UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

JOÃO BATISTA FLORINDO

**Análise de Dados Funcionais Aplicada à
Geração de Descritores de Assinaturas de
Dimensão Fractal Multiescala**

São Carlos
2008

JOÃO BATISTA FLORINDO

Análise de Dados Funcionais Aplicada à Geração de Descritores de Assinaturas de Dimensão Fractal Multiescala

Dissertação apresentada ao Programa de Pós-graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para a obtenção do título de Mestre em Ciência.

Área de Concentração: Física Aplicada

Opção: Física Computacional

Orientador: Prof. Dr. Odemir Martinez Bruno

São Carlos

2008

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC/USP

Florindo, João Batista

Análise de dados funcionais aplicada à geração de descritores de assinaturas de dimensão fractal multiescala./João Batista Florindo; orientador Odemir Martinez Bruno.-- São Carlos, 2008.

100 p.

Dissertação (Mestrado em Ciências - Área de concentração: Física Aplicada - Opção Física Computacional) – Instituto de Física de São Carlos da Universidade de São Paulo.

1. Análise de dados funcionais. 2. Dimensão fractal multiescala.
3. Reconhecimento de padrões .4. Descritores fractais. I. Título.

FOLHA DE APROVAÇÃO

João Batista Florindo

Dissertação apresentada ao Instituto de Física
de São Carlos da Universidade de São Paulo
para obtenção do título de Mestre em Ciências.
Área de Concentração: Física Aplicada – Opção:
Física Computacional.

Aprovado em: 19/01/2009

Comissão Julgadora

Prof. Dr. Odemir Martinez Bruno

Instituição: IFSC/USP

Assinatura

Prof. Dr. Luiz Vitor de Souza Filho

Instituição: IFSC/USP

Assinatura

Prof. Dr. Hélio Pedrini

Instituição: UNICAMP

Assinatura

Dedicatória

Dedico este trabalho à memória de minha saudosa mãe, Hélia, pelo amor dedicado e educação esmera com a qual me constituiu como homem.

Agradecimentos

- Ao meu orientador, Odemir Martinez Bruno, pela oportunidade de realizar este trabalho, pela orientação competente, confiança e amizade.
- Ao Prof. Dr. Mário de Castro Andrade Filho pelo auxílio durante várias etapas desta pesquisa, sobretudo no tocante às ferramentas estatísticas.
- À minha namorada, Rosi, pelo carinho, confiança e apoio psicológico e espiritual.
- Ao meu pai, Luiz, e meu irmão, José Carlos, pelo apoio psicológico e financeiro no início do mestrado.
- Aos meus amigos do Grupo de Oração Universitário pelo apoio espiritual e pela amizade.
- Aos amigos atuais ou que já passaram pela república, Ageu, Claudinei, Thiago, Tiago, Wilson, André, Rafael, Henrique, Lucas, Mário, Carlos, entre outros.
- Aos amigos do laboratório, André, Jarbas, Márcio, Cláudio, Marco, Dalcimar, pelo companheirismo e ajuda em tantas dúvidas tiradas.
- À Fundação de Amparo à Pesquisa no Estado de São Paulo - FAPESP e à sociedade brasileira, pelo financiamento deste trabalho.

“A humildade é o primeiro degrau para a sabedoria.”
— SÃO TOMÁS DE AQUINO (1225 - 1274)

*“A busca de Deus é a busca da felicidade.
O encontro com Deus é a própria felicidade.”*
— SANTO AGOSTINHO (354 - 430)

“Não tentes ser bem sucedido, tenta antes ser um homem de valor.”
— ALBERT EINSTEIN (1879 - 1955)

Resumo

Esta dissertação faz um estudo da aplicação da técnica estatística denominada Análise de Dados Funcionais (ADF) à geração de descritores usados em reconhecimento de padrões, mais especificamente, no reconhecimento de objetos de interesse em imagens. Estes objetos podem ser representados por vetores de características, também chamados de assinaturas, obtidos por uma técnica chamada de Dimensão Fractal Multiescala (DFM). Ocorre que estes vetores apresentam alta dimensionalidade (número de elementos), fazendo-se assim necessário o uso de uma abordagem que reduza este número de valores, sem que haja uma grande perda da informação transmitida pela assinatura. Neste contexto, diversas técnicas de extração de um reduzido conjunto de descritores da assinatura são apresentadas pela literatura. Entre estas, as mais populares são Fourier e *wavelets*, ambas relativamente simples de se apresentar e com resultados satisfatórios. A proposta aqui apresentada é de se utilizar ADF em combinação com DFM na geração de descritores de padrões. Os resultados obtidos com o uso desta abordagem na geração de descritores demonstraram que a técnica possibilita bons resultados, mesmo em situações em que não é possível o uso de muitos descritores. Os experimentos demonstraram que ADF apresenta um bom potencial para aplicação neste tipo de problema, permitindo que o método de classificação alcance bons resultados mesmo com poucos descritores. São sugeridos trabalhos futuros em que ADF possa ser usada, pesquisando-se por métodos ainda mais eficazes.

Palavras-chave: Análise de dados funcionais. Dimensão fractal multiescala. Reconhecimento de padrões. Descritores fractais.

Abstract

This work studies the application of a statistical technique named Functional Data Analysis (FDA) for the generation of descriptors. These descriptors can be used for pattern recognition, more specifically, for the recognition of relevant objects in an image. These objects can be represented by features vectors, also known as signatures, obtained by a technique named Multi-scale Fractal Dimension (MFD). These vectors present a high dimensionality (number of elements), causing to be necessary the use of an approach for the reduction of this number of values, but without a large loss of information carried by the signature. In this context, several techniques for the extraction of a reduced set of signature descriptors are studied in the literature. Among these techniques, the most classic are Fourier and wavelets, both with simple presentation and providing satisfactory results. The proposal presented here is the use of FDA combined with MFD for the generation of pattern descriptors. The results obtained by the use of this approach for the generation of descriptors showed that this technique allows the obtention of good results, even in situations in which is not possible the use of many descriptors. FDA was also applied to the extraction of descriptors of MFD texture signatures. Also in this case, the results were interesting. The experiments showed the FDA presents a good potential for the application to this type of problem, allowing the obtention of good results even by using a few descriptors. It is suggested future works in which FDA can be used, researching for still more efficient methods.

Keywords: Functional data analysis. Multi-scale fractal dimension. Pattern recognition. Fractal descriptors.

Lista de Figuras

Figura 1 -	Fractal de Mandelbrodt. Regra de geração simples produz forma complexa no limite tendendo ao infinito.	23
Figura 2 -	Curva de Koch. objeto fractal cujo cálculo direto da dimensão é simples.	25
Figura 3 -	Curva do cálculo da dimensão fractal por Box-Counting.	27
Figura 4 -	Geração da curva DFM. (a) Imagem original. (b) Curva de $A(r)$ gerada por Bouligand-Minkowsky. (c) Curva DFM.	31
Figura 5 -	Fenômeno de Gibbs. (a) Sinal original. (b) Derivada do sinal em (a) calculada pelo método derivativo de Fourier, sem tratamento especial. (c) Curva em (a) replicada horizontalmente. (d) Derivada convencional de Fourier do sinal em (c). (e) Parte da curva em (d) interessante para o problema. (f) Reflexão horizontal da curva em (e) em volta de seu ponto médio cuja translação e escala geram a derivada de (a). (g) Comparação entre a derivada de Fourier tratada e a derivada analítica.	34
Figura 6 -	Multiescala usada na geração de assinaturas. (a) Quádrica de Koch. (b) Gráfico log-log da área de influência obtida pelo método de Minkowsky em função do raio.	35
Figura 7 -	Curva obtida a partir do gráfico log-log por eliminação dos pontos de baixa amostragem e interpolação por raios médios.	36
Figura 8 -	Curva DFM obtida.	36
Figura 9 -	DFM usada na classificação de imagens de letras do alfabeto. Nota-se que imagens de classes diferentes produzem assinaturas distintas.	38
Figura 10 -	Assinatura DFM menos a assinatura portadora. Nota-se que objetos distintos mas apresentando valores similares para a dimensão fractal (DF) podem ser discriminadas por esta curva DFM subtraída da portadora.	38
Figura 11 -	Curva de amostra da altura de 10 pessoas em função da idade.	41
Figura 12 -	Curva da derivada primeira (velocidade) da função de altura.	42
Figura 13 -	Curva da derivada segunda (aceleração) da função de altura.	42

Figura 14 - Curva do índice mensal de bens não-duráveis nos Estados Unidos ao longo de anos.	44
Figura 15 - Curva de uma função <i>B-spline</i> suave ajustada ao logaritmo decimal do índice de bens não-duráveis.	44
Figura 16 - Gráfico fase-plano do índice de bens não-duráveis. As letras na curva representam os meses do ano.	45
Figura 17 - Bases <i>splines</i> aplicadas à representação de funções analíticas suaves. (a) Funções originais. (b) Funções-base <i>splines</i> . (c) Funções representadas por <i>splines</i>	50
Figura 18 - Descritores de Fourier. (a) Sinais originais obtidos pela aplicação de DFM à forma das letras A e B. (b) 50 descritores de Fourier para as mesmas assinaturas.	60
Figura 19 - Árvore de decomposição no método de <i>wavelets</i>	62
Figura 20 - Esquema demonstrando a aplicação de ADF no presente trabalho.	72
Figura 21 - Base de formas variadas. Cada linha apresenta uma classe de formas.	74
Figura 22 - Base de letras do alfabeto. Cada coluna representa uma letra (classe) afetada por diferentes graus de ruído.	74
Figura 23 - Base shapes99. (a) Classificação usando K-médias. (b) Classificação usando ADFI. (c) Classificação usando RNA.	77
Figura 24 - Base de letras do alfabeto com ruído de nível 1. (a) Classificação usando K-médias. (b) Classificação usando ADFI. (c) Classificação usando RNA.	78
Figura 25 - Base de letras do alfabeto com ruído de nível 2. (a) Classificação usando K-médias. (b) Classificação usando ADFI. (c) Classificação usando RNA.	79
Figura 26 - Base de letras do alfabeto com ruído de nível 3. (a) Classificação usando K-médias. (b) Classificação usando ADFI. (c) Classificação usando RNA.	80
Figura 27 - Base de letras do alfabeto com ruído de nível 4. (a) Classificação usando K-médias. (b) Classificação usando ADFI. (c) Classificação usando RNA.	81
Figura 28 - Base de letras do alfabeto com os 4 níveis de ruídos misturados. (a) Classificação usando K-médias. (b) Classificação usando ADFI. (c) Classificação usando RNA.	82
Figura 29 - Base de texturas de Brodatz. (a) Classificação usando K-médias. (b) Classificação usando ADFI.	83
Figura 30 - Base de formas MPEG-7. (a) Classificação usando K-médias. (b) Classificação usando ADFI.	84

Figura 31 - Análise discriminante flexível aplicada à base de peixes.	84
Figura 32 - Matriz de confusão para a base de letras usando-se descritores de Fourier.	89
Figura 33 - Matriz de confusão para a base de letras usando-se descritores <i>wavelets</i>	89
Figura 34 - Matriz de confusão para a base de letras usando-se descritores ADF.	90
Figura 35 - Componentes Principais usando-se descritores de Fourier.	90
Figura 36 - Componentes Principais usando-se descritores <i>wavelets</i>	91
Figura 37 - Componentes Principais usando-se descritores ADF.	91

Lista de Tabelas

Tabela 1 - Tabela do tempo de processamento em função do número de descritores usados.	76
Tabela 2 - Tabela de melhores resultados para os métodos.	86
Tabela 3 - Tabela de resultados médios com seus respectivos desvios padrões para os métodos.	87
Tabela 4 - Tabela de melhores resultados usando-se ADL.	88

Lista de Abreviaturas

<i>ADF</i>	Análise de Dados Funcionais
<i>DFM</i>	Dimensão Fractal Multiescala
<i>RNA</i>	Rede Neural Artificial
<i>ADL</i>	Análise Discriminante Linear
<i>ADFl</i>	Análise Discriminante Flexível
<i>RP</i>	Reconhecimento de Padrões
<i>ACP</i>	Análise de Componentes Principais
<i>DF</i>	Dimensão Fractal
<i>MLP</i>	Multi-layer Perceptron
<i>ASR</i>	Average Square Residue

Sumário

1	Introdução	18
1.1	Caracterização do Problema	18
1.2	Motivação	19
1.3	Contribuições	19
1.4	Objetivos	20
1.5	Organização	21
2	Dimensão Fractal Multiescala	22
2.1	Definição e Conceito Matemático de Dimensão Fractal	22
2.2	Métodos de Cálculo	25
2.2.1	Box-Counting	26
2.2.2	Bouligand-Minkowsky	27
2.3	Dimensão Fractal Multiescala	29
2.3.1	Transformação espaço-escala	30
2.3.2	Métodos de Diferenciação Numérica	32
2.3.3	Aplicação à Geração de Assinaturas	35
2.4	Dimensão Fractal Multiescala Aplicada ao Reconhecimento de Padrões de Textura	37
3	Análise de Dados Funcionais	40
3.1	O que é um Dado Funcional: Definições e Conceitos Matemáticos	40
3.1.1	Anatomia de um Dado Funcional e Dimensionalidade	41
3.1.2	Gráfico Fase-plano	43

3.2	Representação de Conjuntos de Valores por funções	45
3.2.1	Bases de Fourier	49
3.2.2	Sistema de Funções-base tipo <i>Splines</i>	49
3.3	Suavização de Dados Funcionais por Mínimos Quadrados	52
3.4	Motivação	54
4	Classificação de Formas com Base em Descritores	57
4.1	Descritores de Fourier	57
4.2	Descritores <i>Wavelets</i>	59
4.3	Aplicação de ADF à Geração de Descritores	62
4.3.1	Uso Direto	62
4.3.2	Mudança de Espaço Algébrico	63
5	Métodos de Classificação	65
5.1	K-médias	65
5.2	Rede Neural Artificial	66
5.3	Análise Discriminante Linear	67
5.4	Análise Discriminante Flexível	69
6	Metodologia	71
6.1	Passos Adotados	71
7	Resultados Experimentais	73
7.1	Bases de Dados Utilizadas	73
7.2	Detalhes dos Experimentos	75
7.3	Tempo Computacional em função do número de descritores	76
7.4	Análise dos Resultados	76
7.4.1	Shapes99	76

7.4.2	Letras do Alfabeto	77
7.4.3	Resultados para a análise de textura	82
7.4.4	MPEG-7	83
7.4.5	Contornos de Peixes	84
7.4.6	Melhores Resultados Obtidos	85
7.5	Matrizes de Confusão	88
7.6	Análise de Componentes Principais	90
7.7	Síntese dos Resultados	91
8	Conclusões	93
8.1	Conclusões Sobre o Trabalho Realizado	93
8.2	Considerações Finais	94
8.3	Sugestões para Trabalhos Futuros	94
8.4	Trabalhos gerados	95
	Referências	96

1 Introdução

Neste capítulo inicial são introduzidos os conceitos fundamentais relacionados à área de pesquisa desta dissertação, além da apresentação dos assuntos a serem efetivamente estudados neste trabalho.

1.1 Caracterização do Problema

Em um objeto fractal (1), o valor de sua dimensão fractal (DF) é importante para seu estudo, uma vez que expressa o índice de ocupação do espaço no qual o objeto fractal está inserido. Este valor indica ainda a fractalidade do objeto, que está relacionada à complexidade da entidade fractal.

Embora os objetos do mundo real, como contornos de nuvens e árvores ou mesmo texturas de imagens, não sejam fractais verdadeiros, o valor da dimensão fractal destes objetos pode ser usado como um bom caracterizador do objeto. Para que o cálculo da dimensão fractal de objetos do mundo real se torne possível, entretanto, é necessária a adaptação de alguns métodos clássicos de cálculo (1) para o espaço discreto das representações computacionais dos objetos (2).

Entretanto, como o valor da dimensão fractal é apenas um número para caracterizar todo o objeto de interesse, (3) propõe uma abordagem multiescala, a partir da qual o processo de cálculo da DF gera um vetor de características do objeto, usado para caracterizá-lo com maior precisão, permitindo, por exemplo, seu posterior uso em métodos de classificação. Como exemplo de trabalho que faz uso de DFM para a classificação de objetos pode se citar (3), no qual a técnica é aplicada à classificação de espécies da flora brasileira com base em órgãos foliares.

1.2 Motivação

Como o vetor de características obtidos por DFM corresponde a um conjunto grande de valores, seu processamento por métodos clássicos de classificação se torna muito caro e são necessárias técnicas que reduzam o vetor a um pequeno conjunto de valores chamados de descritores deste vetor. Em (3), usa-se a técnica de descritores de Fourier, já bem conhecida da literatura. Ainda descritores *wavelets* (4) podem ser empregados para o mesmo fim.

Os bons resultados em aplicações práticas demonstrados na literatura encorajaram um aprofundamento maior nesta técnica de DFM. Esta dissertação apresenta e estuda o uso de um novo método para a obtenção de descritores, denominado Análise de Dados Funcionais(ADF) (5)(6). ADF é desenvolvida em (5) como uma técnica que muda a abordagem estatística usada na manipulação de grandes conjuntos de valores. Em vez de se tratarem estes conjuntos como simplesmente vetores de números, sem uma correlação maior entre os componentes, os dados são tratados como sendo uma função analítica que gera os valores presentes no vetor original.

1.3 Contribuições

Simplificadamente, pode se dizer que a idéia de ADF é obter uma função x tal que para cada valor $y(t)$ presente no vetor:

$$y(t) = x(t) + \epsilon, \quad (1.1)$$

em que ϵ é um ruído decorrente da aproximação realizada, uma vez que geralmente é inviável a obtenção exata de x . Para facilitar os cálculos e a representação no computador, x costuma ser representado por uma combinação linear de funções-base:

$$x(t) = \sum_{k=1}^K c_k \phi_k(t), \quad (1.2)$$

em que ϕ são as funções-base e c_k são coeficientes. Os descritores podem ser tomados como sendo os próprios coeficientes diretamente como se faz em (7) ou por meio de processamento efetuado sobre estes coeficientes como descrito em (8). Trabalhos presentes na literatura, como (7) e (8), demonstram que a segunda abordagem apresenta os melhores

resultados e por isso esta estratégia foi usada em todos os experimentos.

1.4 Objetivos

O objetivo deste trabalho é verificar a aplicação de descritores ADF associada ao uso de assinaturas DFM na classificação de objetos do mundo real com base na análise do contorno, forma, textura, entre outras propriedades destes objetos. DFM é usada para gerar uma assinatura do objeto de interesse. Já é conhecido na literatura o uso de descritores de Fourier (9) e *wavelets* (4) para representar com menos componentes esta assinatura. A proposta aqui é aplicar ADF à geração destes descritores. Com isto, espera-se uma vantagem em relação aos métodos já tradicionalmente usados para este fim, uma vez que ADF permite uma visão mais global da assinatura, que é tratada como sendo uma única função analítica e não apenas como um conjunto de pontos no espaço discreto.

Os objetivos principais do estudo são:

- Propor e estudar o uso de ADF na geração de descritores de assinaturas DFM
- Apresentar e verificar uma metodologia para se usar um objeto funcional como dado a ser processado por um método de classificação
- Verificar a precisão de ADF como geradora de descritores, a partir da comparação com outros métodos de descritores
- Fazer a verificação e validação usando-se diferentes métodos de classificação
- Identificar qual método de descritores apresenta maior separação entre as classes e menor separação intra-classes usando-se recursos como o gráfico de componentes principais dos descritores

Para testes foram usadas cinco bases de dados, sendo duas do mundo real e as outras três artificiais, usadas para evidenciar com mais força os bons resultados alcançados pela presente abordagem. Para a validação do uso de ADF e a medida de seu desempenho foram usados, para fins de comparação, descritores de Fourier e de *wavelets*. Os métodos usados para a classificação foram os mais conhecidos e que apresentaram melhores resultados na literatura, ou seja, K-médias, Rede Neural Artificial (RNA), Análise Discriminante Linear (ADL) e Análise Discriminante Flexível (ADFL).

1.5 Organização

Esta dissertação está dividida em 8 capítulos, incluindo esta introdução. No próximo capítulo será abordada a técnica de Dimensão Fractal Multiescala, tanto sob o ponto de vista teórico como em termos da aplicação em reconhecimento de padrões. No terceiro capítulo será abordada a técnica de Análise de Dados Funcionais, mostrando-se seu histórico e embasamento matemático, além de tratar de aplicações já estudadas na literatura e possíveis aplicações futuras. No quarto capítulo o tema tratado é o uso de descritores conhecidos pela literatura, a saber, Fourier e *wavelets*, para se reduzir a dimensionalidade das assinaturas obtidas por DFM. O capítulo quinto trata dos métodos estatísticos usados para validação por meio de comparação entre os métodos já conhecidos. O sexto capítulo aborda a metodologia usada. O sétimo trata dos experimentos e dos resultados obtidos, bem como uma análise destes resultados. O oitavo capítulo trata das considerações finais, da contribuição dada ao mundo científico e da sugestão de trabalhos futuros, bem como dos trabalhos gerados em decorrência deste mestrado.

2 *Dimensão Fractal Multiescala*

Neste capítulo são abordados os aspectos teóricos e da aplicação de conceitos da teoria de fractais e da técnica chamada de dimensão fractal multiescala, usada nesta dissertação.

2.1 Definição e Conceito Matemático de Dimensão Fractal

O conceito de fractal surgiu há mais de um século, mas começou a apresentar um número significativo de aplicações nas últimas décadas, levando ao surgimento da chamada Teoria do Caos, empregada na resolução de sistemas dinâmicos complexos (10), permitindo também o estudo de objetos com formas irregulares presentes na natureza, como as montanhas ou as folhas das árvores. Ocorre que, na geometria tradicional euclideana, todo estudo era feito sobre formas perfeitas, com valor inteiro para a dimensão. Entretanto, há que se notar que a representação da folha de uma árvore por uma elipse por exemplo implica na perda de um significativo conjunto de informações, referentes às irregularidades presentes naquela folha.

O interesse por este novo tratamento para as formas geométricas se deu desde a percepção de que muitas formas da natureza não podiam ser representadas pela geometria euclideana. Estas entidades vinham despertando o interesse de matemáticos como Cantor e Peano desde o século XIX e foi retomado seu estudo formal por Mandelbrodt em 1968 (11). Estas novas formas possuíam uma regra de construção simples e apresentavam auto-similaridade em nível de escala (a forma era construída iterativamente e cada iteração era composta de uma série de cópias fiéis da iteração anterior), o que as tornava muito complexas. Estas formas receberam por Mandelbrodt o nome de fractais e ficaram famosas quando apareceram na representação de fenômenos aperiódicos da natureza, possuindo posteriormente grande aplicação em problemas de reconhecimento de padrões e de análise e processamento de imagens (12).

Figura 1 – *Fractal de Mandelbrodt. Regra de geração simples produz forma complexa no limite tendendo ao infinito.*

Atualmente, várias aplicações podem ser encontradas para a geometria fractal e particularmente para a dimensão fractal (13). Entre estas, encontram-se aplicações teóricas como em teoria dos números, em aplicações como a distribuição de dígitos da expansão em base- m de um inteiro ou aproximação diofantina (1). Ainda na matemática pura, a teoria de fractais aparece na teoria de grafos, no problema de Keakeya (14), na conjectura de Vitushkin (15), no estudo do conjunto de pontos de não diferenciabilidade de funções convexas, entre outros (10). Entre as aplicações mais práticas encontram-se o estudo de sistemas dinâmicos em geral, movimentos brownianos, singularidades em potenciais eletrostáticos e gravitacionais, dinâmica de fluidos e turbulência, geometria de antenas de celulares. Em Economia, aplica-se ao estudo de curvas de preços, taxas de câmbio, entre outros (1). Existem aplicações ainda em sistemas caóticos como os que modelam o padrão de formação de nuvens e fluídos em geral, análise e reconhecimento de padrões em imagens e sinais, análise de texturas (16), caracterização de objetos (3), medição de comprimento de curvas, entre muitos outros.

Os fractais podem ser gerados por replicações ou divisões recorrentes de uma unidade de origem. A Figura 1 ilustra um fractal gerado por uma relação de recorrência no plano complexo, denominado conjunto de Mandelbrodt, cuja fórmula de geração é $Z_{n+1} = Z_n^k + c$, sendo Z a função geradora, n a iteração, k e c constantes reais. Um fato interessante no que diz respeito aos fractais como o de Mandelbrodt é que uma mínima interferência nas condições iniciais de geração do fractal leva a formas totalmente distintas ao final do processo.

Vale ressaltar que na natureza não são conhecidas figuras perfeitamente fractais, mas sim formas que apresentam a característica de auto-semelhança e complexidade alta em maior ou menor grau (2). A impossibilidade de fractais verdadeiros em aplicações práticas se deve basicamente a dois fatores. O primeiro é que a auto-similaridade infinita é impossível, uma vez que os objetos do mundo real apresentam limites na escala de observação,

seja microscópica (raramente se chega à dimensão atômica) ou macroscópica (objetos reais apresentam tamanho finito). O segundo motivo é que a representação digital da imagem do objeto limita o número de detalhes que podem ser processados computacionalmente ao tamanho do pixel.

Esta propriedade de construção a partir de elementos geométricos simples e auto-similaridade passou a ser denominada de fractalidade. Desta forma, esta nova modalidade de geometria associa a cada objeto irregular um número que indica o grau de fractalidade deste objeto. Este número é chamado de dimensão fractal e é um número real, não necessariamente inteiro. Um valor de dimensão não inteiro faz sentido na medida em que se busca o conceito de dimensão da Teoria da Medida (1). Segundo este conceito, a dimensão mede o quanto o comprimento entre dois pontos na entidade geométrica diminui à medida que a escala de observação aumenta. Quando se estende o cálculo da dimensão fractal a formas que não são fractais perfeitos, como será visto adiante, o valor da dimensão fractal pode ser pensado como uma medida da irregularidade (complexidade) ou ainda da ocupação espacial do objeto analisado.

Generalizando-se o cálculo de dimensão euclideana, obtém-se a fórmula genérica expressa abaixo usada para o cálculo da dimensão de qualquer objeto fractal.

$$D = \lim_{u \rightarrow 0} \frac{\ln N}{\ln L/u}, \quad (2.1)$$

em que L é o comprimento da linha do objeto fractal, u é o comprimento de uma linha menor que L e N é o número necessário de objetos de comprimento u para “cobrir” o objeto de comprimento L . O valor da dimensão fractal pode ser calculado de maneira exata sempre que se conhece o processo de formação da entidade geométrica.

Como exemplo de cálculo, tem-se na Figura 2, a curva fractal triádica de Koch. Esta curva é gerada a partir de um segmento de reta, que é dividido em três partes iguais. A parte intermediária é então usada como lado para a construção de um triângulo equilátero. Em seguida, este segmento intermediário é eliminado da figura, sendo substituído então por dois novos segmentos idênticos em escala aos anteriores (os outros dois lados do triângulo). Deste modo, a curva é composta a cada nova iteração sempre por 4 novos segmentos equivalendo a $1/3$ do segmento anterior. Assim, sua dimensão fractal é:

$$D = \lim_{u \rightarrow 0} \frac{\ln N}{\ln L/u} = \lim_{u \rightarrow 0} \frac{\ln 4}{\ln 3/1} = \frac{\ln 4}{\ln 3} \cong 1,26. \quad (2.2)$$

Figura 2 – Curva de Koch. objeto fractal cujo cálculo direto da dimensão é simples.

Isto significa que uma curva de Koch possui uma ocupação espacial maior que uma reta (dimensão 1) e menor que um plano (dimensão 2).

Na prática, este limite é calculado a partir da curva de $\ln(N) \times \ln(L/u)$. A dimensão é calculada como sendo o coeficiente angular da reta que aproxima esta curva.

Esta dimensão, calculada em 2.2, é chamada de dimensão de Hausdorff-Besicovitch e é também conhecida como dimensão “verdadeira”. Ela difere da chamada dimensão topológica d_T , que sempre assume valores inteiros e mede o número de graus de liberdade que um ponto pertencente ao objeto apresenta, assim d_T calculada em um ponto vale 0, em uma curva vale 1, em um plano vale 2, e assim por diante. Vale ressaltar que para estas duas medidas sempre vale a desigualdade $D \geq d_T$. Se um fractal está contido em um espaço topológico de dimensão N , então $0 \leq D \leq N$. No caso de fractais no plano, quanto mais o valor de D se aproxima de d_T , no caso 2, mais a curva fractal preenche o plano no qual ela está contida. Como consequência disto, a medida da dimensão fractal acaba sendo usada para caracterizar o grau de ocupação espacial e complexidade do objeto estudado.

O matemático Benoit Mandelbrot popularizou o conceito de conjunto fractal, como sendo aquele em que a dimensão de Hausdorff-Besicovitch possui um valor não inteiro (daí o nome fractal) e é maior do que o valor da dimensão topológica.

2.2 Métodos de Cálculo

A geometria fractal vem apresentando uma crescente aplicação no cálculo da dimensão de objetos que não são fractais verdadeiros, como ocorre em processamento de imagens. As imagens, na maioria das vezes representativas de objetos da natureza, não são fractais perfeitos, já que possuem auto-similaridade e complexidade apenas em nível finito. Trabalhos como o de (2) questionam o uso do termo “fractal” para este tipo de aplicação,

mas o fato é que a geometria fractal vem sendo usada cada vez mais para se medir a complexidade de imagens (3) e outros objetos do mundo real.

Para este caso em que são analisados objetos “pseudo-fractais”, como imagens, o cálculo direto da dimensão de Hausdorff-Besicovitch torna-se muito complexo ou mesmo inviável e surgem outras definições de medida de dimensão fractal, como a medida Box-Counting (17) e a de Bouligand-Minkowsky (18). Estas novas definições levam a métodos práticos e de fácil implementação para o cálculo. Estes métodos de cálculo da dimensão fractal, no caso, chamada de “não verdadeira”, podem ser divididos em duas categorias: os baseados em Hausdorff-Besikovitch, como Massa Raio (19), Box-Counting (17), *Dividers* (20), entre outros e os baseados em Bouligand-Minkowsky.

2.2.1 Box-Counting

Um método popular e de fácil implementação para o cálculo da dimensão fractal é o de Box-Counting (17). Baseada na medida de mesmo nome, esta técnica é amplamente aplicada ao cálculo da dimensão fractal de imagens.

Se $F \in \mathbb{R}^2$ é um objeto no plano, que pode ser um objeto de interesse em uma imagem, a seguinte regra exponencial é válida:

$$S(r) \propto r^{-d}, \quad (2.3)$$

em que d é a dimensão fractal e $S(r)$ é o número de bolas de raio r necessárias para cobrir o objeto F , sendo que uma bola de raio r é definida como sendo o conjunto de pontos $(x_0, y_0) \in \mathbb{R}^2$ tais que $(x, y) \in \mathbb{R}^2 | \sqrt{((x - x_0)^2 + (y - y_0)^2)} < r$.

Portanto, a dimensão fractal neste caso é fornecida por:

$$D = -\lim_{r \rightarrow 0} \frac{\ln S(r)}{\ln r}. \quad (2.4)$$

Para o cálculo efetivo da dimensão por Box-Counting, a imagem é dividida em caixas quadradas de lado $L \times L$ e calcula-se o número de caixas $N(L)$ contendo pelo menos um ponto do objeto analisado. A caixa inicial é a própria imagem como um todo e a cada passo do método o valor de L é reduzido pela metade. Ao final, analisa-se a curva $\log(N(L)) \times \log(L)$ e o coeficiente angular da reta ajustada à curva é o valor obtido da dimensão fractal.

Figura 3 – Curva do cálculo da dimensão fractal por Box-Counting.

Na prática, o gráfico desta curva de Box-Counting, ilustrado na Figura 3, apresenta três regiões, uma região “euclideana” com dimensão próxima de 1, outra com dimensão “fractal”, ou seja, maior que 1 e outra com dimensão 0. Tal fato se justifica pelos diferentes tamanhos que se usam para as caixas. Quando a caixa possui um tamanho menor, ela contém em seu interior partes do objeto que se assemelham a um segmento de reta (dimensão inteira). À medida que a caixa vai aumentando, detalhes da fractalidade do objeto vão sendo trazidos à tona. Quando finalmente a caixa é a imagem como um todo e todo o objeto está envolvido na caixa, a dimensão do objeto é 0.

Como o método básico é muito simples, a literatura apresenta diversos exemplos de Box-Counting mais elaborados. Um exemplo de variação é tratar separadamente as caixas que são cobertas apenas parcialmente pelo objeto daquelas que são cobertas na íntegra (21). Outra estratégia seria usar um número mínimo de caixas suficientes para cobrir a imagem. Ainda seria possível o uso de círculos, não necessariamente disjuntos, no lugar das caixas, em uma abordagem que se assemelharia mais à de Bouligand-Minkowsky, discutida a seguir.

Um problema no uso de Box-Counting é que pequenas variações na qualidade da imagem, na estrutura da malha de caixas ou no comprimento do lado destas caixas podem interferir drasticamente no resultado final obtido para o valor da dimensão (22).

2.2.2 Bouligand-Minkowsky

O método adotado neste trabalho é o de Bouligand-Minkowsky, por ter se mostrado em trabalhos de pesquisa do grupo como o que apresentou melhor precisão no cálculo da dimensão fractal, bem como a maior consistência nos resultados para o cálculo de objetos fractais verdadeiros, cuja dimensão é conhecida previamente (23)(18).

O cálculo da dimensão fractal de um objeto bidimensional por Bouligand-Minkowsky provê um valor sempre maior ou igual à dimensão de Hausdorff-Besicovitch e envolve a definição do conceito de dilatação de um objeto, também conhecida como área de influência. Sendo $A \in \mathfrak{R}^2$ o objeto do qual se deseja calcular a dimensão, tem-se a dilatação A_r de A dada por

$$A_r = \{x \in \mathfrak{R}^2 \mid \exists y \in A : |x - y| \leq r\}, \quad (2.5)$$

ou ainda por,

$$A_r = \cup_{x \in A} B_r(x), \quad (2.6)$$

em que r é o raio da dilatação e $B_r(x)$ é um disco de raio r . Desta forma, vê-se então que a área de influência é composta pelos pontos cuja distância mínima a qualquer ponto do objeto é de pelo menos r .

A soma do número de pontos pertencentes à área de influência $A(r)$ apresenta uma relação exponencial com o raio da dilatação:

$$A(r) = \mu r^{2-d}, \quad (2.7)$$

em que μ é uma constante real e d é a dimensão fractal.

No cálculo da dimensão fractal então, usa-se uma estratégia denominada “salsicha de Bouligand-Minkowsky”, na qual o objeto em análise é varrido por dilatações com valor variável de r , tendo como centro os pontos do objeto. Os pontos contidos nestes círculos são somados fornecendo a área de influência $A(r)$. A dimensão é fornecida então por:

$$d = 2 - \lim_{r \rightarrow 0} \frac{\ln(A(r))}{\ln(r)}. \quad (2.8)$$

Curvas simples geram valores próximos de 1 para d e curvas mais complexas apresentam valores de d relativamente menores que 2.

Para uma melhora do desempenho computacional do cálculo por esta fórmula, a função $A(r)$ é obtida com o auxílio da Transformada Exata da Distância (TED). Esta transformação associa cada pixel de um objeto em uma imagem à menor distância euclideana entre este pixel e o fundo da imagem (em que se representa o objeto de interesse). Para o cálculo desta transformada duas abordagens são as mais recomendadas (24). A primeira

chamada de Dilatação Exata (25) é mais simples e consiste na criação de uma estrutura que armazena todas as distâncias exatas possíveis no espaço discreto da imagem e os pontos que estão a esta distância do fundo. A outra abordagem é chamada de algoritmo de Saito e Toriwaki (26) e é mais complexa, baseando-se em varreduras independentes de linhas e colunas da imagem. Devido ao custo computacional significativamente mais baixo desta segunda técnica, ela é a usada no cálculo de $A(r)$ neste trabalho.

Com o objetivo de facilitar os cálculos subseqüentes e possibilitar uma visualização da dimensão de forma mais intuitiva, aplica-se a estratégia de usar a curva de $A(r) \times r$ plotada em escala log-log. Neste gráfico observa-se um comportamento quase linear da curva de $A(r)$. O coeficiente angular da reta ajustada à curva em log-log é tomado como o valor da dimensão.

2.3 Dimensão Fractal Multiescala

Um problema que surge quando se calcula a dimensão fractal é que a maioria das aplicações é voltada para objetos que não são fractais verdadeiros. Isto ocorre, por exemplo, com objetos da natureza representados por fractais, como nuvens, folhas ou, de um modo geral, relacionados a experimentos, tais como imagens histológicas, geológicas, médicas, entre outros. Ocorre que estes objetos não possuem complexidade infinita, não sendo desta forma infinitamente auto-similares. O que se observa com estes objetos é que sua fractalidade tende a zero à medida que a escala de observação usada vai sendo aumentada.

Quando se analisa a curva $A(r)$, que agora passa a ser tratada genericamente como a função $u(t)$, do método de Bouligand-Minkowsky na escala logarítmica, nota-se uma grande riqueza de detalhes quando se tem um olhar sobre a curva em diferentes escalas de observação. Para que estes detalhes sejam capturados é que se propõe uma transformação multiescala $U(b, a)$ gerando a curva de Dimensão Fractal Multiescala (DFM) (27)(28)(18). O parâmetro b está associado à variável t e a está associado à escala de observação. Esta transformação é denominada de escala-espacial ou tempo-escala. Desta forma, extrai-se uma informação muito mais completa acerca do objeto do que apenas um número (dimensão) que não expressa as nuances da dimensão fractal em cada ponto da curva u (29).

Vale ressaltar também que o parâmetro a está relacionado de forma simples ao domínio da frequência f , isto é, $a \cong 1/f$, o que faz com que esta transformação seja dual à

transformação tempo-frequência. Esta transformação bidimensional de um objeto unidimensional introduz redundâncias significativas, mas isto acaba sendo necessário para que os detalhes não detectados no objeto original possam ser enfatizados. Para que estas redundâncias possam ser contornadas, o processo mais comum na análise multiescala é a transformação seguida da extração das características que mais interessam para a aplicação.

As abordagens mais comuns para o caso da transformação multiescala são transformação escala-espço, tempo-frequência (filtros de Gabor) e tempo-escala (*wavelets*). A abordagem escolhida neste trabalho, por ser a que apresenta melhores resultados na literatura (30)(31), é a espaço-escala, descrita na seção seguinte.

2.3.1 Transformação espaço-escala

Desenvolvido em (31), esta abordagem vem se tornando cada vez mais popular em aplicações deste tipo de transformação (32).

Na prática, a curva de DFM calculada nesta abordagem é obtida por:

$$DFM = N - \frac{du(t)}{dt} \quad (2.9)$$

em que N é a dimensão do espaço euclidiano em que está imerso o objeto. A Figura 4 ilustra as fases do processo de obtenção da curva DFM de imagens de letras do alfabeto, usando-se a notação de $A(r)$ para a função transformada.

O uso da derivada neste tipo de transformação está associado ao fato de que as informações mais significativas de um sinal estão diretamente relacionadas ou podem ser calculadas a partir da determinação dos pontos de máximo e mínimo locais destes sinais (31). E estes pontos de extremos locais são obtidos a partir dos cruzamentos da derivada do sinal com o eixo das abscissas. Na maior parte das aplicações práticas, a derivada analítica do sinal não é conhecida, fazendo-se necessário o uso de técnicas de diferenciação numérica. Ocorre que as técnicas de derivada se baseiam, para o cálculo, nas vizinhanças de cada ponto analisado, o que leva à idéia de escala de observação de cada ponto.

O uso de derivação numérica, entretanto, traz consigo um problema significativo. O cálculo da primeira derivada costuma enfatizar ruídos do sinal original, sobretudo os de altas frequências. Para minimizar este problema é necessária a aplicação de filtros passa-

Figura 4 – Geração da curva DFM. (a) Imagem original. (b) Curva de $A(r)$ gerada por Bouligand-Minkowsky. (c) Curva DFM.

baixas. Um dos mais simples e eficientes filtros desta espécie é o da gaussiana, usado neste trabalho.

A gaussiana é definida por:

$$g_a(t) = \frac{1}{a\sqrt{2\pi}} \exp\left(\frac{-t^2}{2a^2}\right), \quad (2.10)$$

em que $a > 0$ é um valor real chamado de desvio padrão, que controla o grau de suavização do filtro e está diretamente associado à escala: quanto maior a , maior a escala de observação.

A derivada $u'(t)$ pode ser calculada por meio da propriedade da convolução com a gaussiana:

$$u'(t) * g_a(t) = (u(t) * g_a(t))' = u(t) * g'_a(t), \quad (2.11)$$

em que $*$ representa a operação de convolução. Isto significa que a derivada de $u(t)$ pode ser obtida pela simples convolução com a derivada primeira da gaussiana. Desta forma a transformação espaço-escala pode ser representada por:

$$U'(t, a) = u(t) * g'_a(t). \quad (2.12)$$

Os extremos do sinal podem ser obtidos então pelos cruzamentos por zero do resultado da convolução. O parâmetro a da transformada controla a “largura” do filtro, determinando assim a escala de atuação do filtro. Quanto menor o valor de a menor o grau de suavidade aplicado ao sinal e desta forma maiores detalhes de escala são exibidos. A referência (31) generalizou o conceito do uso de diferentes valores para a e passou a tratar esta variável como sendo contínua, podendo assumir quaisquer valores reais em um determinado intervalo. Assim, para cada valor de a , um conjunto de extremos é encontrado e a evolução deste conjunto com a compõe a transformada espaço-escala de $u(t)$.

Desta forma, a transformação espaço-escala pode ser definida, de maneira mais formal, como sendo o seguinte conjunto:

$$(b_0, a_0) | a_0, b_0 \in \mathfrak{R}, a_0 > 0, b_0 \in U'(t, a_0)_{zeros}, \quad (2.13)$$

em que $U'(t, a_0)_{zeros}$ é conjunto de cruzamentos por zero de $U'(t, a_0)$ e $U(t, a_0)$ é a versão suavizada de $u(t)$ em escala a_0 .

A transformação espaço-escala $U(t, a)$ ainda pode ser definida em termos de uma série de convoluções com a gaussiana. Em geral, no limite, o espaço-escala $U^n(t, a)$ pode ser obtido pelos cruzamentos por zero de $g_a^{n+1}(t)$. O trabalho (33) propõe o cálculo de U como sendo a solução da equação de difusão unidimensional:

$$\frac{\partial U}{\partial a} = k \frac{\partial^2 U}{\partial t^2}, \quad (2.14)$$

sendo k uma constante real e $U(t, 0) = u(t)$ a condição inicial.

2.3.2 Métodos de Diferenciação Numérica

A derivada empregada neste método pode ser calculada geralmente de duas formas: por diferenciação numérica (diferenças finitas), ou por diferenciação de Fourier (no espectro). Testes foram realizados em (23) visando à comparação do resultado da derivada por Fourier e numérica com o valor esperado para curvas com derivada analítica conhecida. Os resultados comprovaram que a diferenciação por Fourier apresenta os resultados mais precisos. Tal resultado se deve ao fato de que a derivada de Fourier é calculada a partir

de uma análise global do sinal, enquanto a derivada numérica utiliza apenas os pontos de vizinhança local para o cálculo. A partir deste fato optou-se pelo uso da diferenciação por Fourier para o cálculo da derivada neste trabalho.

A transformação multiescala por Fourier compreende, em uma primeira fase, o cálculo da derivada da função da área acumulada pelo método de Bouligand-Minkowsky. Esta derivada é calculada com base na propriedade derivativa de Fourier e é obtida por:

$$\frac{dv(r')}{dr'} = F^{-1}(Fv(r')G_a(f)(j2\pi f)), \quad (2.15)$$

em que $v(r')$ e r' são, respectivamente a curva obtida por Bouligand-Minkowsky e o raio, ambos em escala log-log, F e F^{-1} são a transformada de Fourier e sua inversa, f é a frequência e $j = \sqrt{-1}$ é o número imaginário, G_a é a transformada de Fourier da gaussiana g_a .

Para a implementação computacional da técnica, na qual são analisados sinais discretos, a transformada usada deve ser a discreta e, mais precisamente, a transformada discreta rápida (cuja sigla em inglês é FFT). Supondo-se que o sinal analisado $u(t)$ apresenta $t \in [0, V]$ e uma amostragem em N pontos. O intervalo de amostragem será dado então por $T = V/(N - 1)$.

A transformada pode ser obtida então por

$$U(s) = \sum_{n=0}^{N-1} u(n) \exp(-jVns/N), s \in [0, N - 1]. \quad (2.16)$$

Como qualquer transformada de Fourier, FFT gera um sinal periódico, de período N . Após a aplicação de FFT ao sinal original, o sinal obtido apresenta, na primeira parte do domínio, valores negativos e, na segunda parte, valores positivos. Para uma melhor distribuição, este sinal é tomado em módulo e uma operação de deslocamento (*shift* em inglês) é executada, de modo a trazer o conteúdo mais importante para a região central do sinal. Após estas etapas, o intervalo do domínio s deve ser ajustado para $[-\text{floor}(N/2), \text{floor}(N/2) - 1]$. Posteriormente o sinal deve ser estudado em um novo domínio, o espaço de frequências contínuo, definido por $f = s/NT$. A derivada é fornecida então pela multiplicação de $U(f)$ por jVf , seguida pela transformação de Fourier inversa.

Para o uso deste método de diferenciação, o sinal deve passar inicialmente por um pré-tratamento que garantirá uma amostragem uniforme e suficientemente grande. Para

Figura 5 – Fenômeno de Gibbs. (a) Sinal original. (b) Derivada do sinal em (a) calculada pelo método derivativo de Fourier, sem tratamento especial. (c) Curva em (a) replicada horizontalmente. (d) Derivada convencional de Fourier do sinal em (c). (e) Parte da curva em (d) interessante para o problema. (f) Reflexão horizontal da curva em (e) em volta de seu ponto médio cuja translação e escala geram a derivada de (a). (g) Comparação entre a derivada de Fourier tratada e a derivada analítica.

que se garanta uma amostragem significativa ao longo de todo o sinal, os primeiros pontos deste sinal devem ser descartados, já que na escala logarítmica estes pontos encontram-se afastados uns dos outros. Já para a uniformidade, pode-se usar interpolação por média aritmética entre os raios válidos na curva original.

Ainda um outro problema que ocorre com os métodos de diferenciação numérica como o caso de Fourier é que estes métodos não convergem nas extremidades do sinal, devido à não uniformidade da convergência de Fourier nestes pontos de descontinuidade. Este fenômeno é chamado de fenômeno de Gibbs e é resolvido pela aplicação de um processo de replicação e reflexão do sinal original em suas extremidades, como se vê na Figura 5, com o fim de garantir a continuidade do sinal no intervalo $[2N-1, 3N]$ e se evitar o fenômeno de Gibbs (9).

Figura 6 – Multiescala usada na geração de assinaturas. (a) Quádrica de Koch. (b) Gráfico log-log da área de influência obtida pelo método de Minkowsky em função do raio.

2.3.3 Aplicação à Geração de Assinaturas

A transformada multiescala pode ser representada como uma função de $\mathbb{R}^2$ em $\mathbb{R}$ ou $\mathbb{C}$, dependendo do tipo de transformação usada e do sinal original:

$$U(b, a) : \mathbb{R}^2 \rightarrow \mathbb{R} \quad \text{ou} \quad U(b, a) : \mathbb{R}^2 \rightarrow \mathbb{C}. \quad (2.17)$$

Para a transformação espaço-escala, se o sinal original é real, o sinal resultante é real, se o original é complexo, o resultante é complexo.

O conceito de transformação multiescala pode ser aplicado a partir do estudo da curva log-log obtida por qualquer um dos métodos de dimensão fractal conhecidos. Neste trabalho, usa-se o método da salsicha de Minkowsky por Saito e Toriwaki (26). A Figura 6 seguinte ilustra a imagem de um objeto fractal, seguido pelo gráfico log-log da área de influência em função do raio de dilatação, usando-se um valor máximo de 100 pontos para este raio.

Logo em seguida à esta primeira etapa, a curva deste gráfico log-log deve ter seus pontos iniciais eliminados devido à baixa amostragem e o gráfico resultante deve ser interpolado, com o objetivo de permitir que uma amostragem igualmente espaçada possa ser extraída a partir da curva da Figura 6. A Figura 7 mostra a curva resultante.

Em seguida, o método de diferenciação por Fourier, com suavização pela gaussiana é aplicado, gerando a curva DFM, como mostrado na Figura 8.

Figura 7 – Curva obtida a partir do gráfico log-log por eliminação dos pontos de baixa amostragem e interpolação por raios médios.

Figura 8 – Curva DFM obtida.

A curva DFM traz uma riqueza significativa para a representação da forma analisada, uma vez que representa não mais um número, mas uma função que exhibe o grau de fractalidade, ou seja, de ocupação espacial do objeto para cada escala de observação. Tal fato, torna a DFM uma ferramenta poderosa no processamento de imagens digitais, sobretudo quando associada a outras técnicas.

Entre estas aplicações pode-se citar o cálculo do comprimento de curvas irregulares a partir de suas imagens digitais (1). Outra aplicação é a segmentação de imagens, uma vez que o índice de ocupação espacial pode ser usado para a rotulação de partes de uma imagem (30). Ainda uma versão tridimensional de DFM pode ser obtida sem muito esforço e usada na análise de complexidade de texturas (16).

A aplicação focada neste trabalho entretanto será a classificação de objetos representados em imagens digitais por meio de uma transformada DFM. A transformada gera um vetor de valores (características), chamado de assinatura da imagem, que pode ser usado como entrada para métodos de classificação conhecidos da literatura. Para que melhores resultados possam ser alcançados, ferramentas adicionais como descritores de Fourier, Wavelet e ADF são usadas. A Figura 9 ilustra o uso de DFM na classificação de objetos. Ainda a Figura 10 ilustra o uso da diferença entre a curva DFM e a portadora destas curvas (no caso tomada como sendo a média de todas as curvas DFM para todas as amostras). A portadora guarda toda informação redundante e métodos de classificação como os usados aqui extraem esta portadora automaticamente podendo se usar assim a assinatura DFM de maneira direta como entrada para estes métodos. Observa-se que na Figura 10 dois contornos de diferentes imagens de peixes são bem discriminados quando se elimina a portadora, mesmo neste caso em que os valores da dimensão fractal (DF) são idênticos para as duas imagens.

2.4 Dimensão Fractal Multiescala Aplicada ao Reconhecimento de Padrões de Textura

A técnica de DFM pode ser aplicada de maneira semelhante ao reconhecimento de objetos em imagens em tons de cinza apresentando determinado padrão de textura.

Neste caso, a imagem é mapeada para uma superfície em 3 dimensões, sendo que cada ponto de coordenada (x, y) da imagem é mapeada para um ponto de coordenada (x, y, z) , sendo z a intensidade do pixel de coordenada (x, y) . Mais detalhes podem ser encontrados em (30).

Figura 9 – DFM usada na classificação de imagens de letras do alfabeto. Nota-se que imagens de classes diferentes produzem assinaturas distintas.

Figura 10 – Assinatura DFM menos a assinatura portadora. Nota-se que objetos distintos mas apresentando valores similares para a dimensão fractal (DF) podem ser discriminadas por esta curva DFM subtraída da portadora.

Em seguida, uma técnica de cálculo de dimensão fractal em 3 dimensões é aplicada. Neste trabalho, usou-se a técnica de Bouligand-Minkowsky. Neste caso, cada ponto(pixel) da superfície é expandido por esferas de raio r , formando volumes de influência $V(r)$. Em seguida, a função $V(r)$ pode ser processada da mesma forma que $A(r)$ gerando-se assim uma assinatura multiescala da imagem original.

3 *Análise de Dados Funcionais*

Neste capítulo, faz-se um estudo da técnica de Análise de Dados Funcionais, tanto do ponto de vista teórico quanto de suas possíveis aplicações.

3.1 O que é um Dado Funcional: Definições e Conceitos Matemáticos

A análise estatística tradicional se baseia, na maioria das vezes, no uso de uma parte de uma população como amostra de observação para que possam ser obtidas informações acerca da população toda (34). Ocorre que em um número crescente de aplicações esta amostra de observação é composta por um número alto de valores que podem ser representados por curvas (como as séries temporais) ou imagens, ou ambos (5)(6). Estes dados podem ser representados como funções discretas, por exemplo, nas séries temporais e em dados espectrométricos (8). De fato, as curvas e imagens obtidas nestes tipos de aplicações são funções, uma vez que apresentam apenas um valor em cada ponto analisado. A Análise de Dados Funcionais (ADF) é o termo cunhado em (5) para o conjunto de ferramentas usadas para se fazer análise deste tipo de dado apresentado sob a forma de função. Como qualquer ferramenta de análise estatística, o foco de ADF está em representar o dado analisado de forma a facilitar manipulações posteriores e exibir características importantes do dado, além de permitir a análise de padrões e variações na função de representação.

Qualquer conjunto de medidas de uma determinada variável, quando apresenta comportamento suave, pode ser considerado e tratado como uma função ou mais especificamente neste caso como um dado funcional. Como exemplo de tal situação tem-se a Figura 11 que ilustra a medição da estatura de dez pessoas tomada em diferentes idades. Um fato interessante que já pode ser notado em um primeiro momento é que a medida da altura não é tomada em intervalos de idade igualmente espaçados. De fato, a altura é medida trimestralmente até 1 ano, anualmente de 2 a 8 anos e bienalmente daí em diante. Nota-se

Figura 11 — Curva de amostra da altura de 10 pessoas em função da idade.

que a medida apresenta um caráter suave intrínseco e esta suavidade pode ser enfatizada na medida em que se aumenta a frequência da medição. Com isto tem-se na prática que o conjunto de valores medidos da estatura pode ser considerado um dado funcional.

Diversas informações acerca do dado que gerou a medida podem ser obtidas diretamente a partir da análise do dado funcional como será visto mais adiante. Entretanto, em diversas situações, lança-se mão de alguma operação sobre a função obtida. Por exemplo, a Figura 12 mostra o uso da primeira derivada (velocidade na analogia com a Física) da função cujo gráfico está na Figura 11. Na Figura 13 vê-se o gráfico da segunda derivada (aceleração) da mesma função. Um fato notável é que o uso da derivada segunda permite que seja observado um rápido crescimento seguido de um rápido decrescimento na amplitude da curva por volta dos 12 anos (puberdade) e o mesmo comportamento se repete por volta dos 6 anos, fase também importante no crescimento da criança. Este exemplo ilustra a importância que o uso de derivadas, integrais e outras operações funcionais podem representar na identificação de padrões. E aí já se vê um motivo importante para o uso de funções em vez de simples conjuntos de valores. Ainda outros estudos interessantes da amostra podem ser efetuados a partir da função como a análise de correlação e de variação (5).

3.1.1 Anatomia de um Dado Funcional e Dimensionalidade

Várias informações podem ser extraídas a partir do gráfico de funções como o da Figura 11. Entre estas informações pode se citar a localização dos picos e vales, os valores (chamados de níveis) da variável independente t para os quais o gráfico cruza o eixo

Figura 12 – Curva da derivada primeira da função de altura.

Figura 13 – Curva da derivada segunda da função de altura.

das abcissas, entre outros. Estas informações são efetivamente eventos que podem ser observados ao longo de t . Estes eventos podem ser medidos por sua localização, como no caso dos cruzamentos no eixo t , mas também por sua amplitude ou até mesmo largura, como ocorre no caso de picos e vales observados.

Dado o conceito informal de dimensão como sendo o número de parâmetros necessários para descrever completamente uma entidade algébrica, os níveis podem ser considerados como sendo unidimensionais, os cruzamentos como bidimensionais e picos e vales como tridimensionais. A presença de erros amostrais no problema faz com que o número de parâmetros necessários (dimensão) seja aumentado.

A partir deste conceito informal, surge o conceito de resolução de um conjunto de dados, medida inversamente proporcional à menor largura possível para qualquer evento observado na função. Em diversas situações a resolução acaba tendo uma importância maior do que o número de valores da função. O conceito de resolução leva também ao conceito de dimensionalidade, que pode ser definido sem um maior rigor matemático como sendo o número necessário de valores para definir completamente a função. Rigorosamente falando, as funções tratadas em ADF apresentam dimensão infinita, uma vez que *a priori* os valores não podem ser definidos previamente e assim é necessário que todos os valores da função, para cada valor de t , sejam conhecidos, gerando assim um conjunto infinito de valores. Na prática, porém, o uso de um conjunto limitado de valores é suficiente para caracterizar toda a função com um erro que pode ser desprezado. Assim, para cada situação tem-se um valor ideal para a dimensão de cada função analisada.

3.1.2 Gráfico Fase-plano

O gráfico fase-plano, como descrito no trabalho (6), é a representação da segunda derivada desenhada contra a primeira derivada. Em diversos problemas estudados em ADF nos quais a medida de transferência de energia apresenta um papel importante, este gráfico é largamente usado. Este é o caso, por exemplo, do consumo de bens não duráveis, índice usado como medida do desempenho da economia.

A Figura 14 mostra o gráfico do consumo de bens não-duráveis nos Estados Unidos no século XX. Na Figura 15 tem-se o gráfico da função composta por bases *B-splines* ajustada e na Figura 16 é desenhado o gráfico fase-plano do problema. Nota-se que visualmente o gráfico é uma série de elipses com raios variáveis sobrepostos. As informações mais importantes que podem ser extraídas deste tipo de gráfico são:

Figura 14 — Curva do índice mensal de bens não-duráveis nos Estados Unidos ao longo de anos.

Figura 15 — Curva de uma função B-spline suave ajustada ao logaritmo decimal do índice de bens não-duráveis.

Figura 16 – Gráfico fase-plano do índice de bens não-duráveis. As letras na curva representam os meses do ano.

- Uma elipse principal, com o maior raio
- O raio desta elipse: quanto maior este raio maior a transferência de energia envolvida no processo
- A localização horizontal do centro: se à direita, a velocidade de crescimento é positiva e negativa caso contrário
- A localização vertical do centro: se acima de zero, a velocidade de crescimento está aumentando e diminuindo caso contrário
- Mudanças no formato das elipses no decorrer dos anos

3.2 Representação de Conjuntos de Valores por funções

Além de um conjunto de métodos simplesmente, ADF é também uma nova forma de se pensar em estatística na qual se enxerga não um conjunto de valores, mas todo este conjunto como uma única entidade: a função que gerou deu origem àqueles valores, chamada de dado funcional. A técnica apresenta quatro objetivos básicos que permitem sua aplicação em diversas situações. O primeiro deles é possibilitar a manipulação posterior diretamente sobre a função por métodos já conhecidos. O segundo é destacar, por meio da representação gráfica desta função, aspectos expressivos do dado em análise. O terceiro objetivo é estudar a presença de padrões de comportamento ou de variabilidade dentro do conjunto de dados original. O quarto é explicar o comportamento de uma variável de saída a partir dos valores presentes no conjunto de dados de entrada.

As funções manipuladas por ADF são assumidas como sendo suaves (derivada(s) contínua(s)). Para os casos em que o dado a ser analisado não permite um controle preciso no processo de geração ou está afetado por uma grande quantidade de erros, é necessário que este dado seja submetido a um método de suavização antes que o método de ADF em si seja aplicado.

Para que o dado funcional possa ser tratado computacionalmente, como é o objetivo no presente estudo, faz-se necessário o uso de uma representação adequada (discreta) para o computador. A principal técnica usada para este fim é o uso da combinação linear de um conjunto de funções-bases previamente escolhidas. Entre as vantagens que esta abordagem oferece pode-se citar a flexibilidade de se representarem dados apresentando os mais diversos padrões de variação, a facilidade de representação computacional e uma maior facilidade de cálculos que podem ser executados usando-se ferramentas da álgebra de matrizes. Entre as bases de funções que costumam ser usadas em ADF destacam-se as de Fourier, para conjuntos de valores que apresentam comportamento cíclico e as *B-splines* para as demais situações.

O dado funcional é armazenado inicialmente como sendo um conjunto de n pares (t_j, y_j) , sendo y_j o valor da observação realizada em t_j . A variável t_j é qualquer variável contínua em relação à qual as observações são tomadas, podendo ser tempo, frequência, peso, etc.

Quando se diz que uma observação é um funcional quer se dizer que existe uma função suave que gera os valores daquela observação. Desta forma, valores de y adjacentes muito dificilmente vão apresentar valores muito discrepantes entre si. A presença desta suavidade é um forte indicativo para o uso da técnica de ADF em contraposição à análise multivariada clássica. Na prática, a suavidade desta função pode ser medida pela existência pelo menos da derivada primeira da função. As funções do mundo real podem apresentar também um determinado nível de ruídos que torna a obtenção de derivadas um desafio.

Em muitas situações surge um conjunto de dados funcionais, em vez de apenas uma função. Ou seja, cada função x_i consiste de n_i pares (t_{ij}, y_{ij}) , $j = 1, \dots, n_i$. O argumento t_{ij} e o intervalo de amostragem pode variar de função para função. O funcional também pode possuir um domínio multidimensional.

Muito embora a função deva apresentar caráter suave, isto não é uma condição necessária para a observação original $y = (y_1, \dots, y_n)$, uma vez que esta é passível de ruídos intrínsecos à observação. De fato, pode-se escrever:

$$y_j = x(t_j) + \epsilon_j, \quad (3.1)$$

em que x é a função suave e ϵ pode ser chamado de erro de observação, perturbação, ruído, etc. Duas abordagens podem ser usadas para uma representação eficiente do dado original por uma função suave. Na primeira, faz-se uma filtragem de modo a eliminar o máximo possível de ϵ . Na segunda, y_j (incluindo ϵ) é suavizado. Para simplificar usa-se a notação vetorial:

$$\vec{y} = x(\vec{t}) + \vec{e}, \quad (3.2)$$

em que $\vec{y}$, $x(\vec{t})$ e $\vec{e}$ são vetores-coluna de comprimento n . Nesta notação tem-se $\sum_{\vec{e}}$ representando a matriz de covariância do erro $\vec{e}$ e, uma vez que $x(\vec{t}_j)$ são fixos com variância 0, esta matriz pode ser tomada como medida estatística de $\vec{y}$.

Por padrão, assume-se que ϵ_j segue um modelo de distribuição estatística normal com média 0 e variância σ^2 , ou seja:

$$Var(y) = \sum_{\vec{e}} = \sigma^2 I, \quad (3.3)$$

sendo I a matriz identidade de ordem n . Vale destacar que a distribuição normal não é perfeitamente adequada para representar o ruído na maior parte das observações do mundo real. A variância de ϵ_j pode depender de t . Além disso, existe uma correlação entre valores vizinhos de ϵ_j , a chamada auto-correlação. Entretanto, a simplicidade e o baixo custo computacional envolvido fazem da distribuição normal um modelo interessante de aproximação em muitas situações.

Outra questão importante é a resolução do dado, mais precisamente a densidade de valores de t_j em relação à curvatura de y_j , curvatura esta que pode ser medida pela intensidade da segunda derivada $|D^2x(t)|$. Nos pontos em que a curvatura é mais alta uma maior densidade de t_j é recomendada. O número exato de pontos necessários depende da quantidade de erro ϵ_j associada. Quanto menor o erro, menor a densidade exigida.

Como se descreveu anteriormente, as funções são representadas por combinações lineares de funções-base. Estas são funções conhecidas ϕ_k , independentes entre si e satisfazendo a propriedade de que uma combinação linear destas funções é capaz de representar arbitrariamente bem a função original. As funções originais são representadas então por:

$$x(t) = \sum_{k=1}^K c_k \phi_k(t), \quad (3.4)$$

sendo K o número de funções-base usadas na representação e c_k , números reais chamados de coeficientes da representação. Usando-se a notação vetorial, sendo $\vec{c}$ o vetor de coeficientes e $\vec{\phi}$ o vetor de funções-base, tem-se:

$$\vec{x} = \vec{c}' \vec{\phi} = \vec{\phi}' \vec{c} \quad (3.5)$$

A equação acima faz um nexso entre o mundo das funções analíticas de dimensão infinita e o mundo das dimensões finitas dos vetores de coeficientes, de dimensão K . Uma interpolação exata de y é obtida quando $K = n$ e vale exatamente $y(t_j) = x(t_j)$ para todo j . Entretanto, o real objetivo em ADF é a obtenção de uma função suave e com determinadas características semelhantes às encontradas na função original, não sendo necessária a equivalência exata entre os valores da função. Isto faz com que, na prática, um número relativamente pequeno de funções-base seja usado. O uso de um número reduzido de bases apresenta certas vantagens como um número maior de graus de liberdade para testar hipóteses e calcular intervalos de confiança, um menor custo computacional e uma maior probabilidade de os coeficientes virem a se tornar interessantes descritores dos dados originais. O número K de bases a serem usadas determina ainda o grau de suavidade da função obtida (quanto maior o valor de K , mais suave a função obtida).

Sendo V um subconjunto compacto de $\mathbb{R}^p$, p inteiro, a técnica de ADF presume que dado um conjunto de observações (t_j, y_j) , $t_j \in V$, $y_j \in \mathbb{R}$, sempre pode-se encontrar uma função $x \in L^2(V)$, tal que $y_j = x(t_j) + \epsilon_j$, em que ϵ_j é o ruído inerente à observação. Vale lembrar que $L^2(V)$ é o espaço das funções de quadrado integrável cujo domínio é um subconjunto compacto V de $\mathbb{R}^p$, sendo p um inteiro, e a imagem é $\mathbb{R}$. Há que se notar que a função x não é conhecida previamente. Assim se faz necessário o uso de aproximações para a função x e uma forma possível de se obter esta aproximação é por meio da representação de x em um subespaço qualquer de funções-base.

Entre as vantagens desta representação por funções-base pode-se citar a possibilidade de se manipular mais facilmente conjuntos de valores com amostragem irregular ou faltando valores. Outra vantagem é a possibilidade de se aplicarem operações como integração e derivada a estas funções. Ainda uma outra vantagem é o fato de que várias estratégias de análise por ADF usam apenas o vetor de coeficientes, o que simplifica o processamento computacional envolvido. Teoricamente, não há nenhum tipo de restrição

em relação ao tipo de base a ser usado, sendo mais comum o uso de Fourier, exponencial, polinomial, *wavelets*, logarítmica e *spline*. A escolha fica por conta do comportamento do dado original. Por exemplo, é conveniente que se represente um vetor de valores com comportamento periódico por Fourier. Nas subseções seguintes serão analisados dois dos sistemas de bases mais utilizados em aplicações práticas: Fourier e *B-splines*.

3.2.1 Bases de Fourier

O conjunto de funções-base de Fourier, amplamente usado na análise de dados de comportamento periódico, tem sua origem na série de Fourier:

$$x = \phi_0(t) + \phi_1(t) + \phi_2(t) + \dots = c_0 + c_1 \sin(\omega t) + c_2 \cos(\omega t) + c_3 \sin(\omega t) + c_4 \cos(\omega t) + \dots \quad (3.6)$$

em que $\phi_0(t) = 1$, $\phi_{2r-1}(t) = \sin(r\omega t)$, $\phi_{2r}(t) = \cos(r\omega t)$ e ω está relacionado ao período, dado por $2\pi/\omega$. Se for tomado um intervalo qualquer em que t_j é igualmente espaçado e o período é o comprimento deste intervalo. Neste intervalo, a base é ortogonal, ou seja, o produto envolvendo a matriz de bases $\Phi'\Phi$ é uma matriz diagonal.

As bases de Fourier são mais eficientemente aplicadas na representação de funções que apresentam um comportamento periódico intrínseco e que podem ser consideradas estáveis, ou seja, que não apresentam características locais muito sobressalentes, como descontinuidades, nem variações na ordem da curvatura.

3.2.2 Sistema de Funções-base tipo *Splines*

As *splines* são as funções mais usadas como base na representação de dados não periódicos. Esta modalidade de base combina o baixo custo computacional das funções polinomiais com a flexibilidade de se adaptar aos mais diversos padrões de função, mesmo quando se usa um número baixo de bases.

O processo de aproximação de uma função por *spline* inicia-se com a divisão do intervalo de domínio da função original em L sub-intervalos separados pelos pontos $\tau_l, l = 1, \dots, L - 1$ denominados de pontos de junção. Em cada sub-intervalo, a função *spline* é definida por um polinômio de ordem m . Vale ressaltar que a ordem de um polinômio é o número de coeficientes necessários para defini-lo, valendo uma unidade a mais que seu grau, que é o valor da potência mais alta presente no polinômio. A Figura 17 ilustra o

Figura 17 – Bases splines aplicadas à representação de funções analíticas suaves. (a) Funções originais. (b) Funções-base splines. (c) Funções representadas por splines.

processo de uso de uma função do tipo *spline* de ordem baixa.

A transição entre dois polinômios de ordem maior que 1 em um ponto de junção deve ser suave, isto é, os valores dos dois polinômios, bem como de suas derivadas até ordem $m-2$, devem coincidir neste ponto. Outra informação importante acerca da *spline* aproximadora é o seu número de graus de liberdade. Facilmente pode ser verificado que o número de graus de liberdade é igual à ordem dos polinômios usados mais o número de pontos interiores de junção (não incluindo os extremos).

Os pontos de junção podem ser igualmente espaçados, como também pode se desejar uma maior densidade de pontos onde existe um padrão de variação mais significativo.

O sistema de funções-base baseado em *splines* é composto por funções $\phi_k(t)$ satisfazendo os seguintes quesitos:

- Cada função é uma *spline* definida por uma ordem m e uma seqüência τ de pontos de junção;
- Qualquer combinação linear destas funções é também uma *spline*;
- Qualquer *spline* pode ser definida como uma combinação linear destas funções-base.

O sistema de funções mais popular satisfazendo estes requisitos é chamado de *B-splines* e foi desenvolvido por De Boor (35), possuindo pacotes implementados em *R*, *S-PLUS*®, *MATLAB*® (36)(37), entre outros ambientes de programação. Uma função *B-spline* de ordem m , além de ser uma *spline*, possui suporte compacto, isto é, é positiva apenas ao longo de m intervalos adjacentes.

Se uma determinada função é representada em um espaço de K funções-base *B-splines* de ordem m , a matriz de produtos internos entre as bases apresenta apenas $m - 1$ sub-diagonais abaixo e acima da diagonal principal contendo elementos não-nulos. Isto implica que o cálculo de funções B-spline pode ser feito de tal forma que o esforço computacional seja proporcional a K , equivalendo-se a sistemas de bases como Fourier.

Os valores pontuais para a base B-spline costumam ser denotados por $B_k(t, \tau)$, em que k corresponde ao maior ponto de junção em t ou à direita e τ é o conjunto de pontos de junção. No caso em que todos os pontos de junção interiores são discretos, existem $m + L - 1$ funções-base e a função original $S(t)$ pode ser denotada por:

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau). \quad (3.7)$$

Um fato interessante é que o aumento do número K de bases não implica em uma melhor aproximação por *B-splines*. Este fato se justifica porque quando se mantém a ordem fixa, o conjunto de K bases não necessariamente está contido no conjunto de $K + 1$ bases, fato este que é observado quando se aumenta o número de pontos de junção além de K ou quando se aumenta a ordem das bases, mantendo τ .

Ainda um fator de grande importância na representação de funções-base é a escolha da variável de domínio t . As bases aqui usadas, de Fourier e *B-splines*, são normalizadas, ou seja, apresentam valores limitados: entre 0 e 1 para as *B-splines* e entre -1 e 1 para Fourier. Assim, se um número muito grande de bases é usado em um intervalo pequeno, as suas derivadas tendem a apresentar valores extremamente altos. Isto gera problemas sérios sobretudo nos casos em que se necessita manipular derivadas de ordens mais altas realizando-se operações algébricas entre elas. Para contornar esta situação, recomenda-se o uso de intervalos para a variável t com valores próximos do período de oscilação da base usada que apresenta maior variação de amplitude.

3.3 Suavização de Dados Funcionais por Mínimos Quadrados

Nesta subseção, será abordado o método de mínimos quadrados, uma das técnicas mais comumente usadas em ADF, no ajuste de funções-base ao dado originalmente apresentado sob a forma de pares de valores.

O problema consiste essencialmente em se aproximar um conjunto de valores observados $y_j, j = 1, \dots, n$ por um modelo $y_j = x(t_j) + \epsilon_j$. A função $x(t)$ é definida por uma combinação linear de funções-base $\phi(t)$, de tal forma que:

$$x(t) = \sum_k^K c_k \phi_k(t) = \vec{c}' \vec{\phi}, \quad (3.8)$$

em que $\vec{c}$ é o vetor de coeficientes c_k , $\vec{\phi}$ é o vetor de bases e K é o número de bases e conseqüentemente de coeficientes.

Sabe-se da teoria de mínimos quadrados, desenvolvida resumidamente em (5), que uma forma funcional simples pode ser obtida pela expansão usando-se coeficientes obtidos pela minimização de:

$$SMSSE(y|c) = \sum_{j=1}^n [y_j - \sum_k^K c_k \phi_k(t_j)]^2. \quad (3.9)$$

Ou ainda, em linguagem matricial:

$$SMSSE(\vec{y}|\vec{c}) = (\vec{y} - \Phi \vec{c})'(\vec{y} - \Phi \vec{c}), \quad (3.10)$$

ou de forma ainda mais compacta:

$$\|\vec{y} - \Phi \vec{c}\|^2, \quad (3.11)$$

em que Φ é a matriz $c_k \times K$ contendo os valores de $\phi_k(t_j)$.

Como já se conhece do cálculo básico tradicional, o problema de minimização de uma função analítica pode ser resolvido verificando-se os valores para os quais a primeira derivada desta função se anula. Desta forma, o problema passa a ser a solução da equação:

$$2\Phi\Phi'\vec{c} - 2\Phi'\vec{c} = 0 \quad (3.12)$$

e o vetor $\vec{c}$ satisfazendo esta equação é dado por:

$$\vec{c} = (\Phi'\Phi)^{-1}\Phi'\vec{y}. \quad (3.13)$$

O vetor de valores ajustados é dado então por:

$$\vec{\hat{y}} = \Phi\vec{c} = \Phi(\Phi'\Phi)^{-1}\Phi'\vec{y}. \quad (3.14)$$

Este modelo simples de aproximação apresenta resultados satisfatórios quando pode se assumir que ϵ_j é i.i.d. (independente e identicamente distribuído), com média 0 e variância σ^2 .

Em situações mais complexas em que o modelo tradicional para ϵ_j não se aplica, como no caso em que este erro é não-estacionário ou é auto-correlacionado, uma nova expressão para o SMSSE pode ser usada:

$$SMSSE(\vec{y}|\vec{c}) = (\vec{y} - \Phi\vec{c})'W(\vec{y} - \Phi\vec{c}). \quad (3.15)$$

A matriz W é positiva definida e simétrica o que permite que pesos diferentes sejam atribuídos aos elementos do produto. Se $\sum_{\vec{e}}$ é a matriz de variância-covariância de ϵ_j , a matriz W pode ser obtida por:

$$W = \sum_{\vec{e}}^{-1} \quad (3.16)$$

Se o conhecimento de $\sum_{\vec{e}}$ não for possível, a matriz W pode ser definida como uma matriz diagonal cujos valores não-nulos correspondem à variância associada a cada ϵ_j .

O vetor de coeficientes $\vec{c}$ nesta nova abordagem é dado por:

$$\vec{c} = (\Phi'W\Phi)^{-1}\Phi'W\vec{y}. \quad (3.17)$$

3.4 Motivação

O uso de ADF como uma ferramenta estatística surgiu a partir do fato de que as manipulações matemáticas usadas pela estatística multivariada tradicional podem ser estendidas sem muitas dificuldades para o uso com funções.

Desta forma, pode-se definir a média aritmética funcional:

$$\bar{x}(t) = N^{-1} \sum_{i=1}^N x_i(t) \quad (3.18)$$

em que x_i são N funções avaliadas em t . Tem-se ainda a definição da variância funcional:

$$var_X(t) = (N-1)^{-1} \sum_{i=1}^N [x_i(t) - \bar{x}(t)]^2 \quad (3.19)$$

e, assim como na abordagem clássica, o desvio padrão é dado pela raiz da variância.

Para o caso de dois argumentos t_1 e t_2 , a função de covariância pode ser definida por:

$$cov_X(t_1, t_2) = (N-1)^{-1} \sum_{i=1}^N x_i(t_1) - \bar{x}(t_1)x_i(t_2) - \bar{x}(t_2) \quad (3.20)$$

e a função de correlação é:

$$corr_X(t_1, t_2) = \frac{cov_X(t_1, t_2)}{\sqrt{var_X(t_1)var_X(t_2)}} \quad (3.21)$$

Ainda para o caso em que as variáveis estudadas são pares de funções (x_i, y_i) pode-se definir a variância cruzada:

$$cov_{X,Y}(t_1, t_2) = (N-1)^{-1} \sum_{i=1}^N x_i(t_1) - \bar{x}(t_1)y_i(t_2) - \bar{y}(t_2) \quad (3.22)$$

e a correlação cruzada:

$$corr_{X,Y}(t_1, t_2) = \frac{cov_{X,Y}(t_1, t_2)}{\sqrt{var_X(t_1)var_Y(t_2)}} \quad (3.23)$$

Para a aplicação realizada neste estudo, os dados funcionais são representados por funções pertencentes ao espaço de Hilbert. Este é o espaço $L^2(V)$ munido da operação de produto interno definido por $\langle f, g \rangle = \int_V f(x)g(x)dx$. A partir do conceito de produto interno pode se definir a norma de uma função u no espaço de Hilbert como sendo $\|u\| = \sqrt{\langle u, u \rangle}$.

Operações tradicionais da estatística multivariada podem ser redefinidas com objetos funcionais no espaço de Hilbert por meio dos ajustes necessários ao método tradicional. Este é o caso, por exemplo, do método de k-médias, baseado em operações lineares elementares, que é adaptado para uma aplicação com dados funcionais em (38). Ainda em (39) pode se ver uma aplicação de redes do tipo MAO (Mapeamento Auto-Organizável) a objetos funcionais.

No caso em que o dado funcional é submetido a métodos que envolvem apenas operações lineares, a função pode ser representada simplesmente como sendo o conjunto de coeficientes c_k usados no desenvolvimento em funções-base. De fato, se u e v são funções representadas em um determinado espaço de funções-base, vale a seguinte propriedade linear: $c_k(\lambda u + \mu v) = \lambda c_k(u) + \mu c_k(v)$, para todo k inteiro válido, em que λ e μ são números reais e c_k é o conjunto de coeficientes.

Já o presente trabalho apresenta aplicações de métodos de classificação mais elaborados, como Análise de Dados Linear e Flexível e Redes Neurais Artificiais Perceptron Multicamadas. Para estes métodos, que envolvem operações mais complexas como produto interno e distância, um tratamento mais cuidadoso deve ser dado na representação do objeto funcional de maneira a torná-lo manipulável pelo método de análise.

De fato, o produto interno entre duas funções u e v , representadas por q conjuntos de coeficientes c_k e c_l , e por funções base ϕ_k e ϕ_l , respectivamente, é dado por:

$$\langle u, v \rangle = \sum_{k=1}^q \sum_{l=1}^q c_k(u) c_l(v) \langle \phi_k, \phi_l \rangle, \quad (3.24)$$

ou ainda usando-se uma notação mais sucinta:

$$\langle u, v \rangle = c(u)^T \Phi c(v), \quad (3.25)$$

em que T é o operador de transposição e Φ é a matriz de produtos internos $\Phi_{kl} = \langle \phi_k, \phi_l \rangle$, que independe de u e v . Esta fórmula demonstra que o produto interno definido no espaço das funções-base usadas difere do mesmo produto definido no espaço $\mathbb{R}^q$, sendo

q um inteiro, em que se define o vetor de coeficientes do desenvolvimento em funções-base. Seguindo o mesmo princípio, o conceito de norma da diferença (distância) entre funções também é redefinido:

$$||c(u) - c(v)||^2 = (c(u) - c(v))^T \Phi (c(u) - c(v)). \quad (3.26)$$

Para o caso em que o conjunto de funções-base é ortonormal, como ocorre com Fourier, a matriz Φ é a identidade e novamente o vetor de coeficientes $\vec{c}$ pode ser usado diretamente. Entretanto para o caso em que esta propriedade não se verifica, como ocorre com as *B-splines* aqui empregadas, a matriz Φ deve ser levada em conta e calculada.

Para que a influência da matriz Φ possa ser representada no computador, uma solução simples e eficiente é o uso da decomposição de Cholesky, na qual uma matriz quadrada Φ é decomposta em uma outra matriz quadrada U , tal que $\Phi = U^T U$. Desta transformação pode se obter um novo vetor de coeficientes $\bar{\vec{c}} = U c(u)$. Nota-se que a realização de operações lineares sobre o novo vetor continua válida:

$$\bar{\vec{c}}(\lambda u + \mu v) = \lambda \bar{\vec{c}}(u) + \mu \bar{\vec{c}}(v). \quad (3.27)$$

Agora, porém, as operações de produto interno também podem ser realizadas diretamente sobre $\bar{\vec{c}}$:

$$\bar{\vec{c}}(u)^T \bar{\vec{c}}(v) = \sum_{k=1}^q \bar{\vec{c}}_k(u) \bar{\vec{c}}_k(v) = \langle u, v \rangle. \quad (3.28)$$

Isto implica que as operações realizadas no espaço vetorial $\mathfrak{R}^q$, usando-se as coordenadas de $\bar{\vec{c}}$, correspondem exatamente às operações realizadas no espaço das funções-base de u e v . Por fim, isto significa que qualquer objeto funcional pode ser usado como entrada para outro método de análise de dados clássico, bastando para isto que um sistema de funções-base seja escolhido e o vetor $\bar{\vec{c}}$ seja calculado para cada dado funcional.

4 *Classificação de Formas com Base em Descritores*

O capítulo a seguir estuda métodos para reduzir a dimensionalidade de sinais contendo um alto número de valores como costuma ocorrer com as assinaturas obtidas por DFM descritas no capítulo 2.

4.1 Descritores de Fourier

Os descritores de Fourier podem ser obtidos a partir da transformada de Fourier (40). A Transformada de Fourier apresenta um grande número de aplicações em processamento de imagens e sinais e foi desenvolvida com o objetivo de possibilitar a transferência da análise do sinal inicialmente no domínio do espaço para o domínio da frequência (41)(42). A fórmula usada para expressar tal transformação é:

$$F(u) = \int_{-\infty}^{+\infty} f(x)e^{-2\pi jux} dx, \quad (4.1)$$

em que f é a função original no domínio espacial x , F é a função transformada no domínio da frequência u e j é o número imaginário $\sqrt{-1}$. Quando a função $f(x)$ é real como nos casos aqui tratados, a transformada de Fourier é complexa e costuma ser representada por:

$$F(u) = R(u) + I(u)j, \quad (4.2)$$

em que R e I são, respectivamente, as partes real e imaginária da função transformada.

Em aplicações computacionais como as que são aqui estudadas, costuma-se usar a versão discreta da transformada. Para isto, necessita-se de um processo de discretização no qual uma função contínua $f(x)$ passa a ser representada por:

$$f(x) = f(x_0), f(x_0 + \Delta x), \dots, f(x_0 + (N - 1)\Delta x) \quad (4.3)$$

ou ainda,

$$f(x) = f(x_0 + x\Delta x), \quad (4.4)$$

em que Δx são N intervalos de discretização de mesmo comprimento e x assume agora valores inteiros $0, 1, 2, \dots, N - 1$. Neste contexto, a transformada de Fourier no domínio discreto pode ser expressa por:

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) e^{-2\pi u x j / N}, \quad (4.5)$$

sendo $u = 0, 1, \dots, N - 1$.

A transformada de Fourier, sobretudo em sua versão discreta, tornou-se popular em análise de sinais devido à sua simplicidade de implementação e aos bons resultados alcançados nos mais diversos tipos de processamento que podem ser aplicados ao sinal transformado. Estes bons resultados se devem primordialmente ao fato de que a transformada apresenta invariância à rotação, translação e escala devido à facilidade de normalização, além de apresentar tolerância a ruídos. Ainda outro fato interessante é que no domínio da frequência existe uma alta concentração de energia em determinadas regiões, o que permite que todo o espectro seja representado apenas por um reduzido subconjunto de coeficientes da transformada.

A expressão do sinal no domínio da frequência representa seu espectro, que indica também a distribuição de energia do sinal. Ocorre que uma parte significativa desta distribuição de energia está concentrada nos componentes de baixa frequência do sinal transformado ($F(u)$). Estes componentes de baixa frequência são freqüentemente usados como descritores do sinal original, os chamados descritores de Fourier do sinal, e são usados para representar este sinal usando-se poucos elementos, permitindo assim a aplicação de algoritmos de classificação que possuem um aumento drástico em seu custo (tempo) computacional à medida que aumenta o tamanho da entrada (43). Este é o caso das assinaturas DFM aqui usadas que apresentam geralmente uma quantidade alta de valores tornando necessária a redução de sua dimensionalidade.

Neste trabalho, os descritores $D(u)$ são obtidos em duas etapas. Na primeira, efetua-

se o cálculo da transformada de Fourier por meio da Transformada Rápida de Fourier (cuja sigla em inglês é FFT) (9)(44) e calcula-se a magnitude da transformada. Esta magnitude é obtida por:

$$||F(u)|| = [R^2(u) + I^2(u)]^{1/2} \quad (4.6)$$

Os valores de $||F(u)||$ correspondendo às menores frequências são os descritores $D(u)$. Em seguida, estes valores são normalizados, para que a tolerância a ruídos e variações geométricas simples se mantenha. Assim, pode-se resumir o processo na expressão:

$$D(u) = \begin{cases} 0 & \text{se } f = 0; \\ D(u)/D(1) & \text{se } f \neq 0. \end{cases}$$

A Figura 18 ilustra como os descritores de Fourier podem ser usados para caracterizar sinais de diferentes formas. Nota-se que dois sinais distintos, ambos com uma grande quantidade de valores, geraram conjuntos reduzidos de descritores de Fourier significativamente diferentes entre si.

4.2 Descritores *Wavelets*

A transformada de Fourier apresenta alguns problemas relevantes, mas o principal deles é o fato de que a forma como é implementada, fazendo-se uma integração da função original em todo o seu domínio para gerar a transformada, impede que uma mudança no domínio original (do espaço ou tempo no caso de séries temporais) seja capturada com fidelidade. Tal fato faz com que a representação de sinais não-estacionários (com variação no domínio do tempo) fique comprometida.

Com o objetivo de contornar esta situação, desenvolveu-se a transformada *wavelet* que apresenta como principal novidade em relação a Fourier o uso de uma função base de suporte compacto (domínio limitado).

A técnica de *wavelets* tal como é conhecida hoje foi desenvolvida em (45) e é baseada na transformada de Fourier por janelas, desenvolvida por Gabor em 1946 (46), na qual a transformada de Fourier é aplicada em janelas tomadas no domínio do espaço (ou tempo).

A vantagem acrescentada pela abordagem de *wavelets* é a flexibilidade no tamanho desta janela que não é mais constante como ocorria na técnica de Gabor. A transformada

(a)

(b)

Figura 18 — Descritores de Fourier. (a) Sinais originais obtidos pela aplicação de DFM à forma das letras A e B. (b) 50 descritores de Fourier para as mesmas assinaturas.

wavelet é expressa, no caso contínuo, pela seguinte fórmula:

$$F(a, b) = \int f(x) \psi_{a,b}(x) dx \quad (4.7)$$

em que as funções $\psi_{a,b}(x)$ são chamadas de bases *wavelets* e devem cumprir os seguintes requisitos:

$$\psi_{a,b}(x) = \frac{a}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (4.8)$$

e

$$\int \psi(x) dx = 0 \quad (4.9)$$

Nota-se assim que a transformada *wavelet* depende de dois parâmetros a e b que representam respectivamente a escala e o tempo.

A obtenção da transformada contínua exige o cálculo para muitos níveis de escala, o que o torna dispendioso. Sendo assim, nas aplicações práticas usa-se a transformada discreta. A versão discreta da transformada de $f(x)$ é o conjunto de valores dos parâmetros c_k e $d_j k$ na seguinte equação:

$$f(tx) = \sum_{k=-\infty}^{\infty} c_k \phi(x-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi(2^j x - k), \quad (4.10)$$

em que ϕ e ψ são, respectivamente, as *wavelets* pai e mãe.

Na prática, a transformada discreta é aplicada ao sinal em etapas (níveis de decomposição). Em cada um destes níveis aplica-se um filtro passa-baixas que gera um vetor de coeficientes chamado de aproximação e um filtro passa-altas que gera o vetor de detalhe. No nível seguinte, a transformação é aplicada apenas à aproximação do nível anterior ou à aproximação e detalhe separadamente (*wavelet-packet*) (47). Desta forma, constrói-se a chamada árvore de decomposição *wavelet*, na qual em cada nível apresenta-se uma maior compactação da energia (magnitude) do vetor de aproximação. A transformada discreta convencional é obtida, a cada nível, pela concatenação do vetor de aproximação do nível atual com o vetor de detalhes do nível anterior. Já na *wavelet-packet*, em cada nível, esta mesma transformada é a concatenação entre a aproximação e o detalhe daquele nível.

Da mesma forma que na transformada de Fourier, a energia do sinal transformado está

Figura 19 – Árvore de decomposição no método de wavelets.

concentrada em poucos termos de menor frequência, que podem, desta forma, ser tomados como descritores representativos do sinal original como um todo (4). Para a extração destes descritores, várias combinações de aproximação, detalhe e outros parâmetros extraídos da árvore podem ser usadas, tomando-se o cuidado, assim como no caso de Fourier, de se aplicar um processo de normalização a estes descritores.

Na Figura 19 tem-se a árvore de decomposição de *wavelets*, em que A_n e D_n representam, respectivamente, a aproximação e o detalhe em nível n .

4.3 Aplicação de ADF à Geração de Descritores

Neste contexto, o objetivo deste trabalho é o estudo do uso de ADF associada à técnica de DFM para a geração de descritores de imagens, sobretudo representando objetos de interesse em imagens mais complexas. A idéia surge a partir de trabalhos como (8)(7) em que se usa ADF para gerar um conjunto de valores representativos do dado original, cujo tratamento direto seria inviável ou muito dispendioso computacionalmente.

4.3.1 Uso Direto

Esta abordagem é usada no trabalho (7), em que são usados os coeficientes da representação funcional de curvas que medem a qualidade da água em uma determinada região. Como os coeficientes são usados diretamente para aplicação de um método classificativo, esta técnica representa o uso direto de ADF na geração de descritores.

Em (7), o autor traça uma abordagem de agrupamento funcional hierárquico. O objetivo é generalizar o conceito da estatística clássica, na qual os agrupamentos são

construídos a partir da definição de uma medida de distância entre vetores de amostra. Para facilitar a análise ADF, as funções usadas nesse trabalho pertencem ao espaço vetorial de Hilbert $L^2(V)$, isto é, o espaço das funções de quadrado integrável de V , em que V é um compacto de $\mathbb{R}^p$, p inteiro em $\mathbb{R}$. De fato, se existem duas funções $f_i(t)$ e $f_j(t)$ satisfazendo tais condições pode-se definir uma medida de distância entre elas. Uma possibilidade para tal fim é adotar:

$$d_{ij} = (\vec{c}_i - \vec{c}_j)' W (\vec{c}_i - \vec{c}_j) \quad (4.11)$$

em que $\vec{c}_i$ e $\vec{c}_j$ são os vetores de coeficientes e W é a matriz dos produtos internos das bases:

$$W_{ab} = \int_0^T \phi_a(t) \phi_b(t) dt \quad (4.12)$$

Como $d_{ij} = d_{ji}$, esta matriz é simétrica e assim apenas metade da matriz precisa ser calculada e, nos casos de bases ortogonais como Fourier, a matriz W é a própria matriz identidade. De posse desta métrica de distância o próximo passo é a definição de agrupamentos visando uma menor variabilidade (distância entre os elementos) intra-classes e maior variabilidade inter-classes. Para tal podem ser usados métodos simples e conhecidos da literatura, como K-médias e K-medóides (48).

4.3.2 Mudança de Espaço Algébrico

Esta abordagem é proposta pela primeira vez, ao que se sabe, em (8). Esta estratégia se faz necessária no caso da aplicação de ADF na geração de entradas para métodos que não são baseados em distância. Tal é o caso das Redes Neurais Artificiais (RNA).

Neste trabalho, a abordagem de RNA adotada foi a de Multi-Layer Perceptron (MLP), sendo a que apresenta a melhor relação entre precisão de resultados e simplicidade de implementação (49). Neste método, o dado de entrada é manipulado por uma função M , chamada função de ativação, tal que

$$M = T(\lambda_0 + \sum_{i=1}^p \lambda_i x_i) \quad (4.13)$$

em que T é um operador não linear, x_i são componentes do vetor de entrada e λ_i são parâmetros, chamados de pesos.

Desta forma, esta função pode ser adaptada para o uso em qualquer espaço vetorial normado (8). De fato, o somatório em 4.13 pode ser substituído por uma forma linear $l(x)$ do espaço conjugado. Em espaços de Hilbert, esta forma linear é o produto interno $\langle u, v \rangle$.

Embora o vetor de coeficientes do funcional possa ser usado diretamente em fórmulas lineares, seu uso no cálculo de distâncias e produtos internos gera distorções inconvenientes. Para esta situação, tem-se que o produto interno de duas funções u e v representadas pelo vetor de coeficientes $\vec{\alpha}$ do desenvolvimento em B é dado por

$$\langle u, v \rangle = \sum_{k=1}^q \sum_{l=1}^q \vec{\alpha}_k(u) \vec{\alpha}_l(v) \langle \phi_k, \phi_l \rangle, \quad (4.14)$$

sendo

$$\langle \phi_k, \phi_l \rangle = \int_{\mathbb{R}^q} \phi_k(t) \phi_l(t) dt. \quad (4.15)$$

A Fórmula 4.14 pode ser resumidamente expressa em

$$\langle u, v \rangle = \vec{\alpha}(\vec{u})^T \Phi \vec{\alpha}(\vec{v}). \quad (4.16)$$

em que T é o operador de transposição e Φ é matriz cujos elementos são dados por $\Phi(k, l) = \langle \phi_k, \phi_l \rangle$. A matriz Φ é independente de u e v e no caso de bases ortonormais, como Fourier, é a matriz identidade. Para outros tipos de bases que não apresentam esta mesma característica, entretanto, faz-se necessário o cálculo dos valores de Φ . Para facilitar manipulações algébricas, os autores de (8) propõem que se defina um novo vetor de coeficientes que reflete a mudança de espaço algébrico do vetor de números reais para o espaço das bases de representação. Este novo vetor $\beta(u)$ é definido por:

$$\beta(u) = M \alpha(u) \quad (4.17)$$

em que M é a matriz quadrada obtida pela decomposição de Cholesky de Φ (8).

5 *Métodos de Classificação*

Para a verificação da precisão do uso de ADF na geração de descritores é necessário aplicar-se estes descritores como entrada para diferentes métodos de classificação. Além de ADF, outras formas de se gerarem descritores são aplicadas e estes descritores são usados para o mesmo método de classificação. Ao final, uma comparação é realizada entre os resultados para que se obtenham conclusões sólidas acerca do método em estudo. Este capítulo trata dos métodos de classificação usados.

5.1 K-médias

O método básico de classificação aplicado neste trabalho é o método não supervisionado de k-médias. O objetivo de tal técnica é o agrupamento (50)(51) em k grupos das amostras de entrada com base em seus atributos (vetores de características ou, neste caso, descritores). O método é levado a cabo por meio de seguidas iterações que visam aumentar a distância inter-grupos e diminuir a distância intra-grupo. A cada iteração, a técnica visa minimizar:

$$K = \sum_{i=1}^k \sum_{x_j \in C_i} (x_j - \mu_i)^2, \quad (5.1)$$

em que C_i são os agrupamentos formados e μ_i é o vetor de características médio do grupo i , também chamado de centróide. A cada iteração, o valor de cada μ_i , bem como o somatório acima, é recalculado. O particionamento inicial pode ser feito aleatoriamente ou seguindo alguma heurística de inicialização.

Métodos baseados em distância como k-médias ilustram bem o fato de que muitos métodos da estatística multivariada convencional não ficam limitados pela dimensão finita de seus vetores de entrada. No caso de k-médias, o agrupamento dos vetores é atualizado com base em uma combinação linear de outros vetores usados para redefinir os centróides

(48).

5.2 Rede Neural Artificial

Uma Rede Neural Artificial (RNA) é uma técnica computacional que simula o funcionamento do cérebro biológico.

A estrutura básica de uma RNA perceptron, como a usada neste trabalho, consiste da entrada de um sinal que deverá ser classificado e um conjunto de unidades de processamento que multiplicam cada componente da entrada por um valor real chamado de peso e somam os produtos para todos os componentes. Esta soma produz um valor chamado de atividade. A resposta de saída é dada com base no fato de este valor exceder ou não um limite definido previamente. Na maioria das vezes, a arquitetura da RNA é composta de camadas (de entrada, de saída e, possivelmente, uma ou mais camadas intermediárias).

Portanto, uma rede neural deste tipo pode ser modelada pela função f :

$$f(x) = K\left(\sum_i w_i g_i(x)\right), \quad (5.2)$$

sendo K uma função pré-definida, chamada de função de ativação, w_i os pesos e g_i um conjunto de funções que geram a função geral do modelo.

Estes pesos são determinados, na maioria das vezes, com base em estratégias de treinamento usadas com sinais de entrada separados para tal fim. Esta estratégia de treinamento (ou de aprendizado) pode ser supervisionada, no qual a classificação de cada sinal de treinamento é fornecida por um agente externo (geralmente o usuário do sistema). Há ainda o aprendizado não supervisionado, no qual a resposta para cada elemento treinado é determinada pela própria rede. Por fim, tem-se o aprendizado com reforço, no qual um agente externo avalia a resposta da rede (49).

O uso de uma rede com apenas duas camadas (entrada e saída), embora seja possível, traz consigo uma grande limitação à aplicação da técnica, uma vez que todo mapeamento realizado é linear. Desta forma que o autor do trabalho (52) concluiu que o uso deste tipo de rede não possibilita a solução de problemas que não são linearmente separáveis. Com o uso de uma ou mais camadas intermediárias, a rede ganha um caráter mais “inteligente”, uma vez que a rede “aprende” novas formas de mapeamento, o que faz com que a resposta não dependa tão diretamente dos dados de entrada.

A aplicação na prática de redes multicamadas só foi possível na década de 80 quando os autores do trabalho (53) desenvolveram o algoritmo de *backpropagation*, que mostrou que é possível o treinamento de uma rede com camadas intermediárias. Estas redes são chamadas de MLP (sigla de Perceptron Multicamadas em inglês).

5.3 Análise Discriminante Linear

O método de classificação denominado Análise Discriminante Linear (ADL) busca classificar objetos com base em um conjunto de características destes objetos, denominado vetor de características (ou descritores, como são tratados aqui).

Como todo método satisfatório de classificação, ADL deve visar a uma minimização do número de objetos classificados erroneamente, tanto no sentido de objetos atribuídos a uma classe à qual não pertencem, quanto de objetos que pertencem a determinada classe e não são atribuídos a ela. Para tal, o método de análise discriminante deve buscar o melhor conjunto de descritores para cada objeto e a melhor regra para se obterem as respectivas classes com base nestes descritores.

O primeiro problema é tratado com maiores detalhes em (54). Já a regra de classificação varia de acordo com a estratégia de classificação discriminante usada. Para o caso de ADL aplicada a um determinado conjunto de descritores x , usa-se uma função linear em x ou uma combinação de funções lineares de x , chamada de discriminante, cujas variáveis independentes são os descritores e as variáveis de saída fornecem as classes dos objetos correspondentes aos respectivos descritores.

As funções discriminantes devem satisfazer certos requisitos. A função ótima é obtida quando as distribuições de probabilidade de cada classe são cooperativas. Funções lineares costumam ser usadas como um classificador inicial em problemas dos quais não se tem um conhecimento exato acerca das distribuições de probabilidade e não se conhecem informações que sugiram o uso de outro tipo de função discriminante.

Uma vez que a função ideal não pode ser encontrada na maioria dos casos, assim como ocorre com outros métodos de classificação, o encontro da melhor função discriminante se resume à minimização de um determinado critério. O critério mais simples é a taxa de erro na classificação de um dado conjunto cuja classificação correta é conhecida previamente, chamado de conjunto de treinamento. Ocorre, entretanto, que o acerto na classificação deste conjunto de treinamento não pode garantir o acerto no conjunto de dados usado nas aplicações efetivas do método, chamado de conjunto de teste. Assim, vários critérios

mais complexos são desenvolvidos para a solução deste problema.

Para o caso em que o problema apresenta apenas duas classes, o gráfico da função discriminante é uma reta. Neste caso, a função discriminante pode ser dada por:

$$f(x) = p(\vec{x}) + p_0, \quad (5.3)$$

em que $p(x)$ é o vetor de pesos e p_0 é o limiar de peso. O dado com vetor de descritores x é atribuído à classe 1 se $f(x) > 0$ e à classe 2 em caso contrário. A expressão $f(x) = 0$ define o que se costuma chamar de superfície de decisão, que no caso em que $f(x)$ é linear, para duas classes, expressa-se como um plano. Geometricamente falando, neste caso, pode-se dizer que a função discriminante linear divide o espaço de descritores por um plano e a magnitude da função para cada conjunto de descritores pode ser medida como sendo a distância deste conjunto ao plano.

Para o caso em que mais classes estão envolvidas, caso em que o método também é chamado de Análise Discriminante Mista (ADM), outras estratégias podem ser usadas, como a aplicação de uma função discriminante linear eliminando do problema uma classe por vez ou ainda a definição de uma função para cada par de classes. Qualquer uma destas abordagens porém podem levar a situações em que a classificação não é possível. Assim, propõe-se o uso da generalização da equação 5.3 usada para 2 classes. Nesta abordagem, para cada classe i , tem-se uma função discriminante linear:

$$f_i(x) = p^t x_i + p_{i0}, i = 1, \dots, c \quad (5.4)$$

O dado com descritores x é atribuído à classe i se $f_i(x) > f_j(x)$, para todo j . Esta função divide o espaço de características em c regiões de decisão separadas por hiperplanos.

Na prática, para a classificação propriamente dita, ADL usa a regra de Bayes (48). Assim, um objeto x é atribuído a uma determinada classe i de um total de g classes se ela apresenta a maior probabilidade condicional de conter aquele objeto, ou seja, se:

$$P(i|x) > P(j|x), \forall j \neq i. \quad (5.5)$$

Como a obtenção de $P(i|x)$ é mais complexa, aplica-se o Teorema de Bayes para a obtenção desta probabilidade a partir de $P(x|i)$, cujo conhecimento é bem mais simples

quando se tem um conjunto de objetos com classificação conhecida, usados para treinamento. A partir do teorema obtém-se:

$$P(i|x) = \frac{P(x|i).P(i)}{\sum_{\forall j} P(x|j).P(j)} \quad (5.6)$$

Para fins de facilidade matemática e computacional, assume-se que $P(x|i)$ apresenta distribuição normal multivariada, sendo que cada classe apresenta uma matriz de covariância. Assim, por (55), obtém-se:

$$f_i = \mu_i C^{-1} x_k^T - 0.5 \mu_i C^{-1} x_k^T + \log(p_i), \quad (5.7)$$

sendo μ_i a média dos vetores de características da classe i , C é a matriz de covariância do conjunto de classes e p_i é a probabilidade a priori da classe i . p_i pode ser considerada como sendo a frequência relativa de objetos do conjunto de treinamento pertencentes à classe i . O objeto k é atribuído à classe i apresentando maior f_i .

5.4 Análise Discriminante Flexível

Neste método, denominado Análise Discriminante Flexível (ADFL), ADL é aplicada ao conjunto de vetores de descritores, gerando valores que são usados como rótulos para as classes envolvidas no problema. Uma função $g : 1, \dots, C \mapsto \mathbb{R}^1$ mapeia cada um destes rótulos para cada uma das C classes. Na prática, N funções $g_1, g_2, \dots, g_N$ de rotulação são usadas e os mapeamentos gerados são dados por $\eta_n(X) = X^T \beta_n$, com $n = 1, \dots, N$, em que g e η são obtidas por

$$\min_{g, \eta}(ASR) = 1/T \sum_{n=1}^N \left[\sum_{i=1}^T (g_n(c_i - x_i^T \beta_n))^2 \right]. \quad (5.8)$$

Mais detalhes sobre esta fórmula podem ser encontrados em (55). A quantidade ASR é a sigla em inglês de resíduo quadrático médio e este mínimo é calculado a partir do conjunto de treinamento composto pelo vetor de preditores de objetos cujas classes são conhecidas previamente. A idéia do método de ADFL é usar-se uma fórmula de regressão não parametrizada, mais flexível do que $\eta_n(X)$. Assim, conforme descrito no trabalho (55), surge a seguinte fórmula de regressão:

$$ASR(g_n, \eta_{n=1}^N) = 1/T \sum_{n=1}^N \left[\sum_{i=1}^T (g_n(c_i) - \eta_n(x_i))^2 + \lambda L(\eta_n) \right], \quad (5.9)$$

sendo L um parâmetro regularizador usado em alguns tipos de regressão e λ é um parâmetro global da regressão.

A estratégia por ADFI vem suprir a lacuna deixada por ADL para o caso em que o conjunto de dados não é linearmente separável. Tal dificuldade é superada pelo uso de uma fronteira de decisão que não necessita mais ser linear, mas pode ser irregular. Em paralelo a isto, ADFI mantém as vantagens de ADL no que diz respeito à simplicidade da técnica, uma vez que emprega regras simples de processamento e usa o conjunto de descritores de forma direta para a classificação.

6 *Metodologia*

Este capítulo descreve os passos executados em todas as etapas desta pesquisa.

6.1 Passos Adotados

Este trabalho apresenta uma aplicação de ADF à geração de descritores para um vetor de assinatura obtido por DFM. A assinatura é um conjunto com um número alto de valores (próximo de 1000 na maioria dos casos) e apresenta comportamento intrinsecamente suave. Tais fatos capacitam estes vetores numéricos a serem tratados como funções. As funções são geradas a partir do desenvolvimento no sistema de bases de *B-splines*. Este sistema é usado devido à sua flexibilidade inerente e ao fato das assinaturas não apresentarem comportamento periódico. Em seguida, usando-se o procedimento descrito anteriormente, os coeficientes $\bar{c}$ são processados e usados como descritores da assinatura e estes descritores são usados como entrada para métodos de classificação.

Já se conhecem exemplos do uso de descritores de Fourier destas assinaturas em (23). Este trabalho, entretanto, é o primeiro, ao que se sabe, a estudar o uso de ADF para este fim. Em (6) verificam-se aplicações de ADF em vários campos da ciência, mas a maioria destas aplicações está relacionada ao reconhecimento de padrões e à previsão de comportamento de uma determinada variável em relação à outra. Ao que se sabe, aplicações de ADF no tipo de problema abordado nesta dissertação são inéditas.

Os descritores usados nos experimentos aqui realizados para fins de comparação são ADF, Fourier e *wavelets*. Estes descritores são empregados como entrada para os métodos de classificação usados, que foram K-médias, ADF1, ADL e RNA.

A forma como é executado método de K-médias possibilita a aplicação do conceito de K-médias funcional, no qual os vetores são substituídos por funções. A linearidade sugere a possibilidade de se usar diretamente em tal operação o vetor de coeficientes destas funções representadas em um determinado espaço de funções-base mais simples.

Figura 20 – Esquema demonstrando a aplicação de ADF no presente trabalho.

Entretanto, neste trabalho a abordagem usada foi a descrita em (8). Já o modelo de RNA usado é o de perceptron multicamadas, o mais simples dos modelos de retro-alimentação (49). Por fim as taxas de acerto obtidas pelo método de classificação são comparadas para diferentes bases de dados, para os diferentes descritores e métodos de classificação.

A Figura 20 resume todo o processo.

7 *Resultados Experimentais*

Este capítulo descreve cada uma das bases de dados a serem usadas nos experimentos de comparação entre os métodos de classificação e de descritores, além de explicar de forma mais detalhada a realização de cada experimento como, por exemplo, a determinação dos parâmetros envolvidos. Por fim, são apresentados os resultados obtidos, separados pela base de dados usada. Para cada base, foram usados descritores de Fourier, *wavelets* e ADF. Os métodos de classificação usados foram Redes Neurais, Análise Discriminante Flexível, Análise Discriminante Linear e K-médias. Os gráficos mostram a taxa de acerto de cada combinação de descritores e métodos de classificação em função do número de descritores usados.

7.1 Bases de Dados Utilizadas

Os experimentos de comparação entre os métodos foram feitos com a aplicação de testes a cinco bases de dados distintas, sendo duas naturais, de espécies de peixes e texturas de imagens de cenários reais e as outras três artificiais, sendo duas de formas genéricas e uma de letras do alfabeto afetadas por diferentes níveis de ruídos.

A primeira base usada, chamada de *shapes99*, é artificial e composta por nove tipos de formas genéricas, como carro, homem, etc., apresentando onze exemplares para cada um dos nove tipos de forma (classe). A figura 21 a seguir ilustra alguns elementos da base.

A segunda base de dados usada, de letras do alfabeto, é artificial, no sentido de que foi gerada computacionalmente com o objetivo de enfatizar a eficácia dos métodos testados, uma vez que o nível de ruídos na amostra é totalmente controlado.

As imagens usadas nesta base são das 26 letras do alfabeto latino, com fonte Arial, maiúscula, de tamanho 72, com cor branca sobre fundo preto. Para cada uma destas imagens, foram gerados quatro conjuntos de imagens que representam a imagem original

Figura 21 – Base de formas variadas. Cada linha apresenta uma classe de formas.

afetada por quatro níveis diferentes de ruídos aleatórios.

Estes ruídos foram aplicados por meio de um processo simples. Na primeira etapa deste processo, os contornos externo e interno (no caso de letras que o possuírem) são parametrizados em função das coordenadas x e y . Em seguida, à função obtida soma-se um valor aleatório para cada ponto da função (y), valor este que deve pertencer ao intervalo $[-n, n]$, em que n é o nível de ruído (valor inteiro). Por fim, a função obtida é usada para se redesenhar o contorno para que se obtenha a imagem afetada por ruído. Deve-se tomar o cuidado de se preencher o contorno externo com cor branca e o interno com cor preta. O processo é ilustrado em (23). Já a figura 22 mostra alguns elementos da base propriamente dita.

Figura 22 – Base de letras do alfabeto. Cada coluna representa uma letra (classe) afetada por diferentes graus de ruído.

Ao todo, a base é composta por 1378 formas, sendo 53 para cada letra (1 forma da letra original sem ruído e 13 formas para cada nível de ruído).

A terceira base, de Brodatz, apresenta um conjunto de imagens com dimensão 640×640 , cada uma delas apresentando uma textura diferente, baseadas em fotografias de cenários do mundo real. Deste conjunto, foram tomadas 40 imagens e de cada imagem foram extraídas 10 sub-imagens (janelas) de dimensões 100×100 não sobrepostas. Como cada textura é uma classe, a base possui ao todo 40 classes com 10 amostras cada.

A quarta base é chamada de MPEG-7 e é composta de formas genéricas, assim como a *shapes99*, entretanto apresenta um número maior de classes, sendo 70 classes de 20 amostras cada, totalizando 1400 amostras.

A quinta base, baseada em (56), é composta por contornos de imagens de peixes. Cada imagem de peixe corresponde a uma classe e a partir do contorno desta imagem são obtidas mais 9 imagens, usando-se rotação e ruído, seguindo-se o mesmo processo aplicado à base de letras. Ao todo, a base é composta por 1100 classes com 10 amostras cada uma.

7.2 Detalhes dos Experimentos

No caso em que se usou o método de K-médias para a classificação, a taxa de acerto foi medida pelo número de amostras de uma classe, que ao final do processo de classificação, efetivamente ficaram na mesma classe. O sistema foi implementado em *MATLAB*® (37).

Para os experimentos com RNA, foi definida uma meta de erro de 0.01%, um limite de épocas de treinamento de 15000, uma taxa de aprendizado de 0.01, usando-se redes perceptron multicamadas, com retro-alimentação e uma camada intermediária com um número de neurônios próximo da média geométrica entre os números de neurônios nas camadas de saída e entrada (que equivalem ao número de descritores usados e de classes envolvidas no problema, respectivamente). O sistema foi desenvolvido em *MATLAB*® (37).

O conjunto de dados foi dividido em dois grupos de igual tamanho, sendo que os elementos de cada grupo são escolhidos aleatoriamente. Um grupo é usado como teste e outro como treinamento.

No caso em que o método de classificação é ADFI ou ADL, usou-se a mesma divisão de conjuntos de treinamento e teste que no caso de RNA. Em alguns experimentos com

bases muito grande, porém, usou-se uma técnica chamada de *leave-one-out* (48), na qual se usa toda a base para o treinamento com exceção de um elemento que é usado para teste. O processo é repetido para cada elemento da base e a taxa de acerto ao final é obtida pela média de acerto para cada elemento usado no teste. O método de regressão usado foi o MARS (55). O sistema foi implementado em R (36).

7.3 Tempo Computacional em função do número de descritores

A tabela 1 ilustra o tempo computacional para a aplicação do método de ADL à base de letras do alfabeto com descritores de Fourier.

Tabela 1 – *Tabela do tempo de processamento em função do número de descritores usados.*

Número de descritores	Tempo de processamento (s)
10	0.037817
20	0.072094
30	0.087039
40	0.104817
50	0.146728
60	0.182718
70	0.214142
80	0.243574
90	0.297023
100	0.375234

Nota-se que um aumento no número de descritores acarreta um aumento no tempo de processamento, situação esta que pode se tornar crítica no caso de bases de dados com um grande número de amostras, situação que ocorre comumente com bases do mundo real.

7.4 Análise dos Resultados

7.4.1 Shapes99

A Figura 23 apresenta três gráficos comparativos do desempenho dos descritores usados no processo de classificação da base de formas. Nota-se a vantagem do uso de ADF em relação a Fourier para os três métodos de classificação usados, havendo uma pequena

Figura 23 – Base shapes99. (a) Classificação usando K-médias. (b) Classificação usando ADFL. (c) Classificação usando RNA.

aproximação no desempenho de *wavelets* apenas quando o método de classificação usado foi ADFL. Observa-se ainda que os melhores resultados foram alcançados quando a classificação foi efetuada por RNA. Vale ressaltar entretanto que o uso de RNA acarreta uma flutuação maior dos resultados. Tal fato se explica pela forma como foi implementada a rede neural neste caso, usando-se inicialização aleatória para os pesos.

7.4.2 Letras do Alfabeto

A base usada nos próximos resultados é a de letras. Para cada método de classificação, o experimento foi repetido para letras com ruído 1, 2, 3, 4 e com letras contendo todos os

Figura 24 – Base de letras do alfabeto com ruído de nível 1. (a) Classificação usando K-médias. (b) Classificação usando ADFI. (c) Classificação usando RNA.

níveis possíveis de ruídos.

A Figura 24 apresenta os resultados para o caso em que o nível de ruído da base é um. Nota-se que todos os descritores apresentaram taxa de acerto próxima de 100%. Neste caso, houve uma pequena flutuação nos resultados apenas quando o método de classificação foi o de K-médias. No caso de K-médias também houve um pior desempenho de *wavelets*.

Já a Figura 25 apresenta os resultados para o caso em que o nível de ruído da base é dois. Neste caso já se observa uma pequena vantagem de ADF, sobretudo em relação a Fourier. Esta vantagem fica mais evidente sobretudo no caso de classificação por rede

Figura 25 – Base de letras do alfabeto com ruído de nível 2. (a) Classificação usando K-médias. (b) Classificação usando ADFL. (c) Classificação usando RNA.

neural. Observa-se agora também uma maior flutuação dos resultados, principalmente no uso de RNA. Tal fato era esperado uma vez que o nível de dificuldade do processo de classificação aumentou, gerando uma maior quantidade de erros.

A Figura 26 apresenta o experimento com a base de letras com nível 3 de ruído. A vantagem do uso de ADF é um pouco mais evidente. Novamente esta evidência é mais clara quando se usa RNA. Nota-se também que os resultados alcançados com o uso de *wavelets* são mais irregulares. tal fato era esperado uma vez que o aumento do número de descritores não está associado a um aumento na performance do método de classificação, fato este que fica mais evidenciado quando o erro de classificação tende a ser maior.

(a)

(b)

(c)

Figura 26 – Base de letras do alfabeto com ruído de nível 3. (a) Classificação usando K -médias. (b) Classificação usando ADFL. (c) Classificação usando RNA.

Figura 27 – Base de letras do alfabeto com ruído de nível 4. (a) Classificação usando *K*-médias. (b) Classificação usando ADF. (c) Classificação usando RNA.

A Figura 27 apresenta o experimento com a base de letras com nível 4 de ruído. Neste caso, a vantagem dos descritores ADF chega ao seu máximo, quando se usa RNA, evidenciando a utilidade do uso de ADF na classificação de bases ruidosas. Mais uma vez o desempenho de *wavelets* foi bastante irregular e ligeiramente inferior ao de ADF e o desempenho de Fourier também foi sempre inferior.

A Figura 28 apresenta o experimento com a base de letras completa, ou seja, com todos os níveis de ruídos. Mais uma vez, a vantagem dos descritores ADF se confirma sobretudo no experimento com RNA. Neste caso porém o desempenho de *wavelets* se aproximou mais de ADF. Tal fato era previsível devido a uma diminuição do ruído médio

Figura 28 – Base de letras do alfabeto com os 4 níveis de ruídos misturados. (a) Classificação usando K-médias. (b) Classificação usando ADFI. (c) Classificação usando RNA.

da base (em relação à base com ruído 4 por exemplo).

7.4.3 Resultados para a análise de textura

Os gráficos da Figura 29 foram obtidos usando-se a base de Brodatz e os três métodos de descritores estudados. A classificação foi efetuada com K-médias e ADFI. Para esta base, assim como ocorre com a base MPEG-7 e a de contornos de peixes, o método de RNA não é usado devido ao alto custo computacional envolvido no processo de classificação por este método, devido ao alto número de amostras na base. A assinatura DFM original no caso de Brodatz é menor, apresentando apenas 85 componentes para o raio de dilatação

Figura 29 – Base de texturas de Brodatz. (a) Classificação usando K -médias. (b) Classificação usando ADFL.

máximo de 10 usado no método de DFM por Bouligand-Minkowsky. Assim, o número de descritores variou de 1 a 50, com o objetivo de haver uma redução significativa na quantidade de valores da assinatura original.

Nota-se uma vantagem significativa de ADF sobre Fourier e uma pequena vantagem de *wavelets* sobre ADF. Entretanto, vale destacar que, com poucos descritores ADF, já é possível se obter um resultado satisfatório, o que é muito importante para o custo computacional do problema, sobretudo neste caso em que a base de dados contém muitas amostras.

7.4.4 MPEG-7

Os gráficos da Figura 30 mostram a taxa de acerto em função do número de descritores para a base MPEG-7. Os métodos de classificação usados foram K -médias e ADFL. RNA não foi usada devido ao alto custo computacional envolvido na classificação por rede neural em bases de dados com quantidades maiores de amostras, como a MPEG-7. Neste experimento, usou-se *leave-one-out* para a classificação por ADFL. Mais uma vez, ADF leva vantagem sobre Fourier, especialmente quando se usa ADFL. Os desempenhos de ADF e *wavelets* se equivalem.

Figura 30 – Base de formas MPEG-7. (a) Classificação usando *K*-médias. (b) Classificação usando ADFI.

7.4.5 Contornos de Peixes

Na Figura 31, tem-se os experimentos realizados com a base de contornos de imagens de peixes, com classificação por ADFI, usando-se *leave-one-out*. Novamente, nota-se a vantagem de ADF em relação a Fourier e o melhor resultado foi alcançado usando-se descritores ADF. De um modo geral, ADF e *wavelets* tiveram resultados equivalentes, havendo vantagem de *wavelets* com determinados números de descritores (próximos de 20).

Figura 31 – Análise discriminante flexível aplicada à base de peixes.

7.4.6 Melhores Resultados Obtidos

A Tabela 2 apresenta a taxa de acerto para diferentes bases, com diferentes métodos de classificação, com diferentes descritores. Exibe-se o melhor resultado obtido com cada uma destas combinações e para cada resultado o número de descritores que foram necessários para se chegar até aquela taxa de acerto.

Observa-se que na base shapes99 os melhores resultados foram obtidos sempre usando-se RNA na classificação, sendo que o descritor que combinado a RNA apresentou a maior taxa de acerto foi o de ADF. Já no caso da base de letras com ruído misturado o melhor desempenho foi verificado com o uso de ADFI e a taxa mais alta novamente foi obtida com o uso de ADF. Para a base de Brodatz também o melhor método de classificação foi ADFI e a taxa de acerto mais alta foi obtida usando-se *wavelets*. No caso da base MPEG-7 o melhor desempenho se deu com o uso de K-médias, o que era esperado devido ao uso de *leave-one-out* no método ADFI, o que diminui a taxa de acerto em troca de um aumento na fidelidade do resultado. Já no caso da base de peixes, o maior acerto foi obtido com *wavelets*, embora a vantagem não tenha sido significativa em relação a ADF.

De um modo geral, se observou um bom desempenho dos descritores ADF, uma vez que nos casos em que a técnica teve um desempenho pior que *wavelets* ou mesmo Fourier, a vantagem foi pequena, com exceção da aplicação na base de texturas.

A Tabela 3 mostra a taxa de acerto médio para todos os números de descritores usados para cada base, método de classificação e de descritor. Estes dados são importantes uma vez que mostram uma tendência de comportamento de cada método, uma vez que os erros inseridos a cada passo no processo de classificação fazem com que a relação entre as taxas de acerto e o número de descritores não seja tão regular quando poderia se esperar.

Da tabela se observa que mais uma vez, com exceção da base MPEG-7, a melhor técnica de classificação foi ADFI. Os descritores ADF apresentaram melhor desempenho na base shapes99 e de Brodatz. No caso da base de letras, o melhor desempenho foi obtido com descritores *wavelets* e Fourier respectivamente, embora a vantagem sobre ADF tenha sido pouco significativa. Apenas na base de peixes, os descritores *wavelets* apresentaram um desempenho notadamente melhor que os demais métodos. Observa-se também que os maiores desvios padrões foram observados com as técnicas de RNA e ADFI. Este fato pode ser justificado pelos parâmetros escolhidos de forma aleatória para estes métodos, como pesos e conjuntos de treinamento e teste.

Por sua vez, a tabela 4 mostra os melhores resultados obtidos quando a técnica de

Tabela 2 – Tabela de melhores resultados para os métodos.

Base de Dados	Descritor	Classificador	Maior Acerto (%)	Número de Descritores
SHAPES99	Fourier	K-médias	77.78	40
		ADFl	78.06	20
		RNA	83.33	7
	<i>Wavelets</i>	K-médias	78.79	110
		ADFl	81.73	47
		RNA	83.33	123
	ADF	K-médias	82.83	10
		ADFl	81.84	110
		RNA	91.67	150
LETRAS	Fourier	K-médias	89.11	90
		ADFl	97.53	150
		RNA	89.94	70
	<i>Wavelets</i>	K-médias	86.43	59
		ADFl	97.39	123
		RNA	96.66	40
	ADF	K-médias	90.49	120
		ADFl	98.26	100
		RNA	93.19	70
BRODATZ	Fourier	K-médias	55.5	40
		ADFl	72.99	40
	<i>Wavelets</i>	K-médias	62	20
		ADFl	91.1	31
	ADF	K-médias	56.75	9
		ADFl	82.99	40
MPEG7	Fourier	K-médias	57.21	7
		ADFl	23.28	6
	<i>Wavelets</i>	K-médias	52.21	121
		ADFl	53.28	108
	ADF	K-médias	57.07	6
		ADFl	50.93	100
PEIXES	Fourier	ADFl	7.81	7
	<i>Wavelets</i>	ADFl	84.44	108
	ADF	ADFl	82.9	100

Tabela 3 – Tabela de resultados médios com seus respectivos desvios padrões para os métodos.

Base de Dados	Descritor	Classificador	Acerto médio (%)	Desvio padrão		
SHAPES99	Fourier	K-médias	71.3940	4.8768		
		ADFl	74.5224	4.1030		
		RNA	67.1667	21.0258		
	Wavelets	K-médias	74.2036	2.2817		
		ADFl	79.3878	1.2002		
		RNA	75.1603	4.9199		
	ADF	K-médias	77.3191	4.3163		
		ADFl	79.9026	2.6836		
		RNA	77.7462	17.7987		
	LETRAS	Fourier	K-médias	82.1045	7.8490	
ADFl			92.4180	10.4141		
RNA			69.2748	26.7667		
Wavelets		K-médias	74.1040	7.6103		
		ADFl	96.6713	0.6190		
		RNA	82.5917	24.9802		
ADF		K-médias	85.2322	8.6019		
		ADFl	96.0417	4.0627		
		RNA	80.2076	16.2951		
BRODATZ		Fourier	K-médias	50.6786	4.6154	
	ADFl		59.6500	14.1673		
	Wavelets		K-médias	58.4318	2.0282	
	Wavelets	ADFl	76.2220	12.9218		
		ADF	K-médias	53.2955	3.6705	
			ADFl	78.8716	4.3983	
	MPEG7		Fourier	K-médias	56.1917	0.4666
		ADFl		16.1786	2.6831	
		Wavelets		K-médias	56.1648	0.6659
			Wavelets	ADFl	40.6758	7.7189
ADF				K-médias	56.1339	0.5837
		ADF		ADFl	34.7321	8.7264
			PEIXES	Fourier	ADFl	3.2994
Wavelets					ADFl	71.7895
		ADF			ADFl	56.2770

classificação usada foi ADL, com divisão do conjunto de dados entre treinamento e teste (sem *leave-one-out*). Neste caso, também os descritores *wavelets* foram usados segundo o que se costuma encontrar mais comumente na literatura, ou seja, o uso do vetor de aproximação para cada nível de decomposição, além da média e da entropia dos detalhes até um determinado nível. As bases usadas neste caso foram de Brodatz, de letras com ruído misturado, de peixes com rotação e peixes com ruído.

Nota-se que neste experimento, os descritores ADF apresentaram a maior taxa de acerto para todas as bases. E mesmo em casos em que a vantagem de ADF é pequena como na base de peixes com rotação, nota-se que o melhor resultado foi obtido com um número bem menor de descritores ADF (80 contra 108), o que permite uma redução no tempo de processamento, sobretudo no caso de bases de dados com grande quantidade de amostras.

Tabela 4 – Tabela de melhores resultados usando-se ADL.

Base de Dados	Descritor	Maior acerto (%)	Número de descritores
BRODATZ	Fourier	98.5000	30
	Apr. wavelets	93.5000	11
	Det. wavelets	80.5000	8
	Ent. wavelets	85.0000	8
	ADF	99.5000	50
LETRAS	Fourier	95.9479	100
	Apr. wavelets	98.1022	108
	Det. wavelets	90.5248	10
	Ent. wavelets	89.3895	10
	ADF	99.4186	70
PEIXES (ROTAÇÃO)	Fourier	98.6715	20
	Apr. wavelets	99.7106	108
	Det. wavelets	97.0395	10
	Ent. wavelets	75.2193	10
	ADF	99.7796	80
PEIXES (RUÍDO)	Fourier	79.2545	20
	Apr. wavelets	79.4000	32
	Det. wavelets	76.5455	10
	Ent. wavelets	79.1091	10
	ADF	79.9091	10

7.5 Matrizes de Confusão

As Figuras 32, 33 e 34 mostram as chamadas matrizes de confusão para cada tipo de descritor usado, com a base de letras do alfabeto contendo todos os tipos de ruídos. As

matrizes foram tomadas para o método de classificação que proporcionou a maior taxa de acerto, segundo a Tabela 2 anterior, neste caso, o de Análise Discriminante Flexível.

A matriz de confusão mostra, para cada classe, o número de elementos que foram classificados na classe correta, na diagonal principal, e os elementos que foram classificados em outra classe, fora da diagonal. As amostras de teste da base foram tomadas aleatoriamente, assim que o número de elementos por classe varia dependendo da classe.

Observou-se um bom nível de acerto, havendo erro maior no caso de letras com traços mais parecidos, como K e Z, por exemplo, como já era esperado em um processo de classificação deste tipo.

-	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
C	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
L	0	0	0	0	0	0	0	0	0	1	0	21	0	0	0	0	0	0	0	4	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0
P	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	23	0	0	0	0	0	0	0	0
S	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	22	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0
Y	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27

Figura 32 – Matriz de confusão para a base de letras usando-se descritores de Fourier.

-	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B	0	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
C	0	0	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	1	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
I	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2
L	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	2	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	22	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	24	0	0	0	0	0	0
U	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	1
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19

Figura 33 – Matriz de confusão para a base de letras usando-se descritores wavelets.

-	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
C	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	3	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	28	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	24	0	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0
Z	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	25

Figura 34 – Matriz de confusão para a base de letras usando-se descritores ADF.

7.6 Análise de Componentes Principais

A idéia da Análise de Componentes Principais (ACP) é que um pequeno número de combinações lineares de um conjunto limitado de variáveis possa ser usado para representar o conjunto de valores de todas as variáveis originais. Desta forma, é possível uma redução da dimensionalidade do conjunto de dados com um mínimo possível de perda de informação. Os gráficos das Figuras 35 a 37 exibem a distribuição dos dois primeiros componentes principais (mais significativos) para a base shapes99, com os diferentes descritores usados: Fourier, *wavelets* e ADF. Cada classe é representada por um símbolo diferente no gráfico. Algumas classes foram circundadas por uma curva para facilitar a visualização. Observa-se uma maior separação inter-classes e menor intra-classes no caso de ADF, como se espera de um bom método de descritores para classificação.

Figura 35 – Componentes Principais usando-se descritores de Fourier.

Figura 36 – Componentes Principais usando-se descritores wavelets.

Figura 37 – Componentes Principais usando-se descritores ADF.

7.7 Síntese dos Resultados

A partir dos gráficos e tabelas anteriores, podem ser concluídos alguns pontos importantes a respeito dos descritores usados:

- A técnica de ADF apresentou um desempenho melhor que Fourier em todos os experimentos. Isto implica que o uso de Fourier pode ser substituído pelo uso de ADF em outras bases de dados, garantindo-se quase sempre um desempenho melhor na classificação.

- Em bases de dados com menos amostras, ADF levou vantagem sobre *wavelets*. Em bases maiores, ambas tiveram desempenho semelhante. Entretanto, observou-se que o resultado de ADF com poucos descritores foi satisfatório, sobretudo quando não se usou *leave-one-out*. Isto possibilita que o método de classificação forneça bons resultados para o experimento com um baixo custo de processamento, o que é muito útil, principalmente nestes casos em que o tamanho da base já costuma aumentar o tempo de processamento.
- Em bases de dados com alto nível de ruído e sem transformações geométricas associadas, como ocorre com a base de letras, a vantagem de ADF sobre os demais métodos é significativa. Este resultado é importante, uma vez que muitos conjuntos de dados são afetados por diferentes níveis de ruídos, oriundos das mais diversas fontes.
- O uso de ADF se mostrou vantajoso quando usado em combinação com Análise Discriminante Linear.

Em resumo, verifica-se que ADF apresenta um grande potencial para ser usada na geração de descritores de vetores DFM em diferentes tipos de bases de dados, sendo assim de grande importância para a área de reconhecimento de padrões como um todo.

8 *Conclusões*

Neste capítulo, são efetuadas as conclusões gerais acerca do trabalho e são apresentadas sugestões para trabalhos futuros relacionados ao estudo tratado, além dos trabalhos gerados a partir desta dissertação.

8.1 **Conclusões Sobre o Trabalho Realizado**

O estudo realizado nesta dissertação mostrou, analisou e elucidou vários pontos relacionados à aplicação de ADF para a geração de descritores de assinaturas DFM de objetos de interesse em imagens, como contornos, formas, texturas, entre outros.

Os resultados foram obtidos a partir da comparação entre quatro técnicas de classificação (K-médias, ADL, ADFI e RNA) e três técnicas de geração de descritores (Fourier, *wavelets* e ADF). Estes resultados demonstraram que o uso de descritores de ADF apresentou boa precisão no processo de classificação, sobretudo quando a base é afetada por ruído ou quando se usa classificação por ADL.

Em ADF, o vetor todo é tratado como uma função analítica geradora dos valores deste vetor e os descritores são os parâmetros desta função. No caso, estes descritores foram calculados a partir dos coeficientes de cada função-base usada na representação da função geradora. Desta forma, não há uma perda tão significativa de informação mesmo quando se usa um conjunto reduzido de descritores. Tal fato colabora substancialmente para o processo de classificação como um todo, tanto em termos de precisão como de custo de cálculo, ou seja, consegue-se com isso a obtenção de melhores resultados com um menor conjunto de descritores, implicando em menor tempo computacional.

Neste trabalho, os descritores foram extraídos a partir da técnica de DFM, que gerou um vetor de características a partir do objeto de interesse na imagem usando-se da teoria de geometria fractal. Com isto, um objeto complexo na imagem é transformado em um conjunto unidimensional de valores. Este conjunto é reduzido a partir da aplicação das

técnicas de descritores, possibilitando assim o processo de classificação. Desta forma, o uso de ADF associado a DFM pode ser empregado na solução de uma série de problemas de classificação e reconhecimento de padrões, sobretudo quando a base de dados é afetada por ruídos aleatórios, como acontece em muitos conjuntos de dados do mundo real.

O trabalho pode ser resumido no seguinte conjunto de pontos principais estudados:

- Apresentação e estudo de ADF como uma ferramenta que associada a DFM possibilita a obtenção de bons resultados para o processo de classificação de formas, contornos, textura, entre outros objetos de interesse em uma imagem.
- Comparação do uso de ADF ao uso de outras técnicas de descritores já conhecidas da literatura, tais como Fourier e *wavelets*. Esta comparação permitiu observar a vantagem de ADF em relação aos demais métodos, sobretudo quando são usados poucos descritores e quando a base de dados apresenta um alto nível de ruído, situação bastante comum em experimentos do mundo real.
- Estudo comparativo de diferentes técnicas de análise estatística e multivariada como os métodos de classificação por ADFI, RNA e o uso de PCA para uma melhor visualização da capacidade de agrupamento de cada descritor usado.
- Análise da relação entre a precisão do processo de classificação e o número de descritores usados como entrada para os métodos de classificação.

8.2 Considerações Finais

De todas as análises feitas, pode se concluir que a técnica de ADF pode ser usada como uma ferramenta poderosa na geração de descritores, possibilitando que o classificador produza resultados bons mesmo com o uso de poucos descritores ADF (menos de 30, por exemplo), o que garante um aumento significativo no desempenho computacional do processo de classificação como um todo. Em particular, pode ser observado que o uso de DFM com descritores ADF e classificação por RNA, ADFI ou ADL foi um conjunto que apresentou bons resultados para a maior parte das bases usadas em toda a análise.

8.3 Sugestões para Trabalhos Futuros

Devido aos resultados interessantes obtidos com o uso de ADF pode se ver que existe muito ainda a ser pesquisado sobre esta ferramenta estatística, principalmente em suas

aplicações relacionadas à geração de descritores para reconhecimento de padrões. Entre estes possíveis trabalhos, pode-se citar o uso de ADF em bases de aplicações distintas e mais específicas como, por exemplo, em sistemas de recuperação de imagens baseada em conteúdo (57), em bases de imagens coloridas, que podem ser tratadas separadamente em seus componentes de luminância e cromaticidade (42), o uso de ADF associado a outros métodos de assinatura (que não seja por DFM), o uso de ADF com outros tipos de funções-base e com outros parâmetros (mudando-se a ordem, usando derivadas, etc), entre tantas outras formas pelas quais se pode explorar todo o potencial da técnica.

Ainda diversas aplicações interessantes podem ser encontradas em Física. De um modo geral, a metodologia aqui abordada pode ser aplicada a qualquer problema que envolva reconhecimento de padrões. E muitos problemas na Física podem ser resolvidos por reconhecimento de padrões, como no estudo da radiação de fundo, na análise de cadeias de DNA, na detecção de partículas atômicas, no estudo de propriedades dos mais diversos tipos de materiais, entre muitos outros exemplos de aplicação.

8.4 Trabalhos gerados

Este mestrado gerou três artigos que estão em fase final de escrita e devem ser submetidos brevemente para revistas internacionais. O primeiro deles, intitulado *Functional Data Analysis Associated to Multi-scale Fractal Dimension for the Recognition of Shapes* apresenta ADF como técnica para geração de descritores de assinaturas DFM, com aplicação em classificação de formas e contornos, devendo ser submetido à revista *Pattern Recognition Letters* ou outra similar. No segundo, ADF combinado com DFM é usada em comparação com um grande número de outras estratégias usadas na classificação de formas e contornos. Este artigo será intitulado *Multi-scale Fractal Dimension Descriptors Comparison* e deverá ser submetido à revista *Journal of Computational and Applied Mathematics* ou outra similar. Já o terceiro aplica ADF na classificação de texturas de imagens em tons de cinza. Este será intitulado *Functional Data Analysis Associated to Multi-scale Fractal Dimension for Texture Pattern Recognition*, devendo ser submetido à revista *Pattern Recognition* ou outra similar.

Referências

- 1 FALCONER, K. J. *The geometry of fractal sets*. New York, NY, USA: Cambridge University Press, 1986. 162 p.
- 2 CARLIN, M. Measuring the complexity of non-fractal shapes by a fractal method. *Pattern Recognition Letters*, Oxford, v. 21, n. 11, p. 1013–1017, Oct. 2000.
- 3 PLOTZE, R. O. et al. Leaf shape analysis by the multiscale minkowski fractal dimension, a new morphometric method: a study in passiflora l. (passifloraceae). *Canadian Journal of Botany-Revue Canadienne de Botanique*, Montreal, v. 83, n. 3, p. 287–301, Mar. 2005.
- 4 WUNSCH, P.; LAINE, A. F. Wavelet descriptors for multiresolution recognition of handprinted characters. *Pattern Recognition*, Oxford, v. 28, n. 8, p. 1237–1249, Aug. 1995.
- 5 RAMSAY, J. O.; SILVERMAN, B. W. *Applied functional data analysis: methods and case studies*. New York: Springer-Verlag, 2002. 190 p.
- 6 RAMSAY, J. O.; SILVERMAN, B. W. *Functional data analysis*. New York: Springer-Verlag, 2005. 457 p.
- 7 HENDERSON, B. Exploring between site differences in water quality trends: a functional data analysis approach. *Environmetrics*, Chechester, UK, v. 17, n. 1, p. 65–80, Aug. 2005.
- 8 ROSSI, F. et al. Representation of functional data in neural networks. *Neurocomputing*, New York, v. 64, n. 1, p. 183–210, Mar. 2005.
- 9 BRIGHAM, E. O. *The fast Fourier transform and its applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988. 640 p.
- 10 SCHROEDER, M. *Fractals, chaos, power laws: minutes from an infinite paradise*. New York: W. H. Freeman and Company, 1996. 429 p.
- 11 MANDELBROT, B. B. *The Fractal geometry of nature*. New York: Freeman, 1968. 468 p.
- 12 GULICK, D. *Encounters with chaos*. New York: McGraw-Hill International Editions, 1992. 224 p. (Mathematics and Statistics Series).
- 13 PEITGEN, H. O.; SAUPE, D. *The science of fractal images*. Berlin: Springer, 1988. 312 p.

- 14 TAO, T.; VARGAS, A.; VEGA, L. A bilinear approach to the restriction and takeya conjectures. *Journal-American Mathematical Society*, Providence, US, v. 11, n. 4, p. 967–1000, Oct. 1998. Disponível em: <<http://www.ams.org/jams/1998-11-04/S0894-0347-98-00278-1/S0894-0347-98-00278-1.pdf>>. Acesso em: 08 dez. 2008.
- 15 FALCONER, K. J. Digital sundials, paradoxical sets, and Vitushkin's conjecture. *The Mathematical Intelligencer*, New York, v. 9, n. 1, p. 24–27, Jan. 1987.
- 16 BACKES, A. R.; BRUNO, O. M. A new approach to estimate fractal dimension of texture images. In: ICISP-INTERNATIONAL CONFERENCE ON IMAGE AND SIGNAL PROCESSING, 2008, Cherbourg-Octeville, Normandy, France. *Proceedings...* Berlin / Heidelberg: Springer, 2008. p. 136–143.
- 17 BACKES, R. C. C.; COSTA, L. F. The box-counting fractal. dimension: Does it provide an accurate subsidy for experimental shape characterization? if so, how to use it? In: SIBGRAPI - BRAZILIAN SYMPOSIUM ON COMPUTER GRAPHICS AND IMAGE PROCESSING, 1995, São Carlos. *Anais...* California, USA: IEEE Computer Society, 1995. p. 183–191.
- 18 TRICOT, C. *Curves and fractal dimension*. New York: Springer-Verlag, 1995. 323 p.
- 19 CORNFORTH, D.; JELINEK, H.; PEICHL, L. *Fractop: A Tool for Automated Biological Image Classification*. fev. 25 2003. Disponível em: <<http://citeseer.ist.psu.edu/567064.html>; http://life.csu.edu.au/~dcornfor/Fractop_v7.pdf>.
- 20 STOYAN, D.; STOYAN, H. Methods for the empirical determination of fractal dimension. In: _____. *Fractals, random shapes, and point fields*. 1. ed. Chichester: John Wiley & Sons, 1994. cap. 4, p. 39–45.
- 21 MORENCY, C.; CHAPLEAU., R. Fractal geometry for the characterisation of urban-related states: greater montreal case. *HarFA - Harmonic and Fractal Image Analysis*, Brno, Czech Republic, v. 1, n. 1, p. 30–34, Jan. 2003.
- 22 BACKES, A. R.; BRUNO, O. M. Fractal and multi-scale fractal dimension analysis: a comparative study of bouligand-minkowski method. *INFOCOMP Journal of Computer Science*, Lavras, MG, v. 7, n. 2, p. 74–83, Feb. 2008.
- 23 BACKES, A. R. *Implementação e comparação de métodos de estimativa da dimensão fractal e sua aplicação à análise e processamento de imagens*. 2006. 107 f. Dissertação (Mestrado em Ciência da Computação) — Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo, São Carlos, 2006.
- 24 FABBRI, R. et al. 2D euclidean distance transform algorithms: a comparative survey. *ACM Computing Surveys*, New York, NY, USA, v. 40, n. 1, p. 1–44, Feb. 2008.
- 25 BRUNO, O. M.; COSTA, L. da F. A parallel implementation of exact euclidean distance transform based on exact dilations. *Microprocessors and Microsystems*, Amsterdam, NL, v. 28, n. 3, p. 107–113, Apr. 2004.

- 26 SAITO, T.; TORIWAKI, J.-I. New algorithms for Euclidean distance transformations of an n -dimensional digitized picture with applications. *Pattern Recognition*, Oxford, v. 27, n. 11, p. 1551–1565, Nov. 1994.
- 27 MANOEL, E. T. M. et al. Multiscale fractal characterization of three-dimensional gene expression data. In: SIBGRAPI - BRAZILIAN SYMPOSIUM ON COMPUTER GRAPHICS AND IMAGE PROCESSING, 2002, Fortaleza. *Anais...* California, USA: IEEE Computer Society, 2002. p. 269–274.
- 28 EMERSON, C. W.; LAM, N. S. N.; QUATTROCHI, D. A. Multi-scale fractal analysis of image texture and patterns. *Photogrammetric Engineering and Remote Sensing*, Bethesda, MD, USA, v. 65, n. 1, p. 51–62, Jan. 1999.
- 29 BRUNO, O. M. et al. Fractal dimension applied to plant identification. *Information Sciences*, New York, v. 178, n. 12, p. 2722–2733, Jun. 2008.
- 30 COSTA, L. da F.; CESAR, R. M. J. *Shape analysis and classification: theory and practice*. Pennsylvania: CRC Press, 2000. 680 p.
- 31 WITKIN, A. P. Scale space filtering: a new approach to multi-scale descriptions. In: ICASSP - IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 1984, San Diego, California, USA. *Proceedings...* Saint Martin d'Hères, France: GRETSI, 2003. p. 79–95.
- 32 MARSHALL, S. Review of shape coding techniques. *Image and Vision Computing*, Newton, MA, USA, v. 7, n. 4, p. 281–294, Nov. 1989.
- 33 KØENDERINK, J. J.; DOORN, A. J. van. The singularities of the visual mapping. *Biological Cybernetics*, Berlin / Heidelberg, v. 24, p. 51–59, Mar. 1976.
- 34 MEYER, P. L. *Probabilidade - aplicações à estatística*. Rio de Janeiro: LTC, 2000. 426 p.
- 35 BOOR, C. D. The condition of the B-spline basis for polynomials. *SIAM Journal on Numerical Analysis*, Philadelphia, v. 25, n. 1, p. 148–152, Feb. 1988.
- 36 R Development Core Team. *R: A language and environment for statistical computing*. Vienna, Austria, 2005. Disponível em: <<http://www.R-project.org>>.
- 37 The MathWorks, Inc. *The Student Edition of MATLAB: Student User Guide*. pub-PH:adr: Prentice-Hall, 1992.
- 38 JAMES, G. M.; SUGAR, C. A. Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, New York, v. 98, n. 12, p. 397–408, Jun. 2003.
- 39 ROSSI, F.; CONAN-GUEZ, B.; GOLLI, A. E. Clustering functional data with the SOM algorithm. In: ESANN-EUROPEAN SYMPOSIUM ON ARTIFICIAL NEURAL NETWORKS, 2004, Bruges, Belgium. *Proceedings...* Brussels, Belgium: d-side publications, 2004. p. 305–312.
- 40 CASTLEMAN, K. R. *Digital image processing*. Englewood Cliffs, NJ, USA: Prentice Hall, 1996. 667 p.

- 41 SERRA, J. *Image analysis and mathematical morphology*. New York: Academic Press, 1982. 610 p.
- 42 GONZALEZ, R. C.; WOODS, R. *Digital image processing*. Reading, Massachusetts: Addison Wesley, 1992. 954 p.
- 43 CRIMMINS, T. R. A complete set of fourier descriptors for two-dimensional shapes. *IEEE Transactions on Systems, Man and Cybernetics*, New York, v. 12, n. 6, p. 848–855, Nov. 1982.
- 44 BRACEWELL, R. N. *The Fourier transform and its applications*. New York: McGraw-Hill, 2000. 640 p.
- 45 GOUPILLAUD, P.; GROSSMANN, A.; MORLET, J. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, Amsterdam,NL, v. 23, n. 1, p. 85–102, Oct. 1984.
- 46 GABOR, D. Theory of communication. *Journal of the Institute of Electronical Engeneering*, London, UK, v. 26, n. 93, p. 429–457, Nov. 1946.
- 47 WICKERHAUSER, M. V. Lectures on wavelet packet algorithms. In: LECTURE NOTES, INRIA - L'INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE, 1991, Paris. *Proceedings...* Paris: INRIA Press, 1991. p. 31–99.
- 48 DUDA, R. O.; HART, P. E. *Pattern classification*. New York: John Wiley and Sons, 2000. 654 p.
- 49 RIPLEY, B. R. *Pattern recognition and neural networks*. Cambridge,UK: Cambridge Univ. Press, 1996. 416 p.
- 50 HALKIDI, M.; BATISTAKIS, Y.; VAZIRGIANNIS, M. On clustering validation techniques. *Journal of Intelligent Information Systems*, Hingham, MA, USA, v. 17, n. 2-3, p. 107–145, Dec. 2001.
- 51 FRED, A. L. N. Finding consistent clusters in data partitions. In: INTERNATIONAL WORKSHOP ON MULTIPLE CLASSIFIER SYSTEMS, 2., 2001, Cambridge, UK. *Proceedings...* London, UK: Springer-Verlag, 2001. p. 309–318.
- 52 MINSKY, M. L. *Neural nets and the brain-model problem*. Tese (Doutorado) — Princeton University, Princeton, 1954.
- 53 RUMELHART; WIDROW; LEHR. The basic ideas in neural networks. *CACM: Communications of the ACM*, New York, NY, USA, v. 37, n. 3, Mar. 1994.
- 54 YAFFEE, R. A.; MACGEE, B. *Introduction to time series analysis and forecasting: with applications of SAS and SPSS*. Orlando, FL, USA: Academic Press, 2000. 628 p.
- 55 HASTIE, T.; TIBSHIRANI, R.; BUJA, A. Flexible discriminant and mixture models. In: NEURAL NETWORKS AND STATISTICS CONFERENCE, 1998, Edinburgh. *Proceedings...* Oxford, UK: Oxford University Press, 1995. p. 1–23.

- 56 MOKHTARIAN, F.; ABBASI, S.; KITTLER, J. Efficient and robust retrieval by shape content through curvature scale space. In: IDB-MMS-WORKSHOP ON IMAGE DATABASES AND MULTI-MEDIA SEARCH, 1996, Amsterdam,NL. *Proceedings...* New York: Springer, 1996. p. 35–42.
- 57 LEW, M. S. et al. Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, v. 2, n. 1, p. 1–19, fev. 2006.