
Visão artificial e morfometria na análise e
classificação de espécies biológicas

Rodrigo de Oliveira Plotze

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 09/12/2009

Assinatura: _____

Visão artificial e morfometria na análise e classificação de espécies biológicas

Rodrigo de Oliveira Plotze

Orientador: *Prof. Dr. Odemir Martinez Bruno*

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências de Computação e Matemática Computacional.

**São Carlos
2009**

A Fernanda, minha esposa,
você é minha inspiração.

Aos meus pais, Roberto e Marilda,
essa conquista também
é de vocês.

Agradecimentos

Ao professor Odemir M. Bruno,

Obrigado por acreditar no meu sonho. Pela maneira tão clara e simples de demonstrar qual a essência de ser um pesquisador. Pelos sábios ensinamentos e valiosos conselhos. Obrigado pela orientação e confiança. Minha profunda admiração!

Agradeço aos professores do ICMC-USP que foram essenciais na minha formação, em especial a Profa. Agma Traina e ao Prof. João Batista pelas diversas sugestões ao longo do desenvolvimento deste trabalho. A todos os funcionários do ICMC-USP que tornaram o *dia-a-dia* mais fácil e agradável. Vocês que na maioria das vezes estão por trás dos trabalhos, mas nem sempre recebem os méritos. Muito obrigado! A todos os colegas que conheci ao longo desses anos... E foram muitos! Se não fossem vocês e os períodos de descontração este trabalho nunca chegaria ao fim.

A Universidade de São Paulo e ao Instituto de Ciências Matemáticas e de Computação de São Carlos, pela excelente estrutura.

Meus sinceros agradecimentos ao Prof. Giancarlo Oliveira (ESALQ/USP), ao Prof. Luís Bernacci (IAC-Campinas), ao Juliano de Pádua (Embrapa) e a Carla Nardin, pelas valiosas contribuições para realização deste trabalho.

Aos companheiros de caminhada, Fernando Peres, Saulo Rosa, Márcio Cunha, Lucas Baggio, Avelino Pimenta, Jean Scatena, Jaciara Carosia, obrigado pelo apoio.

Aos meus familiares pelo carinho e compreensão. Aos meus pais Roberto e Marilda, por sempre me acompanhar (e torcer) ao longo dessa caminhada. O apoio de vocês foi essencial para esta conquista. À Fernanda, não tenho como expressar o quanto você foi importante para eu chegar até aqui! Sem você essa jornada seria impossível. Aos meus avós, minha profunda admiração. Vocês me ensinaram muito do que eu sei. Ao *Pingo*, meu cachorro, pelos momentos de descontração. Obrigado!

Agradeço a Deus por trilhar meus caminhos durante todos esses anos e, enfim, chegar até aqui. Que tua presença constante sempre ilumine meu futuro.

*Todas as teorias deveriam ser expressas
em uma linguagem tão simples que
até uma criança pudesse
entender.*

Albert Einstein

Resumo

PLOTZE, R.O. **Visão artificial e morfometria na análise e classificação de espécies biológicas**. 2009. 184f. Tese (Doutorado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2009.

A descoberta da história taxonômica e evolutiva das espécies é a principal fronteira das pesquisas científicas em diversas áreas do conhecimento. A biodiversidade dos indivíduos associada a grande variabilidade morfológica, torna essa tarefa um verdadeiro desafio. Os métodos tradicionais baseados na inspeção visual estão ultrapassados. Hoje em dia, os avanços tecnológicos têm colocado à disposição dos pesquisadores um arcabouço de ferramentas para o estudo das espécies. Este trabalho tem como objetivo investigar e desenvolver modelos computacionais capazes de analisar e classificar espécies biológicas por meio de características morfométricas. Para isso, técnicas de análise de imagens foram utilizadas para determinar a variabilidade das espécies em função de três informações de interesse: forma, estrutura tubular e textura. A potencialidade das metodologias foi avaliada por meio das seguintes espécies biológicas: maracujás silvestres do gênero *Passiflora*, eucaliptos do gênero *Eucalyptus* e arroz do gênero *Oryza*. Os experimentos produziram um conjunto de dados que representam uma detalhada descrição sobre a morfometria das espécies. Os resultados demonstraram que as técnicas de visão artificial são importantes para o estudo das espécies. As técnicas de análise de formas indicaram a viabilidade dessas metodologias na classificação das espécies, em particular, as redes complexas, a transformada de *wavelets* e a dimensão fractal multiescala que alcançaram altas taxas de classificações corretas. Os métodos desenvolvidos para análise de estruturas tubulares também demonstraram grande potencialidade na discriminação das espécies, principalmente a assinatura fractal multiescala a partir de pontos biométricos. As técnicas de análise de textura desenvolvidas também contribuíram para o estudo das espécies. Os resultados obtidos com as abordagens sinalizam que a relação entre biologia e computação é essencial para o desenvolvimento de metodologias eficientes. A inferência evolutiva das espécies foi um importante resultado obtido com os dados morfométricos, tanto a partir da forma, quanto da estrutura tubular e da textura. Testes estatísticos comprovaram a correlação entre os dados morfométricos obtidos por visão artificial e os dados moleculares de reconstrução filogenética. A multidisciplinaridade é o ponto central do trabalho, que está inserido na fronteira das áreas de visão artificial, morfometria e biologia. Com isso, essa simbiose resultou em promissoras contribuições para as áreas envolvidas.

Palavras-chave: visão artificial, morfometria, análise de imagens, taxonomia.

Abstract

PLOTZE, R.O. **Artificial vision and morfometry in analysis and classification of biological species**. 2009. 184f. Tese (Doutorado) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2009.

The discovery of taxonomic and evolutionary history of species is the main frontier of scientific research in various knowledge areas. The biodiversity of living things associated with the great morphological variability, makes this task a rightful challenge. The traditional methodologies based on visual inspection are totally outdated. Nowadays, technological advances have made available to researchers a framework of tools for the study of the species. This study aims to investigate and develop computer models to perform analysis and classification of biological species from morphometric features. For this, techniques of image analysis were used to determine the variability of the species in terms of three information of interest: shape, tubular structure and texture. The capability of the methods was evaluated by follows biological species: passion fruits of genus *Passiflora*, eucalyptus of genus *Eucalyptus* and rice of genus *Oryza*. The experiments produced a detailed dataset about the morphometric information of the species. The results showed that the techniques of artificial vision are demonstrably important to the study of the species. The shape analysis techniques indicated the viability of these methodologies in the species classification, in special the complex networks, the *wavelets* transform and the multiscale fractal dimension, have achieved high rates of correct discrimination. The new methods developed for analysis of tubular structures have also show great potential in species classification, especially the multiscale fractal signature from biometrics points. The techniques developed for texture analysis also show significant results in the investigation of the species. The results obtained with the approaches indicate that the relationship biology and computing is essential for the development of efficient methods. The inference of evolutionary tree of species was an important result obtained with the morphometric data collected from form, tubular structure and texture. Statistical tests showed a correlation between the morphometric data obtained by artificial vision and molecular data of phylogenetic reconstruction. A multidisciplinary approach is the focus of the work, which is related to the areas of artificial vision, morphometry and biology. Thus, this symbiosis has resulted in important contributions to the areas involved.

Keywords: artificial vision, morphometry, image analysis, taxonomy.

Lista de Figuras

Figura 1.1: Visão geral da arquitetura do sistema <i>TreeVis</i> (adaptado de (Plotze, 2004)).	24
Figura 1.2: Organização dos capítulos da tese.	27
Figura 3.1: Espécies de maracujás silvestres do gênero <i>Passiflora</i> : (a) <i>P. actinia</i> ; (b) <i>P. amethystina</i> ; (c) <i>P. caerulea</i> ; (d) <i>P. coriacea</i> ; (e) <i>P. foetida fluminensis</i> ; (f) <i>P. foetida foetida</i> ; (g) <i>P. gibertii</i> ; (h) <i>P. maliformis</i> ; (i) <i>P. miersii</i> ; (j) <i>P. organensis</i> ; (k) <i>P. pohlii</i> ; (l) <i>P. sidifolia</i> ; (m) <i>P. suberosa</i> .	37
Figura 3.2: Espécies de eucalipto do gênero <i>Eucalyptus</i> utilizadas nas análises: (a) <i>E. grandis</i> ; (b) <i>E. pellita</i> ; (c) <i>Corymbia torelliana</i> ; (d) <i>E. urophylla</i> ; (e) <i>E. grandis x E. camaldulensis</i> ; (f) <i>E. saligna x E. camaldulensis</i> ; (g) <i>E. urophylla x E. camaldulensis</i> ; (h) <i>E. urophylla x E. globulus</i> ; (i) <i>E. grandis x E. globulus</i> ; (j) <i>E. urophylla x E. grandis</i> .	39
Figura 3.3: Espécies de arroz do gênero <i>Oryza</i> utilizadas nos experimentos: (a) <i>O. alta</i> ; (b) <i>O. barthii</i> ; (c) <i>O. glaberrima</i> ; (d) <i>O. glumaepatula</i> ; (e) <i>O. grandiglumis</i> ; (f) <i>O. latifolia</i> ; (g) <i>O. longistaminata</i> ; (h) <i>O. nivara</i> ; (i) <i>O. officinalis</i> ; (j) <i>O. punctata</i> ; (k) <i>O. rufipogon</i> ; (l) <i>O. sativa</i> .	40
Figura 3.4: Resultado do processo para detecção do contorno das imagens biológicas.	42
Figura 3.5: Metodologia proposta para segmentação do sistema de venação foliar.	45
Figura 4.1: Assinaturas digitais das espécies biológicas utilizadas nos experimentos.	49
Figura 4.2: Análise da curvatura de formas biológicas. (a) representação unidimensional (1D); (b) curvograma do contorno; (c) representação tridimensional (3D).	51
Figura 4.3: Transformada de Fourier de sinais biológicos.	52
Figura 4.4: Estimativa da dimensão fractal multiescala em espécies biológicas.	56
Figura 4.5: Evolução dinâmica da estrutura da rede em função dos limiares T_i ; (a) $T_i = 0.050$; (b) $T_i = 0.100$; (c) $T_i = 0.125$; (d) $T_i = 0.150$; (e) $T_i = 0.175$; (f) $T_i = 0.200$ (adaptado de (Backes, Casanova <i>et al.</i> , 2008)).	58
Figura 4.6: Decomposição de sinal <i>ID</i> por meio da transformada de wavelets.	62
Figura 4.7: Principais famílias de wavelets mães: Daubechies (db), Coiflets (coif) e Symlets (sym).	63
Figura 4.8: Análise de formas biológicas com a transformada de wavelets. Processo de extração de tendência e detalhe para a família <i>db4</i> com o nível 5.	63
Figura 4.9: Extração de características morfométricas de estruturas tubulares.	65
Figura 4.10: Determinação do pixel de bifurcação na estrutura tubular. (a) O pixel ρ não é uma bifurcação, pois $\mathcal{A}\rho \leq 2$; (b) O pixel ρ é uma bifurcação $\mathcal{A}\rho > 2$.	67
Figura 4.11: Extração das características biométricas da estrutura tubular: determinação das bifurcações e dos pontos extremos.	67

Figura 4.12: Determinação dos ângulos entre as ramificações da estrutura tubular: ângulo de abertura (α_1); ângulo de fechamento (α_2).....	71
Figura 4.13: Assinatura fractal multiescala de estruturas tubulares: (a) estrutura tubular; (b) dilatações exatas; (c) curva $\log(A(d)) \times \log(d)$; (d) assinatura fractal multiescala.	72
Figura 4.14: Assinatura fractal multiescala baseada em descritores biométricos: (a) assinatura gerada em função dos pontos biométricos de bifurcação; (b) cálculo da assinatura a partir dos pontos extremos e (c) combinação das características biométricas das bifurcações e dos pontos extremos para produção da assinatura fractal multiescala.....	73
Figura 4.15: Representação de uma estrutura tubular por meio de uma rede complexa.	74
Figura 4.16: Modelagem de uma estrutura tubular por meio de uma rede complexa a partir de bifurcações e pontos extremos: (a) estrutura tubular; (b) características biométricas; (c) rede complexa para $Ti=0.50$	75
Figura 4.17: Histograma de uma imagem em níveis de cinza: (a) imagem original; (b) histograma e (c) distribuição acumulativa da probabilidades dos níveis de cinza.	77
Figura 4.18: Seleção de um subconjunto de pixels de textura com base na distribuição de probabilidade de níveis de cinza: (a) imagem original em níveis de cinza; (b) seleção para $k=1.0$; (c) seleção para $k=0.9$; (d) seleção para $k=0.8$; (e) seleção para $k=0.7$; (f) seleção para $k=0.6$; (g) seleção para $k=0.5$	78
Figura 4.19: Análise de textura com base na distribuição de níveis de cinza e assinatura fractal multiescala. (a) forma biológica em escala de cinza; (b) seleção do subconjunto para $k=1.0$; (c) dilatações exatas; (d) assinatura fractal multiescala; (e) dilatações exatas em detalhe.	79
Figura 4.20: Análise de textura com base na distribuição de níveis de cinza e redes complexas. (a) forma biológica em escala de cinza; (b) seleção do subconjunto para $k=1.0$; (c) $Ti = 0.025$; (d) $Ti = 0.075$; (e) $Ti = 1.250$	79
Figura 5.1: Comparação entre neurônio biológico e neurônio artificial.	88
Figura 5.2: Arquitetura de uma rede neural do tipo <i>perceptron</i> multicamada (MLP).....	90
Figura 5.3: Estrutura de uma matriz de confusão.	92
Figura 6.1: Estrutura organizacional das técnicas de análise de imagens utilizadas nos experimentos.....	97
Figura 6.2: Curvatura multiescala de uma forma biológica para diferentes valores do desvio padrão da gaussiana: (a) $a = 50$; (b) $a = 20$ e (c) $a = 10$	99
Figura 6.3: Descritores de Fourier: (a) $N = 10$; (b) $N = 20$; (c) $N = 40$; (d) $N = 80$; (e) $N = 160$ e (f) $N = 320$	101
Figura 6.4: Funções fractal multiescala obtidas a partir dos valores da ordem de derivação e da suavização gaussiana.	102
Figura 6.5: Principais características extraídas a partir de uma rede complexa: (a) grau médio distribuição de conexões; (b) coeficiente de aglomeração; (c) energia; (d); entropia; (e) grau máximo; (f) grau médio; (g); graus dos vértices para $Ti = 0.15$; (h) distribuição de conexões para $Ti = 0.15$	104
Figura 6.6: Assinatura fractal multiescala de estruturas tubulares para $k=250$ e $\sigma=250$. (a) todos os pontos (b) pontos de bifurcação; (c) pontos extremos; (d) combinação das bifurcações e dos pontos extremos.	107

Figura 6.7: Visão estrutural das técnicas de reconhecimento de padrões utilizadas nos experimentos.....	111
Figura 6.8: Processo de reconhecimento de padrões adotado nos experimentos.	112
Figura 6.9: Processo de classificação do conjunto de dados de treinamento com as técnicas de reconhecimento de padrões	113
Figura 7.1: Resultado da classificação das espécies utilizando SRDA com as características da assinatura digital. (a) maracujá; (b) eucalipto e (c) arroz.....	118
Figura 7.2: Resultado da classificação utilizando LDA com as características da curvatura. (a) maracujá; (b) eucalipto e (c) arroz.....	120
Figura 7.3: Resultado da classificação das espécies com as características dos descritores Fourier. (a) maracujá (SRDA); (b) eucalipto (LDA) e (c) arroz (SRDA).	123
Figura 7.4: Curvas fractal multiescala geradas a partir da análise da forma das espécies de maracujá.	125
Figura 7.5: Resultado da classificação das espécies com as características da dimensão fractal multiescala. (a) maracujá (LDA); (b) eucalipto (SRDA) e (c) arroz (SRDA).	126
Figura 7.6: Representação dos graus máximos das espécies de maracujá.	129
Figura 7.7: Resultado da classificação das espécies com as características das redes complexas. (a) maracujá (LDA); (b) eucalipto (SRDA) e (c) arroz (LDA).....	130
Figura 7.8: Resultado da classificação das espécies com as características da transformada de wavelets. (a) maracujá (LDA); (b) eucalipto (SRDA) e (c) arroz (LDA).....	133
Figura 7.9: Gráficos ROC gerado a partir dos resultados da classificação das espécies por meio da seleção de características: (a) maracujá (LDA); (b) eucalipto (SRDA); (c) arroz (LDA).....	134
Figura 7.10: Gráficos ROC gerado a partir dos resultados da classificação das espécies por meio da seleção de características das redes complexas: (a) maracujá (SRDA); (b) eucalipto (SRDA); (c) arroz (LDA).....	136
Figura 7.11: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados biométricos.	138
Figura 7.12: Representação das assinaturas fractal multiescala geradas a partir dos dados da estrutura tubular das espécies de maracujá.	140
Figura 7.13: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados da assinatura fractal multiescala.	141
Figura 7.14: Representação dos coeficientes de aglomeração das estruturas tubulares das espécies de maracujás.	143
Figura 7.15: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados das redes complexas.	143
Figura 7.16: Gráficos ROC gerados a partir do desempenho dos classificadores com os dados da seleção de características. (a) classificador DM; (b) classificador LDA; (c) classificador SRDA e (d) rede neural artificial.	145
Figura 7.17: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados da matriz de coocorrência.....	147
Figura 7.18: Assinaturas fractal multiescala geradas a partir da análise de textura.	149
Figura 7.19: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados da assinatura fractal para $k = 1.0$ e classificador LDA.	150

Figura 7.20: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados da das redes complexas e classificador LDA.	151
Figura 7.21: Gráficos ROC gerados a partir do desempenho dos classificadores com os dados da seleção de características. (a) classificador DM; (b) classificador LDA; (c) classificador SRDA e (d) rede neural artificial.	153
Figura 7.22: Gráficos ROC gerados a partir do desempenho dos classificadores com os dados da combinação de características. (a) classificador DM; (b) classificador LDA e (c) classificador SRDA.....	155
Figura 7.23: Inferência evolutiva das espécies de maracujás a partir da análise de formas. .	157
Figura 7.24: Inferência evolutiva das espécies de maracujás a partir da análise de estruturas tubulares.	158
Figura 7.25: Inferência evolutiva das espécies de maracujás a partir da análise de textura. .	159
Figura 7.26: Inferência evolutiva das espécies de eucalipto a partir da análise de formas com a transformada de wavelets.	161
Figura 7.27: Separabilidade das espécies de eucalipto: (a) espécies puras; (b) espécies híbridas.	162
Figura 7.28: Inferência evolutiva das espécies de arroz a partir da análise de formas com a transformada de <i>wavelets</i>	163

Lista de Tabelas

Tabela 3.1: Espécies vegetais do gênero <i>Passiflora</i> utilizadas nas análises.	37
Tabela 3.2: Espécies do gênero <i>Eucalyptus</i> utilizadas nas análises.	38
Tabela 3.3: Espécies de arroz do gênero <i>Oryza</i> utilizadas nas análises.	40
Tabela 3.4: Valores em pixels utilizados para normalização das imagens biológicas.	41
Tabela 5.1: Métodos aglomerativos de agrupamento hierárquico.	87
Tabela 5.2: Funções de ativação dos neurônios artificiais.	88
Tabela 6.1: Organização dos experimentos de análise de imagens empregados nas espécies biológicas.	96
Tabela 6.2: Identificadores utilizados para referenciar as espécies biológicas.	96
Tabela 6.3: Experimentos realizados com base na assinatura digital das imagens biológicas.	98
Tabela 6.4: Experimentos realizados com a curvatura das espécies biológicas.	100
Tabela 6.5: Experimentos realizados com os descritores de Fourier.	101
Tabela 6.6: Experimentos realizados com a dimensão fractal multiescala.	103
Tabela 6.7: Experimentos realizados com as redes complexas.	105
Tabela 6.8: Experimentos realizados com a transformada de wavelets.	106
Tabela 6.9: Experimentos realizados com a assinatura fractal multiescala das estruturas tubulares.	107
Tabela 6.10: Experimentos realizados com as redes complexas das estruturas tubulares.	108
Tabela 6.11: Experimentos realizados com análise estatística da textura.	109
Tabela 6.12: Experimentos realizados com análise de textura baseada na distribuição de níveis de cinza.	110
Tabela 7.1: Resultados da classificação das espécies com características da assinatura digital.	117
Tabela 7.2: Detalhamento dos resultados da classificação das espécies com os dados da assinatura digital.	117
Tabela 7.3: Resultados da classificação das espécies com base em características curvatura.	119
Tabela 7.4: Detalhamento dos resultados da classificação das espécies com os dados da curvatura.	119
Tabela 7.5: Resultados da classificação das espécies com as características dos descritores de Fourier.	121
Tabela 7.6: Detalhamento dos resultados da classificação das espécies com os descritores de Fourier.	122

Tabela 7.7: Resultados da classificação com as características da dimensão fractal multiescala.	123
Tabela 7.8: Detalhamento dos resultados da classificação das espécies com a dimensão fractal multiescala.	124
Tabela 7.9: Resultados da classificação das espécies com as características das redes complexas.	127
Tabela 7.10: Detalhamento dos resultados da classificação das espécies com as redes complexas.	128
Tabela 7.11: Resultados da classificação das espécies com as características extraídas das <i>wavelets</i>	131
Tabela 7.12: Resultados da classificação com as características das <i>wavelets</i>	131
Tabela 7.13: Detalhamento dos resultados da classificação das espécies com as <i>wavelets</i>	132
Tabela 7.14: Resultados da classificação das espécies com a seleção de características.	133
Tabela 7.15: Resultados da classificação das espécies com os dados selecionados dos conjuntos das redes complexas.	135
Tabela 7.16: Resultados da classificação das espécies com os dados biométricos.	137
Tabela 7.17: Matriz de confusão e dados estatísticos resultantes do experimento com dados biométricos e o classificador LDA.	137
Tabela 7.18: Resultados da classificação das espécies com a assinatura fractal multiescala.	138
Tabela 7.19: Matriz de confusão e dados estatísticos resultantes do experimento com a assinatura fractal multiescala.	139
Tabela 7.20: Resultados da classificação das espécies com as redes complexas.	141
Tabela 7.21: Matriz de confusão do experimento E131 com redes complexas e estruturas tubulares.	142
Tabela 7.22: Resultados da classificação das espécies com os dados da seleção de características.	144
Tabela 7.23: Resultados da classificação das espécies com os dados da seleção de características.	146
Tabela 7.24: Matriz de confusão e dados estatísticos resultantes do experimento T005.	146
Tabela 7.25: Resultados da classificação por meio da análise de textura a partir da assinatura fractal.	147
Tabela 7.26: Matriz de confusão e dados estatísticos resultantes do experimento com assinatura fractal gerada a partir da distribuição de níveis de cinza.	148
Tabela 7.27: Resultados da classificação por meio da análise de textura a partir das redes complexas.	150
Tabela 7.28: Matriz de confusão e dados estatísticos resultantes do experimento com redes complexas.	151
Tabela 7.29: Resultados da classificação das espécies com os dados da seleção de características.	152
Tabela 7.30: Resultados da classificação das espécies com os dados combinados.	154
Tabela 7.31: Resultados da correspondência entre dados morfométricos e moleculares obtidos com o teste estatístico de Mantel.	156
Tabela 7.32: Principais resultados da análise morfométrica das espécies biológicas.	164

Sumário

1. INTRODUÇÃO.....	21
1.1 PROJETO <i>TREEVIS</i> – TREE VISION SYSTEM	23
1.2 OBJETIVOS	25
1.3 ORGANIZAÇÃO DOS CAPÍTULOS.....	25
2. ASPECTOS BIOLÓGICOS: EVOLUÇÃO, FILOGENIA E MORFOMETRIA....	29
2.1 EVOLUÇÃO DAS ESPÉCIES E BIODIVERSIDADE	29
2.2 ANÁLISES FILOGENÉTICAS.....	31
2.3 ANÁLISES MORFOMÉTRICAS	31
2.4 CONSIDERAÇÕES FINAIS	33
3. AQUISIÇÃO E PROCESSAMENTO DE IMAGENS BIOLÓGICAS	35
3.1 MATERIAL BIOLÓGICO	36
3.1.1 <i>Espécies de Maracujás (Passiflora)</i>	36
3.1.2 <i>Espécies de Eucalipto (Eucalyptus)</i>	38
3.1.3 <i>Espécies de Arroz (Oryza)</i>	39
3.2 AQUISIÇÃO	40
3.3 PROCESSAMENTO	41
3.3.1 <i>Segmentação do Sistema de Venação Foliar</i>	43
3.4 CONSIDERAÇÕES FINAIS	45
4. ANÁLISE DE IMAGENS BIOLÓGICAS.....	47
4.1 ANÁLISE DE FORMAS.....	47
4.1.1 <i>Assinaturas Digitais</i>	48
4.1.2 <i>Curvatura</i>	49
4.1.3 <i>Transformada e Descritores de Fourier</i>	51
4.1.4 <i>Dimensão Fractal Multiescala</i>	53
4.1.5 <i>Redes Complexas</i>	56
4.1.6 <i>Wavelets</i>	60
4.2 ANÁLISE DE ESTRUTURAS TUBULARES	64
4.2.1 <i>Biometria</i>	64
4.2.2 <i>Extração Automática de Características Biométricas</i>	65
4.2.3 <i>Assinatura Fractal Multiescala</i>	71
4.2.4 <i>Modelagem de Estruturas Tubulares por Redes Complexas</i>	73
4.3 ANÁLISE DE TEXTURA	75
4.3.1 <i>Textura Estatística</i>	75
4.3.2 <i>Textura por Distribuição de Níveis de Cinza</i>	77
4.4 CONSIDERAÇÕES FINAIS	80

5. RECONHECIMENTO DE PADRÕES	81
5.1 FUNDAMENTOS.....	81
5.2 ANÁLISE ESTATÍSTICA MULTIVARIADA	82
5.2.1 Classificador Euclidiano por Distância Mínima.....	83
5.2.2 Análise Discriminante Linear	83
5.2.3 Análise Discriminante por Regressão Espectral.....	84
5.2.4 Análise de Componentes Principais	86
5.2.5 Análise de Agrupamentos.....	86
5.3 REDES NEURAIIS ARTIFICIAIS	87
5.3.1 Perceptron Multicamadas (MLP).....	89
5.4 SELEÇÃO DE CARACTERÍSTICAS.....	91
5.5 AVALIAÇÃO DOS CLASSIFICADORES	91
5.6 CONSIDERAÇÕES FINAIS	93
6. EXPERIMENTOS	95
6.1 ESPÉCIES BIOLÓGICAS	95
6.2 ANÁLISE DE IMAGENS.....	97
6.2.1 Análise de Formas: Assinaturas.....	98
6.2.2 Análise de Formas: Curvatura.....	99
6.2.3 Análise de Formas: Descritores de Fourier.....	100
6.2.4 Análise de Formas: Dimensão Fractal Multiescala	101
6.2.5 Análise de Formas: Redes Complexas.....	103
6.2.6 Análise de Formas: Wavelets.....	105
6.2.7 Estruturas Tubulares: Biometria.....	106
6.2.8 Estruturas Tubulares: Assinatura Fractal	107
6.2.9 Estruturas Tubulares: Redes Complexas	108
6.2.10 Textura: Análise Estatística	109
6.2.11 Textura: Distribuição de Níveis de Cinza	110
6.3 RECONHECIMENTO DE PADRÕES.....	111
6.4 CONSIDERAÇÕES FINAIS	114
7. RESULTADOS E DISCUSSÃO.....	115
7.1 RESULTADOS DA CLASSIFICAÇÃO BASEADA NA ANÁLISE DE FORMAS	116
7.2 RESULTADOS DA CLASSIFICAÇÃO BASEADA NA ANÁLISE DE ESTRUTURAS TUBULARES	136
7.3 RESULTADOS DAS CARACTERÍSTICAS MORFOMÉTRICAS DAS TEXTURAS.....	145
7.4 COMBINAÇÃO DE CARACTERÍSTICAS MORFOMÉTRICAS	153
7.5 INFERÊNCIA EVOLUTIVA DAS ESPÉCIES	155
7.6 DISCUSSÃO.....	163
7.6.1 Análise de Formas.....	164
7.6.2 Análise de Estruturas Tubulares	166
7.6.3 Análise de Textura.....	167
7.6.4 Seleção de Características.....	167
7.7 CONSIDERAÇÕES FINAIS	168
8. CONCLUSÃO.....	171
8.1 CONTRIBUIÇÕES	174
8.2 TRABALHOS FUTUROS	176
REFERÊNCIAS BIBLIOGRÁFICAS	177

1

Introdução

As técnicas de visão artificial exercem papel fundamental em muitas pesquisas na área de ciências biológicas. Esses métodos têm como principal objetivo a descoberta de conhecimento a partir de informações visuais. O grande desafio diz respeito à questão da multidisciplinaridade, uma vez que, quase a totalidade dos trabalhos desenvolvidos tem algum tipo de interação com outra área de conhecimento. Por outro lado, essa simbiose entre visão artificial e biologia resulta em importantes progressos para as áreas envolvidas.

O ponto central das pesquisas em biologia está relacionado ao estudo e comparação das características morfométricas dos organismos. O entendimento da biodiversidade, principalmente visando conhecer a evolução e taxonomia das espécies, está associado à compreensão das complexas imagens biológicas (Adams, Rohlf *et al.*, 2004).

A morfometria é uma área multidisciplinar relacionada à visão artificial. As técnicas morfométricas utilizam avançados modelos matemáticos e estatísticos para explicar fenômenos biológicos (Monteiro e Reis, 1999). Essas técnicas representam um novo paradigma para o estudo da evolução das formas biológicas e suas correlações com outras variáveis (Rohlf, 2002). Diversas aplicações de morfometria podem ser encontradas nas áreas de ecologia (Loy, Mariani *et al.*, 1998; Adams e Rohlf, 2000), antropologia (Delson, Harvati *et al.*, 2001; Frieb, 2003), botânica (Sonibare, Jayeola *et al.*, 2004; Plotze, Falvo *et al.*, 2005), entre outras.

Os dados morfométricos proveem uma abrangente descrição das formas biológicas e têm sido amplamente empregados em estudos evolutivos e taxonômicos (Wiens, 2004; Macleod, 2007). A análise filogenética é uma importante abordagem para inferência das mudanças evolutivas das espécies (Futuyma, 2002). A filogenia estuda a história evolutiva de uma espécie, bem como a relação de parentesco ou ancestralidade entre determinados indivíduos. O princípio defendido por *Darwin* de que todos os organismos descenderam (com modificação) de ancestrais comuns é fundamental para as reconstruções filogenéticas de espécies (Darwin, 1859; Futuyma, 2002). Os estudos taxonômicos têm como principal desafio a descoberta e descrição das milhares de espécies que ainda não foram catalogadas. Nos dias atuais, o trabalho taxonômico é essencial, principalmente devido ao aumento constante das taxas de extinção de espécies, além dos distúrbios e degradações dos ecossistemas. Quanto maior o conhecimento taxonômico, melhor será a fundamentação filogenética das espécies (Wheeler, 2007).

Os trabalhos recentes demonstram grande aplicabilidade dos dados morfométricos na filogenia e taxonomia das espécies. No entanto, tradicionalmente, esses dados são coletados por meio de inspeções visuais da morfologia das espécies. As características das espécies são avaliadas qualitativamente e a diversidade é identificada a partir de complexos padrões de variação. Essa metodologia está ultrapassada, além disso, esse procedimento é lento e pouco confiável. A literatura contemporânea é repleta de informações inconsistentes e identificações incorretas (Macleod, 2007).

Com o avanço das técnicas de biologia molecular, a inclusão de dados genéticos tem resultado em importantes contribuições para o estudo das espécies. Para reconstrução filogenética de uma determinada espécie utilizando dados moleculares (DNA, RNA ou proteínas) é imprescindível o correto alinhamento das sequências (Pádua, 2004). Um ponto desfavorável das técnicas moleculares é a análise filogenética de táxons extintos, uma vez que para estudos genéticos as evidências materiais são indispensáveis. Outro fator importante é que quase a totalidade das espécies conhecidas na literatura foi identificada por meio de análises morfológicas, e não por meio de análises moleculares. Dessa forma, os dados morfológicos têm demonstrado maior eficácia no estudo das espécies (Wiens, 2004).

Dentro deste contexto, a análise morfométrica se mostra fundamentalmente importante para o progresso das pesquisas sobre a história taxonômica e evolutiva dos organismos vivos. Os avanços constantes da arquitetura dos computadores, bem como inovações na construção

de sistemas computacionais, têm colocado a disposição dos pesquisadores um arcabouço de ferramentas para o estudo das espécies. Dessa forma, o desenvolvimento de técnicas de visão artificial aplicadas na análise morfométrica das espécies, se faz cada vez mais necessário. As técnicas de visão artificial fornecem todo formalismo matemático necessário para uma análise quantitativa rápida e eficiente das espécies. A confiabilidade das informações a respeito da estrutura organizacional das espécies é indispensável para construção da “árvore da vida”.

1.1 Projeto *TreeVis* – Tree Vision System

O projeto *TreeVis* tem como objetivo o desenvolvimento de um sistema computacional contendo técnicas de visão artificial empregadas na análise e extração de características de espécies biológicas. Esse projeto é desenvolvido por pesquisadores do Instituto de Física de São Carlos (IFSC) e do Instituto de Ciências Matemáticas e de Computação (ICMC) da Universidade de São Paulo do campus de São Carlos. O projeto conta com colaborações de diversos pesquisadores de outros centros tais como, a Empresa Brasileira de Pesquisa Agropecuária (Embrapa), a Escola Superior Luiz de Queiroz (ESALQ/USP), o Instituto Agrônomo de Campinas (IAC), a Universidade Estadual Paulista (UNESP-Assis), dentre outros.

A proposta do sistema *TreeVis* é fornecer uma ferramenta capaz de armazenar, classificar e organizar dados científicos, afim de, realizar a identificação automática das espécies biológicas. Para isso o sistema trabalha com a exploração sistemática de um vasto conjunto de características das espécies. O grande número de atributos necessários é justificado principalmente pela diversidade das espécies, bem como a enorme variabilidade de indivíduos dentro de um mesmo táxon. As principais características exploradas no *TreeVis* estão organizadas em cinco grupos: geometria (Plotze, 2004), cor (Junior, 2008), forma (Falvo, 2005), textura (Casanova, 2008) e complexidade (Plotze, 2004; Backes, 2006).

A arquitetura do sistema *TreeVis* é composta por cinco módulos principais, sendo eles: estrutura de amostragem, controle central, extração de características, treinamento/identificação e base de dados. A Figura 1.1 apresenta uma visão geral do sistema *TreeVis*.

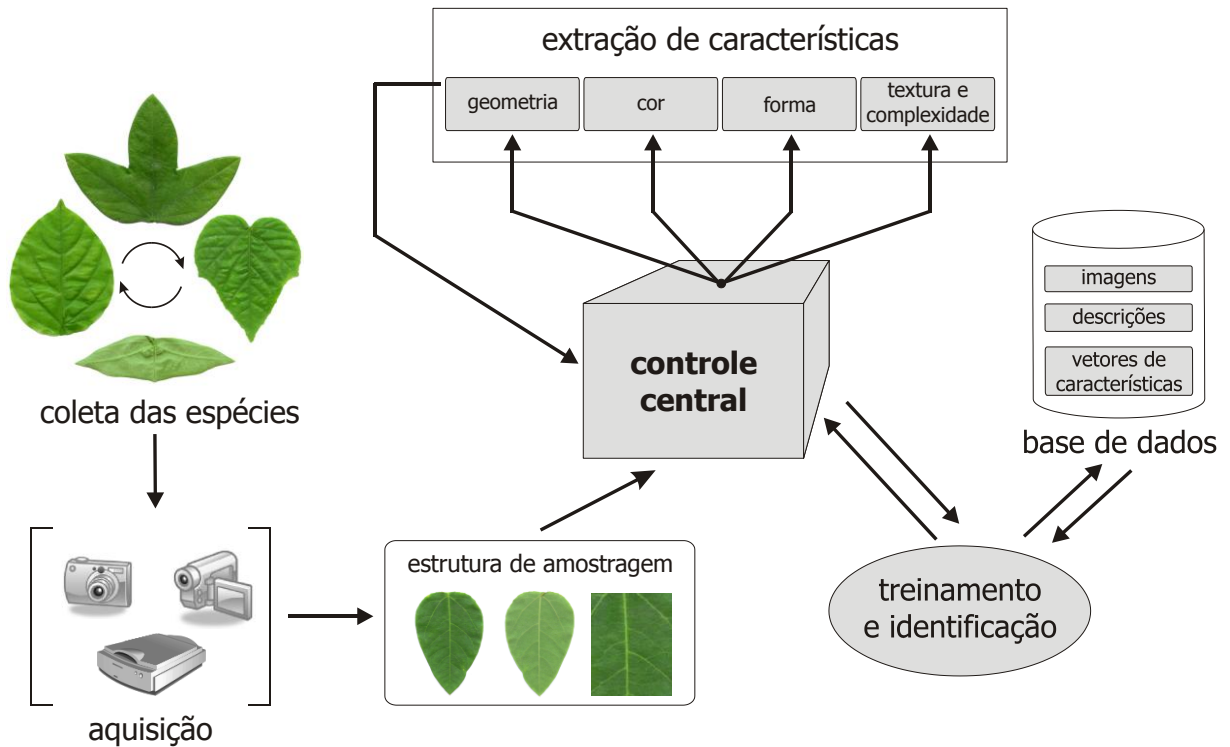


Figura 1.1: Visão geral da arquitetura do sistema *TreeVis* (adaptado de (Plotze, 2004)).

As informações biológicas são inseridas no sistema por meio de imagens digitais, que podem ser fotografadas ou digitalizadas (*scanner*). A alta resolução das imagens é fundamental para o correto funcionamento do sistema. Para uma mesma espécie biológica, diversos tipos de imagens podem ser inseridos no sistema, por exemplo, para uma folha é inserido uma imagem do lado exposto ao sol, uma imagem do lado de sombra, além de uma imagem ampliada (utilizada para análise de textura). A amostra, ou conjunto de amostras, de uma espécie é definido no sistema como estrutura de amostragem.

O módulo principal do sistema é denominado controle central. Ele funciona como maestro do sistema, mediando todas as informações que são trocadas entre os demais módulos. O módulo de extração de características é responsável pela análise das imagens biológicas. Neste módulo, as informações coletadas são estruturadas vetores de características que são utilizados para a identificação das espécies. O módulo de treinamento/identificação aplica técnicas de reconhecimento de padrões, nos vetores de características, visando à classificação das espécies. Como resultado, o sistema é capaz de apresentar gráficos e informações estatísticas a respeito das espécies analisadas.

1.2 Objetivos

Este trabalho está inserido na fronteira entre as áreas de visão artificial, morfometria e biologia. O objetivo principal foi estudar e desenvolver modelos computacionais capazes de extrair e analisar características morfométricas de espécies biológicas. Esses modelos computacionais foram integrados as metodologias já existentes no sistema *TreeVis*.

As espécies biológicas foram analisadas com base em três grupos de técnicas computacionais: formas, estruturas tubulares e textura. Para cada um dos grupos, um conjunto extenso de técnicas foi estudado e novas metodologias foram propostas. Métodos de reconhecimento de padrões foram empregados para avaliar a qualidade das técnicas morfométricas estudadas e desenvolvidas.

A organização das fases de desenvolvimento do trabalho foi dividida em quatro módulos principais: (i) aspectos biológicos – em que foram estudados todos os conceitos biológicos, tais como morfometria, filogenia, taxonomia, além de técnicas estatísticas multivariadas empregadas na taxonomia das espécies; (ii) desenvolvimento – no qual as técnicas de análise morfométrica foram implementadas e testadas, além disso, novos métodos foram propostos; (iii) experimentos – com base no entendimento biológico, e com as técnicas computacionais desenvolvidas, foram realizadas análises morfométricas com espécies biológicas reais; (iv) análise dos resultados – neste módulo as informações coletadas por meio dos experimentos foram avaliadas por meio de técnicas de reconhecimento de padrões.

Para avaliar as técnicas de análise morfométricas e qualificar a potencialidade da metodologia, foram utilizados três gêneros de espécies biológicas: (i) folhas das espécies de maracujás silvestres do gênero *Passiflora*; (ii) folhas das espécies de eucalipto do gênero *Eucalyptus* e (iii) grãos das espécies de arroz do gênero *Oryza*.

1.3 Organização dos Capítulos

No capítulo 2 (Aspectos Biológicos: Evolução, Filogenia e Morfometria) são apresentados os principais aspectos biológicos relacionados ao trabalho. São relatados os conceitos fundamentais da teoria da evolução defendida por *Charles Darwin*, bem como, o progresso das pesquisas até os dias atuais. Também são descritas as técnicas de análise morfométrica do

ponto de vista biológico, desde os modelos tradicionais até os mais recentes conhecidos como “nova morfometria”.

O capítulo 3 (Aquisição e Processamento de Imagens Biológicas) apresenta as informações a respeito das espécies biológicas utilizadas nos experimentos, bem como, os procedimentos utilizados para aquisição e processamento das imagens. Neste capítulo é detalhada a organização das espécies, seus respectivos gêneros, além do número de exemplares para cada espécie. Os aspectos relativos a digitalização do material biológico é descrito, além das técnicas empregadas para filtragem, normalização e segmentação das imagens. Por fim, uma nova metodologia para segmentação do sistema de venação foliar é proposta.

O capítulo 4 (Análise de Imagens Biológicas) detalha as principais técnicas de visão artificial que foram utilizadas no projeto, bem como as novas metodologias propostas. Essas técnicas foram divididas em três grupos de análise: formas, estruturas tubulares e textura. Além disso, é demonstrada a aplicabilidade desses métodos na análise e extração de características de materiais biológicos.

No capítulo 5 (Reconhecimento de Padrões) são descritos os principais conceitos sobre classificação de padrões utilizando análise estatística multivariada e redes neurais artificiais. Além disso, são apresentados métodos para seleção de características e para avaliação dos classificadores. Os métodos de reconhecimento de padrões foram necessários para avaliar a potencialidade das características morfométricas das espécies.

O capítulo 6 (Experimentos) apresenta detalhadamente como foram executados os experimentos com as espécies biológicas. Neste capítulo são descritas as metodologias empregadas para cada tipo de análise, bem como os parâmetros adotados durante os experimentos. Os critérios utilizados para a classificação das espécies também são apresentados, tanto para o reconhecimento de padrões com métodos estatísticos quanto com as redes neurais artificiais.

No capítulo 7 (Resultados e Discussão) são apresentados os resultados obtidos com a classificação das espécies por meio das características extraídas com as técnicas de análise de imagens e os métodos de reconhecimento de padrões. Neste capítulo os resultados foram agrupados por tipo de análise (forma, estrutura tubular e textura), além disso, são relatados os

dados obtidos com a seleção de características e a combinação dos atributos morfométricos. O capítulo também apresenta uma discussão a respeito dos resultados alcançados.

Por fim, o capítulo 8 (Conclusão) sintetiza as considerações e conclusões a respeito do trabalho desenvolvido. As conclusões são descritas tanto do ponto de vista biológico quanto computacional. Além disso, são apresentadas as principais contribuições do projeto, a produção científica gerada por meio de publicações e, os trabalhos futuros. A Figura 1.2 ilustra a organização dos capítulos da tese.

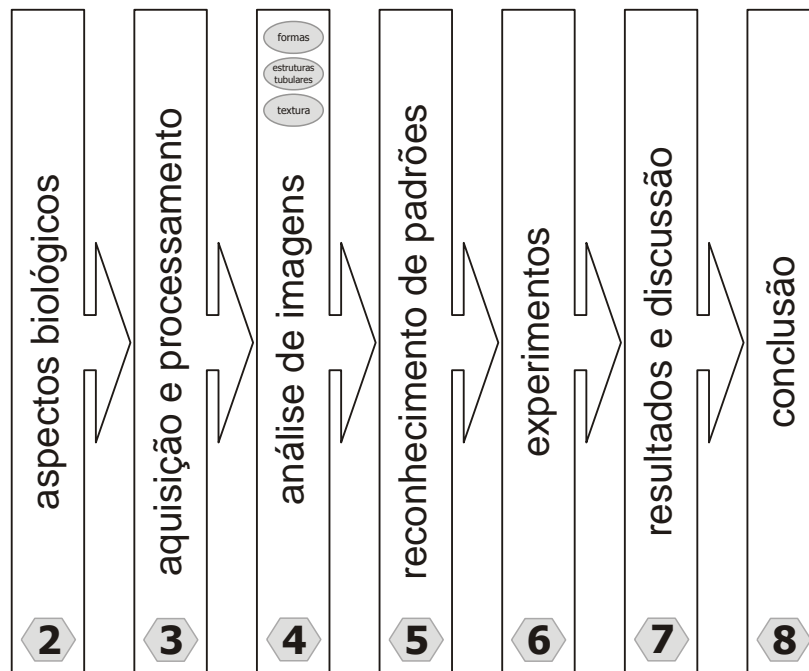


Figura 1.2: Organização dos capítulos da tese.

2

Aspectos Biológicos Evolução, Filogenia e Morfometria

Durante décadas, a descoberta da história evolutiva dos organismos vivos tem demandado esforços de diversos pesquisadores. Desde os primeiros anseios no século XIX, até as modernas técnicas de biologia molecular do século XXI, o sonho da reconstrução da “árvore da vida” é obstinado pela ciência contemporânea. As técnicas de análises morfométricas têm contribuído intensamente para solução desse desafio. Nos últimos anos, avançadas técnicas de visão artificial, associadas a métodos morfométricos e estatísticos, auxiliam na corrida pela reconstrução filogenética das espécies. Neste capítulo é apresentada uma revisão dos principais aspectos biológicos relacionados ao trabalho.

2.1 Evolução das Espécies e Biodiversidade

A teoria da evolução biológica revolucionou a visão da humanidade perante o mundo que vivemos. O primeiro defensor da evolução das espécies foi *Jean-Baptiste de Lamarck* (1744-1829). Infelizmente suas ideias não foram aceitas pela comunidade científica, pois *Lamarck* não conseguiu evidências que comprovassem a evolução das espécies. *Charles Robert Darwin* (1809-1882) durante sua viagem a bordo do *H.M.S. Beagle* (27 de dezembro de 1831 – 2 de outubro de 1836) começou a agrupar essas evidências, além disso, procurou idealizar um mecanismo que justificasse a teoria evolucionária (Futuyma, 2002). Influenciado pelas ideias de um jovem naturalista, *Alfred Russel Wallace* (1823-1913) que independentemente

concebeu a seleção natural das espécies, *Darwin* publicou o célebre livro *A origem das espécies* (Darwin, 1859). A teoria da evolução de *Darwin* procurava explicar e descrever como os organismos da terra mudavam ao longo do tempo e adquiriam uma diversidade de novas formas (Raven e Johnson, 2001).

Desde aquela época mais de 150 anos se passaram, e diversos cientistas ainda procuram mecanismos capazes de inferir a filogenia das espécies. Independente da metodologia utilizada, a principal dificuldade está na necessidade de agrupar evidências favoráveis que comprovem a história evolutiva. Essa complexa tarefa de determinar e descrever as mudanças evolutivas fez com que surgissem diversas escolas taxonômicas (totalmente divergentes) para explicar a filogenia das espécies. Na literatura não existe um consenso sobre qual metodologia é mais eficiente na análise filogenética das espécies, nem mesmo uma divisão coerente entre os diversos pensamentos das escolas taxonômicas. Recentemente, a principal preocupação está relacionada à apresentação com maior clareza dos passos envolvidos nas análises filogenéticas, além dos argumentos utilizados para as tomadas de decisão durante a classificação (Matioli, 2001; Amorin, 2002; Futuyma, 2002).

Os pensamentos divergentes das escolas taxonômicas seguem basicamente duas linhas (Brummitt, 1997): a taxonomia tradicional e a taxonomia filogenética. Para taxonomia tradicional uma espécie é formada apenas por ancestrais comuns (monofiléticos). Dessa forma, o conceito de filogenia e taxonomia das espécies se torna altamente distinto. Por outro lado, a taxonomia filogenética sustenta a equivalência entre evolução das espécies e filogenia. Nesse caso, a estrutura de classificação deve refletir de maneira precisa o conhecimento disponível sobre as relações de parentescos entre as espécies (Rapini, 2000).

Embora discordantes, as escolas de taxonomia enfrentam o mesmo desafio: a necessidade de conhecimento, conservação e utilização sustentável da fauna e flora do ambiente onde vivem os organismos (animais, plantas, etc.). Essa demanda crescente pelo conhecimento acerca da biodiversidade em escala global, regional e local, faz com que cada vez mais sejam necessários métodos para estudar a história evolutiva e taxonômica das espécies.

2.2 Análises Filogenéticas

A reconstrução filogenética tem como base o conceito de ancestralidade entre espécies, assim características observadas em espécies distintas, herdadas a partir de uma ancestral comum, indicam relações genealógicas. O estudo da história evolutiva das espécies por meio de análise filogenética tem sido amplamente empregado em biologia (Marvalid, Sequeira *et al.*, 2002; Soltis e Soltis, 2003; Pádua, 2004; Plotze, Falvo *et al.*, 2005). As relações de parentesco, ou ancestralidade, entre determinados indivíduos pode ser analisada por meio de duas abordagens filogenéticas: morfológica ou molecular.

A filogenia morfológica é fundamentada em dados coletados das formas biológicas por meio das técnicas de morfometria. Os dados morfométricos proveem uma rica descrição morfológica dos organismos biológicos (Adams, Rohlf *et al.*, 2004). As mudanças evolutivas dos mecanismos em desenvolvimento (seleção natural) produzem variações que podem ser analisadas pela filogenia morfológica (Klingenberg, 2002).

Outro método para estudo das relações evolutivas dos organismos é a filogenia molecular. Essa abordagem utiliza modernas técnicas de biologia molecular para a coleta de informações visando à taxonomia das espécies (Matioli, 2001). A reconstrução filogenética baseada em dados moleculares utiliza informações sequenciadas de DNA, RNA ou proteínas. A principal dificuldade dessa técnica é determinar o correto alinhamento das sequências obtidas para análise (Pádua, 2004).

2.3 Análises Morfométricas

As técnicas morfométricas representam um importante ferramental para o estudo das formas biológicas. Esses métodos são capazes de quantificar variações morfológicas, bem como determinar os fatores (ecológicos ou filogenéticos) que levaram a essas mudanças (Bookstein, 1991). Essencialmente, os estudos morfométricos estão ligados à utilização de técnicas de visão artificial (Costa e Cesar, 2000), e têm sido empregados em diferentes áreas das ciências biológicas como sistemática, genética, evolução, ecologia e fisiologia (Lestrel, 2000).

A morfometria tradicional utiliza técnicas de análise multivariada para quantificar as variações de forma e tamanho dos organismos. O principal problema desses métodos é o fato

deles não serem invariantes às transformações geométricas (translação, escala e rotação) das formas analisadas (Monteiro e Reis, 1999).

Assim mesmo, essa metodologia foi bastante empregada até o final da década de 80, quando as técnicas passaram por uma fase conhecida como “revolução da morfometria” (Rohlf, 1990; Bookstein, 1991; Rohlf e Marcus, 1993). A diferença fundamental entre a antiga e a nova morfometria foi a maneira de quantificar as estruturas morfológicas, e como os dados eram analisados. Essa mudança enfatizou os métodos que capturavam as informações geométricas das estruturas e preservavam esses dados durante as análises (Adams, Rohlf *et al.*, 2004).

Existem diversas maneiras de realizar análises morfométricas, uma vez que inúmeros tipos de dados podem ser coletados das formas biológicas. Os tipos de dados convencionais utilizam variáveis que não representam a geometria da forma analisada, por exemplo, distâncias, ângulos, área, etc. Outro tipo de dados bastante empregado em morfometria são os marcos anatômicos (*landmarks*). Esses marcos consideram as coordenadas cartesianas como pontos anatômicos para extração de medidas (Slice, 2007). Os marcos representam os pontos correspondentes em cada forma analisada, e são homólogos entre e dentro de uma espécie. Usualmente, pontos de máxima curvatura, justaposição de tecidos e pontos extremos são considerados como *landmarks* (Bookstein, 1991).

Em alguns casos, devido ao pequeno número de marcos anatômicos, não é possível realizar a análise morfométrica. Nesses casos, a forma analisada pode ser descrita por meio de coordenadas cartesianas ao longo do contorno ou organismo de interesse. Os pontos de contorno não possuem correspondência biológica entre espécies, ou seja, não apresentam homologia (Monteiro e Reis, 1999). Outros tipos de dados também podem ser considerados na análise morfométrica, como descritores de Fourier, textura, coeficientes de *wavelets*, complexidade, dentre outros.

Independente do tipo de dados utilizado, as técnicas de análise morfométrica exercem papel fundamental nas investigações filogenéticas. Essas técnicas podem ser utilizadas como instrumento para descoberta e documentação de novas características morfológicas, bem como, para o estudo e entendimento da funcionalidade das formas biológicas (Macleod e Forey, 2002).

2.4 Considerações Finais

Neste capítulo foram apresentados os principais aspectos biológicos relacionados a este trabalho. A teoria da evolução revolucionou a ciência contemporânea, com isso, o entendimento da estrutura organizacional dos organismos se faz cada vez mais necessária. As análises filogenéticas são fundamentais para o estudo da história evolutiva dos indivíduos. Esses métodos podem ser empregados tanto em dados morfológicos (extraídos pelas técnicas de análise morfométrica), quanto em dados moleculares (obtidos por biologia molecular). As análises morfométricas associadas às técnicas de visão artificial fornecem todo formalismo necessário para investigação das características morfológicas das espécies.

3

Aquisição e Processamento de Imagens Biológicas

Os sistemas de visão artificial fornecem um importante arcabouço para o estudo de espécies biológicas. Dentro deste contexto, um sistema de visão artificial é formado por várias etapas que são empregadas na investigação dos problemas biológicos. Considerando a diversidade de problemas, essas etapas estão intimamente relacionadas ao domínio da aplicação, uma vez que o desenvolvimento de sistemas de visão generalistas se torna inviável.

A primeira etapa de um sistema de visão artificial constitui na aquisição das informações a respeito do material biológico. Esse procedimento pode ser executado por meio de digitalizações, de maneira que todo conteúdo biológico seja transformado em conteúdo digital. Nesse caso, o resultado desse processo são imagens digitais. A qualidade das imagens obtidas é crucial para eficiência de um sistema de visão computacional. O processamento das imagens digitais é uma etapa subsequente a aquisição, cujo objetivo é transformar as imagens de acordo com a necessidade do sistema. Tarefas típicas dessa etapa envolvem a melhoria da qualidade da imagem e a segmentação das áreas de interesse (Gonzalez e Woods, 2001).

Neste capítulo são apresentados todos os aspectos relacionados a aquisição e processamento das imagens digitais utilizadas no trabalho. As informações a respeito do material biológico utilizado são detalhadas, evidenciando e descrevendo cada uma das espécies. Além disso, é apresentada uma nova proposta para a segmentação do sistema de venação foliar aplicada nas espécies de maracujás silvestres.

3.1 Material biológico

A proposta deste trabalho foi estudar e desenvolver modelos computacionais capazes de extrair e analisar características morfométricas de espécies biológicas. Para validação das metodologias foram utilizados três conjuntos de imagens digitais das espécies de maracujá (*Passiflora*), eucalipto (*Eucalyptus*) e arroz (*Oryza*).

3.1.1 Espécies de Maracujás (*Passiflora*)

As espécies vegetais do gênero *Passiflora* possuem cerca de 450 espécies catalogadas, das quais 120 podem ser encontradas no Brasil (Vitta e Bernacci, 2004). Esse gênero faz parte da família *Passifloraceae* e é popularmente conhecido como maracujás. A importância econômica do gênero *Passiflora* é muito grande, no entanto poucas espécies são exploradas comercialmente. Com isso, o conhecimento da história evolutiva do gênero, bem como a taxonomia das espécies pode trazer contribuições importantes para diversas áreas (Pádua, 2004).

O gênero *Passiflora* apresenta uma grande variabilidade morfológica entre indivíduos da mesma espécie. Além disso, algumas espécies são muito semelhantes tanto em análises morfológicas quanto moleculares, tornando a discriminação visual uma tarefa complexa (Plotze, Falvo *et al.*, 2005).

Neste trabalho foram utilizadas 13 espécies do gênero *Passiflora*, obtidas por intermédio de parceiras com pesquisadores da Escola Superior Luiz de Queiroz (ESALQ/USP) e o Instituto Agrônomo de Campinas (IAC). No total foram utilizados 580 exemplares de folhas das espécies de *Passiflora*, sendo que o número de exemplares para cada espécie foi variável devido a dificuldade de obtenção.

A Tabela 3.1 apresenta os nomes científicos das espécies utilizadas e seus respectivos números de exemplares.

Tabela 3.1: Espécies vegetais do gênero *Passiflora* utilizadas nas análises.

Espécie	Núm. de Exemplos
<i>P. actinia</i>	40
<i>P. amethystina</i>	24
<i>P. caerulea</i>	45
<i>P. coriácea</i>	41
<i>P. foetida fluminensis</i>	40
<i>P. foetida foetida</i>	44
<i>P. gibertii</i>	37
<i>P. maliformis</i>	40
<i>P. miersii</i>	56
<i>P. organensis</i>	57
<i>P. pohlii</i>	20
<i>P. sidifolia</i>	80
<i>P. suberosa</i>	56
Total:	580

A Figura 3.1 apresenta um exemplar de cada espécie do gênero *Passiflora* utilizado nos experimentos.

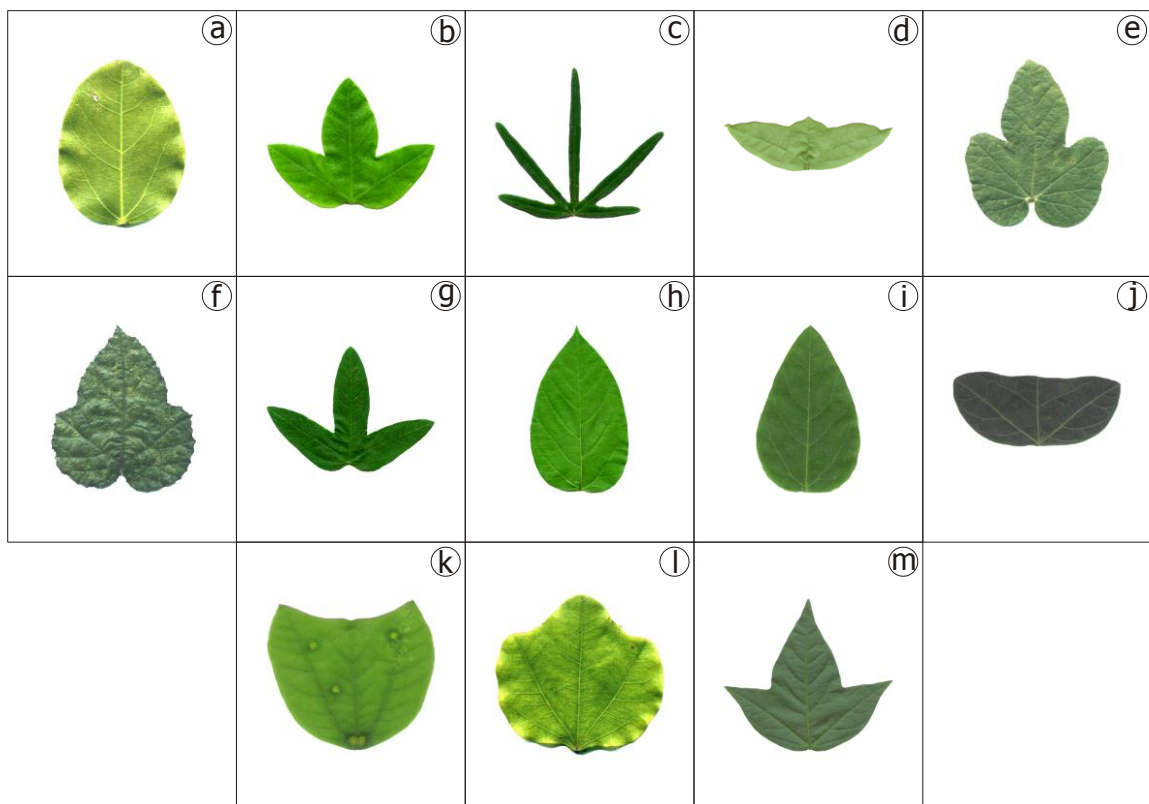


Figura 3.1: Espécies de maracujás silvestres do gênero *Passiflora*: (a) *P. actinia*; (b) *P. amethystina*; (c) *P. caerulea*; (d) *P. coriácea*; (e) *P. foetida fluminensis*; (f) *P. foetida foetida*; (g) *P. gibertii*; (h) *P. maliformis*; (i) *P. miersii*; (j) *P. organensis*; (k) *P. pohlii*; (l) *P. sidifolia*; (m) *P. suberosa*.

3.1.2 Espécies de Eucalipto (*Eucalyptus*)

O gênero *Eucalyptus* é formado por cerca de 600 espécies e está compreendido dentro da família *Myrtaceae* (Brooker, 2000). A grande maioria tem origem Australiana, e sua introdução no Brasil ocorreu na segunda metade do século XIX. No início as espécies foram utilizadas para alimentação de caldeiras dos trens e também como mourões e postes. Posteriormente, algumas espécies foram empregadas na produção de celulose de fibra curta. Esse fato tornou o Brasil mundialmente conhecido na produção de papel e celulose (Missiaggia, 2005; Pádua, 2007).

Nas análises realizadas foram utilizados 285 exemplares de folhas de espécies do gênero *Eucalyptus*, sendo 60 exemplares puros (originários da empresa Aracruz Celulose) e 225 exemplares híbridos. Esses exemplares compreendem 10 espécies diferentes. As imagens digitais das espécies foram obtidas por meio de uma parceria com a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), por meio do Laboratório de Genética Vegetal da Embrapa Recursos Genéticos e Biotecnologia em Brasília.

A Tabela 3.2 apresenta as espécies utilizadas contendo os nomes científicos, tipos (pura ou híbrida) e número de exemplares. Na Figura 3.2 são ilustradas as espécies de eucalipto do gênero *Eucalyptus* utilizadas nos experimentos.

Tabela 3.2: Espécies do gênero *Eucalyptus* utilizadas nas análises.

Espécie	Tipo	Núm. de Exemplares
<i>E. grandis</i>	Pura	10
<i>E. pellita</i>	Pura	12
<i>Corymbia torelliana</i>	Pura	14
<i>E. urophylla</i>	Pura	24
<i>E. grandis x E. camaldulensis</i>	Híbrida	46
<i>E. saligna x E. camaldulensis</i>	Híbrida	12
<i>E. urophylla x E. camaldulensis</i>	Híbrida	29
<i>E. urophylla x E. globulus</i>	Híbrida	36
<i>E. grandis x E. globulus</i>	Híbrida	24
<i>E. urophylla x E. grandis</i>	Híbrida	78
Total:		285

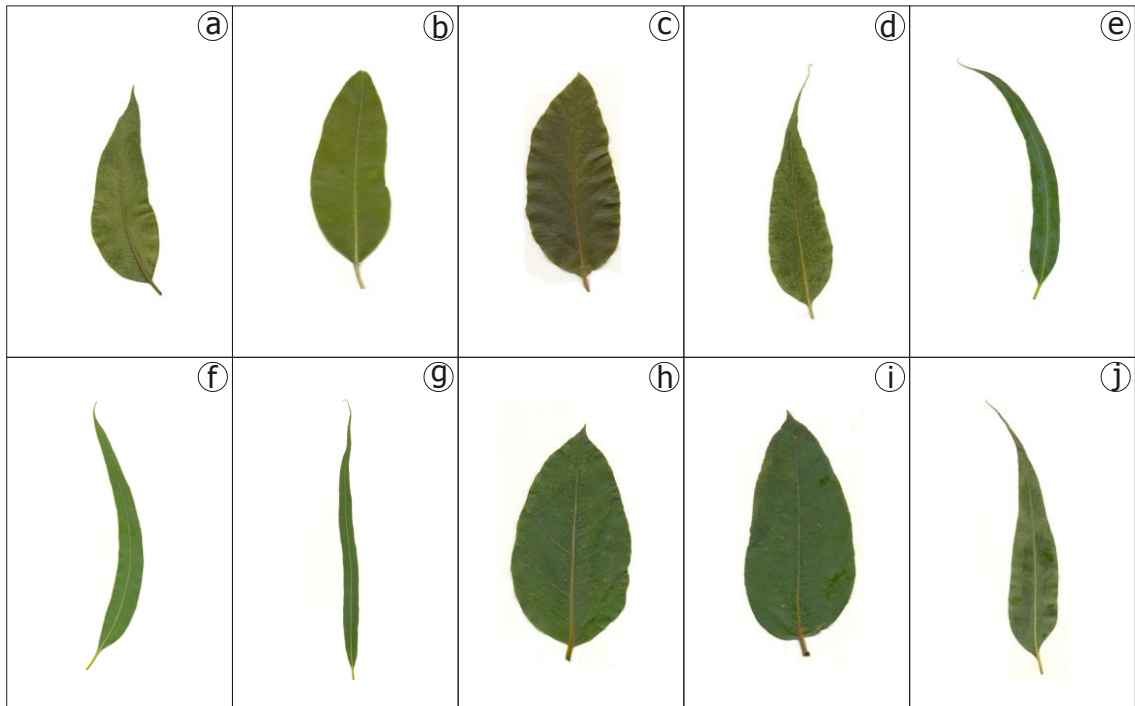


Figura 3.2: Espécies de eucalipto do gênero *Eucalyptus* utilizadas nas análises: (a) *E. grandis*; (b) *E. pellita*; (c) *Corymbia torelliana*; (d) *E. urophylla*; (e) *E. grandis* x *E. camaldulensis*; (f) *E. saligna* x *E. camaldulensis*; (g) *E. urophylla* x *E. camaldulensis*; (h) *E. urophylla* x *E. globulus*; (i) *E. grandis* x *E. globulus*; (j) *E. urophylla* x *E. grandis*.

3.1.3 Espécies de Arroz (*Oryza*)

O arroz possui uma enorme importância econômica, uma vez que é um alimento essencial para população mundial. O gênero *Oryza* pertence a família *Poaceae* e é originário de regiões da Ásia e África. Diversas pesquisas são desenvolvidas por meio de melhoramento genético utilizando o gênero *Oryza*, com o objetivo de aumentar a produtividade e qualidade dos grãos (Buso, 1998; Oliveira, 2002). As pesquisas envolvendo análise morfométrica das espécies também são fundamentais, pois fornecem conhecimento a respeito de espécies silvestres pouco estudadas.

Neste trabalho foram utilizadas 12 espécies de arroz do gênero *Oryza*. O número de exemplares por espécies possui uma distribuição variada, totalizando 177 amostras. As imagens digitais foram obtidas por meio de uma parceria com a Escola Superior Luiz de Queiroz (ESALQ/USP). A Tabela 3.3 apresenta a distribuição das espécies e seus respectivos números de exemplares. Na Figura 3.3 a imagem digital de um exemplar de cada espécie pode ser observada.

Tabela 3.3: Espécies de arroz do gênero *Oryza* utilizadas nas análises.

Espécie	Núm. de Exemplos
<i>O. alta</i>	11
<i>O. barthii</i>	13
<i>O. glaberrima</i>	9
<i>O. glumaepatula</i>	32
<i>O. grandiglumis</i>	18
<i>O. latifolia</i>	16
<i>O. longistaminata</i>	16
<i>O. nivara</i>	14
<i>O. officinalis</i>	12
<i>O. punctata</i>	12
<i>O. rufipogon</i>	16
<i>O. sativa</i>	8
Total:	177

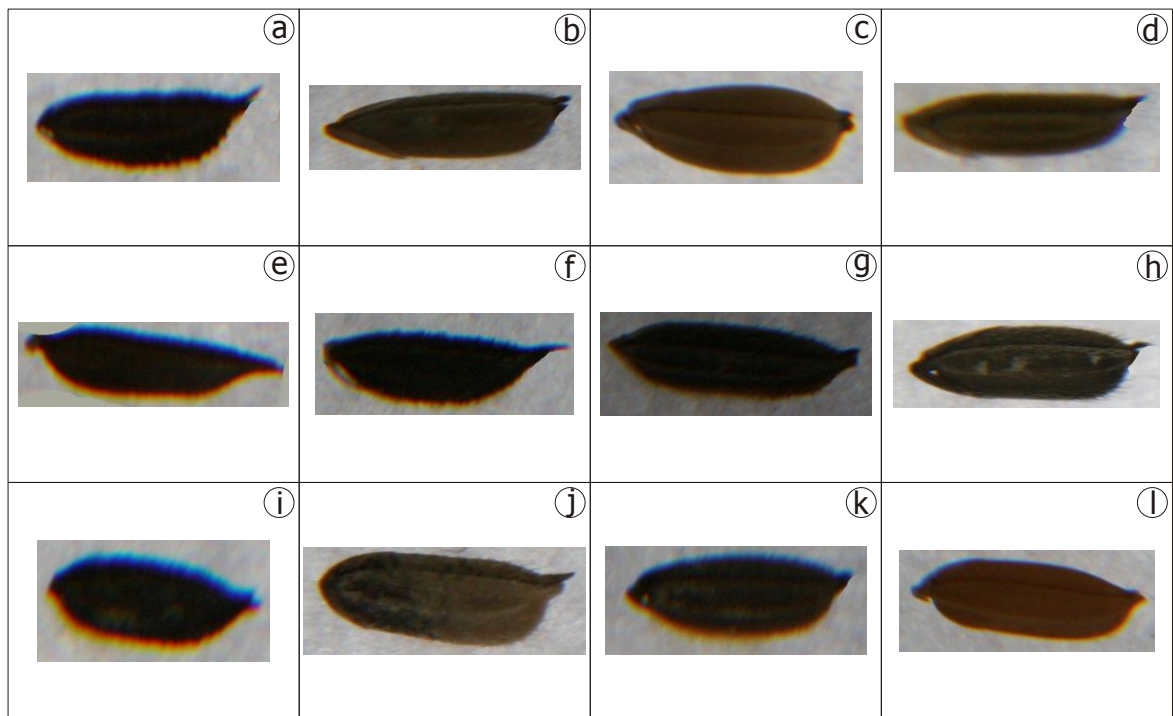


Figura 3.3: Espécies de arroz do gênero *Oryza* utilizadas nos experimentos: (a) *O.alta*; (b) *O.barthii*; (c) *O.glaberrima*; (d) *O.glumaepatula*; (e) *O.grandiglumis*; (f) *O.latifolia*; (g) *O.longistaminata*; (h) *O.nivara*; (i) *O.officinalis*; (j) *O.punctata*; (k) *O.rufipogon*; (l) *O.sativa*.

3.2 Aquisição

Seguindo o modelo de funcionamento do sistema *TreeVis*, o processo de aquisição de imagens teve como objetivo a digitalização do material biológico. Para as espécies foliares de maracujás e eucalipto, as imagens foram adquiridas com um digitalizador comercial

(scanner). Para garantir a eficiência do processamento e análise das espécies, as imagens foram digitalizadas em alta resolução (superior a 300 dpi).

Como os grãos de arroz possuem características volumétricas, uma estratégia diferente foi adotada para aquisição das imagens. Para essas espécies foi utilizada uma câmera fotográfica digital de alta resolução. Para evitar distorções e perda de informações, todas as imagens foram armazenadas no formato bitmap (bmp). No entanto, qualquer formato de imagem pode ser utilizado, desde que não realize compressão com perdas.

3.3 Processamento

Durante a fase de processamento as imagens biológicas são adequadas para posterior análise. Neste procedimento, três tarefas principais são executadas: (i) filtragem; (ii) normalização e (iii) segmentação.

Inicialmente, uma filtragem gaussiana foi aplicada nas imagens com o intuito de minimizar possíveis ruídos originários do processo de aquisição. Em seguida, o tamanho das imagens foi normalizado em função da largura e altura, com o objetivo de desconsiderar fatores relativos a escala, principalmente, por conta do gênero *Oryza* que foi fotografado. Dessa maneira, todas as imagens pertencentes a um mesmo gênero foram definidas com o mesmo tamanho. A Tabela 3.4 apresenta os valores em pixels utilizados na normalização das imagens.

Tabela 3.4: Valores em pixels utilizados para normalização das imagens biológicas.

Gênero	Largura (L)	Altura (A)
<i>Passiflora</i>	600	600
<i>Eucalyptus</i>	800	800
<i>Oryza</i>	760	260

A segmentação de imagens é uma das tarefas mais complexas dentro de um sistema de visão artificial. No caso de imagens do mundo real, o processo de segmentação de regiões de interesse pode ser encarado como um verdadeiro desafio. O principal objetivo da segmentação de imagens é a divisão da imagem em partes que contenham algum tipo de correlação (Sonka, Hlavac *et al.*, 1999; Gonzalez e Woods, 2001). O tipo de relação entre as subdivisões da imagem está intimamente ligado ao problema a ser resolvido. Para as espécies biológicas

utilizadas nas análises, a segmentação foi empregada para produzir dois objetos de interesse: (i) contorno e (ii) estrutura tubular.

A detecção dos contornos das imagens foi realizada com o método de limiarização de *Otsu* (Otsu, 1979; Sezgin e Sankur, 2004), seguido da técnica de detecção de contorno de *Canny* (Canny, 1986; Parker, 1996). Em alguns casos foi utilizado um algoritmo seguidor de contorno (*contour-following*) (Costa e Cesar, 2000). A escolha dessas técnicas foi realizada por meio de testes experimentais. A Figura 3.4 apresenta o resultado do processo para detecção de contorno das imagens de maracujás, eucalipto e arroz.

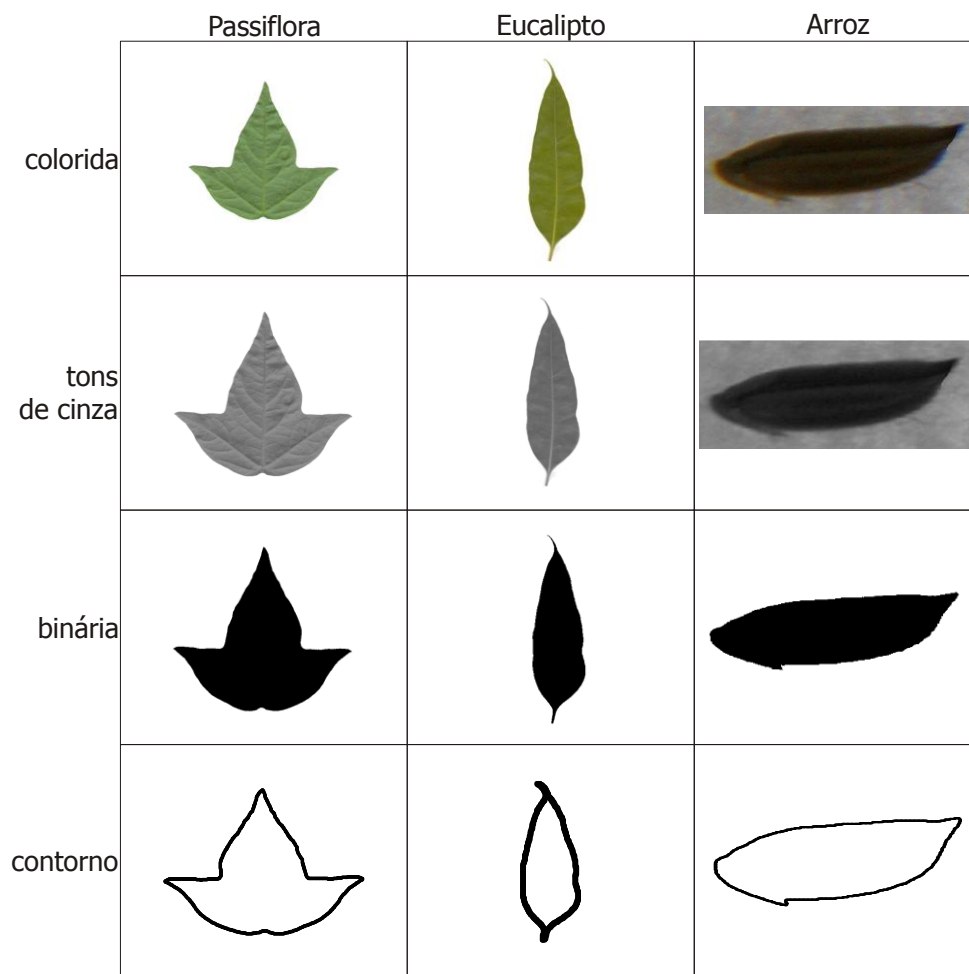


Figura 3.4: Resultado do processo para detecção do contorno das imagens biológicas.

Para a segmentação das estruturas tubulares uma nova abordagem foi proposta combinando processos químicos de clareamento e algoritmos de visão artificial (Plotze, Nardin *et al.*, 2006). Essa metodologia foi empregada na segmentação do sistema de venação foliar (nervuras) das espécies de *Passiflora*.

3.3.1 Segmentação do Sistema de Venação Foliar

O sistema de venação dos órgãos foliares representa uma imensa rede de células interconectadas que transportam água e nutrientes por toda a planta. Esse fato indica a importância de se estudar e conhecer o comportamento interno dos órgãos foliares. Além disso, grande parte do processo de taxonomia de espécies vegetais é baseada em características extraídas do sistema venação. Dessa forma, a segmentação das nervuras é imprescindível para a correta taxonomia das espécies.

A metodologia proposta para segmentação das nervuras dos órgãos foliares foi dividida em duas etapas. A primeira etapa consiste em um processo químico de clarificação, em que a ideia principal é remover a pigmentação do órgão foliar, além de evidenciar o sistema de venação. Entretanto, o processo químico não resulta em uma segmentação eficiente, principalmente quando é necessário realizar algum tipo de estudo no sistema de venação, como por exemplo, extração de medidas morfométricas. Assim, a segunda etapa da metodologia utiliza técnicas de visão artificial para alcançar uma eficiente segmentação do sistema de venação.

Para o processo de clarificação e coloração (etapa 1) da lâmina foliar foram selecionadas folhas frescas e herborizadas de materiais. Entretanto, de alguns materiais foi possível dispor apenas de material herborizado. Todo este procedimento foi realizado no Instituto Agrônomo de Campinas (IAC). As folhas foram colocadas numa solução de NaOH a 5%, que foi trocada a cada três dias aproximadamente, até o material se tornar transparente. A solução de NaOH (5%), foi substituída por água destilada e depois por água sanitária (50%), na qual o material foi deixado de 30 minutos a duas horas, até ficar bem claro.

Após esse tempo a água sanitária foi substituída por água destilada novamente, que foi trocada duas vezes em 30 minutos. A água destilada foi então substituída por álcool (70%) por duas horas, no mínimo. O material foi depois colocado em safranina (1%) em álcool a 70% por 24 horas, sendo lavado posteriormente com álcool 70%, até apenas as nervuras ficarem bem coradas e a lâmina apenas rosada, para que a venação ficasse bem visível. Posteriormente foi utilizado xilol (1:1) ou xilol puro, dependendo da fragilidade da folha, para que os materiais ficassem mais duros possibilitando a digitalização de materiais mais sensíveis. As folhas foram então, digitalizadas com resolução de 300dpi, e para isso utilizou-se duas placas de vidro a fim de não manchar o digitalizador.

As técnicas de visão artificial utilizadas na segunda fase do processo de segmentação são baseadas em dois conceitos: filtragem por difusão anisotrópica e *thresholding* adaptativo. O modelo de filtros anisotrópicos foi proposto por *Perona e Malik* (Perona e Malik, 1990), e sua ideia principal é realizar a suavização seletiva da imagem. Uma característica importante deste filtro é que as imagens suavizadas mantêm as bordas nítidas, além de remover os ruídos e descartar detalhes insignificantes da imagem. *Perona e Malik* definiram um espaço de escalas não lineares anisotrópicas, por meio de uma equação diferencial parcial, como descrito na Equação 3.1:

$$\begin{cases} \frac{\partial F_t(x, y)}{\partial t} = \nabla \cdot [g(\|\nabla F_t(x, y)\|)\nabla F_t(x, y)] \\ F_0(x, y) = f(x, y) \end{cases} \quad (3.1)$$

em que $f(x, y)$ representa uma imagem bidimensional; $F_t(x, y)$ o espaço de escalas dessa imagem; $\|\nabla F_t(x, y)\|$ a magnitude do gradiente e g uma função de *parada-na-aresta* (Perona e Malik, 1990). A formulação discreta para equação não-linear de difusão anisotrópica de *Perona e Malik* é definida na Equação 3.2.

$$I(s, t + 1) = I(s, t) + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(|\nabla I_s, p(t)|) \nabla I_s, p(t) \quad (3.2)$$

Finalmente, a segmentação do sistema de venação é alcançada utilizando a técnica de limiarização adaptativa (Gonzalez e Woods, 2001). Esta abordagem consiste na divisão da imagem original em diversas subimagens, e então são aplicados diferentes valores de limiar (*threshold*) para segmentar cada subimagem. A utilização desta técnica foi necessária, uma vez que as imagens resultantes do filtro *Perona-Malik* contêm grandes variações de tons de cinza (não uniforme). Dessa forma, um limiar global não é capaz de segmentar satisfatoriamente o sistema de venação. A Figura 3.5 ilustra todas as fases do processo de segmentação do sistema de venação dos órgãos foliares.

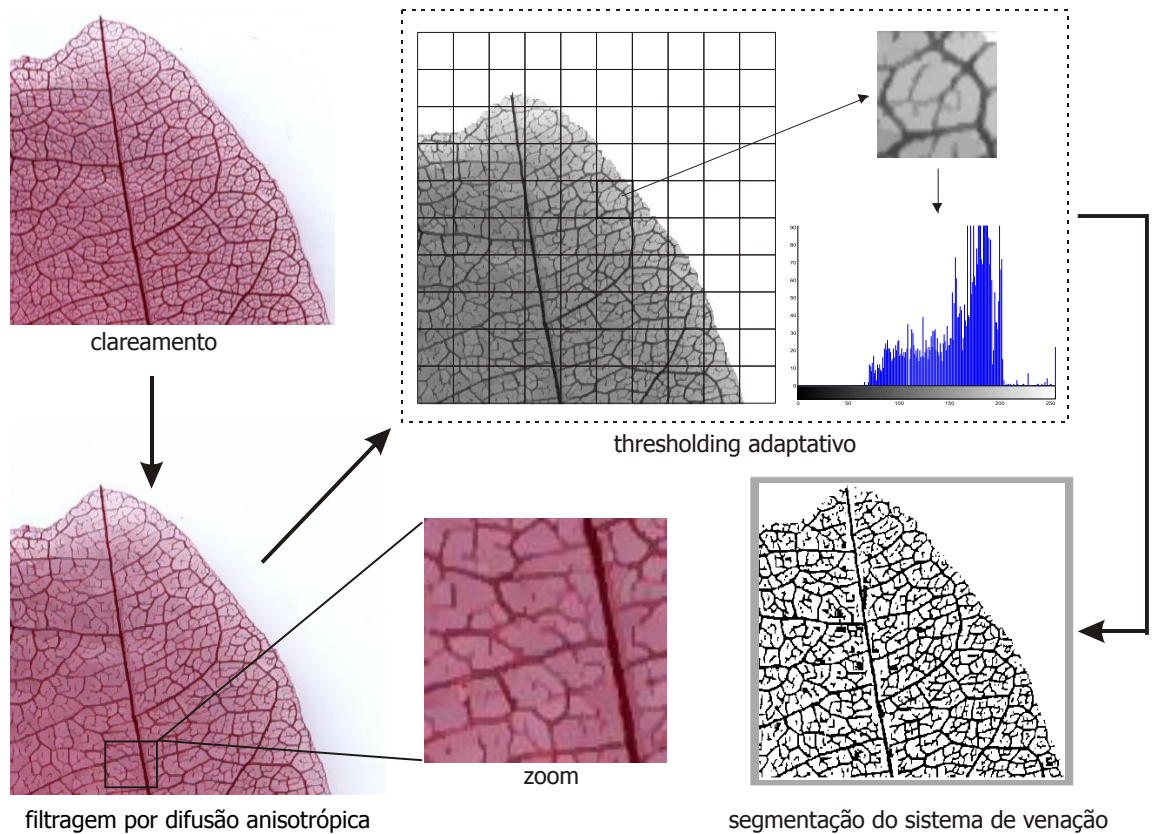


Figura 3.5: Metodologia proposta para segmentação do sistema de venação foliar.

Quando o processo de segmentação é terminado, um algoritmo de afinamento é aplicado na imagem. Afinamento é um processo que reduz uma forma para sua versão simplificada, chamada esqueleto (Lam, Lee *et al.*, 1992; Plotze e Bruno, 2004). O algoritmo de afinamento elimina os resquícios de ruídos do processo de segmentação. Neste momento as lâminas foliares contêm as nervuras devidamente segmentadas, e podem ser utilizadas na análise morfométrica.

3.4 Considerações Finais

Neste capítulo foram apresentadas as espécies biológicas utilizadas nos experimentos, bem como o processo de aquisição e processamento das imagens. Para cada gênero foram descritos os nomes científicos e seus respectivos números de exemplares. A aquisição das imagens foi detalhada, evidenciando a importância da qualidade e resolução para as análises experimentais. Além disso, foram descritas as etapas do processamento das imagens biológicas, desde a filtragem até a segmentação. Uma nova metodologia para segmentação do

sistema de venação foliar (nervura) foi proposta. A técnica foi aplicada nas espécies de *Passiflora* e os resultados da segmentação foram apresentados.

4

Análise de Imagens Biológicas

As técnicas de análise de imagens são capazes de examinar estruturas biológicas em diversas escalas espaciais. Dependendo da escala, várias características numéricas ou semânticas podem ser extraídas. Em escala macroscópica, por exemplo, informações como tamanho, formato, cor e textura podem ser estudadas. Por outro lado, em escalas menores (microscópicas) outros tipos de descrições podem ser obtidas, como por exemplo, a contagem do número de células. Este capítulo apresenta uma descrição das principais técnicas de análise de imagens que foram utilizadas no trabalho. Além disso, novas metodologias foram propostas. Essas técnicas foram empregadas nas investigações morfométricas das espécies biológicas, e foram divididas em três grupos: análise de formas, análise de estruturas tubulares e análise de textura.

4.1 Análise de Formas

As técnicas de análise de formas são essenciais para os sistemas de visão artificial. Essas técnicas são responsáveis pela análise dos objetos de uma cena, transformando as informações visuais em descritores que são a base para os métodos de reconhecimento de padrões (Loncaric, 1998; Mingqiang, Kidiyo *et al.*, 2008). A aplicabilidade é comprovada principalmente na área biológica, sendo utilizada em estudos evolutivos, taxonomia de espécies, análises morfológicas, contagem de células, dentre outras (Costa e Cesar, 2000). Nesses casos, grande parte das análises é realizada em função do contorno das formas

biológicas. No contexto do trabalho, as técnicas de análise de formas foram empregadas nas espécies de maracujá, eucalipto e arroz.

4.1.1 Assinaturas Digitais

A assinatura digital é uma das abordagens mais empregadas na exploração de características das formas biológicas. Basicamente, a assinatura pode ser entendida como a representação de uma forma bidimensional por meio de um sinal unidimensional (Gonzalez e Woods, 2001). A principal vantagem dessa representação é a possibilidade de aplicar técnicas de processamento de sinais, tais como transformada de Fourier, *wavelets*, etc. (Pazoti, 2005).

Diversas técnicas de assinaturas digitais podem ser encontradas na literatura, dentre elas a mais popular é a representação por sinais paramétricos. Nessa abordagem, os componentes das coordenadas do contorno da forma biológica (x, y) são decompostos em dois sinais unidimensionais independentes (um para cada eixo). O tamanho do sinal é proporcional ao perímetro do contorno, além disso, essa técnica de assinatura digital é sensível à translação, escala e rotação (Gonzalez e Woods, 2001). Outra desvantagem, é que a assinatura resultante depende do ponto inicial utilizado para percorrer o contorno. Um excelente algoritmo seguidor de contorno é apresentado em (Costa e Cesar, 2000).

Outra abordagem empregada na investigação do contorno de formas biológicas é a assinatura baseada no centroide. Essa técnica consiste na extração da assinatura digital utilizando as distâncias entre o centroide e os *i-ésimos* pontos que compõem o contorno da forma biológica. Dependendo da métrica de distância utilizada, diferentes assinaturas digitais podem ser obtidas.

A assinatura por projeção é uma técnica também utilizada para representação unidimensional de formas biológicas. Nessa abordagem, a assinatura digital da forma é produzida com o somatório dos pontos (que compõem a imagem) em função de um determinado ângulo. Dessa forma, diversas assinaturas digitais podem ser obtidas dependendo da orientação (ângulo) escolhida. Normalmente, as assinaturas por projeção mais utilizadas são para os ângulos de $\theta = 90^\circ$ (vertical), $\theta = 180^\circ$ (horizontal) e $\theta = 45^\circ$ (Jain, 1989).

A Figura 4.1 apresenta um quadro comparativo entre as principais técnicas para extração de assinaturas digitais. Na Figura é possível notar o resultado das assinaturas para um exemplar da espécie de maracujá, eucalipto e arroz. As técnicas empregadas foram:

assinaturas por coordenadas paramétricas, assinaturas baseadas no centróide a assinaturas por projeção.

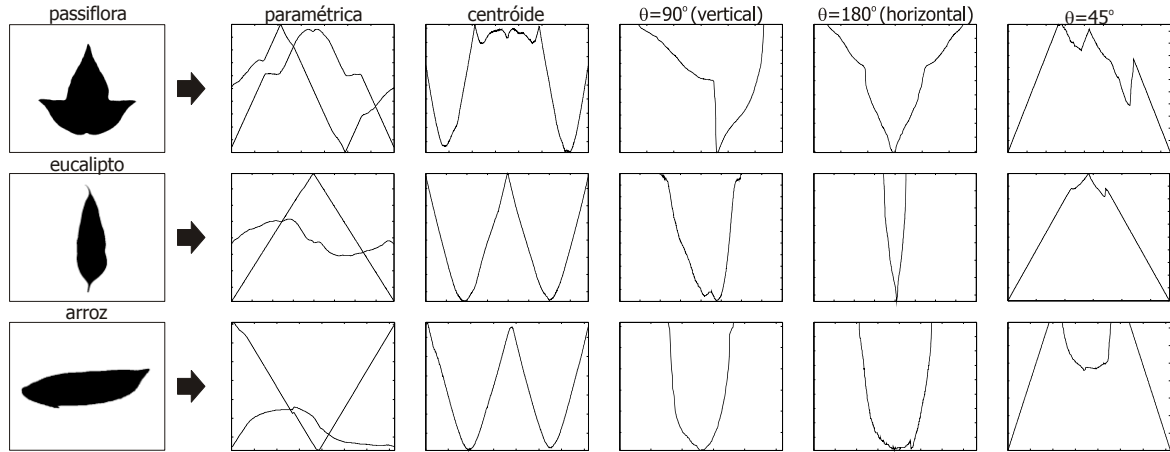


Figura 4.1: Assinaturas digitais das espécies biológicas utilizadas nos experimentos.

4.1.2 Curvatura

A análise de curvatura é uma eficiente abordagem para descrição e representação do contorno de formas biológicas. As características de curvatura são biologicamente relacionadas com as informações utilizadas pelo sistema visual humano na percepção de formas (Attneave, 1954). Diversos aspectos geométricos das formas biológicas podem ser caracterizados por meio da curvatura, sendo alguns deles: máximo valor absoluto da curvatura local = vértice genérico; máxima curvatura local positiva = vértice convexo; máxima curvatura local negativa = vértice côncavo; curvatura constante em zero = segmento de reta; cruzamento em zero = ponto de inflexão; dentre outros (Costa e Cesar, 2000).

A curvatura pode ser definida como a taxa de variação quantitativa do contorno ao longo da forma analisada. O contorno paramétrico de uma forma genérica pode ser expresso como $c(t) = (x(t), y(t))$ assim, a curvatura $k(t)$ de uma curva paramétrica $c(t)$ pode ser definida na Equação 4.1 (Mokhtarian e Mackworth, 1992).

$$k(t) = \frac{\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t)}{(\dot{x}(t)^2 + \dot{y}(t)^2)^{\frac{3}{2}}} \quad (4.1)$$

A primeira derivada da função $(\dot{x}(t), \dot{y}(t))$ define o campo tangencial unitário, e a segunda derivada $(\ddot{x}(t), \ddot{y}(t))$ determina as mudanças de orientação no campo tangencial

(Pazoti, 2005). Como o cálculo das derivadas é obtido a partir da representação discreta da forma analisada, e nesse caso não há equações paramétricas eficientes, algumas soluções alternativas foram propostas.

A abordagem mais simples, conhecida como curvatura baseada no ângulo entre dois vetores, utiliza uma representação vetorial do contorno da forma. Assim, considerando um ponto de contorno $c(n_0) = \{x(n), y(n)\}$, a curvatura pode ser estimada pelo ângulo formado entre os vizinhos (da direita e da esquerda) que estão a uma distância i do ponto $c(n_0)$. Os vetores dos vizinhos da esquerda podem ser expressos como $u_i(n) = (x(n) - x(n - i), y(n) - y(n - i))$ e da direita $v_i(n) = (x(n) - x(n - i), y(n) - y(n - i))$, dessa forma, o cosseno do ângulo entre os vetores pode ser obtido com a Equação 4.2.

$$r_i(n) = \frac{u_i(n) \cdot v_i(n)}{\|u_i(n)\| \cdot \|v_i(n)\|} \quad (4.2)$$

A interpolação de $x(t)$ e $y(t)$ é outra abordagem para estimativa da curvatura. Nesse método, a partir da derivada de $x(n)$ e $y(n)$ por diferenças finitas, é possível determinar a curvatura substituindo os valores obtidos na Equação 4.1.

Outra metodologia bastante eficiente, e amplamente empregada em processamento de imagens, é a estimativa da curvatura por meio da transformada de Fourier utilizando sua propriedade derivativa (Papoulis, 1962; Jain, 1989). Para isso, as derivadas de primeira e segunda ordem são calculadas dos sinais $x(n)$ e $y(n)$. Assim, aplicando as simplificações e substituições necessárias, a curvatura pela transformada de Fourier pode ser estimada por meio da Equação 4.3 (Cesar, 1997; Costa e Cesar, 2000).

$$k(t) = \frac{F^{-1}\{fX(f)\}(t)F^{-1}\{f^2Y(f)\}(t) - F^{-1}\{f^2X(f)\}(t)F^{-1}\{fY(f)\}(t)}{\left[(F^{-1}\{fX(f)\}(t))^2 + (F^{-1}\{fY(f)\}(t))^2 \right]^{\frac{3}{2}}} \quad (4.3)$$

A Figura 4.2 ilustra a análise do contorno de formas biológicas por meio da curvatura. A análise foi empregada para os três gêneros estudados *Passiflora*, *Eucalyptus* e *Oryza*.

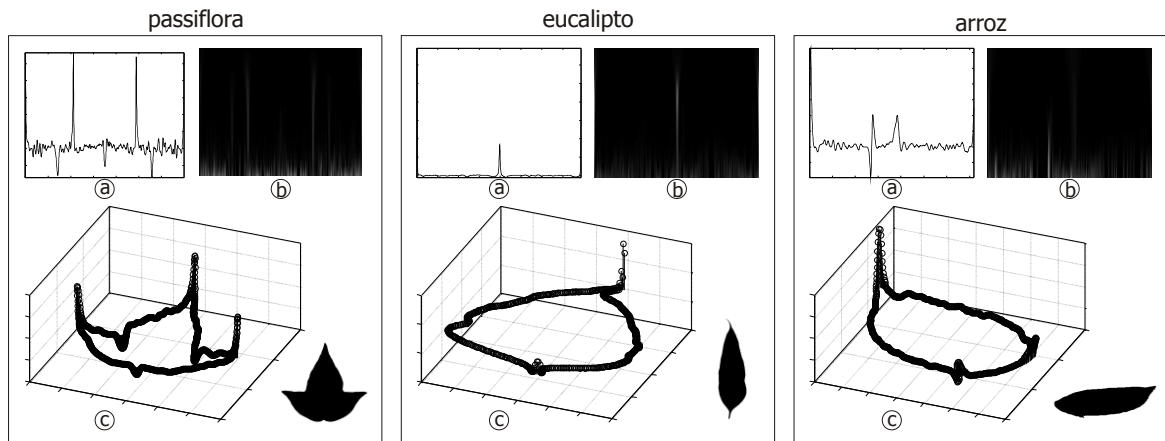


Figura 4.2: Análise da curvatura de formas biológicas. (a) representação unidimensional (1D); (b) curvograma do contorno; (c) representação tridimensional (3D).

4.1.3 Transformada e Descritores de Fourier

A transformada de Fourier (TF) é uma importante ferramenta para investigação de sinais e formas biológicas. Essa técnica tem como objetivo a modificação de um sinal (nD) no domínio do tempo para o domínio da frequência. Em diversas aplicações científicas um sinal $F(u)$, representado no domínio da frequência, demonstra propriedades mais interessantes do que um sinal $f(x)$ no domínio do tempo (Semmlow, 2004). A filtragem de sinais (1D ou 2D), por exemplo, pode ser facilmente realizada no domínio da frequência por meio do teorema da convolução (Gonzalez e Woods, 2001). A transformada de Fourier permite decompor o sinal $f(x)$ em todos os seus componentes de frequência constituintes, sendo representada discretamente por um somatório de senos e cossenos (Costa e Cesar, 2000).

Para uma imagem bidimensional $f(x, y) = \{(x_0, y_0), \dots, (x_{M-1}, y_{N-1})\}$, a representação discreta da transformada de Fourier $f(x, y) \Leftrightarrow F(u, v)$ é expressa na Equação 4.4 (Gonzalez e Woods, 2001).

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(ux/M + vy/N)} \quad (4.4)$$

A transformação inversa, $F(u, v) \Leftrightarrow f(x, y)$, representa a conversão do sinal 2D complexo no domínio da frequência para o domínio do tempo (sinal original). A transformada de Fourier inversa de um sinal bidimensional é formalizada na Equação 4.5 (Gonzalez e Woods, 2001).

$$f(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{j2\pi(ux/M + vy/N)} \quad (4.5)$$

O resultado da *TF* é um sinal complexo $F(u, v)$ composto por uma parte real (R) e uma parte imaginária (I). Dependendo do tipo de análise realizada, pode ser conveniente representar este sinal complexo de outra maneira, como: espectro de Fourier (Equação 4.6), ângulo da fase (Equação 4.7) ou espectro da potência (*power spectrum*) (Equação 4.8).

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

$$\phi(u, v) = \tan^{-1} \left[\frac{I(u, v)}{R(u, v)} \right] \quad (4.7)$$

$$P(u, v) = |F(u, v)|^2 \quad (4.8)$$

A Figura 4.3 ilustra a representação de formas biológicas com a transformada de Fourier. Para isso, o contorno das imagens foi extraído com o algoritmo seguidor de contorno (Costa e Cesar, 2000). Em seguida, a transformada de Fourier 1D foi aplicada, e como resultado é possível obter o espectro de Fourier.

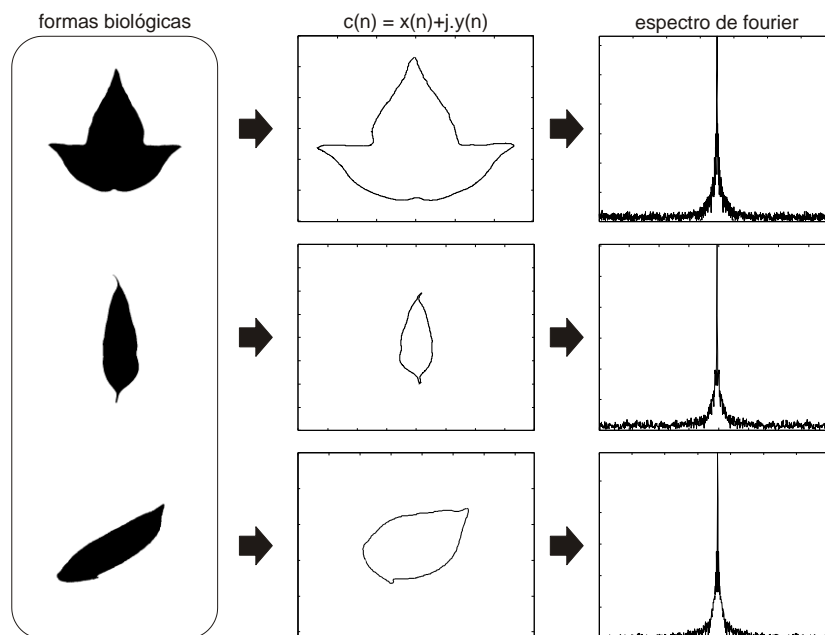


Figura 4.3: Transformada de Fourier de sinais biológicos.

Uma importante aplicação da transformada de Fourier é a extração de características por meio dos Descritores de Fourier (Pratt, 2001; Pedrini e Schwartz, 2008). A principal vantagem deste método é produzir representações invariantes às transformações geométricas de translação, escala e rotação (Pavlidis, 1977; Pandya, 1995). As características globais da imagem (ou de um sinal unidimensional) podem ser extraídas nos primeiros componentes de baixa frequência, enquanto os componentes de alta frequência capturam as características mais detalhadas (Costa e Cesar, 2000).

Seja $c(n) \Leftrightarrow C(n)$ a transformada de Fourier discreta do contorno de uma forma bidimensional. Considerando o contorno representado pelo sinal complexo $c(n) = (xn) + j \cdot y(n)$, o conjunto de descritores de Fourier (D_f) é obtido truncando os coeficientes complexos $C(n)$. Estes descritores, invariantes as transformações geométricas, podem ser definidos pela Equação 4.9 (Costa e Cesar, 2000).

$$D_f(n) = |ND_f(n)| \quad \text{onde} \quad ND_f(n) = \begin{cases} 0 & n = 0 \\ C(n)/C(1) & n \neq 0 \end{cases} \quad (4.9)$$

4.1.4 Dimensão Fractal Multiescala

O mundo a nossa volta é essencialmente constituído por um emaranhado de formas geometricamente complexas (Pentland, 1985). Certamente, esse conjunto caótico e irregular de formas contém explicações para a grande maioria dos fenômenos naturais (Falconer, 1990). Dessa forma, a análise de complexidade compreende um importante ferramental para o estudo morfológico das formas biológicas. Na literatura, diversas abordagens foram propostas para análise de complexidade, dentre elas a mais difundida é a dimensão fractal (Mandelbrot, 1977). Intuitivamente, a dimensão fractal representa a quantificação da aspereza ou rugosidade de uma forma (Pentland, 1985). O estudo evolutivo e morfológico de espécies com a dimensão fractal tem demonstrado excelentes resultados, como na taxonomia de borboletas do gênero *Danaus* por meio da análise de complexidade das asas (Castrejón-Pita, Sarmiento-Galán *et al.*, 2005), em espécies de maracujás silvestres do gênero *Passiflora* (Plotze, Falvo *et al.*, 2005), no estudo paleontológico da estrutura craniana de seres humanos (Arques e Gilbert, 2002), em espécies florestais (Bruno, Plotze *et al.*, 2008), dentre outros.

A geometria fractal é um campo da matemática que tem como objetivo principal representar geometricamente formas da natureza (Mandelbrot, 1983). O arranjo complexo das formas biológicas provê importantes informações para análise morfológica, e pode ser quantificado com a dimensão fractal. Enquanto a dimensão topológica é restrita a valores inteiros, a dimensão fractal permite valores fracionários (Falconer, 1997).

Baseado no método *Bouligand-Minkowski* (Tricot, 1995), a dimensão fractal de uma forma pode ser definida como se segue: Seja S um conjunto que representa a forma analisada, S_r suas respectivas dilatações por um raio d (este processo pode ser realizado por meio da técnica de dilatações exatas (Costa e Cesar, 2000)). Denota-se $A(d)$, as diversas áreas assumidas por S enquanto ocorre o processo de dilatação. Dessa forma, a dimensão fractal pode ser estimada na Equação 4.10.

$$df = 2 - \lim_{d \rightarrow 0} \frac{\log(A(d))}{\log(d)} \quad (4.10)$$

A principal característica das formas fractais é sua auto-similaridade (Schroeder, 1996). Isso significa que uma determinada forma, tanto em escalas maiores (macroscópicas) quanto em escalas menores (microscópicas), é constituída por um mesmo conjunto de primitivas (Pinto, 2001). No caso da forma S , a auto-similaridade perfeita é assumida em escalas espaciais menores, ou seja, quando $d \rightarrow 0$. Entretanto, é importante salientar que a auto-similaridade infinita não pode ser encontrada na natureza, assim, formas biológicas apresentam apenas alguns níveis de auto-similaridade. Este fato se torna mais complexo quando as formas analisadas são representadas discretamente, em que a resolução finita dos dispositivos de aquisição limita os detalhes em escalas menores (Carlin, 2000).

As restrições relativas à complexidade de formas biológicas (não-fractais) são superadas com a representação da Equação 4.10 por meio de uma função logarítmica, em que temos: $\log(A(d)) \times \log(d)$. Nesse caso, a estimativa da dimensão fractal é obtida computando o coeficiente angular resultante da interpolação da função logarítmica.

Na literatura, um amplo número de técnicas para estimativa da dimensão fractal pode ser encontradas (Turner, Blackledge *et al.*, 1998). A principal limitação desses métodos está relacionada ao fato de resultarem em um único valor para representar a complexidade da

forma. Na maioria das vezes, um único descritor é insuficiente para caracterização e análise morfológica das formas biológicas (Bruno, Plotze *et al.*, 2008). Além disso, essas técnicas são incapazes de expressar completamente a auto-similaridade das formas biológicas (Torres, Falcão *et al.*, 2004).

Para superar esses problemas foi introduzido o método para estimativa da dimensão fractal conhecido como *dimensão fractal multiescala* (Costa, Campos; *et al.*, 2001). Em linhas gerais, essa abordagem consiste na estimação da derivada numérica da função logarítmica ($\log(A(d)) \times \log(d)$). A principal vantagem dessa técnica é a representação da complexidade da forma analisada por meio de uma função de escala espacial. Dessa forma, ao invés de um único valor global para complexidade, temos uma abrangente descrição da auto-similaridade da forma ao longo das diversas escalas espaciais.

A diferenciação da função logarítmica pode ser calculada utilizando a propriedade derivativa da transformada de Fourier (Papoulis, 1962). Além da diferenciação no domínio da frequência, é aplicada uma função gaussiana para suavização (Pinto, 2001). Seja $f(x) \Leftrightarrow F(u)$, uma transformação tempo-frequência da função logarítmica $\log(A(d)) \times \log(d)$, k um número real não negativo que define a ordem de derivação, e $G_\sigma(u)$ uma função gaussiana para suavização. Esse processo da dimensão fractal multiescala pode ser definido como apresentado na Equação 4.11.

$$\frac{d^k f(x)}{dx^k} = F(u)D_k(u)G_\sigma(u) \quad (4.11)$$

em que $D_k(u) = (j2\pi u)^k$, e σ representa o desvio padrão da função gaussiana. Dessa maneira, a dimensão fractal multiescala pode ser expressa por meio de uma função $\log(d) \times \log(A')$, em que A' representa a derivada (e suavização) do ponto $(d, A(d))$ da função logarítmica. Analisando a curva da dimensão fractal multiescala é possível determinar padrões de comportamento para formas de uma mesma classe (ou espécie) (Plotze e Bruno, 2007). Essa característica foi explorada para classificação de espécies vegetais de maracujás silvestres (Plotze, Falvo *et al.*, 2005). A Figura 4.4 ilustra o processo de estimativa da dimensão fractal multiescala em espécies biológicas.

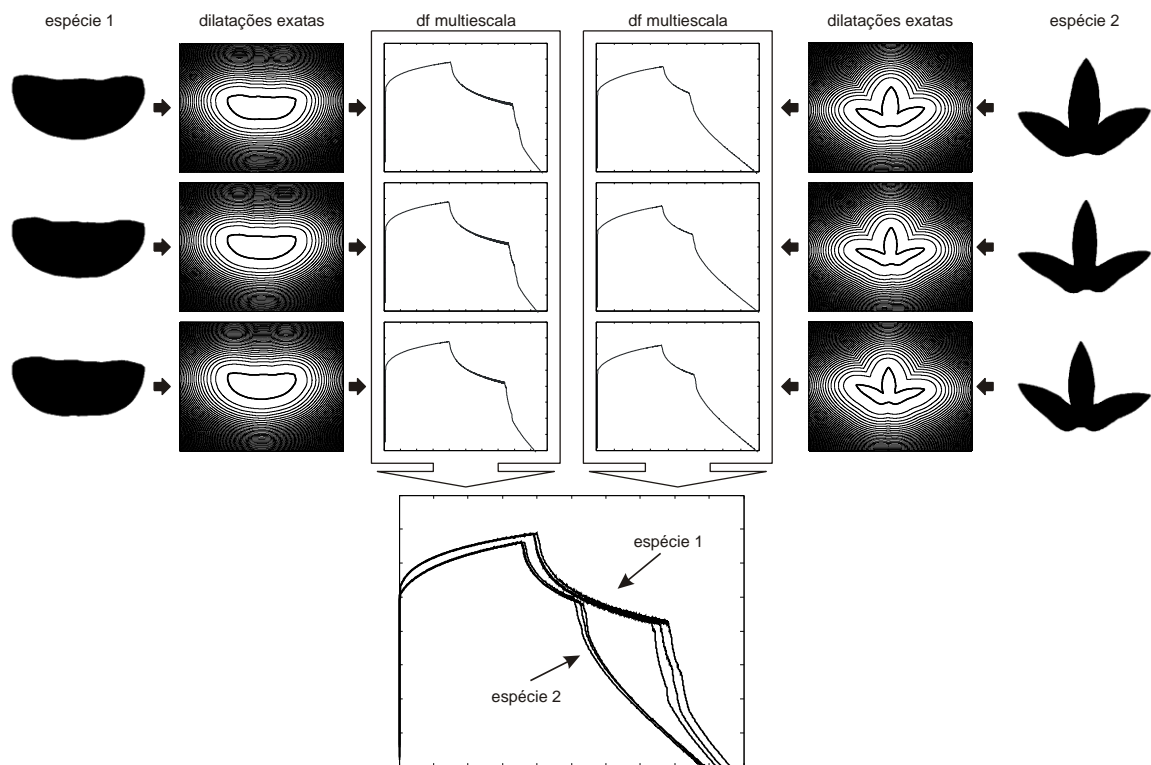


Figura 4.4: Estimativa da dimensão fractal multiescala em espécies biológicas.

4.1.5 Redes Complexas

As redes complexas representam uma importante metodologia para análise de imagens biológicas e estão inseridas na interseção entre as áreas de teoria dos grafos e mecânica estatística (Albert e Barabási, 2002; Costa, Rodrigues *et al.*, 2007). Na literatura, os trabalhos recentes utilizam esses conceitos para análise de formas (Backes, Casanova *et al.*, 2008), análise de textos (Antiqueira, Nunes *et al.*, 2007), neurociência (Ortega, Sola *et al.*, 2008), dentre outros.

Em análise de imagens, os conceitos de redes complexas podem ser utilizados para representação de formas biológicas. Para isso, é necessário que o objeto de interesse na imagem seja caracterizado por meio de um grafo $G = \langle \mathcal{V}, E \rangle$, em que \mathcal{V} representa um conjunto de \mathcal{N} vértices $\{v_1, v_2, \dots, v_{\mathcal{N}}\}$, e E um conjunto de \mathcal{M} arestas $\{e_1, e_2, \dots, e_{\mathcal{M}}\}$ (Rodrigues, 2007).

Seja \mathcal{C} o contorno de uma imagem, tal que $\{c_1, c_2, \dots, c_{\mathcal{N}}\}$ são as coordenadas (x, y) dos \mathcal{N} pixels presentes no contorno. Assim, cada pixel $c_i = (x_i, y_i)$ pode ser expresso como um vértice do grafo, ou seja, $c_i = v_i$. As arestas representam as ligações não-dirigidas

(ou não-direcionadas) entre os pixels, as quais podem ser armazenadas computacionalmente por meio de uma matriz de adjacência. Uma matriz de adjacência \mathcal{A} é uma estrutura utilizada para armazenar as informações a respeito das redes complexas. Esta matriz expressa todas as ligações entre os \mathcal{N} vértices da rede, assim, se dois vértices (v_i, v_j) estão ligados, a posição $a_{i,j}$ na matriz será igual a 1, ou igual a 0, caso contrário.

A cada aresta da rede está associado um peso que representa o valor das ligações entre os vértices. Os pesos são armazenados em uma estrutura conhecida como matriz de pesos \mathcal{W} , e os valores $w_{i,j}$ podem ser calculados por meio da distância euclidiana entre os pixels (c_i, c_j) , tal como representado na Equação 4.12. Os valores dos pesos podem ser normalizados no intervalo [0,1] com a Equação 4.13 (Backes, Casanova *et al.*, 2008).

$$w(c_i, c_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (4.12)$$

$$W = \frac{W}{\max_{w(c_i, c_j) \in W}} \quad (4.13)$$

O resultado da representação do contorno de uma imagem em termos de um grafo é uma *rede regular*, na qual todos os vértices estão conectados entre si. Neste tipo de rede, todos os vértices possuem o mesmo número de conexões, porém, essa estrutura não é considerada uma rede complexa (Wang e Chen, 2003). As redes regulares não contêm informações relevantes para aplicações envolvendo análise de imagens, e com isso, se torna necessária a transformação de uma rede regular em uma rede complexa (Backes, Casanova *et al.*, 2008).

Para isso, um limiar (ou *threshold*) pode ser utilizado para converter uma rede regular em uma rede complexa. Neste processo, um novo grupo de arestas $E^* \subseteq E$ é selecionado sempre que o peso da aresta for menor ou igual a um limiar T_i . Essa operação é realizada em cada elemento da matriz de pesos \mathcal{W} , e com isso, novas matrizes de adjacência são produzidas, como expresso na Equação 4.14 (Backes, Casanova *et al.*, 2008). Essa transformação é representada por $\mathcal{A} = \delta_{T_i}(\mathcal{W})$. As novas redes geradas $G^* = \langle \mathcal{V}, E^* \rangle$ a partir das matrizes de adjacência (\mathcal{A}_{T_i}) são consideradas redes complexas, pois satisfazem o modelo *small-world* proposto por Watts e Strongatz (Watts e Strogatz, 1998).

$$\mathcal{A}_{T_i} = \delta_{T_i}(\mathcal{W}) = \forall w \in \mathcal{W} \begin{cases} a_{i,j} = 0 & \text{se } w_{i,j} \geq T_i, \\ a_{i,j} = 1 & \text{se } w_{i,j} < T_i, \end{cases} \quad (4.14)$$

Diversas propriedades podem ser extraídas das redes complexas a partir das transformações $\delta_{T_i}(\mathcal{W})$. Essas informações contêm aspectos relevantes sobre a caracterização e evolução das redes (Backes e Bruno, 2009). A Figura 4.5 ilustra a evolução dinâmica da estrutura da rede em função da aplicação dos limiares T_i .

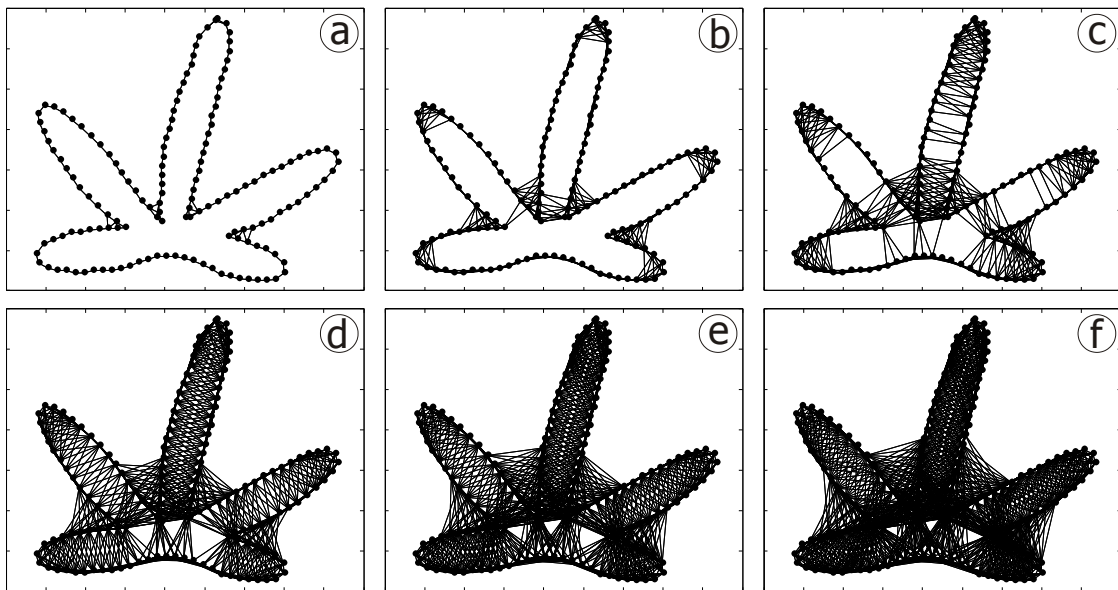


Figura 4.5: Evolução dinâmica da estrutura da rede em função dos limiares T_i ; (a) $T_i = 0.050$; (b) $T_i = 0.100$; (c) $T_i = 0.125$; (d) $T_i = 0.150$; (e) $T_i = 0.175$; (f) $T_i = 0.200$ (adaptado de (Backes, Casanova *et al.*, 2008)).

A evolução dinâmica de uma rede complexa permite a coleta de um conjunto imenso de características. Esses descritores podem ser combinados para formar vetores de características da rede. Os principais descritores de uma rede complexa são obtidos a partir do grau dos vértices. O grau de um vértice representa o número de arestas conectadas ao vértice i da rede, e é denotado por g_i . Esse descritor pode ser calculado como expresso na Equação 4.15. A partir do conjunto de graus dos vértices da rede, outros descritores podem ser extraídos, tais como grau médio g_μ (Equação 4.16), grau máximo g_{max} (Equação 4.17) e grau mínimo g_{min} (Equação 4.18) (Costa, Rodrigues *et al.*, 2007).

$$g_i = \sum_{j=1}^{\mathcal{N}} a_{ij} \quad (4.15)$$

$$g_\mu = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} g_i \quad (4.16)$$

$$g_{max} = \max_i(g_i) \quad (4.17)$$

$$g_{min} = \min_i(g_i) \quad (4.18)$$

Outra propriedade importante que pode ser extraída de uma rede é a correlação entre os graus de diferentes vértices. Essa correlação pode ser expressa em termos da *distribuição das conexões (joint degree distribution)*. A distribuição das conexões, $P(k)$, representa a probabilidade de um vértice i escolhido aleatoriamente ter conectividade k (Rodrigues, 2007). A análise dessa distribuição permite que várias medidas sejam coletadas, tais como, entropia (Equação 4.19), energia (Equação 4.20) e grau médio da distribuição (*average joint degree*) (Equação 4.21) (Costa, Rodrigues *et al.*, 2007).

$$H = - \sum_{i=1}^{\mathcal{N}} P(k) \log_2(P(k)) \quad (4.19)$$

$$E = \sum_{i=1}^{\mathcal{N}} (P(k))^2 \quad (4.20)$$

$$P = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} P(k) \quad (4.21)$$

O coeficiente de aglomeração (*clustering coefficient*) é outra propriedade importante que pode ser calculada de uma rede complexa. Essa característica indica a probabilidade de dois vértices v_A e v_B terem como vizinho um vértice v_C (Rodrigues, 2007). Esse descritor pode ser calculado com a Equação 4.22, em que \mathcal{N}_Δ representa o número de triângulos existentes na rede e \mathcal{N}_3 o número de triplas conectadas. Uma tripla ocorre quando o vértice i é conectado ao vértice j , e j é conectado ao vértice k .

$$C = \frac{3\mathcal{N}_\Delta}{\mathcal{N}_3} \quad (4.22)$$

Outros descritores para redes complexas podem ser encontrados em (Costa, Rodrigues *et al.*, 2007).

4.1.6 Wavelets

Nas últimas décadas, a teoria de *wavelets* emergiu subitamente como uma poderosa metodologia para análise de sinais, e demonstrou enorme aplicabilidade em diversas áreas. Esse método representa uma nova abordagem para decomposição de sinais por meio da superposição de funções base (Sonka, Hlavac *et al.*, 1999). Ao contrário da transformada de Fourier em que as funções base são senoidais, na transformada de *wavelets* (*TW*) as funções base são representadas por pequenas ondas chamadas *wavelets*, que apresentam variações de frequência e duração limitada de tempo (Gonzalez e Woods, 2001).

Basicamente, a decomposição de um sinal $f(x)$ em um conjunto de subsinais é realizada com a translação e dilatação da função base $\Psi(x)$ - conhecida como *wavelet mãe* (Grossmann e Morlet, 1984). Esse processo pode ser entendido como a combinação linear, em diferentes escalas e translações, da *wavelet mãe* $\Psi(x)$ com o sinal $f(x)$. A dilatação da função $\Psi(x)$ é alcançada com um parâmetro a , que expande $\Psi(x/a)$ se $a > 1$, e retrai $\Psi(x)$ se $0 < a < 1$. A localização espacial (translação) da *wavelet mãe* é configurada por meio do parâmetro b - tal que $\Psi(x - b)$. Assim, a dilatação e translação da *wavelet mãe* pode ser expressa como $\Psi((x - b)/s)$ (Parker, 1996). A base ortonormal das *wavelets* é definida pela Equação 4.23 (Sonka, Hlavac *et al.*, 1999).

$$\psi_{(a,b)}(x) = 2^{-(a/2)} \psi(2^{-a} x - b) \quad (4.23)$$

em que o fator de escala (ou dilatação) a indica a largura das *wavelets* $\Psi(a,b)(x)$; e o parâmetro b o índice de localização ao longo do eixo x . Juntamente com a função base, há também uma função de escala como expresso na Equação 4.24.

$$\varphi_{(a,b)}(x) = 2^{-(a/2)} \varphi(2^{-a} x - b) \quad (4.24)$$

Frequentemente, na análise de sinais utilizando *wavelets* as baixas frequências contêm as informações mais relevantes. Convencionalmente esses componentes são chamados de coeficientes de aproximação. Os eventos ocorridos em componentes de baixa frequência são propagados ao longo da evolução do sinal. Por outro lado, nas altas frequências são encontrados componentes específicos, que são conhecidos como coeficientes de detalhe (ou tendência). Os eventos ocorridos na alta frequência duram um pequeno intervalo de tempo. Considerando a transformada de *wavelets* discreta, os coeficientes de aproximação e detalhe podem ser expressos na Equação 4.25 e na Equação 4.26, respectivamente (Gonzalez e Woods, 2001).

$$W\phi(a_0, b) = \frac{1}{\sqrt{M}} \sum_x f(x)\phi_{a_0, b}(x) \quad (4.25)$$

em que a_0 representa uma escala arbitrária inicial.

$$W\psi(a, b) = \frac{1}{\sqrt{M}} \sum_x f(x)\psi_{a, b}(x) \quad (4.26)$$

Assim, um sinal unidimensional $f(x)$ é formado pelos componentes de aproximação ($W\phi$) e detalhe ($W\psi$) como expressos na Equação 4.27. Os componentes $W\phi$ e $W\psi$ também podem ser considerados como filtros de passa-alta e passa-baixa, respectivamente. Similarmente aos descritores de Fourier, os coeficientes de *wavelets* também podem ser empregados na extração de características do sinal ou forma biológica. Os componentes, tanto de aproximação quanto de detalhe, podem ser selecionados a partir de uma determinada escala a , ou localização b .

$$f(x) = \frac{1}{\sqrt{M}} \sum_b W_\phi(a_0, b)\phi_{a_0, b}(x) + \frac{1}{\sqrt{M}} \sum_{a=a_0} \sum_b W_\psi(a, b)\psi_{a, b}(x) \quad (4.27)$$

A Figura 4.6 ilustra a decomposição de um sinal por meio da transformada de *wavelets* utilizando a abordagem multiescala. Nesse exemplo foi utilizada uma *Wavelet* mãe da família

Daubechies (*db4*). Os coeficientes de aproximação são representados pelo conjunto $\{a_0, a_1, a_2, a_3\}$, e os componentes de detalhe $\{d_1, d_2, d_3\}$.

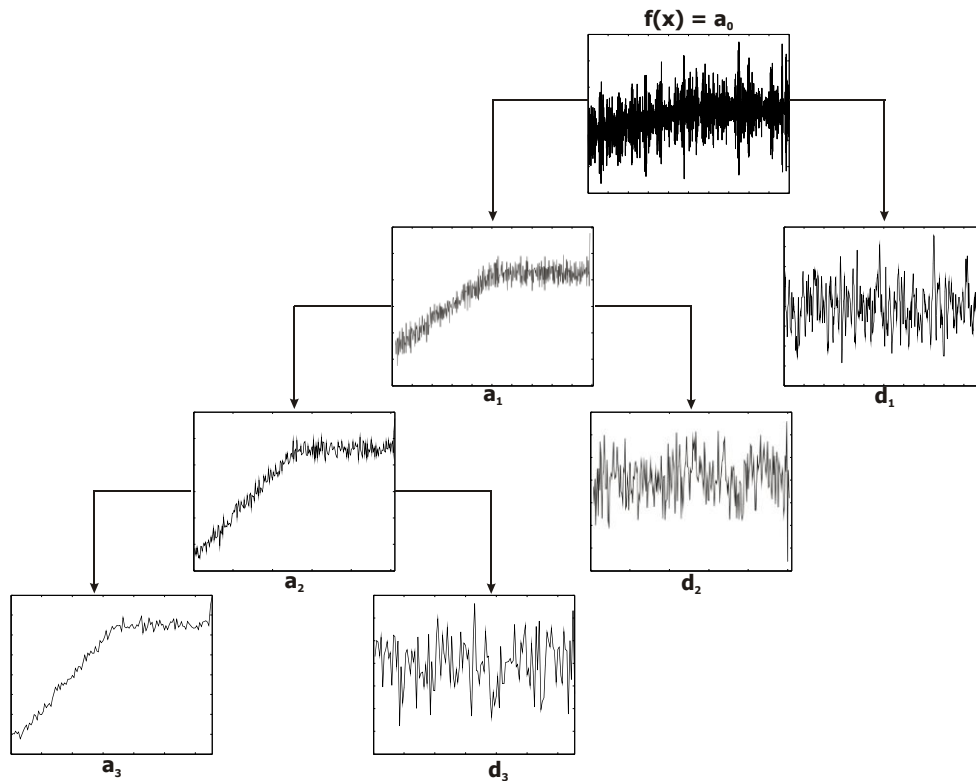


Figura 4.6: Decomposição de sinal *ID* por meio da transformada de wavelets.

A transformada de *wavelets* é formada por diversas famílias de *wavelets* mães $\Psi(x)$, sendo as mais populares (Boggess e Narcowich, 2001): *Haar* – que possui os coeficientes mais simples e suporte compacto, porém não é continuamente diferenciável, gerando assim algumas limitações; *Daubechies* – é uma família ortonormal, com suporte compacto e têm sido bastante empregada em processamento de sinais; *Symlets* – semelhante a família *Daubechies*, também possui suporte compacto, porém menos simétricas. Outras famílias bastante conhecidas são as *wavelets* mães *Coiflets* e *Morlet* (Castleman, 1996). A Figura 4.7 apresenta algumas das principais famílias de *wavelets* mães.

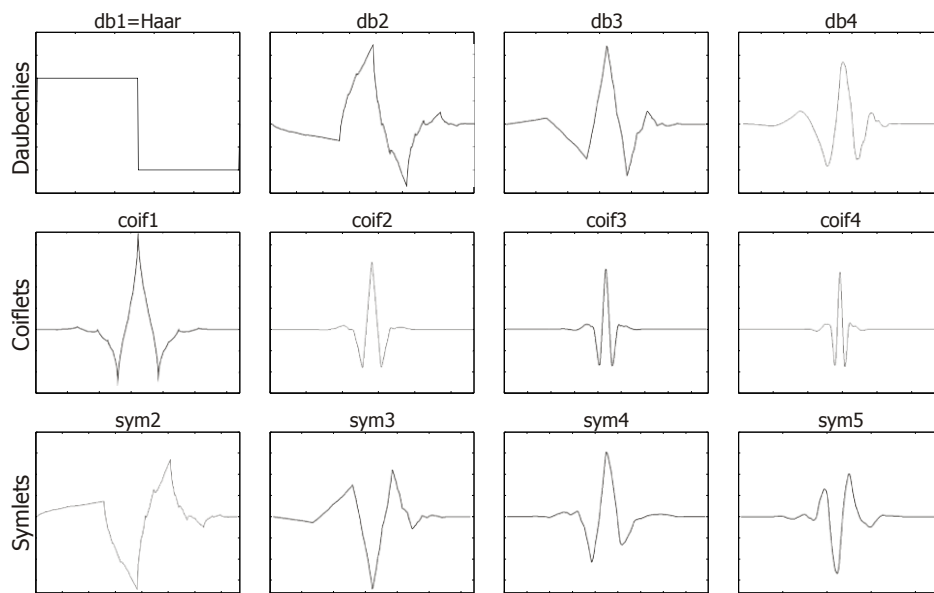


Figura 4.7: Principais famílias de wavelets mães: Daubechies (db), Coiflets (coif) e Symlets (sym).

A Figura 4.8 ilustra o processo de análise de formas biológicas com a transformada de *wavelets*. Inicialmente, o contorno da forma é extraído com o algoritmo seguidor de contorno e representado por meio de um sinal complexo. Uma família de *wavelets* deve ser selecionada para análise, bem como um nível para extração das características de tendência e detalhe. Na Figura 4.8, foram definidas a família *db4* e o nível 5.

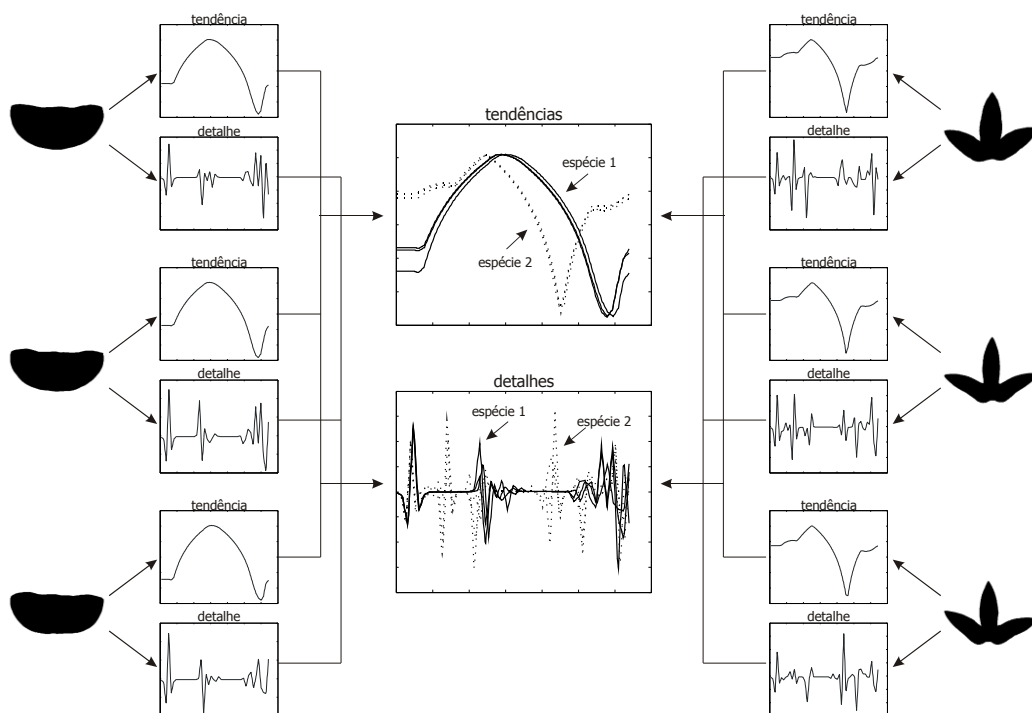


Figura 4.8: Análise de formas biológicas com a transformada de wavelets. Processo de extração de tendência e detalhe para a família *db4* com o nível 5.

4.2 Análise de Estruturas Tubulares

Na natureza podemos encontrar uma ampla diversidade de estruturas tubulares que exercem um papel fundamental em diversos ramos da biologia (Clarke, Barman *et al.*, 2006). Essas estruturas podem ser descritas como uma complexa e irregular rede de caminhos interconectados, que na maioria das vezes, transportam informações indispensáveis para o funcionamento dos organismos. Podemos citar como exemplos de estruturas tubulares os vasos sanguíneos dos seres vivos, o sistema de venação das plantas, as ramificações dos neurônios, os troncos e galhos das árvores, os esqueletos dos vertebrados, dentre outros (Milner, Raz *et al.*, 2007). Essas estruturas são ricas em informações morfométricas que podem ser empregadas no estudo e caracterização de formas biológicas. Além disso, a enorme semelhança entre essas estruturas possibilita o desenvolvimento de análises morfométricas genéricas para diversas aplicações, ou seja, um método para morfometria do sistema de venação foliar pode também ser utilizado na caracterização dos vasos sanguíneos da retina.

Dentro desse contexto, esta seção apresenta um novo conjunto de técnicas que foram desenvolvidas para a análise morfométrica de estruturas tubulares. Para validação da metodologia proposta, as técnicas foram empregadas na extração de características do sistema de venação foliar das espécies de *Passiflora*. As características das imagens tanto das espécies de eucalipto, quanto de arroz impossibilitaram a utilização da metodologia.

4.2.1 Biometria

A organização hierárquica das estruturas tubulares possibilita o desenvolvimento e aplicação de uma infinidade de métodos de análise morfométrica. Essas técnicas podem ser empregadas na extração de características locais ou globais da estrutura tubular. Na análise local são coletadas informações de uma região específica da estrutura, como por exemplo, o ângulo de uma determinada ramificação, a distância entre a ramificação e os seus pontos extremos, dentre outras. Por outro lado, a análise global permite uma descrição geométrica completa da estrutura tubular, com isso, diversas características podem ser extraídas, tais como: número de bifurcações, número total de ramificações, tamanho da ramificação principal, total de pontos extremos (pontos de alta curvatura), distância média entre as bifurcações, distância média entre os pontos extremos, etc. A Figura 4.9 ilustra algumas das principais características que podem ser extraídas das estruturas tubulares.

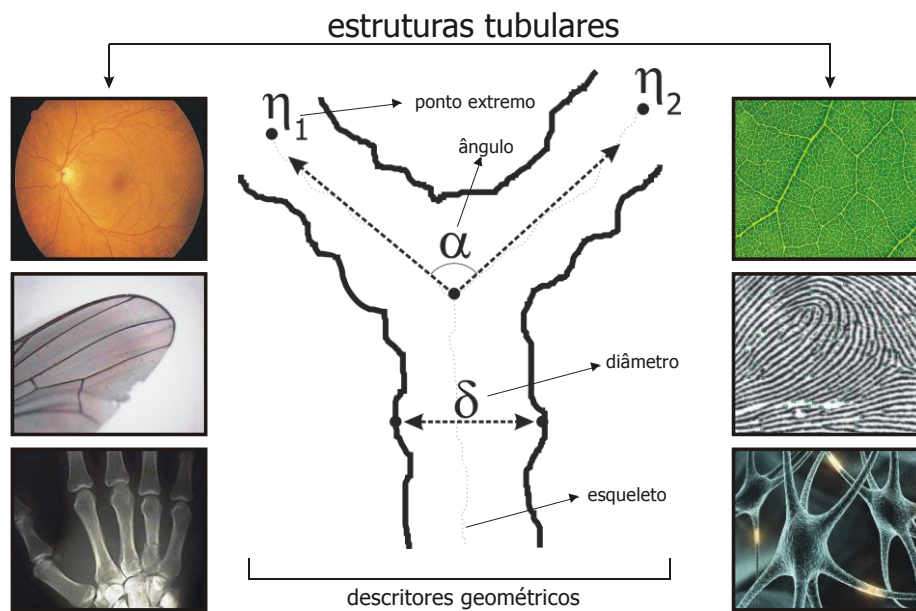


Figura 4.9: Extração de características morfométricas de estruturas tubulares.

Diversas informações estatísticas podem ser extraídas com base nas principais características morfométricas das estruturas tubulares: diâmetro, ângulo e pontos extremos. Essas características representam uma completa e quantitativa descrição das formas biológicas, uma vez que contêm dados sobre as espessuras, bifurcações, e arranjos geométricos. Além disso, essas informações podem ser comparadas e correlacionadas, e dessa maneira, pode ser realizada uma investigação evolutiva e taxonômica das estruturas tubulares das espécies.

Em botânica, os métodos biométricos são fundamentais para identificação de espécies vegetais e são baseados em rigorosos protocolos de taxonomia (McLellan e Endler, 1998; Bockhoff, 2001; Wing, Wilf *et al.*, 2001). No entanto, as características são coletadas manualmente, o que torna o procedimento de identificação complexo e pouco confiável. Neste sentido, uma nova metodologia foi proposta para extração de medidas biométricas do sistema de venação foliar (Plotze e Bruno, 2009).

4.2.2 Extração Automática de Características Biométricas

As características biométricas das imagens são extraídas com um algoritmo que navega no sistema de venação foliar. Para que a navegação seja possível, inicialmente as imagens são filtradas para redução dos ruídos e o sistema de venação foliar é segmentado com a

abordagem descrita na Seção 3.3.1 (Plotze, Nardin *et al.*, 2006). A abordagem utilizada para segmentação é baseada no processo de clareamento (diafanização) do material biológico, seguida da filtragem por difusão anisotrópica.

Com o sistema de venação segmentado, o próximo passo é a navegação no interior da estrutura tubular. Para isso, o algoritmo determina o ponto inicial do sistema de venação foliar. Essa busca pode ser executada de duas maneiras: da parte superior da imagem para baixo (da esquerda para direita) ou da parte inferior para cima (da direita para esquerda). Esse método é similar aos algoritmos para extração de contorno (Pratt, 2001). O primeiro pixel encontrado é definido como *ponto inicial*.

Com a determinação do *ponto inicial*, o algoritmo de navegação é iniciado. A ideia principal do método é visitar todos os pontos (pixels) da estrutura de venação, coletando informações morfológicas dos pixels. Considerando uma imagem binária na qual os pixels de fundo (*background*) são representados pelo valor 0 (zero), e os pixels de interesse (*foreground*) são representados por 1 (um), e a estrutura de interesse (nervuras) possui um pixel de largura, cada pixel pode ser classificado com base nos seguintes critérios: (i) bifurcação; (ii) ponto extremo ou (iii) intermediário.

Um pixel de *bifurcação* determina o início de uma nova ramificação no interior da estrutura tubular. No sistema de venação foliar, as bifurcações são responsáveis pela interconexão das nervuras primárias, secundárias e terciárias. Esses pixels são definidos como se segue:

Definição 1: Seja I uma imagem binária, na qual cada pixel contém até oito vizinhos. Um pixel p é classificado como bifurcação, quando o número de transições de 0 (zero) para 1 (um), na sequência: $p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8$ e p_1 , é maior que 2 (dois). Em outras palavras, um pixel p é uma bifurcação quando $\mathcal{A}(p) > 2$. A Figura 4.10 ilustra a definição de um pixel de bifurcação.

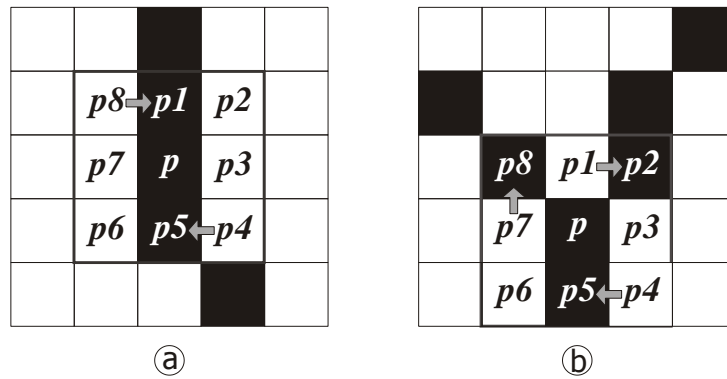


Figura 4.10: Determinação do pixel de bifurcação na estrutura tubular. (a) O pixel p não é uma bifurcação, pois $\mathcal{A}(p) \leq 2$; (b) O pixel p é uma bifurcação $\mathcal{A}(p) > 2$.

O conceito de pixel *extremo* pode ser entendido como sendo o último pixel presente na estrutura tubular. Nas nervuras, o pixel extremo representa o final de uma ramificação. Esses pixels são definidos como se segue:

Definição 2: Seja p um pixel de interesse (*foreground*), e $\mathcal{B}(p)$ o número de vizinhos adjacentes a p com valor diferente de zero. Se o número de vizinhos de p é igual a 1 (um), ou seja, $\mathcal{B}(p) = 1$, então p é classificado como ponto extremo (ou terminal).

A Figura 4.11 ilustra graficamente os conceitos de pixel de bifurcação e de pixel extremo. Na representação matricial ilustrada na Figura, as posições em branco referem-se aos pixels de fundo, as posições em preto a estrutura tubular, e as posições denotadas por \boxtimes representam as bifurcações e os pontos extremos.

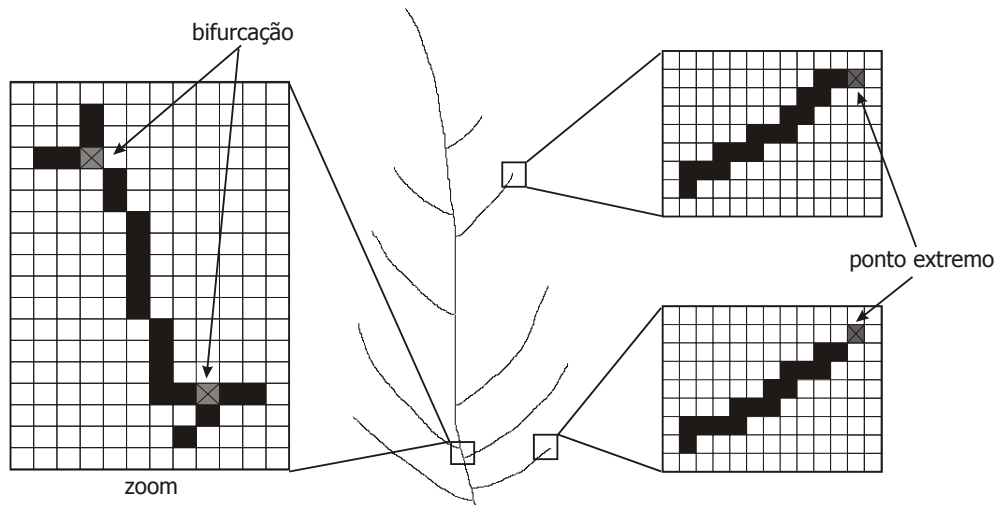


Figura 4.11: Extração das características biométricas da estrutura tubular: determinação das bifurcações e dos pontos extremos.

Todos os pixels que não pertencem ao conjunto de bifurcações e aos pontos extremos são denotados como pixels *intermediários*.

O algoritmo de navegação realiza uma exploração sistemática da estrutura tubular. Para garantir que todos os pixels sejam inspecionados, e que o algoritmo não entre em repetições infinitas, os pixels explorados são rotulados como já visitados. A partir do ponto inicial, o processo de exploração da estrutura ocorre como descrito no Algoritmo 1.

Algoritmo 1: Exploração sistemática da estrutura tubular.

```
01. função ExplorarEstruturaTubular(vetor de pixels de interesse[])
02. {
03.     enquanto (existir pixels para serem visitados)
04.     {
05.         sinalizar o pixel  $p$  como visitado
06.
07.         para todos os vizinhos de  $p$ 
08.         {
09.             se ( $p$  não foi visitado )
10.                 adicionar  $p$  a pilha
11.         }
12.
13.         se ( pilha.tamanho() == 0 )
14.         {
15.             todos os pixels da estrutura já foram visitados
16.         }
17.         senão se ( pilha.tamanho() == 1 )
18.         {
19.             mover para o próximo vizinho e sinalizá-lo como visitado
20.
21.             se  $B(p) = 1$  // ponto extremo
22.                 adicionar ao vetor de pontos extremos
23.         }
24.         senão se ( pilha.tamanho() > 1 )
25.         {
26.             se  $A(p) > 2$  // bifurcação
27.                 adicionar ao vetor de bifurcações
28.             senão
29.                 mover para o próximo vizinho e sinalizá-lo como visitado
30.
31.             para todos os elementos da pilha
32.                 ExplorarEstruturaTubular() // chamada recursiva
33.         }
34.     }
35. }
```

O funcionamento do algoritmo é baseado em uma *pilha*, na qual são armazenados todos os vizinhos do pixel atual (p) que não foram visitados. Se o pixel p_i é um vizinho do pixel atual e não foi visitado, então ele é adicionado a pilha. Quando o número de elementos da pilha é igual a zero (tamanho==0), significa que todos os pixels da estrutura (ou de uma das ramificações) já foram visitados (linha 13).

O tamanho da pilha igual a 1 (um), demonstra que o pixel atual possui um único vizinho (linha 17). Nesse caso, o pixel atual é rotulado como visitado e seu próximo vizinho é selecionado. O próximo vizinho poderá ser classificado como ponto extremo, caso $B(p) = 1$, então a coordenada desse pixel é adicionada ao vetor de pontos extremos (linha 22).

Outra situação que pode ocorrer durante a exploração da estrutura tubular é quando o número de elementos da pilha é maior que 1 (um). Neste caso, o pixel atual tem mais de um vizinho (linha 24). Para essa condição podemos ter duas possibilidades: (i) se o pixel atual não é uma bifurcação $\mathcal{A}(p) \leq 2$, então o algoritmo deverá determinar entre os pixels possíveis qual será o próximo pixel e (ii) se o pixel atual é uma bifurcação $\mathcal{A}(p) > 2$, o algoritmo adiciona o pixel atual ao vetor de bifurcações (linha 27). Quando uma bifurcação é encontrada, o algoritmo precisa explorar detalhadamente todos os vizinhos do pixel atual que não foram visitados. Esses vizinhos representam as ramificações na estrutura tubular, assim, chamadas recursivas da função *ExplorarEstruturaTubular()* são disparadas, tendo como ponto inicial a coordenada de cada um desses vizinhos (linha 32). Esse fato garante que todas as ramificações da estrutura serão exploradas. O critério de parada da função recursiva (*backtracking*) é quando o número de elementos da pilha é igual a zero.

Durante o processo de exploração sistemática da estrutura tubular, diversas informações biométricas podem ser coletadas. Essas características podem ser agrupadas em três categorias: (i) distâncias; (ii) tamanho; e (iii) ângulos.

Na primeira categoria, características biométricas podem ser extraídas com base na distância entre determinados pontos da estrutura tubular. Utilizando as informações contidas no vetor de bifurcações e no vetor de pontos extremos, as seguintes características podem ser extraídas: número de bifurcações; distância média entre todas as bifurcações; desvio padrão da distância entre as bifurcações; variância da distância entre as bifurcações; distâncias, médias, desvio padrão e variância entre todas as bifurcações e o ponto de origem da estrutura tubular (denominado na morfologia foliar como pecíolo); dentre outras.

A segunda categoria é responsável por coletar medidas biométricas relacionadas ao tamanho. Neste caso, são utilizadas informações do caminho percorrido pelo algoritmo durante a exploração da estrutura. Algumas medidas que podem ser extraídas são: a área da estrutura (em pixels); o número de ramificações; tamanho médio das ramificações; desvio padrão do tamanho das ramificações; tamanho da ramificação principal; dentre outras. Os protocolos de taxonomia definem a ramificação (nervura) principal, que é o caminho entre a base (pecíolo) da estrutura até o topo, como sendo uma característica fundamental para taxonomia das espécies (Wing, Wilf *et al.*, 2001).

O terceiro grupo de medidas corresponde aos ângulos formados entre as ramificações da estrutura. Essas informações podem ser extraídas para todas as ramificações (análise global), ou a partir de uma ramificação específica (análise local). A extração dos ângulos entre as ramificações é executada da seguinte maneira: (i) Seja p_1 um pixel de bifurcação ($\mathcal{A}(p) > 2$) definido como a raiz da ramificação; (ii) \mathcal{L}_n uma lista contendo todos os pixels vizinhos de p_1 (em que n é igual o número de vizinhos); (iii) para cada ramificação dos vizinhos p_1 navegar um número pré-determinado de pixels até encontrar os pixels p_n ; (iv) formar vetores entre os pixels p_1, p_2, \dots, p_n .

Para uma bifurcação contendo duas ramificações temos a seguinte situação: p_1 como sendo a raiz da ramificação, p_2 e p_3 representando os pixels encontrados durante a navegação nas ramificações. Com isso, podemos definir o vetor \vec{u} entre os pontos p_1 e p_2 , e \vec{v} o vetor entre os pontos p_1 e p_3 . Se a coordenada y do pixel p_2 é maior que a coordenada y de p_1 ($p_{2,y} > p_{1,y}$) então o ângulo de abertura α_1 é extraído; entretanto, se $p_{2,y} < p_{1,y}$ o ângulo de fechamento α_2 é extraído. Os ângulos são determinados utilizando a notação vetorial, conforme expresso na Equação 4.28.

$$\cos \alpha = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \|\vec{v}\|} \quad 0 \leq \alpha \leq \pi \quad (4.28)$$

A Figura 4.12 ilustra graficamente a extração dos ângulos de abertura e fechamento. Com esses dados coletados, diversas informações estatísticas podem ser obtidas, tais como média entre os ângulos de abertura; média entre os ângulos de fechamento; desvio padrão dos ângulos, dentre outras.

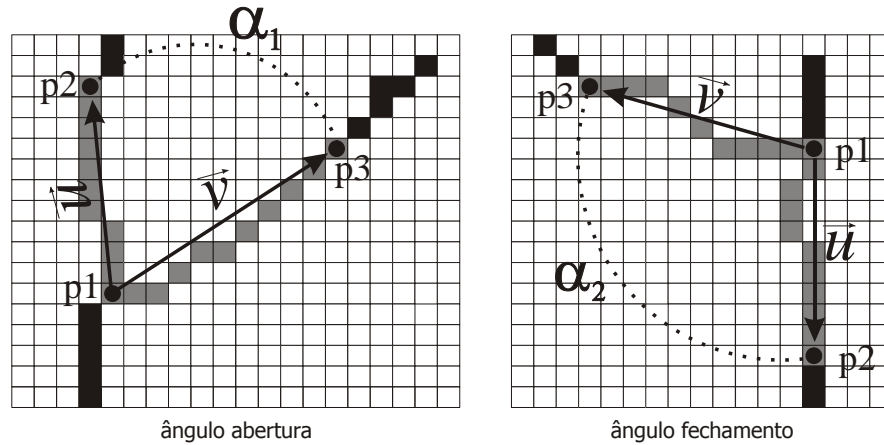


Figura 4.12: Determinação dos ângulos entre as ramificações da estrutura tubular: ângulo de abertura (α_1); ângulo de fechamento (α_2).

4.2.3 Assinatura Fractal Multiescala

A dimensão fractal multiescala é uma técnica utilizada em análise de imagens biológicas que apresenta ótimos resultados (Castrejón-Pita, Sarmiento-Galán *et al.*, 2005; Plotze, Falvo *et al.*, 2005). No entanto, a maioria dos trabalhos na literatura emprega essa metodologia na análise do contorno de formas biológicas. A abordagem baseada em contorno, apesar de eficiente em alguns casos, não é capaz de caracterizar adequadamente qualquer tipo de forma biológica. Além disso, quando comparada a outras metodologias, a dimensão fractal multiescala do contorno não demonstra resultados satisfatórios (Backes e Bruno, 2009).

Dentro deste contexto, a assinatura fractal multiescala é uma nova abordagem para análise de estruturas tubulares. Como essas estruturas estão presentes nos mais variados tipos de organismos na natureza, essa abordagem tem grande aplicabilidade em problemas biológicos. Além disso, analogamente, uma estrutura tubular tem as mesmas características de um esqueleto resultante do processo de afinamento (Plotze e Bruno, 2004). Assim, além das estruturas tubulares, essa metodologia pode ser empregada na análise de esqueletos.

Seja I uma imagem digital, e E_T uma estrutura tubular devidamente segmentada, a assinatura fractal multiescala pode ser calculada como ilustrado na Figura 4.13. Na literatura, trabalhos demonstram que a assinatura fractal multiescala de estruturas tubulares contém relevantes informações para caracterização de formas biológicas (Plotze e Bruno, 2005).

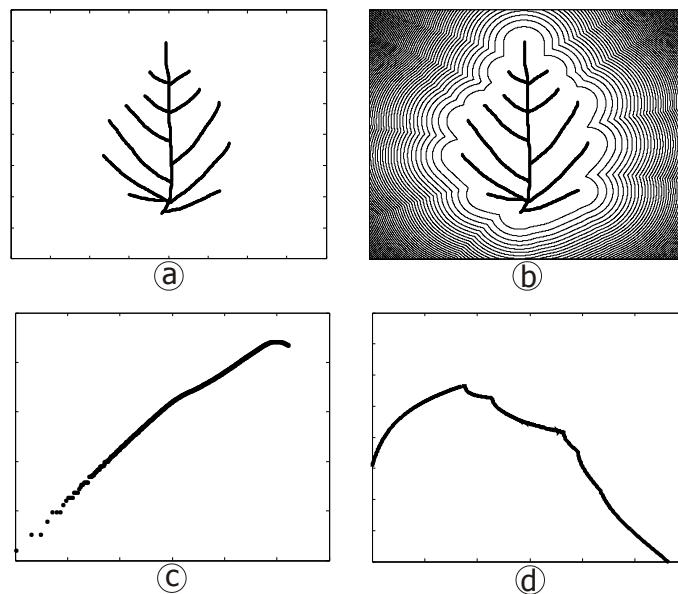


Figura 4.13: Assinatura fractal multiescala de estruturas tubulares: (a) estrutura tubular; (b) dilatações exatas; (c) curva $\log(A(d)) \times \log(d)$; (d) assinatura fractal multiescala.

Uma contribuição importante da assinatura fractal multiescala é a capacidade de representar imagens bidimensionais (de estruturas tubulares) por meio de uma assinatura unidimensional. Esse fato possibilita que diversas técnicas de processamento e análise de sinais sejam aplicadas na assinatura. Além disso, essa metodologia reduz a dimensionalidade das informações analisadas, o que é um fator fundamental em processos de reconhecimento de padrões.

Além da assinatura fractal multiescala, este trabalho apresenta uma nova abordagem para análise da complexidade de estruturas tubulares. Nesta nova metodologia, os pontos biométricos da estrutura tubular (bifurcações e pontos extremos) são considerados como informações de interesse para o cálculo da assinatura fractal multiescala. Com isso, os rigorosos conceitos dos protocolos de taxonomia de espécies são agregados às técnicas de visão artificial.

A assinatura fractal baseada em descritores biométricos permite que a análise da complexidade da estrutura tubular seja realizada de três maneiras: (i) em função dos pontos biométricos de bifurcação; (ii) a partir dos pontos extremos da estrutura e (iii) combinando as informações biométricas das bifurcações e dos pontos extremos. A Figura 4.14 apresenta a assinatura fractal multiescala baseada em descritores geométricos.

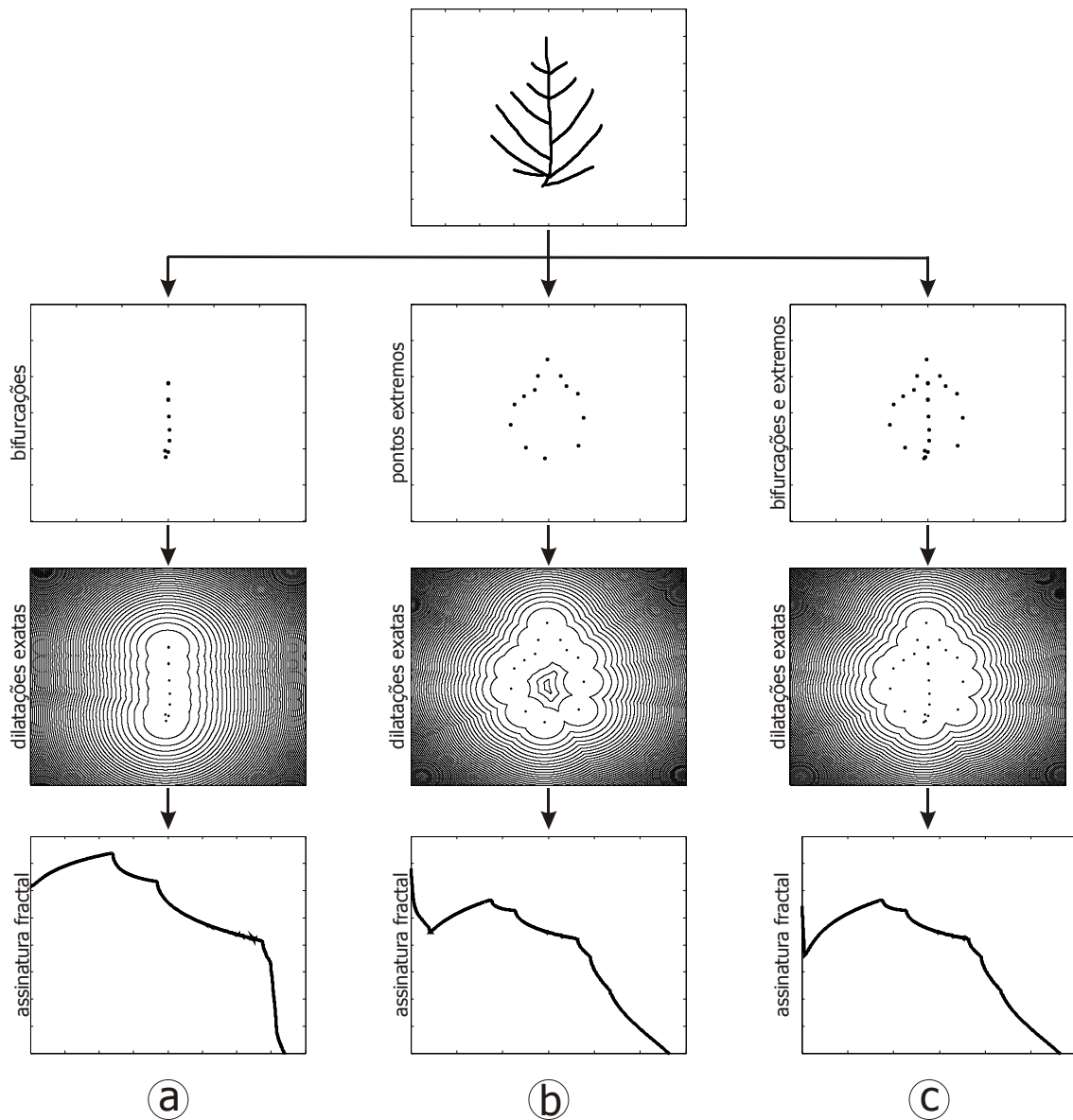


Figura 4.14: Assinatura fractal multiescala baseada em descritores biométricos: (a) assinatura gerada em função dos pontos biométricos de bifurcação; (b) cálculo da assinatura a partir dos pontos extremos e (c) combinação das características biométricas das bifurcações e dos pontos extremos para produção da assinatura fractal multiescala.

4.2.4 Modelagem de Estruturas Tubulares por Redes Complexas

As redes complexas tem demonstrado grande potencial para análise e descrição de imagens, permitindo a representação de informações visuais a partir de um grafo. A evolução dinâmica de uma rede complexa pode ser acompanhada a partir de um limiar, com isso, um conjunto de diversas propriedades pode ser extraído da rede.

Para estruturas tubulares, uma nova metodologia foi proposta com objetivo de mapear os pontos de interesse da estrutura por meio de uma rede complexa. Considerando E_T como o conjunto de \mathcal{N} pontos de interesse de uma estrutura tubular, denotado por $E_T = \{p_1, p_2, \dots, p_N\}$ sendo $p_i = (x_i, y_i)$, uma rede complexa pode ser construída transformando E_T em vértices da rede e aplicando um limiar T_i nas matrizes de pesos das arestas. A Figura 4.15 ilustra uma rede complexa de uma estrutura tubular gerada a partir de um limiar T_i .

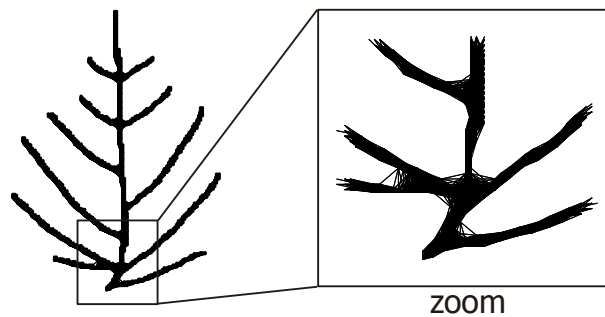


Figura 4.15: Representação de uma estrutura tubular por meio de uma rede complexa.

Outra abordagem para modelagem de uma estrutura tubular por meio de uma rede complexa foi proposta, na qual os pontos de interesse da estrutura são resumidos aos pontos biométricos de bifurcação. Assim, considerando E_T como o conjunto de pontos de interesse da estrutura tubular, e B_i como os pontos de bifurcação, tal que a seguinte relação seja satisfeita $B_i \subseteq E_T$, a estrutura tubular pode ser modelada transformando B_i nos vértices da rede complexa.

Com base na proposta descrita anteriormente, em que a rede complexa é modelada a partir dos pontos de bifurcação de uma estrutura tubular, duas outras abordagens podem ser propostas: (i) representação da estrutura tubular por meio de uma rede complexa utilizando apenas os pontos extremos e (ii) construção de uma rede complexa a partir da combinação das características biométricas de bifurcação e dos pontos extremos. O potencial das abordagens propostas foi avaliado por meio de experimentos com problemas reais de análise de imagens de estruturas tubulares. A Figura 4.16 ilustra a modelagem de uma estrutura tubular por meio de uma rede complexa a partir das bifurcações e dos pontos extremos.

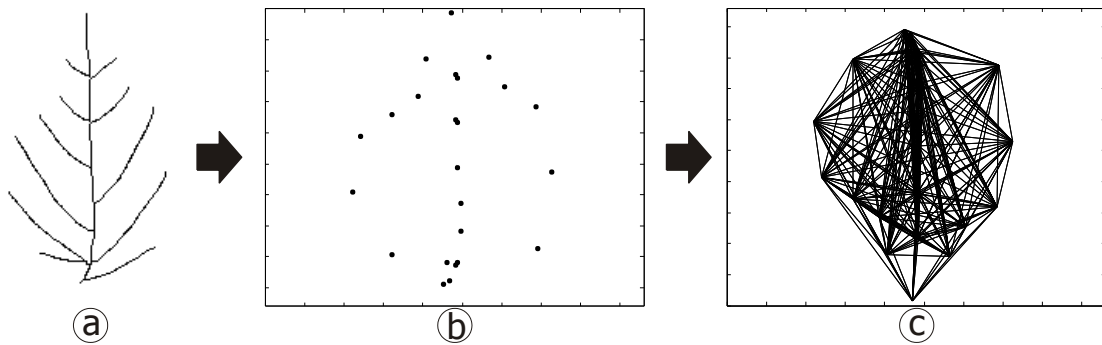


Figura 4.16: Modelagem de uma estrutura tubular por meio de uma rede complexa a partir de bifurcações e pontos extremos: (a) estrutura tubular; (b) características biométricas; (c) rede complexa para $T_i=0.50$

4.3 Análise de Textura

Nas últimas décadas, as técnicas de classificação e segmentação de textura em imagens digitais têm recebido grande atenção da comunidade acadêmica. Tanto os sistemas de visão biológica quanto artificial utilizam as características de textura para percepção de informações do mundo real (Mittal, Mital *et al.*, 1999). Estudos psicofisiológicos demonstraram que a discriminação da textura acontece paralelamente em grandes regiões do campo visual, e ocorre no início do processamento visual de informações (Campbell e Robson, 1968). A análise de textura é uma tarefa complexa, principalmente devido a grande diversidade de texturas naturais e artificiais, por isso, uma grande variedade de abordagens e aplicações foram propostas na literatura. As técnicas de análise de textura podem ser classificadas em cinco categorias principais, que dependem inerentemente das características extraídas para identificação da textura (Tuceryan e Jain, 1993). Nesta seção são apresentados os conceitos a respeito da textura estatística, além disso, são discutidas duas novas propostas para análise de textura baseadas em dimensão fractal multiescala e redes complexas.

4.3.1 Textura Estatística

Nas técnicas estatísticas, as texturas são descritas por regras que determinam a distribuição e a relação dos níveis de cinza da imagem. Nessa abordagem a textura representa uma medida quantitativa das intensidades de uma determinada região da imagem (Shapiro e Stockman, 2001).

A função de autocorrelação é um método de análise de textura estatística baseado nas frequências espaciais (Sonka, Hlavac *et al.*, 1999). Essa técnica avalia a organização espacial da textura por meio de um coeficiente de correlação. Se as primitivas de textura são grandes, a função de autocorrelação diminui lentamente com o aumento da distância, caso contrário, a função diminui rapidamente quando a textura é composta por pequenas primitivas. A função de autocorrelação é expressa na Equação 4.29.

$$Cff(p, q) = \frac{MN}{(M-p)(N-q)} \frac{\sum_{i=1}^{M-p} \sum_{j=1}^{N-q} f(i, j)f(i+p, j+q)}{\sum_{i=1}^M \sum_{j=1}^N f^2(i, j)} \quad (4.29)$$

Os parâmetros p e q representam diferentes posições na imagem para variadas direções i, j ; enquanto M e N são as dimensões da imagem.

Uma das abordagens estatísticas mais utilizadas para análise de textura é a matriz de coocorrência, introduzida por *Haralick* (Haralick, Shanmugam *et al.*, 1973). Essa técnica tem como objetivo caracterizar a textura da imagem por meio da análise das probabilidades de transição dos níveis de cinza. Os elementos da matriz são formados pela contagem dos níveis de cinza da imagem em função de uma determinada métrica de distância, para um ângulo de inclinação específico (Nixon e Aguado, 2002). Computada a matriz de coocorrência, diversas características podem ser extraídas, sendo algumas delas: segundo momento angular ou energia (Equação 4.30), contraste (Equação 4.31), correlação (Equação 4.32), entropia (Equação 4.33), dentre outras.

$$Ener = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p_{i,j}^2 \quad (4.30)$$

$$Con = \sum_{n=0}^{G-1} N^2 \left\{ \sum_{|i-j|=n} p_{i,j} \right\} \quad (4.31)$$

$$Cor = \frac{1}{\sigma_x \sigma_y} \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} ij p_{i,j} - \mu_x \mu_y \quad (4.32)$$

$$Ent = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p_{i,j} \log p_{i,j} \quad (4.33)$$

O número total de níveis de cinza da imagem corresponde ao parâmetro G ; enquanto μ e σ correspondem as médias e desvio padrão da distribuição dos pixels p_{ij} .

4.3.2 Textura por Distribuição de Níveis de Cinza

A análise de textura baseada na distribuição de níveis de cinza considera a intensidade de cada pixel presente na imagem como característica para descrição da textura. Essa metodologia permite que diversas informações sejam coletadas tais como, média, variância, desvio padrão, simetria, etc. (Pedrini e Schwartz, 2008).

O histograma dos níveis de cinza de uma imagem também pode ser utilizado para extração de características da textura (Ojala, Pietikainen *et al.*, 1996). O histograma é uma função discreta que indica a quantidade pixels que contém um determinado nível de cinza (Gonzalez e Woods, 2001). Essa função pode ser calculada em termos da distribuição de probabilidade de níveis de cinza como expresso na Equação 4.34, na qual h_i representa o número de ocorrências de pixels com intensidade i e n o número de pixels da textura (Pedrini e Schwartz, 2008). A Figura 4.17 ilustra o histograma de uma imagem em níveis de cinza.

$$P_i = \frac{h_i}{n} \quad (4.34)$$

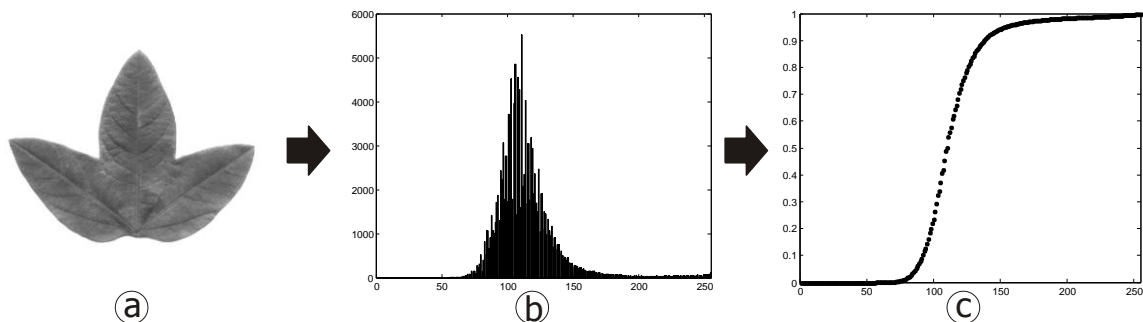


Figura 4.17: Histograma de uma imagem em níveis de cinza: (a) imagem original; (b) histograma e (c) distribuição acumulativa da probabilidades dos níveis de cinza.

Baseado na distribuição de níveis de cinza de uma imagem, uma nova abordagem foi proposta para extração de características em texturas. Inicialmente, um subconjunto de pixels da imagem é selecionado com base na probabilidade de ocorrência de uma determinada intensidade. Seja P a distribuição de probabilidades de níveis de cinza de uma imagem I , e k um limiar utilizado para seleção, um pixel $p \in I$ será selecionado sempre que $P_i \geq k$. Em outras palavras, um pixel da imagem será selecionado sempre que a distribuição de probabilidade de níveis de cinza deste pixel for maior ou igual que o limiar. Este procedimento gera um novo subconjunto de pixels da textura da imagem que pode ser utilizado para extração de características. A Figura 4.18 ilustra a metodologia proposta para extração de características a partir da distribuição de níveis de cinza da imagem. Na Figura é possível observar que o valor adotado como limiar produz um efeito multiescala na imagem.

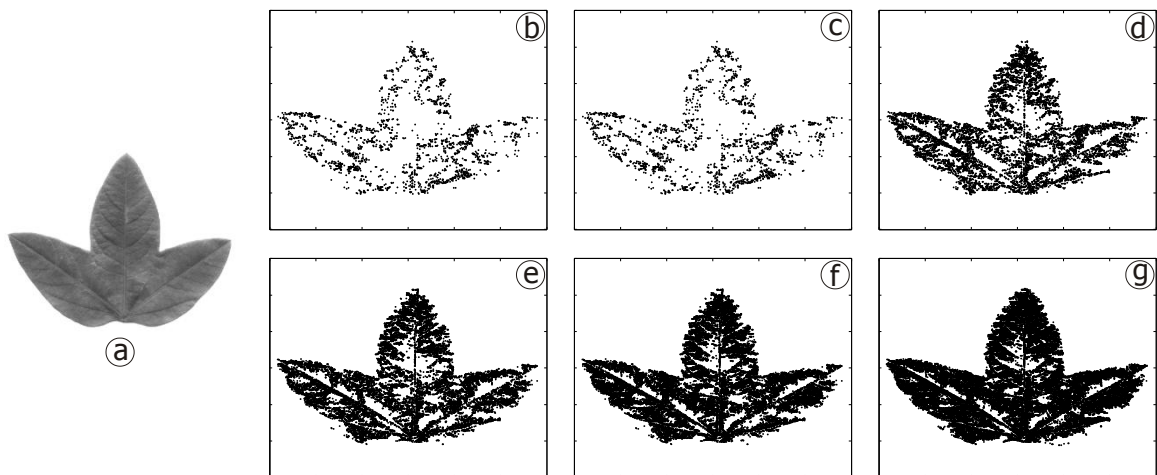


Figura 4.18: Seleção de um subconjunto de pixels de textura com base na distribuição de probabilidade de níveis de cinza: (a) imagem original em níveis de cinza; (b) seleção para $k=1.0$; (c) seleção para $k=0.9$; (d) seleção para $k=0.8$; (e) seleção para $k=0.7$; (f) seleção para $k=0.6$; (g) seleção para $k=0.5$;

O próximo passo da metodologia agrega os pixels selecionados à assinatura fractal multiescala descrita na seção 4.2.3. Esses pixels representam a estrutura de interesse da imagem e são utilizados durante o cálculo da assinatura fractal multiescala. Como o valor do limiar influencia no conjunto de pixels de interesse, diversas assinaturas podem ser calculadas a partir da textura. As informações da assinatura fractal podem ser utilizadas para caracterização da textura. A Figura 4.19 ilustra a metodologia para extração de características de textura a partir da assinatura fractal multiescala e a distribuição de níveis de cinza.

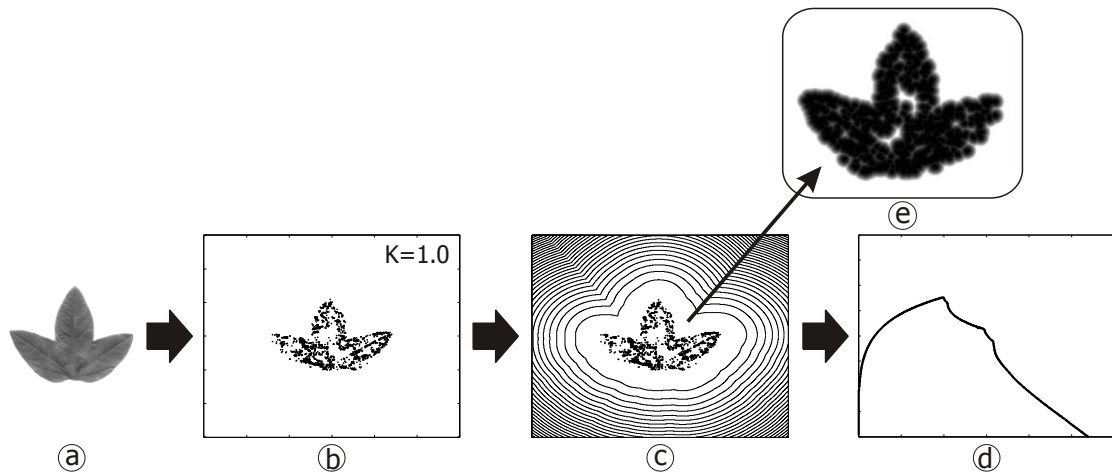


Figura 4.19: Análise de textura com base na distribuição de níveis de cinza e assinatura fractal multiescala. (a) forma biológica em escala de cinza; (b) seleção do subconjunto para $k=1.0$; (c) dilatações exatas; (d) assinatura fractal multiescala; (e) dilatações exatas em detalhe.

Similarmente, os mesmo conceitos descritos podem ser utilizados para extrair características de texturas associando a distribuição de níveis de cinza as redes complexas. Neste caso, o conjunto de pixels selecionados a partir de um determinado limiar é utilizado como vértices da rede complexa. Da mesma maneira, inúmeras informações podem ser coletadas a partir da estrutura da rede complexa. A Figura 4.20 apresenta a análise de textura por meio da associação da distribuição de níveis de cinza e das redes complexas.

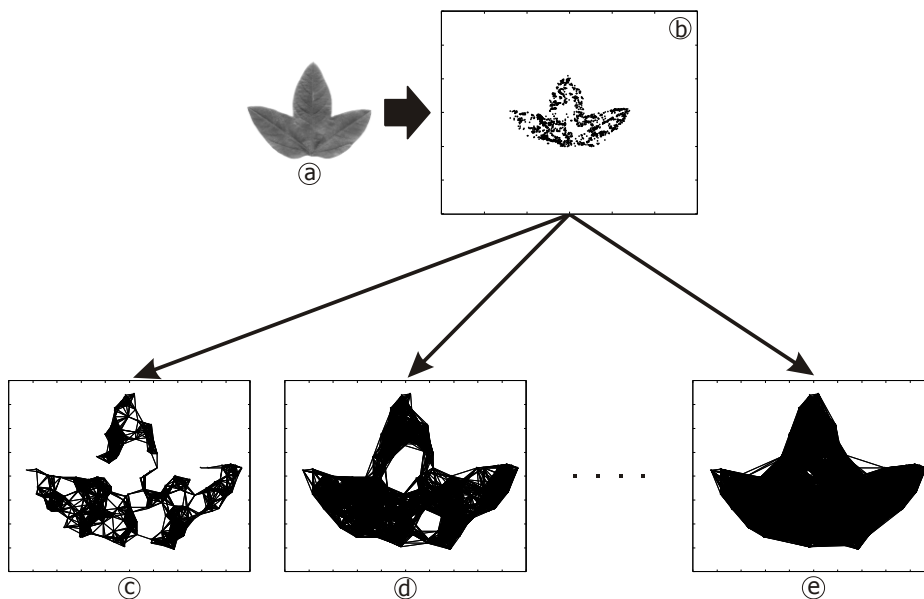


Figura 4.20: Análise de textura com base na distribuição de níveis de cinza e redes complexas. (a) forma biológica em escala de cinza; (b) seleção do subconjunto para $k=1.0$; (c) $T_i = 0.025$; (d) $T_i = 0.075$; (e) $T_i = 1.250$;

4.4 Considerações Finais

Neste capítulo foram apresentadas as principais metodologias de visão artificial utilizadas no trabalho. Diversas técnicas foram descritas e novas abordagens foram propostas. As técnicas de análise de imagens foram divididas em três grupos: formas, estruturas tubulares e textura. As técnicas de análise de formas foram utilizadas para o estudo do contorno das formas biológicas. O grupo de análise de estruturas tubulares possui técnicas morfométricas empregadas na investigação de características internas, como por exemplo, no sistema de venação foliar. Por fim, as técnicas de análise de texturas podem ser empregadas em estudos macro ou microscópicos das espécies. Todas as características extraídas fornecem a base para o estudo evolutivo e taxonômico das espécies. Essa investigação é complementada com as técnicas de reconhecimento de padrões apresentadas no próximo capítulo.

5

Reconhecimento de Padrões

A habilidade de reconhecer padrões é uma característica intrínseca dos seres humanos. A classificação de padrões em categorias similares é uma tarefa natural, que não precisa de aprendizado, e é realizada desde os primeiros meses de vida. Dentro do contexto de um sistema de visão artificial, as técnicas de reconhecimento de padrões são empregadas na classificação dos conjuntos de dados produzidos pelas técnicas de análise de imagens. Existem inúmeras abordagens para reconhecimento de padrões, e neste capítulo são apresentadas as metodologias utilizadas no trabalho.

5.1 Fundamentos

A área de reconhecimento de padrões é responsável pelo estudo de como as máquinas podem observar o ambiente a sua volta, aprendendo a distinguir padrões de interesse, e principalmente tomar decisões lógicas sobre as categorias desses padrões (Jain, Duin *et al.*, 2000). O grande desafio é desenvolver máquinas tão eficientes quanto às *máquinas* humanas.

Grande parte da metodologia de reconhecimento de padrões é originária das ciências naturais, como biologia, botânica, etc. Essas técnicas são fundamentais para diversas linhas de pesquisas. Em biologia, por exemplo, a classificação de organismos é largamente estudada. *Aristóteles* foi o primeiro a construir um sistema eficiente para classificar espécies do reino

animal. Em botânica, *Linnaeus* formulou a base metodológica para a classificação de plantas conhecida como taxonomia moderna.

O processo de reconhecer padrões representa um método eficiente de organizar grandes conjuntos de informações, de maneira que esses dados possam ser facilmente entendidos e rapidamente recuperados (Everitt, Landau *et al.*, 2001). Considerando um padrão, o reconhecimento ou classificação consiste em uma das seguintes tarefas: classificação supervisionada - em que o padrão é identificado como membro de uma categoria (classe) pré-definida; ou classificação não-supervisionada - no qual o padrão é associado a uma classe inicialmente desconhecida (Jain, Duin *et al.*, 2000).

Em ambos os tipos de abordagem, supervisionada ou não-supervisionada, a técnica de reconhecimento de padrões necessita de propriedades discriminantes de cada padrão a ser classificado. Assim, um padrão é representado por um conjunto de d características, estruturado em um vetor de características d -dimensional. Essas informações, na maioria das vezes, são armazenadas em uma matriz multivalorada X contendo $n \times m$ elementos. A entrada $x_{ij} \in X$ contém a j -ésima característica do padrão i .

Em alguns casos os vetores de características que compõem a matriz X necessitam ser normalizados. Esse fato é altamente importante para reduzir arbitrariedades impostas por diferentes unidades de medida que representam as características (Costa e Cesar, 2000). Usualmente, um conjunto de dados pode ser normalizado aplicando a Equação 5.1, onde μ e σ representam a média e o desvio padrão da característica j . Para essa função de normalização os dados são escalonados no intervalo de $[-2, +2]$ (Costa e Cesar, 2000).

$$\tilde{f}(i, j) = \frac{f(i, j) - \mu_j}{\sigma_j} \quad (5.1)$$

5.2 Análise Estatística Multivariada

Os métodos de análise multivariada são utilizados para avaliar simultaneamente conjuntos de dados com múltiplas dimensões (Raykov e Marcoulides, 2008). Essas técnicas aplicam avançados modelos estatísticos, visando à classificação de grandes quantidades de dados

complexos (Hair, Anderson *et al.*, 1998). Em aplicações de visão artificial, esses conjuntos de dados são formados por vetores de características gerados a partir de técnicas de análise de imagens.

Neste trabalho cinco técnicas de análise estatística multivariada foram empregadas na classificação das características morfométricas extraídas das espécies biológicas: (i) classificador euclidiano por distância mínima (DM); (ii) análise discriminante linear (LDA); (iii) análise discriminante por regressão espectral (SRDA); (iv) análise de componentes principais (PCA) e (v) análise de agrupamentos (*clustering*).

5.2.1 Classificador Euclidiano por Distância Mínima

A classificação por distância mínima utiliza uma função discriminante para deliberar a associação de um padrão a uma determinada classe (Jain, 1989). A construção do classificador é realizada por meio do cálculo do centro de massa, ou centroide, entre todos os padrões pertencentes a cada uma das classes do conjunto de dados (Jahne, 2002). Seja \vec{X}_c^i o vetor de características de um padrão i e K_c o número de padrões da classe c , o centro de massa pode ser calculado por meio da média entre todos os padrões pertencentes a mesma classe c , como expresso na Equação 5.2 (Fischer, Shahbazkis *et al.*, 2002).

$$\vec{A}_c = \frac{1}{K_c} \sum_{i=0}^{K_c-1} \vec{X}_c^i \quad (5.2)$$

Considerando \vec{X}_d o vetor de características do padrão d , e \vec{A}_i o centro de massa da classe i , tal que $i \in \{0, 1, \dots, n_c - 1\}$ para n_c classes, então d é associado a classe j , para $j \in \{0, 1, \dots, n_c - 1\}$ quando a distância euclidiana entre \vec{A}_i e \vec{X}_d for a menor dentre todas as outras classes. Alternativamente, outras métricas de distância podem ser utilizadas como regra de decisão (Costa e Cesar, 2000).

5.2.2 Análise Discriminante Linear

A análise discriminante é uma técnica estatística que tem como principal objetivo a classificação de um padrão, ou um conjunto de padrões, em uma classe previamente (*a priori*) conhecida (Hardle e Simar, 2007). A classificação é realizada por meio de funções

discriminantes que são utilizadas para descrever as diferenças entre as classes (Rencher, 2002). Quando a classificação das classes é linearmente separável, podemos utilizar a análise discriminante linear (*Linear Discriminant Analysis - LDA*). Essa abordagem pode ser empregada tanto para classificação, quanto para redução de dimensionalidade do conjunto de dados (Fukunaga, 1990).

Para problemas envolvendo múltiplas classes, é necessário considerar que cada classe possui uma distribuição normal e a mesma matriz de covariância. Com isso, é possível utilizar como regra de decisão a função discriminante linear de Fischer (Timm, 2002). Seja $g_i(x) = \{g_1(x), g_2(x), \dots, g_c(x)\}$ o conjunto de funções discriminantes lineares para um grupo de c classes. Assim, um padrão x será associado a uma classe ω_i quando a função discriminante assumir o maior valor para x , como expresso na Equação 5.3. Para casos em que o maior valor é equivalente a maior probabilidade, então o erro mínimo bayseano é obtido conforme expresso na Equação 5.4 (Webb, 2002).

$$g_i(x) = \max \{g_j(x), j = 1, 2, \dots, c\} \quad (5.3)$$

$$g_i(x) = \max \{g_j(x), j = 1, 2, \dots, c\} \Leftrightarrow p(\omega_i|x) = \max \{p(\omega_i|x), j = 1, 2, \dots, c\} \quad (5.4)$$

5.2.3 Análise Discriminante por Regressão Espectral

A análise discriminante linear (LDA) é um dos métodos mais utilizados para reconhecimento de padrões, sendo empregada em diversos problemas de classificação e redução de dimensionalidade. No entanto, o grande volume de dados manipulado durante o cálculo das funções discriminantes, torna a LDA uma técnica com alto custo computacional de tempo e memória.

A análise discriminante por regressão espectral (*Spectral Regression Discriminant Analysis - SRDA*) é uma metodologia alternativa baseada na análise espectral de grafos, que minimiza o tempo gasto para análise discriminante (Cai, He *et al.*, 2008). Nessa abordagem os vetores de transformação são obtidos por meio da resolução de um conjunto de problemas de regressão linear (Cai, D., He, X. *et al.*, 2007). A potencialidade da metodologia é comprovada em trabalhos envolvendo redução de dimensionalidade de grandes conjuntos de dados (Cai,

He *et al.*, 2008), recuperação de imagens por conteúdo (Cai, D., He, Xiaofei *et al.*, 2007), dentre outros.

Seja X um conjunto de dados formado por m padrões $\{x_1, x_2, \dots, x_m\}$, com n características, distribuídos em c classes. Considerando m_k o número de elementos da k -ésima classe, o primeiro passo da técnica *SRDA* consiste em computar os autovetores y como expreso na Equação 5.5.

$$y_k = [\underbrace{0, \dots, 0}_{\sum_{i=1}^{k-1} m_i}, \underbrace{1, \dots, 1}_{m_k}, \underbrace{0, \dots, 0}_{\sum_{i=k+1}^c m_i}]^T \quad k = 1, \dots, c \quad (5.5)$$

Utilizando $y_0 = [1, 1, \dots, 1]^T$, um vetor contendo apenas elementos iguais a 1 (um), como o primeiro vetor e o processo de *Gram-Schmidt* (Golub e Loan, 1996) para ortogonalização de $\{y_k\}$, é possível obter $c - 1$ vetores ortogonais como representado na Equação 5.6.

$$\{\bar{y}_k\}_{k=1}^{c-1}, (\bar{y}_i^T y_0 = 0, \bar{y}_i^T \bar{y}_j = 0, i \neq j) \quad (5.6)$$

O próximo passo do método é calcular os vetores de transformação a_k por meio dos quadrados mínimos, como expreso na Equação 5.7. Finalmente, a combinação dos $c - 1$ vetores $\{a_k\}$ forma a base dos vetores do *SRDA*. Sendo $A = [a_1, a_2, \dots, a_{c-1}]$ uma matriz representada por $(n+1) \times (c-1)$ elementos, os vetores de características cotendo $c - 1$ dimensões podem ser obtidos com a Equação 5.8.

$$a_k = \min_a \left(\sum_{i=1}^m (a^T x_i - \bar{y}_i^k)^2 + \alpha \|a\|^2 \right) \quad (5.7)$$

em que \bar{y}_i^k representa o i -ésimo elemento de \bar{y}_k .

$$X \rightarrow Z = A^T \begin{bmatrix} X \\ 1 \end{bmatrix} \quad (5.8)$$

5.2.4 Análise de Componentes Principais

A análise de componentes principais (*Principal Component Analysis – PCA*) é um método estatístico que busca maximizar a variância da combinação linear entre os vetores de características (Rencher, 2002). Essa técnica utiliza a matriz de covariância do conjunto de dados original para calcular autovetores conhecidos como componentes principais. Esses vetores contêm as características mais discriminantes, ou mais correlacionadas, do espaço de características (Jain, Duin *et al.*, 2000).

Seja X um conjunto de dados composto por m padrões, tal que $X = \{x_1, x_2, \dots, x_m\}$, e matriz de covariância C . Considere $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ os autovalores da matriz C , com seus respectivos autovetores normalizados $e_i = \{e_1, e_2, \dots, e_m\}^T$, a j -ésima componente principal Y pode ser obtida com a Equação 5.9 (Mingoti, 2005).

$$Y_j = e_j^T X = e_{j1}X_1 + e_{j2}X_2 + \dots + e_{jm}X_m \quad \text{para } j = \{1, 2, \dots, m\} \quad (5.9)$$

5.2.5 Análise de Agrupamentos

As técnicas de análise de agrupamentos (do inglês *cluster analysis*, ou *clustering*) têm como objetivo organizar padrões com base nas características que eles possuem. Os padrões similares são agrupados de acordo com um critério predeterminado de seleção. O resultado são agrupamentos, ou classes, internamente homogêneas (dentro do cluster) e externamente heterogêneas (entre os clusters) (Hair, Anderson *et al.*, 1998). Análise de agrupamentos são métodos de classificação não-supervisionados, em que duas abordagens podem ser utilizadas: hierárquica e não-hierárquica (Kohn, 1998).

A análise de agrupamento hierárquico (*hierarchical clustering*) representa a relação de similaridade ou dissimilaridade entre padrões e agrupamentos (Kohn, 1998). Essa metodologia tem como objetivo unir M padrões em n classes de maneira progressiva (Costa e Cesar, 2000). O processo de agrupamento é realizado com base na distância (ou similaridade) entre as características que representam os padrões. As técnicas de agrupamento hierárquico podem ser divididas em duas categorias de métodos (Everitt, Landau *et al.*, 2001): aglomerativos e divisivos. Devido à importância para o trabalho apenas a abordagem aglomerativa será descrita.

Os métodos aglomerativos são as técnicas de análise de agrupamento hierárquicas mais utilizadas. O agrupamento dos padrões é realizado como se segue (Hair, Anderson *et al.*, 1998): inicialmente, cada padrão é representado como um cluster de um único elemento. Nas etapas seguintes, os padrões (ou clusters) mais próximos são combinados em um novo cluster. Assim, em cada etapa o número de clusters é reduzido. O processo termina quando todos os padrões são combinados em um único cluster.

Na literatura podem ser encontrados diversos métodos aglomerativos, cada qual utiliza uma métrica para determinar as similaridades (ou dissimilaridades) entre os agrupamentos, os principais são apresentados na Tabela 5.1.

Tabela 5.1: Métodos aglomerativos de agrupamento hierárquico.

Método	Definição da distância entre os clusters
<i>Single Linkage</i>	Menor distância entre dois padrões em clusters distintos.
<i>Complete linkage</i>	Maior distância entre dois padrões em clusters distintos.
<i>Unweighted pair-group average (UPGMA)</i>	Distância média entre todos os pares de padrões em dois diferentes clusters.
<i>Weighted pair-group average (WPGMA)</i>	Idêntico ao método UPGMA, contudo o número de elementos do cluster é levado em consideração.
<i>Unweighted pair-group centroid (UPGMC)</i>	Distância entre os centroides de dois clusters distintos.
<i>Weighted pair-group centroid (WPGMC)</i>	Idêntico ao método UPGMC, contudo o número de elementos do cluster é levado em consideração.
<i>Ward's method</i>	Utiliza uma medida de dispersão (soma dos quadrados) para determinar a distância entre os clusters.

5.3 Redes Neurais Artificiais

As redes neurais artificiais (RNA) são técnicas de reconhecimento de padrões capazes de resolver problemas altamente complexos (Freeman e Skapura, 1991). Baseadas no modelo de funcionamento dos sistemas neurais biológicos, as redes neurais artificiais tem como principal característica a habilidade de aprender relações não-lineares de entrada e saída. A Figura 5.1 apresenta um neurônio biológico comparado a um modelo neural artificial. O neurônio artificial é inspirado no modelo proposto por *McCulloch e Pitts* (McCulloch e Pitts, 1943).

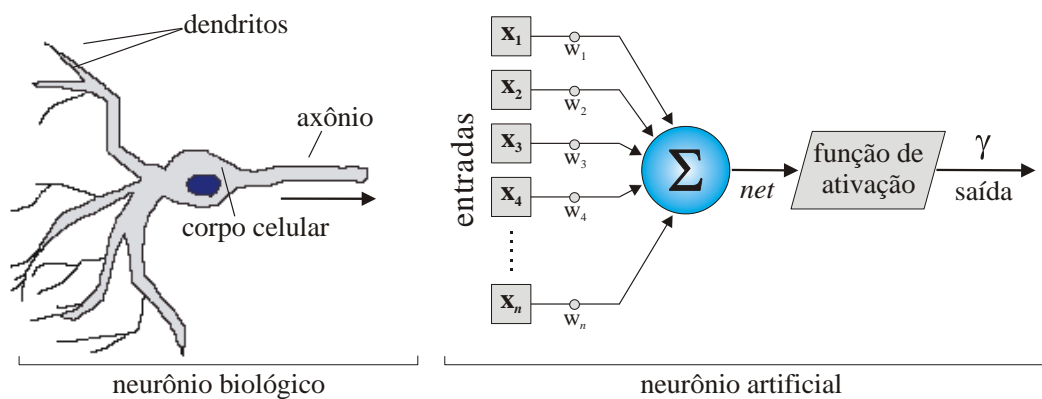


Figura 5.1: Comparação entre neurônio biológico e neurônio artificial.

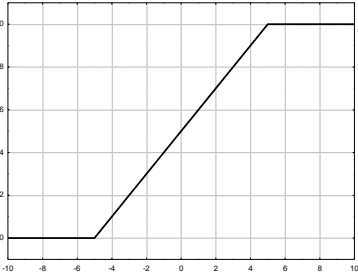
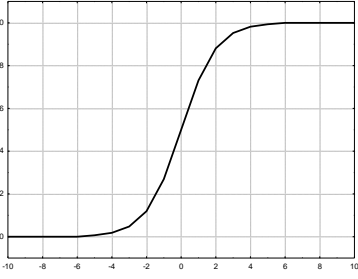
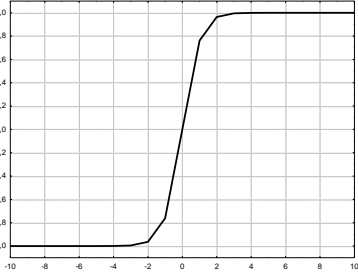
As entradas do modelo artificial (x_1, x_2, \dots, x_n) tem como função receber informações originárias de outros neurônios. Biologicamente essas entradas executam a tarefa dos dendritos. O processamento de informações no neurônio artificial (também conhecido como unidade de processamento ou nó da rede) é realizado com uma combinação linear entre as entradas e os pesos da rede, como expresso na Equação 5.10.

$$net_i = \sum_{j=1}^n x_j \cdot w_{ij} \quad (5.10)$$

Todas as entradas da rede estão associadas a pesos w_i que refletem a importância de cada entrada X_n . O resultado da combinação linear é o valor net_i . A saída γ de cada neurônio artificial está ligada a uma função de ativação, cujo objetivo é restringir o resultado produzido nas unidades de processamento. A Tabela 5.2 apresenta diferentes tipos de funções de ativação que podem ser utilizadas para produzir a saída dos neurônios artificiais. O intervalo resultante de cada neurônio de saída depende da função selecionada.

Tabela 5.2: Funções de ativação dos neurônios artificiais.

Função	Formulação	Representação
limiar		

<p>linear</p>	
<p>sigmóide</p>	
<p>tangente hiperbólica</p>	

As redes neurais artificiais podem ser entendidas como grafos orientados, em que os vértices representam os neurônios e as arestas definem as conexões eles. A principal família de RNA's é conhecida como *feed-forward*, em que os grafos não possuem loops (Jain, Mao *et al.*, 1996). Dentro dessa família, a arquitetura de rede mais utilizada é a *perceptron* multicamadas (*Multilayer Perceptron – MLP*), que redes capazes de resolver problemas não linearmente separáveis (Haykin, 1998).

5.3.1 Perceptron Multicamadas (MLP)

O *perceptron* multicamadas (*Multilayer Perceptron - MLP*) é o tipo mais popular de rede neural da classe *feed-forward*. Cada camada da rede consiste em um conjunto de unidades de processamento (neurônios) que recebem como entrada informações dos neurônios da camada anterior, e fornecem como saída às informações processadas para os neurônios da camada posterior (Braga, Carvalho *et al.*, 2000).

A estrutura da rede pode ser dividida em três partes principais: camada de entrada (C_E); camada(s) oculta(s) (C_O) e camada de saída (C_S). Os neurônios da C_E são responsáveis pela alimentação de informações da rede. Eles fornecem informações para a primeira C_{O-1} , entretanto C_E não realizam nenhum tipo de processamento. As camadas ocultas $\{C_{O-1}, \dots, C_{O-k}\}$ são responsáveis por todo processamento da rede. As informações são propagadas direcionalmente camada-a-camada, ou seja, de C_{O-1} até C_{O-k} . As saídas dos neurônios da camada C_{O-k} alimentam os neurônios da C_S , que são responsáveis pela apresentação dos resultados da rede. Os neurônios da C_S também não realizam nenhum tipo de processamento. A Figura 5.2 ilustra uma rede neural do tipo *perceptron* multicamada.

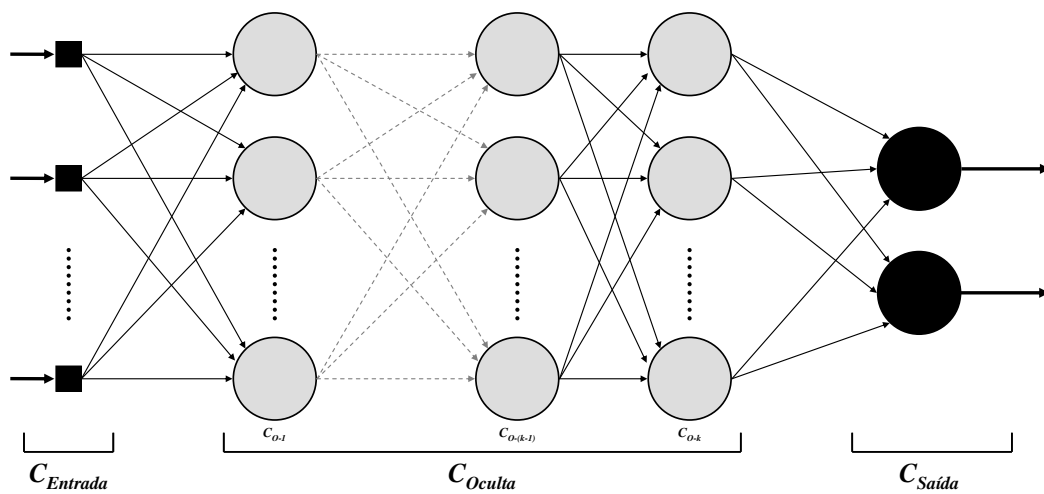


Figura 5.2: Arquitetura de uma rede neural do tipo *perceptron* multicamada (MLP).

O processamento das informações de uma rede neural multicamadas é realizado com um algoritmo de aprendizado. O método mais popular é conhecido como algoritmo *back-propagation*. Este algoritmo utiliza as informações de entrada $\{x_1, x_2, \dots, x_n\}$ e as saídas desejadas $\{d_1, d_2, \dots, d_n\}$ para ajustar os pesos da rede por meio de um mecanismo de correção de erros. O aprendizado da rede ocorre em duas fases chamadas *backward* e *forward*, e o principal objetivo é minimizar uma função de erro. A Equação 5.11 apresenta a função de erro médio quadrático (*mean square error*).

$$M_{SE} = \frac{1}{2} \sum_{i=1}^n (d_j - o_j)^2 \quad (5.11)$$

em que $\{o_1, o_2, \dots, o_n\}$ representam as saídas obtidas pela rede após o treinamento.

5.4 Seleção de Características

As técnicas de seleção de características têm como objetivo a seleção de subconjuntos de dados a partir de um conjunto contendo um determinado número de características. Assim, para um conjunto contendo d características, tenta-se buscar um subconjunto m tal que $m < d$ (Castañón, 2006). Na prática essas técnicas reduzem a dimensionalidade do conjunto, sempre objetivando a escolha das características que sejam altamente discriminantes (Jain, Duin *et al.*, 2000).

As técnicas de seleção de características podem ser empregadas para minimizar a *maldição da dimensionalidade* (*curse of dimensionality*). Em problemas de visão artificial, as técnicas de análise de imagens permitem a extração de um vasto número de características. No entanto, a utilização dessas inúmeras características em processos de reconhecimento de padrões, principalmente envolvendo análise multivariada e redes neurais artificiais, podem resultar em classificações incorretas. De fato, o número excessivo de características pode comprometer significativamente o desempenho do classificador. Para auxiliar nesse processo, as técnicas de seleção de características podem ser empregadas para reduzir o número de características utilizadas no processo de reconhecimento de padrões.

Na literatura é possível encontrar diversas metodologias para seleção de características (Jain e Zongker, 1997), dentre elas, uma abordagem baseada na análise de componentes principais (PCA) (Le e Satoh, 2005). A seleção de características por meio do PCA utiliza como base a variância total explicada pela j -ésima componente principal. Assim, considerando os autovalores da matriz de covariância $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$, para um conjunto composto por m padrões, a proporção da variância total do conjunto de dados X pode ser calculada como expresso na Equação 5.12 (Mingoti, 2005).

$$\frac{\lambda_j}{\sum_{i=1}^m \lambda_i} \quad \text{para } i, j = \{1, 2, \dots, m\} \quad (5.12)$$

5.5 Avaliação dos Classificadores

Os diversos conjuntos de dados biológicos utilizados neste trabalho, bem como os vários tipos de classificadores, fazem com que sejam necessárias metodologias sistemáticas para avaliação

dos resultados. Além disso, esses critérios estabelecem normas para garantir a qualidade dos resultados obtidos.

Independente do tipo de classificador utilizado, uma das fases fundamentais do processo de classificação é a divisão do conjunto de dados em treinamento e teste. Diversas abordagens podem ser empregadas para separação do conjunto, as quais são denominadas métodos de validação cruzada. A utilização dessas técnicas minimiza problemas comumente encontrados em reconhecimento de padrões, como por exemplo, a generalização do classificador (Jain, Duin *et al.*, 2000). Os métodos mais conhecidos de validação cruzada são: *k-fold*, *leave-one-out* e *holdout* (Haykin, 1998).

A avaliação do desempenho de um classificador pode ser estimada por meio de um conjunto de informações extraídas do resultado da classificação. Para isso, uma estrutura denominada *matriz de confusão* pode ser construída com as informações obtidas a partir das saídas do classificador. Considerando um problema no qual é necessário classificar um padrão como pertencendo, ou não, a uma determinada classe, quatro saídas podem ser consideradas: (i) verdadeiro-positivo; (ii) falso-negativo; (iii) verdadeiro-negativo e (iv) falso positivo (Fawcett, 2006). A Figura 5.3 ilustra a estrutura de uma matriz de confusão.

		resultado correto	
		Verdadeiro Positivo (VP)	Falso Positivo (FP)
resultado obtido		Falso Negativo (FN)	Verdadeiro Negativo (VN)

Figura 5.3: Estrutura de uma matriz de confusão.

Considerando P total de amostras positivas (pertencente à classe), e N o total de amostras negativas (não pertencente a classe), diversas informações estatísticas podem ser obtidas a partir da matriz de confusão, como por exemplo: acurácia do classificador (Equação 5.13); a taxa de verdadeiro positivo, também chamada de sensibilidade ou *recall* (Equação 5.14); a taxa de falso positivo (Equação 5.15); precisão (Equação 5.16), especificidade (Equação 5.17) e *F-Measure* (Equação 5.18) (Fawcett, 2006).

$$AC \approx \frac{VP + VN}{P + N} \quad (5.13)$$

$$T_{VP} \approx \frac{VP}{P} \quad (5.14)$$

$$T_{FP} \approx \frac{FP}{N} \quad (5.15)$$

$$PR \approx \frac{VP}{VP + FP} \quad (5.16)$$

$$ES \approx \frac{VN}{VN + FP} \quad (5.17)$$

$$F \approx \frac{2}{\frac{1}{Pr} + \frac{1}{T_{VP}}} \quad (5.18)$$

Um gráfico ROC (*Receiver Operation Characteristics*) é uma técnica empregada para visualização do desempenho de um classificador. Esse gráfico é comumente utilizado para descrever a relação entre as taxas de acerto (verdadeiros positivos) e erro (falso positivos) de um classificador. A construção do gráfico é baseada na taxa de verdadeiro positivo (T_{VP}) que é esboçada no eixo Y , e na taxa de falso positivo (T_{FP}) que é representada no eixo X . Assim, um ponto no espaço de ROC é discretizado na forma (T_{VP}, T_{FP}) . Uma descrição detalhada da construção de gráficos ROC pode ser obtida em (Fawcett, 2006).

5.6 Considerações Finais

Neste capítulo foram apresentados os principais conceitos relacionados ao reconhecimento de padrões. Essas técnicas têm como objetivo principal associar um padrão a uma determinada classe. Foram descritos métodos de análise estatística multivariada, como o classificador euclidiano por distância mínima, a análise discriminante linear (e por regressão espectral) e a análise de componentes principais, além das redes neurais artificiais. Um método para seleção de características baseado na proporção de variância do PCA foi apresentado. Por fim, um conjunto de métricas para avaliação de um classificador foi detalhado, envolvendo validação cruzada, matriz de confusão e gráficos ROC.

6

Experimentos

Este capítulo apresenta a descrição dos experimentos realizados para análise e classificação de espécies biológicas. Os experimentos têm como foco principal a análise e o reconhecimento de padrões baseados nas características morfométricas extraídas das espécies. Para isso, diversas informações biológicas foram coletadas das espécies por meio das técnicas de análise de imagens. Além disso, todo conhecimento extraído das espécies foi avaliado por meio de métodos de reconhecimento de padrões. Todo referencial teórico a respeito das técnicas de análise de imagens e reconhecimento de padrões utilizadas nos experimentos foi descrito no capítulo 4 e capítulo 5, respectivamente. Além disso, as novas técnicas propostas para análise de imagens são relatadas no capítulo 4, dentre elas a assinatura fractal multiescala para estruturas tubulares. Neste capítulo, inicialmente, são descritos os experimentos envolvendo a análise das imagens biológicas, em que cada método empregado é relatado. Por fim, é apresentado como o processo de reconhecimento de padrões foi realizado.

6.1 Espécies Biológicas

Para análise e reconhecimento de padrões a partir de características morfométricas, três gêneros de espécies biológicas distintas foram utilizados:

- maracujá (*Passiflora*): 580 exemplares / 13 espécies;
- eucalipto (*Eucalyptus*): 285 exemplares / 10 espécies;

- arroz (*Oryza*): 177 exemplares / 12 espécies;

A distribuição do número de exemplares por espécie é variável e uma descrição completa a respeito de cada espécie é apresentada no capítulo 3.

Para determinação dos experimentos realizados com cada espécie, foi considerada a representação do objeto de interesse (contorno, estrutura tubular e textura). Nas espécies de maracujás, por exemplo, os objetos de interesse considerados foram o contorno, a estrutura tubular (nervuras) e a textura, enquanto para as espécies de arroz e eucalipto apenas o contorno foi utilizado. A organização dos experimentos em função dos objetos de interesse presentes nas imagens é apresentada na Tabela 6.1.

Tabela 6.1: Organização dos experimentos de análise de imagens empregados nas espécies biológicas.

Análise de Imagens	<i>Passiflora</i>	<i>Eucalyptus</i>	<i>Oryza</i>
Análise de Formas	X	X	X
Estruturas Tubulares	X		
Textura	X		

A Tabela 6.2 apresenta os identificadores utilizados para representar as espécies biológicas utilizadas nos experimentos. Na descrição dos resultados esses identificadores serão utilizados para referenciar as espécies.

Tabela 6.2: Identificadores utilizados para referenciar as espécies biológicas.

Identificador	Espécies		
	<i>Passiflora</i>	<i>Eucalyptus</i>	<i>Oryza</i>
C ₀₁	<i>P.actinia</i>	<i>E. grandis</i>	<i>O. alta</i>
C ₀₂	<i>P.foetida fluminensis</i>	<i>E. pellita</i>	<i>O. barthii</i>
C ₀₃	<i>P.foetida foetida</i>	<i>Corymbia torelliana</i>	<i>O. glaberrima</i>
C ₀₄	<i>P.sidfolia</i>	<i>E. urophylla</i>	<i>O. glumaepatula</i>
C ₀₅	<i>P.suberosa</i>	<i>E. grandis x E. camaldulensis</i>	<i>O. grandiglumis</i>
C ₀₆	<i>P.miersii</i>	<i>E. saligna x E. camaldulensis</i>	<i>O. latifolia</i>
C ₀₇	<i>P.amethystina</i>	<i>E. urophylla x E. camaldulensis</i>	<i>O. longistaminata</i>
C ₀₈	<i>P.caerulea</i>	<i>E. urophylla x E. globulus</i>	<i>O. nivara</i>
C ₀₉	<i>P.coriacea</i>	<i>E. grandis x E. globulus</i>	<i>O. officinalis</i>
C ₁₀	<i>P.gibertii</i>	<i>E. urophylla x E. grandis</i>	<i>O. punctata</i>
C ₁₁	<i>P.maliformis</i>		<i>O. rufipogon</i>
C ₁₂	<i>P.organensis</i>		<i>O. sativa</i>
C ₁₃	<i>P.pohlii</i>		

6.2 Análise de Imagens

As técnicas de análise de imagens foram utilizadas para a extração de medidas morfométricas das espécies biológicas. Os métodos utilizados (e desenvolvidos) foram organizados em três grupos: formas, estruturas tubulares e textura. Todos os métodos foram implementadas utilizando o programa *Matlab*® versão *R2008b*, e executados em um computador Intel Core2Duo 2.1Ghz com 3gb de memória. Além disso, alguns algoritmos foram escritos em linguagem C/C++ e Borland Delphi®. A Figura 6.1 apresenta uma visão estrutural das técnicas de análise de imagens empregadas nos experimentos.

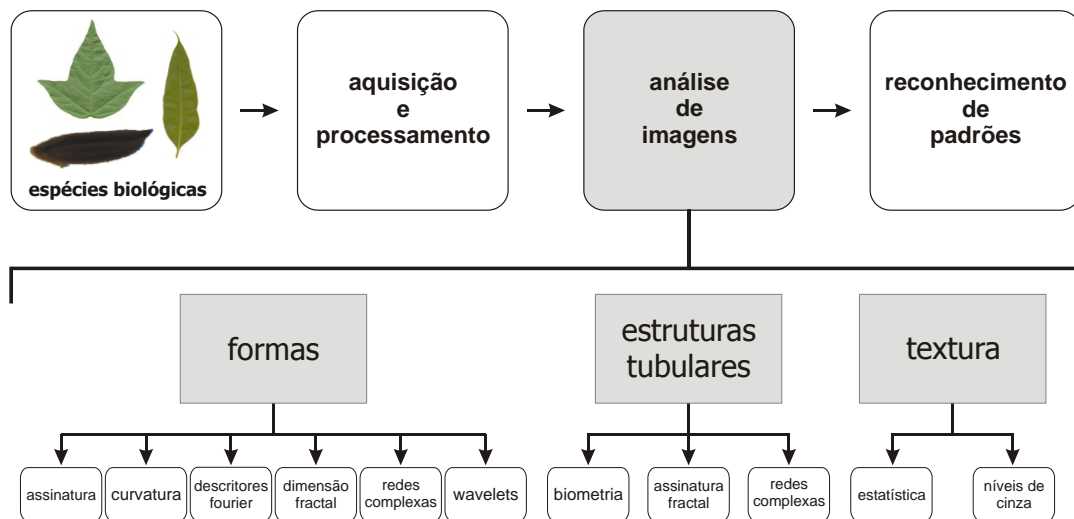


Figura 6.1: Estrutura organizacional das técnicas de análise de imagens utilizadas nos experimentos.

Os métodos de análise de imagens utilizados nos experimentos são, em grande parte, parametrizáveis. Esse fato permite extrair com uma mesma técnica, um conjunto distinto de características morfométricas. Para facilitar a organização e análise dos resultados obtidos, cada experimento efetuado foi identificado com um número único precedido de uma letra que representa o tipo do experimento: *F* para análise de formas; *E* para análise de estruturas tubulares e *T* para análise de textura. As próximas seções descrevem as sistemáticas utilizadas durante a execução dos experimentos, bem como os parâmetros empregados em cada técnica, além do número de medidas morfométricas que foram obtidas.

6.2.1 Análise de Formas: Assinaturas

O primeiro experimento realizado utilizou como base as informações extraídas a partir da representação unidimensional do contorno. Considerando as técnicas para cálculo da assinatura digital, duas abordagens foram utilizadas: assinatura baseada no centroide e assinatura por projeção.

Para assinatura baseada no centroide, inicialmente o contorno das imagens biológicas foi extraído utilizando o algoritmo seguidor de contorno descrito em (Costa e Cesar, 2000). Com isso, o contorno foi decomposto em dois sinais unidimensionais independentes X e Y (um para cada eixo). Considerando (C_x, C_y) a coordenada do centro de massa da forma analisada, a assinatura baseada no centroide pode ser obtida com a Equação 6.1.

$$C = \sqrt{(C_x - X_i)^2 + (C_y - Y_i)^2} \quad (6.1)$$

A assinatura por projeção de uma imagem digital é obtida com o somatório dos pontos a partir de um ângulo específico. Nos experimentos, três ângulos foram considerados: $\theta = 45^\circ$ (diagonal); $\theta = 90^\circ$ (vertical) e $\theta = 180^\circ$ (horizontal).

A Tabela 6.3 sintetiza os experimentos que foram realizados com base na assinatura das imagens biológicas. A primeira coluna (*ID*) representa um identificador único de cada experimento; a coluna *Título* uma descrição sucinta a respeito do método; a coluna *Parâmetros* descreve as variáveis de entrada passadas para o método e, por fim, o *Número de Características* morfométricas extraídas para cada tipo de imagem biológica analisada.

Tabela 6.3: Experimentos realizados com base na assinatura digital das imagens biológicas.

ID	Título	Parâmetros	Número de Características		
			<i>Passiflora</i>	<i>Eucalyptus</i>	<i>Oryza</i>
F001	Assinatura Centroide	-	3766	881	1500
F002	Assinatura Projeção_045	$\theta = 045^\circ$	849	525	721
F003	Assinatura Projeção_090	$\theta = 090^\circ$	600	440	260
F004	Assinatura Projeção_180	$\theta = 180^\circ$	600	300	760

6.2.2 Análise de Formas: Curvatura

A extração de características morfométricas com base na curvatura foi realizada a partir de duas abordagens: curvatura por definição e curvatura multiescala por meio da propriedade derivativa da transformada de Fourier.

Na curvatura por definição foi utilizada a metodologia proposta por (Mokhtarian e Mackworth, 1992). Essa abordagem utiliza a derivada de primeira e segunda ordem para estimar a curvatura de um contorno paramétrico, como expresso na Equação 4.1 (vide capítulo 4, seção 4.1.2).

A curvatura multiescala é calculada utilizando a propriedade derivativa da transformada de Fourier, como apresentado na Equação 4.3. No entanto, essa abordagem acentua os ruídos das altas frequências do contorno da imagem (Costa e Cesar, 2000). Assim, para minimizar este problema, a diferenciação numérica utilizando Fourier deve ser combinada com um filtro gaussiano de passa-baixa. O conjunto de frequências do filtro pode ser ajustado com o desvio padrão da gaussiana a . Esse parâmetro permite a geração de curvaturas multiescala da forma analisada. A filtragem gaussiana no domínio da frequência para um sinal f pode ser obtida com a Equação 6.2 (Costa e Cesar, 2000).

$$G(f) = \exp\left(\frac{-(2\pi)^2 f^2}{2\tau^2}\right), \text{ para } \tau = 1/a \quad (6.2)$$

Nos experimentos, os seguintes valores para o parâmetro a foram considerados $a = \{50, 20, 10, 2, 1, 0.5\}$, assim temos $\tau = \{0.02, 0.05, 0.10, 0.50, 1.00, 2.00\}$. Esses valores foram empiricamente definidos. A Figura 6.2 ilustra as curvaturas multiescala que podem ser obtidas a partir da variação do parâmetro a . Na Figura, a escala do eixo Y foi normalizada no intervalo $[-1,1]$.

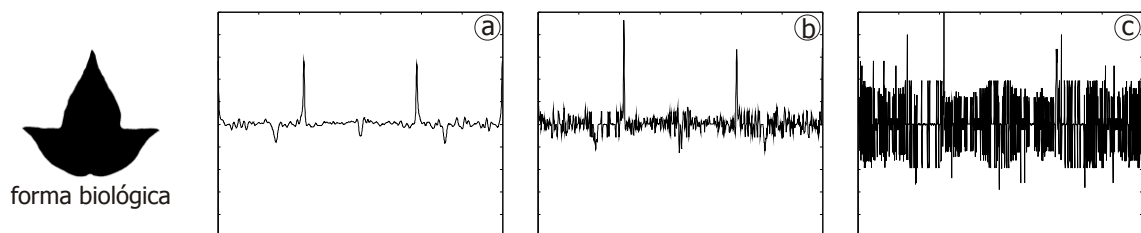


Figura 6.2: Curvatura multiescala de uma forma biológica para diferentes valores do desvio padrão da gaussiana: (a) $a = 50$; (b) $a = 20$ e (c) $a = 10$.

A Tabela 6.4 apresenta os experimentos realizados com a curvatura das espécies biológicas. Na Tabela são descritos os parâmetros utilizados no método de curvatura multiescala, e seus respectivos parâmetros.

Tabela 6.4: Experimentos realizados com a curvatura das espécies biológicas.

ID	Título	Parâmetros	Número de Características		
			<i>Passiflora</i>	<i>Eucalyptus</i>	<i>Oryza</i>
F005	Curvatura Definição	-	3766	879	1498
F006	Curvatura Multiescala_0.02	$\tau = 0.02$	3766	881	1500
F007	Curvatura Multiescala_0.05	$\tau = 0.05$	3766	881	1500
F008	Curvatura Multiescala_0.10	$\tau = 0.10$	3766	881	1500
F009	Curvatura Multiescala_0.50	$\tau = 0.50$	3766	881	1500
F010	Curvatura Multiescala_1.00	$\tau = 1.00$	3766	881	1500
F011	Curvatura Multiescala_2.00	$\tau = 2.00$	3766	881	1500

6.2.3 Análise de Formas: Descritores de Fourier

Os descritores de Fourier foram utilizados para extração de características morfométricas das espécies biológicas. Para isso, o contorno da forma foi extraído com o algoritmo seguidor de contorno e, em seguida, transformado em um sinal complexo ($u = x + j * y$). A transformada rápida de Fourier (*FFT*) foi aplicada no sinal, com isso, o sinal foi convertido para o domínio da frequência, ou seja, $u \Leftrightarrow U$.

Para extração dos descritores de Fourier foi utilizado o espectro da potência, conforme expresso na Equação 4.8 (vide seção 4.1.3). O sinal foi normalizado para torna-lo invariante as transformações geométricas (Equação 4.9). As características foram extraídas a partir dos componentes de baixa frequência, os quais descrevem as propriedades mais detalhadas da forma biológica.

Para cada forma biológica analisada, seis vetores de características foram extraídos contendo N descritores de Fourier. Os números de descritores considerados foram determinados empiricamente, sendo $N = \{10,20,40,80,160,320\}$. A Figura 6.3 ilustra os vetores de características extraídos para os diversos valores de N .

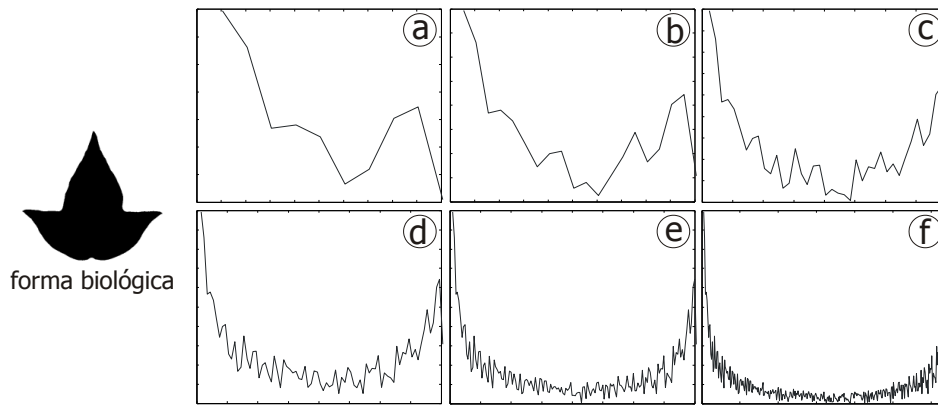


Figura 6.3: Descritores de Fourier: (a) N = 10; (b) N = 20; (c) N = 40; (d) N = 80; (e) N = 160 e (f) N = 320.

Na Tabela 6.5 são apresentados os experimentos realizados com os descritores de Fourier e os respectivos parâmetros utilizados como número de descritores.

Tabela 6.5: Experimentos realizados com os descritores de Fourier.

ID	Título	Parâmetros	Número de Características		
			<i>Passiflora</i>	<i>Eucalyptus</i>	<i>Oryza</i>
F012	Descritores Fourier_10	N = 10	10	10	10
F013	Descritores Fourier_20	N = 20	20	20	20
F014	Descritores Fourier_40	N = 40	40	40	40
F015	Descritores Fourier_80	N = 80	80	80	80
F016	Descritores Fourier_160	N = 160	160	160	160
F017	Descritores Fourier_320	N = 320	320	320	320

6.2.4 Análise de Formas: Dimensão Fractal Multiescala

A complexidade das formas biológicas foi analisada por meio da dimensão fractal multiescala. Inicialmente o contorno da forma foi detectado utilizando o método de *Canny*. Em seguida, o contorno S foi dilatado por um raio d utilizando a técnica de dilatações exatas. As diversas áreas assumidas por S durante o processo de dilatação foi representado por $A(d)$. Essa relação foi expressa por meio de uma função logarítmica: $\log(A(d)) \times \log(d)$.

A estimativa da dimensão fractal multiescala ocorre por meio da derivada numérica da função logarítmica. Para isso, é necessário definir um parâmetro k referente à ordem de derivação, além do parâmetro σ que representa o desvio padrão utilizado na suavização

gaussiana. Como resultado deste processo é gerada uma função $\log(d) \times \log(A')$, em que A' representa a derivada e suavização do ponto $(d, A(d))$ da função logarítmica.

Diversos valores foram testados experimentalmente para a ordem de derivação e para o desvio padrão da gaussiana. Ao final, foi estabelecido um conjunto de valores para $k = \{100, 250, 500\}$ e para $\sigma = \{100, 250, 500\}$. A combinação entre os valores da ordem de derivação e do desvio padrão produz 12 (doze) configurações possíveis. Todas essas configurações foram utilizadas como vetores de características para cada forma analisada. A Figura 6.4 ilustra a relação entre a ordem de derivação e o desvio padrão da gaussiana no resultado da dimensão fractal multiescala. O número de características de cada experimento foi reamostrado com uma função de interpolação linear, de maneira que todos os experimentos resultassem no mesmo número de características.

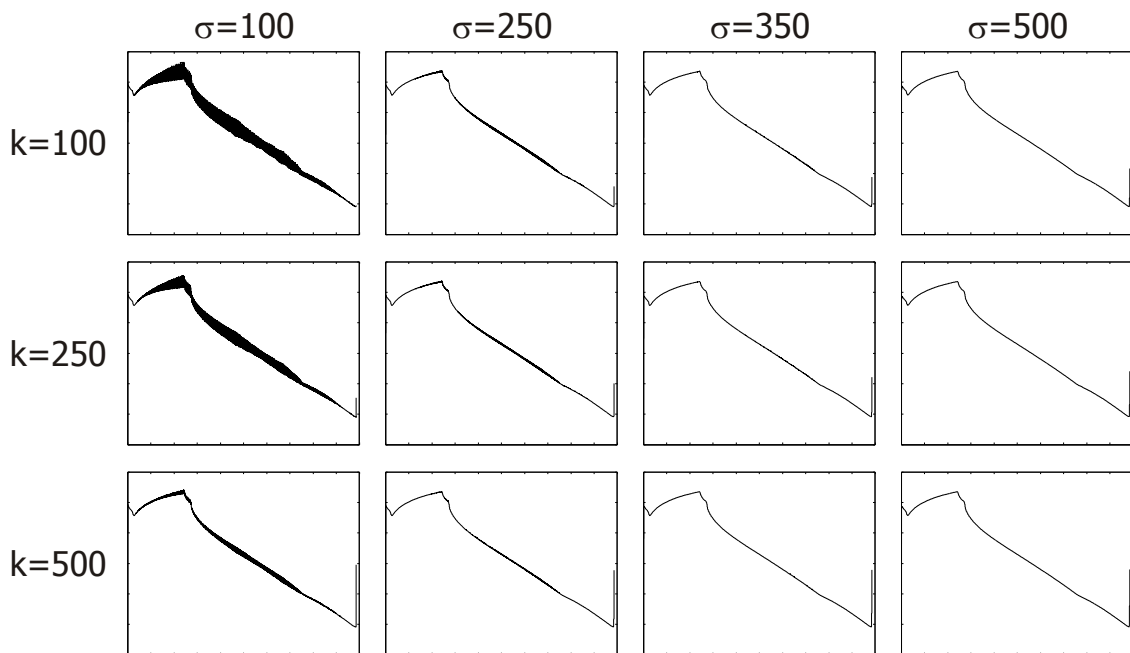


Figura 6.4: Funções fractal multiescala obtidas a partir dos valores da ordem de derivação e da suavização gaussiana.

A Tabela 6.6 descreve os diversos experimentos realizados com a dimensão fractal multiescala das formas biológicas.

Tabela 6.6: Experimentos realizados com a dimensão fractal multiescala.

ID	Título	Parâmetros		Número de Características		
		k	σ	<i>Passiflora</i>	<i>Eucalyptus</i>	<i>Oryza</i>
F018	DF Multiescala_100_100	100	100	2000	1800	1250
F019	DF Multiescala_100_250	100	250	2000	1800	1250
F020	DF Multiescala_100_350	100	350	2000	1800	1250
F021	DF Multiescala_100_500	100	500	2000	1800	1250
F022	DF Multiescala_250_100	250	100	2000	1800	1250
F023	DF Multiescala_250_250	250	250	2000	1800	1250
F024	DF Multiescala_250_350	250	350	2000	1800	1250
F025	DF Multiescala_250_500	250	500	2000	1800	1250
F026	DF Multiescala_500_100	500	100	2000	1800	1250
F027	DF Multiescala_500_250	500	250	2000	1800	1250
F028	DF Multiescala_500_350	500	350	2000	1800	1250
F029	DF Multiescala_500_500	500	500	2000	1800	1250

6.2.5 Análise de Formas: Redes Complexas

A utilização das redes complexas para análise de formas fornece um rico conjunto de informações sobre as espécies biológicas. O primeiro passo deste procedimento é a detecção do contorno com o algoritmo seguidor de contorno. Em seguida, uma matriz de pesos \mathcal{W} é calculada a partir da distância euclidiana entre cada coordenada dos pixels pertencentes ao contorno. Essa matriz é normalizada com a Equação 4.13.

Os fundamentos das redes complexas são baseados na teoria dos grafos, assim, o mapeamento de uma forma em um grafo não-dirigido ocorre por meio da representação de cada pixel do contorno em um vértice do grafo. Os pesos associados a cada aresta do grafo são armazenados na matriz de pesos \mathcal{W} .

A matriz de adjacência \mathcal{A} é calculada a partir de um conjunto de limiares T_i aplicados na matriz de pesos. Esse procedimento seleciona um subconjunto de arestas formando com isso uma rede complexa para cada limiar T_i . No trabalho foram considerados 38 limiares $T_i = \{0.025, 0.050, \dots, 0.925, 0.950\}$. Para cada rede complexa construído a partir de um limiar T_i as seguintes informações foram extraídas:

- Graus dos vértices (g_i)
- Grau máximo (g_{max}) dos vértices
- Grau médio (g_μ) dos vértices
- Distribuição das conexões (*joint degree distribution*)
- Entropia (H)
- Energia (E)
- Grau médio de distribuição (*average joint degree*)
- Coeficiente de aglomeração (*clustering coefficient*)

Todas as características extraídas a partir das diversas configurações das redes complexas formaram os vetores de características utilizados para classificação das espécies. A Figura 6.5 ilustra algumas das principais características que podem ser extraídas a partir de uma rede complexa. Essas características podem ser utilizadas separadamente como vetores de características, ou combinadas por meio de técnicas de seleção de características. No caso da seleção de características, para cada característica extraída são selecionados os atributos com maior potencialidade de discriminação.

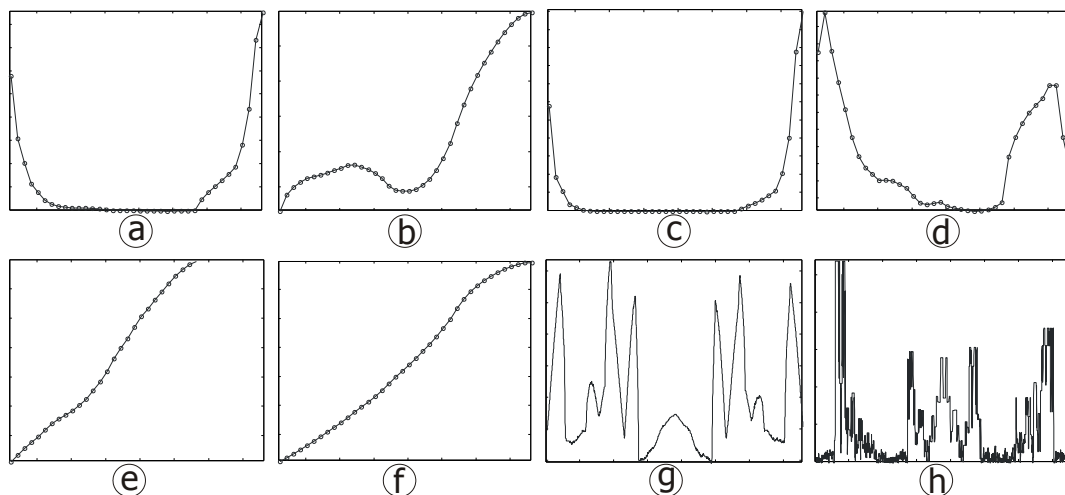


Figura 6.5: Principais características extraídas a partir de uma rede complexa: (a) grau médio distribuição de conexões; (b) coeficiente de aglomeração; (c) energia; (d); entropia; (e) grau máximo; (f) grau médio; (g); graus dos vértices para $T_i = 0.15$; (h) distribuição de conexões para $T_i = 0.15$.

O número total de experimentos realizados com as redes complexas foi 82. Assim, para facilitar a visualização apenas alguns destes experimentos são apresentados, os quais podem ser visualizados na Tabela 6.7.

Tabela 6.7: Experimentos realizados com as redes complexas.

ID	Título	Parâmetros	Número de Características		
			<i>Passiflora</i>	<i>Eucalyptus</i>	<i>Oryza</i>
F030	Rede Complexa Grau_0.025	$T_i = 0.025$	3766	881	1500
F031	Rede Complexa Grau_0.050	$T_i = 0.050$	3766	881	1500
...
F067	Rede Complexa Grau_0.950	$T_i = 0.950$	3766		
F068	Rede Complexa Grau Máximo	-	38	38	38
F069	Rede Complexa Grau Médio	-	38	38	38
F070	Rede Complexa Dist. Conexões_0.025	$T_i = 0.025$	3766	881	1500
F071	Rede Complexa Dist Conexões_0.050	$T_i = 0.050$	3766	881	1500
...
F107	Rede Complexa Dist. Conexões_0.950	$T_i = 0.950$		881	1500
F108	Rede Complexa Energia	-	38	38	38
F109	Rede Complexa Entropia	-	38	38	38
F110	Rede Complexa Grau Médio Dist.	-	38	38	38
F111	Rede Complexa Coef. Aglomera.	-	38	38	38

6.2.6 Análise de Formas: Wavelets

Na análise de formas biológicas por meio da transformada de *wavelets* foram consideradas as famílias *Daubechies* (db) e *Coiflets* (coif). Estas famílias foram empiricamente selecionadas. Na família *Daubechies*, um conjunto de quarenta e cinco *wavelets* foi utilizado $\{db1 (Haar), db2, db3, \dots, db45\}$, enquanto na família *Coiflets* foram consideradas as *wavelets* $\{coif1, coif2, coif3, coif4 e coif5\}$. A extração de características morfométricas utilizou tanto os coeficientes de tendência quanto de detalhe. Esses coeficientes foram extraídos dos cinco primeiros níveis de cada uma das *wavelets*.

O volume de informações coletado a partir da transformada de *wavelets* foi grande, o que possibilitou uma análise detalhada do contorno das formas biológicas. A Tabela 6.8 apresenta os experimentos realizados com a transformada de *wavelets* das espécies biológicas. O título de cada experimento descreve a família da *wavelets* utilizada, o nível (n), e o coeficiente (tendência ou detalhe).

Tabela 6.8: Experimentos realizados com a transformada de wavelets.

ID	Título	Parâmetros	Número de Características		
			<i>Passiflora</i>	<i>Eucalyptus</i>	<i>Oryza</i>
F112	<i>Wavelets_coif1_n1_ten</i>	<i>nível=1</i>	1855	443	752
F113	<i>Wavelets_coif1_n1_det</i>	<i>nível=1</i>	1855	443	752
...
F160	<i>Wavelets_coif5_n5_ten</i>	<i>nível=5</i>	145	55	74
F161	<i>Wavelets_coif5_n5_det</i>	<i>nível=5</i>	145	55	74
F162	<i>Wavelets_db1_n1_ten</i>	<i>nível=1</i>	1883	441	750
F163	<i>Wavelets_db1_n1_det</i>	<i>nível=1</i>	1883	441	750
...
F610	<i>Wavelets_db45_n5_ten</i>	<i>nível=5</i>	200	61	129
F611	<i>Wavelets_db45_n5_det</i>	<i>nível=5</i>	200	61	129

6.2.7 Estruturas Tubulares: Biometria

Os experimentos envolvendo a extração de características de estruturas tubulares consideraram um grupo específico de espécies do gênero *Passiflora*. Estas espécies foram escolhidas devido à grande variabilidade que apresentam quanto à forma de suas folhas e sistema de venação. Assim, para esse grupo de experimentos, onze espécies foram selecionadas sendo que para cada espécie quatro exemplares foram utilizados.

A análise biométrica das nervuras utilizou as características extraídas a partir das bifurcações e dos pontos extremos da estrutura. As informações biométricas utilizadas para classificação das espécies foram: número de bifurcações, número de pontos extremos; distância média e desvio padrão entre todas as bifurcações; distância média e desvio padrão entre todos os pontos extremos; área do sistema de venação e tamanho médio das nervuras. Assim, para cada exemplar de *Passiflora* oito características biométricas foram coletadas e combinadas em um vetor de características. Este experimento foi identificado como *E001*.

Para extração das características biométricas, inicialmente o sistema de venação foi segmentado utilizando a metodologia descrita na seção 3.3.1, e em seguida as informações foram coletadas com o algoritmo apresentado na seção 4.2.2. Todas as informações coletadas formaram o vetor contendo as características biométricas da estrutura.

6.2.8 Estruturas Tubulares: Assinatura Fractal

Para a análise das estruturas tubulares por meio da assinatura fractal multiescala, foram considerados quatro grupos de experimentos. No primeiro, a assinatura fractal foi calculada a partir de toda estrutura tubular. Nesse caso as dilatações exatas são empregadas em todos os pixels pertencentes ao sistema de venação. No segundo, foram utilizadas apenas as bifurcações para extrair a assinatura fractal. No terceiro, a assinatura fractal foi obtida considerando apenas as informações biométricas dos pontos extremos. Por fim, o último grupo agregou as informações biométricas das bifurcações e dos pontos extremos para o cálculo da assinatura fractal. A Figura 6.6 apresenta um exemplo das assinaturas fractais multiescala obtida a partir dos quatro grupos de experimentos.

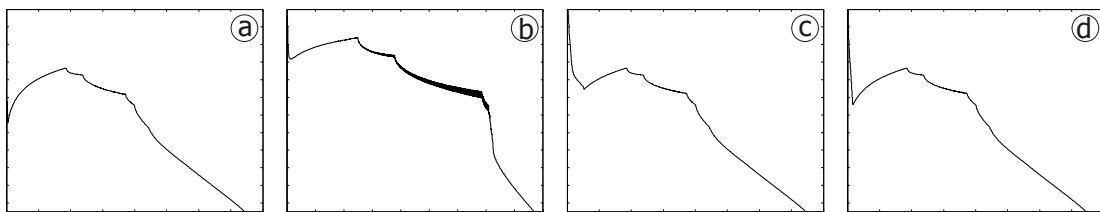


Figura 6.6: Assinatura fractal multiescala de estruturas tubulares para $k=250$ e $\sigma=250$. (a) todos os pontos (b) pontos de bifurcação; (c) pontos extremos; (d) combinação das bifurcações e dos pontos extremos.

Para a ordem de derivação e o desvio padrão da função gaussiana foram utilizados os mesmo valores empregados no experimento de análise de formas, sendo $k = \{100, 250, 500\}$ e $\sigma = \{100, 250, 350, 500\}$. Na Tabela 6.9 são descritos os experimentos executados com as estruturas tubulares das espécies de *Passiflora*.

Tabela 6.9: Experimentos realizados com a assinatura fractal multiescala das estruturas tubulares.

ID	Título	Parâmetros		Num. Características <i>Passiflora</i>
		k	σ	
E002	AssinaturaFractal_TodosPontos_100_100	100	100	2000
...
E013	AssinaturaFractal_TodosPontos_500_500	500	500	2000
E014	AssinaturaFractal_Bifurcacoes_100_100	100	100	2000
...
E025	AssinaturaFractal_Bifurcacoes_500_500	500	500	2000
E026	AssinaturaFractal_PtosExtremos_100_100	100	100	2000

...
E037	AssinaturaFractal_PtosExtremos_500_500	500	500	2000
E038	AssinaturaFractal_Bif_Extremos_100_100	100	100	2000
...
E049	AssinaturaFractal_Bif_Extremos_500_500	500	500	2000

6.2.9 Estruturas Tubulares: Redes Complexas

A extração de medidas morfométricas das estruturas tubulares por meio das redes complexas utilizou os mesmos padrões dos experimentos com assinatura fractal. Foram consideradas quatro estruturas de interesse: (i) todos os pontos da nervura; (ii) apenas pontos de bifurcação; (iii) apenas pontos extremos e (iv) pontos de bifurcações combinados aos pontos extremos.

As redes complexas foram geradas utilizando os mesmos limiares empregados nos experimentos de análise de formas, sendo $T_i = \{0.025, 0.050, \dots, 0.925, 0.950\}$. Com isso, as seguintes características foram extraídas das redes: graus dos vértices (g_i); grau máximo (g_{max}) dos vértices; grau médio (g_μ) dos vértices; distribuição das conexões (*joint degree distribution*); entropia (H); energia (E); grau médio de distribuição (*average joint degree*) e coeficiente de aglomeração (*clustering coefficient*). A Tabela 6.10 descreve os experimentos realizados com as redes complexas das estruturas tubulares.

Tabela 6.10: Experimentos realizados com as redes complexas das estruturas tubulares.

ID	Título	Parâmetros	Num. Características <i>Passiflora</i>
E050	RedesComplexas_TodosPontos_Grau_0.025	$T_i = 0.025$	1005
...
E087	RedesComplexas_TodosPontos_Grau_0.950	$T_i = 0.950$	1005
E088	RedesComplexas_TodosPontos_GrauMax	-	38
E089	RedesComplexas_TodosPontos_GrauMed	-	38
E090	RedesComplexas_TodosPontos_DistCon_0.025	$T_i = 0.025$	1005
...
E127	RedesComplexas_TodosPontos_DistCon_0.950	$T_i = 0.950$	1005
E128	RedesComplexas_TodosPontos_Energia	-	38

E129	RedesComplexas_TodosPontos_Entropia	-	38
E130	RedesComplexas_TodosPontos_GrauMed_Dist	-	38
E131	RedesComplexas_TodosPontos_Coef_Aglom	-	38
E132	RedesComplexas_Bifurcacoes_Grau_0.025	$T_i = 0.025$	20
...
E213	RedesComplexas_Bifurcacoes_Coef_Aglom		38
E214	RedesComplexas_PtosExtremos_Grau_0.025	$T_i = 0.025$	22
...
E295	RedesComplexas_PtosExtremos_Coef_Aglom		38
E296	RedesComplexas_Bifurc_Extremos_Grau_0.025	$T_i = 0.025$	42
...
E377	RedesComplexas_Bifurc_Extremos_Coef_Aglom	-	38

6.2.10 Textura: Análise Estatística

Para análise estatística da textura foram consideradas apenas as espécies de *Passiflora*. Esse tipo de análise tem como base o cálculo da matriz de coocorrência. Assim, para cada exemplar das espécies estudadas a matriz foi calculada considerando os seguintes ângulos: 0° , 45° , 90° e 135° . A partir dessas matrizes as seguintes informações foram extraídas: energia, contraste, correlação e entropia. Esses valores foram combinados, formando o vetor de características que descreve as informações a respeito da textura da imagem biológica. A Tabela 6.11 descreve os experimentos realizados para análise da textura das espécies biológicas.

Tabela 6.11: Experimentos realizados com análise estatística da textura.

ID	Título	Parâmetros	Num.Características
			<i>Passiflora</i>
T001	TexturaEstatistica_000	$\theta = 0$	4
T002	TexturaEstatistica_045	$\theta = 45$	4
T003	TexturaEstatistica_090	$\theta = 90$	4
T004	TexturaEstatistica_135	$\theta = 135$	4
T005	TexturaEstatistica_AngulosCombinados	-	16

T107	TexturaRedeComplexa_DistCon_0950	1.0	-	-	0.950	1400
T108	TexturaRedeComplexa_Energia	1.0	-	-	-	38
T109	TexturaRedeComplexa_Entropia	1.0	-	-	-	38
T110	TexturaRedeComplexa_GrauMedDist	1.0	-	-	-	38
T111	TexturaRedeComplexa_CoefAglom	1.0	-	-	-	38
T112	TexturaRedeComplexa_Grau_0.025	0.8	-	-	0.0250	1400
...	-	-
T192	TexturaRedeComplexa_CoefAglom	0.8	-	-	-	38

6.3 Reconhecimento de Padrões

Os experimentos realizados com as técnicas de análise de imagens produziram diversos conjuntos de dados contendo as informações morfométricas das espécies biológicas. Esses dados foram utilizados como entrada para os métodos de reconhecimento de padrões. O processo de reconhecimento de padrões foi baseado nas técnicas de análise estatística multivariada e redes neurais artificiais. Além disso, métodos de seleção de características foram empregados para reduzir a dimensionalidade e selecionar as características mais discriminantes do conjunto. A Figura 6.7 apresenta uma visão estrutural das técnicas de reconhecimento de padrões utilizadas nos experimentos.

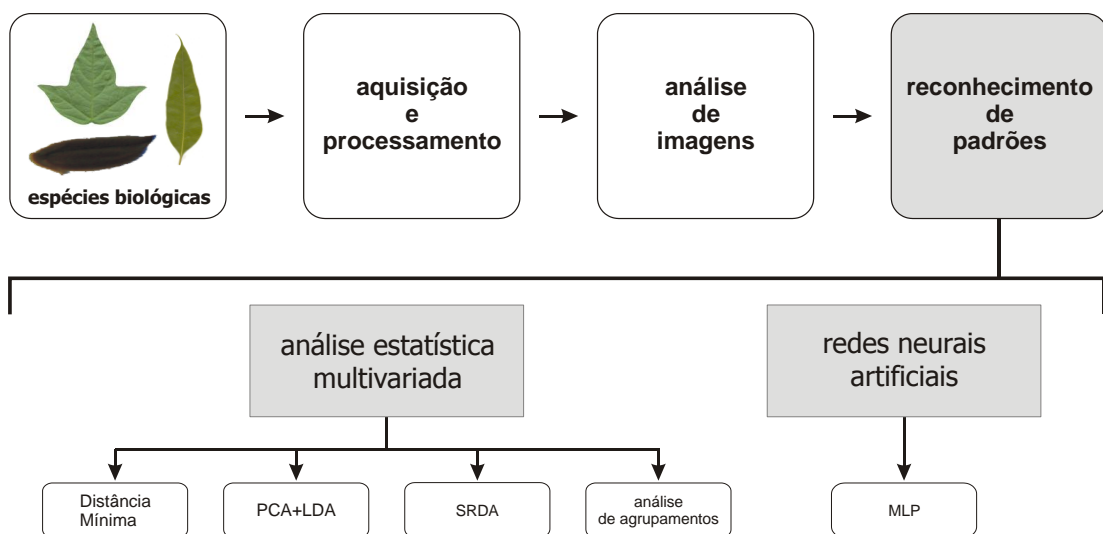


Figura 6.7: Visão estrutural das técnicas de reconhecimento de padrões utilizadas nos experimentos.

Para o reconhecimento de padrões baseado nas informações morfométricas das espécies, inicialmente, o banco de dados de características foi normalizado no intervalo $[-2, +2]$, conforme descrito em (Costa e Cesar, 2000). Esse critério garante que as diferentes unidades de medida dos vetores de características não influenciem no processo de classificação. Para aumentar a acurácia do processo de classificação o método de validação cruzada *k-fold* foi utilizado para particionar os dados em conjunto de treinamento e teste.

O método de validação cruzada *k-fold* divide o conjunto de dados em k partições mutuamente exclusivas. Os dados selecionados para cada partição são definidos aleatoriamente. Em cada iteração do método, uma das k partições é utilizada como conjunto de teste e as outras $k - 1$ são usadas como conjunto de treinamento. Nos experimentos foi adotado $k = 10$ para o número de partições. O classificador é construído a partir dos exemplos selecionados para o conjunto de treinamento. E para avaliação do classificador são utilizados os elementos do conjunto de teste. Este procedimento é realizado 10 vezes (validação cruzada). A Figura 6.8 ilustra o processo de reconhecimento de padrões, em que é destacado o método de validação cruzada.

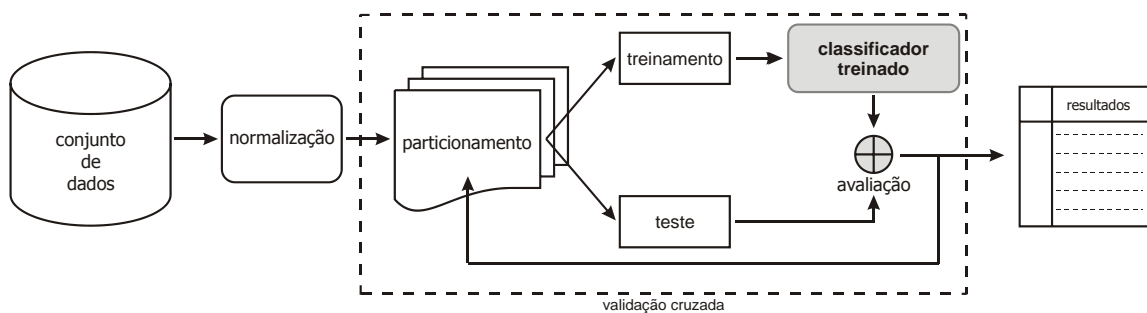


Figura 6.8: Processo de reconhecimento de padrões adotado nos experimentos.

No processo de treinamento do classificador foram utilizados os métodos de reconhecimento de padrões descritos no capítulo 5. A Figura 6.9 ilustra o processo de classificação das características morfométricas extraídas das espécies.

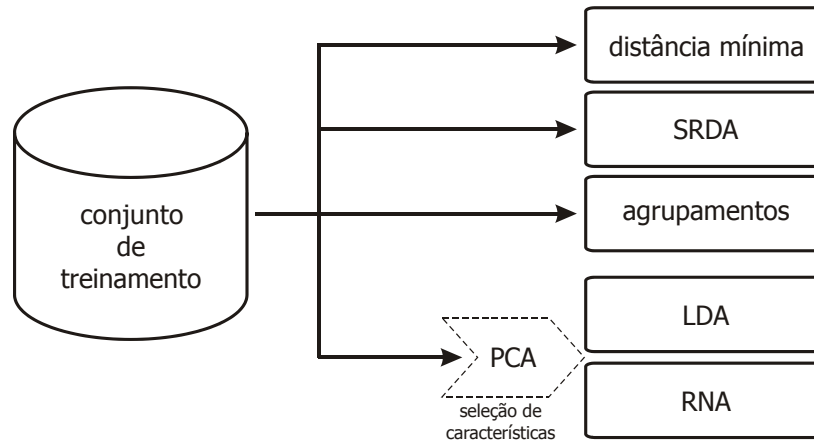


Figura 6.9: Processo de classificação do conjunto de dados de treinamento com as técnicas de reconhecimento de padrões

O classificador euclidiano por distância mínima foi construído com base nas distâncias entre os centros de massa das espécies. Assim, para cada conjunto de treinamento, de todos os experimentos realizados, foi calculado o centroide a partir dos vetores de características extraídos de cada espécie. A distância euclidiana entre os centroides foi utilizada como função discriminante das espécies.

A análise discriminante por regressão espectral (SRDA) é um método robusto para classificação de conjuntos de dados com grande dimensionalidade. O classificador foi construído a partir da resolução de um conjunto de problemas de regressão linear. Este classificador foi aplicado em todos os conjuntos de dados obtidos das espécies.

A análise de agrupamentos foi realizada com base no método de agrupamento hierárquico UPGMA com as métricas de distância euclidiana e *chebychev*. Esse método considera a distância média entre todos os pares de padrões de dois diferentes clusters como critério de classificação. Com os agrupamentos foi possível observar a taxonomia das espécies em função da organização hierárquica dos dendrogramas.

Na análise discriminante linear (LDA), para garantir a confiabilidade da classificação e minimizar o problema do *mau da dimensionalidade*, o método foi combinado com a Análise de Componentes Principais (PCA). A técnica de PCA foi aplicada anteriormente ao LDA para selecionar os k componentes principais com maior variabilidade (Belhumeur, Hespanha *et al.*,

1997). Foram selecionados k componentes principais, tal que k é menor ou igual ao número de espécies.

As redes neurais artificiais (RNA) foram utilizadas para classificação de conjuntos específicos de dados. Esses conjuntos foram formados com base na combinação de múltiplas características das espécies. Por exemplo, um conjunto foi criado a partir de dados selecionados por PCA contendo apenas as características com variância total superior 85%; outro conjunto foi construído considerando a variância total superior a 95%; dentre outros. Para o treinamento da rede foi utilizado o algoritmo *back-propagation* em uma rede do tipo *perceptron* multicamadas (MLP). A topologia da rede foi definida experimentalmente, em que o número de entradas é igual ao número de características analisadas; o número de camadas ocultas foi determinado como dois (para solução de problemas não linearmente separáveis); e o número de saídas igual ao número de espécies (por exemplo, para as espécies de arroz foram utilizadas 12 saídas). Como parâmetros, a rede foi configurada com a função de transferência sigmoide; taxa de aprendizado igual a 0.1; e número de épocas de treinamento igual a 5000.

Os resultados do processo de reconhecimento de padrões foram apresentados por meio de matrizes de confusão, gráficos ROC, além de dados estatísticos como acurácia, sensibilidade (*recall*), precisão, dentre outros.

6.4 Considerações Finais

Neste capítulo foi apresentada uma descrição detalhada dos experimentos realizados para análise e classificação de espécies biológicas. Foram descritos cada um dos experimentos executados para análise de formas, análise de estruturas tubulares e análise de textura. Além disso, foram apresentados os parâmetros utilizados pelas técnicas de extração de características, bem como o número de atributos extraídos. No próximo capítulo são relatados os resultados obtidos a partir destes experimentos.

7

Resultados e Discussão

Este capítulo apresenta os resultados dos experimentos realizados para análise e classificação das espécies biológicas. Um extenso conjunto de características morfométricas foi extraído com base em três grupos de interesse: formas, estrutura tubular e textura. Os vetores de características coletados com as técnicas de análise de imagens foram utilizados como entrada para os métodos de reconhecimento de padrões. Esses métodos foram empregados para avaliar a qualidade das características extraídas, bem como para realizar a classificação das espécies. Para discriminação das espécies foram utilizados métodos baseados em análise estatística multivariada e redes neurais artificiais.

Os resultados foram organizados seguindo a ordem das técnicas de análise de imagens utilizadas nos experimentos. Assim, inicialmente são apresentados os resultados da classificação das espécies com as características extraídas das técnicas de análise de formas, em seguida, os resultados da classificação com as características obtidas das estruturas tubulares e por fim, a classificação baseada nas informações coletadas da textura.

Para apresentação dos resultados são utilizadas tabelas em que são descritas as porcentagens de acerto (%*A*) e erro (%*E*). As tabelas demonstram comparativamente os resultados das classificações de cada uma das técnicas utilizadas, em função de cada um dos métodos de reconhecimento de padrões, e para cada uma das espécies. Com isso, é possível identificar qual metodologia de análise de imagens e reconhecimento de padrões é mais adequada para classificação de uma determinada espécie.

Os experimentos com as melhores porcentagens de acerto foram selecionados para uma descrição mais detalhada. Estes experimentos foram identificados na tabela por meio de linhas tracejadas. Para visualização mais abrangente dos resultados, foram utilizadas matrizes de confusão, além de diversas informações estatísticas, tais como: acurácia (AC), taxa de verdadeiro positivo (T_{VP}), taxa de falso positivo (T_{FP}), especificidade (ES), precisão (PR) e F -Measure (F). A partir das taxas de verdadeiro positivo e falso positivo, foi criado um gráfico ROC. Nos gráficos ROC, os pontos mais próximos a (0,1), ou seja, maior T_{VP} e menor T_{FP} , representam a classificação perfeita.

7.1 Resultados da Classificação baseada na Análise de Formas

As técnicas de análise formas foram utilizadas para extração de características morfométricas com base no contorno da imagem biológica. Neste contexto as técnicas utilizadas foram: (a) assinatura; (b) curvatura; (c) descritores de Fourier; (d) dimensão fractal multiescala; (e) redes complexas e (f) *wavelets*. Além disso, o item (g) descreve os resultados obtidos com a seleção das características com maior potencialidade discriminante. Os resultados do processo de classificação são apresentados como se segue:

A. Resultados da Assinatura Digital

A Tabela 7.1 apresenta os resultados da classificação das espécies biológicas por meio das características morfométricas extraídas a partir da assinatura digital. São descritas as porcentagens de acerto e erro para cada classificador utilizado, bem como, para cada espécie biológica estudada.

Nesta tabela são descritos os resultados dos experimentos:

- F001 - assinatura centroide;
- F002 – assinatura projeção 45°;
- F003 – assinatura projeção 90°;
- F004 – assinatura projeção 180°;

Tabela 7.1: Resultados da classificação das espécies com características da assinatura digital.

Assinatura		Maracujá (<i>Passiflora</i>)			Eucalipto (<i>Eucalyptus</i>)			Arroz (<i>Oryza</i>)		
		DM	LDA	SRDA	DM	LDA	SRDA	DM	LDA	SRDA
F001	%A	75,23	83,18	89,75	42,56	50,46	55,81	22,31	48,08	46,15
	%E	24,77	16,82	10,25	57,44	49,54	44,19	77,69	51,92	53,85
F002	%A	75,11	81,14	80,68	65,58	62,79	65,12	35,39	45,38	50,00
	%E	24,89	18,86	19,32	34,42	37,21	34,88	64,61	54,62	50,00
F003	%A	82,95	85,68	84,09	68,37	70,00	65,12	40,39	41,92	57,69
	%E	17,05	14,32	15,91	31,63	30,00	34,88	59,61	58,08	42,31
F004	%A	80,91	83,75	82,95	50,46	57,21	72,09	45,38	50,00	61,54
	%E	19,09	16,25	17,04	49,54	42,79	27,91	54,62	50,00	38,46

Os resultados obtidos com as assinaturas digitais demonstram a viabilidade da utilização da metodologia na análise de espécies biológicas. No caso das espécies de maracujás, por exemplo, a porcentagem de acerto no processo de classificação foi igual a 89,75. Além disso, por ser uma técnica simples é importante destacar que os resultados para as espécies de eucalipto (70,00%) e arroz (61,54%) também foram promissores. A Tabela 7.2 apresenta detalhadamente os resultados dos melhores experimentos realizados com a assinatura digital. Na Tabela são descritas as porcentagens de acerto (%A) e erro (%E) para cada espécie, além da acurácia (AC) e *F-measure* (F). Ao final de cada coluna são apresentadas as médias e os desvios padrão.

Tabela 7.2: Detalhamento dos resultados da classificação das espécies com os dados da assinatura digital.

	Maracujá (<i>Passiflora</i>)				Eucalipto (<i>Eucalyptus</i>)				Arroz (<i>Oryza</i>)			
	%A	%E	AC	F	%A	%E	AC	F	%A	%E	AC	F
C ₀₁	83,33	16,67	98,86	90,91	0,00	100,00	93,02	0,00	20,00	80,00	73,08	22,22
C ₀₂	100,00	0,00	100,00	100,00	50,00	50,00	93,02	40,00	50,00	50,00	96,15	66,67
C ₀₃	100,00	0,00	100,00	100,00	100,00	0,00	97,67	80,00	50,00	50,00	96,15	66,67
C ₀₄	100,00	0,00	100,00	100,00	100,00	0,00	97,67	88,89	50,00	50,00	88,46	40,00
C ₀₅	50,00	50,00	95,45	66,67	85,71	14,29	93,02	80,00	50,00	50,00	96,15	66,67
C ₀₆	87,50	12,50	94,32	73,68	50,00	50,00	95,35	50,00	100,00	0,00	100,00	100,00
C ₀₇	100,00	0,00	98,86	88,89	25,00	75,00	90,70	33,33	100,00	0,00	96,15	80,00
C ₀₈	100,00	0,00	100,00	100,00	80,00	20,00	95,35	80,00	50,00	50,00	92,31	50,00
C ₀₉	66,67	33,33	97,73	80,00	75,00	25,00	95,35	75,00	100,00	0,00	100,00	100,00
C ₁₀	100,00	0,00	98,86	92,31	75,00	25,00	93,02	85,71	0,00	100,00	96,15	0,00
C ₁₁	83,33	16,67	97,73	83,33	-	-	-	-	100,00	0,00	96,15	80,00
C ₁₂	100,00	0,00	98,86	94,74	-	-	-	-	100,00	0,00	92,31	75,00
C ₁₃	100,00	0,00	98,86	85,71	-	-	-	-	-	-	-	-
μ	90,06	9,94	98,43	88,94	64,07	35,93	94,42	61,29	64,17	35,83	93,59	62,27
σ	15,92	15,92	1,77	10,76	32,61	32,61	2,25	29,28	35,02	35,02	7,21	29,88

A Figura 7.1 apresenta comparativamente o desempenho da classificação das espécies por meio de um gráfico ROC (*Receiver Operation Characteristics*). No gráfico ROC a linha diagonal tracejada representa o desempenho de um classificador aleatório, essa linha serve de referência para avaliação do classificador. Os pontos no espaço ROC acima da diagonal representam uma melhor classificação do que os pontos abaixo da diagonal. Um classificador perfeito é aquele que produz como resultado um ponto próximo a (0,1). Isto significa que este classificador tem uma taxa nula (igual a zero) de falsos positivos, e uma taxa de 100% de acerto de verdadeiros positivos. Analisando a Figura 7.1 é possível notar que o desempenho da classificação das espécies de arroz foi inferior às espécies de eucalipto e maracujá.

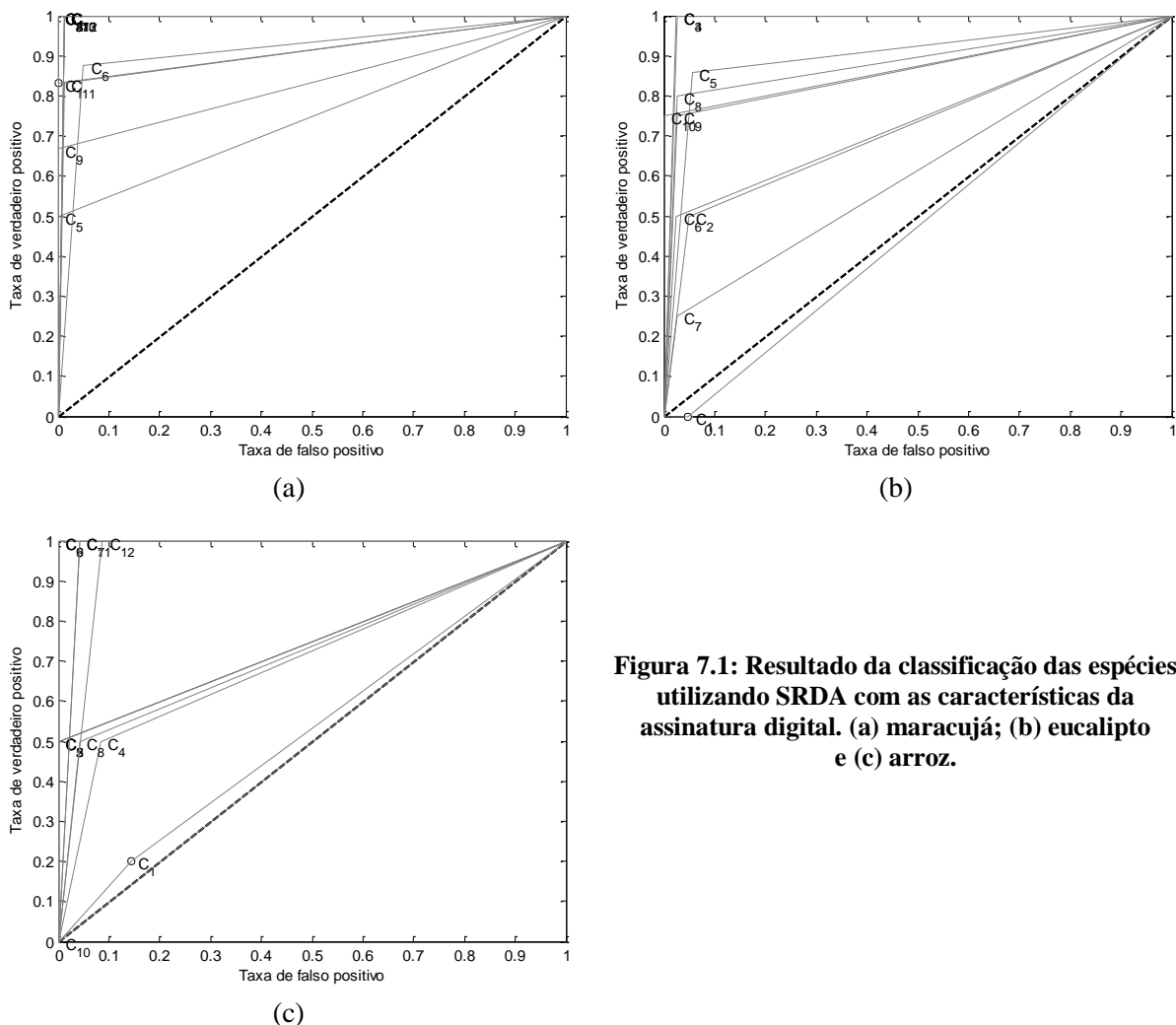


Figura 7.1: Resultado da classificação das espécies utilizando SRDA com as características da assinatura digital. (a) maracujá; (b) eucalipto e (c) arroz.

B. Resultados da Curvatura

Os resultados da análise da curvatura foram obtidos a partir de duas abordagens: curvatura por definição (F005) e curvatura multiescala (F006 até F011). Os diversos resultados da curvatura multiescala foram produzidos em virtude dos valores utilizados como desvio padrão da

função gaussiana de suavização. Na Tabela 7.3 são apresentados os resultados obtidos com a classificação das espécies por meio das informações coletadas com a análise da curvatura.

Tabela 7.3: Resultados da classificação das espécies com base em características curvatura.

Curvatura	Maracujá (<i>Passiflora</i>)			Eucalipto (<i>Eucalyptus</i>)			Arroz (<i>Oryza</i>)			
		DM	LDA	SRDA	DM	LDA	SRDA	DM	LDA	SRDA
F005	%A	73,86	74,43	78,64	37,91	45,35	46,51	18,08	22,31	19,23
	%E	26,14	25,57	21,36	62,09	54,65	53,49	81,92	77,69	80,77
F006	%A	65,91	67,16	70,45	46,05	28,84	46,51	17,69	18,85	26,92
	%E	34,09	32,84	29,54	53,95	71,16	53,49	82,31	81,15	73,08
F007	%A	48,52	34,89	34,09	36,05	20,93	32,56	21,15	11,54	15,39
	%E	51,48	65,11	65,91	63,95	79,07	67,44	78,85	88,46	84,61
F008	%A	28,98	12,96	19,32	26,98	13,72	16,28	20,38	9,61	7,69
	%E	71,02	87,04	80,68	73,02	86,28	83,72	79,62	90,39	92,31
F009	%A	73,86	84,09	77,27	52,09	53,49	53,49	17,31	27,69	26,92
	%E	26,14	15,91	22,73	47,91	46,51	46,51	82,69	72,31	73,08
F010	%A	73,10	81,99	76,51	51,33	52,33	52,73	16,55	26,93	26,16
	%E	26,90	18,01	23,49	48,67	47,67	47,27	83,45	73,07	73,84
F011	%A	72,63	81,52	76,04	50,86	55,81	52,26	16,08	26,46	25,69
	%E	27,37	18,48	23,96	49,14	44,19	47,74	83,92	73,54	74,31

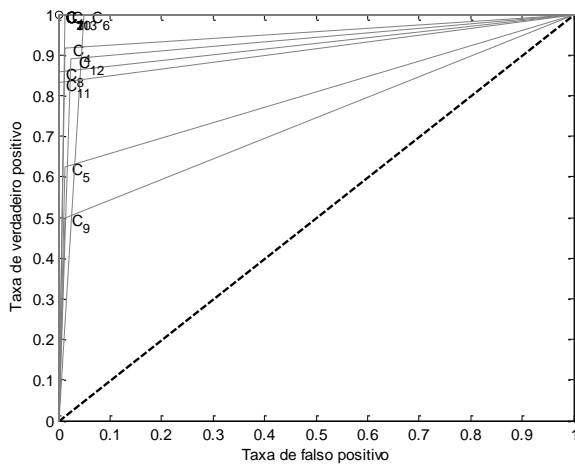
Para a curvatura, os melhores resultados foram obtidos com a abordagem multiescala. Nas espécies de maracujá o parâmetro de suavização da função gaussiana foi igual a $\tau=0.50$, enquanto para as espécies de eucalipto o valor adotado foi $\tau = 0.10$ e para as espécies de arroz $\tau = 0.50$. O fraco desempenho da classificação das espécies de arroz é justificado pela pequena variabilidade do contorno, o que torna este processo de classificação complexo. Dessa forma, o descritor de curvatura pode ser considerado uma propriedade ruim para discriminação de espécies de arroz. A Tabela 7.4 apresenta detalhadamente os dados estatísticos a respeito da classificação das espécies com a curvatura.

Tabela 7.4: Detalhamento dos resultados da classificação das espécies com os dados da curvatura.

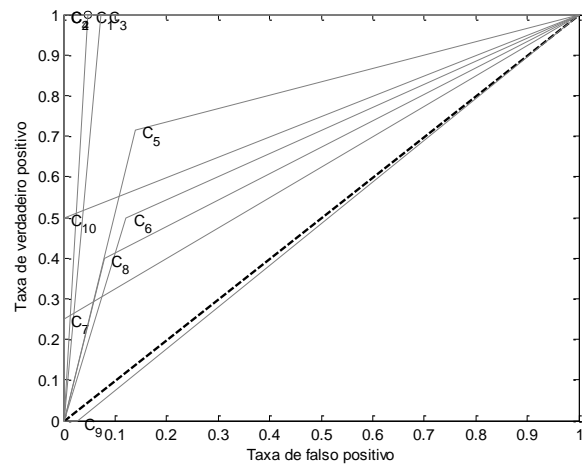
	Maracujá (<i>Passiflora</i>)				Eucalipto (<i>Eucalyptus</i>)				Arroz (<i>Oryza</i>)			
	%A	%E	AC	F	%A	%E	AC	F	%A	%E	AC	F
C ₀₁	100,00	0,00	100,00	100,00	100,00	0,00	95,35	50,00	20,00	80,00	78,08	25,97
C ₀₂	100,00	0,00	100,00	100,00	100,00	0,00	100,00	100,00	5,00	95,00	88,85	6,45
C ₀₃	100,00	0,00	100,00	100,00	100,00	0,00	93,02	57,14	20,00	80,00	85,38	17,39
C ₀₄	91,67	8,33	97,73	91,67	100,00	0,00	100,00	100,00	35,00	65,00	91,92	40,00
C ₀₅	62,50	37,50	95,45	71,43	71,43	28,57	83,72	58,82	20,00	80,00	87,69	20,00
C ₀₆	100,00	0,00	95,45	80,00	50,00	50,00	86,05	25,00	50,00	50,00	92,31	50,00
C ₀₇	100,00	0,00	100,00	100,00	25,00	75,00	93,02	40,00	25,00	75,00	90,00	27,78

C ₀₈	85,71	14,29	98,86	92,31	40,00	60,00	86,05	40,00	20,00	80,00	86,54	18,60
C ₀₉	50,00	50,00	95,45	60,00	0,00	100,00	88,37	0,00	80,00	20,00	88,85	35,56
C ₁₀	100,00	0,00	100,00	100,00	50,00	50,00	86,05	66,67	40,00	60,00	92,69	29,62
C ₁₁	83,33	16,67	98,86	90,91	-	-	-	-	50,00	50,00	87,31	37,74
C ₁₂	88,89	11,11	96,59	84,21	-	-	-	-	16,67	83,33	85,77	21,28
C ₁₃	100,00	0,00	98,86	85,71	-	-	-	-	-	-	-	-
μ	89,39	10,61	98,25	88,94	63,64	36,36	91,16	53,76	31,81	68,19	87,95	27,51
σ	16,12	16,12	1,89	12,55	36,24	36,24	5,98	30,89	20,48	20,48	3,98	11,90

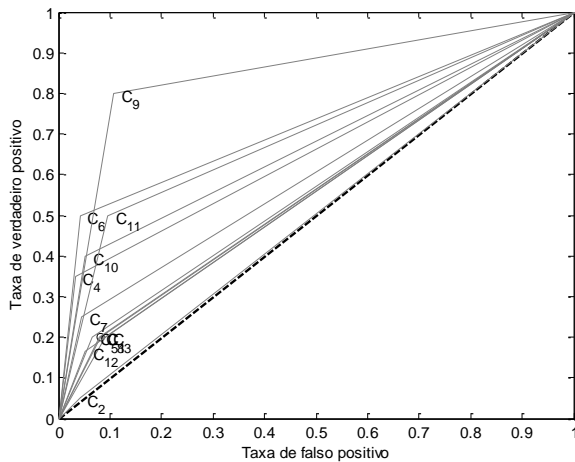
Os resultados obtidos com as espécies de maracujá e eucalipto são promissores e sinalizam a viabilidade destas características na classificação das espécies. Comparativamente, a Figura 7.2 apresenta as taxas de verdadeiro positivo e falso positivo resultantes do processo de classificação das espécies. Estes dados são expressos por meio de pontos no espaço ROC. No gráfico da Figura 7.2b é possível notar que algumas espécies de eucalipto dificultaram o processo de classificação, essas espécies identificadas de C₀₆ até C₁₀ são híbridas, enquanto as espécies puras (de C₀₁ até C₀₄) foram perfeitamente classificadas.



(a)



(b)



(c)

Figura 7.2: Resultado da classificação utilizando LDA com as características da curvatura. (a) maracujá; (b) eucalipto e (c) arroz.

C. Resultados dos Descritores de Fourier

Com os descritores de Fourier foi possível transformar o contorno da imagem biológica em um sinal unidimensional. Dessa maneira, vetores de características foram extraídos a partir dos componentes de baixa frequência da transformada de Fourier. A Tabela 7.5 apresenta os resultados obtidos com a classificação das espécies com os descritores de Fourier. O número de descritores utilizado em cada experimento é descrito na seção 6.2.3.

Tabela 7.5: Resultados da classificação das espécies com as características dos descritores de Fourier.

Descritores de Fourier	Maracujá (<i>Passiflora</i>)			Eucalipto (<i>Eucalyptus</i>)			Arroz (<i>Oryza</i>)			
	DM	LDA	SRDA	DM	LDA	SRDA	DM	LDA	SRDA	
F012	%A	72,05	71,25	75,91	47,21	65,12	56,51	32,69	38,46	52,69
	%E	27,95	28,75	24,09	52,79	34,88	43,49	67,31	61,54	47,31
F013	%A	83,07	65,00	84,32	51,40	55,58	53,49	33,08	16,92	50,39
	%E	16,93	35,00	15,68	48,60	44,42	46,51	66,92	83,08	49,61
F014	%A	80,57	85,34	88,64	45,35	44,19	34,88	40,00	36,92	57,31
	%E	19,43	14,66	11,36	54,65	55,81	65,12	60,00	63,08	42,69
F015	%A	77,84	83,86	86,36	51,63	36,74	25,58	39,23	34,61	53,85
	%E	22,16	16,14	13,64	48,37	63,26	74,42	60,77	65,39	46,15
F016	%A	76,70	84,20	84,09	48,60	35,81	39,53	35,00	32,69	57,69
	%E	23,29	15,80	15,91	51,40	64,19	60,47	65,00	67,31	42,31
F017	%A	73,07	78,18	79,55	43,26	33,95	34,88	29,23	33,85	42,31
	%E	26,93	21,82	20,46	56,74	66,05	65,12	70,77	66,15	57,69

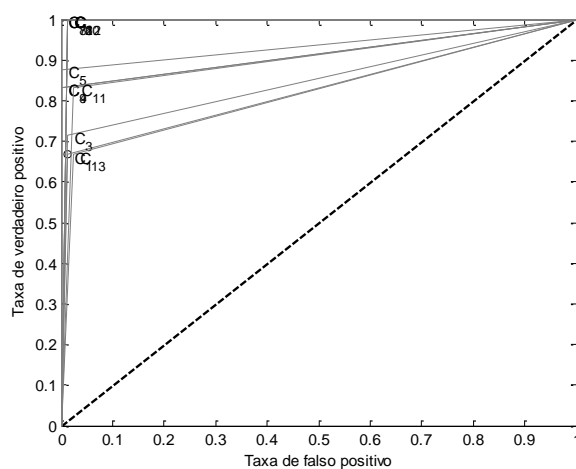
Os experimentos com os descritores de Fourier indicaram que um número elevado de características não produz resultados promissores, como por exemplo, os 320 descritores utilizados no experimento F017. Para todas as espécies estudadas, as melhores porcentagens de classificação corretas foram alcançadas com no máximo 40 descritores. Este fato é muito importante para discriminação das espécies, uma vez não necessita de um conjunto de dados com alta dimensionalidade e acelera o processo de classificação.

A Tabela 7.6 apresenta uma análise descritiva do processo de classificação das espécies com os descritores de Fourier. Para as espécies de maracujá e arroz o melhor resultado foi obtido com 40 descritores, enquanto para as espécies de eucalipto 10 descritores foram utilizados.

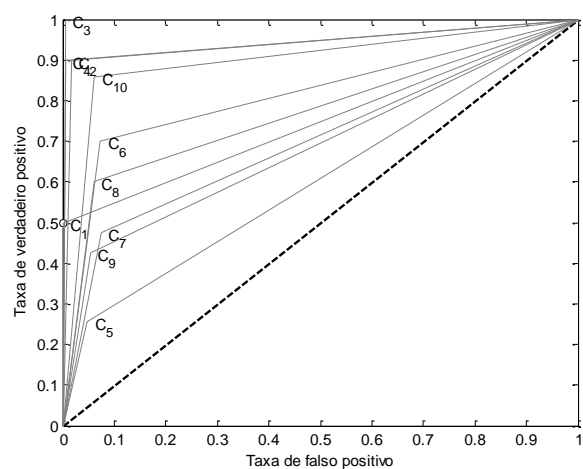
Tabela 7.6: Detalhamento dos resultados da classificação das espécies com os descritores de Fourier.

	Maracujá (<i>Passiflora</i>)				Eucalipto (<i>Eucalyptus</i>)				Arroz (<i>Oryza</i>)			
	%A	%E	AC	F	%A	%E	AC	F	%A	%E	AC	F
C ₀₁	66,67	33,33	96,59	72,73	50,00	50,00	98,84	0,00	40,00	60,00	84,62	50,00
C ₀₂	100,00	0,00	98,86	92,31	90,00	10,00	97,91	82,05	50,00	50,00	96,15	66,67
C ₀₃	71,43	28,57	96,59	76,92	100,00	0,00	99,53	96,00	0,00	100,00	84,62	0,00
C ₀₄	83,33	16,67	97,73	90,91	90,00	10,00	98,60	91,59	100,00	0,00	92,31	66,67
C ₀₅	87,50	12,50	98,86	93,33	25,71	74,29	83,95	0,00	50,00	50,00	92,31	50,00
C ₀₆	100,00	0,00	98,86	94,12	70,00	30,00	91,63	43,18	100,00	0,00	96,15	80,00
C ₀₇	100,00	0,00	100,00	100,00	47,50	52,50	88,37	0,00	100,00	0,00	96,15	80,00
C ₀₈	100,00	0,00	100,00	100,00	60,00	40,00	90,00	56,82	0,00	100,00	88,46	0,00
C ₀₉	83,33	16,67	98,86	90,91	42,50	57,50	89,77	41,14	100,00	0,00	100,00	100,00
C ₁₀	100,00	0,00	98,86	92,31	85,83	14,17	91,63	85,07	0,00	100,00	92,31	0,00
C ₁₁	83,33	16,67	96,59	76,92	-	-	-	-	100,00	0,00	96,15	80,00
C ₁₂	100,00	0,00	98,86	94,74	-	-	-	-	66,67	33,33	96,15	80,00
C ₁₃	66,67	33,33	96,59	57,14	-	-	-	-	-	-	-	-
μ	87,87	12,13	98,25	87,10	66,15	33,85	93,02	49,58	58,89	41,11	92,95	54,44
σ	13,27	13,27	1,28	12,50	24,78	24,78	5,36	39,12	42,03	42,03	4,87	35,57

A Figura 7.3 apresenta o resultado da classificação das espécies com os dados dos descritores de Fourier. Na Figura é importante notar o resultado da classificação das espécies de arroz, em que, de um total de 12 espécies, apenas três espécies tiveram desempenho ruim (abaixo da linha diagonal). Este fato prejudicou o desempenho global do classificador, porém demonstra que os descritores de Fourier são características morfométricas significantes para classificação das espécies.



(a)



(b)

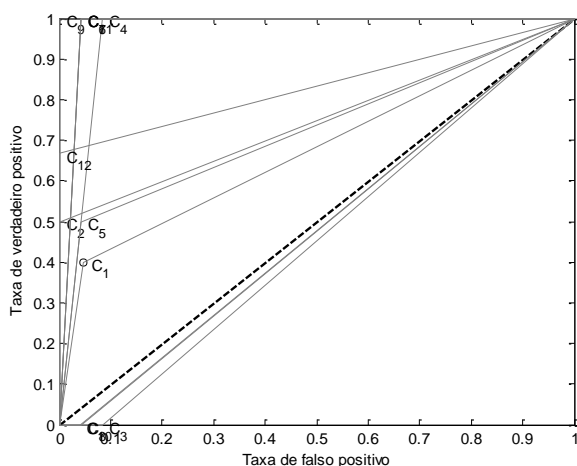


Figura 7.3: Resultado da classificação das espécies com as características dos descritores Fourier. (a) maracujá (SRDA); (b) eucalipto (LDA) e (c) arroz (SRDA).

(c)

D. Resultados da Dimensão Fractal Multiescala

A análise por meio da dimensão fractal multiescala possibilitou um estudo abrangente a respeito da complexidade do contorno das imagens biológicas. Diversas configurações foram testadas para geração da curva fractal, tanto com a mudança do fator de derivação, quanto com a alteração do desvio padrão da função de suavização gaussiana. Os experimentos envolvendo a dimensão fractal multiescala são identificados de F018 até F029. A Tabela 7.7 apresenta os resultados obtidos com a análise do contorno por meio da dimensão fractal multiescala.

Tabela 7.7: Resultados da classificação com as características da dimensão fractal multiescala.

Dimensão Fractal		Maracujá (<i>Passiflora</i>)			Eucalipto (<i>Eucalyptus</i>)			Arroz (<i>Oryza</i>)		
		DM	LDA	SRDA	DM	LDA	SRDA	DM	LDA	SRDA
F018	%A	70,91	84,32	79,55	40,00	46,51	55,81	34,61	47,69	50,00
	%E	29,09	15,68	20,46	60,00	53,49	44,19	65,39	52,31	50,00
F019	%A	70,45	89,32	89,32	40,93	52,09	60,47	34,23	51,92	62,32
	%E	29,54	10,68	10,68	59,07	47,91	39,53	65,77	48,08	37,68
F020	%A	70,11	85,00	84,09	41,40	57,44	62,79	34,23	50,00	53,85
	%E	29,89	15,00	15,91	58,60	42,56	37,21	65,77	50,00	46,15
F021	%A	70,23	84,66	86,36	41,40	56,98	62,79	34,23	50,39	53,65
	%E	29,77	15,34	13,64	58,60	43,02	37,21	65,77	49,61	46,35
F022	%A	70,91	84,09	79,55	40,46	43,49	51,16	34,23	49,23	53,55
	%E	29,09	15,91	20,46	59,54	56,51	48,84	65,77	50,77	46,45
F023	%A	70,34	84,89	85,40	40,93	52,56	60,47	34,23	53,08	57,69
	%E	29,66	15,11	14,60	59,07	47,44	39,53	65,77	46,92	42,31
F024	%A	70,23	85,00	85,23	41,40	57,67	62,69	34,23	50,39	53,05
	%E	29,77	15,00	14,77	58,60	42,33	37,31	65,77	49,61	46,95
F025	%A	70,23	84,77	86,36	41,40	56,98	62,79	34,23	50,00	53,25

	%E	29,77	15,23	13,64	58,60	43,02	37,21	65,77	50,00	46,75
F026	%A	70,45	84,55	82,95	39,77	49,53	60,47	33,85	47,69	63,31
	%E	29,54	15,46	17,04	60,23	50,47	39,53	66,15	52,31	38,46
F027	%A	70,11	84,66	85,23	43,40	53,72	65,21	34,23	50,77	53,15
	%E	29,89	15,34	14,77	55,60	46,28	34,88	65,77	49,23	46,85
F028	%A	70,23	85,00	85,23	41,20	56,98	62,59	34,23	50,00	53,85
	%E	29,77	15,00	14,77	58,80	43,02	37,41	65,77	50,00	46,15
F029	%A	70,23	84,77	86,36	42,40	57,21	61,19	34,23	50,39	53,65
	%E	29,77	15,23	13,64	57,60	42,79	38,81	65,77	49,61	46,35

A Tabela 7.8 apresenta as informações estatísticas da classificação das espécies com as características da dimensão fractal multiescala. Na Tabela é descrito o melhor resultado de cada gênero estudado.

Tabela 7.8: Detalhamento dos resultados da classificação das espécies com a dimensão fractal multiescala.

	Maracujá (<i>Passiflora</i>)				Eucalipto (<i>Eucalyptus</i>)				Arroz (<i>Oryza</i>)			
	%A	%E	AC	F	%A	%E	AC	F	%A	%E	AC	F
C ₀₁	100,00	0,00	100,00	100,00	0,00	100,00	95,35	0,00	60,00	40,00	92,31	75,00
C ₀₂	100,00	0,00	98,86	92,31	100,00	0,00	97,67	80,00	100,00	0,00	96,15	80,00
C ₀₃	71,43	28,57	97,73	83,33	100,00	0,00	95,35	66,67	0,00	100,00	92,31	0,00
C ₀₄	100,00	0,00	100,00	100,00	75,00	25,00	95,35	75,00	50,00	50,00	92,31	50,00
C ₀₅	75,00	25,00	96,59	80,00	71,43	28,57	93,02	76,92	0,00	100,00	92,31	0,00
C ₀₆	75,00	25,00	95,45	75,00	100,00	0,00	93,02	57,14	100,00	0,00	100,00	100,00
C ₀₇	100,00	0,00	98,86	88,89	25,00	75,00	88,37	28,57	100,00	0,00	96,15	80,00
C ₀₈	100,00	0,00	100,00	100,00	80,00	20,00	90,70	66,67	50,00	50,00	88,46	40,00
C ₀₉	83,33	16,67	98,86	90,91	0,00	100,00	88,37	0,00	100,00	0,00	96,15	66,67
C ₁₀	83,33	16,67	97,73	83,33	75,00	25,00	93,02	85,71	0,00	100,00	96,15	0,00
C ₁₁	100,00	0,00	95,45	75,00	0,00	0,00	0,00	0,00	100,00	0,00	88,46	57,14
C ₁₂	88,89	11,11	98,86	94,12	0,00	0,00	0,00	0,00	66,67	33,33	92,31	66,67
C ₁₃	66,67	33,33	98,86	80,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
μ	87,97	12,03	98,25	87,91	62,64	37,36	93,02	53,67	60,56	39,44	93,59	51,29
σ	12,83	12,83	1,58	9,16	39,63	39,63	3,10	32,41	41,61	41,61	3,41	34,52

Os resultados obtidos com a dimensão fractal multiescala demonstram a grande potencialidade dessa metodologia na classificação de espécies biológicas. Para as espécies de maracujá, trabalhos na literatura já indicavam a qualidade dessas características na discriminação das espécies (Plotze, Falvo *et al.*, 2005). No entanto, para as espécies de eucalipto e arroz estes são os primeiros resultados.

A Figura 7.4 apresenta todas as curvas fractal multiescala produzidas no experimento F019 com as espécies de maracujá. Em cada gráfico é possível observar comparativamente as

curvas fractal para cada uma das espécies utilizadas no experimento. A identificação das classes pode ser visualizada no canto superior direito. O eixo Y dos gráficos representa a fractalidade da forma analisada, enquanto o eixo X o número de pontos da função fractal.

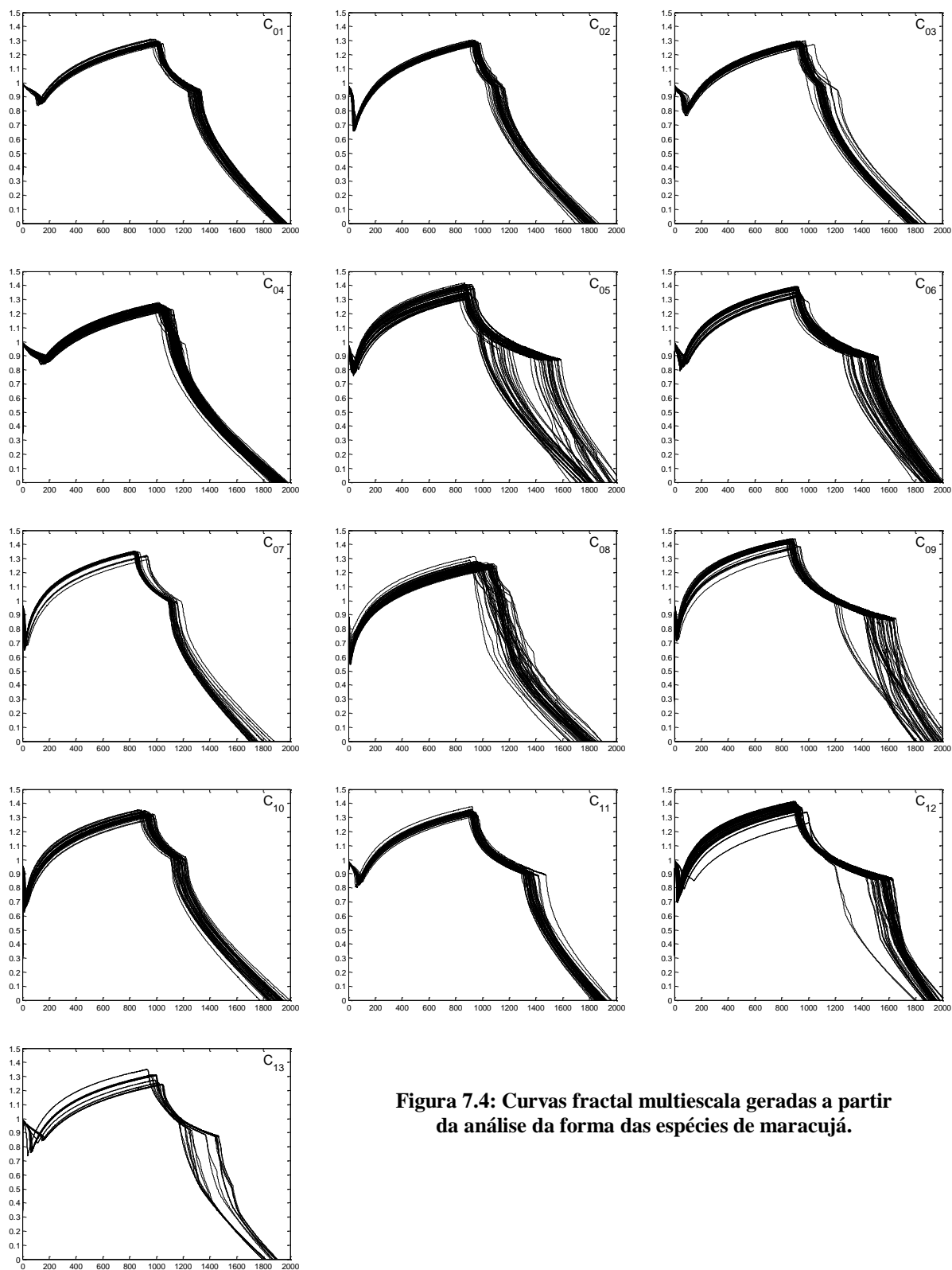


Figura 7.4: Curvas fractal multiescala geradas a partir da análise da forma das espécies de maracujá.

Para uma análise comparativa, foram gerados gráficos ROC que demonstram o processo de classificação para cada gênero: maracujá, eucalipto e arroz. Os resultados são apresentados na Figura 7.5.

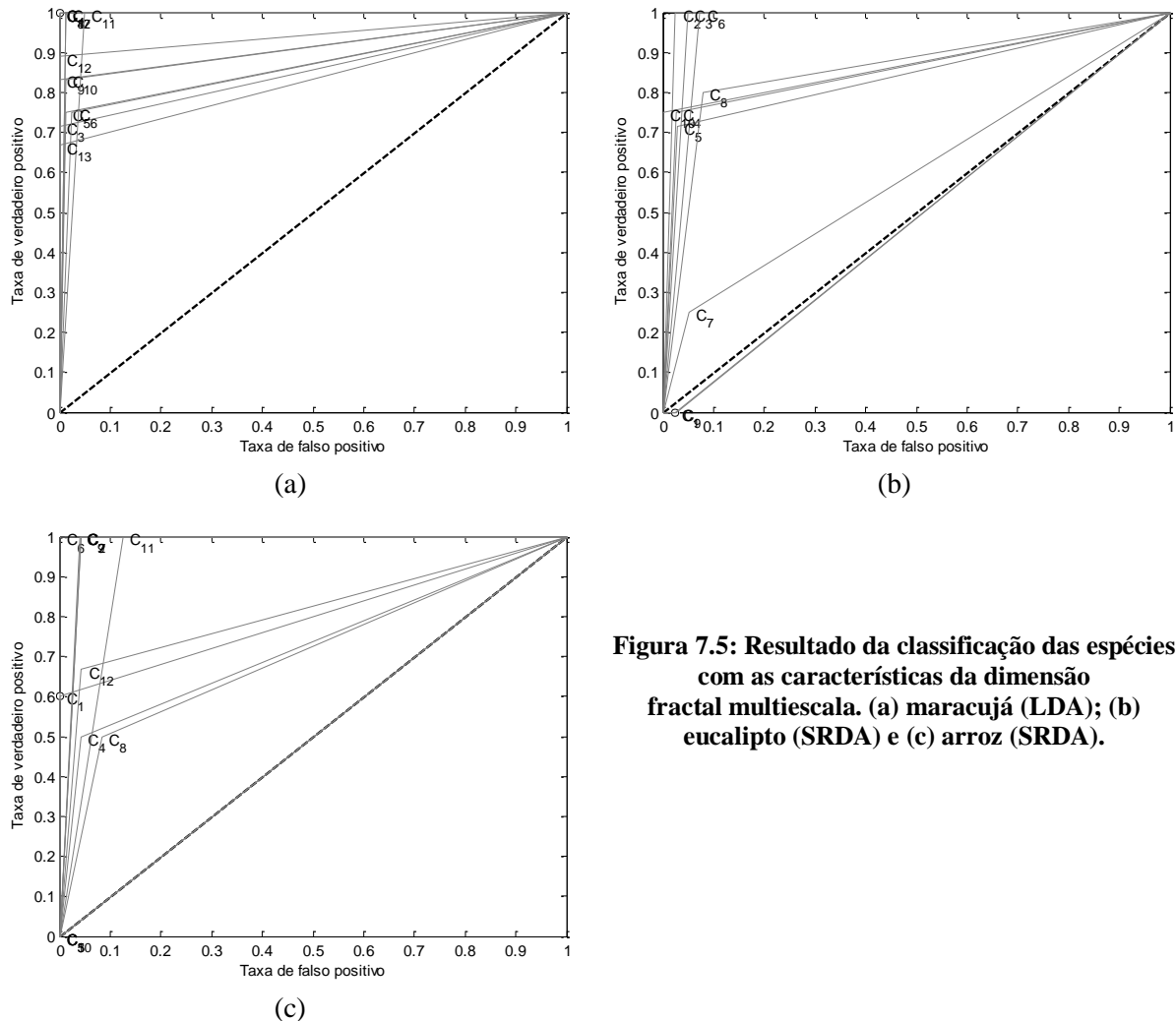


Figura 7.5: Resultado da classificação das espécies com as características da dimensão fractal multiescala. (a) maracujá (LDA); (b) eucalipto (SRDA) e (c) arroz (SRDA).

E. Resultados das Redes Complexas

Os experimentos envolvendo redes complexas produziram um grande conjunto de informações das espécies biológicas. A Tabela 7.9 apresenta os resultados obtidos com as redes complexas. Devido ao grande número de experimentos, apenas os resultados com porcentagem de acerto superior a 80% são apresentados. Além disso, é importante salientar que para as espécies de maracujá mais de 60 experimentos (de um total de 82) produziram resultados superiores a 80% de classificações corretas.

Tabela 7.9: Resultados da classificação das espécies com as características das redes complexas.

Redes Complexas		Maracujá (<i>Passiflora</i>)			Eucalipto (<i>Eucalyptus</i>)			Arroz (<i>Oryza</i>)		
		DM	LDA	SRDA	DM	LDA	SRDA	DM	LDA	SRDA
F035	%A	79,43	81,14	85,23	70,23	72,56	79,07	29,23	32,31	42,31
	%E	20,57	18,86	14,77	29,77	27,44	20,93	70,77	67,69	57,69
F040	%A	77,27	78,86	86,36	51,86	52,56	72,09	27,31	31,54	34,61
	%E	22,73	21,14	13,64	48,14	47,44	27,91	72,69	68,46	65,39
F044	%A	75,23	84,20	86,36	44,42	45,12	58,14	27,69	57,69	38,46
	%E	24,77	15,80	13,64	55,58	54,88	41,86	72,31	42,31	61,54
F054	%A	77,27	80,23	85,23	43,26	47,44	55,81	31,54	36,92	38,46
	%E	22,73	19,77	14,77	56,74	52,56	44,19	68,46	63,08	61,54
F057	%A	74,66	79,89	85,23	37,68	40,46	55,81	31,54	38,08	42,21
	%E	25,34	20,11	14,77	62,32	59,54	44,19	68,46	61,92	57,79
F061	%A	78,98	76,59	87,50	30,47	41,86	34,88	26,92	30,38	42,31
	%E	21,02	23,41	12,50	69,53	58,14	65,12	73,08	69,62	57,69
F062	%A	79,09	78,07	85,23	27,68	37,44	32,56	27,31	30,77	30,77
	%E	20,91	21,93	14,77	72,32	62,56	67,44	72,69	69,23	69,23
F068	%A	91,93	93,18	91,59	41,86	43,95	46,74	31,54	54,23	60,25
	%E	8,07	6,82	8,41	58,14	56,05	53,26	68,46	45,77	39,75
F069	%A	76,02	83,98	87,50	49,53	53,02	58,14	31,92	49,61	43,46
	%E	23,98	16,02	12,50	50,47	46,98	41,86	68,08	50,39	56,54
F070	%A	80,23	79,77	86,36	42,56	36,98	48,84	13,46	19,23	26,92
	%E	19,77	20,23	13,64	57,44	63,02	51,16	86,54	80,77	73,08
F071	%A	78,41	79,09	86,36	45,81	42,56	55,81	18,46	20,00	26,92
	%E	21,59	20,91	13,64	54,19	57,44	44,19	81,54	80,00	73,08
F109	%A	77,27	88,41	89,77	40,93	49,07	37,21	45,00	49,61	53,85
	%E	22,73	11,59	10,23	59,07	50,93	62,79	55,00	50,39	46,15

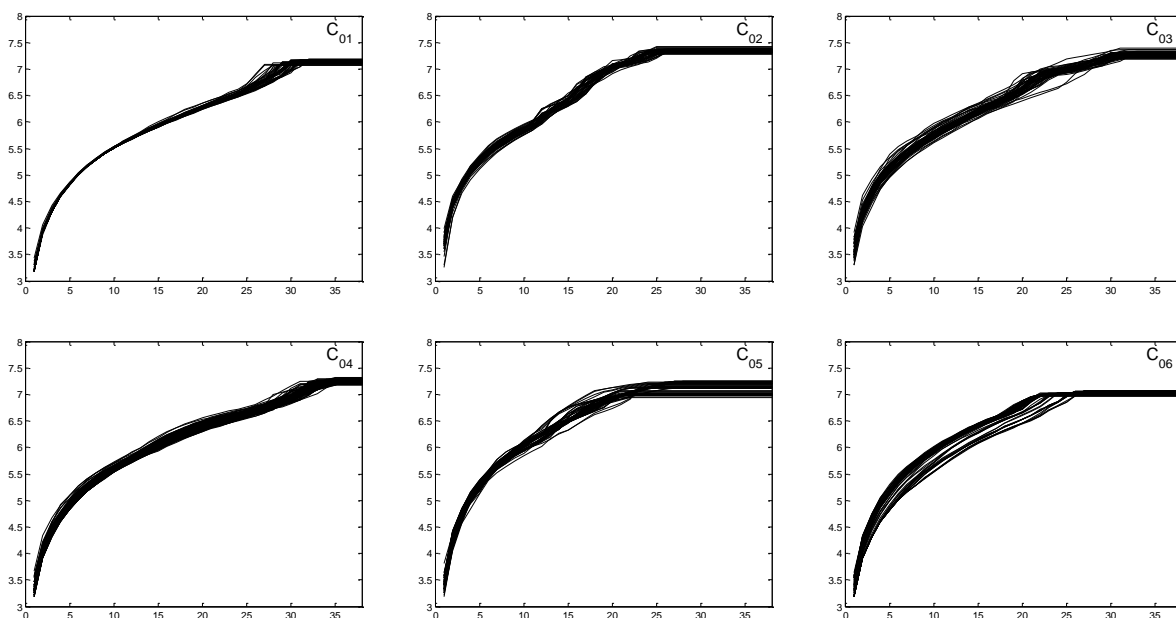
Nos experimentos com redes complexas o melhor resultado foi alcançado com as características extraídas do grau máximo da rede. Este experimento foi realizado com as espécies de maracujás. Além disso, para as espécies de eucalipto a taxa de classificações corretas foi muito satisfatória, uma vez que, foi próxima a 80% de acerto. Com as espécies de arroz o melhor resultado alcançado foi de 60,25% de acerto com as características do grau máximo da rede

A Tabela 7.10 apresenta detalhadamente os melhores resultados obtidos com as redes complexas. É importante observar a potencialidade deste descritor para a classificação das espécies de maracujás, uma vez que os resultados superaram 90% de discriminações corretas.

Tabela 7.10: Detalhamento dos resultados da classificação das espécies com as redes complexas.

	Maracujá (<i>Passiflora</i>)				Eucalipto (<i>Eucalyptus</i>)				Arroz (<i>Oryza</i>)			
	%A	%E	AC	F	%A	%E	AC	F	%A	%E	AC	F
C ₀₁	100,00	0,00	98,86	92,31	100,00	0,00	97,67	66,67	40,00	60,00	80,77	44,44
C ₀₂	100,00	0,00	98,86	92,31	50,00	50,00	97,67	66,67	50,00	50,00	96,15	66,67
C ₀₃	100,00	0,00	100,00	100,00	100,00	0,00	100,00	100,00	100,00	0,00	88,46	57,14
C ₀₄	100,00	0,00	100,00	100,00	100,00	0,00	100,00	100,00	50,00	50,00	96,15	66,67
C ₀₅	75,00	25,00	96,59	80,00	85,71	14,29	97,67	92,31	0,00	100,00	84,62	0,00
C ₀₆	87,50	12,50	97,73	87,50	100,00	0,00	93,02	57,14	100,00	0,00	100,00	100,00
C ₀₇	100,00	0,00	100,00	100,00	50,00	50,00	93,02	57,14	0,00	100,00	92,31	0,00
C ₀₈	85,71	14,29	98,86	92,31	100,00	0,00	90,70	71,43	100,00	0,00	100,00	100,00
C ₀₉	83,33	16,67	98,86	90,91	0,00	100,00	90,70	0,00	0,00	100,00	96,15	0,00
C ₁₀	100,00	0,00	97,73	85,71	91,67	8,33	97,67	95,65	100,00	0,00	96,15	66,67
C ₁₁	100,00	0,00	100,00	100,00	-	-	-	-	100,00	0,00	100,00	100,00
C ₁₂	100,00	0,00	100,00	100,00	-	-	-	-	100,00	0,00	92,31	75,00
C ₁₃	66,67	33,33	98,86	80,00	-	-	-	-	-	-	-	-
μ	92,17	7,83	98,95	92,39	77,74	22,26	95,81	70,70	61,67	38,33	93,59	56,38
σ	11,46	11,46	1,08	7,46	33,86	33,86	3,60	30,19	43,66	43,66	6,21	38,20

Na Figura 7.6 é ilustrada uma análise comparativa das funções (em escala logarítmica) geradas a partir do grau máximo extraído das redes complexas. Os gráficos foram produzidos com os dados do experimento F068, e em cada representação são agrupadas as funções de cada espécie. O identificador das espécies é apresentado no canto superior direito.



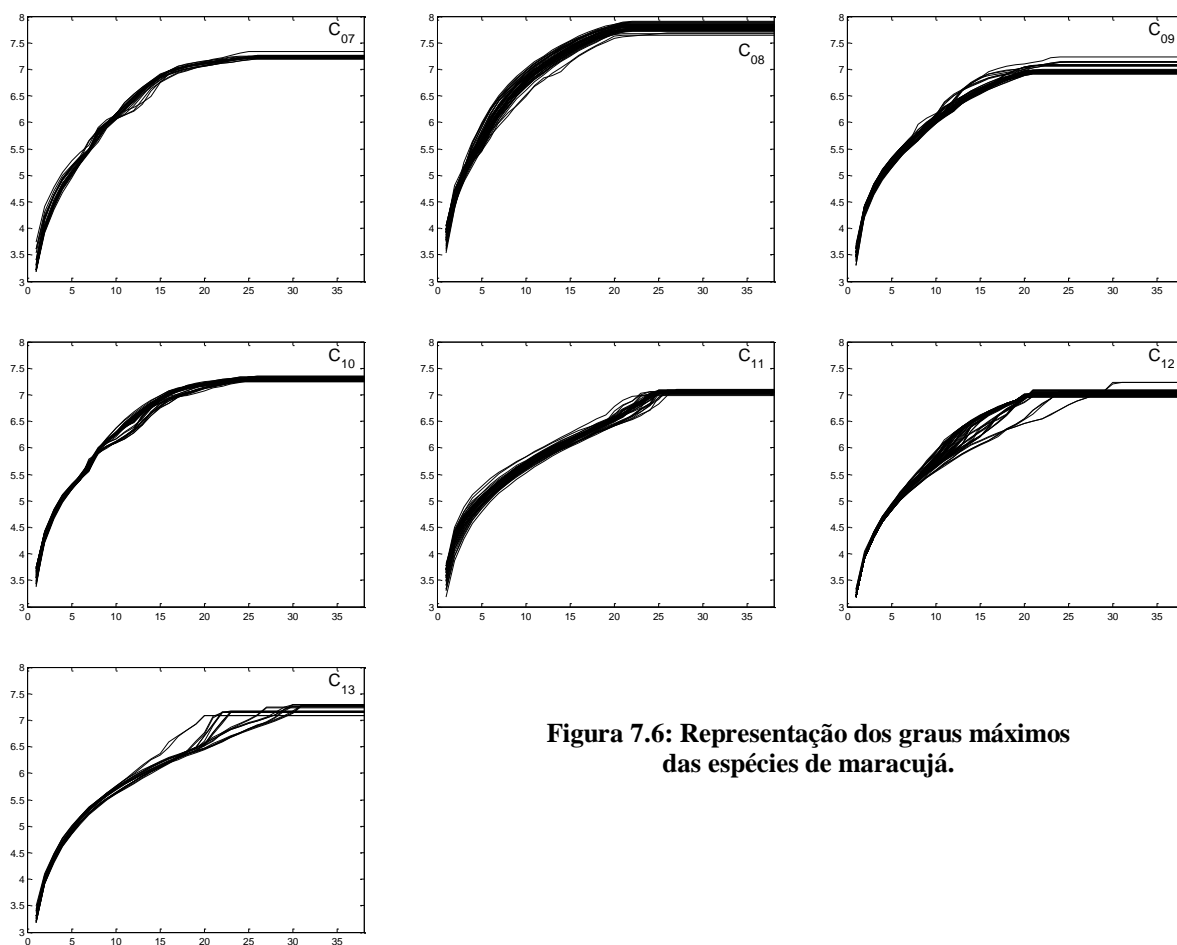


Figura 7.6: Representação dos graus máximos das espécies de maracujá.

Na Figura 7.7 é ilustrado comparativamente o desempenho dos classificadores para as espécies de maracujá, eucalipto e arroz. O gráfico ROC esboça as taxas de verdadeiro positivo e falso positivo dos melhores experimentos. Na Figura é importante observar a potencialidade das redes complexas na discriminação de espécies de eucalipto. Apesar da dificuldade de classificação de espécies puras e híbridas, os resultados sinalizam a viabilidade dessa metodologia na classificação de imagens de eucaliptos.

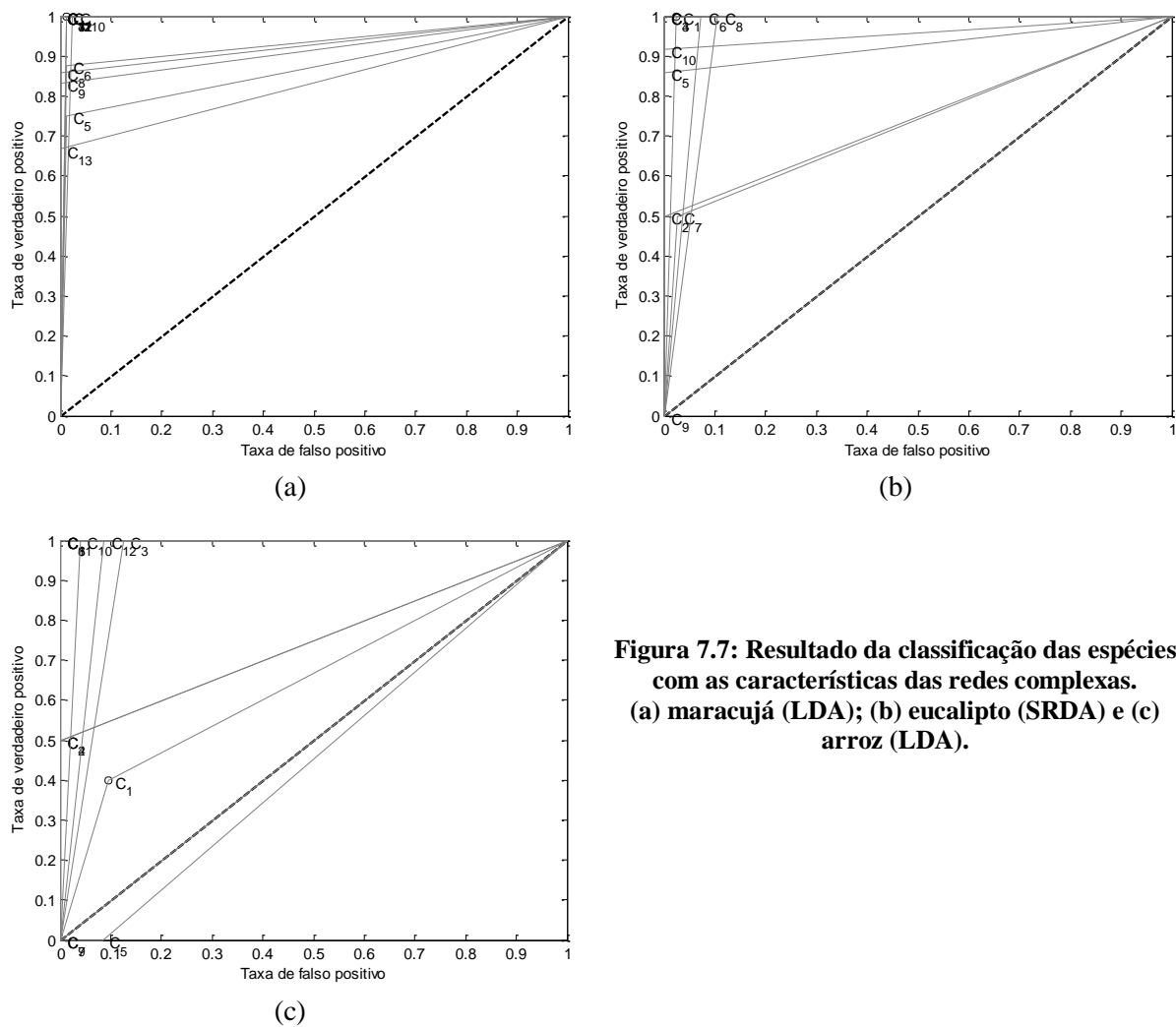


Figura 7.7: Resultado da classificação das espécies com as características das redes complexas. (a) maracujá (LDA); (b) eucalipto (SRDA) e (c) arroz (LDA).

F. Resultados das Wavelets

Os experimentos com a transformada de *wavelets* consideraram as famílias *Daubechies* e *Coiflets* para extração de características das espécies. Estes experimentos produziram vetores de características a partir dos coeficientes de tendência e detalhe.

Na Tabela 7.11 são apresentadas as porcentagens médias de acerto e erro para cada família. Com os resultados é possível concluir que os coeficientes de tendência são em média mais adequados para classificação das espécies. Estes resultados consideram os dados dos experimentos com maracujá, eucalipto e arroz.

Tabela 7.11: Resultados da classificação das espécies com as características extraídas das *wavelets*.

Família	Total de Vetores de Características	Coefficiente	%A (média)	%E (média)
<i>Coiflets</i>	50	tendência	84,78 ± 1.8516	15,22
		detalhe	47,06 ± 14.8110	52,94
<i>Daubechies</i>	450	tendência	85,45 ± 3.0027	14,55
		detalhe	44,49 ± 15.0227	55,51

Na Tabela 7.12 são apresentados os melhores experimentos selecionados dentre todos os resultados obtidos com a transformada de *wavelets*. Nos resultados são descritos tanto experimentos envolvendo coeficientes de tendência, quanto de detalhe.

Tabela 7.12: Resultados da classificação com as características das *wavelets*.

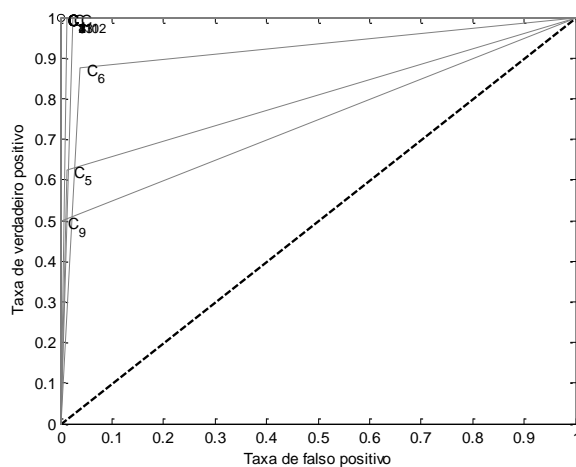
<i>Wavelets</i>		Maracujá (<i>Passiflora</i>)			Eucalipto (<i>Eucalyptus</i>)			Arroz (<i>Oryza</i>)		
		DM	LDA	SRDA	DM	LDA	SRDA	DM	LDA	SRDA
F164	%A	80,78	86,55	91,44	44,19	62,09	72,33	18,46	31,29	32,27
	%E	19,22	13,45	8,56	55,81	37,91	27,67	81,54	68,71	67,73
F487	%A	74,65	83,95	91,80	44,26	64,65	68,27	18,46	33,46	30,77
	%E	25,35	16,05	8,20	55,74	35,35	31,73	81,54	66,54	69,23
F497	%A	73,07	91,96	85,99	43,02	65,35	73,95	18,46	31,54	26,92
	%E	26,93	8,04	14,01	56,98	34,65	26,05	81,54	68,46	73,08
F499	%A	75,02	85,08	92,40	50,93	59,30	69,53	19,23	36,25	34,61
	%E	24,98	14,92	7,60	49,07	40,70	30,47	80,77	63,75	65,39
F507	%A	76,12	92,88	87,10	43,26	67,21	66,98	20,00	50,00	30,77
	%E	23,88	7,12	12,90	56,74	32,79	33,02	80,00	50,00	69,23
F539	%A	76,02	86,95	92,75	46,05	64,42	66,51	19,62	33,08	30,77
	%E	23,98	13,05	7,25	53,95	35,58	33,49	80,38	66,92	69,23
F541	%A	75,50	86,02	91,87	51,16	60,47	69,77	19,23	36,00	34,61
	%E	24,50	13,98	8,13	48,84	39,53	30,23	80,77	64,00	65,39

Na Tabela 7.13 são detalhados resultados selecionados da Tabela 7.12. É importante observar a alta taxa de acerto na classificação de espécies de maracujás, sendo superior a 90%. Os resultados com as espécies de eucalipto também foram promissores, alcançando uma taxa superior a 70%. Para as espécies de arroz o desempenho foi mediano, uma vez que obteve uma taxa de acerto igual a 50%.

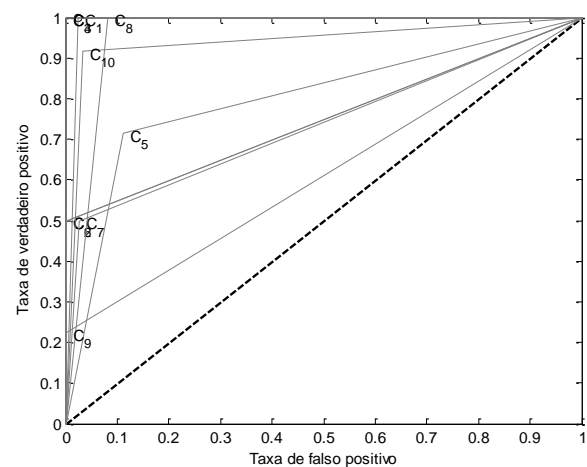
Tabela 7.13: Detalhamento dos resultados da classificação das espécies com as wavelets.

	Maracujá (<i>Passiflora</i>)				Eucalipto (<i>Eucalyptus</i>)				Arroz (<i>Oryza</i>)			
	%A	%E	AC	F	%A	%E	AC	F	%A	%E	AC	F
C ₀₁	100,00	0,00	100,00	100,00	100,00	0,00	97,67	66,67	40,00	60,00	76,92	40,00
C ₀₂	100,00	0,00	100,00	100,00	50,00	50,00	97,67	66,67	100,00	0,00	100,00	100,00
C ₀₃	100,00	0,00	100,00	100,00	100,00	0,00	100,00	100,00	0,00	100,00	88,46	0,00
C ₀₄	100,00	0,00	100,00	100,00	100,00	0,00	100,00	100,00	50,00	50,00	88,46	40,00
C ₀₅	62,50	37,50	95,45	71,43	71,43	28,57	86,05	62,50	0,00	100,00	88,46	0,00
C ₀₆	87,50	12,50	95,45	77,78	50,00	50,00	97,67	66,67	100,00	0,00	100,00	100,00
C ₀₇	100,00	0,00	100,00	100,00	50,00	50,00	93,02	57,14	100,00	0,00	96,15	80,00
C ₀₈	100,00	0,00	100,00	100,00	100,00	0,00	92,79	76,37	50,00	50,00	92,31	50,00
C ₀₉	50,00	50,00	96,59	66,67	22,50	77,50	92,79	0,00	100,00	0,00	96,15	66,67
C ₁₀	100,00	0,00	98,86	92,31	91,67	8,33	95,35	91,67	0,00	100,00	96,15	0,00
C ₁₁	100,00	0,00	100,00	100,00	0,00	0,00	0,00	0,00	50,00	50,00	84,62	33,33
C ₁₂	100,00	0,00	97,73	90,00	0,00	0,00	0,00	0,00	33,33	66,67	92,31	50,00
C ₁₃	100,00	0,00	100,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
μ	92,31	7,69	98,78	92,17	73,56	26,44	95,30	68,77	51,94	48,06	91,67	46,67
σ	16,57	16,57	1,82	12,19	28,68	28,68	4,28	28,77	40,26	40,26	6,73	35,70

A Figura 7.8 ilustra os gráficos ROC gerado a partir das saídas dos classificadores. No gráfico da Figura 7.8a é possível notar a grande potencialidade das características da transforma de *wavelets* na classificação das espécies de maracujá. Na Figura 7.8b apenas a espécie C₀₉ obteve um desempenho mediano, enquanto na Figura 7.8c algumas espécies tiveram desempenho abaixo da linha diagonal.



(a)



(b)

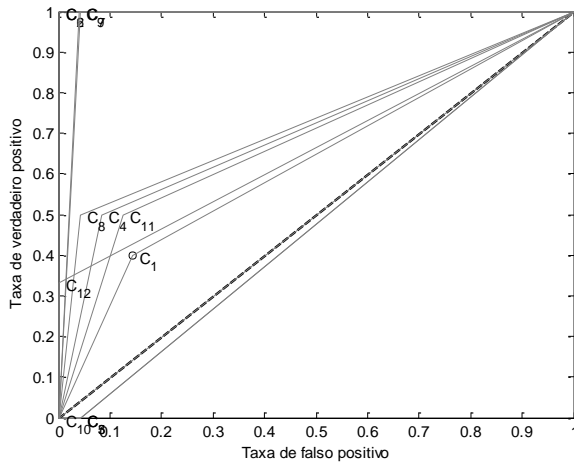


Figura 7.8: Resultado da classificação das espécies com as características da transformada de wavelets. (a) maracujá (LDA); (b) eucalipto (SRDA) e (c) arroz (LDA).

(c)

G. Seleção de Características

Para seleção das características extraídas por meio das técnicas de análise formas foi utilizada uma abordagem baseada na variância total resultante da Análise de Componentes Principais. A técnica de PCA foi aplicada nos conjuntos de dados obtidos em cada experimento, e em seguida, a variância total foi calculada. A variância tem importância fundamental no processo de classificação de padrões, pois em geral é mais fácil distinguir padrões em que seus vetores não são correlacionados e que apontam para a direção de maior variância dos dados (Campos, 2001).

Utilizando a abordagem descrita para seleção das características, dois novos conjuntos de dados foram gerados contendo as características (dentre todos os conjuntos) com variância superior a: (i) 85% - identificado como S001 e (ii) 95% - identificado como S002. Essas características foram classificadas utilizando os classificadores já adotados (DM, LDA e SRDA), além das redes neurais artificiais. A Tabela 7.14 apresenta os resultados do processo de classificação das características.

Tabela 7.14: Resultados da classificação das espécies com a seleção de características.

Seleção de características		Maracujá (<i>Passiflora</i>)				Eucalipto (<i>Eucalyptus</i>)				Arroz (<i>Oryza</i>)			
		DM	LDA	SRDA	RNA	DM	LDA	SRDA	RNA	DM	LDA	SRDA	RNA
S001	%A	73,07	88,07	87,27	53,76	46,28	49,30	61,16	31,75	23,46	53,46	42,31	20,34
	%E	26,93	11,93	12,73	46,24	53,72	50,70	38,84	68,25	76,54	46,54	57,69	79,66
S002	%A	74,89	72,84	67,95	50,48	40,00	51,86	47,21	21,93	27,31	54,62	44,62	31,81
	%E	25,11	27,16	32,05	49,52	60,00	48,14	52,79	78,07	72,69	45,39	55,38	68,19

No experimento S001 o número de características selecionadas foi: 20 para maracujás, 259 para eucalipto e 22 para arroz. No experimento S002, foram selecionadas: 12 características das espécies de maracujá, 254 das espécies de eucalipto e 8 das espécies de arroz. A Figura 7.9 ilustra comparativamente as curvas ROC para os melhores resultados obtidos para cada gênero. É importante notar que a redução de dimensionalidade dos conjuntos de dados, principalmente de maracujá, não prejudicou a qualidade da classificação das espécies.

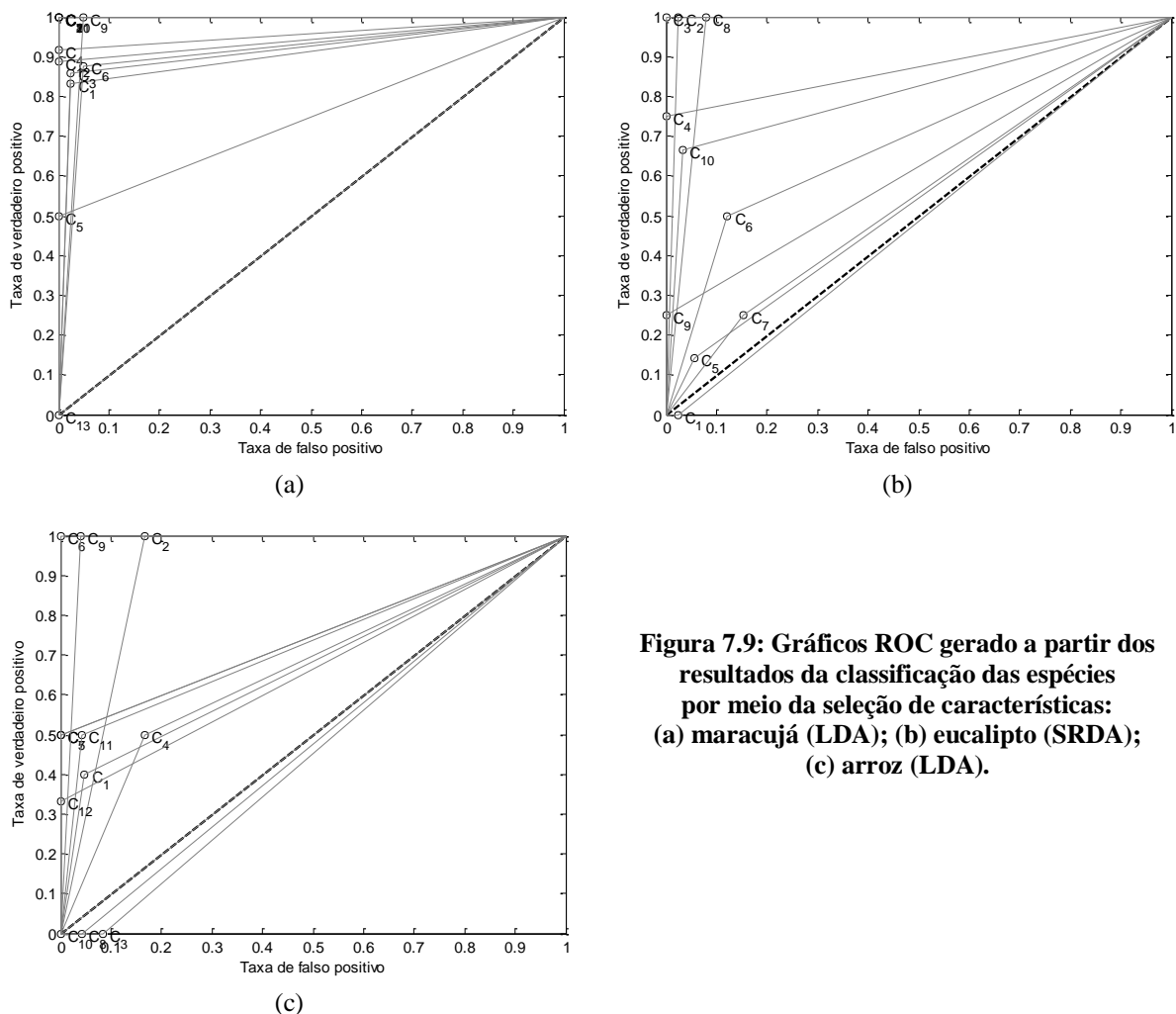


Figura 7.9: Gráficos ROC gerado a partir dos resultados da classificação das espécies por meio da seleção de características: (a) maracujá (LDA); (b) eucalipto (SRDA); (c) arroz (LDA).

Os resultados obtidos com os dados das redes complexas demonstraram grande potencialidade na discriminação das espécies. Com isso, o método de seleção de características foi utilizado especificamente nesses dados, e dois novos conjuntos foram gerados: (i) contendo características (dentre todos os dados gerados pelas redes complexas)

com variância superior a 95% - identificado como S003 e (ii) contendo a primeira componente principal (PC1) dos dados sobre os graus das redes (S004).

Esses conjuntos foram classificados utilizando os classificadores tradicionais (DM, LDA e SRDA) e uma rede neural artificial. Os resultados são apresentados na Tabela 7.15.

Tabela 7.15: Resultados da classificação das espécies com os dados selecionados dos conjuntos das redes complexas

Seleção de características		Maracujá (<i>Passiflora</i>)				Eucalipto (<i>Eucalyptus</i>)				Arroz (<i>Oryza</i>)			
		DM	LDA	SRDA	RNA	DM	LDA	SRDA	RNA	DM	LDA	SRDA	RNA
S003	%A	75,68	74,43	67,95	50,48	30,93	40,70	42,56	27,09	23,85	56,15	44,62	31,81
	%E	24,32	25,57	32,05	49,52	69,07	59,30	57,44	72,91	76,15	43,85	55,38	68,19
S004	%A	70,80	74,43	78,18	45,02	20,47	22,33	21,40	32,95	20,39	34,62	38,85	15,54
	%E	29,21	25,57	21,82	54,98	79,54	77,67	78,61	67,05	79,62	65,39	61,15	84,46

A Figura 7.10 apresenta os gráficos ROC para os resultados selecionados de cada gênero. Para as espécies de maracujás, o melhor resultado foi obtido com o experimento S004 e o classificador SRDA, no qual foram utilizadas 12 características. O melhor resultado das espécies de eucalipto foi para o classificador SRDA e o experimento S003, com 212 características. E, para as espécies de arroz o melhor resultado foi alcançado com o experimento S003 com o classificador LDA com 18 características.

Os resultados obtidos com a seleção de características aplicada aos dados das técnicas de análise de formas demonstram a importância da mineração do conjunto de dados. Os experimentos S001 e S002 alcançaram taxas satisfatórias de classificações corretas, mesmo utilizando um conjunto de dados reduzido (selecionado). Para a seleção de características a partir do conjunto das redes complexas (S003 e S004), os resultados foram medianos, tendo em vista que as classificações foram inferiores a 80%.

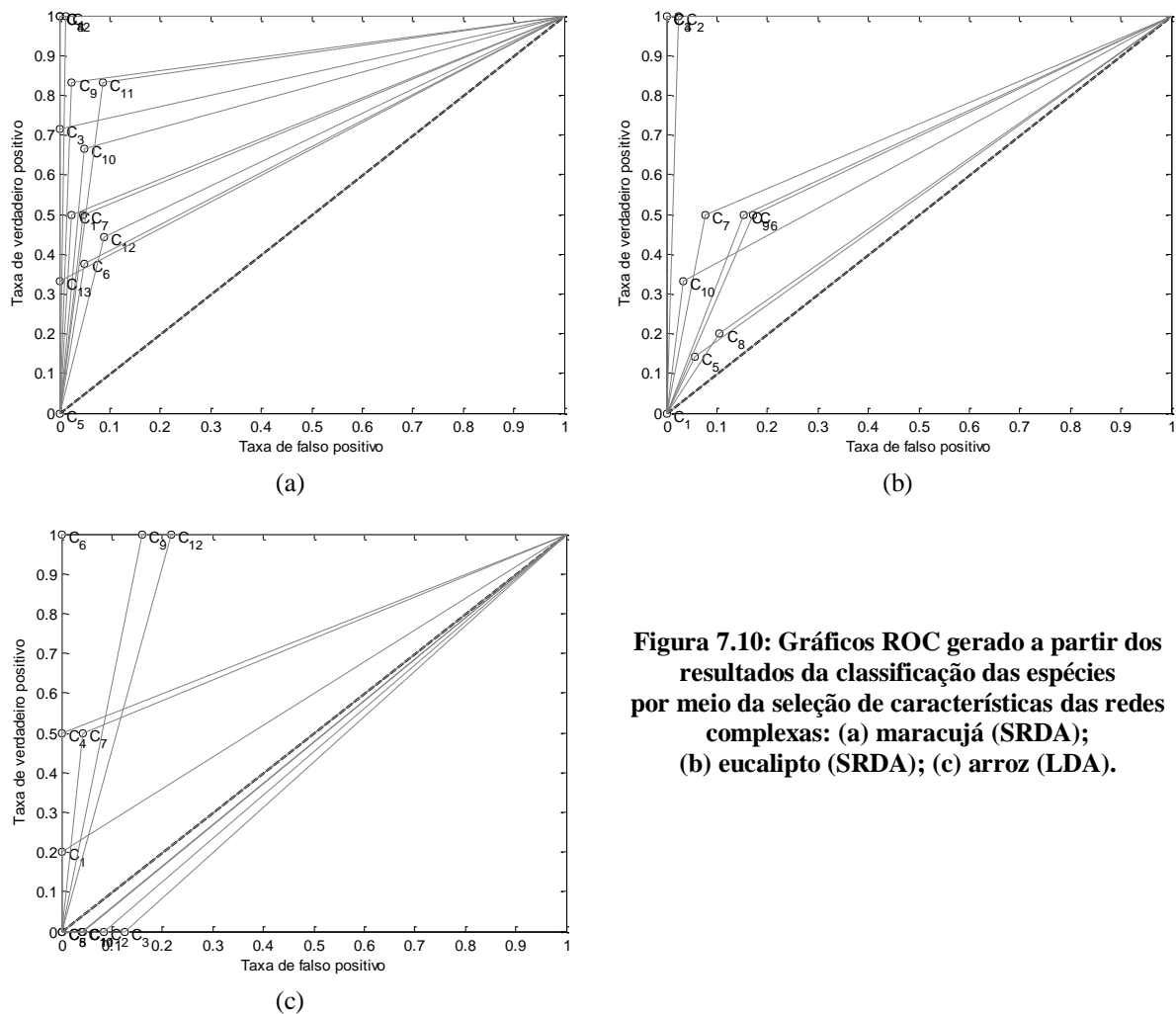


Figura 7.10: Gráficos ROC gerado a partir dos resultados da classificação das espécies por meio da seleção de características das redes complexas: (a) maracujá (SRDA); (b) eucalipto (SRDA); (c) arroz (LDA).

7.2 Resultados da Classificação baseada na Análise de Estruturas Tubulares

As técnicas de análise de estruturas tubulares foram utilizadas exclusivamente nas espécies de maracujás silvestres. Essas técnicas possibilitaram o estudo do sistema de venação das espécies, e os seguintes métodos foram utilizados: (a) biometria; (b) assinatura fractal e (c) redes complexas. Por fim, a técnica de seleção de características (d) foi aplicada aos conjuntos para a escolha das melhores características.

Nestes experimentos foram utilizadas 11 espécies de maracujás do gênero *Passiflora*, contendo quatro exemplares por espécie. As espécies são identificadas por C_{01} – *P.actinia*, C_{02} – *P.amethystina*, C_{03} – *P.caerulea*, C_{04} – *P.coriácea*, C_{05} – *P.foetida*, C_{06} – *P.gibertii*,

C_{07} – *P.maliformis*, C_{08} - *P.miersii*, C_{09} – *P.organensis*, C_{10} – *P.pohlii* e C_{11} – *P.suberosa*. Os resultados da classificação das espécies de maracujás são apresentados como se segue:

A. Biometria

Na Tabela 7.16 são descritos os resultados da classificação das espécies de maracujás por meio das características biométricas. Para cada classificador, são apresentadas as porcentagens de acerto e erro. O experimento para extração de medidas biométricas foi identificado como E001.

Tabela 7.16: Resultados da classificação das espécies com os dados biométricos.

Biometria		Maracujá (<i>Passiflora</i>)			
		DM	LDA	SRDA	RNA
E001	%A	87,60	90,08	83,47	36,14
	%E	12,40	9,92	16,53	63,86

Este experimento obteve um resultado muito promissor, pois atingiu uma taxa de acerto superior a 90%. Além disso, as características biométricas extraídas do sistema de venação foram definidas a partir de rigorosos protocolos de taxonomia (Wing, Wilf *et al.*, 2001). Este fato demonstra a aplicabilidade de técnicas de visão artificial no auxílio de profissionais durante a taxonomia das espécies. A Tabela 7.17 ilustra a matriz de confusão e suas respectivas informações estatísticas do classificador LDA.

Tabela 7.17: Matriz de confusão e dados estatísticos resultantes do experimento com dados biométricos e o classificador LDA.

	C ₀₁	C ₀₂	C ₀₃	C ₀₄	C ₀₅	C ₀₆	C ₀₇	C ₀₈	C ₀₉	C ₁₀	C ₁₁	%A	%E
C ₀₁	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₂	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	100,0
C ₀₃	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₄	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₅	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₆	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₇	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₈	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	100,0	0,0
C ₀₉	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	100,0	0,0
C ₁₀	0,0	0,0	0,0	0,0	0,0	0,0	0,0	7,7	0,0	92,3	0,0	92,3	7,7
C ₁₁	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	100,0	0,0
											μ	90,1	9,9

AC = 98,21% ±0,038 T_{VP} = 90,21% ±0,300 T_{FP} = 0,99% ±0,031 ES = 99,01% ±0,031 PR = 95,02% ±0,165 F = 93,12% ±0,153

O resultado da classificação das espécies a partir do classificador LDA pode ser visualizado no gráfico ROC ilustrado na Figura 7.11. Nos resultados é possível observar que a classe 10 não obteve uma boa classificação, no entanto, as demais classes foram classificadas corretamente, as quais estão agrupadas próximas à coordenada (0,1).

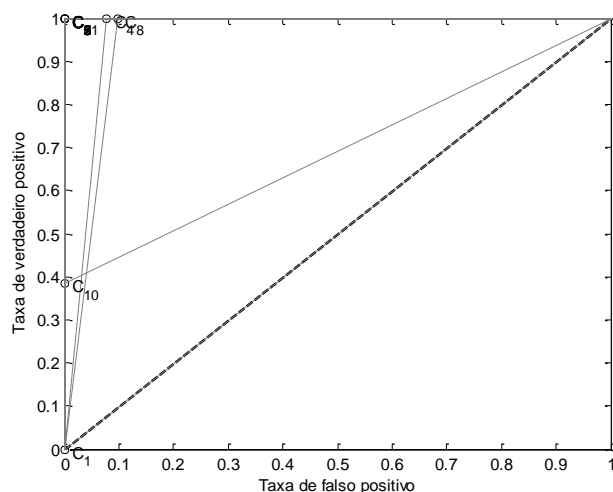


Figura 7.11: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados biométricos.

B. Assinatura Fractal Multiescala

A Tabela 7.18 apresenta os resultados obtidos com a assinatura fractal das estruturas tubulares. Os experimentos foram realizados com base em todos os pixels da estrutura, apenas as bifurcações, apenas os pontos extremos e combinando as bifurcações e os pontos extremos. Esses experimentos foram identificados de E002 até E049, no entanto, a Tabela apresenta apenas alguns resultados selecionados.

Tabela 7.18: Resultados da classificação das espécies com a assinatura fractal multiescala.

Assinatura Fractal	Maracujá (<i>Passiflora</i>)				
	DM	LDA	SRDA	RNA	
E005	%A	58,68	92,56	72,45	25,23
	%E	41,32	7,44	27,55	74,77
E008	%A	60,32	91,31	74,53	26,82
	%E	39,68	8,69	25,47	73,18
E026	%A	58,68	94,22	71,83	33,64
	%E	41,32	5,78	28,17	66,36
E028	%A	59,50	93,39	74,66	22,05
	%E	40,50	6,61	25,34	77,95
E042	%A	57,85	92,56	70,54	27,05
	%E	42,15	7,44	29,46	72,95
E046	%A	59,68	91,74	71,22	31,36
	%E	40,32	8,26	28,78	68,64

No experimento E026 é possível observar o melhor resultado obtido com a classificação das espécies, com taxa de acerto superior a 94%. Neste experimento a assinatura fractal multiescala foi gerada a partir dos pontos extremos das estruturas tubulares das espécies. Os parâmetros adotados para geração da curva fractal foram $k = 100$ e $\sigma = 100$. A Tabela 7.19 apresenta a matriz de confusão e informações estatísticas a respeito do experimento E026.

Tabela 7.19: Matriz de confusão e dados estatísticos resultantes do experimento com a assinatura fractal multiescala.

	C ₀₁	C ₀₂	C ₀₃	C ₀₄	C ₀₅	C ₀₆	C ₀₇	C ₀₈	C ₀₉	C ₁₀	C ₁₁	%A	%E
C ₀₁	61,4	0,0	0,0	0,0	0,0	0,0	0,0	38,6	0,0	0,0	0,0	61,4	38,6
C ₀₂	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₃	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₄	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₅	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₆	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₇	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₈	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	100,0	0,0
C ₀₉	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	100,0	0,0
C ₁₀	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	100,0	0,0
C ₁₁	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	100,0	0,0
											μ	94,2	5,8

AC = 98,95% ±0,023 T_{VP} = 96,49% ±0,116 T_{FP} = 0,57% ±0,019 ES = 99,43% ±0,019 PR = 96,32% ±0,122 F = 95,52% ±0,100

Na Figura 7.12 são apresentados os gráficos das assinaturas fractal multiescala para cada uma das 11 espécies utilizadas nos experimentos. No canto superior direito é possível observar a identificação de cada espécie. É importante observar o comportamento padronizado das curvas em relação às espécies. Este fato indica que a assinatura fractal multiescala é uma abordagem potencialmente discriminante, uma vez, que visualmente é possível identificar as espécies. Além disso, a assinatura representa a redução de dimensionalidade da estrutura tubular, originalmente bidimensional, para um sinal unidimensional. No eixo Y dos gráficos são representadas as fractalidades, enquanto no eixo X o número de pontos da assinatura fractal.

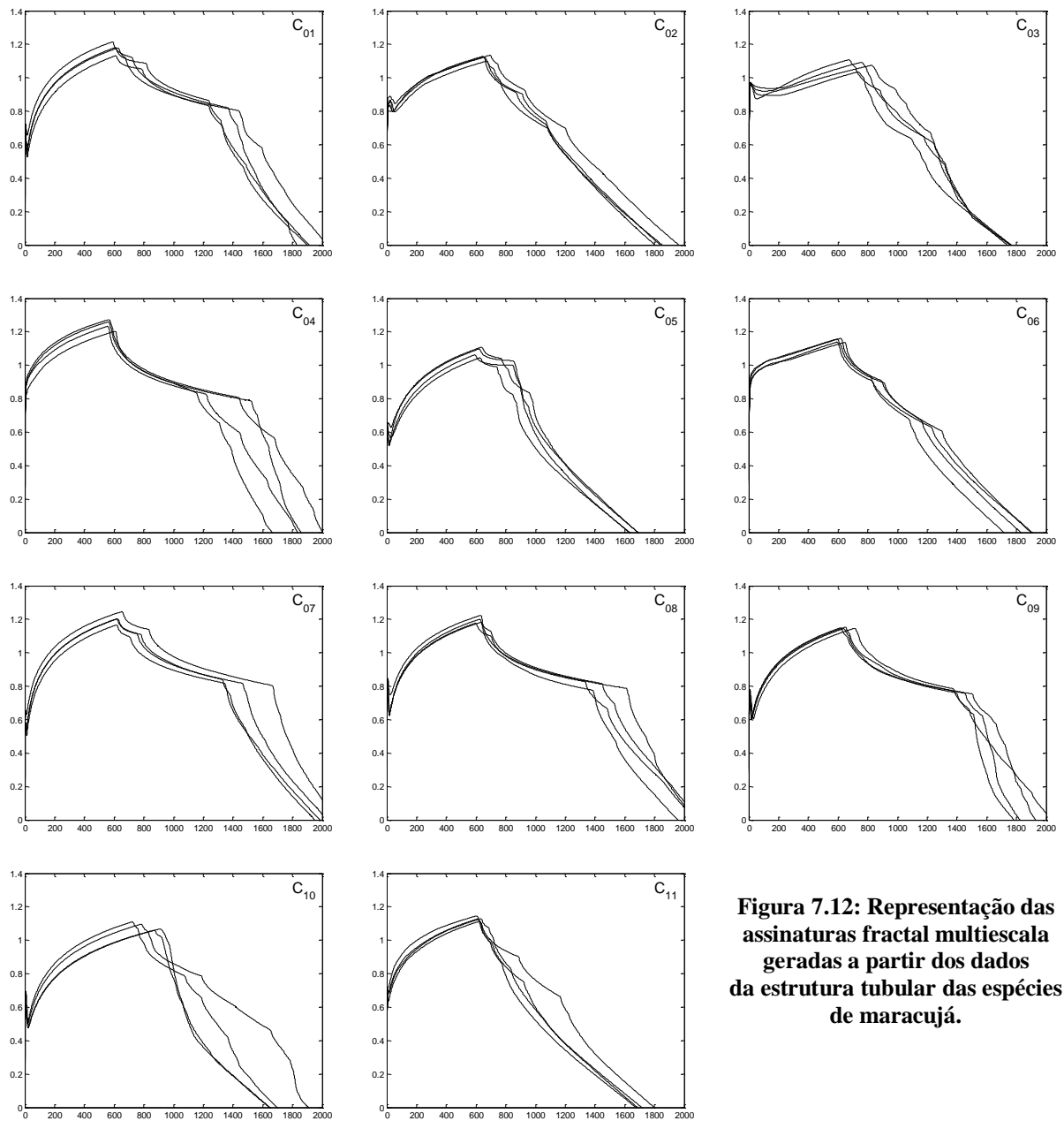


Figura 7.12: Representação das assinaturas fractal multiescala geradas a partir dos dados da estrutura tubular das espécies de maracujá.

A Figura 7.13 apresenta o gráfico ROC produzido a partir da matriz de confusão do experimento E026. No gráfico é possível observar a alta taxa de classificações corretas, uma vez que a maioria das classes está agrupada próximo ao ponto (0,1).

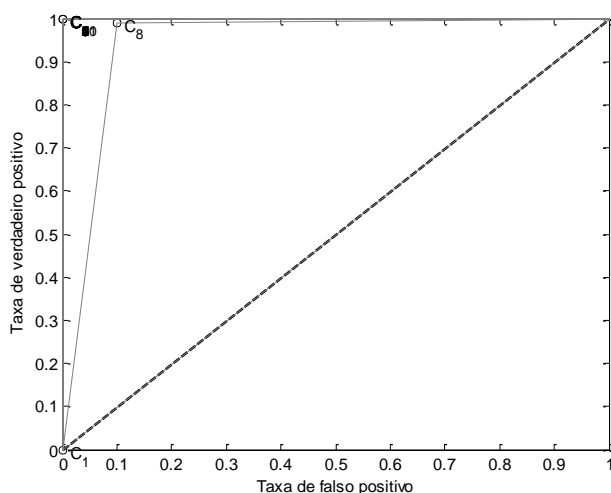


Figura 7.13: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados da assinatura fractal multiescala.

C. Redes Complexas

Os experimentos envolvendo redes complexas e estruturas tubulares consideraram como informação de interesse todos os pixels do sistema de venação, apenas as bifurcações, apenas os pontos extremos e a combinação entre bifurcações e pontos extremos. Com a metodologia de redes complexas foi possível extrair um conjunto vasto de dados a respeito das estruturas tubulares. Os resultados apresentados evidenciam as melhores porcentagens de acerto, e são representados na Tabela 7.20. Os experimentos com redes complexas foram identificados de E050 até E377. Na Tabela os resultados estão ordenados pelo identificador do experimento.

Tabela 7.20: Resultados da classificação das espécies com as redes complexas.

Redes Complexas		Maracujá (<i>Passiflora</i>)			
		DM	LDA	SRDA	RNA
E087	%A	48,76	80,99	91,33	17,95
	%E	51,24	19,01	8,67	82,05
E008	%A	61,16	85,95	91,56	23,86
	%E	38,84	14,05	7,44	76,14
E131	%A	79,34	91,74	90,22	29,32
	%E	20,66	8,26	9,78	70,68
E147	%A	48,76	89,22	90,12	28,64
	%E	51,24	10,78	9,88	71,36
E235	%A	79,34	76,86	89,61	32,73
	%E	20,66	23,14	10,39	67,27
E237	%A	67,77	71,90	90,12	44,32
	%E	32,23	28,10	9,88	55,68
E297	%A	38,02	43,80	90,85	30,68
	%E	61,98	56,20	9,15	69,32
E303	%A	50,41	66,94	90,17	36,36
	%E	49,59	33,06	9,83	63,64

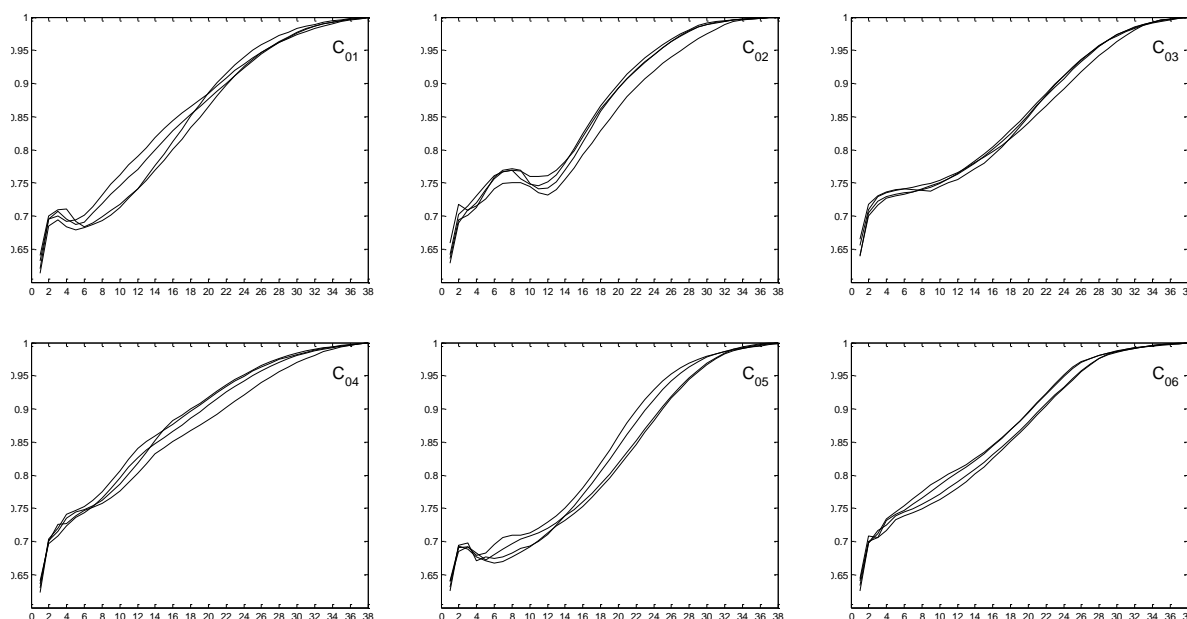
O experimento E131 foi realizado com base nas informações coletadas a partir de todos os pontos da estrutura tubular. O resultado foi alçando a partir do coeficiente de aglomeração da rede complexa e o classificador LDA. A Tabela 7.21 apresenta detalhadamente os resultados obtidos com a matriz de confusão.

Tabela 7.21: Matriz de confusão do experimento E131 com redes complexas e estruturas tubulares.

	C ₀₁	C ₀₂	C ₀₃	C ₀₄	C ₀₅	C ₀₆	C ₀₇	C ₀₈	C ₀₉	C ₁₀	C ₁₁	%A	%E
C ₀₁	97,7	0,0	0,0	0,0	0,0	0,0	2,3	0,0	0,0	0,0	0,0	97,7	2,3
C ₀₂	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₃	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₄	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₅	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₆	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₇	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₈	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	100,0	0,0
C ₀₉	0,0	83,3	0,0	0,0	0,0	0,0	0,0	0,0	16,7	0,0	0,0	16,7	83,3
C ₁₀	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	100,0	0,0
C ₁₁	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	100,0	0,0
											μ	91,7	8,3

AC = 98,50% $\pm 0,031$ $T_{VP} = 92,22\% \pm 0,251$ $T_{FP} = 0,81\% \pm 0,025$ ES = 99,19% $\pm 0,025$ PR = 94,94% $\pm 0,152$ F = 90,11% $\pm 0,227$

A Figura 7.14 ilustra graficamente os coeficientes de aglomeração obtidos para cada uma das espécies de maracujá. Os gráficos foram gerados a partir do experimento E131. É importante observar na Figura o comportamento das curvas geradas a partir dos dados das redes complexas. Os primeiros 10 pontos da curva representam uma assinatura com alto grau de discriminação das espécies.



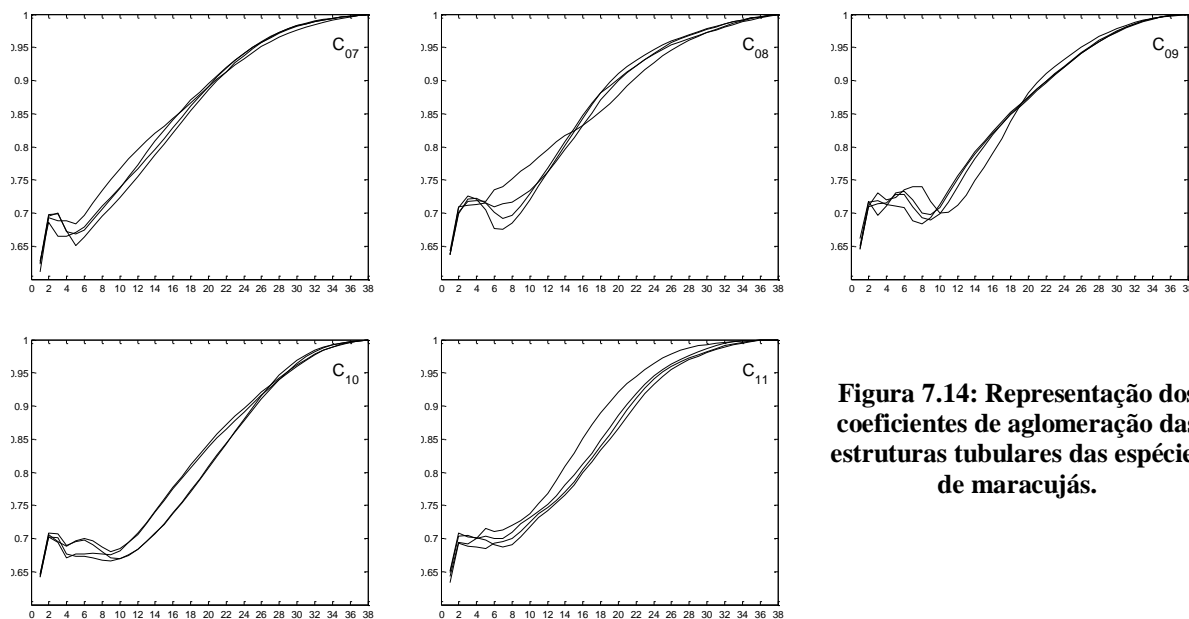


Figura 7.14: Representação dos coeficientes de aglomeração das estruturas tubulares das espécies de maracujás.

A Figura 7.15 apresenta o gráfico ROC gerado a partir do experimento E131. No gráfico é possível notar o alto poder de discriminação das espécies, uma vez que grande parte das espécies está agrupada próxima ao canto superior esquerdo (classificação perfeita).

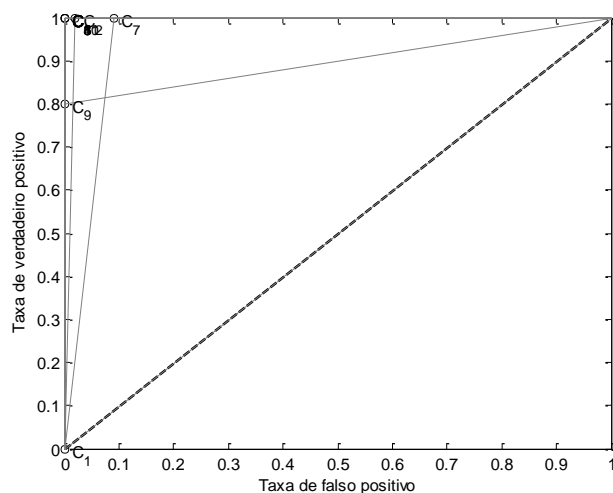


Figura 7.15: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados das redes complexas.

D. Seleção de Características

A seleção de características foi utilizada para selecionar os atributos mais discriminantes dentre as técnicas utilizadas para análise das estruturas tubulares. Foram adotados quatro critérios para seleção das características: (i) variância total superior a 85% dentre todas as características – identificado como S005; (ii) variância total superior a 95% dentre todas as

características (S006); (iii) primeiro componente principal (PC1) de cada técnica baseada na assinatura fractal (S007) e (iv) primeiro componente principal (PC1) de cada técnica baseada nas redes complexas (S008). As características selecionadas foram classificadas com os métodos já utilizados (DM, LDA, SRDA e RNA). Os resultados são apresentados na Tabela 7.22.

Tabela 7.22: Resultados da classificação das espécies com os dados da seleção de características.

Seleção de características		Maracujá (<i>Passiflora</i>)			
		DM	LDA	SRDA	RNA
S005	%A	64,46	94,21	88,29	13,18
	%E	35,54	5,79	11,71	86,82
S006	%A	57,85	76,86	92,56	14,55
	%E	42,15	23,14	7,44	85,45
S007	%A	38,02	85,12	84,30	10,23
	%E	61,98	14,88	15,70	89,77
S008	%A	86,78	91,74	91,23	20,68
	%E	13,22	8,26	8,77	79,32

A Figura 7.16 apresenta comparativamente o desempenho dos classificadores com os dados obtidos da seleção de características. Na Figura 7.16a ilustra o desempenho do classificador DM que alcançou uma porcentagem de acerto de 86,78% com os dados obtidos na seleção S008. A Figura 7.16b apresenta o resultado do classificador LDA, com 94,21% de acerto para a seleção S005. A Figura 7.16c é descrito o resultado do classificador SRDA, o qual alcançou 92,56% de discriminações corretas. E, por fim, a Figura 7.16d demonstra o resultado das RNA's com 20,68% de acerto. O fraco desempenho da rede neural está relacionado ao número de exemplos utilizados para treinamento e teste da rede. Como a análise de estruturas tubulares utilizou um número restrito de exemplares, não foi possível obter melhores resultados com esse classificador. Por outro lado, é importante destacar os excelentes resultados obtidos com os classificadores estatísticos: DM, LDA e SRDA.

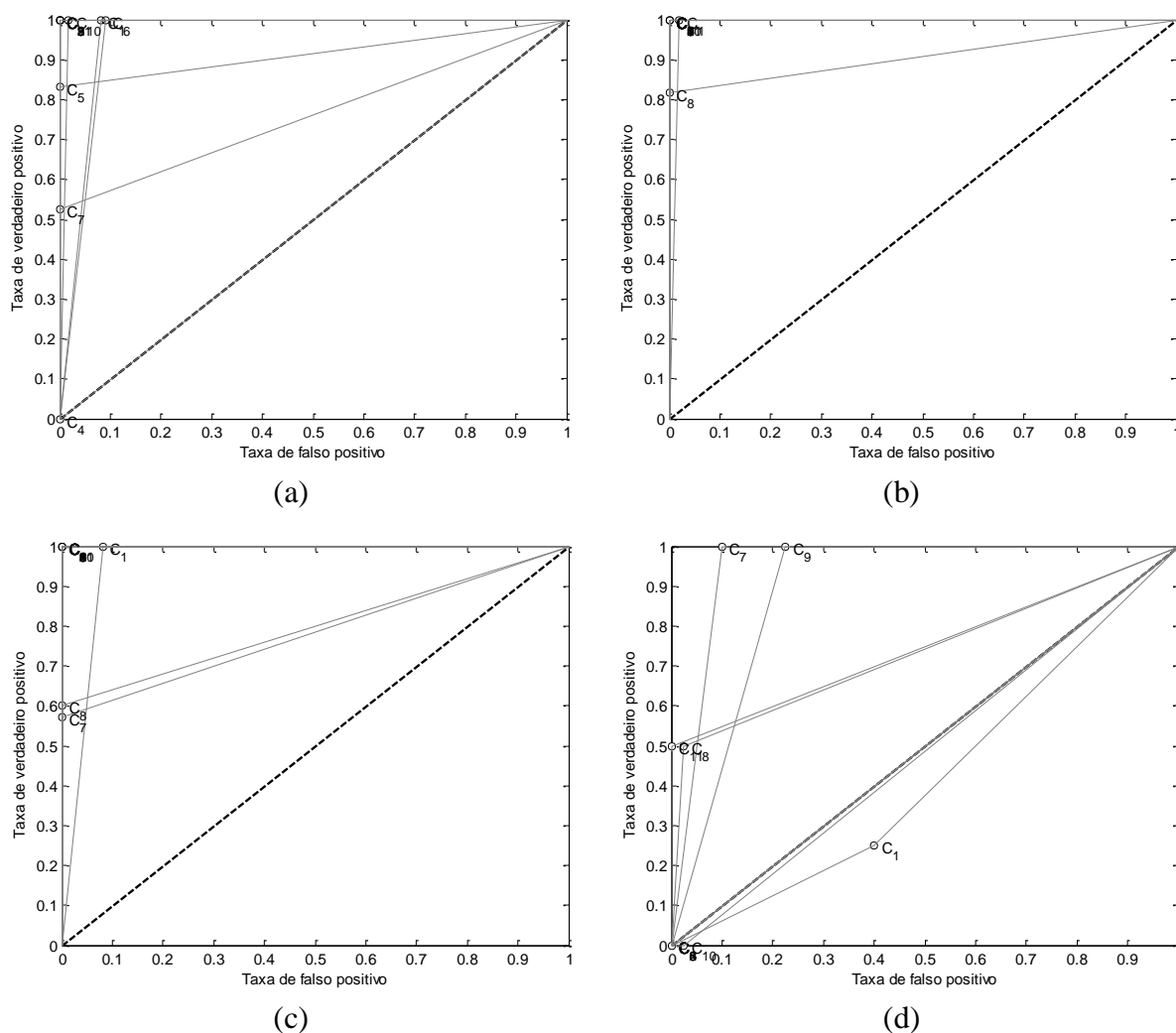


Figura 7.16: Gráficos ROC gerados a partir do desempenho dos classificadores com os dados da seleção de características. (a) classificador DM; (b) classificador LDA; (c) classificador SRDA e (d) rede neural artificial.

7.3 Resultados das características morfométricas das texturas

Para análise da textura das espécies biológicas foi utilizado a matriz de coocorrência (*Haralick*), além das abordagens propostas baseadas na distribuição de níveis de cinza associadas a assinatura fractal e redes complexas. Nestes experimentos foram consideradas treze espécies de maracujás silvestres do gênero *Passiflora*. Dentro deste contexto, as técnicas utilizadas foram: (a) matriz de coocorrência; (b) assinatura fractal a partir da distribuição de níveis de cinza; (c) redes complexas. Por fim, é utilizada a seleção de características (item (d)). Os resultados são apresentados como se segue:

A. Matriz de Coocorrência

A matriz de coocorrência foi utilizada para extração de características relativa a textura das espécies biológicas. As características consideradas foram energia, contraste, correção e entropia. A Tabela 7.23 apresenta os resultados da classificação das espécies. Os experimentos com a matriz de coocorrência são identificados como T001 até T005.

Tabela 7.23: Resultados da classificação das espécies com os dados da seleção de características.

Matriz Coocorrência		Maracujá (<i>Passiflora</i>)		
		DM	LDA	SRDA
T001	%A	54,37	62,41	54,89
	%E	45,63	37,59	45,11
T002	%A	53,23	64,42	55,33
	%E	46,77	35,58	44,67
T003	%A	51,05	63,72	56,21
	%E	48,95	36,28	43,79
T004	%A	53,58	64,34	54,37
	%E	46,42	35,66	45,63
T005	%A	55,77	71,33	60,51
	%E	44,23	28,67	39,49

A Tabela 7.24 apresenta detalhadamente a matriz de confusão e informações estatísticas do resultado que obteve melhor desempenho. No experimento T005 contém os dados combinados das matrizes de coocorrência dos experimentos T001, T002, T003 e T004.

Tabela 7.24: Matriz de confusão e dados estatísticos resultantes do experimento T005.

	C ₀₁	C ₀₂	C ₀₃	C ₀₄	C ₀₅	C ₀₆	C ₀₇	C ₀₈	C ₀₉	C ₁₀	C ₁₁	C ₁₂	C ₁₃	%A	%E	
C ₀₁	50,0	33,3	16,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	50,0	50,0	
C ₀₂	0,0	86,1	13,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	86,1	13,9	
C ₀₃	0,0	16,7	83,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	83,3	16,7	
C ₀₄	16,7	0,0	8,3	75,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	75,0	25,0	
C ₀₅	0,0	0,0	0,0	0,0	50,0	37,5	0,0	0,0	0,0	12,5	0,0	0,0	0,0	50,0	50,0	
C ₀₆	0,0	0,0	0,0	0,0	12,5	87,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	87,5	12,5	
C ₀₇	0,0	0,0	0,0	0,0	0,0	0,0	75,0	0,0	0,0	0,0	0,0	25,0	0,0	75,0	25,0	
C ₀₈	0,0	0,0	0,0	0,0	0,0	28,6	28,6	42,9	0,0	0,0	0,0	0,0	0,0	42,9	57,1	
C ₀₉	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	100,0	0,0	
C ₁₀	0,0	0,0	0,0	0,0	0,0	0,0	16,7	0,0	16,7	33,3	33,3	0,0	0,0	33,3	66,7	
C ₁₁	0,0	0,0	0,0	0,0	0,0	0,0	0,0	16,7	0,0	0,0	83,3	0,0	0,0	83,3	16,7	
C ₁₂	0,0	0,0	0,0	0,0	0,0	0,0	12,5	0,0	12,5	0,0	0,0	75,0	0,0	75,0	25,0	
C ₁₃	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	100,0	0,0	
														μ	71,3	28,7

$$AC = 95,59\% \pm 0,019$$

$$T_{VP} = 72,42\% \pm 0,216$$

$$T_{FP} = 2,38\% \pm 0,019$$

$$ES = 97,62\% \pm 0,019$$

$$PR = 72,68\% \pm 0,162$$

$$F = 70,39\% \pm 0,158$$

A Figura 7.17 ilustra graficamente as taxas de verdadeiro positivo e falso positivo por meio de um gráfico ROC. Os resultados obtidos com a análise de textura não alcançam resultados significantes, no entanto, serão muito importantes para uma análise comparativa com as outras metodologias.

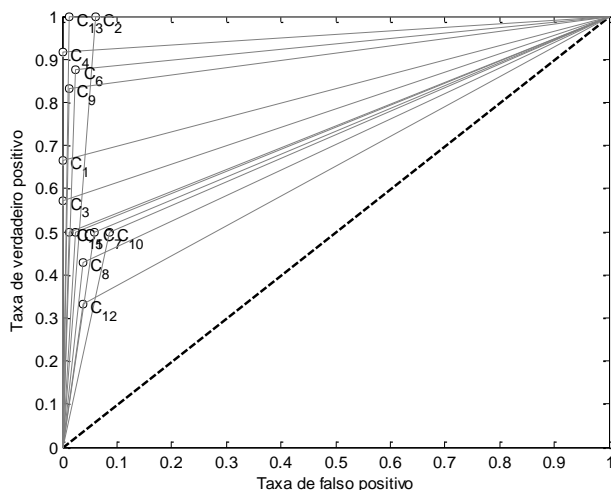


Figura 7.17: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados da matriz de coocorrência.

B. Distribuição de Níveis de Cinza e Assinatura Fractal

A análise de textura por meio da assinatura fractal multiescala utilizou como base as informações da distribuição de níveis de cinza das espécies. Foram adotados dois limiares para seleção dos subconjuntos de pixels $k = \{1.0, 0.8\}$. Os experimentos foram identificados de T006 até T029. A Tabela 7.25 apresenta os resultados obtidos com a classificação das espécies com a abordagem descrita. Na Tabela são sintetizadas as maiores taxas de classificações corretas.

Tabela 7.25: Resultados da classificação por meio da análise de textura a partir da assinatura fractal.

Assinatura Fractal	Maracujá (<i>Passiflora</i>)			
	DM	LDA	SRDA	
T009	%A	69,24	86,71	79,21
	%E	30,76	13,29	20,79
T012	%A	69,84	86,54	80,05
	%E	30,16	13,46	19,95
T015	%A	69,95	87,06	76,36
	%E	30,05	12,94	23,64
T017	%A	70,52	87,33	79,87
	%E	29,48	12,68	20,13
T017	%A	70,12	87,59	78,68
	%E	29,88	12,41	21,32

Na Tabela 7.26 são descritas a matriz de confusão e as informações estatísticas do experimento selecionado. Para a assinatura fractal gerada a partir da distribuição de níveis de cinza, a maior taxa de acerto foi alcançada para ordem de derivação igual a 500, desvio padrão da gaussiana igual a 250 e limiar de seleção do subconjunto igual a 1.0 – este experimento foi identificado como T017.

Tabela 7.26: Matriz de confusão e dados estatísticos resultantes do experimento com assinatura fractal gerada a partir da distribuição de níveis de cinza.

	C ₀₁	C ₀₂	C ₀₃	C ₀₄	C ₀₅	C ₀₆	C ₀₇	C ₀₈	C ₀₉	C ₁₀	C ₁₁	C ₁₂	C ₁₃	%A	%E
C ₀₁	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₂	0,0	66,7	33,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	66,7	33,3
C ₀₃	0,0	14,3	85,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	85,7	14,3
C ₀₄	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₅	0,0	0,0	0,0	0,0	75,0	12,5	12,5	0,0	0,0	0,0	0,0	0,0	0,0	75,0	25,0
C ₀₆	0,0	0,0	0,0	0,0	9,1	72,7	0,0	0,0	0,0	0,0	18,2	0,0	0,0	72,7	27,3
C ₀₇	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₈	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₉	0,0	0,0	0,0	0,0	10,0	10,0	0,0	0,0	80,0	0,0	0,0	0,0	0,0	80,0	20,0
C ₁₀	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	100,0	0,0
C ₁₁	16,7	0,0	0,0	0,0	0,0	16,7	0,0	0,0	0,0	0,0	66,7	0,0	0,0	66,7	33,3
C ₁₂	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	88,9	11,1	88,9	11,1
C ₁₃	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	33,3	66,7	66,7	33,3
													μ	87,6	12,4

AC = 98,09%
±0,016

T_{VP} = 84,79%
±0,142

T_{FP} = 1,03%
±0,009

ES = 98,97%
±0,009

PR = 84,98%
±0,130

F = 84,69%
±0,129

A Figura 7.18 apresenta as assinaturas fractal multiescala geradas a partir dos dados do experimento T017. Os gráficos demonstram comparativamente o comportamento das curvas com base nas informações extraídas das espécies. Cada gráfico agrupa exemplares de uma mesma espécie, e a identificação da espécie pode ser observada no canto superior direito. Na Figura é importante observar que a maioria das curvas possui um comportamento padronizado, o que indica a potencialidade da metodologia na identificação das espécies. No entanto, algumas curvas, como por exemplo, a C05 e C09, apresentam comportamento não uniforme. O eixo Y dos gráficos representa a fractalidade das espécies, enquanto o eixo X o número de pontos da assinatura fractal.

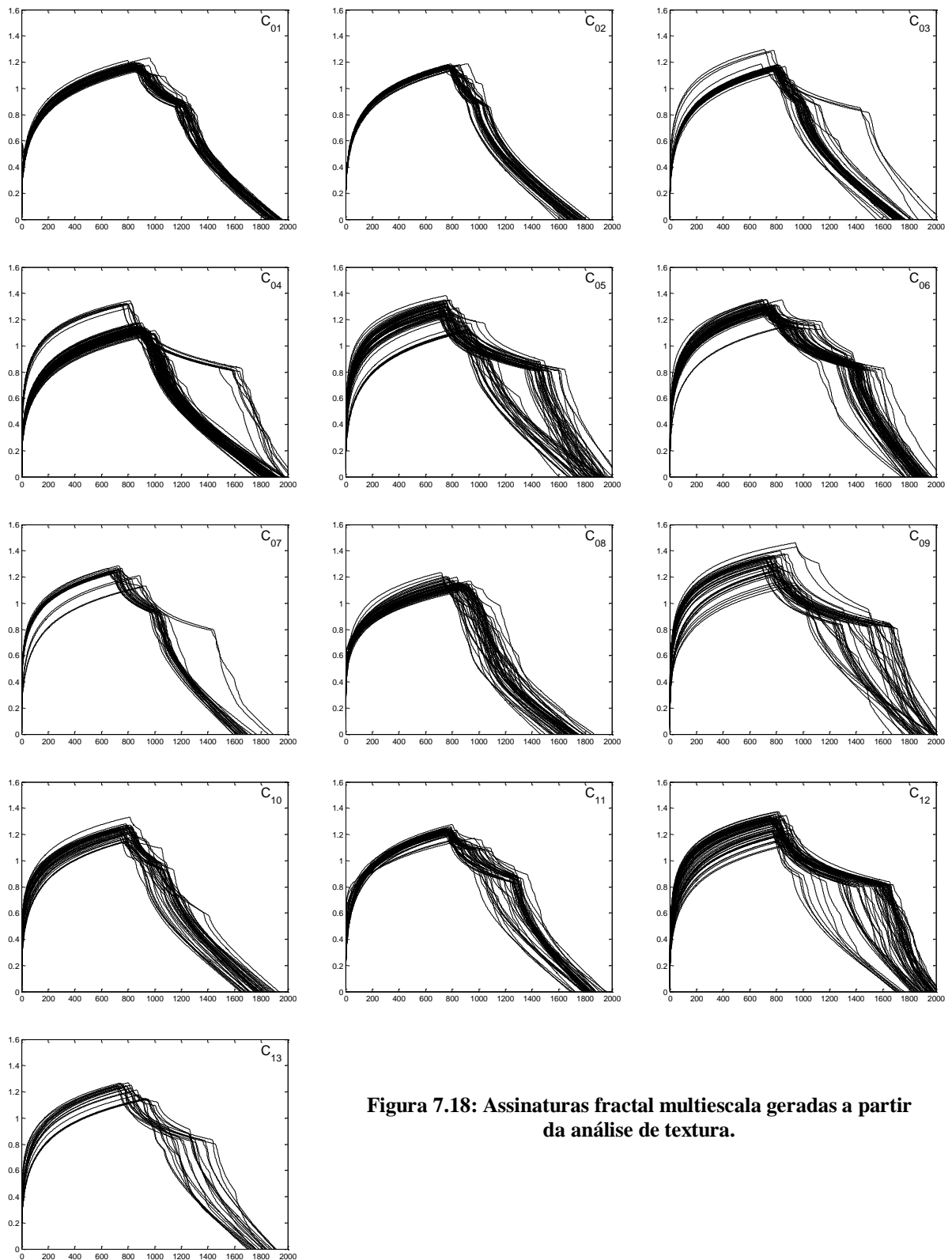


Figura 7.18: Assinaturas fractal multiescala geradas a partir da análise de textura.

A metodologia proposta de análise de textura baseada na assinatura fractal demonstrou grande potencialidade na classificação das espécies de maracujás. Além disso, esta abordagem sinaliza a viabilidade da utilização de características de textura na classificação das espécies.

O gráfico ROC da Figura 7.19 ilustra comparativamente a classificação das espécies por meio da metodologia descrita.

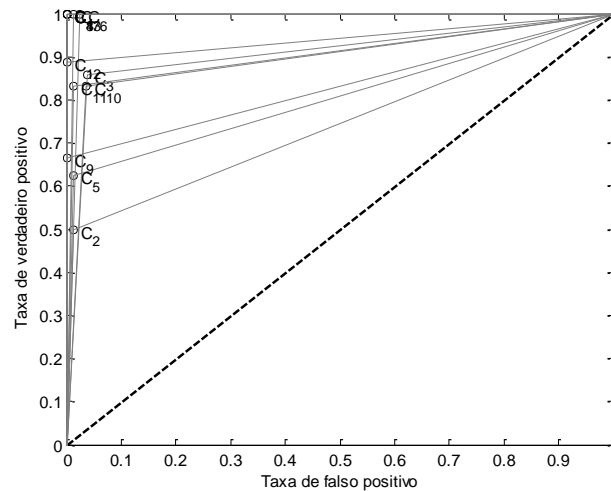


Figura 7.19: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados da assinatura fractal para $k = 1.0$ e classificador LDA.

C. Distribuição de Níveis de Cinza e Redes Complexas

A Tabela 7.27 sintetiza os resultados obtidos com a análise de textura com as redes complexas. Os resultados compreendem os experimentos com taxa de acerto potencialmente discriminantes. Os experimentos utilizando esta abordagem foram identificados de T030 até T192.

Tabela 7.27: Resultados da classificação por meio da análise de textura a partir das redes complexas.

Redes Complexas	Maracujá (<i>Passiflora</i>)			
	DM	LDA	SRDA	
T115	%A	58,13	68,44	68,13
	%E	41,87	31,56	31,87
T117	%A	40,21	68,01	65,87
	%E	59,79	31,99	34,13
T140	%A	43,36	63,37	62,20
	%E	56,64	36,63	37,80
T182	%A	62,50	65,03	66,13
	%E	37,50	34,97	33,87
T185	%A	64,60	65,30	68,44
	%E	35,40	34,70	31,56

Dentre os experimentos realizados com redes complexas para análise de textura, o melhor resultado foi alcançado com as características extraídas a partir da entropia da rede complexa. A Tabela 7.28 apresenta a matriz de confusão e respectivas informações estatísticas do experimento T115. Neste experimento o critério adotado para seleção do subconjunto foi $k = 0.8$. Os resultados alcançados foram instigantes, uma vez que a taxa de acerto foi inferior a 70%.

Tabela 7.28: Matriz de confusão e dados estatísticos resultantes do experimento com redes complexas.

	C ₀₁	C ₀₂	C ₀₃	C ₀₄	C ₀₅	C ₀₆	C ₀₇	C ₀₈	C ₀₉	C ₁₀	C ₁₁	C ₁₂	C ₁₃	%A	%E
C ₀₁	66,9	0,0	0,0	0,0	0,0	11,0	0,0	0,0	0,0	0,0	11,0	11,0	0,0	66,9	33,1
C ₀₂	0,0	71,4	14,3	14,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	71,4	28,6
C ₀₃	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₄	0,0	6,7	13,3	80,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	80,0	20,0
C ₀₅	0,0	0,0	0,0	0,0	37,5	0,0	0,0	12,5	50,0	0,0	0,0	0,0	0,0	37,5	62,5
C ₀₆	0,0	25,0	0,0	0,0	0,0	12,5	12,5	12,5	0,0	25,0	12,5	0,0	0,0	12,5	87,5
C ₀₇	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0	0,0	0,0	0,0	0,0	0,0	100,0	0,0
C ₀₈	14,3	0,0	0,0	0,0	0,0	0,0	0,0	85,7	0,0	0,0	0,0	0,0	0,0	85,7	14,3
C ₀₉	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	66,7	16,7	0,0	16,7	0,0	66,7	33,3
C ₁₀	16,7	0,0	0,0	0,0	16,7	0,0	0,0	0,0	0,0	66,7	0,0	0,0	0,0	66,7	33,3
C ₁₁	0,0	0,0	0,0	0,0	0,0	0,0	16,7	16,7	0,0	0,0	66,7	0,0	0,0	66,7	33,3
C ₁₂	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	11,1	88,9	0,0	88,9	11,1
C ₁₃	0,0	0,0	0,0	33,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	33,3	33,3	33,3	66,7
														μ	68,4
															31,6

AC = 95,14% $T_{VP} = 67,40\% \pm 0,261$ $T_{FP} = 2,64\% \pm 0,012$ ES = 97,36% $\pm 0,012$ PR = 68,36% $\pm 0,141$ F = 64,44% $\pm 0,177$

Na Figura 7.20 são apresentados os resultados da classificação por meio de um gráfico ROC. No gráfico é possível visualizar que o desempenho da classificação das texturas das espécies com as redes complexas foi promissor, uma vez que nenhuma das espécies foi disposta na diagonal inferior do gráfico ROC.

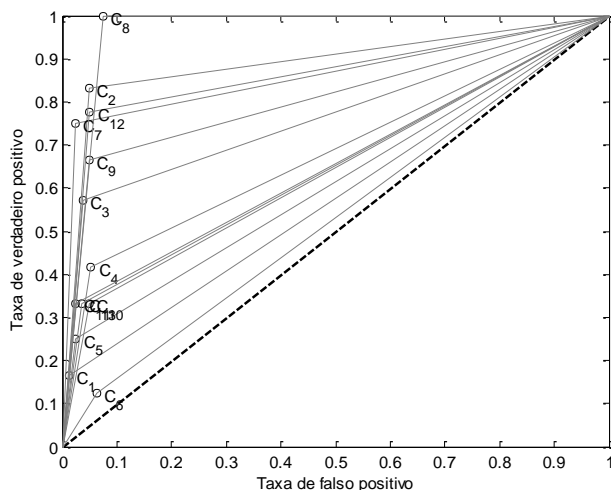


Figura 7.20: Gráfico ROC gerado a partir dos resultados da classificação das espécies de maracujá com os dados das redes complexas e classificador LDA.

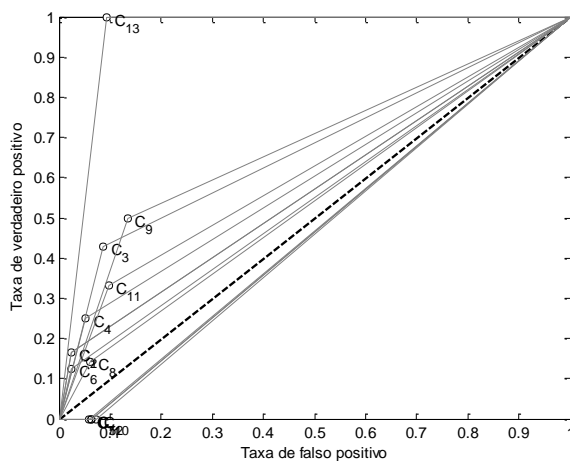
D. Seleção de Características

Para seleção de características dos dados resultantes dos experimentos de análise de textura, foram utilizados quatro critérios: (i) seleção de características com variância superior a 85% - identificado como S009; (ii) variância superior a 95% (S010); (iii) primeira componente principal dos dados da assinatura fractal (S011) e (iv) PC1 dos conjuntos das redes complexas (S012). A Tabela 7.29 descreve o resultado das classificações.

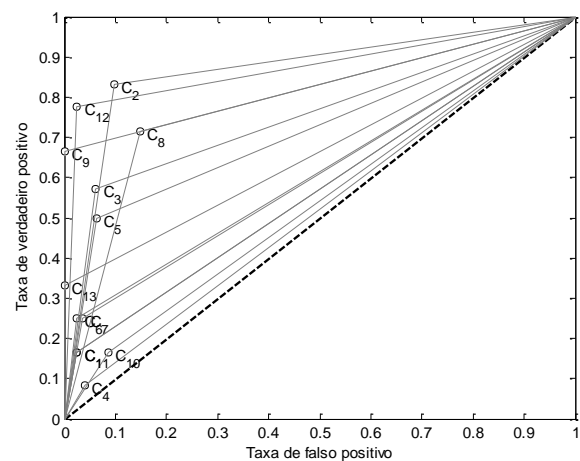
Tabela 7.29: Resultados da classificação das espécies com os dados da seleção de características.

Seleção de características		Maracujá (<i>Passiflora</i>)			
		DM	LDA	SRDA	RNA
S009	%A	42,22	58,83	75,09	20,86
	%E	57,78	41,17	24,91	79,14
S010	%A	39,77	53,93	59,70	24,40
	%E	60,23	46,07	40,30	75,60
S011	%A	28,32	62,24	28,32	38,97
	%E	71,68	37,76	71,68	61,03
S012	%A	53,23	65,82	85,63	39,07
	%E	46,77	34,18	14,37	60,93

Na Figura 7.21 os resultados são comparados por meio de gráficos ROC. Os gráficos demonstram a relação entre as taxas de verdadeiro positivo e falso positivo obtidas pelos classificadores. É importante observar que a seleção de características alcançou melhores resultados com os dados exclusivamente das redes complexas. Este fato demonstra a importância da seleção de características na mineração dos dados mais discriminantes.



(a)



(b)

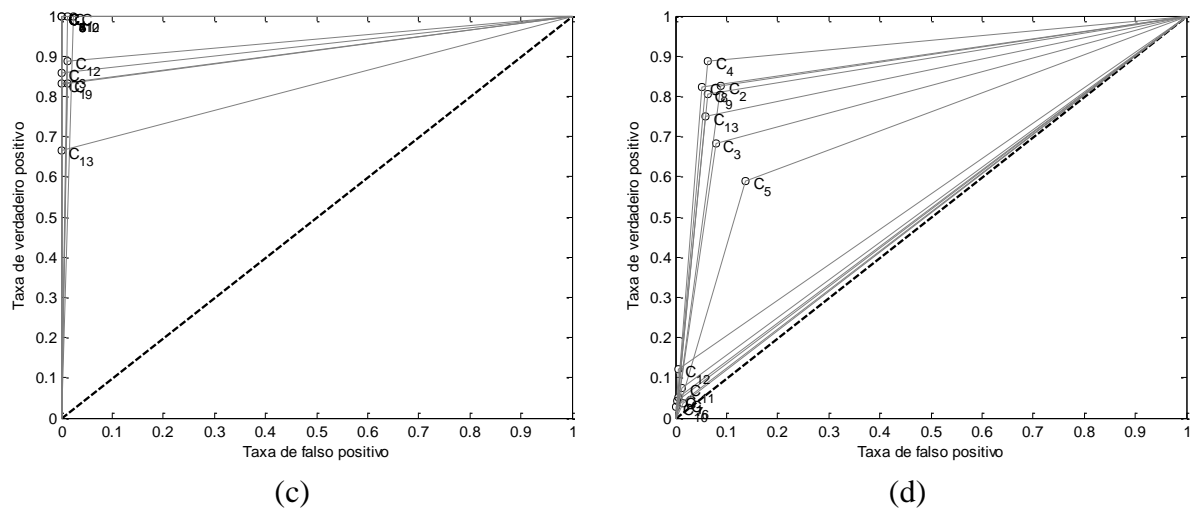


Figura 7.21: Gráficos ROC gerados a partir do desempenho dos classificadores com os dados da seleção de características. (a) classificador DM; (b) classificador LDA; (c) classificador SRDA e (d) rede neural artificial.

7.4 Combinação de características morfométricas

Os experimentos realizados com as técnicas de análise de imagens extraíram informações morfométricas das espécies a partir da forma, estrutura tubular e textura. As técnicas de reconhecimento de padrões foram empregadas nesses dados para avaliar a potencialidade de cada técnica na classificação das espécies. Essa avaliação foi realizada para cada um dos experimentos, e os dados demonstraram quantitativamente a qualidade de cada uma das características.

Buscando uma análise mais abrangente dos resultados, as características extraídas pelas técnicas de análise de imagens foram combinadas empiricamente e classificadas por meio das técnicas de reconhecimento de padrões. Dentro deste contexto, as seguintes combinações foram realizadas: (i) forma \times estrutura tubular – identificado como C001; (ii) forma \times textura (C002); (iii) estrutura tubular \times textura (C003) e (iv) forma \times estrutura tubular \times textura (C004).

Os dados para cada combinação foram selecionados a partir das primeiras componentes principais de cada experimento realizado. Em virtude das características dos experimentos, apenas onze espécies de maracujás foram consideradas (quatro exemplares por espécie), dentre os três grupos: forma, estrutura tubular e textura. A Tabela 7.30 apresenta os

resultados das classificações com base nos dados combinados. Os classificadores utilizados foram DM, LDA e SRDA.

Tabela 7.30: Resultados da classificação das espécies com os dados combinados.

Combinação de características		Maracujá (<i>Passiflora</i>)		
		DM	LDA	SRDA
C001	%A	81,82	93,25	82,14
	%E	18,18	6,75	17,86
C002	%A	68,59	66,94	61,55
	%E	38,02	33,06	38,45
C003	%A	85,95	91,74	87,32
	%E	14,05	8,26	12,68
C004	%A	80,99	93,39	88,35
	%E	19,01	6,61	11,65

Os resultados obtidos com a combinação das características de formas, estruturas tubulares e texturas demonstram a importância de uma análise morfométrica detalhada na classificação das espécies. Nos experimentos, independente do classificador utilizado, a discriminação correta das espécies foi superior a 85%. Este fato sinaliza para necessidade de análises minuciosas de organismos biológicos, as quais não devem considerar apenas uma (ou algumas) características de morfométrica. Além disso, para que essa abordagem seja eficiente, é necessário que um processo prévio de seleção de características seja realizado nos conjuntos de dados. Com isso, a dimensionalidade dos conjuntos não prejudicará (ou influenciará) no processo de classificação.

Na Figura 7.22 é possível observar os resultados dos classificadores para abordagem baseada na combinação das características morfométricas. É importante observar na Figura, que todas as espécies estão próximas ao ponto (0,1) no espaço ROC que indica uma classificação perfeita.

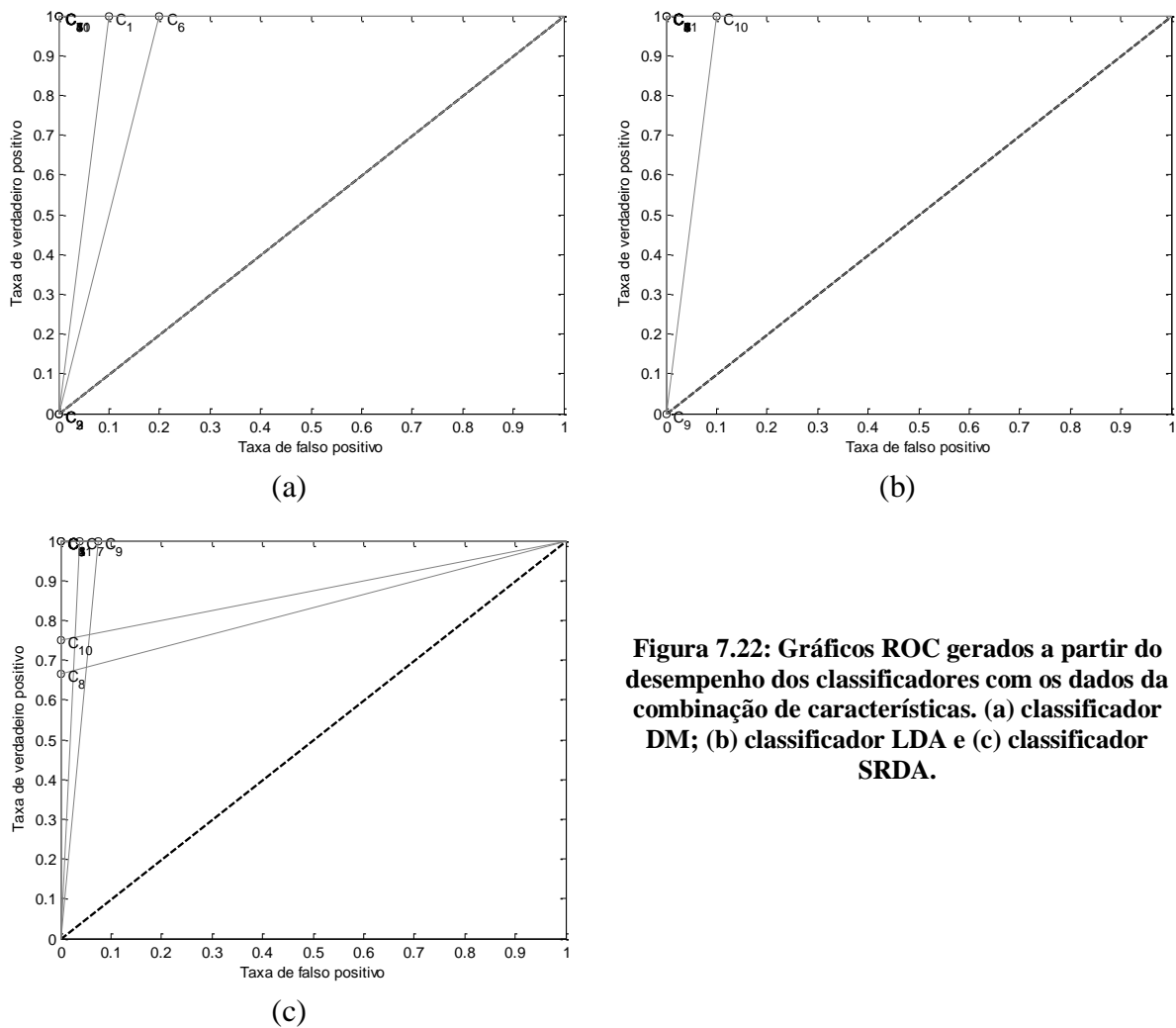


Figura 7.22: Gráficos ROC gerados a partir do desempenho dos classificadores com os dados da combinação de características. (a) classificador DM; (b) classificador LDA e (c) classificador SRDA.

7.5 Inferência Evolutiva das Espécies

Os resultados obtidos com as técnicas de análise de imagens indicaram grande viabilidade da utilização destas metodologias na taxonomia das espécies biológicas. Os dados morfométricos foram capaz de distinguir corretamente as espécies com base na forma, estrutura tubular e textura.

Um importante resultado foi alcançando por meio da correlação dos dados morfométricos com dados moleculares utilizados em estudos evolutivos das espécies de *Passiflora*. Estes dados de reconstrução filogenética estão disponíveis em (Muschner, Lorenz *et al.*, 2003; Pádua, 2004). Para essa avaliação foram consideradas espécies de maracujás com grande variabilidade (*P.amethystina*, *P.caerulea*, *P.coriacea*, *P.foetida*, *P.gibertii*, *P.maliformis*, *P.miersii*, *P.organensis*, *P.pohlii* e *P.suberosa*), e a correlação entre as

abordagens foi calculada com o teste estatístico de Mantel (Mantel, 1967; Sokal e Rohlf, 1995). O índice positivo do teste de Mantel indica que as amostras com características morfométricas diferentes (grande distância entre clusters) tendem a ter baixa correlação, enquanto que as características similares possuem alta covariância (Diniz-Filho, 2000). A Tabela 7.31 apresenta os resultados do teste de Mantel para os grupos de técnicas de análise de imagens utilizadas nos experimentos, além disso, são evidenciadas as técnicas visão artificial que demonstraram maior correspondência com os dados da reconstrução filogenética. A métrica de distância *chebychev* foi utilizada para estimar a correlação entre as duas abordagens.

Tabela 7.31: Resultados da correspondência entre dados morfométricos e moleculares obtidos com o teste estatístico de Mantel

Técnicas de Análise de Imagens	Mantel
Análise de Formas	
<i>Wavelets Daubechies</i> (db13 n5 detalhe)	0,3137
<i>Wavelets Daubechies</i> (db33 n5 detalhe)	0,2740
<i>Wavelets Coiflets</i> (coif1 n3 detalhe)	0,2130
Análise de Estruturas Tubulares	
Assinatura Fractal ($\delta = 250$; $\sigma=500$)	0,5894
Redes Complexas Pontos Extremos (Dist.Conexões $Ti = 0.375$)	0,5831
Redes Complexas Bifurações e Pts Extremos (Grau $Ti = 0.25$)	0,5709
Análise de Textura	
Redes Complexas Grau $Ti = 0.1$ e $k = 0.8$	0,3309
Redes Complexas Grau Máximo $k = 0.8$	0,2566
Redes Complexas Grau $Ti = 0.3$ e $k = 0.8$	0,2426

Os resultados indicam que as metodologias de análise de imagens foram capazes de não apenas distinguir as espécies biológicas, como também identificar a variabilidade dentro de cada uma das espécies. Para as técnicas de análise de estruturas tubulares a alta correlação entre as abordagens demonstra a importância do sistema de venação no estudo evolutivo das espécies.

Para demonstrar a correspondência entre as abordagens (morfométrica e molecular), dendrogramas foram gerados a partir dos dados morfométricos extraídos pelas técnicas de análise de imagens. Foram consideradas as técnicas com maior correlação entre morfometria e filogenia (conforme descrito na Tabela 7.31). Os dendrogramas também foram importantes

para visualizar o agrupamento das espécies. Diversas metodologias foram testadas para geração dos dendrogramas (*single linkage*, *complete linkage*, *UPGMA*, *WPGMA*, *UPGMC*, *WPGMC* e *Ward's*), além das funções de distâncias euclidiana e *chebychev*.

A Figura 7.23 apresenta o dendrograma gerado com a análise da forma das espécies de maracujás por meio da técnica de *wavelets* (db13). Este dendrograma foi construído com o método UPGMA e a métrica de distância euclidiana. No dendrograma é importante notar a separabilidade das espécies a partir do ponto de corte (linha tracejada), uma vez que as espécies a partir deste ponto ($\cong 0,43$) foram 100% classificadas.

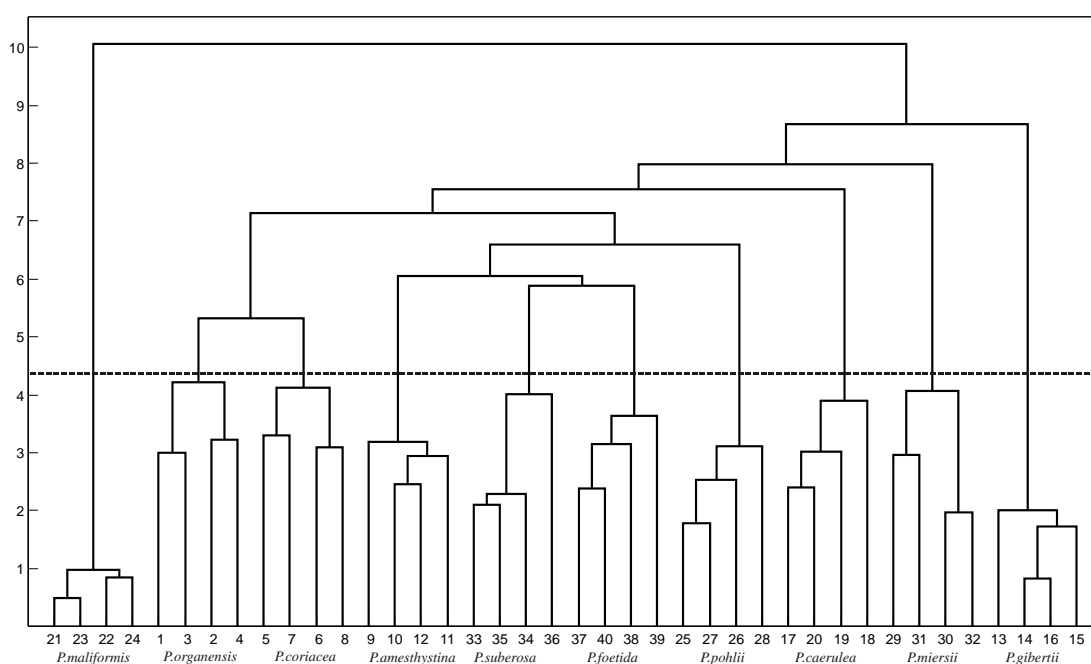


Figura 7.23: Inferência evolutiva das espécies de maracujás a partir da análise de formas.

Para avaliar a relação entre a matriz de distâncias dos dados e o dendrograma gerado, foi utilizado o coeficiente de correlação cofenético (*cophenetic correlation coefficient - CCF*) (Theodoridis e Koutroumbas, 2003). Este coeficiente permite comparar um dendrograma com uma matriz de distância, ou também avaliar a correlação entre dois dendrogramas, e é definido como se segue (Jain e Dubes, 1988): Seja $d(i, j)$ a matriz de distâncias entre os padrões (ou espécies) i e j . Denota-se por $dc(i, j)$ a matriz cofenética, que representa o nível no dendrograma em que os padrões i e j foram colocados no mesmo cluster. A correlação entre as matrizes $d(i, j)$ e $dc(i, j)$ pode ser expresso na Equação 7.1. O resultado é um valor no intervalo $[-1, +1]$, o qual quanto mais próximo de $+1$, maior é a correlação entre a matriz de distância e o dendrograma gerado.

$$CCF = \frac{(1/M) \sum d(i, j) \cdot dc(i, j) - (m_D, m_C)}{\sqrt{(1/M) \sum d(i, j)^2 - m_D} \cdot \sqrt{(1/M) \sum dc(i, j)^2 - m_C}} \quad (7.1)$$

$$\{(i, j) : 1 \leq i < j \leq n\}$$

sendo $M = n(n-1)/2$ o total de entradas da diagonal superior; $m_D = (1/M) \sum d(i, j)$, $m_C = (1/M) \sum dc(i, j)$ a soma de todos elementos da diagonal superior.

Para o dendrograma gerado a partir da técnica de *wavelets* o coeficiente cofenético obtido foi igual a 0,8979.

Um dendrograma também foi gerado com os dados morfométricos obtidos com a análise de estruturas tubulares, e é apresentado na Figura 7.24. Neste caso, o dendrograma foi construído a partir dos dados da assinatura fractal dos pontos extremos, com o método UPGMA e distância euclidiana. O coeficiente de correlação cofenético obtido foi igual a 0,9593. Novamente, considerando o ponto de corte no dendrograma, aproximadamente $\cong 2,50$, todas as espécies foram discriminadas corretamente.

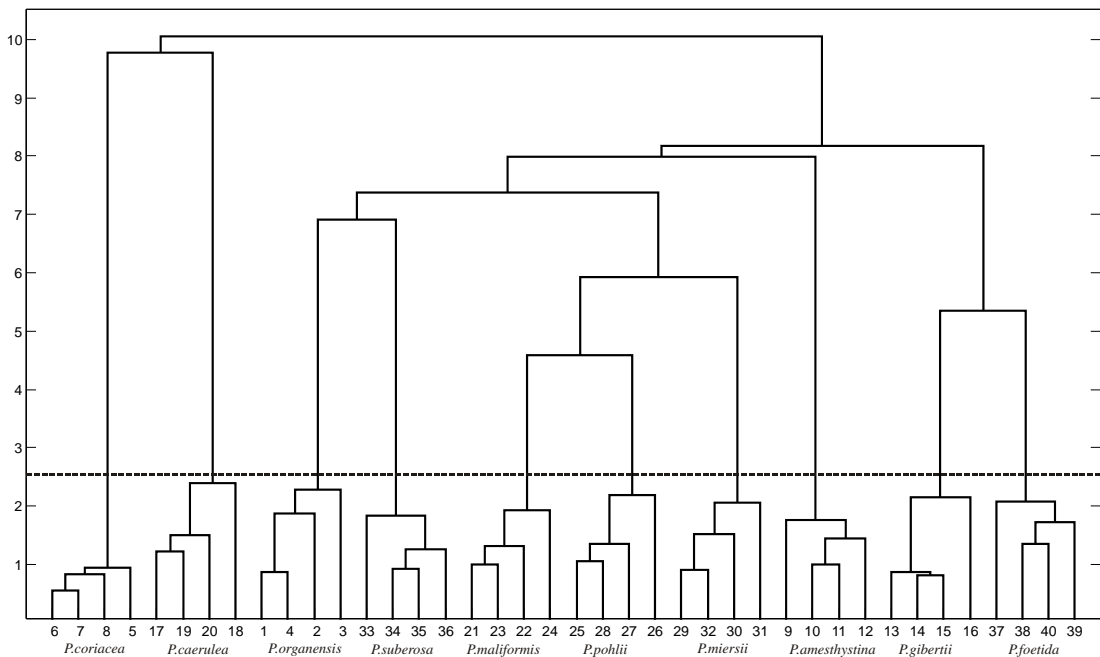


Figura 7.24: Inferência evolutiva das espécies de maracujás a partir da análise de estruturas tubulares.

Para as técnicas de análise de textura a melhor correlação foi obtida com os dados das redes complexas. O dendrograma foi construído para a rede gerada com o subconjunto selecionado a partir do parâmetro 0.8. Neste dendrograma o ponto de corte foi definido como $\cong 20,00$, no entanto, as espécies não foram corretamente classificadas. Na Figura 7.25 é apresentado o dendrograma gerado com o método UPGMA e métrica de distância euclidiana. É importante observar na Figura que os exemplares 17, 31 e 32 foram classificados em espécies diferentes das quais eles realmente pertencem. O exemplar 17, por exemplo, foi agrupado a espécies *P.maliformis*, porém este exemplar pertence à espécie *P.caerulea*. O coeficiente cofenético do dendrograma apresentado foi igual a 0,8978.

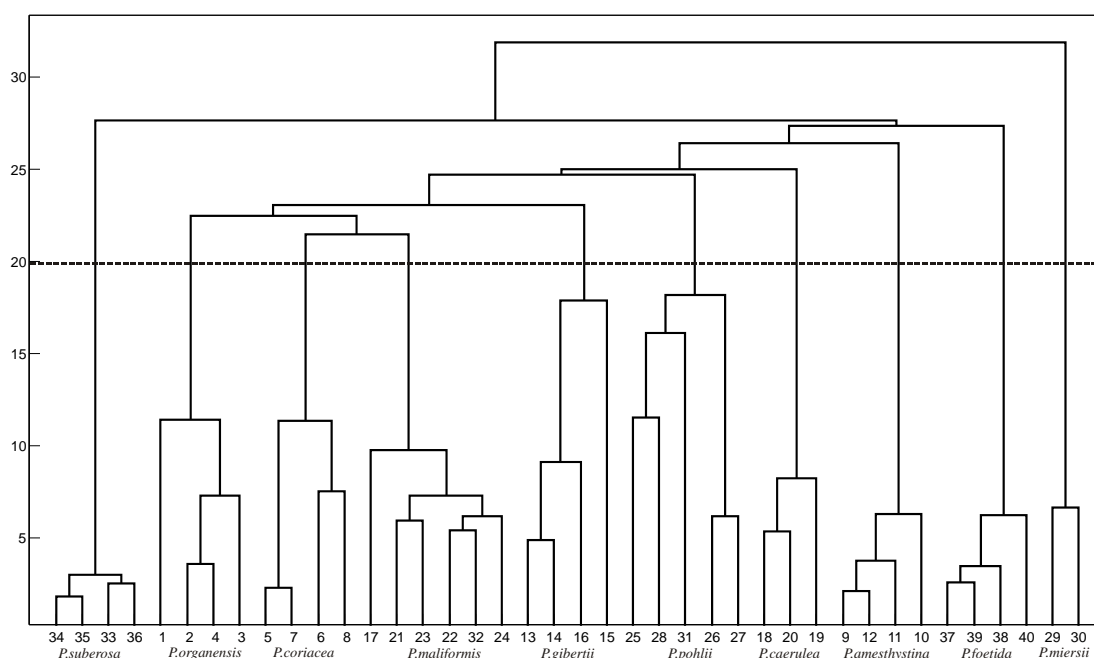


Figura 7.25: Inferência evolutiva das espécies de maracujás a partir da análise de textura.

As técnicas de análise de imagens aplicadas ao gênero *Passiflora* produziram um rico conjunto de dados a respeito da morfometria das espécies. O alto grau de similaridade entre as abordagens morfométricas e filogenéticas, demonstram o grande potencial destas metodologias de visão artificial no estudo evolutivo das espécies.

Para as espécies de eucalipto não foi possível realizar a correlação dos dados morfométricos e filogenéticos. No entanto, dendrogramas também foram gerados para visualização do agrupamento das espécies. Como o número de experimentos realizados com as espécies de eucalipto também foi grande, os dados utilizados para construção do dendrograma foram selecionados com base no resultado do processo de reconhecimento de

padrões. Para isso, o valor do coeficiente de correlação de *Matthews* (Baldi, Brunak *et al.*, 2000) foi utilizado como critério. Esta medida é utilizada para representar com um único valor numérico a qualidade das classificações. O valor retornado pertence ao intervalo -1 e $+1$, em que um coeficiente de $+1$ representa a classificação perfeita, o coeficiente 0 uma classificação aleatória, e -1 uma classificação inversa. O coeficiente de *Matthews* pode ser calculado com a Equação 7.2.

$$CCM = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (7.2)$$

em que *VP* representa o número de verdadeiros positivos, *VN* o número de verdadeiros negativos, *FP* o número de falso positivos e *FN* o número de falso negativos.

Para os dados morfométricos das espécies de eucalipto, produzidos a partir das técnicas de análise de formas, o maior coeficiente de correlação de *Matthews* encontrado foi igual a 0,8262. Este valor foi obtido a partir do resultado da classificação das espécies com os dados da transformada de *wavelets* com a família *Daubechies* (db43, nível 4). O experimento considerou apenas os coeficientes de tendência e o classificador utilizado foi o LDA. A Figura 7.26 ilustra o dendrograma construído com as informações da transformada de *wavelets*.

Como o número de exemplares das espécies de eucalipto é grande, a visualização das informações no dendrograma tornou-se prejudicada, uma vez, que no eixo *X* são representados cada um dos 285 exemplares. Para facilitar a visualização, o dendrograma foi construído a partir de quatro exemplares selecionados aleatoriamente de cada espécie. O método de agrupamento utilizado foi o UPGMA (distância euclidiana) e coeficiente cofenético igual a 0,9932.

Um fato interessante pode ser visualizado no dendrograma das espécies de eucalipto. As espécies utilizadas nos experimentos podem ser divididas em dois grupos: puras e híbridas. O dendrograma gerado descreve fielmente esta separação entre as espécies. Outro resultado importante que pode ser visualizado no dendrograma é o padrão de agrupamento das espécies híbridas. Na Figura é possível notar que as espécies híbridas, identificadas de 17 a 28, formadas pela espécie *E.camaldulensis* foram agrupadas próximas umas das outras. Essa relação de similaridade também pode ser observada com os híbridos formados pela espécie

E.globulus (identificados de 29 a 36). Estes fatores podem ser fortes evidências entre a correlação de dados morfométricos e filogenéticos, e que certamente serão explorados.

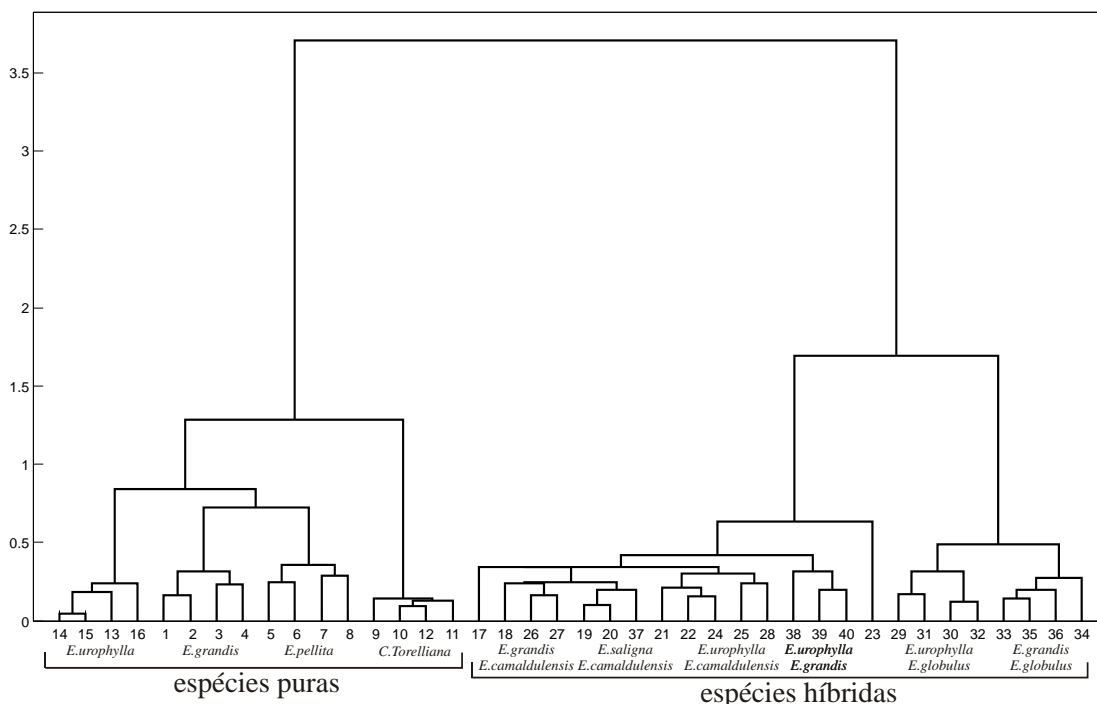


Figura 7.26: Inferência evolutiva das espécies de eucalipto a partir da análise de formas com a transformada de wavelets.

Analisando as espécies de eucalipto separadamente, é possível observar a potencialidade da metodologia na classificação das espécies. Além disso, a relação de similaridade entre as espécies híbridas pode ser comprovada.

A Figura 7.27 apresenta o resultado do classificador LDA aplicado às espécies puras e híbridas separadamente. Na Figura 7.27a é demonstrada a eficiência da classificação das espécies puras, na qual 100% das espécies foram corretamente classificadas. A Figura 7.27b ilustra a classificação das espécies híbridas, em que é importante notar a relação de similaridade entre as espécies que possuem ancestrais em comum.

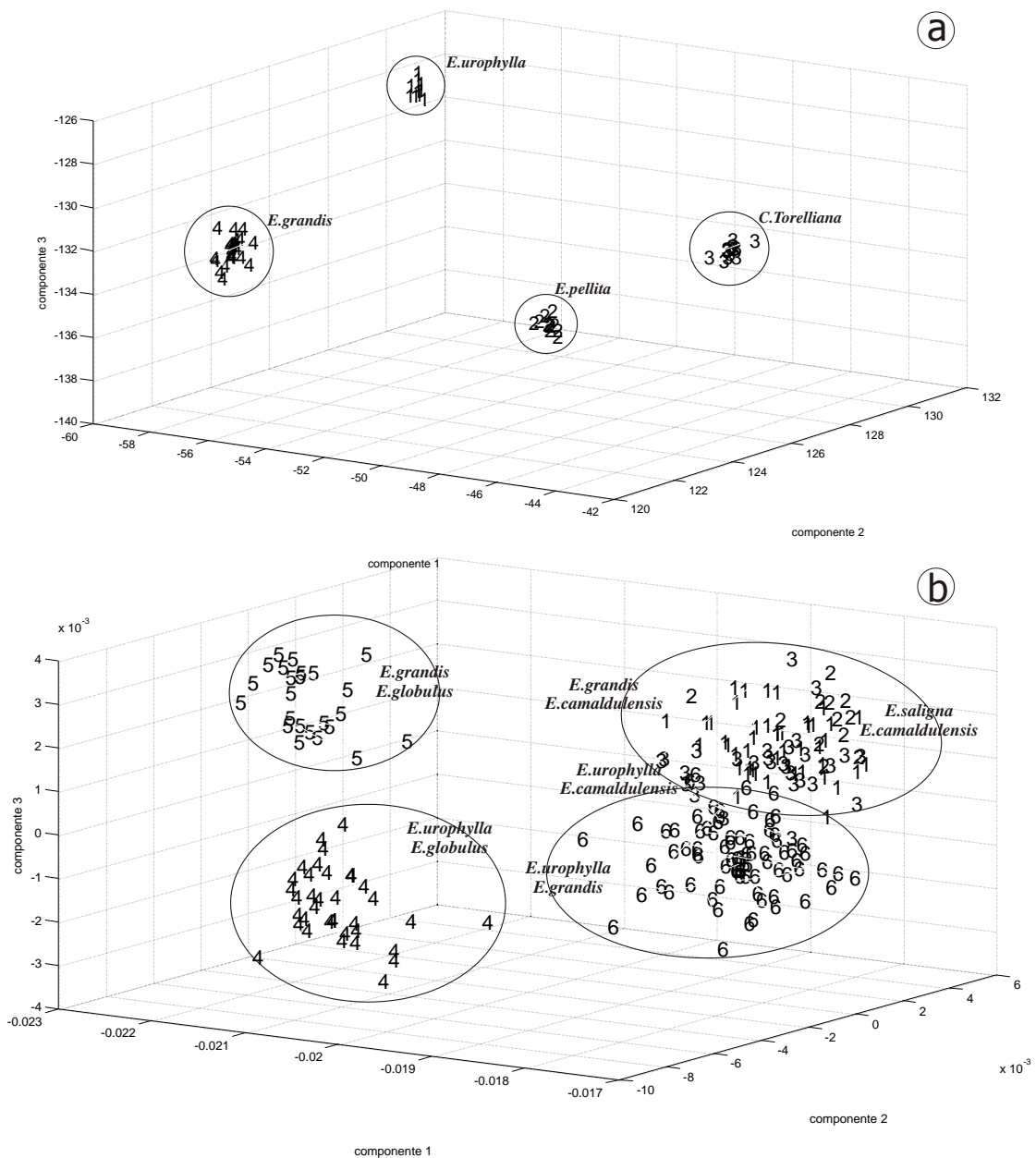


Figura 7.27: Separabilidade das espécies de eucalipto: (a) espécies puras; (b) espécies híbridas.

Os dados morfométricos das espécies de arroz também foram utilizados para construção de dendrogramas. Neste caso, o maior coeficiente de correlação de *Matthews* obtido foi igual a 0,6419. Este valor foi alcançado para as características morfométricas da transformada de *wavelets* com a família *Daubechies* (db2, nível 2) e o classificador LDA. Para geração do dendrograma foi utilizado o método UPGMA e a métrica euclidiana, o qual resultou no coeficiente cofenético igual a 0,9376.

Como nas espécies de eucalipto, quatro amostras de cada espécie foram escolhidas aleatoriamente para construção dos dendrogramas. Com isso, o número de exemplares apresentados nos dendrogramas é menor que o número do conjunto original. Os exemplares de 1 até 4 representam amostras da espécie C_{01} , de 5 até 8 da espécie C_{02} , de 9 até 12 da espécie C_{03} , e assim, sucessivamente. A Figura 7.28 apresenta o dendrograma das espécies de arroz do gênero *Oryza*.

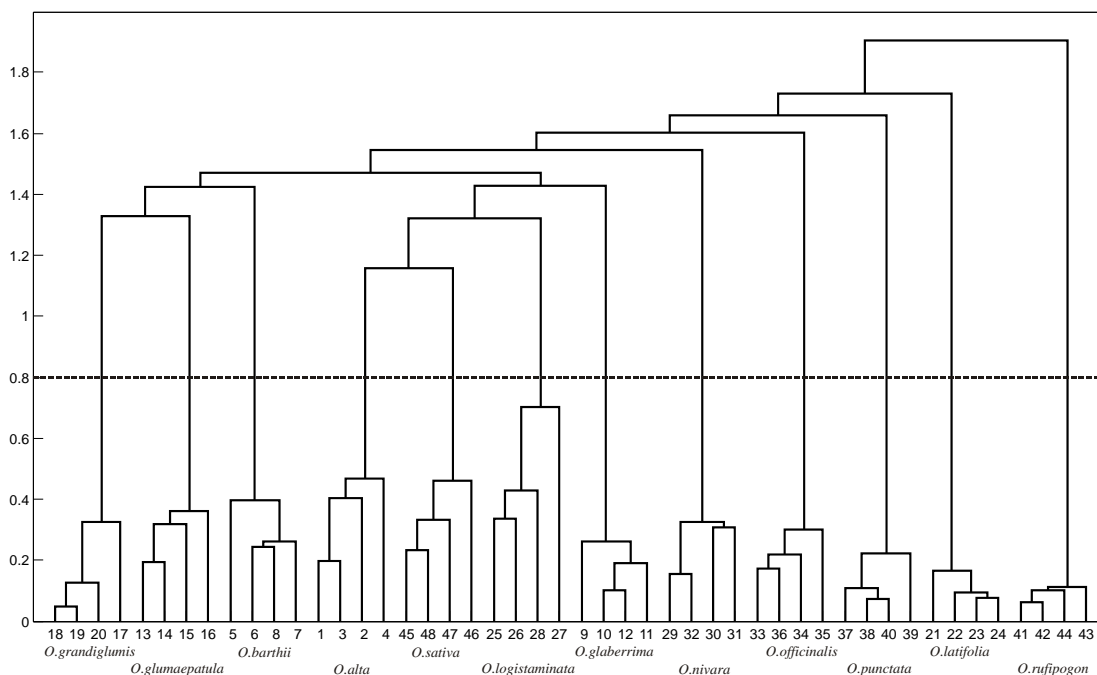


Figura 7.28: Inferência evolutiva das espécies de arroz a partir da análise de formas com a transformada de *wavelets*.

A análise do dendrograma indica grande separabilidade das espécies, uma vez, que no ponto de corte $\cong 0,8$ todas as doze espécies são corretamente classificadas. Embora, não foi possível correlacionar os dados morfométricos e filogenéticos, os resultados obtidos com a análise de agrupamento são uma contribuição importante para o estudo das espécies. A dificuldade de discriminação das espécies pode ser facilitada pelos métodos de análise de imagens.

7.6 Discussão

As técnicas de análise de imagens utilizadas nos experimentos produziram um grande conjunto de dados a respeito das características morfométricas das espécies. Associado aos

métodos de reconhecimento de padrões, os resultados demonstraram grande potencialidade na classificação das espécies. A Tabela 7.32 sintetiza os principais resultados obtidos com a classificação das espécies por meio das características morfométricas. As informações estatísticas apresentadas são: porcentagem de acerto (%A), porcentagem de erro (%E), acurácia (AC) e *F-measure* (F).

Tabela 7.32: Principais resultados da análise morfométrica das espécies biológicas.

Técnicas de Análise de Imagens	Maracujá (<i>Passiflora</i>)				Eucalipto (<i>Eucalyptus</i>)				Arroz (<i>Oryza</i>)			
	%A	%E	AC	F	%A	%E	AC	F	%A	%E	AC	F
Análise de Formas												
Assinatura Digital	89,75	10,25	98,43	88,94	72,09	27,91	94,42	61,29	61,54	38,46	93,59	62,27
Curvatura	84,09	15,91	98,25	88,94	55,81	44,19	91,16	53,76	27,69	72,31	87,95	27,51
Descritores Fourier	88,64	11,36	98,25	87,10	65,12	34,88	93,02	49,58	57,31	42,69	92,95	54,44
Dimensão Fractal Multiescala	89,32	10,68	98,25	87,91	65,21	34,79	93,02	53,67	63,31	36,69	93,59	51,29
Redes Complexas	93,18	6,82	98,95	92,39	79,07	20,93	95,81	70,70	57,69	42,31	93,59	56,38
Wavelets	92,88	7,12	98,78	92,17	72,33	27,67	95,30	68,77	50,00	50,00	91,67	46,67
Análise de Estrut. Tubulares												
Biometria	90,08	9,92	98,20	93,12								
Assinatura Fractal	94,22	5,78	98,95	95,52								
Redes Complexas	91,74	8,26	98,50	90,11								
Análise de Textura												
Matriz de Coocorrência	71,33	28,67	95,59	70,39								
Assinatura Fractal	87,59	12,41	98,09	84,69								
Redes Complexas	68,44	31,56	95,14	64,44								

Os experimentos realizados para análise e classificação de espécies biológicas foram concentrados em três informações de interesse: forma, estrutura tubular e textura. Os resultados apresentaram individualmente a potencialidade de cada abordagem na solução de problemas biológicos. Além disso, as metodologias foram combinadas com as técnicas de seleção de características.

7.6.1 Análise de Formas

Analisando os resultados obtidos com as técnicas de análise de formas, é possível concluir que todas as características morfométricas são adequadas para classificação das espécies de maracujás. Todos os melhores resultados para essas espécies foram superiores a 84%. O desempenho da classificação das espécies de eucalipto foi motivante, mesmo considerando

que o melhor resultado obtido foi inferior a 80%. Para essas espécies, é importante observar que os exemplares híbridos prejudicaram o processo de classificação, como foi demonstrado na Figura 7.27. Considerando apenas as espécies puras, o resultado do pode ser considerado excelente, no qual 100% das espécies foram corretamente classificadas. A complexidade da discriminação das espécies de arroz foi demonstrada por meio dos resultados obtidos. Embora com 63,31% de classificações corretas, é importante salientar que essas espécies possuem poucas informações de interesse que podem ser exploradas.

Considerando as técnicas utilizadas para análise de formas, os resultados obtidos com as redes complexas foram promissores. Esta metodologia demonstrou ser muito eficaz para as espécies de maracujá e eucalipto. Além disso, é importante destacar que para as espécies de maracujá, dentre os 82 experimentos realizados com as redes complexas, mais de 60 experimentos alcançaram taxa de acerto (classificações corretas) superior a 80%.

A transformada de *wavelets* também demonstrou grande aplicabilidade na classificação das espécies de maracujá, principalmente com a família *Daubechies* e com os coeficientes de tendência. Os experimentos com a transformada de *wavelets*, em que foram analisadas diversas funções e níveis, realizaram uma análise detalhada do contorno das espécies biológicas. Outro fato importante é que o dendrograma das espécies de arroz, demonstrando a correta classificação de cada uma das amostras, foi construído a partir da transformada de *wavelets*. Além disso, baseado no coeficiente de correlação de *Matthews*, a transformada de *wavelets* foi a melhor metodologia para classificação das espécies de arroz.

A dimensão fractal multiescala calculada a partir dos pontos de contorno, demonstrou resultados muito promissores na classificação das espécies. As curvas fractais multiescala, geradas com os dados morfométricos, representam uma assinatura unidimensional das espécies. Nos experimentos realizados, todos os pontos da curva foram considerados na classificação das espécies, no entanto, abordagens descritas na literatura demonstram que a seleção de pontos específicos também pode trazer resultados satisfatórios (Plotze, Falvo *et al.*, 2005). No entanto, a seleção de pontos específicos pode ser trabalhosa, uma vez que métodos automáticos para seleção desses pontos são inviáveis.

Os demais descritores utilizados para análise de formas, assinatura digital, curvatura e descritores de Fourier, são metodologias tradicionais para este tipo de problema. Os resultados obtidos comprovaram a aplicabilidade dessas metodologias também em espécies de maracujá, eucalipto e arroz. No entanto, é importante notar que a curvatura não pode ser

considerada uma técnica efetiva para análise das espécies de arroz, em virtude do resultado ruim do processo de classificação.

Analisando os dados estatísticos, a acurácia das técnicas de análise de formas foi superior a 87%. Para as espécies de *Passiflora*, por exemplo, esse valor foi superior a 98%. Esta informação representa a proporção de classificações corretas, tanto verdadeiros positivos quanto verdadeiros negativos. Além disso, a *F-measure* também indica a potencialidade das metodologias na classificação das espécies, uma vez que as técnicas utilizadas no gênero *Passiflora* foram superiores a 87%. Estatisticamente, esta medida é utilizada para representar com um único valor a relação entre precisão e revocação (sensibilidade). Os resultados podem ser interpretados como a média harmônica entre esses dois valores, e os melhores resultados são próximos a 100%, enquanto os piores são próximos a 0%.

7.6.2 Análise de Estruturas Tubulares

As técnicas de análise de estruturas tubulares foram aplicadas nas espécies de *Passiflora*. Embora o conjunto de amostras utilizado nos experimentos foi restrito, os resultados obtidos sinalizam a eficiência das metodologias na taxonomia das espécies.

Dentro deste contexto, três metodologias foram propostas: (i) um algoritmo para extração automática de medidas biométricas; (ii) a assinatura fractal a partir de pontos biométricos (bifurcações, pontos extremos e a combinação dos pontos) e (iii) a geração de redes complexas utilizando os pontos biométricos. Essas metodologias têm como base as informações utilizadas por taxonomistas nos protocolos de classificação das espécies. Com isso, são indicadas para apoiar o processo de taxonomia das espécies.

A assinatura fractal multiescala calculada a partir de pontos biométricos foi a metodologia que obteve a melhor classificação dentre todas as técnicas utilizadas. Esse resultado foi alcançado utilizando os pontos extremos das espécies. Outro fato importante é que a assinatura fractal pode ser utilizada em outros tipos de análises de estruturas tubulares. Além disso, essa metodologia pode ser aplicada para transformação de esqueletos, resultantes do processo de afinamento, em assinaturas unidimensionais.

As redes complexas também alcançaram resultados motivantes com as estruturas tubulares. A porcentagem de classificações corretas foi superior a 90%. Essa metodologia também pode ser utilizada em qualquer tipo de problema envolvendo estruturas tubulares,

inclusive na análise de esqueletos. As curvas geradas a partir do coeficiente de aglomeração das redes complexas representam uma alternativa para visualização do comportamento da rede.

As medidas estatísticas de acurácia e *F-measure* comprovam a potencialidade das metodologias na classificação das espécies. Para todas as técnicas a acurácia foi superior a 98%, enquanto a *F-measure* foi superior a 90%.

7.6.3 Análise de Textura

As técnicas de análise de textura foram empregadas nas espécies de *Passiflora*. Analisando os resultados obtidos, essas metodologias alcançaram um desempenho regular, uma vez que algumas técnicas classificaram corretamente $\cong 70\%$ das espécies.

Para a análise de textura, duas novas abordagens foram propostas utilizando como base a assinatura fractal e as redes complexas. Os resultados obtidos com a assinatura fractal foram motivantes e alcançaram uma taxa de acerto superior a 87%. O desempenho da abordagem com redes complexas foi regular, e obteve 68,44% de classificações corretas.

A análise de textura de imagens biológicas é um procedimento complexo, uma vez que está diretamente relacionada as características morfológicas das espécies. Para as imagens de arroz, por exemplo, a análise de textura não pode ser empregada, devido às características intrínsecas das imagens dessas espécies.

Embora todas as técnicas utilizadas não tenham alcançado um resultado motivante, é importante observar que a assinatura fractal pode ser indicada para análise de textura. Além disso, as medidas estatísticas, tanto de acuraria quanto de *F-measure*, demonstram a viabilidade dessas metodologias no estudo de espécies biológicas.

7.6.4 Seleção de Características

Para uma análise mais detalhada das informações morfométricas das espécies, as técnicas de seleção de características foram empregadas com o propósito de escolher os atributos com maior potencialidade discriminante dentre todas as técnicas utilizadas.

Para seleção das características foram consideradas a primeira componente (PC1) e a variância total, resultantes da análise de componentes principais (PCA). Assim, as características morfométricas foram escolhidas para o *conjunto de seleção* a partir desses critérios adotados.

Com as técnicas de análise de formas, a seleção de características alcançou resultados promissores. Para as espécies de *Passiflora* a taxa de acerto foi igual a 88,07%, para as espécies de eucalipto 61,16% e arroz igual a 54,62%. Um fato importante que deve ser destacado é que o resultado das espécies de *Passiflora* foi obtido com apenas 20 características selecionadas, dentre todas as metodologias utilizadas para análise de formas. Com isso, o conjunto de dados foi significativamente reduzido.

Para a análise de estruturas tubulares a seleção de características alcançou uma taxa de acerto igual a 94,21%, enquanto para a análise de textura o resultado foi igual a 85,63%. Nas duas abordagens os resultados foram muito motivantes.

Um experimento importante foi realizado combinando as características morfométricas extraídas. Neste experimento foi possível confrontar a potencialidade de cada metodologia de análise de imagens. Agrupando os dados morfométricos de forma e estrutura tubular a taxa de acerto na classificação das espécies foi de 93,25%. Para o agrupamento entre dados de forma e textura o acerto foi igual a 61,55%. Com os dados de estrutura tubular e textura a taxa de acerto foi de 91,74%. E, finalmente, combinando as três metodologias (forma, estrutura tubular e textura), as classificações corretas foram iguais a 93,39%. Estes resultados indicam que as características morfométricas de forma e estrutura tubular são essenciais no processo de análise e classificação das espécies.

7.7 Considerações Finais

Neste capítulo foram apresentados os resultados dos experimentos realizados para análise e classificação de espécies vegetais. Os experimentos foram realizados com espécies biológicas reais de maracujá, eucalipto e arroz. Os resultados foram descritos em função das técnicas de análise de imagens empregadas: análise de formas, análise de estruturas tubulares e análise de textura. Para apresentação dos resultados foram utilizados dados estatísticos de porcentagens de acerto e erro, além de diversas medidas comumente utilizadas em reconhecimento de padrões, tais como: porcentagem de acerto, porcentagem de erro, acurácia, dentre outras. Os

gráficos ROC auxiliaram na visualização dos resultados. As informações morfométrica foram analisadas com as técnicas de reconhecimento de padrões: DM, LDA, SRDA e RNA. Além disso, técnicas de seleção de características baseadas em PCA forneceram dados para classificação das espécies. Os resultados obtidos demonstraram grande aplicabilidade das informações morfométricas, obtidas por visão artificial, na análise e classificação de espécies biológicas. Outro fato importante foi observado com a correlação dos dados morfométricos e moleculares, em que foi comprovada a correspondência entre as duas abordagens.

8

Conclusão

A investigação da história taxonômica e evolutiva dos organismos é a principal fronteira das pesquisas científicas em diversas áreas do conhecimento. A biodiversidade das espécies associada a grande variabilidade de características taxonômicas, torna essa tarefa um verdadeiro desafio. Os métodos tradicionais, baseados na inspeção visual das espécies por taxonomistas treinados, estão totalmente ultrapassados. Atualmente, modernas técnicas de análise filogenética, que utilizam características morfométricas e moleculares, têm contribuído intensamente para reconstrução da árvore evolutiva das espécies. A análise filogenética baseada em dados morfológicos utiliza avançadas técnicas de visão artificial para extração de características das espécies. Além disso, o formalismo dos métodos computacionais contribui potencialmente para a descrição e modelagem da história evolutiva das espécies.

Dentre deste contexto, o objetivo deste trabalho foi estudar e desenvolver modelos computacionais capazes de extrair e analisar características morfométricas de espécies biológicas. Para isso, técnicas de análise de imagens foram utilizadas para avaliar a variabilidade das espécies em função de três informações de interesse: forma, estrutura tubular e textura.

Nas técnicas de análise de formas, foram consideradas as seguintes metodologias: assinatura digital, curvatura, descritores de Fourier, dimensão fractal multiescala, redes complexas e a transformada de *wavelets*. Essas técnicas produziram um conjunto detalhado de informações a respeito do contorno das imagens biológicas. As técnicas de análise de

estruturas tubulares foram utilizadas para examinar o sistema de venação foliar, também conhecido como nervuras. Neste grupo, diversas metodologias foram propostas para descrição do sistema de venação, as quais foram baseadas em protocolos de taxonomia. Esse fato foi muito importante para demonstrar a correlação entre as áreas, e como é possível aplicar conceitos biológicos no desenvolvimento de sistemas de visão artificial. Além disso, é importante destacar que essas metodologias são adequadas para qualquer tipo de estrutura tubular, tais como esqueletos, vasos sanguíneos, etc.

Para análise de estruturas tubulares foram desenvolvidas técnicas capazes de coletar automaticamente informações biométricas das espécies, as quais são capazes de navegar ao longo da estrutura. A assinatura fractal multiescala a partir de pontos biométricos foi apresentada como uma nova proposta para investigação morfométrica das espécies. Essa abordagem considera pontos biométricos como informações de interesse, e com base nessas informações determina a complexidade da estrutura. Outra metodologia proposta associa a teoria de redes complexas aos pontos biométricos, a qual transforma os pontos biométricos em vértices da rede. Essa metodologia demonstrou grande aplicabilidade na discriminação das espécies biológicas. É importante também destacar que uma nova metodologia para segmentação do sistema de venação foliar foi proposta, baseada em um processo químico de clarificação e técnicas de visão artificial. Essa metodologia foi desenvolvida em parceria com o Instituto Agrônomo de Campinas (IAC).

Na análise de textura foi implementado o método baseado em matrizes de coocorrência. Esta técnica foi utilizada como referência para comparação dos resultados com as abordagens desenvolvidas. Para análise de textura, duas metodologias foram propostas fundamentadas na distribuição de níveis de cinza da imagem biológica. Essas metodologias utilizam como base subconjuntos de pixels da imagem, que são selecionados a partir de um limiar aplicado a distribuição de probabilidade de níveis de cinza. Em seguida, a assinatura fractal multiescala é aplicada aos pixels selecionados e a complexidade da textura é estimada. Seguindo este mesmo modelo, as redes complexas foram utilizadas para mapear os pixels selecionados em vértices da rede. Com isso, diversas informações foram coletadas sobre a evolução da rede complexa.

Para avaliação das técnicas de análise de imagens foram utilizados três conjuntos de espécies biológicas de maracujá, eucalipto e arroz. Essas espécies foram obtidas por meio de parceiras com a ESALQ/USP, IAC e Embrapa. Além disso, os colaboradores envolvidos

foram fundamentais nas discussões biológicas inerentes ao trabalho. A multidisciplinaridade é intrínseca, uma vez que o trabalho está inserido na fronteira entre as áreas de visão artificial, morfometria e biologia.

Diversos experimentos foram realizados utilizando os conjuntos de dados extraídos com as técnicas de análise de imagens. Esses dados constituem em uma detalhada descrição morfométrica das espécies. Para avaliação da qualidade das informações coletadas, técnicas de reconhecimento de padrões foram empregadas, visando a classificação das espécies.

Os resultados obtidos demonstraram que as técnicas de análise de formas são adequadas para a análise e classificação de espécies biológicas. Dentre as técnicas utilizadas, as redes complexas e a transformada de *wavelets* merecem destaque, uma vez que alcançaram altas taxas de classificações corretas. É importante destacar que algumas espécies, como por exemplo, os maracujás do gênero *Passiflora*, apresentaram grande semelhança morfométrica e molecular. Porém, os métodos computacionais foram capazes de identificar esses padrões de variabilidade. Outro fato que deve ser considerado é dificuldade para discriminação das espécies de eucalipto, uma vez que o conjunto utilizado é formado por espécies puras e híbridas. As espécies de arroz, também reforçaram a dificuldade de classificação das espécies, principalmente devido a qualidade das imagens.

As técnicas de análise de estrutura tubulares demonstraram grande potencialidade na discriminação das espécies. Os melhores resultados foram obtidos por essas técnicas, as quais são fundamentadas em rigorosos protocolos de taxonomia. As abordagens propostas neste trabalho, tanto baseadas em assinatura fractal multiescala quanto em redes complexas, realizaram satisfatoriamente a análise e classificação das espécies. Os métodos de análise de textura também sinalizaram a viabilidade dessas técnicas na investigação de problemas biológicos. Os resultados foram adequados para o estudo e contribuíram apropriadamente para a taxonomia das espécies.

A inferência evolutiva das espécies foi um importante resultado obtido com os dados morfométricos, tanto a partir da forma, quanto da estrutura tubular e textura. Para as espécies de *Passiflora*, a reconstrução da história evolutiva dos indivíduos foi comprovada pela correlação entre os dados morfométricos e moleculares. Testes estatísticos indicaram a correspondência entre as duas abordagens, além disso, os dendrogramas gerados com os dados morfométricos representam grande similaridade com a classificação botânica atualmente aceita para o gênero *Passiflora*. Para as espécies de eucalipto, a análise de

agrupamentos demonstrou a separabilidade entre as espécies puras e híbridas. Adicionalmente, outro fato importante foi o agrupamento das espécies híbridas com ancestrais comuns. Com as espécies de arroz, apesar da classificação não alcançar resultados superiores a 80% de acerto, a análise de agrupamentos produziu dendrogramas nos quais espécies são corretamente discriminadas. Este fato indica a potencialidade das metodologias na taxonomia das espécies de arroz.

Neste trabalho foi demonstrado que as técnicas de visão artificial são comprovadamente importantes para o estudo taxonômico e evolutivo de espécies biológicas. As técnicas computacionais investigadas e desenvolvidas contribuíram para o entendimento da árvore evolutiva, e realizaram um detalhamento minucioso das características morfométricas das espécies. Os desafios impostos ao longo do trabalho estão intimamente ligados a multidisciplinaridade das áreas envolvidas. Este fato indica que a simbiose entre áreas de pesquisas necessita de participações colaborativas entre pesquisadores, o que neste trabalho foi um fator fundamental. Como consequência, o trabalho traz resultados importantes para as áreas envolvidas e fortalece o comprometimento da área de computação no progresso das pesquisas na ciência contemporânea.

8.1 Contribuições

Para desenvolvimento deste trabalho foram estudadas e desenvolvidas técnicas de visão artificial com o objetivo de analisar e classificar espécies biológicas. A multidisciplinaridade inerente resultou em contribuições para as áreas envolvidas, o que demonstra a aplicabilidade das técnicas computacionais em outras áreas. As principais contribuições originárias deste trabalho foram:

- desenvolvimento (e implementação) de uma metodologia semi-automática para segmentação do sistema de venação foliar. Esta técnica associa o processo químico de diafanização às técnicas de visão artificial para segmentação da estrutura de venação (Plotze, Nardin *et al.*, 2006);
- desenvolvimento (e implementação) de um algoritmo automático para extração de medidas biométrica de estruturas tubulares. Esta metodologia pode ser aplicada para qualquer tipo de estrutura e permite a coleta de diversas informações biométricas (Plotze e Bruno, 2009);

- proposta de uma nova metodologia para extração de medidas morfométricas de estruturas tubulares baseada na assinatura fractal multiescala e pontos biométricos. Os resultados demonstraram a eficácia da metodologia na análise e classificação de espécies de *Passiflora* (Plotze e Bruno, 2007);
- proposta de uma nova metodologia que associa as redes complexas as características biométricas das estruturas tubulares. Esta abordagem transforma os pontos biométricos em vértices da rede, e é utilizada para extração de medidas morfométricas;
- proposta de uma nova metodologia para análise de textura baseada na distribuição de probabilidade de níveis de cinza e na assinatura fractal multiescala. Esta abordagem alcançou resultados promissores na classificação das espécies;
- proposta de uma nova metodologia que utiliza as redes complexas associada a distribuição de probabilidade de cinza para realizar a análise de textura;
- estudo taxonômico e evolutivo de espécies de maracujá do gênero *Passiflora* por meio de características morfométricas de forma, estrutura tubular e textura. Os resultados foram correlacionados com dados moleculares e indicaram a correspondência entre as abordagens;
- estudo detalhado da dimensão fractal multiescala e suas aplicações na classificação de espécies biológicas (Bruno, Plotze *et al.*, 2008);
- investigação abrangente das técnicas morfométricas visando a análise e classificação de espécies biológicas, com contribuições para os gêneros *Passiflora*, *Eucalyptus* e *Oryza*;
- indícios de inferência evolutiva de espécies de eucalipto e arroz por meio de características morfométricas. Este fato será pesquisado em trabalhos futuros.
- desenvolvimento de técnicas morfométricas que podem ser utilizadas por taxonomistas para auxiliar a discriminação de espécies

8.2 Trabalhos futuros

A multidisciplinaridade do trabalho permite que diversos trabalhos sejam realizados no futuro envolvendo as áreas de visão artificial, morfometria e biologia. Neste sentido, algumas investigações são propostas:

- análise de estrutura tubulares de outros de tipos de espécies biológicas;
- mineração do conjunto de dados morfométricos por meio de novas técnicas de seleção de características;
- estudo evolutivo das espécies de eucalipto do gênero *Eucalyptus*;
- estudo evolutivo das espécies de arroz do gênero *Oryza*;
- aplicação dos métodos desenvolvidos de análise de textura em outros tipos de imagens, tanto biológicas quanto artificiais;
- aplicação das abordagens desenvolvidas para análise de estruturas tubulares na extração de características de esqueletos resultantes da técnica de afinamento (*thinning*);

Referências Bibliográficas

- (Adams e Rohlf, 2000) ADAMS, D. C.; ROHLF, F. J. Ecological Character Displacement in *Plethodon*: Biomechanical Differences Found from a Geometric Morphometric Study. **Proceeding of the National Academy of Sciences - PNAS**, v. 97, n. 8, p. 4106-4111, 2000.
- (Adams, Rohlf *et al.*, 2004) ADAMS, D. C.; ROHLF, F. J.; SLICE, D. E. Geometric Morphometrics: Ten Years of Progress Following the Revolution. **Italian Journal of Zoology**, v. 71, n. 1, p. 5-16, 2004.
- (Albert e Barabási, 2002) ALBERT, R.; BARABÁSI, A. L. Statistical Mechanics of Complex Networks. **Reviews of Modern Physics**, n. 74, p. 48-98, 2002.
- (Amorin, 2002) AMORIN, D. D. S. **Fundamentos da Sistemática Filogenética**. Ribeirão Preto: Holos, 2002.
- (Antiqueira, Nunes *et al.*, 2007) ANTIQUEIRA, M.; NUNES, M. G. V.; OLIVEIRA, O. N.; COSTA, L. F. Strong Correlations Between Text Quality and Complex Network Features. **Physica A**, n. 373, p. 811-820, 2007.
- (Arques e Gilbert, 2002) ARQUES, J. M.; GILBERT, J. Fractal Analysis and Its Applications in Human Palaentology. **Human Evolution**, v. 17, n. 1, p. 45-53, 2002.
- (Attneave, 1954) ATTNEAVE, F. Some Informational Aspects of Visual Perception. **Psychological Review**, v. 1, n. 61, p. 183-193, 1954.
- (Backes, 2006) BACKES, A. R. **Implementação e Comparação de Estimativa da Dimensão Fractal e sua Aplicação à Análise e Processamento de Imagens**. (Mestrado). Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 83 p., 2006.
- (Backes e Bruno, 2009) BACKES, A. R.; BRUNO, O. M. Shape Classification using Complex Network and Multi-Scale Fractal Dimension. **Pattern Recognition Letters**, n. (doi:10.1016/j.patrec.2009.08.007), 2009.
- (Backes, Casanova *et al.*, 2008) BACKES, A. R.; CASANOVA, D.; BRUNO, O. M. A Complex Network-Based Approach for Boundary Shape Analysis. **Pattern Recognition**, n. 42, p. 54-67, 2008.

- (Baldi, Brunak *et al.*, 2000) BALDI, P.; BRUNAK, S.; CHAUVIN, Y.; ANDERSEN, C. A. F.; NIELSEN, H. Assessing the Accuracy for Prediction Algorithms for Classification: An Overview. **Bioinformatics**, v. 16, p. 412-424, 2000.
- (Belhumeur, Hespanha *et al.*, 1997) BELHUMEUR, J.; HESPANHA, P.; KRIEGMAN, D. J. Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 7, p. 711-720, 1997.
- (Bockhoff, 2001) BOCKHOFF, R. C. Plant Morphology: The Historic Concepts of Wilhelm Troll, Walter Zimmermann and Agnes Arber. **Ann. Botany**, v. 88, p. 1153-1172, 2001.
- (Boggess e Narcowich, 2001) BOGGESS, A.; NARCOWICH, F. J. **A First Course in Wavelets with Fourier Analysis**. New Jersey: Prentice Hall, 2001.
- (Bookstein, 1991) BOOKSTEIN, F. L. **Morphometric Tools for Landmark Data: Geometry and Biology**. New York: Cambridge University Press, 1991.
- (Braga, Carvalho *et al.*, 2000) BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro: LTC, 2000.
- (Brooker, 2000) BROOKER, M. I. H. A New Classification of the genus Eucalyptus L 'Hér. (Myrtaceae). **Australian Systematic Botany**, v. 13, n. 1, p. 79-148, 2000.
- (Brummitt, 1997) BRUMMITT, R. K. Taxonomy versus Cladonmy, a Fundamental Controversy in Biological Systematics. **Taxon**, v. 46, n. 1, p. 723-734, 1997.
- (Bruno, Plotze *et al.*, 2008) BRUNO, O. M.; PLOTZE, R. O.; FALVO, M.; CASTRO, M. Fractal Dimension Applied to Plant Identification. **Information Science**, v. 178, p. 2722-2733, 2008.
- (Buso, 1998) BUSO, G. S. C. **Análise genética de espécies silvestres de arroz (Oryza spp.) nativas do Brasil: estrutura de populações, diversidade genética e relações filogenéticas utilizando marcadores moleculares** (Doutorado). Embrapa Recursos Genéticos e Biotecnologia, Empresa Brasileira de Pesquisa Agropecuária, Brasília, 80 p., 1998.
- (Cai, He *et al.*, 2007) CAI, D.; HE, X.; HAN, J. Spectral Regression: A Unified Approach for Sparse Subspace Learning. Int. Conf. on Data Mining (ICDM'07), Omanha. 2007.
- (Cai, He *et al.*, 2007) CAI, D.; HE, X.; HAN, J. Spectral Regression: A Unified Subspace Learning Framework for Content-Based Image Retrieval. ACM Multimedia 2007, Germany. 2007.
- (Cai, He *et al.*, 2008) CAI, D.; HE, X.; HAN, J. SRDA: An Efficient Algorithm for Large Scale Discriminant Analysis. **IEEE Transactions on Knowledge and Data Engineering**, v. 20, n. 1, p. 1-12, 2008.

- (Campbell e Robson, 1968) CAMPBELL, F. W.; ROBSON, J. G. Application of Fourier Analysis to the Visibility of Gratings. **Journal of Physiology**, v. 1, n. 197, p. 551-556, 1968.
- (Campos, 2001) CAMPOS, T. E. **Técnicas de Seleção de Características com Aplicações em Reconhecimento de Faces**. (Mestrado). Instituto de Matemática e Estatísticas, Universidade de São Paulo, São Paulo, 120 p., 2001.
- (Canny, 1986) CANNY, J. A Computational Approach to Edge Detection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 8, n. 6, p. 679-698, 1986.
- (Carlin, 2000) CARLIN, M. Measuring the Complexity of Non-Fractal Shapes by a Fractal Method. **Pattern Recognition Letters**, v. 21, p. 1013-1017, 2000.
- (Casanova, 2008) CASANOVA, D. **Análise de Textura Aplicada no Reconhecimento de Órgãos Foliáres**. (Mestrado). Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 90 p., 2008.
- (Castañón, 2006) CASTAÑÓN, C. A. B. **Análise e Reconhecimento Digital de Formas Biológicas para o Diagnóstico Automático de Parasitas do Gênero Eimeria**. (Doutorado). Universidade de São Paulo, São Paulo, 161 p., 2006.
- (Castleman, 1996) CASTLEMAN, K. R. **Digital Image Processing**. New York: Prentice-Hall, 1996.
- (Castrejón-Pita, Sarmiento-Galán *et al.*, 2005) CASTREJÓN-PITA, A. A.; SARMIENTO-GALÁN, A.; CASTREJÓN-PITA, J. R.; CASTREJÓN-GARÍCA, R. Fractal Dimension in Butterflies' Wings: A Novel Approach to Understanding Wing Patterns? **Journal of Mathematical Biology**, v. 50, n. 1, p. 584-594, 2005.
- (Cesar, 1997) CESAR, R. M. **Análise Multi-escala de Formas Bidimensionais**. mestrado Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, p., 1997.
- (Clarke, Barman *et al.*, 2006) CLARKE, J.; BARMAN, S.; REMAGNINO, P.; BAILEY, K.; KIRKUP, D.; MAYO, S.; WILKIN, P. Venation Pattern Analysis of Leaf Images In: COMPUTING, A. I. V., International Symposium on Visual Computing, Lake Tahoe, Nevada. 2006.
- (Costa, Campos; *et al.*, 2001) COSTA, L. F.; CAMPOS, A. G.; MANOEL, E. T. M. An Integrated Approach to Shape Analysis: Results and Perspectives. International Conference on Quality Control by Artificial Vision, Le Creusot, France. p. 23-34, 2001.
- (Costa e Cesar, 2000) COSTA, L. F.; CESAR, R. M. **Shape Analysis and Classification: Theory and Practice**. Pennsylvania: CRC Press, 2000.
- (Costa, Rodrigues *et al.*, 2007) COSTA, L. F.; RODRIGUES, F. A.; TRAVIESO, G.; VILLAS BOAS, P. R. Characterization of Complex Networks: A Survey of Measurements. **Advances in Physics**, v. 56, n. 1, p. 167-242, 2007.

- (Darwin, 1859) DARWIN, C. R. **On the Origin of Species by Means of Natural Selection**. London: J. Murray, 1859.
- (Delson, Harvati *et al.*, 2001) DELSON, E.; HARVATI, K.; REDDY, D.; MARCUS, L. F.; MOWBRAY, K.; SAWYER, G. J.; JACOB, T.; MÁRQUES, S. The Sambungmacan 3 Homo erectus Calvaria: A Comparative Morphometric and Morphological Analysis. **The Anatomical Record**, v. 262, n. 1, p. 380-397, 2001.
- (Diniz-Filho, 2000) DINIZ-FILHO, J. A. F. **Métodos Filogenéticos Comparativos**. Ribeirão Preto: Holos, 2000.
- (Everitt, Landau *et al.*, 2001) EVERITT, B. S.; LANDAU, S.; LEESE, M. **Cluster Analysis**. 4th. New York: Oxford University Press Inc, 2001.
- (Falconer, 1990) FALCONER, K. **Fractal Geometry: Mathematical Foundations and Applications**. New York: John Wiley & Sons, 1990.
- (Falconer, 1997) FALCONER, K. **Techniques in Fractal Geometry**. New York: John Wiley & Sons, 1997.
- (Falvo, 2005) FALVO, M. **Identificação de Espécies Vegetais por meio da Análise do Contorno Foliar - Uma Abordagem Bio-Inspirada**. (Mestrado). Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 130 p., 2005.
- (Fawcett, 2006) FAWCETT, T. An Introduction to ROC Analysis. **Pattern Recognition Letters**, v. 27, p. 861-874, 2006.
- (Fischer, Shahbazkis *et al.*, 2002) FISCHER, S.; SHAHBAZKIS, H. R.; BUNKE, H. **Automatic Diatom Identification**. Singapore: World Scientific Publishing, 2002.
- (Freeman e Skapura, 1991) FREEMAN, J. A.; SKAPURA, D. M. **Neural Networks: Algorithms, Applications and Programming Techniques**. New York: Addison-Wesley, 1991.
- (Frieb, 2003) FRIEB, M. An Application of the Relative Warps Analysis to Problems in Humam Paleontology - With Notes on Raw Data Quality. **Image Anal Stereol**, v. 22, n. 1, p. 63-72, 2003.
- (Fukunaga, 1990) FUKUNAGA, K. **Introduction to Statistical Pattern Recognition (2nd Edition)**. San Diego: Academic Press, 1990.
- (Futuyma, 2002) FUTUYMA, D. J. **Biologia Evolutiva**. 2nd. Ribeirão Preto: Funpec, 2002.
- (Golub e Loan, 1996) GOLUB, G. H.; LOAN, C. F. V. **Matrix Computations (3rd Edition)**. Baltimore: Johns Hopkins University Press, 1996.
- (Gonzalez e Woods, 2001) GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 2nd. New Jersey: Prentice Hall, 2001.

- (Grossmann e Morlet, 1984) GROSSMANN, A.; MORLET, J. Decomposition of Hardy Functions into Square Interegrable Wavelets of Constant Shape. **SIAM Journal of Mathematical Analysis**, v. 15, n. 4, p. 723-736, 1984.
- (Hair, Anderson *et al.*, 1998) HAIR, J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. **Multivariate Data Analysis**. Fifth Edition. New Jersey: Prentice Hall, 1998.
- (Haralick, Shanmugam *et al.*, 1973) HARALICK, R.; SHANMUGAM, K.; DINSTEN, I. Textural Features for Image Classification. **IEEE Transactions on Systems, Man, Cybernetics**, v. 3, n. 1, p. 610-621, 1973.
- (Hardle e Simar, 2007) HARDLE, W.; SIMAR, L. **Applied Multivariate Statistical Analysis (2nd Edition)**. New York: Springer, 2007.
- (Haykin, 1998) HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. 2nd. Prentice Hall, 1998.
- (Jahne, 2002) JAHNE, B. **Digital Image Processing**. 5th. New York: Sprintg-Verlag, 2002.
- (Jain e Zongker, 1997) JAIN, A.; ZONGKER, D. Feature Selection: Evaluation, Application, and Small Samples Performance. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 19, n. 2, p. 153-158, 1997.
- (Jain, 1989) JAIN, A. K. **Fundamentals of Digital Image Processing**. New Jersey: Prentice Hall, 1989.
- (Jain e Dubes, 1988) JAIN, A. K.; DUBES, R. C. **Algorithms for Clustering Data**. New Jersey: Prentice Hall, 1988.
- (Jain, Duin *et al.*, 2000) JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical Pattern Recognition: A Review. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 1, p. 4-37, 2000.
- (Jain, Mao *et al.*, 1996) JAIN, A. K.; MAO, J.; MOHIUDDIN, K. Artificial Neural Networks: A Tutorial. **Computer**, v. 29, n. 3, p. 31-44, 1996.
- (Junior, 2008) JUNIOR, J. J. D. M. S. **Identificação de Espécies Vegetais por meio de Análise de Imagens Microscópicas de Folhas**. (Mestrado). Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 90 p., 2008.
- (Klingenberg, 2002) KLINGENBERG, C. P. Morphometrics and the Role of the Phenotype in Studies of the Evolution of Developmental Mechanisms. **Gene**, v. 287, n. 1, p. 3-10, 2002.
- (Kohn, 1998) KOHN, A. F. **Reconhecimento de Padrões: Uma Abordagem Estatística**. São Paulo: Universidade de São Paulo, 1998.

- (Lam, Lee *et al.*, 1992) LAM, L.; LEE, S.-W.; SUEN, C. Y. Thinning Methodologies - A Comprehensive Survey. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 14, n. 9, p. 869-885, 1992.
- (Le e Satoh, 2005) LE, D. D.; SATOH, S. An Efficient Feature Selection Method for Object Detection. International Conference on Advances in Pattern Recognition (ICAPR'05), UK. p. 461-468, 2005.
- (Lestrel, 2000) LESTREL, P. E. **Morphometrics for the Life Sciences**. Singapore: World Scientific, 2000.
- (Loncaric, 1998) LONCARIC, S. A Survey of Shape Analysis Techniques. **Pattern Recognition**, v. 31, n. 8, p. 983-1001, 1998.
- (Loy, Mariani *et al.*, 1998) LOY, A.; MARIANI, L.; BERTELLETTI, M.; TUNESI, L. Visualizing Allometry: Geometric Morphometrics in the Study of Shape Changes in the Early Stage of the Two-Banded Sea Bream, *Diplodus vulgaris* (Perciformes, Sparidae). **Journal of Morphology**, v. 237, p. 137-146, 1998.
- (Macleod, 2007) MACLEOD, N. **Automated Taxon Identification in Systematics**. Boca Raton: CRC Press, 2007.
- (Macleod e Forey, 2002) MACLEOD, N.; FOREY, P. L. **Morphology, Shape and Phylogeny**. London: CRC Press, 2002.
- (Mandelbrot, 1977) MANDELBROT, B. B. **Fractals, Form, Chance and Dimension**. San Francisco: W.H. Freeman, 1977.
- (Mandelbrot, 1983) MANDELBROT, B. B. **The Fractal Geometry of Nature**. New York: Freeman, 1983.
- (Mantel, 1967) MANTEL, N. The Detection of Disease Clustering and a Generalized Regression Approach. **Cancer Research**, v. 27, n. 2, p. 209-220, 1967.
- (Marvalid, Sequeira *et al.*, 2002) MARVALID, A. E.; SEQUEIRA, A. S.; O'BRIEN, C. W.; FARREL, B. D. Molecular and Morphological Phylogenetics of Weevils (Coleoptera Curculionoidea): Do Niche Shifts Accompany Diversification? **Systematic Biology**, v. 51, n. 5, p. 761-785, 2002.
- (Matioli, 2001) MATIOLI, S. R. **Biologia Molecular e Evolução**. Ribeirão Preto: Holos, 2001.
- (Mcculloch e Pitts, 1943) MCCULLOCH, W. S.; PITTS, W. A logical calculus of ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, p. 115-133, 1943.
- (McLellan e Endler, 1998) MCLELLAN, T.; ENDLER, J. A. The Relative Success of Some Methods for Measuring and Describing the Shape of Some Complex Objects. **Systematic Biology**, v. 47, n. 2, p. 264-281, 1998.

- (Milner, Raz *et al.*, 2007) MILNER, D.; RAZ, S.; HEL-OR, H.; KEREN, D.; NEVO, E. A New Measure of Symmetry and its Application to Classification of Bifurcating Structures. **Pattern Recognition**, v. 40, p. 2237-2250, 2007.
- (Mingoti, 2005) MINGOTI, S. A. **Análise de Dados Através de Métodos de Estatística Multivariada**. Belo Horizonte: Editora UFMG, 2005.
- (Mingqiang, Kidiyo *et al.*, 2008) MINGQIANG, Y.; KIDIYO, K.; R., J. A Survey of Shape Feature Extractoin Techniques. In: YIN, P. Y. (Ed.). **Pattern Recognition Techniques, Technology and Applications**. Vienna: In-Tech, 2008.
- (Missiaggia, 2005) MISSIAGGIA, A. A. **Mapeamento Genético de QTL para Qualidade da madeira e Florescimento Precoce e Estudos de Expressão Gênica Alelo Específica em Eucalyptus spp.** (Doutorado). Escola Superior Luiz de Queiroz (ESALQ), Universidade de São Paulo, Piracicaba, 236 p., 2005.
- (Mittal, Mital *et al.*, 1999) MITTAL, N.; MITAL, D. P.; CHAN, K. L. Features for Texture Segmentation Using Gabor Filters. IEE Seventh International Conference on Image Processing and Its Applications, p. 353-357, 1999.
- (Mokhtarian e Mackworth, 1992) MOKHTARIAN, F.; MACKWORTH, A. K. A Theory of Multi-Scale Curvature-Based Shape Representation for Planar Curves. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 14, n. 8, p. 789-805, 1992.
- (Monteiro e Reis, 1999) MONTEIRO, L. R.; REIS, S. F. D. **Princípios de Morfometria Geométrica**. Ribeirão Preto: Holos, 1999.
- (Muschner, Lorenz *et al.*, 2003) MUSCHNER, V. C.; LORENZ, A. P.; CERVI, A. C.; BONATTO, S. L.; SOUZA-CHIEZ, T. T.; SALZANO, F. M.; FREITAS, L. B. A First Molecular Phylogenetic Analysis of Passiflora (Passifloraceae). **American Journal of Botany**, v. 90, p. 1229-1238, 2003.
- (Nixon e Aguado, 2002) NIXON, M. S.; AGUADO, A. S. **Features Extraction and Image Processing**. Oxford: Newnes, 2002.
- (Ojala, Pietikainen *et al.*, 1996) OJALA, T.; PIETIKAINEN, M.; HARWOOD, D. A Comparative Study of Texture Measures With Classification based on Featured Distributions **Pattern Recognition**, v. 29, n. 1, p. 51-59, 1996.
- (Oliveira, 2002) OLIVEIRA, G. C. X. **A Molecular Phylogenetic Analysis of Oryza L. based on Chloroplast DNA**. (Doutorado). Washington University, Saint Lous, 100 p., 2002.
- (Ortega, Sola *et al.*, 2008) ORTEGA, G. J.; SOLA, R. G.; PASTOR, J. Complex Network Analysis of Human ECoG Data. **Neuroscience Letters**, n. 447, p. 129-133, 2008.
- (Otsu, 1979) OTSU, N. A Threshold Selection Method from Gray-Level Histograms. **IEEE Transactions on Systems, Man and Cybernetics**, p. 62-66, 1979.

- (Pádua, 2004) PÁDUA, J. G. **Análises Genéticas de Espécies do Gênero Passiflora L. com base em Abordagens Filogenéticas, Morfométricas e em Marcadores Microssatélites.** Doutorado Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 127 p., 2004.
- (Pádua, 2007) PÁDUA, J. G. **Uso de DNA Barcodes, SNP de Ancestralidade e Dimensão Fractal Multiescala para Reconstruir a Introdução e Composição Genética de Clones Híbridos Elite de Eucalyptus no Brasil.** (Relatório Técnico). Embrapa Recursos Genéticos e Biotecnologia, Empresa Brasileira de Pesquisa Agropecuária, Brasília, 32 p., 2007.
- (Pandya, 1995) PANDYA, A. S. **Pattern Recognition with Neural Networks in C++.** Boca Raton: CRC, 1995.
- (Papoulis, 1962) PAPOULIS, A. **The Fourier Integral and Its Applications.** New York: McGraw-Hill, 1962.
- (Parker, 1996) PARKER, J. R. **Algorithms for Image Processing and Computer Vision.** New York: John Wiley & Sons, 1996.
- (Pavlidis, 1977) PAVLIDIS, T. **Structural Pattern Recognition.** Berlim: Springer-Verlag, 1977.
- (Pazoti, 2005) PAZOTI, M. A. **CitruVis - Um Sistema de Visão Artificial para Identificar o Fungo Guignardia Citricarpa, Causador da Mancha Preta (MPC) em Citros.** Mestrado Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, p., 2005.
- (Pedrini e Schwartz, 2008) PEDRINI, H.; SCHWARTZ, W. R. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações.** São Paulo: Thomson Learning, 2008.
- (Pentland, 1985) PENTLAND, A. P. On Describing Complex Surface Shapes. **Image and Vision Computing**, v. 3, n. 4, p. 153-162, 1985.
- (Perona e Malik, 1990) PERONA, P.; MALIK, J. Scale-Space and Edge Detection Using Anisotropic Diffusion. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 12, n. 7, p. 629-639, 1990.
- (Pinto, 2001) PINTO, S. C. D. **Estimação da Dimensão Fractal de Imagens de SPM.** Mestrado Instituto de Física de São Carlos, São Paulo, São Carlos, p., 2001.
- (Plotze, 2004) PLOTZE, R. O. **Identificação de Espécies Vegetais através da Análise da Forma Interna de Órgãos Foliare.** (Mestrado). Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 152 p., 2004.
- (Plotze e Bruno, 2004) PLOTZE, R. O.; BRUNO, O. M. Estudo e Comparação de Algoritmos de Esqueletonização para Imagens Binárias. IV Congresso Brasileiro de Computação (CBCComp), Itajaí. p. 59-64, 2004.

- (Plotze e Bruno, 2005) PLOTZE, R. O.; BRUNO, O. M. Identificação de Espécies Vegetais Através da Análise da Forma Interna de Órgãos Foliares. XVIII Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI), Natal. p. 1-10, 2005.
- (Plotze e Bruno, 2007) PLOTZE, R. O.; BRUNO, O. M. Análise de Formas e Reconhecimento de Padrões por meio da Assinatura Fractal Multiescala. **Journal of Computer Science (Infocomp)**, v. 6, p. 91-101, 2007.
- (Plotze e Bruno, 2009) PLOTZE, R. O.; BRUNO, O. M. Automatic Leaf Structure biometry: Computer Vision Techniques and their Applications in Plant Taxonomy. **International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)**, v. 23, n. 2, p. 247-262, 2009.
- (Plotze, Falvo *et al.*, 2005) PLOTZE, R. O.; FALVO, M.; PÁDUA, J. G.; BERNACCI, L. C.; VIEIRA, M. L. C.; OLIVEIRA, G. C. X.; BRUNO, O. M. Leaf Shape Analysis by the Multiscale Minkowsky Fractal Dimension, a New Morphometric Method: A Study in Passiflora L. (Passifloraceae). **Canadian Journal of Botany**, v. 83, p. 287-301, 2005.
- (Plotze, Nardin *et al.*, 2006) PLOTZE, R. O.; NARDIN, C. F.; BERNACCI, L. C.; BRUNO, O. M. Segmentação do Sistema de Venação Foliar e suas Aplicações na Biometria e Taxonomia das Espécies. II Workshop de Visão Computacional (WVC), São Carlos. p. 37-40, 2006.
- (Pratt, 2001) PRATT, W. K. **Digital Image Processing**. New York: John Wiley and Sons, 2001.
- (Rapini, 2000) RAPINI, A. **Sistemática: Estudos em Asclepiadoideae (Apocynaceae) da Cadeia do Espinhaço de Minas Gerais**. Doutorado Instituto de Biociências, Universidade de São Paulo, São Paulo, 80 p., 2000.
- (Raven e Johnson, 2001) RAVEN, P. H.; JOHNSON, G. B. **Biology**. 6th. McGraw-Hill, 2001.
- (Raykov e Marcoulides, 2008) RAYKOV, T.; MARCOULIDES, G. A. **Introduction to Applied Multivariate Analysis**. New York: Routledge, 2008.
- (Rencher, 2002) RENCHER, A. C. **Methods of Multivariate Analysis (2nd Edition)**. New York: John Wiley & Sons, 2002.
- (Rodrigues, 2007) RODRIGUES, F. A. **Caracterização, Classificação e Análise de Redes Complexas**. (Doutorado). Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 175 p., 2007.
- (Rohlf, 1990) ROHLF, F. J. Morphometrics. **Annual Review of Ecology and Systematics**, v. 21, n. 1, p. 229-316, 1990.
- (Rohlf, 2002) ROHLF, F. J. Geometric Morphometrics in Systematics. In: MACLEOD, N. e FOREY, P. (Ed.). **Morphology, Shape and Phylogenetics**. London: Taylor & Francis, 2002.

- (Rohlf e Marcus, 1993) ROHLF, F. J.; MARCUS, L. F. A Revolution in Morphometrics. **Trends in Ecology and Evolution**, v. 8, n. 4, p. 129-132, 1993.
- (Schroeder, 1996) SCHROEDER, M. **Fractals, Chaos, Power Laws - Minutes from and Infinite Paradise**. New York: W.H. Freeman and Company, 1996.
- (Semmlow, 2004) SEMMLOW, J. L. **Biosignal and Biomedical Image Processing**. New York: Marcel Dekker, 2004.
- (Sezgin e Sankur, 2004) SEZGIN, M.; SANKUR, B. Survey over image thresholding techniques and quantitative performance evaluation. **J. Electron. Imaging**, v. 13, p. 146-165, 2004.
- (Shapiro e Stockman, 2001) SHAPIRO, L. G.; STOCKMAN, G. C. **Computer Vision**. Prentice Hall, 2001.
- (Slice, 2007) SLICE, D. E. Geometric Morphometrics. **Annual Review of Anthropology**, n. 36, p. 261-381, 2007.
- (Sokal e Rohlf, 1995) SOKAL, R. R.; ROHLF, F. J. **Biometry**. New York: Freeman, 1995.
- (Soltis e Soltis, 2003) SOLTIS, P. S.; SOLTIS, D. E. Applying the Bootstrap in Phylogeny Reconstruction. **Statistical Science**, v. 18, n. 2, p. 256-267, 2003.
- (Sonibare, Jayeola *et al.*, 2004) SONIBARE, M. A.; JAYEOLA, A. A.; EGUNYOMI, A. A. Morphometric Analysis of the Genus *Ficus* Linn. (moraceae). **African Journal of Biotechnology**, v. 3, n. 4, p. 229-235, 2004.
- (Sonka, Hlavac *et al.*, 1999) SONKA, M.; HLAVAC, V.; BOYLE, R. **Image Processing Analysis, and Machine Vision**. Pacific Grove: PWS Publishing, 1999. 770
- (Theodoridis e Koutroumbas, 2003) THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**. 2nd. San Diego: Elsevier - Academic Press, 2003.
- (Timm, 2002) TIMM, N. H. **Applied Multivariate Analysis**. New York: Springer, 2002.
- (Torres, Falcão *et al.*, 2004) TORRES, R. D. S.; FALCÃO, A. X.; COSTA, L. F. A Graph-Based Approach for Multiscale Shape Analysis. **Pattern Recognition**, v. 37, p. 1163-1174, 2004.
- (Tricot, 1995) TRICOT, C. **Curves and Fractal Dimension**. New York: Springer-Verlag, 1995.
- (Tuceryan e Jain, 1993) TUCERYAN, M.; JAIN, A. K. Texture Analysis. In: CHEN, C. H.; PAU, L. F., *et al* (Ed.). **Handbook of Pattern Recognition and Computer Vision**. Singapore: World Scientific Publishing, 1993.
- (Turner, Blackledge *et al.*, 1998) TURNER, M. J.; BLACKLEDGE, J. M.; ANDREWS, P. R. **Fractal Geometry in Digital Imaging**. Cambridge: Academic Press, 1998.

- (Vitta e Bernacci, 2004) VITTA, F. A.; BERNACCI, L. C. A New Species and Two Overlooked Species of Passiflora (Passifloraceae) from Brazil. **Brittonia**, n. 56, p. 89-95, 2004.
- (Wang e Chen, 2003) WANG, X. F.; CHEN, G. Complex Networks: Small-World, Scale-Free and Beyond. **IEEE Circuits and Systems Magazine**, v. 3, n. 1, p. 6-20, 2003.
- (Watts e Strogatz, 1998) WATTS, D. J.; STROGATZ, S. H. Collective Dynamics of Small-World Networks. **Nature**, v. 393, n. 6684, p. 400-442, 1998.
- (Webb, 2002) WEBB, A. R. **Statistical Pattern Recognition (2nd Edition)**. New York: John Wiley & Sons, 2002.
- (Wheeler, 2007) WHEELER, Q. D. Digital Innovation and Taxonomy's Finest Hour. In: MACLEOD, N. (Ed.). **Automated Taxon Identification in Systematics**. Boca Raton: CRC Press, 2007.
- (Wiens, 2004) WIENS, J. J. The Role of Morphological Data in Phylogeny Reconstruction. **Systematic Biology**, v. 53, n. 4, p. 653-661, 2004.
- (Wing, Wilf *et al.*, 2001) WING, S.; WILF, P.; JOHSON, K.; HICKEY, L. J.; ELLIS, B.; ASH, A. Manual of Leaf Architecture—Morphological Description and Categorization of Dicotyledonous and Net-veined Monocotyledonous Angiosperms. Leaf Architecture Working Group - Smithsonian Institution, Washington, 2001. Disponível em: < <http://www.yale.edu/peabody/collections/pb/MLA.pdf>. >. Acesso em: agosto/2009.

