recent approaches to modelling spatial time series. In: ANDERSON, O.D., ed. *Time series analysis: Theory and Practice*, 5., Amsterdam, Elsevier, 1984. Proceedings of the International Conference held at Nottingham University. p. 287-302.
- HEITZ, F.; MAITRE, H.; COUESSIN, C. Application of autorregressive models to fine arts painting analysis. *Signal Processing*, 13(1):1-14, July 1987.
- HUANG, T.S. Stability of two-dimensional recursive filters. *IEEE Transactions on Audio and Electroacustics*, AU-20(2):158-163, June 1972.
- II, S.S.; BANON, G.J.F. Comparação do desempenho entre vários filtros para redução do ruído "speckle" em imagens de radar. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 5., Natal, RN, 11-15 out. 1988, *Anais*. São José dos Campos, INPE, 1988. v. 3, p. 565-571.
- JAIN, A. K. Advances in mathematical models for image processing. *Proceedings of the IEEE*, 69(5):122-148, May 1981.

- JAIN, A.K. *Fundamentals of digital image processing*. Englewood Cliffs, NJ, Prentice Hall, 1989.
- KAY, S.M. *Modern spectral estimation: theory and application*. Englewood Cliffs, NJ, Prentice Hall, 1988.
- KETTIG, R.L.; LANDGREBE, D.A. Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics*, GEO-14(1):19-26, Jan. 1976.
- MACCORMIC, B.H.; JAYARAMAMURTHY, S.N. Time series model for texture synthesis. *International Journal of Computer and Information Sciences*, 3(4):329-343, 1974.
- MARPLE, S.L. *Digital spectral analysis*. Englewood Cliffs, NJ, Prentice-Hall, 1987.
- MARZETTA, T.L. Two dimensional linear prediction: autocorrelation arrays, minimum-phase, prediction error filters, and reflections coefficient arrays. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 28(6):725-733, Dec. 1980.
- MASCARENHAS, N.D.A.; VELASCO, F.R.D. *Processamento digital de imagens*. Buenos Aires, Kapelusz-Ebai, 1989.
- MELSA, J.L.; COHN, D.L. *Decision and estimation theory*. New York, NY, McGraw-Hill, 1978.
- MORETTIN, P.A.; TOLOI, C.M.C. *Previsão de séries temporais*. São Paulo, Atual Editora, 1985.
- NELSON, C.R. *Applied time series analysis for managerial forecasting*. San Francisco, CA, Holden Day, 1973.
- OPPENHEIN, A.V.; SCHAFER, R.W. *Digital signal processing*. Englewood Cliffs, NJ, Prentice Hall, 1975.

- PEREIRA, B.B. *Series temporais multivariadas*. Curso apresentado durante o VI Simpósio Nacional de Probabilidade e Estatística, Rio de Janeiro, e realizado na UFRJ, 1984.
- PRIESTLEY, M.B. *Spectral analysis and time series*. London, Academic Press, 1981. (Univariate Series, v. 1).
- ROBINSON, E.A.; TREITEL, S. *Geophysical signal analysis*. Englewood Cliffs, NJ, Prentice-Hall, 1980.
- SOUZA, C.R.; CAMARGOS, L.R.; KUMAR, R.; MASCARENHAS, N.D.A. *Algoritmos para classificação de amostras e seleção de atributos*. São José dos Campos, INPE, 1980. 87p. (INPE-1718-PRE/124).
- TERRIEN, C.W. Relations between 2-D and multichannel linear prediction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-29(3):454-456, June, 1981.
- TERRIEN, C.W. An estimation-theoretic approach to terrain image segmentation. *Computer Vision, Graphics and Image Processing*, 22(3):313-326, June 1983.
- TERRIEN, C.W. *The analysis of multichannel two-dimensional random signals*. Monterrey, CA, Naval Postgraduate School, Oct. 1986. (NPS-87-002).
- TERRIEN, C.W.; QUATIERY, T.F.; DUDGEON, D.E. Statistical model based algorithms for image analysis. *Proceedings of the IEEE*, 74(4):532-551 Apr. 1986.
- TIAO, G.C.; BOX, G.E.P. Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76(376):802-816, Dec. 1981.

- TOU, J.T. Pictorial feature extraction and recognition via image modeling.  
In: ROSENFELD, A., ed. *Image Modelling*. London, Academic Press,  
1981.
- TOU, J.T.; KAO, D.B.; CHANG, Y.S. Pictorial texture analysis and synthesis.  
In: INTERNATIONAL JOINT CONFERENCE ON PATTERN RECOGNITION,  
3., Coronado, CA, USA, Nov. 8-11, 1976. *Proceedings*. Long Beach, CA,  
IEEE, 1976, p. 590a-590p.
- TUBBS, J.D.; COBERLY, W.A. Spatial correlation and its effect upon  
classification results in Landsat. In: INTERNATIONAL SIMPOSIUM ON  
REMOTE SENSING OF ENVIRONMENT, 12., Ann Arbour, MI, 1978.  
*Proceedings*. Ann Arbour, MI, ERIM, 1978, p. 775-781.
- WIGGINS, R.A.; ROBINSON, E.A. Recursive solution to the multichannel  
filtering problem. *Journal of Geophysical Research*, 70(8):1885-1891,  
Apr. 1965.
- WILLSKY, A.S. *Digital signal processing and control and estimation  
theory*. Cambridge, MA, MIT Press, 1979.

## APÊNDICE A

### CALCULO DA DIVERGÊNCIA ENTRE DUAS CLASSES

A expressão da divergência para as classes  $H_1$  e  $H_2$ , para o caso gaussiano, é dada pela Equação 5.24, aqui reproduzida:

$$\begin{aligned} D_{12} &= \frac{1}{2} \text{tr}[(\Sigma_1 - \Sigma_2)(\Sigma_2^{-1} - \Sigma_1^{-1})] + \frac{1}{2} \text{tr}[(\Sigma_1^{-1} + \Sigma_2^{-1})(\underline{m}_1 - \underline{m}_2)(\underline{m}_1 - \underline{m}_2)^T] = \\ &= \frac{1}{2} \text{trR} + \frac{1}{2} \text{trM} \end{aligned} \quad (\text{A.1})$$

onde os vetores de média e matrizes de covariança são dados pelas Equações 5.16, 5.21 e 5.22:

$$\underline{m}_1 = \begin{bmatrix} 0 \\ \phi_2^T \delta \end{bmatrix} \quad \underline{m}_2 = \begin{bmatrix} -\phi_1^T \delta \\ 0 \end{bmatrix}$$

$$\Sigma_1 = \sigma^2 \begin{bmatrix} 1 - \rho_1^2 & 1 - \rho_1^2 \\ 1 - \rho_1^2 & 1 - 2\rho_1\rho_2 + \rho_2^2 \end{bmatrix}$$

$$\Sigma_2 = \sigma^2 \begin{bmatrix} 1 - 2\rho_1\rho_2 + \rho_1^2 & 1 - \rho_2^2 \\ 1 - \rho_2^2 & 1 - \rho_2^2 \end{bmatrix}$$

O primeiro termo (R) da divergência é devido à diferença existente entre as rugosidades das duas texturas. Usando 5.21 e 5.22 temos:

$$\Sigma_1 - \Sigma_2 = \sigma^2 \begin{bmatrix} 2\rho_1\rho_2 - 2\rho_1^2 & \rho_2^2 - \rho_1^2 \\ \rho_2^2 - \rho_1^2 & 2\rho_2^2 - 2\rho_1\rho_2 \end{bmatrix} \quad (\text{A.2})$$

Calculando os determinantes das matrizes de covariâncias

$$\begin{aligned}
 |\Sigma_1| &= \sigma^2(1-\rho_1^2)(\rho_1-\rho_2)^2 \\
 |\Sigma_2| &= \sigma^2(1-\rho_2^2)(\rho_1-\rho_2)^2
 \end{aligned}
 \tag{A.3}$$

e usando 5.21 obtém-se o inverso de  $\Sigma_1$ ,

$$\Sigma_1^{-1} = \frac{1}{\sigma^2(1-\rho_1^2)(\rho_1-\rho_2)^2} \begin{bmatrix} 1-2\rho_1\rho_2+\rho_2^2 & -1+\rho_1^2 \\ -1+\rho_1^2 & 1-\rho_1^2 \end{bmatrix}$$

$\Sigma_2$  pode ser calculado de modo semelhante, obtendo-se então

$$\Sigma_2^{-1} - \Sigma_1^{-1} = \begin{bmatrix} \frac{1}{\sigma^2(1-\rho_1^2)} & 0 \\ 0 & \frac{1}{\sigma^2(1-\rho_2^2)} \end{bmatrix}
 \tag{A.4}$$

OBSERVAÇÃO A.1 - Observando-se a expressão da diferença das matrizes de covariância, que também aparece quando se igualam as funções discriminantes, para o caso de classificação de máxima verossimilhança, conclui-se que a superfície de separação entre as duas classes, no espaço de atributos, é uma hipérbole com eixos paralelos aos eixos coordenados porque o determinante dessa diferença é negativo, e os termos fora da diagonal principal são nulos, como se pode facilmente constatar. (Ver Seção 4). Se diferentes variâncias fossem consideradas para as classes, os eixos da hipérbole não mais seriam paralelos aos eixos coordenados, pois os termos fora da diagonal principal não seriam mais nulos.

O termo R pode agora ser calculado, usando A.2 e A.4:

$$R = \begin{bmatrix} \frac{2\rho_1\rho_2 - 2\rho_1^2}{(1-\rho_1^2)} & 0 \\ 0 & \frac{2\rho_2^2 - 2\rho_1\rho_2}{(1-\rho_2^2)} \end{bmatrix}$$

e finalmente o termo da divergência devido à rugosidade é dado por

$$\frac{1}{2} \text{tr } R = (\rho_1 - \rho_2) \left( \frac{\rho_1}{1-\rho_1^2} - \frac{\rho_2}{1-\rho_2^2} \right) \quad (\text{A.5})$$

Para o cálculo do termo M, necessitam-se dos seguintes termos:

$$\Sigma_1^{-1} + \Sigma_2^{-1} = \frac{1}{X} \begin{bmatrix} (1-\rho_2^2)[2-2\rho_1^2+(\rho_1-\rho_2)^2] & -2(1-\rho_1^2)(1-\rho_2^2) \\ -2(1-\rho_1^2)(1-\rho_2^2) & (1-\rho_1^2)[2-2\rho_2^2+(\rho_1-\rho_2)^2] \end{bmatrix} \quad (\text{A.6})$$

onde

$$X = \sigma^2(1-\rho_1^2)(1-\rho_2^2)(\rho_1-\rho_2)^2$$

Usando-se 5.15 e 5.16

$$\underline{m}_1 - \underline{m}_2 = \begin{bmatrix} \underline{\phi}_1^T \delta \\ \underline{\phi}_2^T \delta \end{bmatrix} \quad (\text{A.7})$$

$$(\underline{m}_1 - \underline{m}_2)(\underline{m}_1 - \underline{m}_2)^T = \delta^2 \begin{bmatrix} (1-\rho_1)^2 & (1-\rho_1)(1-\rho_2) \\ (1-\rho_1)(1-\rho_2) & (1-\rho_2)^2 \end{bmatrix} \quad (\text{A.8})$$

O fator M é dado por A.6 e A.8, assim  $\frac{1}{2} \text{tr} M$  - divergência das médias - depois de simplificações necessárias, é dado por

$$\frac{1}{2}\text{trM} = \frac{\delta^2}{2\sigma^2} \left[ 2 + \frac{(1-\rho_1)^2}{(1-\rho_1^2)} + \frac{(1-\rho_2)^2}{(1-\rho_2^2)} \right] \quad (\text{A.9})$$

Assim, de A.5 e A.9, a divergência,  $D_{12}$ , entre as duas classes é dada por

$$D_{12} = (\rho_1 - \rho_2) \left[ \frac{\rho_1}{1-\rho_1^2} - \frac{\rho_2}{1-\rho_2^2} \right] + \frac{\delta^2}{2\sigma^2} \left[ 2 + \frac{(1-\rho_1)^2}{(1-\rho_1^2)} + \frac{(1-\rho_2)^2}{(1-\rho_2^2)} \right] =$$
$$D_{12} = (\rho_1 - \rho_2) \left[ \frac{\rho_1}{1-\rho_1^2} - \frac{\rho_2}{1-\rho_2^2} \right] + \frac{\delta^2}{2\sigma^2} \left[ 2 + \frac{(1-\rho_1)}{(1+\rho_1)} + \frac{(1-\rho_2)}{(1+\rho_2)} \right] \quad (\text{A.10})$$