*Identificação de espécies vegetais
através da análise da forma interna
de órgãos foliares*

Rodrigo de Oliveira Plotze

Identificação de espécies vegetais através da análise da forma interna de órgãos foliares

Rodrigo de Oliveira Plotze

Orientador: *Prof. Dr. Odemir Martinez Bruno*

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

“VERSÃO REVISADA APÓS A DEFESA”

Data da Defesa:	29/10/2004
Visto do Orientador:	

USP – São Carlos
Dezembro/2004

Aos meus pais, **Roberto e Marilda**, pelo
amor e apoio em todos os momentos
da minha vida.

À **Fernanda**, essa jornada teria sido mais
difícil se não estivéssemos juntos.

Agradecimentos

Inicialmente gostaria de expressar minha sincera admiração pelo **Prof. Odemir Bruno**, e agradecê-lo pela orientação e confiança durante esse período. Obrigado pelos valiosos ensinamentos e principalmente por me mostrar o que é ser um verdadeiro pesquisador.

Agradeço aos colegas que conquistei no ICMC durante esse período. Em especial ao Roberto Platz, Christian Bones, Maurício Falvo, Mario Pazoti, Júlio Torelli, entre outros. Se não fossem vocês e os períodos de descontração esse trabalho nunca chegaria ao fim.

Agradeço também aos professores e funcionários do ICMC-USP. Especialmente a Profa. Agma Traina e Prof. João Batista pelas inúmeras sugestões sobre o desenvolvimento desse trabalho. E as meninas da pós-graduação Elizabeth, Laura e Ana Paula. Vocês que na maioria das vezes estão por trás desse trabalho, mas, nem sempre recebem os méritos justos. Meus sinceros agradecimentos.

Não poderia deixar de agradecer aos meus eternos mestres: Profs. Joaquim Felipe, José Pacheco, Gustavo Batista, Saulo Rosa e todos aqueles que contribuíram para minha formação. Em especial gostaria de agradecer ao, antes professor e hoje amigo Mario Kubo, por sempre me incentivar em seguir a carreira acadêmica. Agradeço também aos meus colegas de graduação que depois se tornaram companheiros na pós-graduação: Jean Domingos, Lucas Baggio, Matheus Pires e Thiago Gaspar.

Agradeço ao Prof. Giancarlo Oliveira, Prof. Luís Bernacci e Juliano Gomes de Pádua, pelas valiosas contribuições para realização desse trabalho.

À minha família meu enorme agradecimento. Especialmente aos meus pais Roberto e Marilda, sem vocês eu não estaria aqui. Obrigado por sempre deixar eu escolher meus caminhos e, nas minhas escolhas sempre me apoiar. Agradeço também a minha namorada Fernanda. Você que abriu mão de muitos momentos de convívio e sofreu pela minha ausência quando o dever do estudo me chamava. Também não posso deixar de citar minha profunda admiração por todos meus avôs e avós. À minha Vó Terezinha, que me ensinou muito do que eu sei, meu infinito amor.

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

Índice

<i>Lista de Figuras</i>	<i>iii</i>
<i>Lista de Tabelas.....</i>	<i>x</i>
<i>Resumo</i>	<i>xii</i>
<i>Abstract.....</i>	<i>xiii</i>
1.INTRODUÇÃO.....	1
<i>1.1 Justificativa e Motivação</i>	<i>1</i>
<i>1.2 Objetivos</i>	<i>2</i>
<i>1.3 Organização do Texto</i>	<i>3</i>
2. SISTEMA TREEVIS	5
<i>2.1 Biodiversidade, Herbários e o TreeVis.....</i>	<i>5</i>
<i>2.2 O Sistema TreeVis.....</i>	<i>7</i>
<i>2.3 Arquitetura do TreeVis</i>	<i>9</i>
<i>2.4 Considerações Finais.....</i>	<i>11</i>
3. ANÁLISE DE IMAGENS E BIOMETRIA.....	12
<i>3.1 Introdução a Análise e Processamento de Imagens</i>	<i>12</i>
<i>3.2 Dimensão Fractal.....</i>	<i>14</i>
<i>3.2.1 Método Box-counting.....</i>	<i>18</i>
<i>3.2.2 Método Bouligand-Minkowsky</i>	<i>19</i>
<i>3.2.3 Método Multiescala</i>	<i>20</i>
<i>3.3 Esqueletonização.....</i>	<i>23</i>
<i>3.3.1 Algoritmo de Hilditch</i>	<i>28</i>
<i>3.3.2 Algoritmo de Stentiford.....</i>	<i>29</i>
<i>3.3.3 Algoritmo de Zhang e Suen.....</i>	<i>30</i>
<i>3.3.4 Algoritmo de Holt</i>	<i>32</i>
<i>3.3.5 Algoritmo Multiescala</i>	<i>35</i>
<i>3.4 Biometria</i>	<i>39</i>
<i>3.5 Considerações Finais.....</i>	<i>43</i>

4. CARACTERIZAÇÃO DE ÓRGÃOS FOLIARES	44
4.1 <i>Análise do Conjunto de Nervuras das Folhas</i>	45
4.2 <i>Análise do Esqueleto das Folhas.....</i>	48
4.3 <i>Análise de Complexidade.....</i>	49
4.4 <i>Considerações Finais.....</i>	52
5. RECONHECIMENTO DE PADRÕES	53
5.1 <i>Fundamentos.....</i>	53
5.2 <i>Métodos para Reconhecimento de Padrões</i>	56
5.2.1 <i>Classificação Estatística.....</i>	56
5.2.2 <i>Classificação por Distância Mínima</i>	57
5.2.3 <i>Classificação por Redes Neurais.....</i>	59
5.3 <i>Análise dos Componentes Principais</i>	63
5.4 <i>Considerações Finais.....</i>	65
6. RESULTADOS EXPERIMENTAIS.....	66
6.1 <i>Espécies de Órgãos Foliares.....</i>	66
6.2 <i>Sistema Computacional</i>	68
6.3 <i>Experimento I: Complexidade de Órgãos Foliares</i>	72
6.3.1 <i>Complexidade das Nervuras.....</i>	75
6.3.2 <i>Complexidade dos Esqueletos</i>	88
6.4 <i>Experimento II: Biometria de Órgãos Foliares.....</i>	98
6.4.1 <i>Biometria das Nervuras.....</i>	99
6.4.2 <i>Biometria dos Esqueletos.....</i>	109
6.5 <i>Experimento III: Comparação dos Métodos</i>	119
6.5.1 <i>Comparação: Complexidade e Biometria das Nervuras.....</i>	120
6.5.2 <i>Comparação: Complexidade e Biometria dos Esqueletos</i>	122
6.5.3 <i>Comparação: Todas as características extraídas</i>	125
6.6 <i>Análise e Discussão dos Resultados</i>	127
6.7 <i>Considerações Finais.....</i>	127
7. CONCLUSÕES.....	132
7.1 <i>Contribuições do Trabalho</i>	134
7.2 <i>Trabalhos Futuros</i>	135
ANEXO	136
REFERÊNCIAS.....	146

Lista de Figuras

Figura 2.1: Visão geral da arquitetura do sistema TreeVis.....	9
Figura 2.2: Ambiente de trabalho do Sistema TreeVis. (a) espécie arbórea analisada; (b) análise de curvatura; (c) análise de textura no domínio da frequência; (d) detecção de borda; (e) histograma e (f) tabela de vetor de características (FALVO, PLOTZE <i>et al.</i> , 2004).	10
Figura 3.1: Ilustração dos conceitos fundamentais adotados nas próximas seções.	14
Figura 3.2: Iterações para criação da curva de Koch.	16
Figura 3.3: Exemplo de uma imagem mapeada com diferentes tamanhos de caixas ε . .	18
Figura 3.4: Estimação da dimensão fractal através do método <i>Box-counting</i> . (a) imagem original; (b) gráfico $\log(N(\varepsilon)) \times \log(\varepsilon)$	19
Figura 3.5: Estimação da dimensão fractal através do método de <i>Minkowsky</i> . (a) imagem original; (b) imagem após sucessivas dilatações; (c) gráfico bi-log contendo a reta resultante da regressão linear	20
Figura 3.6: Processo da estimativa da dimensão fractal multiescala. (a) imagem original; (b) dilatações sucessivas da imagem; (c) gráfico bi-log $\log A(d) \times \log(d)$; (d) gráfico da dimensão fractal multiescala.....	22
Figura 3.7: Ilustração de propriedades extraídas do gráfico da dimensão fractal multiescala: pico de fractalidade ($F_{\max}$), escala de máxima fractalidade ($D_{\max}$) e largura de alta fractalidade (L) (adaptado de (COSTA e BIANCHI, 2002))......	23
Figura 3.8: Exemplo da técnica de esqueletonização em uma imagem de impressão digital: (a) imagem original; (b) imagem esqueletonizada (SMITH, 1999).	24
Figura 3.9: Transformação do eixo medial em uma forma simples. (a) imagem original; (b) imagem após a MAT (GONZALEZ e WOODS, 2002).	25
Figura 3.10: Os 8 vizinhos do <i>pixel</i> P1.	26
Figura 3.11: Número de transição de 0 para 1 na sequência de P2, P3, ..., P9, P2. (a) uma transição; (b) duas transições.	27
Figura 3.12: Número de vizinhos com valor diferente de zero do ponto P1. (a) Dois vizinhos; (b) Cinco vizinhos.	27

Figura 3.13: Ponto extremo, o <i>pixel</i> tem apenas um <i>pixel</i> preto como seu vizinho.	27
Figura 3.14: Exemplo do resultado do processo de esqueletonização através do algoritmo de <i>Hilditch</i> . (a) imagem original; (b) imagem esqueletonizada.	29
Figura 3.15: Os quatro tipos de máscaras utilizadas para erosão das imagens através do algoritmo de <i>Stentiford</i>	29
Figura 3.16: Exemplo da aplicação do algoritmo de <i>Stentiford</i> ; (a) imagem original; (b) imagem esqueletonizada.	30
Figura 3.17 Processo de esqueletonização de <i>Zhang-Suen</i> em uma imagem de folha; (a) imagem original; (b) imagem esqueletonizada.	32
Figura 3.18: Vizinhança do <i>pixel</i> C utilizando coordenadas cardinais.	33
Figura 3.19: As máscaras utilizadas pelo algoritmo de <i>Holt</i> com o conceito de remoção em escada.	34
Figura 3.20: Resultado do algoritmo de esqueletonização de <i>Holt</i> . (a) imagem original; (b) imagem esqueletonizada com o método inicialmente proposto; (c) imagem esqueletonizada com o conceito de remoção em escada.	34
Figura 3.21: O problema principal problema da SAT em objetos com regiões ruidosas. (a) SAT de uma região sem ruídos; (b) SAT da mesma região com um pequeno ruído.	35
Figura 3.22: Processo de rotulação sucessiva dos elementos do contorno do objeto. (a) representação matricial de um objeto; (b) resultado do processo de rotulação sucessiva.	36
Figura 3.23: Propagação dos rótulos do contorno de um objeto através do conceito de dilatação exata.	36
Figura 3.24: Esqueletonização multiescala. (a) imagem original; (b) extração do contorno e rotulação sucessiva; (c) propagação dos rótulos através de dilatações exatas; (d) esqueleto resultante para limiar $T = 3$; (e) $T = 5$; (f) $T = 10$; (g) $T = 15$; (h) $T = 20$; (i) $T = 25$; (j) $T = 50$; (k) $T = 75$ e (l) $T = 100$	38
Figura 3.25: Esqueletos obtidos através da esqueletonização multiescala. (a) esqueleto interno; (b) esqueleto externo.	38
Figura 3.26: Imagem resultante do algoritmo capaz de encontrar bifurcações em estruturas tubulares.	40
Figura 3.27: Extração dos pontos extremos de uma estrutura tubular.	42
Figura 3.28: Extração da ramificação principal de uma estrutura tubular.	43
Figura 4.1: Exemplo de um órgão foliar no qual o conjunto de nervuras está bem evidenciado.	45
Figura 4.2: Extração do número de bifurcações do conjunto de nervuras (adaptado de (WING, WILF <i>et al.</i> , 1999)).	46

Figura 4.3 : Extração de características a partir dos ângulos entre as bifurcações dos órgãos foliares.....	47
Figura 4.4: Espaçamento entre o conjunto de nervuras da folhas. (a) espaçamento uniforme; (b) espaçamento irregular (WING, WILF <i>et al.</i> , 1999).	47
Figura 4.5: Exemplo das nervuras basais dos órgãos foliares (WING, WILF <i>et al.</i> , 1999).	48
Figura 4.6: Resultado do processo de esqueletonização em órgãos foliares. (a) imagem original; (b) respectivo esqueleto; (c) imagem original; (d) respectivo esqueleto.	49
Figura 4.7 : Exemplos de imagens artificialmente geradas que ilustram o conjunto de nervuras das folhas; (a) imagem 1; (b) imagem 2.	50
Figura 4.8: Resultado do método de dilatações exatas em nervuras de órgãos foliares; (a) dilatações sucessivas do objeto da Figura 4.7; (b) dilatações sucessivas do objeto da Figura 4.7b.	50
Figura 4.9: Gráficos bi-log do processo de dilatações exatas; (a) gráfico gerado a partir das dilatações da 4.7a; (b) gráfico gerado a partir das dilatações da Figura 4.7b.....	51
Figura 4.10: Gráfico bi-log $\log(d) \times \log(A')$ da análise multiescala da dimensão fractal; (a) gráfico multiescala da Figura 4.7a; (b) gráfico multiescala da Figura 4.7b;.....	51
Figura 5.1: Visão geral das etapas de um sistema computacional evidenciando a fase de reconhecimento de padrões.....	55
Figura 5.2: Exemplo prático de reconhecimento de padrões de espécies de flores. (a) dados relativos às características das sépalas (b) dados referentes às características das pétalas. (baseado em (GONZALEZ e WOODS, 2002))	56
Figura 5.3: Arquitetura de um classificador estatístico.	57
Figura 5.4: Resultados da classificação por distância mínima com os dados das espécies de flores. Distância entre o conjunto de amostras e os centróides das espécies (a) <i>I.setosa</i> ; (b) <i>I.virginica</i> e (c) <i>I.versicolor</i>	58
Figura 5.5: Comparação entre um neurônio biológico e o modelo artificial proposto por McCulloch-Pits	59
Figura 5.6: Funções de ativação das redes neurais artificiais. (a) função linear; (b) função de escada e (c) função sigmoidal	60
Figura 5.7: Perceptron multicamada com uma camada escondida, quatro variáveis de entrada e três variáveis de saída.....	61
Figura 5.8: Redução de dimensionalidade utilizando a técnica de PCA do conjunto de dados de flores. (a) redução dos dados para três dimensões; (b) redução dos dados para duas dimensões.	64
Figura 6.1: Espécies de maracujás silvestres do gênero <i>Passiflora</i> utilizadas nos experimentos. (a) <i>Passiflora coriacea</i> ; (b) <i>P. foetida</i> ; (c) <i>P. miersii</i> ; (d) <i>P. organensis</i> ; (e)	

P. pohlii; (f) P. suberosa; (g) P. amethystina; (h) P. caerulea; (i) P. giberti e (j) P. alata.	67
Figura 6.2: Espécies arbóreas da Mata Atlântica e do Cerrado utilizadas nos experimentos.	68
Figura 6.3: Diagrama da arquitetura do sistema computacional desenvolvido para realização dos experimentos de identificação das espécies.	70
Figura 6.4: Digrama dos experimentos realizados para identificação das espécies	72
Figura 6.5: Fases do experimento sobre complexidade dos órgãos foliares.....	73
Figura 6.6: Pontos de interesse extraídos da curva fractal multiescala. As coordenadas x e y de cada ponto são consideradas como características.	74
Figura 6.7: Gráficos da dimensão fractal multiescala das espécies de Passiflora. (a) Passiflora coriacea; (b) P. foetida; (c) P. miersii; (d) P. organensis; (e) P. pohlii; (f) P. suberosa; (g) P. amethystina; (h) P. caerulea; (i) P. giberti e (j) P. alata.	75
Figura 6.8: Resultados experimentais da análise baseada na distância entre os clusters. Em verde os clusters com separação satisfatória; em amarelo os clusters parcialmente sobrepostos e em vermelho os clusters sobrepostos. Comparação das distâncias entre os clusters das Espécies: (a) Passiflora coriacea; (b) P. foetida; (c) P. miersii; (d) P. organensis; (e) P. pohlii; (f) P. suberosa; (g) P. amethystina; (h) P. caerulea; (i) P. giberti e (j) P. alata.	77
Figura 6.9: Distância entre os centróides e as quarenta amostras utilizadas nos experimentos com Passiflora. Distâncias entre as amostras e o centróides das Espécies: (a) Passiflora coriacea; (b) P. foetida; (c) P. miersii; (d) P. organensis; (e) P. pohlii; (f) P. suberosa; (g) P. amethystina; (h) P. caerulea; (i) P. giberti e (j) P. alata.	78
Figura 6.10: Redução de dimensionalidade das características extraídas das espécies de 6D para 4D. Visualização dos resultados através da técnica de coordenadas paralelas. Espécies: (a) Passiflora coriacea; (b) P. foetida; (c) P. miersii; (d) P. organensis; (e) P. pohlii; (f) P. suberosa; (g) P. amethystina; (h) P. caerulea; (i) P. giberti e (j) P. alata. .	79
Figura 6.11: Gráficos da dimensão fractal multiescala das espécies arbóreas da Mata Atlântica e do Cerrado. (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9 e (j) Espécie 10..	81
Figura 6.12: Comparação da distância entre os centróides das espécies da Mata Atlântica e de Cerrado. Distância entre as espécies e o centróide da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.	83
Figura 6.13: Análise das distância entre todas as amostras utilizadas no experimento e o clusters da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.	83
Figura 6.14: Redução de dimensionalidade de 6D para 4D através da técnica de PCA. Coordenadas paralelas da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.	84

Figura 6.15: Análise dos clusters por distância mínima das 20 espécies utilizadas no experimento. (a) comparação da distância da espécie 1 aos demais clusters; (b) comparação da distância da espécie 2 aos demais clusters; e assim por diante até (t) comparação da distância da espécie 20 aos demais clusters..... 87

Figura 6.16: Gráficos da dimensão fractal multiescala dos esqueletos das espécies de *Passiflora*. (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*. 89

Figura 6.17: Resultados da análise de clusters das espécies de *Passiflora*. Distância entre os clusters e o centróide da espécie: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*..... 90

Figura 6.18: Distância entre os centróides e as quarenta amostras utilizadas nos experimentos com *Passiflora*. Distâncias entre as amostras e o centróides das Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*..... 90

Figura 6.19: Redução de dimensionalidade das características extraídas das espécies de 6D para 4D. Visualização dos resultados através da técnica de coordenadas paralelas. Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*. . 91

Figura 6.20: Gráficos da dimensão fractal multiescala das espécies arbóreas da Mata Atlântica e do Cerrado. (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9 e (j) Espécie 10... 93

Figura 6.21: Classificação das espécies baseada na distância entre os centróides. Comparação das distâncias entre os clusters das Espécies: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9 e (j) Espécie 10. 94

Figura 6.22: Distância entre os centróides e as cinquenta amostras utilizadas nos experimentos com *Passiflora*. Distâncias entre as amostras e o centróides das Espécies: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9 e (j) Espécie 10..... 95

Figura 6.23: Redução de dimensionalidade das características extraídas das espécies de 6D para 4D. Visualização dos resultados através da técnica de coordenadas paralelas. Espécies: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9 e (j) Espécie 10. 95

Figura 6.24: Análise dos clusters por distância mínima das 20 espécies utilizadas no experimento. (a) distância da espécie 1 aos demais clusters; (b) distância da espécie 2 aos demais clusters; e assim até (t) distância da espécie 20 aos demais clusters. 97

Figura 6.25: Resultados experimentais da análise baseada na distância entre os clusters. Comparação das distâncias entre os clusters das Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*. 100

Figura 6.26: Distância entre os centróides e as quarenta amostras utilizadas nos experimentos com *Passiflora*. Distâncias entre as amostras e o centróides das Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*. 101

Figura 6.27: Redução de dimensionalidade das características extraídas das espécies de 19D para 10D. Visualização dos resultados através da técnica de coordenadas paralelas. Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*. 102

Figura 6.28: Comparação da distância entre os centróides das espécies da Mata Atlântica e de Cerrado. Distância entre as espécies e o centróide da : (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10. 104

Figura 6.29: Análise das distâncias entre todas as amostras utilizadas no experimento e os clusters da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10. 105

Figura 6.30: Visualização dos resultados da técnica de PCA através de coordenadas paralelas. (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10. 106

Figura 6.31: Resultados da análise de clusters (agrupamento) das 20 espécies utilizadas no experimento. (a) distância da espécie 1 aos demais clusters; (b) distância da espécie 2 aos demais clusters; e assim até (t) distância da espécie 20 aos demais clusters. 108

Figura 6.32: Análise baseada na distância entre os clusters. Comparação das distâncias entre os clusters das Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*. 110

Figura 6.33: Distâncias entre as amostras e o centróides das Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*. 111

Figura 6.34 Visualização dos resultados da redução de dimensionalidade de 19D para 10D. (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*. 112

Figura 6.35: Comparação da distância entre os centróides das espécies da Mata Atlântica e de Cerrado. Distância entre as espécies e o centróide da : (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10. 114

Figura 6.36: Análise das distâncias entre todas as amostras utilizadas no experimento e os clusters da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10. 115

Figura 6.37: Redução de dimensionalidade de 19D para 10D através da técnica de PCA. Coordenadas paralelas da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4;

(e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.	116
--	-----

Figura 6.38: Resultados da análise de clusters das 20 espécies utilizadas no experimento. (a) distância entre o centróide da espécie 1 e as demais espécies; (b) distância entre o centróide da espécie 2 e as demais espécies; e assim até (t) distância entre o centróide da espécie 20 e as demais espécies.	118
---	-----

Figura 6.39: Análise dos clusters por distância mínima das 20 espécies utilizadas no experimento. (a) comparação da distância da espécie 1 aos demais clusters; (b) comparação da distância da espécie 2 aos demais clusters; e assim até (t) comparação da distância da espécie 20 aos demais clusters.	122
--	-----

Figura 6.40: Resultados da análise de clusters. (a) distância entre a espécie 1 e as demais espécies; (b) distância entre a espécie 2 e as demais espécies; e assim até (t) distância entre a espécie 20 e as demais espécies.	124
--	-----

Figura 6.41: Análise das distância entre os centróides dos agrupamentos. Distância entre os centróides das espécies e o centróide da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9 e (j) Espécie 10.	126
---	-----

Figura 6.42: Diagrama dos experimentos executados para identificação das espécies vegetais.	127
---	-----

Lista de Tabelas

Tabela 3.1: Condições para remoção de <i>pixels</i> com algoritmo de <i>Hilditch</i>	28
Tabela 3.2: Condições da primeira iteração para remoção de <i>pixel</i> com algoritmo de <i>Zhang-Suen</i>	31
Tabela 3.3: Condições da segunda iteração para remoção de <i>pixel</i> com algoritmo de <i>Zhang-Suen</i>	31
Tabela 5.1: Matriz de confusão ilustrando os resultados da rede neural utilizada para classificação das espécies de flores Íris (50% treinamento / 50% testes).....	62
Tabela 5.2: Matriz de confusão ilustrando os resultados da rede neural utilizada para classificação das espécies de flores Íris (60% treinamento / 40% testes).....	63
Tabela 5.3: Matriz de confusão ilustrando os resultados da rede neural utilizada para classificação das espécies de flores Íris (80% treinamento / 20% testes).....	63
Tabela 6.1: Resultado da classificação por redes neurais dos dados de complexidade das nervuras - espécies de <i>Passiflora</i> (75% treinamento/25% testes).	80
Tabela 6.2: Resultado da classificação por redes neurais dos dados de complexidade das nervuras – espécies de <i>Passiflora</i> (50% treinamento/50% testes).	80
Tabela 6.3: Resultado da classificação por redes neurais dos dados de complexidade das nervuras – espécies da Mata Atlântica e do Cerrado (80% treinamento/20% testes).....	85
Tabela 6.4: Resultado da classificação por redes neurais dos dados de complexidade das nervuras – espécies da Mata Atlântica e do Cerrado (40% treinamento/60% testes).....	85
Tabela 6.5: Resultado da classificação por redes neurais dos dados de complexidade das nervuras – todas as espécies (75% treinamento/25% testes).	88
Tabela 6.6: Resultado da classificação por redes neurais dos dados de complexidade dos esqueletos – espécies de <i>Passiflora</i> (75% treinamento/25% testes).....	92
Tabela 6.7: Resultado da classificação por redes neurais dos dados de complexidade dos esqueletos – espécies de <i>Passiflora</i> (50% treinamento/50% testes).....	92
Tabela 6.8: Resultado da classificação por redes neurais dos dados de complexidade dos esqueletos – espécies da Mata Atlântica e do Cerrado (80% treinamento/20% testes)..	96

Tabela 6.9: Resultado da classificação por redes neurais dos dados de complexidade dos esqueletos – todas as espécies (75% treinamento/25% testes).	98
Tabela 6.10: Resultado da classificação por redes neurais dos dados de biometria das nervuras – espécies de <i>Passiflora</i> (75% treinamento/25% testes).....	103
Tabela 6.11: Resultado da classificação por redes neurais dos dados de biometria das nervuras – espécies de <i>Passiflora</i> (50% treinamento/50% testes).....	103
Tabela 6.12: Resultado da classificação por redes neurais dos dados de biometria das nervuras – espécies da Mata Atlântica e do Cerrado (80% treinamento/20% testes)...	106
Tabela 6.13: Resultado da classificação por redes neurais dos dados de biometria das nervuras – espécies da Mata Atlântica e do Cerrado (40% treinamento/60% testes)...	107
Tabela 6.14: Resultado da classificação por redes neurais dos dados de biometria das nervuras – todas as espécies (75% treinamento/25% testes).	109
Tabela 6.15: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – espécies de <i>Passiflora</i> (75% treinamento/25% testes).....	112
Tabela 6.16: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – espécies de <i>Passiflora</i> (50% treinamento/50% testes).....	113
Tabela 6.17: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – espécies da Mata Atlântica e do Cerrado (80% treinamento/20% testes).116	
Tabela 6.18: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – espécies da Mata Atlântica e do Cerrado (40% treinamento/60% testes).117	
Tabela 6.19: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – todas as espécies (75% treinamento/25% testes).	119
Tabela 6.20: Resultado da classificação por redes neurais dos dados de complexidade e biometria das nervuras – todas as espécies (75% treinamento/25% testes).	122
Tabela 6.21: Resultado da classificação por redes neurais dos dados de complexidade e biometria dos esqueletos – todas as espécies (75% treinamento/25% testes).	124
Tabela 6.22: Resultado da classificação por redes neurais de todos os dados coletados de todas as espécies (75% treinamento/25% testes).	126

Resumo

PLOTZE, R.O. **Identificação de espécies vegetais através da análise da forma interna de órgãos foliares**. 2004. 152p. Dissertação de Mestrado – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.

A diversidade de espécies presentes no riquíssimo reino vegetal, torna o processo de identificação de órgãos foliares uma tarefa muito complexa. A biodiversidade das espécies, associada aos modelos tradicionais de taxonomia, transforma essa tarefa em um verdadeiro desafio para os pesquisadores. Nesse trabalho é apresentado como as técnicas computacionais de análise de imagens e extração de características, podem contribuir na identificação de espécies vegetais. Para isso, os métodos desenvolvidos e implementados se concentram em informações extraídas da parte interna de órgãos foliares, mais especificamente do conjunto de nervuras e do esqueleto. A coleta das informações é realizada através de métodos que extraem características a partir da complexidade (dimensão fractal) e biometria dos órgãos foliares. A eficiência das técnicas implementadas foi avaliada em casos reais de identificação de espécies, nos quais foram utilizados dois conjuntos de imagens de órgãos foliares: árvores da Mata Atlântica e do Cerrado brasileiro e, espécies de maracujás nativos do gênero *Passiflora*. As espécies são classificadas utilizando técnicas de reconhecimento de padrões como: redes neurais artificiais, análise de agrupamentos e análise de componentes principais.

Palavras-chave: análise de imagens, extração de características, dimensão fractal, biometria, identificação de folhas, taxonomia.

Abstract

PLOTZE, R.O. **Plant species identification based on venation system shape analysis.** 2004. 152p. Dissertação de Mestrado – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.

The plant species diversity makes their correct identification a very complex task. The traditional taxonomy models, associated with species biodiversity, has been transformed this task in a challenger for the researches. This work presents a new approach to plant species identification, based on internal characteristics of leaf form. The data are collected by computer vision and shape analysis techniques, which extracts features from complexity (fractal dimension) and biometry of plant species. The methodology efficiency was evaluated with real cases of species identification: digital images of Mata Atlântica and Brazilian Cerrado species; and passion fruit species of genus *Passiflora*. The species classifications are performed using pattern recognition techniques as clustering and artificial neural networks.

Key-words: image analysis, feature extraction, fractal dimension, biometry, leaf identification, taxonomy.

Introdução

1.1 Justificativa e Motivação

A grande variedade de espécies vegetais, com suas inúmeras características taxonômicas, torna a problemática da identificação morfológica um dos principais desafios para a ciência contemporânea. A biodiversidade das espécies transforma os modelos tradicionais de taxonomia em uma tarefa árdua e especializada, na qual todas as fases são realizadas manualmente. Desse modo, o número de profissionais qualificados não são capazes de suprir as necessidades de levantamento e classificação das milhares de espécies do riquíssimo reino vegetal.

A identificação de espécies vegetais em florestas, além de ser um dos alicerces para o estudo e pesquisa científica em diversas áreas da ciência associadas à botânica, apresenta ainda importância vital para a manutenção e preservação ecológica. Além disso, fatores ligados à economia florestal como cultivo, extrativismo, pesquisa de matérias primas para medicamentos e fitoterapia, são muito dependentes das descobertas de novas espécies, bem como de suas características e propriedades. Por outro lado, a importância da taxonomia das espécies se depara com um grande problema: a exploração incessante dos recursos naturais associada à falta de conscientização ecológica. Essa prática tem provocado danos irreparáveis à natureza, fazendo com que espécies vegetais raras desapareçam antes mesmo de serem estudadas.

Para minimizar as dificuldades encontradas na identificação de espécies arbóreas, os herbários vegetais são ferramentas imprescindíveis para auxiliar o trabalho dos taxonomistas. O objetivo dos herbários é concentrar todas as espécies coletadas na natureza, bem como suas características, em um único local facilitando sua análise e pesquisa. Entretanto, mesmo utilizando os herbários para catalogar e identificar novas espécies, todo esse trabalho continua sendo realizado manualmente. Além disso, durante o processo de preparação manual, diversas propriedades importantes sobre as espécies são perdidas, tais como características sobre a coloração da área foliar, propriedades relativas a texturas e brilho, dentre outras.

Nesse contexto surge a necessidade do desenvolvimento de modelos matemáticos e computacionais, que sejam capazes de auxiliar os taxonomistas na tarefa de identificação das espécies vegetais. O intuito de se desenvolver novos métodos, não se concentra em momento algum na substituição dos profissionais de botânica, mais especificamente os taxonomistas, na tarefa de caracterização das espécies. Esses instrumentos, ou novas tecnologias, têm como único e principal objetivo auxiliar esses profissionais, servindo como ferramental durante a complexa tarefa de identificação vegetal.

O desenvolvimento e implementação de novas ferramentas podem contribuir muito para taxonomia vegetal, que é uma ciência responsável pela síntese, organização, classificação e nomenclatura das espécies de plantas. Além disso, outras áreas também podem ser beneficiadas como, genética, citologia, fisiologia, ecologia, paleobotânica, dentre outras. Em contrapartida, a solução de problemas biológicos, através de métodos computacionais, acabam por si mesmo resultando em novas técnicas e progressos em diversas áreas relacionadas à computação como: visão computacional, análise e processamento de imagens, dentre outras.

1.2 Objetivos

O objetivo principal desse trabalho é a identificação de espécies vegetais através das características internas de seus órgãos foliares. As informações extraídas para classificação das espécies concentraram-se em dois grupos de estruturas: conjuntos de nervuras (também conhecido como sistema de venação) e esqueleto.

Para coleta de informações das espécies, foram utilizados algoritmos de visão computacional e análise de imagens, que extraíram características relativas a complexidade (dimensão fractal) e biometria das estruturas (nervura e esqueleto) dos órgãos foliares. Com esses dados (vetores de características), métodos de reconhecimento de padrões foram empregados para caracterização das espécies.

A organização das fases de desenvolvimento do projeto foi dividida em 3 (três) módulos principais:

- implementação e estudo dos algoritmos: nesse módulo as técnicas de processamento e análise de imagens, necessárias para desenvolvimento do projeto foram codificadas e, em alguns casos, novos métodos foram desenvolvidos;
- experimentos: a eficiência dos métodos computacionais implementados e desenvolvidos, foi analisada em casos reais de identificação de espécies arbóreas;
- análise dos resultados: esse módulo se concentra na avaliação dos resultados dos experimentos de classificação das espécies. Com isso, através das porcentagens de acerto e/ou erros, é possível comparar a eficiência dos métodos desenvolvidos.

Para validação dos métodos desenvolvidos foram utilizados durante os experimentos dois conjuntos de imagens de órgãos foliares: (a) imagens de folhas de espécies arbóreas nativas das Mata Atlântica e do Cerrado brasileiro, coletadas no Instituto Florestal de Moji-Mirim; (b) imagens de folhas de maracujás nativos do gênero *Passiflora*, obtidos através de parceria entre o ICMC/USP e a ESALq/USP.

1.3 Organização do Texto

No Capítulo 2 (Sistema TreeVis) é apresentada uma descrição do projeto do sistema TreeVis. Esse sistema visa o estudo e desenvolvimento de técnicas de visão computacional, empregadas na identificação de espécies arbóreas.

No Capítulo 3 (Análise de Imagens e Biometria), há uma revisão bibliográfica sobre as técnicas de análise de imagens e extração de características utilizadas no

projeto. Os principais tópicos abordados são: dimensão fractal (complexidade), esqueletonização e biometria.

O Capítulo 4 (Caracterização de Órgãos Foliares) descreve como as metodologias apresentadas no Capítulo 3, são utilizadas para caracterização dos órgãos foliares. A caracterização, ou análise, dos órgãos foliares através das técnicas computacionais foi dividida em três módulos: (i) análise do conjunto de nervuras; (ii) análise do esqueleto e (iii) análise da complexidade por dimensão fractal.

Com as informações (vetores de características) extraídas dos órgãos foliares, se faz necessário a utilização de métodos para o agrupamento desses dados em categorias, ou mais especificamente em espécies. O Capítulo 5 (Reconhecimento de Padrões) descreve os métodos de reconhecimento de padrões, ou classificação, utilizados no projeto para discriminação dos órgãos foliares em espécies.

No Capítulo 6 (Resultados Experimentais) são apresentados os diversos experimentos realizados visando a identificação das espécies. Nesse capítulo são descritos os resultados dos métodos de reconhecimento de padrões empregados nas características extraídas dos órgãos foliares e, são apresentadas as porcentagens de acerto e erro na classificação das espécies.

Por fim, no Capítulo 7 (Conclusões) são discutidas as conclusões obtidas através dos resultados experimentais, bem como novos trabalhos que poderão ser desenvolvidos dando continuidade a este projeto. Além disso, são relatadas as contribuições trazidas com o desenvolvimento deste projeto.

Sistema TreeVis

Este capítulo apresenta uma descrição do projeto TreeVis. O sistema tem como principais objetivos o estudo e desenvolvimento de técnicas de visão computacional, empregadas na extração e análise de medidas biométricas de estruturas foliares. A proposta do TreeVis não visa substituir a função dos profissionais, mas sim beneficiá-los através da utilização do sistema em diversas áreas científicas como: botânica, fisiologia vegetal, citologia, paleobotânica, ecologia, dentre outras. O sistema pretende complementar as metodologias já utilizadas, facilitando o acesso as informações sobre órgãos foliares, além de acelerar o trabalho de reconhecimento das espécies. Um ponto importante sobre o TreeVis é fato dele ser um sistema digital, sendo assim, pode ser facilmente replicável e distribuído por diversas universidades, laboratórios e institutos de pesquisa.

2.1 Biodiversidade, Herbários e o TreeVis

A necessidade de conhecimento, conservação e utilização sustentável da fauna, flora e do ambiente onde vivem animais e plantas, são um dos principais desafios enfrentados por pesquisadores e cientistas neste início de século. Essa demanda crescente por conhecimento acerca da biodiversidade, em escalas global, regional e nacional, emergiu muito após a Conferência das Nações Unidas sobre Meio Ambiente e Desenvolvimento, realizada em 1992 (PEIXOTO e MORIM, 2003). A tarefa de coleta, identificação,

descrição e estudo das informações sobre a biodiversidade é um trabalho diretamente ligado a Taxonomia Biológica.

Um dos primeiros trabalhos sobre taxonomia vegetal, desenvolvido por *Linnaeus* na segunda metade do século XVIII, fez com que a identificação e reconstrução de relações entre espécies de plantas tenham sido largamente estudadas até hoje (WING, WILF et al., 1999). Além disso, desde aquela época a taxonomia vegetal foi utilizada como suporte indispensável para uma grande variedade de propósitos. Se por um lado a taxonomia vegetal é imprescindível, por outro ela enfrenta sérios problemas ligados à exploração continuada dos recursos naturais. O Brasil que é considerado o país de maior diversidade biológica do mundo, sofre diariamente com a falta de conscientização ecológica que vem acarretando danos irreparáveis. Muitas vezes, espécies pouco conhecidas, ou até mesmo nunca estudadas, desaparecem antes mesmo de serem catalogadas. Mediante esses e outros problemas, a necessidade de inventariar as diversidades de espécies do mundo, bem como analisar e sintetizar as informações delas extraídas se fazem cada vez mais necessárias.

Os herbários são ferramentas indispensáveis para esse trabalho. Eles se caracterizam por uma coleção científica de exemplares de ramos férteis (galhos com folhas, podendo possuir flores ou frutos característicos), coletados na natureza, prensados, dessecados e montados sobre papel cartão. Associado a cada exemplar existe uma ficha padronizada, contendo todas as informações sobre a espécie, com o rigor da taxonomia. Através dos herbários, é possível examinar e estudar espécies provenientes de diferentes locais e ecossistemas. Além disso, é com o auxílio dos herbários que o taxonomista pode comparar um exemplar da natureza com milhares de espécies já catalogadas para realizar sua identificação (BRUNO, 2000). O maior herbário do mundo, segundo o *Index Herbariorum* (lista contendo os maiores herbários cadastrados), contém mais de 8 milhões de espécies. No Brasil, todos os herbários em conjunto (cerca de 114 espalhados pelo país) guardam um acervo de pouco mais de 5 milhões de espécimes (PEIXOTO e MORIM, 2003).

Nos herbários existentes, a identificação de novos espécimes é realizada manualmente, de forma, que são coletados na natureza, dessecados e comparados com as espécies já catalogadas no herbário. Para realizar tal tarefa, são necessários profissionais com grandes conhecimentos e habilidades na área de taxonomia, além de ser bastante trabalhoso. Outro ponto desfavorável na tarefa de identificação manual de

espécies, ocorre pela perda de propriedades importantes sobre as espécies durante o processo de identificação como: coloração, textura, brilho, entre outras (BRUNO, 2000).

Diferente disso, o sistema TreeVis concentra-se nas informações extraídas das folhas para realizar a identificação, ao invés de comparar-se manualmente com as fichas de catalogação como ocorre nos herbários tradicionais. Além disso, o funcionamento do TreeVis é baseado na análise sistemática dos órgãos foliares das espécies, atuando como um herbário digital inteligente (BRUNO e COSTA, 2001; FALVO, PLOTZE et al., 2004).

2.2 O Sistema TreeVis

A tarefa de identificação de espécies arbóreas, bem como sua catalogação e estudo, é um desafio para os pesquisadores frente a abundante biodiversidade. A proposta do sistema TreeVis, *Tree Vision System*, é fornecer aos taxonomistas uma ferramenta capaz de armazenar, classificar e organizar dados científicos, a fim de, realizar a identificação automática de espécies, funcionando como um herbário digital.

Ao contrário dos herbários tradicionais, em que todo processo é realizado de maneira manual, no TreeVis a identificação dos exemplares é feita através de informações contidas nas folhas. Dessa forma, ao invés do taxonomista comparar manualmente todas as fichas contendo as informações das espécies, como ocorre nos herbários tradicionais, o usuário necessita apenas entrar com as imagens dos órgãos foliares no sistema.

Embora as flores e frutos provêem características importantes para taxonomia de espécies, existem situações como por exemplo em algumas espécies de vegetais, em que estes órgãos não estão disponíveis para estudo (WING, WILF *et al.*, 1999). Além disso, as flores e frutos não estão presentes em toda época do ano, tornando a tarefa de identificação mais difícil.

A natureza dos vegetais, do ponto de vista de suas características e peculiaridades, faz com que o processo de reconhecimento através das folhas seja um verdadeiro desafio para um sistema de visão computacional. As variações morfológicas existentes dentro de uma mesma espécie de folha é um dos principais problemas. Outros

fatores como, maturidade ou mesmo o processo de formação das folhas (exposição ao sol, chuva, etc.), faz com que essas tenham grandes variações quanto ao tamanho, coloração, textura, formato, etc. Além disso, indivíduos de uma mesma espécie podem possuir significativas diferenças quando situados em regiões ou ecossistemas diferentes (BRUNO, 2000).

Buscando tentar superar tais desafios, o TreeVis é baseado em informações que apresentam grande potencial para identificação das espécies como:

- exploração sistemática das características das folhas através de um vasto número de atributos visuais: o grande número de atributos necessários para reconhecimento das espécies se justifica principalmente pela diversidade das espécies, bem como a variabilidade dentro de um mesmo conjunto;
- sinergismo entre os atributos: o potencial de caracterização das espécies utilizando os atributos extraídos de maneira isolada é significativamente menos representativo do que quando esses atributos são empregados de maneira combinada;
- análise estatística: a qualidade dos atributos extraídos é avaliada através da análise estatística dos dados.
- metodologia de utilização: a necessidade do botânico taxonomista é essencial em todas as fases do reconhecimento das espécies através do TreeVis. É esse profissional quem garantirá a consistência das informações que são inseridas e aprendidas pelo sistema.

A proposta inicial do sistema TreeVis é empregá-lo no reconhecimento de espécies arbóreas funcionando como um herbário digital, contudo, outras aplicações podem ser consideradas como: (i) laboratório para experimentos morfofisiológicos – atuando como uma ferramenta de coleta e análise de dados; (ii) sistema biotecnológico – o TreeVis pode ser utilizado para acompanhar o desenvolvimento de uma cultura a fim de monitorar ações sofridas por deficiências nutricionais ou ataques por pragas (insetos, fungos ou viroses); dentre outras (FALVO, PLOTZE et al., 2004).

2.3 Arquitetura do TreeVis

A Figura 2.1 apresenta uma visão geral do sistema TreeVis, o qual é composto de cinco módulos: Estrutura de Amostragem, Controle Central, Extração de Características, Treinamento/Identificação e Base de Dados.

Figura 2.1: Visão geral da arquitetura do sistema TreeVis.

As informações são inseridas no sistema através de um dispositivo de digitalização de imagens. Para isso, é utilizado um scanner de mesa comercial que permite digitalizar imagens com até 9.600 DPI (*dot per inch* – pontos por polegada). Geralmente as informações contidas em cada um dos lados das folhas são diferentes, é necessário digitalizar ambos os lados, definidos como: lado de sol (exposta à luz solar direta) e lado de sombra (lado oposto). Além disso, também é necessário digitalizar uma determinada região da folha de maneira ampliada, que é utilizada posteriormente para análise de textura. A amostra, ou conjunto de amostras, inseridas no sistema é denominada *Estrutura de Amostragem*.

O módulo *Controle Central* é responsável por encaminhar as amostras contidas na *Estrutura de Amostragem* para o módulo de *Extração de Características*. O *Controle Central* funciona como um maestro do sistema, mediando as informações trocadas entre os módulos: *Estrutura de Amostragem*, *Extração de Características* e *Treinamento/Identificação*.

O módulo de *Extração de Características* é responsável pela mineração das informações das amostras, que formarão o vetor de características de cada órgão foliar. Para isso, o módulo contém diversas técnicas de visão, análise e processamento de imagens, divididas em quatro submódulos: *Geometria*, *Cor*, *Forma* e

Textura/Complexidade. Depois de gerado o vetor de características da amostra, essas informações são enviadas para o módulo de *Controle Central* que as encaminha para o módulo de *Treinamento/Identificação*.

A classificação das espécies é executada pelo módulo de *Treinamento/Identificação*, ocorrendo através da busca e comparação do vetor de características recebido com os demais vetores armazenados na *Base de Dados*. Para isso, são utilizadas diversas abordagens de reconhecimentos de padrões como: classificação por similaridade, redes neurais, análise multivariada, etc.

Finalmente, o sistema termina sua execução quando a análise e o processamento de cada amostra da *Estrutura de Amostragem* é concluído. O resultado de todo processamento do sistema é uma lista ordenada contendo as espécies mais prováveis e sua respectiva probabilidade de acerto na identificação. Para cada um dos itens da lista, é possível visualizar informações referentes às espécies como: nome popular, científico, família e imagens das amostras. A Figura 2.2 apresenta o ambiente de trabalho do Sistema TreeVis.

Figura 2.2: Ambiente de trabalho do Sistema TreeVis. (a) espécie arbórea analisada; (b) análise de curvatura; (c) análise de textura no domínio da freqüência; (d) detecção de borda; (e) histograma e (f) tabela de vetor de características (FALVO, PLOTZE *et al.*, 2004).

2.4 Considerações Finais

Neste capítulo foi apresentado um projeto complexo para identificação de espécies arbóreas, o sistema TreeVis. A proposta do sistema é funcionar como herbário digital, auxiliando o trabalho dos taxonomistas na identificação das espécies. O funcionamento do sistema TreeVis é baseado na exploração de informações que realmente contribuem para identificação das espécies como: (i) extração de um vasto conjunto de atributos visuais; (ii) sinergismo; (iii) análises estatísticas e (iv) metodologia da utilização.

Análise de Imagens e Biometria

As técnicas de análise de imagens são capazes de extrair informações para descoberta de conhecimento em imagens digitais. Para isso elas extraem propriedades dos objetos contidos nas imagens. Na literatura inúmeros métodos de análise de imagens podem ser encontrados, bem como, as mais variadas aplicações. A etapa de análise de imagens para identificação de espécies arbóreas se concentrou em três grupos principais de características: (i) dimensão fractal; (ii) esqueleto e (iii) biometria. Este capítulo apresenta uma descrição dos métodos computacionais utilizados para extração de características dos órgãos foliares. As informações extraídas por estes métodos, ou mais especificamente o vetor de características, formam a base para etapa de reconhecimento de padrões.

3.1 Introdução a Análise e Processamento de Imagens

Os métodos de análise e/ou processamento de imagens geralmente estão acoplados a sistema de visão computacional e utilizam como “matéria prima” o conteúdo de imagens digitais. Essas técnicas são capazes de extrair informações de uma imagem digital, viabilizando sua análise computacional. Alguns conceitos fundamentais sobre o assunto, bem como terminologias são apresentadas a seguir.

O termo imagem digital pode ser compreendido como sendo uma matriz cujos índices de linhas e colunas identificam uma coordenada na imagem. O valor de cada

elemento da matriz corresponde a intensidade de luz daquela coordenada na imagem. Os elementos dessa matriz são conhecidos como *pixels*, que corresponde a abreviatura de *picture elements* (GONZALEZ e WOODS, 2002).

As imagens digitais utilizadas no projeto de identificação de órgãos foliares foram imagens coloridas, adquiridas com resolução de 300 dpi. Para o processo de análise e extração de características, essas imagens foram transformadas para imagens binárias. Esse tipo de imagem é uma das maneiras mais simples e úteis para representação de imagens através do computador (COSTA e CESAR, 2000). De forma geral, uma imagem binária β utiliza apenas dois valores inteiros para representar seus *pixels*: zero (branco) ou preto (um). Convencionalmente o valor 1 está associado ao objeto de interesse (*foreground*) e, o valor 0 corresponde ao fundo (*background*) da imagem.

Os *pixels* da imagem binária β , referenciados na forma $p(x, y)$, podem possuir até oito *pixels* adjacentes, esse conceito é conhecido como vizinhança de um *pixel*. Duas definições alternativas são comumente consideradas para vizinhança: *vizinhança-de-4* e *vizinhança-de-8* (COSTA e CESAR, 2000). As coordenadas (ímpares) dos quatro vizinhos de $p_0(x, y)$ são: $p_1 = (x, y + 1)$; $p_3 = (x + 1, y)$; $p_5 = (x, y - 1)$; $p_7 = (x - 1, y)$. E as coordenadas pares de p_0 são: $p_2 = (x + 1, y + 1)$; $p_4 = (x + 1, y - 1)$; $p_6 = (x - 1, y - 1)$; $p_8 = (x - 1, y + 1)$. A *vizinhança-de-8* de $p_0(x, y)$ é formada pela união entre as coordenadas ímpares e pares.

A distância entre as coordenadas de dois *pixels* de uma imagem é um conceito importante em processamento e análise de imagens, conhecido como métricas de distância (GONZALEZ e WOODS, 2002). Diversas medidas de distância podem ser encontradas na literatura, contudo a métrica utilizada nesse trabalho foi a distância euclidiana. Considerando dois *pixels* p e q com coordenadas (x_p, y_p) e (x_q, y_q) respectivamente, a distância euclidiana entre eles é definida pela Equação 3.1.

$$De(p, q) = \left(\sqrt{(x_p - x_q)^2 + (y_p - y_q)^2} \right) \quad (3.1)$$

Na Figura 3.1 alguns conceitos fundamentais de análise e processamento de imagens são ilustrados. Eles servirão de base para entendimento das próximas seções. Uma imagem binária é representada pelo fundo branco (zero) e o objeto de interesse preto (um). A *vizinhança-de-4* de um *pixel* p_0 é ilustrada em vermelho e, a *vizinhança-de-8* de p_0 é representada em amarelo. A distância euclidiana entre dois *pixel* $p(x, y)$ e $q(x, y)$ é representada em azul.

0	0	0	0	0	0	0	0	0	0
1	1	p7	0	0	0	1	1	0	0
0	p5	p0	p1	1	1	p6	p7	p8	0
0	1	p3	1	1	1	p5	p0	p1	0
0	0	1	1	1	1	p4	p3	p2	0
0	0	1	p	1	1	1	1	1	0
0	1	1	1	0	0	1	1	1	0
0	1	1	1	0	0	1	q	1	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0

Figura 3.1: Ilustração dos conceitos fundamentais adotados nas próximas seções.

3.2 Dimensão Fractal

Um dos principais desafios enfrentado por matemáticos, desde o início do século 19, tem sido desenvolver maneiras para descrever fenômenos da natureza. Por exemplo, como é possível representar formas irregulares tais como árvores, nuvens, paisagens, processos físicos, etc.? Uma das respostas para essa pergunta transcende a Geometria Euclidiana e ficou conhecida como Geometria Fractal. Célebres matemáticos como *Hausdorff*, *Koch*, *Sierpinski* e *Cantor*, já sugeriam figuras e objetos com dimensão fracionária, no entanto, o conceito se consolidou apenas em 1975 com a teoria da geometria fractal desenvolvida por *Benoit Mandelbrot* (MANDELBROT, 1977).

Na geometria euclidiana as dimensões são inteiras, ou seja, um ponto possui dimensão 0 (0D), uma reta dimensão 1 (1D), um plano dimensão 2 (2D) e um sólido

conseqüentemente dimensão 3 (3D). Por outro lado na geometria fractal é diferente, os objetos são representados por dimensões intermediárias, ou seja, os objetos podem assumir dimensões com valores não inteiros. O matemático *Mandelbrot* definiu a dimensão desses objetos como *Dimensão Fractal* (MANDELBROT, 1983). Imagine uma reta comum entre dois pontos, faça buracos na reta e repita essa operação infinitamente. Esse processo foi idealizado pelo matemático alemão *Georg Cantor*, cujo resultado é um objeto conhecido como “*poeira de Cantor*”, que apresenta dimensão fractal entre 0 e 1.

Teoricamente, é possível determinar a dimensão fractal de um objeto se conhecermos seu processo de formação (PINTO, 2001). Um segmento de reta, por exemplo, que possui uma propriedade chamada auto-similaridade, pode ser dividido em N partes idênticas tais que sejam escalonadas pelo fator $s = 1/N$ de um todo. Da mesma forma, um objeto 2D, pode ser dividido em N áreas auto-similares tais que sejam escalonadas pelo fator $s = 1/N^{1/2}$. Genericamente, um objeto auto-similar pode ser dividido em N cópias menores dele mesmo, tal que seja escalonada pelo fator s , como formalizado na Equação 3.2 (MANDELBROT, 1977):

$$s = \frac{1}{\sqrt[D]{N}} \therefore N(s) = \frac{1}{s^D} \therefore N(s) = s^{-D} \quad (3.2)$$

Assim, sua dimensão fractal pode ser estimada através da Equação 3.3:

$$D = \frac{\log(N)}{\log(1/s)} \quad (3.3)$$

Um fato importante que deve ser ressaltado sobre objetos fractais, é que eles são idealizações que podem não existir na natureza e nem mesmo ser corretamente representadas pelos computadores. As principais razões disso são: (i) a auto-similaridade infinita não pode ser representada na natureza e (ii) as limitações do processo de digitalização das imagens tendem a extinguir pequenos detalhes da imagem. Dessa forma, é necessário assumir que formas fractais possuem fractalidade limitada (COSTA e CESAR, 2000).

Para ilustrar como a dimensão fractal de objetos regulares pode ser estimada utilizaremos a curva de Koch (*Koch's Curve*). Essa curva pode ser construída a partir de um segmento de reta dividida em três partes idênticas. Em seguida, a parte intermediária é substituída por dois segmentos, contendo o mesmo tamanho e a mesma forma das três partes iniciais (1/3 da sua parte original). Como resultado da regra de construção, temos um segmento da curva de Koch composto por quatro subsegmentos. O processo completo para construção da curva fractal é obtido aplicando a regra de construção *ad infinitum* aos quatro subsegmentos iniciais (COSTA e CESAR, 2000). Assim, a dimensão fractal (DF) da curva de *Koch* pode ser expressa através da Equação 3.4:

$$DF = \frac{\log(4)}{\log(3)} \cong 1.26 \quad (3.4)$$

A Figura 3.2 demonstra as iterações de um dos lados do triângulo inicial, da curva de *Koch*. O processo iterativo de construção da curva de Koch apresenta uma das características principais dos fractais, a auto-similaridade (SHROEDER, 1996).

Figura 3.2: Iterações para criação da curva de Koch.

Na literatura podem ser encontradas inúmeras aplicações da dimensão fractal em diferentes áreas, como: medicina (DOUGHERTY e HENEBRY, 2001), física (LÓPEZ e VÍLCHEZ-CABRERIZO, 2001), neuromorfometria (COSTA, MANOEL *et al.*, 2002), banco de dados (clustering) (BARBARÁ e CHEN, 2000), botânica (PLOTZE, FALVO *et al.*, 2003), análise de páginas web (CONSULARO, 2000), compressão de

imagens (TURNER, BLACKLEDGE *et al.*, 1998), densidade de multidões (NILCEU, COSTA *et al.*, 1999), dentre outras.

Um amplo número de técnicas e abordagens para estimação da dimensão fractal têm sido reportados na literatura nas últimas décadas (TURNER, BLACKLEDGE *et al.*, 1998):

- *Walking-Divider*: método introduzido por *Shelberg* que é baseado na relação entre o tamanho e o número de segmentos de reta necessários para cobrir a curva fractal (TURNER, BLACKLEDGE *et al.*, 1998);
- *Box-counting*: proposto *Voss* (VOSS, 1988), porém modificado por diversos autores como *Sarkar* (SARKAR e CHAUDHURI, 1992, 1994), este método envolve a contagem do número caixas necessárias para cobrir toda a área da imagem;
- *Prism counting*: baseado na idéia do método *Box-counting* este método foi proposto por *Clarke*, no qual a área da imagem é coberta por triângulos ao invés de caixas (TURNER, BLACKLEDGE *et al.*, 1998);
- *Divider-Step*: método proposto por *Voss* (VOSS, 1988) é uma das abordagens mais simples para estimação da dimensão fractal de objetos. Uma abordagem mais recente do método pode ser encontrada em (CARLIN, 2000).
- *Bouligand-Minkowsky*: é um método que analisa a relação entre a área do objeto e o espaço que ele ocupa (TRICOT, 1995).
- *Multiescala*: introduzido por Costa (COSTA, 1999) permite acompanhar as variações da dimensão fractal do objeto em função da variação da escala do espaço que ele ocupa.

Diversas abordagens e aplicações para dimensão fractal foram propostas ao longo dos anos. Nas próximas seções serão descritos detalhadamente três métodos para estimativa da dimensão fractal: *Box-counting*, *Bouligand-Minkowsky* e *Multiescala*.

3.2.1 Método *Box-counting*

O método *Box-counting* é um dos mais populares algoritmos para estimar a dimensão fractal em imagens (TURNER, BLACKLEDGE *et al.*, 1998). A técnica é baseada na divisão da imagem em caixas quadradas de tamanho ε e, na contagem do número $N(\varepsilon)$ dessas caixas que contenham pelo menos uma porção do objeto em análise. Quando o tamanho das caixas é alterado, quantidades distintas de caixas podem ser computadas. Conseqüentemente, haverá um número menor de caixas quanto maior for o tamanho de ε e, um número cada vez maior de caixas quanto menor for o tamanho de ε . A seqüência de variação dos tamanhos das caixas começa a partir da imagem inteira e, usualmente são reduzidas pela metade de um nível para outro. A Figura 3.3 apresenta um exemplo de uma imagem mapeada com diferentes tamanhos de caixas.

Figura 3.3: Exemplo de uma imagem mapeada com diferentes tamanhos de caixas ε .

A dimensão fractal pelo método *Box-counting* é baseada na relação entre o tamanho das caixas e o número de caixas contadas, sendo definida pela Equação 3.5 (PEITGEN e SAUPE, 1988; VOSS, 1988):

$$d \sim \frac{\log N(\varepsilon)}{-\log \varepsilon} \quad (3.5)$$

A Figura 3.4 ilustra a estimativa da dimensão fractal de um objeto através do método *Box-counting*, que foi aproximadamente 1.0861. Os tamanhos das caixas ε utilizadas para contagem $N(\varepsilon)$ foram: 64, 32, 16, 8, 4 e 2 *pixels*. Na Figura 3.4a é apresentado o objeto em análise e na Figura 3.4b o gráfico $\log(N(\varepsilon)) \times \log(\varepsilon)$. Além do respectivo valor da dimensão fractal do objeto.

(a)

(b)

Figura 3.4: Estimação da dimensão fractal através do método *Box-counting*.
(a) imagem original; (b) gráfico $\log(N(\varepsilon)) \times \log(\varepsilon)$

Como ilustrado na Figura 3.4b, a dimensão fractal de um objeto pode ser obtida calculando todos os respectivos $N(\varepsilon)$ dos diversos valores de ε . Em seguida, deve-se calcular a regressão linear desse conjunto de pontos. Assim, o módulo do coeficiente angular da reta da regressão linear é o valor da dimensão fractal do respectivo objeto.

3.2.2 Método *Bouligand-Minkowsky*

O método *Bouligand-Minkowsky*, ou simplesmente *Minkowsky*, é uma abordagem alternativa para estimativa da dimensão fractal. Ele é baseado na relação entre área de influência do objeto em análise e o espaço que ele ocupa. A área de influência pode ser definida como um conjunto de pontos localizados a uma distância d do objeto. Dessa forma, a área pode ser calculada dilatando o objeto através de um disco de diâmetro d , esse procedimento é conhecido como *Minkowsky's Sausage* (TRICOT, 1995).

Para determinar a dimensão fractal através do método de *Minkowsky*, é necessário analisar como a área de influência cresce quando o disco de diâmetro d aumenta. Quando o objeto é um simples ponto, a área de influência tende a crescer continuamente, por outro lado, quando sua forma é preenchida a área de influência tende a saturar (COSTA e CESAR, 2000).

O método de *Minkowsky* obtém a estimativa da dimensão fractal analisando os pontos do gráfico bi-log da área de influência *versus* o diâmetro d , ou mais especificamente: $\log(d) \times \log A(d)$. Semelhante ao método *Box-counting*, o método de *Minkowsky* calcula o coeficiente angular da reta, obtida através da regressão linear do

gráfico bi-log, para obter o valor da dimensão fractal do objeto. Assim, com o método de *Minkowsky* a dimensão fractal df é determinada através da Equação 3.6:

$$df \approx 2 - \frac{\log A(d)}{\log(d)} \quad (3.6)$$

A Figura 3.5 ilustra o processo de estimação da dimensão fractal através do método de *Minkowsky*. A Figura 3.5a apresenta a imagem original (cuja df é aproximadamente 1.1187) e, na Figura 3.5b a respectiva imagem após sucessivas dilatações. Na Figura 3.5c o gráfico bi-log $\log(d) \times \log A(d)$ é apresentado com a reta resultante da regressão linear.

Figura 3.5: Estimação da dimensão fractal através do método de *Minkowsky*.
(a) imagem original; (b) imagem após sucessivas dilatações; (c) gráfico bi-log contendo a reta resultante da regressão linear

3.2.3 Método Multiescala

O método de estimação da dimensão fractal multiescala possibilita a análise da variação da dimensão fractal de um objeto, em função da variação do espaço métrico no qual este objeto se encontra. Diferentes dos métodos tradicionais para estimação da dimensão fractal que resultam em um único valor, a dimensão fractal multiescala é uma função de escala espacial (COSTA, 1999; MANOEL, COSTA *et al.*, 2002).

O modo no qual objetos são analisados, em determinados momentos ou lugares, é a base do conceito de escala espacial. Dependendo da escala em que o objeto é estudado podemos extrair diversos tipos de informações. Considerando uma fruta, por exemplo, visualmente várias características podem ser extraídas como: formato, qualidade para consumo, cor, dentre outras. Por outro lado, a mesma fruta agora

examinada microscopicamente revela outros tipos de características, mas qual a relação entre esse mesmo objeto analisado em escalas tão diferentes? Existem inúmeras relações pois propriedades e fenômenos ocorrendo em uma dessas escalas poderão estar relacionados a efeitos em outras. Para ilustrar essa relação tomamos como exemplo a alteração da pigmentação de organismos vegetais, como são os casos das frutas. Esse é um processo que ocorre nas células da casca das frutas (em escala microscópica), porém seu resultado pode ser observado visualmente sem a necessidade de um microscópio. Essa riqueza de informações ao longo das diversas escalas espaciais dos objetos, que são relevantes para estudos de determinados problemas, motivou a criação do método multiescala (COSTA e BIANCHI, 2002).

A utilização da dimensão fractal para caracterizar objetos reais na natureza é limitada, principalmente porque esses objetos não são essencialmente fractais (ou auto-similares) (COSTA, 2003). Com isso, qualquer objeto com tamanho finito possui fractalidade próxima de zero quando observado em escalas cada vez maiores, ou seja, todo objeto quando observado em escalas maiores parece com um ponto (dimensão fractal = zero). Nesses casos a abordagem multiescala é capaz de trazer informações mais substanciais do que os métodos usuais de dimensão fractal, principalmente porque é capaz de analisar o objeto em função da variação da escala do espaço que ele ocupa.

O processo para estimar a dimensão fractal através da abordagem multiescala se inicia com a dilatação sucessiva do objeto em análise por um raio d (método de dilatações exatas). Essa etapa pode ser considerada semelhante ao que ocorre no método de *Minkowsky*. Considerando as áreas $A(d)$ assumida pelo objeto enquanto ele é dilatado em função do raio, é possível calcular um gráfico bi-log que é obtido plotando $\log A(d) \times \log(d)$. Ao contrário do que ocorre nos métodos tradicionais que fazem a regressão linear sobre a curva do gráfico bi-log para determinar o valor da dimensão fractal, o método multiescala é diferente. O principal problema desses métodos é que como a curva não resulta em uma reta de regressão perfeita, valores substancialmente distintos podem ser obtidos dependendo do intervalo ($\log(d)$) escolhido. Utilizando a derivada numérica (primeira derivada) de $\log A(d)$, o método multiescala supera completamente essa subjetividade além de prover informações adicionais sobre características do objeto em análise (MANOEL, COSTA *et al.*, 2002). Dessa forma, para obtenção da análise multiescala da dimensão fractal é necessário que seja obtido

um gráfico $\log(d) \times \log(A')$, no qual A' é o resultado da derivada numérica do ponto $(d, A(d))$ do gráfico bi-log. A Figura 3.6 ilustra o processo para estimativa da dimensão fractal através do método multiescala.

Figura 3.6: Processo da estimativa da dimensão fractal multiescala. (a) imagem original; (b) dilatações sucessivas da imagem; (c) gráfico bi-log $\log A(d) \times \log(d)$; (d) gráfico da dimensão fractal multiescala.

O gráfico da dimensão fractal multiescala (Figura 3.6d), além de apresentar uma abrangente caracterização da complexidade, propriedades importantes do objeto em estudo podem ser analisadas como (MANOEL, COSTA et al., 2002):

- Pico de fractalidade ($F_{\max}$): indica o valor máximo de complexidade observado ao longo da função fractal multiescala em todas as escalas espaciais;
- Escala de máxima fractalidade ($D_{\max}$): Quando duas formas possuem valores de $F_{\max}$ iguais, porém o valor de $D_{\max}$ em uma delas é menor, então essa forma se caracteriza por ter complexidade em menor escala. Assim $D_{\max}$ indica a escala espacial na qual a curva de dimensão fractal multiescala atinge seu valor máximo;
- Largura de alta fractalidade (L): a largura da curva pode ser utilizada como parâmetro de complexidade do objeto analisado. A curva pode ser cortada em uma determinada altura com valor fixo, pré-definido, ou pela metade. Quando o valor de L é elevado indica que o objeto possui complexidade alta em diversas escalas espaciais.

Essas propriedades ($F_{\max}$, $D_{\max}$ e L) podem ser conciliadas para descrever a complexidade do objeto analisado, originando informações mais completas do que as obtidas com os métodos tradicionais. A Figura 3.7 ilustra os três importantes parâmetros utilizados para caracterização de objetos: pico de fractalidade, escala de máxima fractalidade e largura de alta fractalidade.

Figura 3.7: Ilustração de propriedades extraídas do gráfico da dimensão fractal multiescala: pico de fractalidade ($F_{\max}$), escala de máxima fractalidade ($D_{\max}$) e largura de alta fractalidade (L) (adaptado de (COSTA e BIANCHI, 2002)).

3.3 Esqueletonização

No início da evolução dos computadores uma das suas principais aplicações era o reconhecimento de padrões. Entretanto, a grande quantidade de informações e o baixo poder computacional, motivaram diversos pesquisadores ao desenvolvimento de algoritmos capazes de reduzir a quantidade de dados a ser processada. Esses métodos computacionais são conhecidos como algoritmos de esqueletonização. Geralmente esses algoritmos são utilizados para obter o esqueleto de uma região através do seu afinamento. Afinamento é o processo de redução de uma forma para uma versão simplificada que ainda retém as características essenciais do objeto original (BAJA e THIEL, 1996). A versão afinada de uma forma é chamada de esqueleto (GONZALEZ e WOODS, 2002). Na literatura há muita discordância sobre a nomenclatura utilizada no processo de esqueletonização. Alguns autores usam termos diferentes, como por exemplo: eixo médio, afinamento, esqueletização, etc. Contudo, afinamento e esqueletonização tornaram-se os mais utilizados.

Os trabalhos pioneiros da área foram concentrados no reconhecimento de padrões em caracteres manuscritos na década de 1950 (LAM, LEE *et al.*, 1992). Hoje em dia diversas aplicações podem ser encontradas para os algoritmos de

esqueletonização como: neuromorfometria (FALCÃO, COSTA *et al.*, 2002), análise de impressões digitais (VAJNA, ROVATTI *et al.*, 2000), reconhecimento de caracteres (HE e YAN, 2000; MELHI, IPSON *et al.*, 2001), etc. A Figura 3.8 apresenta um exemplo da técnica de esqueletonização empregada em uma imagem de impressão digital.

Figura 3.8: Exemplo da técnica de esqueletonização em uma imagem de impressão digital: (a) imagem original; (b) imagem esqueletonizada (SMITH, 1999).

O esqueleto de uma região pode ser definido pela transformação do eixo medial (*medial axis transformation*) – MAT, também conhecida como *symmetry axis transform* – SAT, proposta por Blum (BLUM, 1967). Fundamentalmente a MAT de uma forma específica corresponde a todas possíveis posições do centro de um círculo que satisfaçam as seguintes condições (PAVLIDIS, 1982): (i) ser bitangente à forma, ou seja, a curva deve tocar a forma em dois pontos distintos; (ii) estar completamente dentro da forma.

Podemos imaginar a MAT pela idéia de propagação do fogo no campo: um gramado em chamas. Se o início do fogo começar simultaneamente em todos os pontos ao longo da borda do campo e, assumirmos que o fogo se propaga em tempos constantes. As posições onde o fogo se extingue correspondem a MAT da forma.

A MAT trata todos os *pixels* limites da imagem, ou seja, *pixels* da borda, como uma frente de onda (PARKER, 1997). Cada um desses *pixels* excita seus vizinhos com um intervalo de tempo proporcional à distância. Assim, ele também torna-se parte da frente de onda. O conceito de “perto” depende da definição de distância, e, portanto, o resultado de uma operação MAT é influenciado pela medida de distância adotada. A sua implementação envolve o cálculo da distância de todos os pontos interiores de uma

imagem para todos os pontos da borda da mesma. As ondas passam através de cada ponto somente uma vez, e quando duas ondas se encontram, cancelam-se uma a outra, produzindo o eixo médio ou o esqueleto (GONZALEZ e WOODS, 2002). A Figura 3.9 apresenta uma forma simples, um retângulo, e sua respectiva MAT.

Figura 3.9: Transformação do eixo medial em uma forma simples.
(a) imagem original; (b) imagem após a MAT (GONZALEZ e WOODS, 2002).

Os algoritmos de esqueletonização excluem de forma sucessiva diversas camadas da extremidade (borda) de um padrão até que apenas o esqueleto permaneça. A exclusão de um ponto p dependerá dos *pixels* da vizinhança deste ponto. De acordo com o modo de como se examinam os *pixels*, estes algoritmos podem ser classificados como (LAM, LEE *et al.*, 1992):

- Iterativos;
- Não-iterativos.

Os algoritmos iterativos são aqueles que produzem o esqueleto através da exclusão repetitiva das características da borda do objeto. Eles podem ser subdivididos em duas categorias: seqüenciais e paralelos (LAM, LEE *et al.*, 1992). Nos algoritmos seqüenciais, os *pixels* são examinados para exclusão em uma seqüência fixa em cada iteração, e a exclusão de p na n -ésima iteração depende de todas as operações que tenham sido realizadas até aquele momento. Isto é, depende do resultado da $(n-1)$ -ésima iteração, bem como dos *pixels* já processados na n -ésima iteração. Por outro lado, nos algoritmos paralelos, a exclusão na n -ésima iteração depende apenas dos *pixels* da iteração $n-1$. Por isso, todos os *pixels* podem ser analisados independentemente, de forma paralela a cada iteração.

Um fator importante que deve ser considerado sobre a classificação dos algoritmos de esqueletonização iterativos, em seqüenciais ou paralelos, é que mesmo algoritmos que pertencem a uma mesma classe (seqüencial ou paralela) podem ser

muito diferentes. De forma geral, a diferença maior está nos testes implementados para garantir a conectividade da forma (objeto) a ser esqueletonizada.

Os algoritmos de afinamento não-iterativos são aqueles que extraem o esqueleto de um objeto a partir de uma varredura na imagem (NEUSIUS e OLSZEWSKI, 1994). Eles produzem o esqueleto do objeto diretamente em um único passo sem examinar individualmente todos os *pixels*. De certa forma esses algoritmos são mais intuitivos que os métodos iterativos e geram esqueletos que conservam propriedades globais, além disso, conseguem manter a conectividade do objeto durante o processo de afinamento.

Como foi apresentando, desde a década de 50, diversos algoritmos foram desenvolvidos e inúmeras abordagens foram propostas. Na literatura, centenas de artigos podem ser encontrados, nos quais a grande maioria se concentra na variação de um algoritmo de esqueletonização já desenvolvido, procurando relatar novos aspectos da performance dos métodos. Nas próximas seções alguns desses métodos serão descritos levando em consideração a contribuição de novos conceitos para área e, a aplicabilidade nos experimentos com órgãos foliares. Antes de apresentarmos os algoritmos de esqueletonização, algumas definições se fazem necessárias:

Definição 1: A vizinhança de um *pixel* $P1$ é representada como $N(P1)$. É definida como o conjunto de todos os vizinhos de $P1$, chamados de vizinhança de $P1$. A Figura 3.10 ilustra o *pixel* $P1$ e seus respectivos vizinhos.

P9 (i-1,j-1)	P2 (i-1,j)	P3 (i-1,j+1)
P8 (i,j-1)	P1 (i,j)	P4 (i,j+1)
P7 (i+1,j-1)	P6 (i+1,j)	P5 (i+1,j+1)

Figura 3.10: Os 8 vizinhos do *pixel* $P1$.

Definição 2: $A(P1)$ representa o número de transições de $P1$, ou seja, $A(P1)$ é o número de transições de 0 para 1 na sequência de $P2, P3, P4, P5, P6, P7, P8, P9$ e $P2$. Essa condição executa um teste de conectividade. De fato, considerando as imagens da Figura 3.11a ou Figura 3.11b em que $A(P1) > 1$, observa-se que mudando o valor de $P1$ para branco o objeto apresentará uma descontinuidade.

Figura 3.11: Número de transição de 0 para 1 na sequência de P2, P3, ..., P9, P2.
(a) uma transição; (b) duas transições.

Definição 3: $B(P1)$ é o número de vizinhos com valor diferente de zero do ponto P1. A Figura 3.12 apresenta dois exemplos do número de vizinhos de P1. Essa condição assegura que nenhum *pixel* isolado seja apagado.

Figura 3.12: Número de vizinhos com valor diferente de zero do ponto P1.
(a) Dois vizinhos; (b) Cinco vizinhos.

Definição 4: Um ponto P1 é um ponto extremo se $B(P1) = 1$. Ou seja um ponto extremo é um *pixel* preto que tem apenas outro *pixel* preto como seu vizinho. Esses *pixels* não podem ser removidos. A Figura 3.13 ilustra a situação de um ponto extremo.

Figura 3.13: Ponto extremo, o *pixel* tem apenas um *pixel* preto como seu vizinho.

3.3.1 Algoritmo de Hilditch

O método de esqueletonização de *Hilditch* (RUTOVITZ, 1966) é um algoritmo baseado nos *pixels* da imagem, ou seja, é um método iterativo. O funcionamento do algoritmo consiste na aplicação de um conjunto de regras para decidir se o valor do *pixel* deve ser mudado de preto para branco.

Inicialmente devemos considerar uma janela 3x3 cujo ponto central é chamado de P1 com seus respectivos vizinhos. Denota-se por $A(P1)$ o número de transições de zero para um na seqüência ordenada de $p2...p9$, $p2$ e, $B(P1)$ o número de vizinhos diferentes de zero de P1. Então, a cada passada do algoritmo (em paralelo) devemos remover os *pixels* que satisfaçam as condições da Tabela 3.1.

TABELA 3.1: Condições para remoção de *pixels* com algoritmo de *Hilditch*

C1	$2 \leq B(p1) \leq 6$
C2	$A(p1) = 1$
C3	$(p2 \vee p4 \vee p9 = 0) \vee (A(p2) \neq 1)$
C4	$(p2 \vee p4 \vee p6 = 0) \vee (A(p4) \neq 1)$

A execução do algoritmo termina quando ao examinar todo conjunto de *pixels* nenhum valor (de preto para branco) é alterado. Ele é considerado um algoritmo paralelo-sequencial. É paralelo pois durante a varredura da imagem todos os *pixels* podem ser considerados ao mesmo tempo e, com isso, é decidido se um *pixel* deve ou não ser removido. E ao mesmo tempo é um algoritmo sequencial porque o conjunto de condições é repetido várias vezes, até que nenhuma mudança ocorra (de preto para branco). A Figura 3.14 apresenta o resultado da aplicação do algoritmo de *Hilditch* sobre duas imagens binárias.

Figura 3.14: Exemplo do resultado do processo de esqueletonização através do algoritmo de *Hilditch*. (a) imagem original; (b) imagem esqueletonizada.

3.3.2 Algoritmo de *Stentiford*

O algoritmo de *Stentiford* (STENTIFORD, 1983) introduziu uma nova abordagem para os algoritmos de esqueletonização: o conceito de máscaras. Para executar a esqueletonização de uma imagem binária o algoritmo utiliza quatro mascaras que devem ser aplicadas sucessivamente de forma ordenada na imagem. Na Figura 3.15 são ilustrados os quatro diferentes tipos de mascaras utilizados pelo algoritmo de *Stentiford*.

Figura 3.15: Os quatro tipos de máscaras utilizadas para erosão das imagens através do algoritmo de *Stentiford*.

Nas máscaras, o círculo branco representa que o valor do *pixel* é 0, por outro lado, o círculo preto representa que o valor do *pixel* é 1. No caso do valor “X” não importa qual o valor do *pixel*, seja ele branco ou preto. Essas máscaras devem percorrer a imagem na seguinte ordem: *M1* – da esquerda para a direita e de cima para baixo; *M2* – de baixo para cima e da esquerda para a direita; *M3* – da direita para a esquerda e de baixo para cima; *M4* – de cima para baixo e da direita para a esquerda;

O algoritmo de *Stentiford* pode ser descrito em seis passos: (1) percorrer a imagem até encontrar um *pixel* que se encaixe na máscara *M1*; (2) Se este *pixel* não for um ponto extremo e se o seu número de vizinhos $B(P1)$ for igual a 1, marcar este ponto

para que seja apagado mais tarde; (3) Repetir os passos 1 e 2 para todos os *pixels* que se encaixem na máscara M1; (4) Repetir os passos, 1,2 e 3 para as outras máscaras M2, M3 e M4 nesta ordem; (5) Se algum ponto estiver marcado para ser removido, seu valor deve ser mudado para branco; (6) Se algum ponto foi removido no passo 5, repetir todos os passos a partir do passo 1. Senão, o processo termina. A Figura 3.16 apresenta uma imagem eskeletonizada com o algoritmo de *Stentiford*.

Figura 3.16: Exemplo da aplicação do algoritmo de *Stentiford*;
(a) imagem original; (b) imagem eskeletonizada

Da mesma forma do algoritmo de *Hilditch*, o método de *Stentiford* funciona somente para alguns tipos de imagens. Dependendo da imagem o objeto resultante pode apresentar descontinuidades nos *pixels*, o que é uma característica ruim para um algoritmo de eskeletonização. A causa provável desta descontinuidade é alguma falha no processo que verifica o número de conectividade de um *pixel* ou na forma com que as máscaras são aplicadas na imagem.

3.3.3 Algoritmo de Zhang e Suen

Em 1984, *Zhang e Suen* (ZHANG e SUEN, 1984) publicaram um artigo no qual foi proposto um novo algoritmo paralelo de eskeletonização. Este trabalho trouxe resultados surpreendentes quando comparado a outros métodos da sua época. Anos mais tarde, muitos pesquisadores, inclusive o próprio *Zhang*, sugeriram novos testes e metodologias que melhoraram ainda mais o algoritmo de *Zhang-Suen*.

O algoritmo de *Zhang-Suen* consiste em sucessivas aplicações ao contorno da imagem de duas regras, sendo que os pontos do contorno são quaisquer pontos com valor 1 que tenham ao menos um dos seus oito vizinhos iguais a 0. Como o algoritmo

proposto *Zhang-Suen* é paralelo, vale lembrar que os *pixels* são examinados para exclusão baseados apenas na iteração anterior. Assim, o algoritmo é composto por duas iterações: Na primeira iteração o *pixel* P1 será excluído se todas as quatro condições da Tabela 3.2 forem satisfeitas.

TABELA 3.2: Condições da primeira iteração para remoção de *pixel* com algoritmo de *Zhang-Suen*

C1	$2 \leq B(p1) \leq 6$
C2	$A(p1) = 1$
C3	$(p2 \vee p4 \vee p6 = 0)$
C4	$(p4 \vee p6 \vee p8 = 0)$

Na segunda iteração, as linhas C3 e C4 são substituídas por suas rotações de 180°, sendo assim, o *pixel* P1 será excluído se as condições da Tabela 3.3 forem satisfeitas.

TABELA 3.3: Condições da segunda iteração para remoção de *pixel* com algoritmo de *Zhang-Suen*

C1	$2 \leq B(p1) \leq 6$
C2	$A(p1) = 1$
C3	$(p2 \vee p4 \vee p8 = 0)$
C4	$(p2 \vee p6 \vee p8 = 0)$

A Figura 3.17 apresenta o processo de esqueletonização de uma imagem com o algoritmo de *Zhang-Suen*. A Figura 3.17a apresenta a imagem original e a Figura 3.17b, o resultado obtido pelo processo.

Figura 3.17 Processo de esqueletonização de *Zhang-Suen* em uma imagem de folha; (a) imagem original; (b) imagem esqueletonizada.

3.3.4 Algoritmo de Holt

Durante vários anos, a maioria dos algoritmos de esqueletonização funcionava através da aplicação sucessiva de um conjunto de regras em uma imagem. Em 1987, *Holt* (HOLT, STEWART *et al.*, 1987) sugeriu um novo algoritmo que não envolvia iteração, além disso, era um dos mais rápidos na época. A idéia proposta por *Holt* foi transformar os dois conjuntos de regras do método de *Zhang-Suen* em expressões lógicas. A Equação 3.6 apresenta o conjunto de condições da primeira iteração do algoritmo de *Zhang* representado através de expressão lógica.

$$v(C) \wedge (\neg edge(C) \vee (v(L) \wedge v(S) \wedge (v(N) \vee v(O)))) \quad (3.6)$$

Da mesma forma, a Equação 3.7 apresenta a segunda iteração descrita na através de uma expressão lógica.

$$v(C) \wedge (\neg edge(C) \vee (v(O) \wedge v(N) \wedge (v(S) \vee v(L)))) \quad (3.7)$$

Se o resultado das expressões lógicas for falso, o *pixel* deve ser removido, caso contrário o estado do *pixel* não deve ser alterado. O significado das funções descritas nas expressões lógicas é dado como se segue:

- *Função $v()$* : representa o valor do *pixel*. O resultado será verdadeiro se o *pixel* representar parte do objeto, ou seja, possuir o valor preto. Caso contrário, a função recebe falso quando o *pixel* representar parte do fundo da imagem, ou seja, possuir valor branco.
- *Função $edge()$* : o resultado será verdadeiro se o *pixel* pertencer a borda do objeto e falso, caso contrário. Um *pixel* que pertence à borda da imagem é um *pixel* que possui conectividade igual a 1, ou seja, $A(P1) = 1$.

Para representar a vizinhança dos pontos *Holt* utilizou as coordenadas cardinais ao contrário dos números empregados pela maioria dos autores. A Figura 3.18 ilustra a vizinhança de um *pixel* C utilizando coordenadas cardinais.

NO (i-1,j-1)	N (i-1,j)	NE (i-1,j+1)
O (i,j-1)	C (i,j)	L (i,j+1)
SO (i+1,j-1)	S (i+1,j)	SE (i+1,j+1)

Figura 3.18: Vizinhança do *pixel* C utilizando coordenadas cardinais.

Por fim, *Holt* combinou as duas expressões lógicas apresentadas: Equação 3.6 (primeira iteração de *Zhang-Suen*) e Equação 3.7 (segunda iteração de *Zhang-Suen*). O resultado da junção das duas expressões lógicas é apresentado na Equação 3.8.

$$v(C) \wedge (\neg edge(C) \vee (edge(L) \wedge v(N) \wedge v(S)) \vee v(edge(S) \wedge v(O) \wedge v(L)) \vee (edge(L) \wedge edge(SE) \wedge edge(S))) \quad (3.8)$$

Na tentativa de melhorar o resultado da expressão acima, um novo conceito foi introduzido por *Holt*: a remoção em escada (*staircase removal*). O processo de remoção em escada explora a seguinte propriedade: metade dos *pixels* que apresentam uma forma semelhante a uma escada pode ser removida sem afetar o formato ou a conectividade do objeto. Na Figura 3.19, o *pixel* central das máscaras pode ser removido se um dos valores do *pixel* X for igual a zero (branco).

0	1	X	X	1	0	0	X	X	X	X	X	X
1	1	X	X	X	1	1	X	1	1	1	1	X
X	X	0	X	0	X	X	X	1	0	0	1	X
máscara 1			máscara 2			máscara 3			máscara 4			

Figura 3.19: As máscaras utilizadas pelo algoritmo de *Holt* com o conceito de remoção em escada.

Da mesma maneira que ocorreu com as iterações do algoritmo de *Zhang-Suen*, o novo conceito (remoção em escada) foi mapeado através de expressões lógicas. A Equação 3.9 apresenta a expressão lógica que captura a situação de remoção em escada para as máscaras 1 e 2. E a Equação 3.10 apresenta a expressão lógica capaz de executar o processo de remoção em escada para as máscaras 3 e 4:

$$v(C) \wedge \neg(v(N) \wedge ((v(L) \wedge \neg v(NE) \wedge \neg v(SO) \wedge (\neg v(O) \vee \neg v(S)) \vee v(v(O) \wedge \neg v(NO) \wedge \neg v(SE) \wedge (\neg v(L) \vee \neg v(S))))) \quad (3.9)$$

$$v(C) \wedge \neg(v(S) \wedge ((v(L) \wedge \neg v(NE) \wedge \neg v(SO) \wedge (\neg v(O) \vee \neg v(N)) \vee v(v(O) \wedge \neg v(NO) \wedge \neg v(SE) \wedge (\neg v(L) \vee \neg v(N))))) \quad (3.10)$$

A Figura 3.20 apresenta um exemplo do algoritmo de esqueletonização de *Holt* aplicado em imagens binárias. A Figura 3.20b demonstra o resultado do algoritmo proposto inicialmente e, a Figura 3.20c o resultado do método com o conceito de remoção em escada.

Figura 3.20: Resultado do algoritmo de esqueletonização de *Holt*.
(a) imagem original; (b) imagem esqueletonizada com o método inicialmente proposto;
(c) imagem esqueletonizada com o conceito de remoção em escada.

O algoritmo proposto por *Holt* utilizando a fórmula de remoção em escada constitui uma poderosa ferramenta para esqueletonização de imagens, superando os resultados do algoritmo de *Zhang-Suen* (ZHANG e SUEN, 1984), sendo mais rápido e simples de implementar.

3.3.5 Algoritmo Multiescala

A maioria dos principais algoritmos de esqueletonização é baseada nos conceitos introduzidos por Blum em 1967 (BLUM, 1967): *medial axis transform* (MAT) também conhecida como *symmetry axis transform* (SAT). Particularmente a SAT de um objeto contém inúmeras propriedades que podem ser utilizadas por uma enorme gama de aplicações, porém ela contém um grave problema. Os algoritmos de esqueletonização baseados nessa abordagem são altamente sensíveis a pequenas distorções ou detalhes do contorno do objeto, ou seja, são facilmente susceptíveis a ruídos da forma. A Figura 3.21 ilustra como um pequeno ruído na borda do objeto pode influenciar no resultado do processo de esqueletonização.

Figura 3.21: O problema principal da SAT em objetos com regiões ruidosas. (a) SAT de uma região sem ruídos; (b) SAT da mesma região com um pequeno ruído.

Com o objetivo de solucionar os problemas apresentados pelos métodos de esqueletonização baseados em SAT, uma nova abordagem de algoritmos de esqueletonização foi introduzida por Costa: a esqueletonização multiescala (COSTA, 2000). A técnica proposta é um método simples e relativamente eficiente para calcular a SAT em imagens binárias, o qual utiliza o conceito de dilatações exatas (COSTA e CESAR, 2000).

O algoritmo esqueletonização multiescala começa rotulando, de forma sucessiva, o contorno do objeto com valores inteiros. A escolha do ponto inicial e o sentido (horário ou anti-horário) na rotulação dos elementos do contorno podem gerar

resultados (esqueletos) relativamente diferentes. Algoritmos para extração de contorno, ou seguidores de contorno, podem ser empregados nessa etapa (um método de *contour-following* é descrito em (COSTA e CESAR, 2000)). A Figura 3.22 ilustra o processo de rotulação do contorno de um objeto. Na Figura 3.22a é apresentada uma representação matricial de um objeto qualquer e, na Figura 3.22b o resultado do processo de rotulação sucessiva do contorno.

Figura 3.22: Processo de rotulação sucessiva dos elementos do contorno do objeto.
(a) representação matricial de um objeto; (b) resultado do processo de rotulação sucessiva.

Em seguida, o próximo passo do algoritmo de esqueletonização é a propagação dos elementos (rótulos) do contorno do objeto através do conceito de dilatações exatas. A matriz img_lbl resultante do processo de dilatação exata (empregado na matriz da Figura 3.22b) é representada na Figura 3.23.

	1	1	2	3	3		
1	1	1	2	3	3	4	
10	1	1	2	3	3	4	4
10	10	1	2	3	4	4	5
10	9	9	8	4	4	5	5
9	9	8	8	6	6	5	5
	8	8	7	7	6	6	5
		7	7	7	7	6	

Figura 3.23: Propagação dos rótulos do contorno de um objeto através do conceito de dilatação exata.

Uma vez concluída a propagação dos rótulos por toda imagem, a diferença entre cada elemento de img_lbl e seus respectivos quatro vizinhos é determinada e, inserida na matriz img_dif . Este procedimento pode ser resumidamente explicado no Algoritmo 3.1 (adaptado (COSTA e CESAR, 2000)), no qual N_x e N_y indicam o tamanho da imagem e, o elemento μ corresponde ao número de elementos de contorno do objeto.

Algoritmo de Esqueletonização multiescala

```

01. Rotular o contorno do objeto com elementos inteiros;
02. Propagar esses rótulos com dilatações exatas
03. Para  $x = 2$  até  $N_x - 1$  faça
04.   Para  $y = 2$  até  $N_y - 1$  faça
05.   Início
06.      $vDif = Abs(img\_lbl[x,y] - img\_lbl[x-1,y]);$  // pixel superior
07.      $vMax = Máximo(vMax, Mínimo(vDif, \mu - vDif));$ 
08.      $vDif = Abs(img\_lbl[x,y] - img\_lbl[x,y+1]);$  // pixel direito
09.      $vMax = Máximo(vMax, Mínimo(vDif, \mu - vDif));$ 
10.      $vDif = Abs(img\_lbl[x,y] - img\_lbl[x+1,y]);$  // pixel inferior
11.      $vMax = Máximo(vMax, Mínimo(vDif, \mu - vDif));$ 
12.      $vDif = Abs(img\_lbl[x,y] - img\_lbl[x,y-1]);$  // pixel esquerdo
13.      $vMax = Máximo(vMax, Mínimo(vDif, \mu - vDif));$ 
14.     Se  $vMax \leq \mu$  então
15.        $Img\_dif[x,y] = vMax$ 
16.     Senão
17.        $Img\_dif[x,y] = Abs(\mu - vMax)$ 
18.   Fim

```

Por fim, um conjunto de SAT's espaço-escala pode ser obtido aplicando um simples limiar T_i , com diferentes valores inteiros, na matriz img_dif . A Figura 3.24 ilustra graficamente as etapas do processo de esqueletonização multiescala.

Figura 3.24: Esqueletonização multiescala. (a) imagem original; (b) extração do contorno e rotulação sucessiva; (c) propagação dos rótulos através de dilatações exatas; (d) esqueleto resultante para limiar $T = 3$; (e) $T = 5$; (f) $T = 10$; (g) $T = 15$; (h) $T = 20$; (i) $T = 25$; (j) $T = 50$; (k) $T = 75$ e (l) $T = 100$.

O método de esqueletonização multiescala permite obter os esqueletos internos e externos de um objeto. Para separar os dois tipos de esqueletos uma máscara pode ser empregada na imagem resultante da aplicação do limiar T_i . A Figura 3.25 apresenta o esqueleto interno e externo obtido através da esqueletonização multiescala.

Figura 3.25: Esqueletos obtidos através da esqueletonização multiescala. (a) esqueleto interno; (b) esqueleto externo.

Os experimentos realizados com o algoritmo multiescala demonstraram a eficiência e robustez do método. Contudo dois pontos devem ser relevados: (i) alto custo computacional – consequência das várias etapas do algoritmo (propagação de

rótulos, dilatações exatas, etc.) e (ii) escolha de limiar T_i adequado – dependendo do limiar utilizado o resultado do processo (esqueleto) é substancialmente diferente, sendo necessário a interação com usuário durante as etapas do método. A definição de um valor para limiar varia de acordo com cada aplicação e principalmente com o tipo de informação que se deseja explorar no objeto.

3.4 Biometria

A extração de medidas biométricas, ou simplesmente biometria, são métodos capazes de coletar e estudar medidas de estruturas em órgãos de seres vivos, bem como, analisar a funcionalidade dessas medidas. A implementação e implantação dessas técnicas em sistemas computacionais auxiliam a tarefa dos biometricistas (profissionais que estudam ou aplicam métodos biométricos) na realização de diversas tarefas do dia-a-dia.

No projeto de identificação de órgãos foliares, os métodos biométricos implementados foram empregados na extração de características das folhas, essencialmente das nervuras e do esqueleto. Entretanto, é importante salientar que devido a semelhança entre nervuras e esqueletos com outros elementos da natureza - denominados estruturas tubulares - os métodos desenvolvidos poderão ser empregados em diversas outras aplicações. Na natureza podemos encontrar uma ampla diversidade de estruturas tubulares. Essas estruturas exercem papéis fundamentais em diversos ramos da biologia, caracterizando células, tecidos, órgãos, sistemas de irrigação entre outros. Como exemplos, podemos citar: as raízes das plantas, os troncos e galhos das árvores, os vasos sanguíneos dos seres vivos, os esqueletos dos vertebrados, as ramificações dendríticas dos neurônios, etc. Dessa forma, os métodos implementados são capazes de analisar e descrever geometricamente diversos tipos de estruturas e assim, contribuir para diferentes áreas da biologia. Nos próximos parágrafos esses métodos serão tratados genericamente como biometria de estruturas tubulares, lembrando que nesse projeto eles foram empregados especificamente para biometria de nervuras e esqueletos.

Os métodos desenvolvidos e implementados foram divididos em três categorias: (i) distâncias; (ii) comprimentos e (iii) ângulos. Na primeira categoria, os métodos biométricos referentes às distâncias têm como objetivo extrair medidas com base na

distância entre determinados pontos das estruturas tubulares. Por exemplo, uma medida que pode ser extraída é a distância média entre todas as bifurcações da estrutura tubular. A segunda categoria é responsável por extrair medidas referentes ao comprimento. Um exemplo de característica extraída nessa categoria é o comprimento de uma determinada ramificação, ou, o comprimento da ramificação principal. Por fim, a terceira categoria é responsável por extrair medidas com bases nos ângulos. Com esses métodos é possível extrair, por exemplo, os ângulos entre duas ramificações da estrutura tubular.

As principais características que podem ser extraídas com os métodos biométricos da categoria de distâncias são:

- número de bifurcações: a extração dessa característica é baseada no número de transições $A(P)$ de 0 para 1 dos oito vizinhos de um determinado *pixel* P . Se o número de transições do *pixel* P é maior que 2, ou seja $A(P) > 2$, então esse *pixel* é uma bifurcação. Com o número de bifurcações extraído da estrutura tubular, outras informações também podem ser coletadas, como por exemplo, a distância média entre as posições das bifurcações, desvio padrão, etc. A Figura 3.26 ilustra o resultado do algoritmo que procura as bifurcações em uma estrutura tubular.

Figura 3.26: Imagem resultante do algoritmo capaz de encontrar bifurcações em estruturas tubulares.

- pontos de origem e destino: o ponto de origem de uma estrutura tubular representa um dos pontos principais dessa estrutura. No caso dos órgãos foliares, o ponto de origem pode ser interpretado como o pecíolo da folha. O ponto de destino representa o outro extremo dessa estrutura.

Para extração dessas características, na maioria dos casos, pode ser utilizado um algoritmo simples que: para o ponto de origem varre a imagem binarizada da estrutura, de baixo para cima e da esquerda para direita até encontrar o primeiro *pixel* preto; no caso do ponto de destino a varredura ocorre de cima para baixo e da esquerda para direita. Da mesma maneira que ocorre com o número de bifurcações, extraindo os pontos de origem e destino outras informações podem ser coletadas como: distância média entre as posições das bifurcações e o ponto de origem; desvio padrão das distâncias entre as bifurcações e o ponto de origem; distância entre o ponto de origem e destino; distância entre o centróide da área foliar e a origem, dentre outras.

- pontos extremos: o conceito de pontos extremos de uma estrutura tubular, pode ser entendido como o *último pixel* de uma determinada ramificação. A base para extração dessa característica é o número de vizinhos $B(P)$ de um determinado *pixel* P . Se o número de vizinhos de um *pixel* P é igual a 1, ou seja $B(P) = 1$, então esse *pixel* é um ponto extremo ou ponto terminal. Assim, a partir dos pontos extremos, outras informações podem ser coletadas como: distância média entre o centróide e os pontos extremos; desvio padrão das distâncias entre o centróide e os pontos extremos; distância média entre os pontos extremos; distância média entre o ponto de origem e os pontos extremos; desvio padrão entre as distâncias do ponto de origem aos pontos extremos, dentre outras. A Figura 3.27 apresenta o resultado do algoritmo que extrai (encontra) os pontos extremos de uma estrutura tubular.

Figura 3.27: Extração dos pontos extremos de uma estrutura tubular.

Os métodos biométricos que coletam medidas de comprimentos da estruturas tubulares, podem extrair as seguintes características:

- tamanho da estrutura tubular: o tamanho de uma estrutura tubular pode ser considerado como a área dessa estrutura. De forma geral, o tamanho de uma estrutura tubular é igual ao número de *pixels* pretos dessa estruturas.
- número de ramificações: a contagem do número de ramificações de uma estrutura tubular é de extrema importância para o caso das folhas. Com essa informação é possível descobrir o número de nervuras primárias e secundárias que um órgão foliar possui. Além do número de ramificações, outras informações relevantes sobre as ramificações podem ser coletadas como: tamanho médio das ramificações, desvio padrão do tamanho médio das ramificações, dentre outras.
- tamanho da ramificação principal: essa informação também é muito importante para o caso das nervuras, pois representa o tamanho da nervura principal da folha. Informações sobre nervura principal de um órgão foliar são imprescindíveis para caracterização das espécies. A Figura 3.28 apresenta o resultado do algoritmo que encontra a ramificação principal de uma estrutura tubular.

Figura 3.28: Extração da ramificação principal de uma estrutura tubular.

Na categoria que extraí características a partir dos ângulos, as principais informações coletadas são referentes aos ângulos de abertura e fechamento entre determinadas ramificações. Esses dados podem ser extraídos a partir de qualquer bifurcação previamente identificada na estrutura tubular, contudo, para identificação das espécies foliares foram considerados os ângulos entre as ramificações que partem da ramificação principal.

3.5 Considerações Finais

Neste capítulo foram apresentadas às metodologias principais sobre análise de imagens e biometria, necessárias para desenvolvimento do projeto de identificação de espécies arbóreas. Inicialmente foram relatados conceitos sobre dimensão fractal, no qual três métodos para sua estimação foram descritos: *Box-counting*, *Bouligand-Minkowsky* e multiescala. Uma revisão sobre algoritmos de esqueletonização de imagens binárias foi apresentada na seção 3.3. Alguns algoritmos de esqueletonização foram descritos, levando em consideração a contribuição de novos conceitos para área de pesquisa. A seção 3.4 apresentou os algoritmos para extração de medidas biométricas. Esses algoritmos são capazes de extrair importantes informações sobre estruturas tubulares que, no caso das espécies arbóreas, são imprescindíveis para sua caracterização. Os métodos e algoritmos discutidos ao longo desse capítulo foram utilizados para extração de características e análise de imagens das folhas.

Caracterização de Órgãos Foliares

O crescimento do uso de imagens digitais em diversas áreas de aplicação tem demandado, cada vez mais, o desenvolvimento de métodos para análise de imagens. Necessariamente, esses métodos precisam ser altamente confiáveis e capazes de extrair informações úteis, além de interpretar resultados. A aplicabilidade é comprovada em diversas áreas de conhecimento como: agricultura (SOILLE, 2000), medicina (CHAMARTHY, STANLEY *et al.*, 2004), robótica (WINTERS e SANTOS-VICTOR, 2002), indústria (SUN, 2004), dentre outras.

O conteúdo das imagens analisadas, nas diferentes áreas de aplicação, é completamente diferente, entretanto, os tipos de análise executados possuem muitas características em comum. Na maioria das vezes, o processo de análise envolve algumas etapas fundamentais como: distinção de partes de interesse na imagem (células, tumores, etc); medição de propriedades dessas partes e utilização dos atributos para classificar ou descrever partes (ROSENFELD, 2001).

Em botânica, sistemas de visão computacional capazes de efetuar análise de imagens podem contribuir substancialmente para o desenvolvimento de pesquisas na área. Hoje em dia, estudos relacionados a análise morfológica, padrão de crescimento e taxonomia de órgãos foliares tem consumido esforços de vários pesquisadores (BOCKHOFF, 2001; HOFER, GOURLAY *et al.*, 2001). Entretanto, as informações que formam a base das pesquisas são coletadas manualmente.

Neste capítulo serão apresentados como os métodos de análise e biometria, podem ser empregados na extração de informações de órgãos foliares, visando sua caracterizando. De maneira geral a extração de características e análise dessas imagens foi dividida em três categorias principais: (i) análise das nervuras, na qual informações quanto ao sistema de venação das folhas são extraídas tais como número de ramificações, ângulo entre as bifurcações, etc. e (ii) análise do esqueleto, neste caso as informações obtidas são extraídas a partir do resultado do processo de esqueletonização da área foliar e (iii) análise da complexidade, no qual foram extraídas propriedades a partir da estimativa da dimensão fractal dos órgãos foliares.

4.1 Análise do Conjunto de Nervuras das Folhas

O conjunto de nervuras dos órgãos foliares, também conhecidos como sistema de venação ou nervação, podem ser descritos como uma rede de células interconectadas que transportam água e nutrientes por toda planta. Isso indica a importância de se estudar e conhecer o comportamento interno dos órgãos vegetais. A Figura 4.1 apresenta uma folha na qual o conjunto de nervuras está bem evidenciado.

Figura 4.1: Exemplo de um órgão foliar no qual o conjunto de nervuras está bem evidenciado.

Grande parte do processo de taxonomia de espécies vegetais é baseada em características extraídas dos órgãos foliares. Segundo (WING, WILF *et al.*, 1999), o primeiro passo para descrever as características do conjunto de nervuras de uma folha é reconhecer suas categorias discretas ou ordens das nervuras. As três principais ordens das nervuras são: nervuras primárias, nervuras secundárias e nervuras terciárias. As nervuras primárias, ou nervuras principais, podem ser comparadas ao tronco das árvores. Geralmente começam na base da folha (pecíolo) e sua espessura vai diminuindo ao longo da área foliar até chegar no topo. As nervuras secundárias começam na base da folha ou a partir de uma bifurcação (ramificação) da nervura

principal. As nervuras terciárias, na maioria das vezes, ocupam grande parte da área foliar e se caracterizam por interconectar as nervuras primárias e secundárias. Dessa forma, a partir do conjunto de nervuras diversas propriedades fundamentais para sua caracterização podem ser extraídas. A seleção dessas características é realizada rigorosamente através de manuais ou protocolos para taxonomia das espécies. A seguir é apresentada uma descrição das principais características utilizadas nos experimentos para identificação das espécies, bem como, a forma como elas são extraídas.

- número de bifurcações: fornece informações quanto ao número de bifurcações ou ramificações do conjunto de nervuras. A Figura 4.2 ilustra como o número de bifurcações, em vermelho, é calculado no conjunto de nervuras das folhas.

Figura 4.2: Extração do número de bifurcações do conjunto de nervuras (adaptado de (WING, WILF *et al.*, 1999)).

- ângulos das nervuras: a medida do ângulo entre as nervuras pode ser utilizada como propriedade para identificação das espécies. Os taxonomistas classificam os ângulos das nervuras em agudos, obtusos e perpendiculares. A Figura 4.3 apresentada como a medida dos ângulos é obtida através da análise dos órgãos foliares. Os ângulos α e β representam orientação e abertura da nervura, respectivamente.

Figura 4.3 : Extração de características a partir dos ângulos entre as bifurcações dos órgãos foliares.

- espaçamento entre as nervuras: outra característica importante para a caracterização de espécies é o espaçamento entre as nervuras. Os taxonomistas dividem o espaçamento entre as nervuras em quatro categorias: (i) uniforme, (ii) irregular, (iii) diminuindo em direção à base e (iv) aumentando em direção à base. A Figura 4.4 apresenta dois exemplos de categorias nas quais os espaçamentos entre as nervuras são visivelmente diferentes (WING, WILF *et al.*, 1999).

Figura 4.4: Espaçamento entre o conjunto de nervuras da folhas.
(a) espaçamento uniforme; (b) espaçamento irregular (WING, WILF *et al.*, 1999).

- número de nervuras basais: usualmente, as nervuras primárias se originam perto da base da folha ou acima do pecíolo. O número de nervuras que iniciam a partir desse local é um parâmetro importante para caracterização das espécies. A Figura 4.5 apresenta uma folha na qual quatro nervuras se originam a partir do pecíolo.

Figura 4.5: Exemplo das nervuras basais dos órgãos foliares (WING, WILF *et al.*, 1999).

Como foi apresentando as nervuras dos órgãos foliares possuem um conjunto imenso de características que podem ser exploradas. Tanto características pontuais como número de bifurcações ou ângulo de uma determinada bifurcação, como características estatísticas como a média ou desvio padrão dos ângulos das bifurcações da folha, etc. Entretanto, esse trabalho é executado de forma manual por taxonomistas ou profissionais de botânica e áreas correlatas. Dessa forma, métodos computacionais podem ser capazes de auxiliar esse processo de extração de características de órgãos foliares e com isso facilitar o trabalho dos taxonomistas.

4.2 Análise do Esqueleto das Folhas

A técnica de esqueletonização é utilizada para obter o eixo médio de um objeto cujo resultado desse processo é conhecido como esqueleto. O formato do esqueleto está intimamente ligado a forma externa do objeto que foi esqueletonizado. No caso dos órgãos foliares, o esqueleto de uma folha está relacionado com o formato da sua área foliar. Conseqüentemente, da mesma maneira que importantes características podem ser extraídas a partir das nervuras, poderosas propriedades podem ser extraídas dos esqueletos das folhas. Essas informações podem complementar as características extraídas dos conjuntos de nervuras.

A Figura 4.6 apresenta o resultado do processo de esqueletonização de duas imagens de folhas com área foliar diferentes. Utilizando os esqueletos gerados é possível extrair informações discriminantes entre os dois órgãos foliares.

Figura 4.6: Resultado do processo de esqueletonização em órgãos foliares.
(a) imagem original; (b) respectivo esqueleto; (c) imagem original; (d) respectivo esqueleto.

Com base no esqueleto de órgãos foliares diversas medidas podem ser obtidas tais como: (i) comprimento do esqueleto; (ii) número de bifurcações; (iii) localização das bifurcações; (iv) ângulos das bifurcações; (v) esqueleto principal, dentre outras. De forma geral, todas as características extraídas do conjunto de nervuras também podem ser extraídas dos esqueletos das folhas.

4.3 Análise de Complexidade

A dimensão fractal é uma importante característica para identificação de órgãos foliares (PLOTZE, FALVO *et al.*, 2004). Esse aspecto é justificado pelo fato da dimensão fractal conter informações quanto a complexidade desses órgãos. Com a dimensão fractal multiescala é possível constatar a variação da dimensão fractal de uma folha em função do espaço métrico (área da imagem) que ela se encontra. Essa análise é muito importante, pois é invariante a transformações geométricas: translação, rotação e escala. A análise multiescala de um objeto (folha) é executada a partir do estudo do gráfico bi-log $\log(d) \times \log(A')$, em que A' é o resultado da derivada numérica do ponto $(d, A(d))$ do gráfico bi-log do método de *Minkowsky*.

Por exemplo, a Figura 4.7 apresenta duas imagens do conjunto de nervuras das folhas no qual o método de análise multiescala da dimensão fractal será aplicado.

Figura 4.7 : Exemplos de imagens artificialmente geradas que ilustram o conjunto de nervuras das folhas; (a) imagem 1; (b) imagem 2.

O primeiro passo da dimensão fractal multiescala é aplicar o método de dilatação exata nos pontos do conjunto de nervuras. Todos os pontos do objeto em análise sofrem dilatações sucessivas a partir de um raio d . A Figura 4.8 ilustra o resultado do processo de dilatações exatas aplicados nas Figura 4.7a e Figura 4.7b.

Figura 4.8: Resultado do método de dilatações exatas em nervuras de órgãos foliares; (a) dilatações sucessivas do objeto da Figura 4.7; (b) dilatações sucessivas do objeto da Figura 4.7b.

Com a matriz resultante do processo de dilatações exatas, o gráfico bi-log pode ser calculado. Os pontos plotados no gráfico representam o diâmetro d *versus* a área de influência, que podem ser representados na forma $\log(d) \times \log A(d)$. A Figura 4.9 apresenta dois gráficos bi-log: (i) na Figura 4.9a corresponde ao gráfico gerado a partir

da matriz de dilatações exatas da Figura 4.7a e, a Figura 4.9b o gráfico referente a Figura 4.7b.

Figura 4.9: Gráficos bi-log do processo de dilatações exatas; (a) gráfico gerado a partir das dilatações da 4.7a; (b) gráfico gerado a partir das dilatações da Figura 4.7b.

Por fim, a análise multiescala da dimensão fractal dos objetos ilustrados na Figura 4.7 é obtida através da derivada numérica do gráfico bi-log das dilatações sucessivas. Através do gráfico $\log(d) \times \log(A')$ é possível obter parâmetros importantes para caracterização de objetos, mais especificamente identificação de órgãos foliares. Alguns parâmetros são: (i) pico de fractalidade ($F_{\max}$); (ii) escala de máxima fractalidade ($D_{\max}$); (iii) largura de alta fractalidade (L) e (iv) pontos singulares (picos ou vales) em determinadas posições da função multiescala. A Figura 4.10 ilustra o gráfico bi-log $\log(d) \times \log(A')$ gerado a partir do gráfico bi-log das dilatações exatas.

Figura 4.10: Gráfico bi-log $\log(d) \times \log(A')$ da análise multiescala da dimensão fractal; (a) gráfico multiescala da Figura 4.7a; (b) gráfico multiescala da Figura 4.7b;

Concluindo, com dimensão fractal multiescala é possível extrair diversas informações capazes de discriminar órgãos foliares. Essas características podem ser utilizadas de maneira isolada ou em conjunto, complementando as outras propriedades biométricas extraídas a partir das nervuras ou do esqueleto.

4.4 Considerações Finais

A extração de informações através de técnicas de biometria e complexidade, são utilizadas nos experimentos para formar o vetor de características da folha a ser identificada. Cada uma das categorias de extração de características tem suas peculiaridades e contribui essencialmente para o processo de identificação das espécies. A diversidade dos órgãos foliares, mais especificamente do conjunto de nervuras, torna o processo de análise e extração de características uma tarefa muito trabalhosa. Por outro lado, a riqueza de informações que podem ser obtidas, através de métodos computacionais semi-automatizados ou até mesmo automatizados, colabora com especialistas e pesquisadores de diversas áreas. A esqueletonização traz propriedades fundamentais para caracterização das espécies, principalmente pelo fato do esqueleto estar intimamente relacionado com a forma da folha. Através da dimensão fractal multiescala é possível coletar informações quanto à complexidade das folhas, que também são importantes para identificação das espécies.

Reconhecimento de Padrões

O processo de análise e extração de características dos órgãos foliares não estaria completo sem a classificação dessas formas em classes. O reconhecimento de padrões, ou classificação de formas, é uma das tarefas mais importantes após a etapa de extração de características (JAIN, DUIN *et al.*, 2000; GONZALEZ e WOODS, 2002) além disso, é um dos principais assuntos relacionados à análise de formas (COSTA e CESAR, 2000). No problema da identificação dos órgãos foliares, os métodos de reconhecimento de padrões são utilizados para tentar descobrir, baseado nas características extraídas, a qual espécie uma determinada folha pertence. Este capítulo apresenta uma introdução ao reconhecimento de padrões e as diferentes abordagens e métodos utilizados nos experimentos com órgãos foliares.

5.1 Fundamentos

A tarefa de reconhecer padrões é uma habilidade muito desenvolvida nos seres humanos, contudo ainda esta longe de ser completamente compreendida (YOUNG, 1994). O ser humano é hábil em reconhecer rostos, vozes, caligrafias e, até mesmo, estados de humor de pessoas desconhecidas. O grande desafio proposto aos pesquisadores da área de reconhecimento de padrões, é desenvolver máquinas que tenham capacidade de simular tais comportamentos.

Enquanto classificar formas pode parecer simples para o ser humano, por outro lado, pode ser impossível quando executado por sistemas computacionais. Na maioria das vezes, o processo de reconhecimento de padrões é uma tarefa complexa que exige diversas técnicas e abordagens, além disso, está restrita ao meio de aplicação. O reconhecimento dos padrões ocorre através de características extraídas dos objetos (imagens ou vídeos), através de sistemas de análise de imagens e visão computacional. Quando esses padrões possuem características em comum, ou seja, as características são similares dentro de um grupo de padrões, denota-se que esses padrões pertencem a uma mesma classe. Com base nas informações disponíveis, os sistemas de reconhecimento podem ser capazes de determinar como os objetos são divididos em classes (SONKA, HLAVAC *et al.*, 1999).

Na literatura podemos encontrar duas abordagens principais para reconhecimento de padrões/classificação de formas (COSTA e CESAR, 2000; JAIN, DUIN *et al.*, 2000):

- (a) classificação supervisionada: o padrão de entrada é identificado como membro de uma classe pré-definida. Um ou mais exemplos de padrões são utilizados como conjunto de treinamento, que formarão a base de conhecimento para classificação dos próximos padrões.
- (b) classificação não-supervisionada: o padrão de entrada é associado a uma classe desconhecida. Nessa abordagem, para um conjunto de padrões de entrada, é necessário descobrir algum tipo de relação, através das suas características, que os classifique.

A Figura 5.1 apresenta uma visão geral das etapas de um sistema de visão computacional até a fase de reconhecimento de padrões.

Figura 5.1: Visão geral das etapas de um sistema computacional evidenciando a fase de reconhecimento de padrões.

Nos dois tipos de abordagens (supervisionada e não-supervisionada), a fase de reconhecimento de padrões necessita de propriedades discriminantes de cada objeto a ser classificado. Essas propriedades são conhecidas como características, do inglês *features*, também chamadas de descritores. O conjunto de características de um objeto é chamado vetor de características (*feature vector*) (GONZALEZ e WOODS, 2002). Assim, um objeto χ pode ser descrito através de seu vetor de características como: $\chi = (x1, x2, x3, \dots, xn)$, no qual x_i representam suas características. O conjunto contendo todos os vetores de características forma o espaço de características dos objetos.

A Figura 5.2 ilustra um exemplo prático de reconhecimento de padrões, baseado nos dados reportados em (FISHER, 1936), disponíveis em (WARD, 2004). O problema em questão foi classificar três tipos de flores da espécie *Íris*: *Íris setosa*, *Íris virginica* e *Íris versicolor*, segundo suas características biométricas. Um conjunto de 150 amostras de flores foi coletado e, para cada uma delas, quatro características foram extraídas: comprimento da sépala, largura da sépala, comprimento da pétala e largura da pétala. Com isso, foi possível constituir dois vetores de características com duas dimensões: $\chi1 = [s1(c,l), s2(c,l), \dots, s150(c,l)]$ contendo as características referentes às sépalas e, $\chi2 = [p1(c,l), p2(c,l), \dots, p150(c,l)]$ as características relativas às pétalas. Ilustrando graficamente os vetores $\chi1$ e $\chi2$, cada flor torna-se um ponto no espaço bidimensional, como apresentado na Figura 5.2. A Figura 5.2a ilustra os dados referentes às características das sépalas e, a Figura 5.2b os dados relativos às pétalas.

Figura 5.2: Exemplo prático de reconhecimento de padrões de espécies de flores. (a) dados relativos às características das sépalas (b) dados referentes às características das pétalas. (baseado em (GONZALEZ e WOODS, 2002))

Analisando os dados ilustrados nos gráficos é simples notar que, dependendo do vetor de características selecionado, o resultado da classificação pode não ser satisfatório. A escolha de características, ou seleção de atributos, é um dos pontos críticos dos sistemas de reconhecimento de padrões (CASTLEMAN, 1979; COSTA e CESAR, 2000; GONZALEZ e WOODS, 2002). Na Figura 5.2a, notamos que as características da sépala não foram capazes de separar (classificar) corretamente os três tipos de flores. Por outro lado, na Figura 5.2b, as características da pétala separam perfeitamente a espécie *Iris setosa* das outras duas, mas razoavelmente as espécies *I. virginica* da *I. versicolor*.

5.2 Métodos para Reconhecimento de Padrões

5.2.1 Classificação Estatística

Na abordagem estatística, cada padrão é representado em termos de n características, que são visualizadas como um ponto no espaço n -dimensional (JAIN, 1989). Dessa forma, cada objeto é formado por um vetor de características, $x = (x_1, x_2, \dots, x_n)$ em que cada $x_i (1 \leq i \leq n)$ é uma característica (DUDA e HART, 1973). O objetivo é escolher entre os vetores de características permitidos, categorias diferentes procurando compactar a ocupação de regiões do espaço de características n -dimensional. A eficiência da representação espacial (conjunto de características) é determinada pelos padrões das diferentes classes que podem ser separadas.

O classificador estatístico é um dispositivo com n entradas e apenas uma saída. Na entrada do dispositivo é associado o vetor de características do objeto a ser classificado, e a saída corresponde a classe que o objeto pertence. Um classificador para N classes $\varpi_1, \varpi_2, \dots, \varpi_n$ gera como saída o identificador de apenas uma dessas classes. Assim, uma classe ϖ_i (a i -ésima classe de um conjunto δ , de n classes) é conjunto que contém padrões com alguma particularidade em comum (SONKA, HLAVAC *et al.*, 1999). A Figura 5.3 ilustra um modelo geral para um classificador estatístico.

Figura 5.3: Arquitetura de um classificador estatístico.

5.2.2 Classificação por Distância Mínima

A classificação por distância de funções é um dos primeiros conceitos em reconhecimento automático de padrões (JAIN, 1989). Esta técnica é uma ferramenta efetiva para a solução de problemas em que cada padrão da classe apresenta limitado grau de variabilidade.

Para cada classe é assumido um centro de massa, conhecido como centróide (COSTA e CESAR, 2000). Assim, um objeto χ pertence a uma determinada classe ϖ , quando a distância entre χ e o centróide da classe ϖ , for menor que todas as distâncias entre ϖ e as outras classes do espaço de características (SONKA, HLAVAC *et al.*, 1999). Para o cálculo da distância, diversas métricas podem ser utilizadas, como por exemplo, à distância euclidiana (COSTA e CESAR, 2000).

Para exemplificar a abordagem de classificação por distância mínima, utilizamos os dados das espécies de flores (*Íris setosa*, *Íris virginica* e *Íris versicolor*). O primeiro passo do processo de classificação por distância mínima é o cálculo dos vetores médios (centróides) que representam cada classe de padrões, que podem ser obtidos através da Equação 5.1.

$$m_j = \frac{1}{N_j} \sum_{x \in \omega_j} x_j \quad j = 1, 2, \dots, M \quad (5.1)$$

Em que N_j é o número de vetores de características da classe ω_j . Dessa forma, um vetor de característica χ é atribuído a uma classe ω quando, a distância entre eles for menor que a distância de χ a todas as outras classes. A Figura 5.4 apresenta os resultados dessa técnica para o caso das espécies de flores. Para simplificar a visualização dos resultados foram utilizadas apenas dez amostras de cada espécie de flor ao invés das cento e cinquenta. Na Figura 5.4a são ilustradas as distâncias entre as trinta amostras de flores (dez de cada espécie) e o vetor médio da classe 1 (*Íris setosa*). Na Figura 5.4b, são apresentadas as distâncias entre as amostras e o vetor médio da classe 2 (*Íris virginica*) e, na Figura 5.4c, as distâncias entre as amostras e o vetor médio da classe 3 (*Íris versicolor*).

Figura 5.4: Resultados da classificação por distância mínima com os dados das espécies de flores. Distância entre o conjunto de amostras e os centróides das espécies
 (a) *I. setosa*; (b) *I. virginica* e
 (c) *I. versicolor*.

Os resultados da Figura 5.4a apresentam uma excelente separação (classificação) entre os elementos da classe 1 (*I. setosa*) e as demais classes. Contudo, as Figura 5.4b e 5.4c demonstram uma sobreposição parcial entre os elementos da classe 2 (*I. virginica*) e classe 3 (*I. versicolor*).

5.2.3 Classificação por Redes Neurais

As redes neurais artificiais (RNA) consistem em um modelo de processamento de informações que possuem características em comum com os sistemas neurais biológicos. Essas redes possuem uma estrutura massivamente paralela. As principais características de uma rede neural são: (i) habilidade de apreender relações complexas não-lineares de entrada/saída, (ii) treinamento sequencial de procedimentos e (iii) adaptação da rede de acordo com os dados de entrada (JAIN, DUIN *et al.*, 2000).

As famílias de redes neurais mais frequentemente utilizadas para classificação de características são (JAIN, DUIN *et al.*, 2000): as redes *feed-forward*, que incluem *perceptron* multicamadas (MLP) e *radial-basis function* (RBF). Essas redes são organizadas em camadas e tem conexões unidirecionais entre elas. Outra família popular de redes são as *Self-Organizing Map* (SOM), ou redes de *Kohonen*, que são principalmente utilizadas para agrupamento de dados (*clustering*) e mapeamento de características. O processo de aprendizado envolve atualização da arquitetura da rede e dos pesos de conexão, assim a rede pode ser eficientemente executada para tarefa de classificação/*clustering*.

Um dos primeiros trabalhos sobre redes neurais artificiais foi apresentado em (MCCULLOCH e PITTS, 1943), e ficou conhecido como neurônio de McCulloch-Pits. O modelo de processamento lógico proposto por eles, cujo objetivo era tentar simular os procedimentos executados pelos neurônios do cérebro humano, foi inicialmente utilizado para construir expressões lógicas finitas. A Figura 5.5 apresenta um neurônio biológico comparado ao modelo artificial proposto McCulloch-Pits.

Figura 5.5: Comparação entre um neurônio biológico e o modelo artificial proposto por McCulloch-Pits

O conjunto de entradas $(x_1, x_2, \dots, x_n)$ do neurônio processador tem a função de receber as informações vindas de outros neurônios (ou inicialmente do vetor de

características). O processamento das informações consiste basicamente de uma combinação linear das entradas, conforme apresentado na Equação 5.2 (RAUBER, 1998).

$$net = w_1x_1 + w_2x_2 + \dots + w_nx_n \Rightarrow \sum_{i=1}^n w_ix_i \quad (5.2)$$

Todas as entradas estão associadas a pesos w_i que refletem a importância da entrada x_n . O resultado da combinação linear é o valor net . A saída do neurônio está ligada a um determinado limiar μ : se o valor de net ultrapassar μ então o neurônio “dispara” o valor 1 na saída binária γ , caso contrário, se não ultrapassar o limiar o valor da saída $\gamma = 0$. A comparação entre o valor net com o limiar μ é realizada através das funções de ativação (HAYKIN, 2001). A Figura 5.6 apresenta diferentes tipos de funções de ativação. No caso da função linear a saída produzida pelo neurônio é linear contínua. Na função de escada, a saída é binária e na função sigmoideal a saída produzida é não-linear contínua.

Figura 5.6: Funções de ativação das redes neurais artificiais.
(a) função linear; (b) função de escada e (c) função sigmoideal

As pesquisas sobre redes neurais artificiais, também conhecida como neurocomputação, culminaram com a criação do modelo *perceptron* (ROSENBLATT, 1958). O *perceptron* é capaz de classificar classes que são linearmente separáveis e foi utilizado inicialmente para reconhecer exemplos de caracteres. Porém, devido a diversos problemas enfrentados pelos pesquisadores (mais detalhes podem ser obtidos em (MINSKY e PAPERT, 1969)), as pesquisas foram interrompidas até a década de 80 (RICH e KNIGHT, 1983).

Nesse período um novo modelo de redes neurais artificiais foi proposto o *perceptron* multicamadas ou *Multi Layer Perceptron – MLP* (RUMELHART, HINTON *et al.*, 1986), que possui pelo menos três camadas: uma de entrada, uma ou mais escondidas (intermediárias) e uma de saída. Esse tipo de rede é utilizada geralmente em problemas não linearmente separáveis, sendo uma evolução para as redes *perceptron* que resolviam apenas problemas linearmente separáveis (HAYKIN, 2001). A Figura 5.7 ilustra um *perceptron multicamada* com uma camada escondida, quatro variáveis de entrada e três variáveis de saída.

Figura 5.7: Perceptron multicamada com uma camada escondida, quatro variáveis de entrada e três variáveis de saída.

Diversos algoritmos podem ser encontrados na literatura para treinar redes neurais do tipo MLP, contudo o mais conhecido é o *back-propagation* (RUMELHART, HINTON *et al.*, 1986). Esse algoritmo utiliza informações das entradas e saídas desejadas para ajustar os pesos da rede, através de um mecanismo de correção de erros. No *back-propagation* o treinamento ocorre em duas fases chamadas de *backward* e *forward*. A fase *forward* é utilizada para definir a saída da rede a partir de um determinado padrão de entrada. E a fase *backward* utiliza a saída desejada e a saída fornecida pela rede para atualizar os pesos de suas conexões (BRAGA, CARVALHO *et al.*, 2000).

O desempenho e a eficiência dos resultados de um algoritmo de redes neurais artificiais pode ser medido através das taxas de erro. Esse cálculo é baseado na diferença entre o resultado alcançado pelo algoritmo de RNA e o resultado esperado. Duas métricas são comumente utilizadas para esse propósito: *Sum of Square Errors – SSE* e *Mean Square Errors – MSE* (HAYKIN, 2001).

Para ilustrar o funcionamento das redes neurais artificiais aplicadas ao reconhecimento de padrões, utilizaremos a base de dados de características de flores das espécies: *Íris setosa*, *Íris virginica* e *Íris versicolor*. Para treinamento da rede foi utilizado o algoritmo *back-propagation* e, a topologia adotada foi 4x3x2 (quatro neurônios na camada de entrada, três neurônios na camada oculta e dois neurônios na camada de saída). Na camada de saída são utilizados apenas dois neurônios porque o resultado da RNA é binário, assim dois dígitos são capazes de representar as três classes de flores: 00 = classe 1; 01 = classe 2 e 10 = classe 3.

Para o treinamento e teste da RNA com os dados das espécies de flores, as características foram distribuídas de três maneiras diferentes: (a) Experimento 1: 50% do conjunto foi utilizado para treinamento e 50% para testes; (b) Experimento 2: 60% para treinamento e 40% para testes e (c) Experimento 3: 80% para treinamento e 20% para testes. A execução dos treinamentos e testes foi realizada utilizando o *Neural Networks Toolbox* do programa Matlab©. Os resultados dos experimentos da rede neural foram visualizados através de matrizes de confusão. Nesse tipo de matriz, o número de predições corretas para cada classe é representado ao longo da diagonal principal. E, nas demais posições da matriz são representados os casos classificados erroneamente (predições incorretas). Na Tabela 5.1 são apresentados os resultados do Experimento 1 através da matriz de confusão. Durante os testes, a taxa de erro (MSE) encontrada foi de 0,03894. Em média, a rede classificou corretamente 92% das espécies.

TABELA 5.1: Matriz de confusão ilustrando os resultados da rede neural utilizada para classificação das espécies de flores *Íris* (50% treinamento / 50% testes)

	<i>I. setosa</i>	<i>I. virginica</i>	<i>I. versicolor</i>	% acerto	% erro
<i>I. setosa</i>	25	0	0	100%	0%
<i>I. virginica</i>	0	25	0	100%	0%
<i>I. versicolor</i>	0	6	19	76%	24%
			% geral	92%	8%

No Experimento 2 a rede neural foi capaz de classificar corretamente 95% dos casos. A taxa de erro obtida durante os testes foi de 0,01603. Os resultados desse experimento são ilustrados na Tabela 5.2.

TABELA 5.2: Matriz de confusão ilustrando os resultados da rede neural utilizada para classificação das espécies de flores Íris (60% treinamento / 40% testes)

	<i>I. setosa</i>	<i>I. virginica</i>	<i>I. versicolor</i>	% acerto	% erro
<i>I. setosa</i>	20	0	0	100%	0%
<i>I. virginica</i>	0	19	1	95%	5%
<i>I. versicolor</i>	0	2	18	90%	10%
			% geral	95%	5%

No Experimento 3 a rede neural utilizou 80% do conjunto para treinamento e os 20% restantes para testes. Nesse caso a porcentagem de acerto da rede para as espécies de flores chegou a 100%. A taxa de erro (MSE) obtida foi de 0,00001.

TABELA 5.3: Matriz de confusão ilustrando os resultados da rede neural utilizada para classificação das espécies de flores Íris (80% treinamento / 20% testes)

	<i>I. setosa</i>	<i>I. virginica</i>	<i>I. versicolor</i>	% acerto	% erro
<i>I. setosa</i>	10	0	0	100%	0%
<i>I. virginica</i>	0	10	0	100%	0%
<i>I. versicolor</i>	0	0	10	100%	0%
			% geral	100%	0%

Em todos os experimentos o acerto no processo de classificação das redes neurais foi superior a 90%. Esses resultados podem ser considerados muito satisfatórios. É importante salientar que outras topologias e algoritmos de treinamento também podem ser empregados com esse mesmo conjunto de dados.

5.3 Análise dos Componentes Principais

Análise de componentes principais (*Principal Component Analysis – PCA*) é uma técnica muito utilizada em reconhecimento de padrões para redução de dimensionalidade (JAIN, DUIN *et al.*, 2000). O conceito de dimensionalidade é

utilizado para especificar o número de características de um determinado padrão, ou seja, a dimensão do espaço de características.

De modo geral, a técnica de PCA consiste em remover redundâncias entre as características extraídas dos objetos. Para isso, o método calcula os autovalores e autovetores da matriz de covariância do conjunto de vetores do espaço de características. As informações mais importantes, ou mais discriminantes, do conjunto de dados estão em cada um dos autovetores que são chamados de *principal components*.

A aplicação da técnica de PCA pode ser ilustrada utilizando o conjunto de dados das flores. Cada amostra do conjunto de 150 flores contém quatro características, ou seja, o espaço de características desses dados possui quatro dimensões. A visualização de informações com mais de três dimensões não é uma tarefa trivial sendo necessário a utilização de técnicas de visualização mais complexas (OLIVEIRA e LEVKOWITZ, 2003). Dessa forma, é possível utilizar a técnica de PCA tentar reduzir a dimensionalidade do conjunto, que inicialmente possuía quatro dimensões, para três ou, até mesmo, duas dimensões. A Figura 5.8, apresenta os resultados da utilização da técnica de PCA no conjunto de dados das flores. Na Figura 5.8a o conjunto foi reduzido para três dimensões, e na Figura 5.8b, o conjunto foi reduzido para duas dimensões.

Figura 5.8: Redução de dimensionalidade utilizando a técnica de PCA do conjunto de dados de flores. (a) redução dos dados para três dimensões; (b) redução dos dados para duas dimensões.

Descrições mais abrangentes do procedimento de Análise de Componentes Principais podem ser encontradas em: (JOHNSON e WICHERN, 1988; OLIVEIRA, 1997; HAYKIN, 2001).

5.4 Considerações Finais

Neste capítulo foram apresentados diversos conceitos fundamentais sobre reconhecimentos de padrões, bem como, alguns exemplos de técnicas utilizadas nos experimentos com órgãos foliares. De forma geral, o principal objeto do processo de reconhecimento de padrões é conseguir identificar algum tipo de similaridade a partir de um conjunto de dados. No caso dos órgãos foliares, as técnicas buscam classificar as espécies segundo características extraídas a partir da análise de complexidade e biometria. Além das técnicas de classificação apresentadas como: classificação estatística, classificação por distância mínima e classificação por redes neurais, métodos para redução de dimensionalidade também podem ser utilizados. Os resultados das técnicas de reconhecimento de padrões aplicadas nos conjuntos de dados dos órgãos foliares, serão apresentados no capítulo de experimentos.

Resultados Experimentais

Neste capítulo são apresentados e discutidos os experimentos realizados para identificação das espécies vegetais. Os experimentos foram divididos em três grupos principais: (i) complexidade dos órgãos foliares – no qual são apresentados os resultados da identificação de espécies, através de características extraídas da dimensão fractal multiescala; (ii) biometria dos órgãos foliares – em que são apresentados os resultados dos métodos desenvolvidos para extração de medidas biométricas em órgãos foliares, bem como a potencialidade desses métodos na identificação das espécies e (iii) comparação dos métodos – nesse grupo as características extraídas tanto de complexidade quanto de biometria são correlacionadas, buscando a discriminação das espécies.

6.1 Espécies de Órgãos Foliares

A proposta do projeto foi estudar e implementar métodos de análise de imagens e extração de características capazes de identificar órgãos foliares através de propriedades de sua estrutura interna. Foram consideradas apenas informações da parte interna das folhas como nervuras e esqueletos. Para validação dos experimentos e verificação do potencial das características extraídas foram utilizados dois conjuntos de imagens de órgãos foliares:

- (i) **imagens de folhas de maracujás nativos:** esse conjunto contém quarenta amostras de 10 espécies diferentes de maracujá nativo, do gênero *Passiflora*

(*Passifloraceae*). As amostras foram obtidas através de uma parceria entre o ICMC-USP e a Escola Superior de Agronomia Luiz de Queiroz (ESALQ-USP). Quatro amostras de cada espécie de maracujá foram utilizadas nos experimentos, totalizando 40 imagens de órgãos foliares. A Figura 6.1 apresenta uma amostra de cada espécie de *Passiflora* utilizada nos experimentos.

Figura 6.1: Espécies de maracujás silvestres do gênero *Passiflora* utilizadas nos experimentos.
 (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*;
 (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*.

(ii) imagens de folhas da Mata Atlântica e do Cerrado: esse conjunto de órgãos foliares foi coletado no Instituto Florestal de Mogi-Mirim (SP) e contém 10 espécies de arbóreas da Mata Atlântica e do Cerrado brasileiro. Esse conjunto possui várias espécies de imagens de órgãos foliares como jequitibá, pitanga, dentre outras. Nos experimentos, cinco amostras de cada espécie foram utilizadas, totalizando 50 imagens de órgãos foliares. A Figura 6.2 apresenta as espécies da Mata Atlântica e do Cerrado utilizadas nos experimentos de identificação.

Figura 6.2: Espécies arbóreas da Mata Atlântica e do Cerrado utilizadas nos experimentos.

A biodiversidade das espécies vegetais presentes na natureza se torna um desafio para sistemas computacionais para identificação das espécies. Um exemplo que ilustra essa dificuldade é o caso das espécies de maracujás silvestres *Passiflora*. O gênero está distribuído por todo continente americano e apresenta cerca de 450 espécies, das quais quase duas centenas podem ser encontradas no Brasil (VANDERPLANK, 2000). A morfologia foliar das espécies possui uma variabilidade muito grande. Além disso, algumas espécies são amplamente semelhantes tanto do ponto de vista morfológico quanto molecular, tornando assim a discriminação visual uma tarefa praticamente impossível (PÁDUA, VIEIRA *et al.*, 2002).

6.2 Sistema Computacional

Para realizar os experimentos de identificação de espécies arbóreas foi necessário o desenvolvimento de um sistema computacional. Esse sistema tem como características principais a capacidade de processar e analisar amostras de órgãos foliares. O ambiente escolhido para desenvolvimento do sistema foi o *Borland Delphi* na plataforma *Microsoft Windows*, porém ele pode ser facilmente convertido para outras plataformas como *Linux*. O sistema foi implementado de maneira modular e seguindo o paradigma de orientação a objetos. Esse fato é de vital importância para facilitar a integração com o Sistema TreeVis. De maneira geral, a arquitetura do sistema foi dividida em 4 módulos, como ilustrado no diagrama da Figura 6.3:

- **aquisição:** neste módulo as imagens de órgãos foliares são digitalizadas e armazenadas para posterior processamento. Todas as imagens foram adquiridas através de um scanner comercial com resolução de 300 dpi. As imagens digitalizadas foram normalizadas em função do tamanho da nervura principal das folhas, dessa forma as amostras de todas as espécies parecem possuir aproximadamente o mesmo tamanho. O processo de normalização das imagens foi adotado para que os métodos de extração de características desenvolvidos fossem invariantes as transformações geométricas como escala e rotação.
- **processamento:** os conjuntos de imagens digitalizadas passam por um pré-processamento cujo objetivo é adequar as imagens para posterior análise. As imagens são filtradas e, os ruídos originários do processo de aquisição são removidos. Nesse módulo os objetos de interesse da estrutura interna dos órgãos foliares são segmentados: as nervuras e o esqueleto. Todas as outras informações da área foliar são descartadas. A implementação de um método automatizado para segmentar as nervuras é uma tarefa complexa, principalmente pela diversidade de formas e cores dos órgãos foliares da natureza. Para solução do problema foi adotada uma abordagem semi-automatizada, descrita posteriormente. As implementações dos algoritmos de esqueletonização para extração dos esqueletos das folhas foram, em geral, simples e rápidas. O método mais trabalhoso foi o algoritmo multiescala, pois envolve outras técnicas de processamento de imagens como extração de contornos e dilatações exatas. Foram implementados os seguintes algoritmos de esqueletonização: *Hilditch*, *Stentiford*, *Zhang-Suen*, *Holt* e multiescala. Uma descrição detalhada desses métodos foi apresentada no capítulo 3 (análise de imagens e biometria).
- **análise de imagens e extração de características:** esse módulo contém métodos computacionais para analisar e extrair características das imagens de órgãos foliares. De maneira geral, os principais métodos implementados foram: (i) dimensão fractal - *Box-counting*, *Minkowsky* e multiescala e (ii) biometria – diversas técnicas para extração de medidas biométricas foram desenvolvidas, como contagem do número de

bifurcações, cálculo dos ângulos entre as bifurcações, etc. Os métodos implementados geram informações (características) que formam o vetor de característica de cada órgão foliar, esse vetor é utilizado posteriormente para identificação das espécies.

- **reconhecimento de padrões:** o objetivo do módulo de reconhecimento de padrões é tentar discriminar as espécies arbóreas a partir das informações extraídas pelo módulo de análise e extração de características. Como o objetivo principal do projeto foi analisar e extrair características, as técnicas de reconhecimento de padrões utilizadas não foram implementadas. Para verificar o potencial das características extraídas na identificação das espécies, foram utilizadas as ferramentas do software científico Matlab®. Para executar tal tarefa, as características extraídas no módulo III foram exportadas do sistema de análise em arquivos texto (ASCII) e, importadas no Matlab®. Além de diversas técnicas de reconhecimento de padrões como *clustering*, redes neurais e PCA, o Matlab® possui um excelente ferramental para visualização de dados científicos.

Figura 6.3: Diagrama da arquitetura do sistema computacional desenvolvido para realização dos experimentos de identificação das espécies.

Uma das principais dificuldades enfrentadas no desenvolvimento do projeto foi durante a fase de processamento das imagens de órgãos foliares. Nessa fase os objetos de interesse (nervura e esqueleto) são segmentados. O processo de segmentar as

nervuras dos órgãos foliares não é uma tarefa trivial, no qual o desenvolvimento de algoritmos capazes de realizar tal procedimento é altamente complexo. Como este não era o objetivo principal do projeto, diversas soluções foram adotadas na tentativa de solucionar o problema de segmentação das nervuras. Inicialmente técnicas tradicionais de processamento de imagens foram empregadas como: limiarização, detecção de bordas, filtros morfológicos, etc., porém os resultados não foram satisfatórios. Uma outra abordagem foi utilizada, na qual o processo de segmentação deveria ser realizado antes do processo de digitalização (aquisição) das imagens. Para isso, diversos produtos químicos foram pesquisados com objetivo de segmentar as nervuras. Algumas soluções foram testadas experimentalmente como: álcool e água sanitária. Nesses casos, os resultados foram interessantes, contudo, devido a corrosividade das soluções químicas as estruturas das folhas eram prejudicadas. Alguns profissionais da área de botânica foram consultados buscando alguma forma de segmentar as nervuras. Em contato com professores do Departamento de Botânica de Universidade Federal de São Carlos, eles sugeriram que fosse utilizado um processo conhecido em morfologia vegetal como diafanização. A diafanização ou clareamento é um processo químico que faz a despigmentação dos órgãos foliares (CARDOSO, 2000). Seus resultados são muito satisfatórios, contudo essa técnica demanda altos custos, além disso, os experimentos são bastante demorados. Por fim, a solução encontrada foi utilizar uma abordagem semi-automatizada. Foram utilizados métodos computacionais convencionais de processamento de imagens para segmentar as nervuras. Os locais que esses métodos não geravam resultados satisfatórios foram segmentados (corrigidos) manualmente.

Com o sistema computacional implementado e os principais problemas resolvidos, os experimentos com casos reais de identificação de espécies foram iniciados. Os experimentos foram divididos em duas categorias: (i) complexidade – na qual são extraídas características a partir da dimensão fractal multiescala e (ii) biometria: em que as características extraídas são baseadas em medidas biométricas. Nas duas categorias, a qualidade das características extraídas para identificação de espécies, é determinada através dos resultados das técnicas de reconhecimento de padrões.

Esses experimentos foram realizados com os dois conjuntos de imagens apresentados na seção 6.1: imagens de maracujás silvestres e imagens da Mata Atlântica e do Cerrado brasileiro. Em cada uma das categorias (complexidade e biometria) os

testes experimentais foram empregados tanto em nervuras, quanto nos esqueletos dos órgãos foliares. Por fim, as características extraídas pelas duas categorias são combinadas visando uma comparação dos métodos e verificação generalizada das informações extraídas. A Figura 6.4 apresenta um digrama que ilustra como os experimentos com órgãos foliares foram executados. Nas próximas seções os resultados desses experimentos são relatados e discutidos.

Figura 6.4: Digrama dos experimentos realizados para identificação das espécies

6.3 Experimento I: Complexidade de Órgãos Foliares

O experimento realizado com base em medidas de complexidade das folhas, tem como objetivo principal a caracterização das espécies através de informações extraídas da dimensão fractal. Dentre os métodos para estimativa da dimensão fractal estudados, a técnica escolhida foi a dimensão fractal multiescala. A escolha do método foi baseada em um estudo comparativo entre as técnicas e sua potencialidade na identificação de espécies arbóreas (PLOTZE, FALVO *et al.*, 2003).

De modo geral, o experimento de complexidade utilizando informações da dimensão fractal multiescala, pode ser dividido em três fases principais: (i) processamento; (ii) extração de características e (iii) reconhecimento de padrões. A Figura 6.5 apresenta um diagrama com as fases do experimento de complexidade dos órgãos foliares.

Figura 6.5: Fases do experimento sobre complexidade dos órgãos foliares

Na primeira fase, o banco de dados de imagens (90 imagens: 40 imagens de *Passiflora* e 50 da Mata Atlântica e do Cerrado) é processado com objetivo de adequar as imagens para fase seguinte de extração de características. O resultado dessa fase são dois conjuntos de imagens com os objetos de interesse da parte interna dos órgãos foliares segmentados:

- **nervuras:** algumas deficiências provocadas pelo método de segmentação fazem com que as nervuras apresentem diâmetros diferentes em determinadas posições da folha. Para solucionar este problema, de maneira que essa diferença não influencie no processo de extração de características, as nervuras foram reduzidas para 1 (um) *pixel* de largura utilizando um algoritmo de afinamento.
- **esqueletos:** os esqueletos dos órgãos foliares foram obtidos através da eskeletonização multiescala. Outros métodos também foram testados, contudo, os resultados mais satisfatórios foram os da eskeletonização multiescala. O valor do limiar (T) para determinação do esqueleto no método multiescala foi estabelecido empiricamente. Assim, os esqueletos utilizados nos experimentos estão entre um intervalo de $T = 5$ a $T = 25$.

A segunda fase consiste em extrair características a partir das imagens resultantes da fase de processamento. Essas características são coletadas através de informações da análise multiescala da dimensão fractal, tanto das nervuras quanto do esqueleto. Todas as imagens do banco de dados de espécies são analisadas através do método multiescala da dimensão fractal. Com base na curva resultante ($\log(d) \times \log(A')$) seis características são extraídas em três pontos de interesse (P1, P2 e P3) diferentes. Esses pontos foram escolhidos devido ao comportamento padronizado

das curvas da dimensão fractal multiescala. A Figura 6.6 ilustra a localização dos três pontos de interesse no gráfico da dimensão fractal multiescala. O primeiro ponto (P1) é coletado em um vale mais acentuado que aparece no início da curva fractal multiescala. O ponto P2 corresponde ao ponto máximo da curva ou, mais especificamente ao pico de fractalidade ($F_{\max}$). E o ponto P3 é relativo ao segundo pico de máxima fractalidade encontrado na curva multiescala. As coordenadas x e y de cada ponto de interesse são utilizadas como características constituindo as seis propriedades extraídas da curva: $P1_x$, $P1_y$, $P2_x$, $P2_y$, $P3_x$ e $P3_y$. A principal vantagem em utilizar pontos de interesse ao invés de toda a curva está relacionada à redução de dimensionalidade. Como a curva da dimensão fractal multiescala é constituída por um vetor de milhares de elementos, fica inviável adotá-la como vetor de características, pois nesse caso o espaço de características apresentaria uma dimensão da ordem de milhar. Por outro lado utilizando os pontos de interesse, o número de informações (dimensões) necessárias para tentar caracterizar o objeto é bem inferior, com isso, métodos convencionais de reconhecimento de padrões podem ser empregados para identificar os objetos.

Figura 6.6: Pontos de interesse extraídos da curva fractal multiescala. As coordenadas x e y de cada ponto são consideradas como características.

A fase de reconhecimento de padrões tem como objetivo realizar a identificação das espécies com base nas características extraídas na segunda fase. As principais técnicas de reconhecimento de padrões empregadas nessa fase foram: classificação por distância mínima, análise de clusters, redes neurais e análise de componentes principais. As técnicas de reconhecimento de padrões foram fundamentais para avaliação dos métodos de extração de características. Com elas foi possível verificar o potencial das características extraídas, a partir da dimensão fractal multiescala, na identificação das espécies foliares. As próximas seções apresentam os resultados dos experimentos de complexidade dos órgãos foliares, tanto para nervuras quanto para os esqueletos.

6.3.1 Complexidade das Nervuras

A. – RESULTADOS DAS ESPÉCIES DE *PASSIFLORA*

As quarenta amostras de órgãos foliares das espécies de *Passiflora* foram segmentadas e suas respectivas nervuras foram extraídas. As amostras estão divididas em dez classes, contendo quatro amostras por classe. A partir do conjunto de nervuras dessas amostras a dimensão fractal multiescala foi determinada. A Figura 6.7 apresenta as curvas da dimensão fractal multiescala das nervuras das espécies de *Passiflora*. Os gráficos estão organizados por espécies, sendo assim cada um deles apresenta quatro curvas correspondentes a uma determinada espécie de *Passiflora*.

Figura 6.7: Gráficos da dimensão fractal multiescala das espécies de *Passiflora*. (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*.

Os resultados da análise multiescala da dimensão fractal, ilustrados na Figura 6.7, demonstram visualmente que as curvas multiescala são capazes de caracterizar órgãos foliares, produzindo assinaturas digitais das espécies. Todos os pontos das curvas podem ser utilizados como vetores de características ou, apenas alguns desses pontos podem ser considerados. Nesse experimento foram definidos três pontos de interesse (P1, P2 e P3), para cada qual suas coordenadas x e y formaram o vetor de características com seis dimensões. Na Tabela A.1 do Anexo, todas as características coletadas das espécies de *Passiflora* podem ser visualizadas.

Baseado nas informações extraídas dos órgãos foliares a primeira análise realizada nos dados foi referente à distância entre os agrupamentos (clusters) de cada espécie. Para isso, foram calculados os centros de massa (centróides) de cada espécie e determinada a distância euclidiana entre eles. Os seguintes critérios foram adotados para avaliar os resultados dos clusters: (i) clusters separados – quando a distância entre eles é superior a 0,4; (ii) clusters parcialmente sobrepostos – quando a distância entre eles é menor que 0,4 e maior que 0,2 e (iii) clusters sobrepostos – quando a distância entre os centróides é menor que 0,2.

A Figura 6.8 ilustra os resultados experimentais da análise baseada na distância entre os clusters. Cada gráfico apresenta o cluster de uma espécie (em azul), comparado individualmente com as outras nove espécies. Os valores das barras verticais representam a distância entre os centróides dos clusters. Nos gráficos, as barras em verde representam clusters separados, enquanto as barras em amarelo indicam clusters parcialmente sobrepostos e, as barras vermelhas clusters sobrepostos. Na grande maioria dos gráficos as barras verdes (clusters separados) são predominantes e, são poucas as sobreposições entre os clusters. Conseqüentemente isto indica que as informações extraídas da dimensão fractal multiescala são satisfatoriamente capazes de discriminar espécies arbóreas.

Para realizar uma análise mais eficiente e comprovar os resultados obtidos, foram calculadas as distâncias entre cada centróide e as quarenta amostras de *Passiflora* utilizadas no experimento. Os resultados são ilustrados nos gráficos da Figura 6.9. O eixo x corresponde ao conjunto de amostras utilizadas e o eixo y a distância entre elas e o centróide de uma determinada espécie. As amostras estão agrupadas por espécie, assim as barras de 1-4 correspondem às amostras da espécie 1; de 5-8 da espécie 2 e assim por diante. No gráfico da Figura 6.9a, por exemplo, são apresentadas as distâncias entre as amostras e o centróide da espécie *Passiflora coriacea*. Em cinza claro estão destacadas as amostras da espécie analisada. Conseqüentemente para uma correta e eficiente classificação, a distância entre as amostras em destaque (cinza claro) precisam ser as quatro menores distâncias entre todas as quarenta amostras analisadas. Na maioria dos casos a discriminação entre as espécies foi satisfatória, ou seja, não houve

sobreposição entre os clusters das espécies. Contudo em alguns casos é possível notar essa sobreposição como nos gráficos da Figura 6.9h, Figura 6.9i e Figura 6.9j.

Com objetivo de tentar melhor, ainda mais, os resultados obtidos com análise de clusters, a técnica de análise de componentes principais – PCA foi aplicada no conjunto de dados. O método permite reduzir o número de variáveis do vetor de características analisado, com base na combinação linear de todos os vetores de características e suas correlações. É importante salientar que, na maioria das vezes, a redução de dimensionalidade pode comprometer o conjunto de dados e resultar em perda de informações. Utilizando o método PCA, no espaço de características extraído a partir da dimensão fractal multiescala, foi possível verificar que não ocorreram melhoras significativas nos resultados. O conjunto de dados original com 6D foi reduzido para dimensões menores como 5D, 4D e 3D. A Figura 6.10 ilustra o resultado da redução de dimensionalidade de 6D para 4D. Para visualização dos resultados foi utilizada a técnica conhecida como coordenadas paralelas (INSELBERT e DIMSDALE, 1990; OLIVEIRA

e LEVKOWITZ, 2003). Nessa técnica, as dimensões correspondem a uma série de eixos paralelos nos quais os vetores de características são representados. Cada eixo é relativo a um dos componentes do vetor de características. A visualização é obtida ligando os eixos uns aos outros através de retas. Na Figura 6.10a os eixos em cinza escuro corresponde à espécie *Passiflora coriacea* e, os eixos em cinza claro as demais espécies.

Figura 6.10: Redução de dimensionalidade das características extraídas das espécies de 6D para 4D. Visualização dos resultados através da técnica de coordenadas paralelas. Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*.

As informações extraídas das nervuras dos órgãos foliares foram submetidas a experimentos envolvendo redes neurais artificiais. Para treinamento da rede foi utilizado o algoritmo *back-propagation*, no qual a topologia adotada foi 6x5x4 (seis neurônios de

entrada, cinco ocultos e quatro neurônios na camada de saída). Os quarenta vetores de características das dez espécies de *Passiflora* foram distribuídos de duas maneiras diferentes: (a) Experimento 1: 75% do conjunto foi utilizado para treinamento e 25% para testes e (b) Experimento 2: 50% para treinamento e 50% para testes.

A Tabela 6.1 apresenta os resultados do Experimento 1. A classe C1 representa a espécie *Passiflora coriacea*, C2 – *P.foetida*, C3 – *P.miersii*, C4 – *P.organensis*, C5 – *P.pohlii*, C6 – *P. suberosa*, C7 – *P. amethystina*, C8 – *P.caerulea*, C9 – *P.giberti* e C10 – *P.alata*. A rede classificou corretamente 90% (erro MSE = 0,07232).

TABELA 6.1: Resultado da classificação por redes neurais dos dados de complexidade das nervuras - espécies de *Passiflora* (75% treinamento/25% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	1	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	1	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	1	0	0	0	0	0	0	0	100%	0%
C4	0	0	0	1	0	0	0	0	0	0	100%	0%
C5	0	0	0	0	1	0	0	0	0	0	100%	0%
C6	0	0	0	0	0	1	0	0	0	0	100%	0%
C7	0	0	0	0	0	0	1	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	1	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	0	1	0%	100%
C10	0	0	0	0	0	0	0	0	0	1	100%	0%
										%	90%	10%

Na Tabela 6.2 são apresentados os resultados do Experimento 2 (50% do conjunto para treinamento e 50% para testes). Nesse caso a rede também classificou corretamente 90% das espécies com uma taxa de erro (MSE) de 0,06231.

TABELA 6.2: Resultado da classificação por redes neurais dos dados de complexidade das nervuras – espécies de *Passiflora* (50% treinamento/50% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	2	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	2	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	2	0	0	0	0	0	0	0	100%	0%
C4	0	0	0	2	0	0	0	0	0	0	100%	0%
C5	1	0	0	0	1	0	0	0	0	0	50%	50%
C6	0	0	0	0	0	2	0	0	0	0	100%	0%
C7	0	0	0	0	0	0	1	0	1	0	50%	50%
C8	0	0	0	0	0	0	0	2	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	2	0	100%	0%
C10	0	0	0	0	0	0	0	0	0	2	100%	0%
										%	90%	10%

Analisando os resultados obtidos com a rede neural, podemos concluir que o desempenho foi satisfatório, com média de classificação superior a 90%. É importante ressaltar que a quantidade de informações (vetores de características) utilizadas para treinamento e testes da RNA não é o ideal, conseqüentemente um número maior de amostras pode trazer resultados ainda mais significativos.

B. – RESULTADOS DAS ESPÉCIES DA MATA ATLÂNTICA E DO CERRADO

As espécies da Mata Atlântica e do Cerrado brasileiro estão divididas em dez classes contendo cinco amostras por classe (espécie). Da mesma maneira que ocorreu com as espécies de *Passiflora* essas espécies foram submetidas a análise multiescala através da dimensão fractal. A Figura 6.11 apresenta os gráficos contendo as curvas da dimensão fractal multiescala das espécies da Mata Atlântica e do Cerrado.

Figura 6.11: Gráficos da dimensão fractal multiescala das espécies arbóreas da Mata Atlântica e do Cerrado. (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9 e (j) Espécie 10.

A partir dos gráficos da dimensão fractal multiescala foram extraídos três pontos de interesse: P1, P2 e P3. Os dados coletados podem ser visualizados na Tabela A.2 do Anexo. Com essas informações, foi realizada uma análise relativa à distância entre os agrupamentos de cada espécie. Como já discutido anteriormente, esse tipo de classificação é baseado na distância entre os centros de massa (centróides) de uma determinada espécie e as demais espécies do conjunto. Os mesmos critérios adotados para as espécies de *Passiflora* foram utilizados com as espécies da Mata Atlântica e do Cerrado: (i) clusters separados – distância superior a 0,4; (ii) clusters parcialmente sobrepostos – distância de 0,2 a 0,4 e (iii) clusters sobrepostos – distância inferior a 0,2.

Os resultados da classificação por distância mínima são apresentados na Figura 6.12. Em cada um desses gráficos, o cluster de uma determinada espécie (em azul), é comparado individualmente com as outras nove espécies. As barras representam as distâncias entre esses clusters: verde – para clusters separados; amarelo – para clusters parcialmente sobrepostos; e vermelho – para clusters sobrepostos. Examinando os resultados apresentados nos gráficos é possível notar a predominância de clusters separados. Consequentemente é possível verificar a potencialidade das características extraídas para identificação das espécies, uma vez, que resultados muito satisfatórios também foram encontrados com espécies de *Passiflora*.

Figura 6.12: Comparação da distância entre os centróides das espécies da Mata Atlântica e de Cerrado. Distância entre as espécies e o centróide da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.

Na Figura 6.13 são apresentados os resultados da análise da distância entre o cluster de uma determinada espécie e as cinquenta amostras utilizadas no experimento. Para que a classificação seja eficiente as cinco amostras em destaque (cinza claro) precisam possuir as menores distâncias entre todo conjunto de amostras. A maioria dos gráficos mostra que a classificação, das espécies da Mata Atlântica e do Cerrado, foi satisfatória em grande parte dos casos.

Figura 6.13: Análise das distância entre todas as amostras utilizadas no experimento e o clusters da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.

A técnica de análise de componentes principais PCA foi aplicada no conjunto de dados. O objetivo principal foi melhorar os resultados da análise de clusters, e com isso

remover possíveis redundâncias entre as características extraídas. O vetor de características que inicialmente possuía 6D (coordenadas x e y dos três pontos de interesse) foi reduzido para 4D. Os resultados são apresentados na Figura 6.14.

Figura 6.14: Redução de dimensionalidade de 6D para 4D através da técnica de PCA. Coordenadas paralelas da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.

Por fim, a ultima análise realizada com as características das espécies da Mata Atlântica e do Cerrado foi através das redes neurais artificiais. Uma rede foi montada com a topologia 6x5x4 e treinada com algoritmo *back-propagation*. O conjunto de dados, contendo cinquenta vetores de características de dez espécies, foi distribuído em duas categorias: (a) Experimento 1 – 80% do conjunto foi utilizado para treinamento e 20% para testes e (b) Experimento 2 – 40% para treinamento e 60% para testes. Na

Tabela 6.3 são apresentados os resultados do primeiro experimento. Nesse caso a rede classificou corretamente 90% das espécies, com taxa de erro = 0,0543.

TABELA 6.3: Resultado da classificação por redes neurais dos dados de complexidade das nervuras – espécies da Mata Atlântica e do Cerrado (80% treinamento/20% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	1	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	1	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	1	0	0	0	0	0	0	0	100%	0%
C4	0	0	0	1	0	0	0	0	0	0	100%	0%
C5	0	0	0	0	1	0	0	0	0	0	100%	0%
C6	0	0	0	0	0	1	0	0	0	0	100%	0%
C7	0	0	0	0	0	0	1	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	1	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	0	0	0%	100%
C10	0	0	0	0	0	0	0	0	0	1	100%	100%
										%	90%	10%

No Experimento 2 a rede foi classificou corretamente 66,7% das espécies, com uma taxa de erro = 0,19367. A Tabela 6.4 apresenta os resultados do Experimento 2.

TABELA 6.4: Resultado da classificação por redes neurais dos dados de complexidade das nervuras – espécies da Mata Atlântica e do Cerrado (40% treinamento/60% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	3	0	0	0	0	0	0	0	0	0	100%	0%
C2	1	2	0	0	0	0	0	0	0	0	67%	33%
C3	0	0	2	0	0	0	0	0	1	0	67%	33%
C4	0	0	0	2	0	1	0	0	0	0	67%	33%
C5	1	0	0	0	1	0	1	0	0	0	33%	67%
C6	0	0	0	1	0	2	0	0	0	0	67%	33%
C7	0	0	0	0	0	0	2	1	0	0	67%	33%
C8	0	0	0	0	0	0	0	2	0	1	67%	33%
C9	0	0	0	0	0	0	0	0	1	2	33%	67%
C10	0	0	0	0	0	0	0	0	0	3	100%	0%
										%	66,7%	33,3%

C – CORRELAÇÃO ENTRE AS ESPÉCIES

Nesse experimento os vetores de características extraídos a partir da dimensão fractal multiescala dos dois conjuntos de espécies, tanto de *Passiflora* quanto da Mata Atlântica e de Cerrado brasileiro, foram reunidos em um único conjunto de dados.

Dessa forma, para esse experimento foram utilizados noventa vetores de características de vinte espécies arbóreas diferentes. O conjunto de informações foi submetido a dois tipos de classificadores: (i) análise de clusters do espaço de características por distância mínima e (ii) redes neurais artificiais treinadas com o algoritmo *back-propagation*.

A Figura 6.15 apresenta os resultados da análise de clusters por distância mínima. Os critérios adotados para coloração dos gráficos foram os mesmos dos anteriores, sendo eles: (i) clusters separados – para os centróides com distância superior a 0,4 a cor selecionada foi verde; (ii) clusters parcialmente sobrepostos – para os centróides com distâncias entre 0,2 e 0,4, esses receberam a cor amarela e (c) clusters sobrepostos – com distâncias inferiores a 0,2, a cor selecionada foi vermelha. Na Figura 6.15a são apresentadas as distâncias entre o centróide da espécie 1 e os demais centróides. Na Figura 6.15b as distâncias entre o centróide da espécie 2 e os demais. E assim até a Figura 6.15t. Analisando os resultados apresentados no gráfico é facilmente notada a plena discriminação das classes pela predominância dos clusters separados (em verde). Novamente é importante ressaltar que esse fato se deve a potencialidade das características extraídas a partir da dimensão fractal multiescala.

Figura 6.15: Análise dos clusters por distância mínima das 20 espécies utilizadas no experimento. (a) comparação da distância da espécie 1 aos demais clusters; (b) comparação da distância da espécie 2 aos demais clusters; e assim por diante até (t) comparação da distância da espécie 20 aos demais clusters.

O treinamento das redes neurais foi realizado através do algoritmo *back-propagation*, utilizando uma topologia de 6x5x5. O conjunto de dados (noventa vetores de características de vinte classes distintas) foi dividido em 75% para treinamento e 25% para teste da rede. A Tabela 6.5 apresenta os resultados da classificação por redes neurais. A rede foi capaz de classificar corretamente 60% das espécies com uma taxa de erro (MSE) nos testes de 0,1524. A coluna %A representa as porcentagens de acertos (classificações corretas) das espécies e, a coluna %E representa as porcentagens de erros (classificações incorretas).

TABELA 6.5: Resultado da classificação por redes neurais dos dados de complexidade das nervuras – todas as espécies (75% treinamento/25% testes).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	%A	%E	
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0%	100%	
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0%	100%	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0%	100%	
9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	100%	0%	
11	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	50%	50%	
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	50%	50%	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0%	100%	
15	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	50%	50%	
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	100%	0%	
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	50%	50%	
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	100%	0%	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	100%	0%	
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	50%	50%	
																					%	60%	40%

6.3.2 Complexidade dos Esqueletos

A. – RESULTADOS DAS ESPÉCIES DE *PASSIFLORA*

O conjunto de imagens de *Passiflora* está dividido em grupos contendo 10 espécies diferentes, totalizando quarenta amostras. Para realização dos experimentos envolvendo complexidade do esqueleto todas as amostras foram eskeletonizadas através do método de eskeletonização multiescala e foi empregado um limiar $T = 10$. Nas imagens resultantes do processo de eskeletonização foi aplicado o método de análise multiescala da dimensão fractal. A Figura 6.16 apresenta os gráficos da análise multiescala utilizando imagens obtidas através da eskeletonização multiescala.

Na Figura 6.17 são apresentados os resultados da análise de clusters por distância mínima. Os vetores de característica extraídos podem ser visualizados na Tabela A.3 do Anexo. Seguindo os mesmos critérios adotados anteriormente, os clusters foram classificados em: (i) clusters separados – distância superior a 0,4; (ii) clusters parcialmente sobrepostos – distância entre 0,2 e 0,4 e (iii) clusters sobrepostos – distância inferior a 0,2. O gráfico da Figura 6.17a representa a distância entre o cluster 1 (da espécie *P.coriacea*) e as 9 demais espécies.

A Figura 6.18 apresenta a distâncias entre os clusters e todas as amostras utilizadas no experimento. Na Figura 6.18a é ilustrada a distância entre as amostras e o cluster da espécie *P.coriacea*; e assim até a Figura 6.18j que apresenta a distância entre as amostras e o cluster da espécie *P.alata*.

A Figura 6.19 apresenta os resultados obtidos através da aplicação da técnica de PCA no conjunto de dados. Os vetores de características extraídos através da dimensão fractal multiescala nas imagens de esqueleto foram reduzidos de 6D para 4D. Da mesma maneira que ocorreu com os outros conjuntos de dados, os resultados não melhoraram substancialmente.

Figura 6.19: Redução de dimensionalidade das características extraídas das espécies de 6D para 4D. Visualização dos resultados através da técnica de coordenadas paralelas.

Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*.

O conjunto de dados extraídos a partir da dimensão fractal multiescala nas imagens eskeletonizadas, foi submetido a experimentos envolvendo redes neurais. O

algoritmo utilizado para treinamento foi o *back-propagation* e, com a topologia 6x5x4. A Tabela 6.6 apresenta os resultados do experimento no qual 75% do conjunto foi utilizada para treinamento e 25% para testes. A taxa de erro MSE durante os testes foi de 0,09583 e a rede foi capaz de classificar corretamente 80% das espécies.

TABELA 6.6: Resultado da classificação por redes neurais dos dados de complexidade dos esqueletos – espécies de *Passiflora* (75% treinamento/25% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	1	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	1	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	0	0	0	0	0	0	0	1	0%	10%
C4	0	0	0	1	0	0	0	0	0	0	100%	0%
C5	0	0	0	0	1	0	0	0	0	0	100%	0%
C6	0	0	0	0	0	1	0	0	0	0	100%	0%
C7	0	0	0	0	0	0	1	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	1	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	1	0	100%	0%
C10	0	0	0	0	0	0	0	0	1	0	0%	100%
										%	80%	20%

Na Tabela 6.7 são representados os resultados no qual 70% das espécies foram classificadas corretamente (MSE = 0,11345). Nesse experimento 50% do conjunto foi utilizado como treinamento e 50% para testes.

TABELA 6.7: Resultado da classificação por redes neurais dos dados de complexidade dos esqueletos – espécies de *Passiflora* (50% treinamento/50% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	0	0	1	1	0	0	0	0	0	0	0%	100%
C2	0	2	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	1	1	0	0	0	0	0	0	50%	50%
C4	0	0	0	2	0	0	0	0	0	0	100%	0%
C5	0	0	0	0	2	0	0	0	0	0	100%	0%
C6	0	0	0	0	0	2	0	0	0	0	100%	0%
C7	0	0	0	0	0	0	1	0	1	0	50%	50%
C8	0	0	0	0	0	1	0	1	0	0	50%	50%
C9	0	0	0	0	0	0	1	0	1	0	100%	0%
C10	0	0	0	0	0	0	0	0	0	2	100%	0%
										%	70%	30%

B. – RESULTADOS DAS ESPÉCIES DA MATA ATLÂNTICA E DO CERRADO

A Figura 6.20 apresenta os gráficos da análise multiescala da dimensão fractal obtidos a partir das imagens esqueletonizadas da Mata Atlântica e de Cerrado. O limiar adotado para esqueletonização foi de $T = 10$. Em cada gráfico uma espécie diferente é representada (cinco amostras por espécie).

Figura 6.20: Gráficos da dimensão fractal multiescala das espécies arbóreas da Mata Atlântica e do Cerrado.
(a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4;
(e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8;
(i) Espécie 9 e (j) Espécie 10.

Na Figura 6.21 são ilustrados os resultados referentes a análise de agrupamentos (clusters) realizada no conjunto de características extraídas. Os vetores de características podem ser visualizados na Tabela A.4 do Anexo. Os critérios adotados para avaliar os resultados foram os mesmos utilizados anteriormente: (i) clusters separados (verde) – distância superior a 0,4; (ii) clusters parcialmente sobrepostos (amarelo) – distância entre 0,2 e 0,4 e (iii) clusters sobrepostos (vermelho) – distância inferior a 0,2.

Nos gráficos da Figura 6.22 são analisadas as distâncias entre os centróides das espécies e as cinquenta amostras utilizadas nos experimentos. Na Figura 6.22a é representada a distância entre a Espécie 1 e as cinquenta amostras utilizadas.

A Figura 6.23 apresenta os resultados da técnica de PCA empregada no conjunto de dados. A dimensionalidade do conjunto foi reduzida de 6D para 4D.

O resultado das redes neurais não foi muito satisfatório para o conjunto de dados de complexidade dos esqueletos. A rede classificou corretamente apenas 50% das espécies e a taxa de erro MSE foi de 0,21094. A Tabela 6.8 apresenta os resultados desse experimento, no qual 80% do conjunto foi utilizado como treinamento e 20% para testes.

TABELA 6.8: Resultado da classificação por redes neurais dos dados de complexidade dos esqueletos – espécies da Mata Atlântica e do Cerrado (80% treinamento/20% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	0	1	0	0	0	0	0	0	0	0	0%	100%
C2	0	1	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	0	0	0	0	0	0	1	0	0%	100%
C4	0	0	0	0	1	0	0	0	0	0	0%	100%
C5	0	0	1	0	0	0	0	0	0	0	0%	100%
C6	1	0	0	0	0	0	0	0	0	0	0%	100%
C7	0	0	0	0	0	0	1	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	1	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	1	0	100%	0%
C10	0	0	0	0	0	0	0	0	0	1	100%	0%
										%	50%	50%

C – CORRELAÇÃO ENTRE AS ESPÉCIES

No experimento envolvendo a correlação das espécies, todas as informações extraídas das espécies foram reunidas em um único conjunto de dados. Esse conjunto possui noventa amostras divididas em vinte classes de espécies (dez espécies de *Passiflora* e dez espécies da Mata Atlântica e do Cerrado).

A Figura 6.24 apresenta os resultados da análise de agrupamentos por distância mínima. Os critérios adotados para avaliação foram os mesmos utilizados anteriormente (clusters separados, clusters parcialmente sobrepostos e clusters sobrepostos).

Figura 6.24: Análise dos clusters por distância mínima das 20 espécies utilizadas no experimento. (a) distância da espécie 1 aos demais clusters; (b) distância da espécie 2 aos demais clusters; e assim até (t) distância da espécie 20 aos demais clusters.

Com as redes neurais foram classificadas corretamente 50% das espécies, utilizando o algoritmo *back-propagation* (75% do conjunto para treinamento e 25% para testes). A Tabela 6.9 apresenta os resultados da tentativa de classificação por redes neurais. A taxa de erro MSE encontrada durante os testes foi de 0,19933.

TABELA 6.9: Resultado da classificação por redes neurais dos dados de complexidade dos esqueletos – todas as espécies (75% treinamento/25% testes).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	%A	%E	
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0%	100%	
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0%	100%	
10	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0%	100%	
12	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	50%	50%	
13	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
14	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	50%	50%	
15	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0%	100%	
16	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	50%	50%	
17	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	50%	50%	
18	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	50%	50%	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	100%	0%	
20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	50%	50%	
																					%	50%	50%

6.4 Experimento II: Biometria de Órgãos Foliares

O objetivo dos experimentos envolvendo biometria dos órgãos foliares, é classificar as espécies através de medidas biométricas extraídas tanto das nervuras, quanto dos esqueletos. Da mesma maneira dos experimentos utilizando complexidade, os experimento com biometria podem ser divididos em três fases principais: (i) processamento; (ii) extração de características e (iii) reconhecimento de padrões.

Na fase de processamento, a base de imagens contendo 90 imagens digitalizadas de espécies arbóreas (40 imagens de espécies de *Passiflora* e 50 imagens de espécies da Mata Atlântica e do Cerrado), é adequada para que posteriormente sejam extraídas suas características. O resultado dessa fase são dois subconjuntos de 90 imagens. O primeiro

contém as imagens das nervuras segmentadas e, o segundo possui as imagens esqueletonizadas da área foliar das espécies.

A próxima fase se concentra na extração de características biometrias dos conjuntos de imagens processados. Essas informações biométricas são extraídas utilizando os algoritmos apresentados na seção 3.4. Os algoritmos foram empregados tanto nas imagens segmentadas das nervuras quanto nas imagens esqueletonizadas. Foram consideradas 19 (dezenove) características biométricas: número de bifurcações; distância média entre as bifurcações e o ponto de origem; desvio padrão das distâncias entre as bifurcações e o ponto de origem; distância média entre todas as bifurcações; distância entre o ponto de origem e o ponto de destino; distância entre o ponto de origem e o centróide; número de pontos extremos das nervuras ou dos esqueletos; distância média entre os centróides e os pontos extremos; desvio padrão das distâncias entre os centróides e os pontos extremos; distância média entre o ponto de origem e os pontos extremos; desvio padrão das distâncias entre os pontos de origem e os pontos extremos; tamanho (área) da nervura ou esqueleto; número de ramificações; tamanho médio das ramificações; desvio padrão das ramificações; número de ângulos extraídos; tamanho da ramificação principal; média dos ângulos das ramificações que se originam na ramificação principal e desvio padrão das medidas dos ângulos das ramificações que se originam na ramificação principal. Essas características foram escolhidas com base em critérios adotados pelos taxonomistas para classificação das espécies, mais informações podem ser encontradas em (WING, WILF *et al.*, 1999).

Na fase de reconhecimento de padrões foram utilizadas as mesmas técnicas empregadas nos experimentos com complexidade: classificação por distância mínima (análise de clusters); análise de componentes principais (PCA) e redes neurais artificiais. Nas próximas seções os resultados dos experimentos com biometria são apresentados e discutidos.

6.4.1 Biometria das Nervuras

A. – RESULTADOS DAS ESPÉCIES DE *PASSIFLORA*

As espécies de *Passiflora* estão distribuídas entre quarenta amostras divididas em dez classes diferentes. As nervuras dessas amostras foram previamente segmentadas e, a

partir dessas imagens informações biométricas foram extraídas. Os dados podem ser visualizados na Tabela A.5 do Anexo.

Na Figura 6.25 são apresentados os resultados referentes ao método de classificação baseado na análise de clusters. Nesse método são calculados os centros de massa (centróide) de cada espécie e determinada a distância euclidiana entre eles. Para avaliação dos resultados foram adotados os seguintes critérios: (i) clusters separados (verde) – quando a distância entre os centróides foi superior a 1,0; (ii) clusters parcialmente sobrepostos (amarelo) – para distâncias entre os centróides superiores a 0,6 e inferiores a 1,0 e (iii) clusters sobrepostos (vermelho) – quando a distância entre os centróides foi inferior a 0,6. A Figura 6.25a representa as distâncias entre o centróide da espécie *Passiflora coriacea* e as demais espécies. Na Figura 6.25b o centróide da espécie *Passiflora foetida* e as demais espécies. E assim até a Figura 6.25j que ilustra a distância entre o centróide da espécie *Passiflora alata* e as demais espécies.

Figura 6.25: Resultados experimentais da análise baseada na distância entre os clusters. Comparação das distâncias entre os clusters das Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*.

Na Figura 6.26 são representadas as distâncias entre cada uma das quarenta amostras de *Passiflora* e, os centróides das espécies. Nos gráficos, as amostras estão agrupadas por espécie, sendo que as barras verticais de 1-4 representam a espécie *Passiflora coriacea*, de 5-8 a espécie *Passiflora foetida* e assim por diante. As barras em cinza claro ilustram a espécie que está sendo analisada. Para uma classificação eficiente, a distância das amostras da espécie destacada precisam ser as quatro menores distâncias entre todas analisadas. A Figura 6.26a ilustra a distância entre todas as amostras e o centróide da espécie *Passiflora coriacea*. Nesse caso é possível notar a correta classificação da espécie, pois a distância das suas amostras (em cinza claro) são as menores distâncias entre todas analisadas.

A Figura 6.27 ilustra os resultados alcançados através da utilização da técnica de análise de componentes principais no conjunto de dados. O vetor de características que

inicialmente possuía 19 (dezenove) dimensões foi reduzido para 10 (dez) dimensões. Para visualização dos resultados foi utilizada a técnica de coordenadas paralelas. A Figura 6.27a apresenta os resultados da redução de dimensionalidade da espécie *Passiflora coriacea*. Os eixos em cinza escuro representam a espécie analisada e os eixos em cinza claro as demais espécies.

As informações biométricas extraídas das espécies também foram utilizadas na classificação por redes neurais. A rede foi treinada com algoritmo *back-propagation* com uma topologia de 19x10x4 (dezenove neurônios na camada entrada; dez neurônios na camada oculta e quatro neurônios na camada de saída). A Tabela 6.10 apresenta os resultados da classificação por redes neurais. Nesse experimento, 75% do conjunto foi

utilizado como treinamento e 25% para testes. A rede foi capaz de classificar corretamente 90% das espécies, e a taxa de erro MSE encontrada durante os testes foi de 0,02349. A classe C1 representa a espécie *Passiflora coriacea*, C2 – *P.foetida*, C3 – *P.miersii*, C4 – *P.organensis*, C5 – *P.pohlii*, C6 – *P. suberosa*, C7 – *P. amethystina*, C8 – *P.caerulea*, C9 – *P.giberti* e C10 – *P.alata*.

TABELA 6.10: Resultado da classificação por redes neurais dos dados de biometria das nervuras – espécies de *Passiflora* (75% treinamento/25% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	1	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	1	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	1	0	0	0	0	0	0	0	100%	0%
C4	0	0	0	1	0	0	0	0	0	0	100%	0%
C5	0	0	0	0	1	0	0	0	0	0	100%	0%
C6	0	0	0	0	0	1	0	0	0	0	100%	0%
C7	0	0	0	0	0	0	1	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	1	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	1	0	100%	0%
C10	0	0	0	0	0	0	0	0	1	0	0%	100%
										%	90%	10%

A Tabela 6.11 representa os resultados das redes neurais em que 50% do conjunto foi utilizado para treinamento e 50% para testes. A rede classificou corretamente 80% das espécies com uma taxa de erro MSE de 0,09843

TABELA 6.11: Resultado da classificação por redes neurais dos dados de biometria das nervuras – espécies de *Passiflora* (50% treinamento/50% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	2	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	2	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	1	0	0	0	1	0	0	0	50%	50%
C4	0	0	0	2	0	0	0	0	0	0	100%	0%
C5	0	0	0	0	2	0	0	0	0	0	100%	0%
C6	0	0	0	0	0	1	0	0	0	1	50%	50%
C7	0	0	0	0	0	0	2	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	2	0	0	100%	0%
C9	0	0	0	0	0	0	1	0	1	0	50%	50%
C10	0	0	0	0	0	0	0	0	1	1	50%	50%
										%	80%	20%

B. – RESULTADOS DAS ESPÉCIES DA MATA ATLÂNTICA E DO CERRADO

As espécies da Mata Atlântica e do Cerrado estão distribuídas em cinquenta amostras divididas em dez classes diferentes. Esse conjunto de imagens teve suas nervuras segmentadas e, posteriormente suas informações biométricas foram extraídas. Os vetores de características de biometria podem ser visualizados na Tabela A.6 do Anexo.

A Figura 6.28 apresenta o resultado da análise de clusters empregada no conjunto de dados biométricos das espécies da Mata Atlântica e do Cerrado. Os critérios adotados para avaliação dos resultados foram os mesmo adotados anteriormente: (i) clusters separados (verde) – para distâncias superiores a 1,0; (ii) clusters parcialmente sobrepostos (amarelo) – para distâncias entre 0,6 e 1,0 e (iii) clusters sobrepostos (vermelho) – para distâncias inferiores a 0,6. A Figura 6.28a representa a distância entre o centróide da Espécie 1 e os centróides das demais espécies. Na Figura 6.28 é ilustrada a distância entre o centróide da Espécie 2 e os demais centróides e, assim por diante.

Figura 6.28: Comparação da distância entre os centróides das espécies da Mata Atlântica e de Cerrado. Distância entre as espécies e o centróide da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.

A Figura 6.29 apresenta a distância entre cada centróide e todas as amostras, da Mata Atlântica e do Cerrado, analisadas. Em cinza claro estão destacadas as amostras da espécie que está sendo analisada. Na Figura 6.29 é ilustrada a distância entre todas as amostras e o centróide da Espécie 1. Na Figura 6.29b a distância entre as amostras e o centróide da Espécie 2, e assim por diante.

Figura 6.29: Análise das distâncias entre todas as amostras utilizadas no experimento e os clusters da: (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.

A Figura 6.30 apresenta o resultado da técnica de PCA empregada no conjunto de dados. A dimensionalidade do conjunto que inicialmente era de 19D foi reduzida para 10D. Para visualização foi utilizada a técnica de coordenadas paralelas.

Figura 6.30: Visualização dos resultados da técnica de PCA através de coordenadas paralelas. (a) Espécie 1; (b) Espécie 2; (c) Espécie 3; (d) Espécie 4; (e) Espécie 5; (f) Espécie 6; (g) Espécie 7; (h) Espécie 8; (i) Espécie 9; (j) Espécie 10.

A Tabela 6.12 representa os resultados da classificação por redes neurais. O conjunto de dados biométricos foi dividido em 80% para treinamento (utilizando o algoritmo *back-propagation*) da rede e 20% para testes. A rede foi capaz de classificar corretamente 90% das espécies e a taxa de erro MSE encontrada foi de 0,05938.

TABELA 6.12: Resultado da classificação por redes neurais dos dados de biometria das nervuras – espécies da Mata Atlântica e do Cerrado (80% treinamento/20% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	1	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	1	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	1	0	0	0	0	0	0	0	100%	0%
C4	0	0	0	1	0	0	0	0	0	0	100%	0%
C5	0	0	0	0	1	0	0	0	0	0	100%	0%
C6	0	0	0	0	0	1	0	0	0	0	100%	0%
C7	0	0	0	0	0	0	0	1	0	0	0%	100%
C8	0	0	0	0	0	0	0	1	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	1	0	100%	0%
C10	0	0	0	0	0	0	0	0	0	1	100%	0%
										%	90%	10%

A Tabela 6.13 apresenta os resultados da rede utilizando 40% do conjunto para treinamento e 60% para testes. Nesse caso a taxa de erro MSE encontrada foi de 0,09763 e a rede foi capaz de classificar corretamente 80% das espécies.

TABELA 6.13: Resultado da classificação por redes neurais dos dados de biometria das nervuras – espécies da Mata Atlântica e do Cerrado (40% treinamento/60% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	2	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	2	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	1	0	0	0	0	1	0	0	50%	50%
C4	0	0	0	2	0	0	0	0	0	0	100%	0%
C5	1	0	0	0	1	0	0	0	0	0	50%	50%
C6	0	0	0	0	0	1	1	0	0	0	50%	50%
C7	0	0	0	0	0	0	2	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	2	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	1	1	50%	50%
C10	0	0	0	0	0	0	0	0	0	2	100%	0%
										%	80%	20%

C – CORRELAÇÃO ENTRE AS ESPÉCIES

Nesse experimento os vetores de características de todas as espécies, tanto de *Passiflora* quanto da Mata Atlântica e do Cerrado foram reunidos em um único grupo. Dessa forma, o conjunto de dados utilizado para classificação contém noventa amostras distribuídas em vinte classes diferentes.

A Figura 6.31 apresenta os resultados da classificação através da técnica de análise de clusters. Os critérios para avaliação são os mesmo citados anteriormente (clusters separados; clusters parcialmente sobrepostos e clusters sobrepostos). A Figura 6.31a representa a distância entre o centróide da Espécie 1 e as dezenove demais espécies. A Figura 6.31b ilustra a distância da Espécie 2 e as demais espécies, e assim por diante.

A Tabela 6.14 apresenta os resultados da classificação por redes neurais aplicada no conjunto das vinte espécies. Para esse conjunto, que contém medidas biométricas das nervuras de órgãos foliares, a rede classificou corretamente 90% das espécies. A taxa de erro MSE encontrada durante os testes foi de 0,05392.

TABELA 6.14: Resultado da classificação por redes neurais dos dados de biometria das nervuras – todas as espécies (75% treinamento/25% testes).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	%A	%E	
1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	100%	0%	
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	100%	0%	
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	100%	0%	
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	100%	0%	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	100%	0%	
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	100%	0%	
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	100%	0%	
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	100%	0%	
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	100%	0%	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	100%	0%	
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	100%	0%	
																					%	90%	10%

6.4.2 Biometria dos Esqueletos

A. – RESULTADOS DAS ESPÉCIES DE *PASSIFLORA*

Para a realização dos experimentos de biometria dos esqueletos de órgãos foliares, o conjunto de imagens das espécies de *Passiflora* foi eskeletonizada através de método de eskeletonização multiescala. O critério adotado para o limiar foi de $T = 10$. A partir das imagens eskeletonizadas foram aplicados os métodos para extração de medidas biométricas. As informações coletadas podem ser visualizadas na Tabela A.7 do Anexo.

A Figura 6.32 apresenta os resultados da análise de clusters empregada nos vetores de características extraídos através de biometria. O critério adotado para avaliação dos resultados foi: (i) clusters separados (verde) – para distâncias superiores a 1,0; (ii) clusters parcialmente sobrepostos (amarelo) para distância entre 0,6 e 1,0 e (iii)

clusters sobrepostos (vermelho) – para distâncias inferiores a 0,6. A Figura 6.32a apresenta a distância entre o centróide da espécie *Passiflora coriacea* e as demais espécies. As demais figuras seguem o mesmo modelo: Figura 6.32b – distância entre o centróide da espécie *Passiflora foetida* e as demais espécies. E assim até a Figura 6.32j que ilustra a distância entre o centróide da espécie *Passiflora alata* e as demais espécies.

A Figura 6.33 apresenta uma análise mais detalhada dos agrupamentos. Nesse caso, cada amostra foi comparada ao centróide de uma determinada espécie. Na Figura 6.33a é representada a distância entre todas as amostras e o centróide da espécie *Passiflora coriacea*. A Figura 6.33b representa a distância entre o centróide da espécie *Passiflora foetida* e as demais espécies. E assim até a Figura 6.33j que ilustra a distância entre o centróide da espécie *Passiflora alata* e as demais espécies. As barras em destaque (cinza claro) ilustram a espécie que está sendo comparada e, conseqüentemente são as barras que devem possuir a menor distância.

Figura 6.33: Distâncias entre as amostras e o centróides das Espécies: (a) *Passiflora coriacea*; (b) *P. foetida*; (c) *P. miersii*; (d) *P. organensis*; (e) *P. pohlii*; (f) *P. suberosa*; (g) *P. amethystina*; (h) *P. caerulea*; (i) *P. giberti* e (j) *P. alata*.

Na Figura 6.34 são visualizados os resultados do processo de redução de dimensionalidade do conjunto de dados. O conjunto de vetores de características que inicialmente possuía 19D foi reduzido para 10D através da técnica de PCA. Os resultados da redução foram visualizados através da técnica de coordenadas paralelas. A Figura 6.34a apresenta as coordenadas paralelas da espécie *Passiflora coriacea*. A Figura 6.34b as coordenadas paralelas da espécie *P. foetida*. E assim até a Figura 6.34j que representa as coordenadas paralelas da espécie *Passiflora alata*.

Outro método de classificação foi aplicado no conjunto de dados: as redes neurais artificiais. A Tabela 6.15 apresenta os resultados da rede que utilizou 75% do conjunto para treinamento e 25% para testes. Nessa configuração a rede classificou corretamente 80% das espécies (erro MSE = 0,10392).

TABELA 6.15: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – espécies de *Passiflora* (75% treinamento/25% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	0	1	0	0	0	0	0	0	0	0	0%	100%
C2	0	1	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	1	0	0	0	0	0	0	0	100%	0%
C4	0	0	0	1	0	0	0	0	0	0	100%	0%
C5	0	0	0	0	1	0	0	0	0	0	100%	0%
C6	0	0	0	0	0	1	0	0	0	0	100%	0%
C7	0	0	0	0	0	0	1	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	1	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	1	0	100%	0%
C10	0	0	0	0	0	0	1	0	0	0	0%	100%
										%	80%	20%

A Tabela 6.16 apresenta os resultados das redes neurais utilizando 50% do conjunto para treinamento e 50% para testes. A rede classificou corretamente 70% das espécies e a taxa de erro encontrada durante os testes foi de 0,12983.

TABELA 6.16: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – espécies de *Passiflora* (50% treinamento/50% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	1	0	0	1	0	0	0	0	0	0	50%	50%
C2	0	2	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	1	0	0	0	0	0	0	1	50%	50%
C4	1	0	0	1	0	0	0	0	0	0	50%	50%
C5	0	0	0	0	2	0	0	0	0	0	100%	0%
C6	0	0	0	0	0	1	0	1	0	0	50%	50%
C7	0	0	0	0	0	0	2	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	2	0	0	100%	0%
C9	2	0	0	0	0	0	0	0	0	0	0%	100%
C10	0	0	0	0	0	0	0	0	0	2	100%	0%
										%	70%	30%

B. – RESULTADOS DAS ESPÉCIES DA MATA ATLÂNTICA E DO CERRADO

O processo extração de medidas biométricas, nos órgãos foliares da Mata Atlântica e do Cerrado, ocorreu da mesma maneira que as espécies de *Passiflora*. Inicialmente todo conjunto de imagens passou pelo processo de esqueletonização, através da esqueletonização multiescala (com limiar $T=10$), e em seguida, as medidas biométricas foram extraídas. Os vetores de característica coletados podem ser visualizados na Tabela A.8 do Anexo.

A Figura 6.35 apresenta os resultados da classificação das espécies através da análise de clusters. Os critérios para avaliação do método foram: clusters separados (distância $> 1,0$); clusters parcialmente (0,6 $\leq$ distância $\leq 1,0$) sobrepostos e clusters sobrepostos (distância $< 0,6$). A Figura 6.35a representa a distância entre o centróide do cluster da Espécie 1 e as demais espécies. A Figura 6.35b representa a distância entre o centróide do cluster da Espécie 2 e as demais espécies. E assim até a Figura 6.35j que representa a distância do centróide da Espécie 10 e as demais espécies.

Na Figura 6.36 são apresentadas as distâncias entre os centróides das espécies e, todas as cinquenta amostras utilizadas no experimento. A Figura 6.36a representa a distância entre o centróide da Espécie 1 e as cinquenta amostras. Na Figura 6.36b a distância entre o centróide da Espécie 2 e as demais espécies. E assim até a Figura 6.36j que ilustra a distância entre o centróide da Espécie 10 e as demais espécies.

A Figura 6.37 apresenta o resultado da técnica de análise de componentes principais (PCA) aplicada no conjunto de dados. A dimensionalidade dos dados foi reduzida de 19D para 10D. A visualização dos resultados foi realizada através do método de coordenadas paralelas. A Figura 6.37a representa as coordenadas paralelas da Espécie 1. Na Figura 6.37b é ilustrada as coordenadas paralelas da Espécie 2. E assim até a Figura 6.37j que representa as coordenadas da Espécie 10.

A Tabela 6.17 apresenta o resultado da classificação por redes neurais artificiais. O conjunto de dados de medidas biométricas, obtidas através da esqueletonização multiescala, das espécies da Mata Atlântica e do Cerrado foi dividido em 80% para treinamento da rede e 20% para testes. A rede classificou corretamente 70% das espécies com taxa de erro MSE = 0,119328.

TABELA 6.17: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – espécies da Mata Atlântica e do Cerrado (80% treinamento/20% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	1	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	1	0	0	0	0	0	0	0	0	100%	0%
C3	0	0	0	0	0	0	0	0	0	1	0%	100%
C4	0	0	0	1	0	0	0	0	0	0	100%	0%
C5	0	0	1	0	0	0	0	0	0	0	0%	100%
C6	1	0	0	0	0	0	0	0	0	0	0%	100%
C7	0	0	0	0	0	0	1	0	0	0	100%	0%
C8	0	0	0	0	0	0	0	1	0	0	100%	0%
C9	0	0	0	0	0	0	0	0	1	0	100%	0%
C10	0	0	0	0	0	0	0	0	0	1	100%	0%
										%	70%	30%

Na segunda rede treinada 40% do conjunto foi utilizado para treinamento e, 60% para testes. Nessa configuração a rede classificou corretamente 43% das espécies com taxa de erro MSE = 0,24098. A Tabela 6.18 apresenta esses resultados.

TABELA 6.18: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – espécies da Mata Atlântica e do Cerrado (40% treinamento/60% testes).

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	% A	% E
C1	3	0	0	0	0	0	0	0	0	0	100%	0%
C2	0	2	0	0	0	0	0	0	0	1	67%	33%
C3	1	0	1	0	0	0	0	0	0	1	33%	67%
C4	0	0	1	1	0	0	0	1	0	0	33%	67%
C5	0	2	0	0	1	0	0	0	0	0	33%	67%
C6	1	0	0	0	0	2	0	0	0	0	67%	33%
C7	1	0	0	0	0	0	0	0	2	0	0%	100%
C8	0	0	0	0	0	0	0	2	0	1	67%	33%
C9	0	0	0	0	0	0	0	0	0	3	0%	100%
C10	0	0	0	0	0	0	0	0	2	1	33%	67%
										%	43%	57%

C – CORRELAÇÃO ENTRE AS ESPÉCIES

No experimento envolvendo a correlação das espécies, todo conjunto de dados (espécies de *Passiflora* e Mata Atlântica e do Cerrado) foi reunido em um único conjunto. Dessa forma para tentativa de classificação das espécies foi utilizado um conjunto de noventa amostras distribuídas em vinte espécies diferentes.

A Figura 6.38 apresenta os resultados da análise de agrupamentos (clusters) das espécies. A Figura 6.38a apresenta a distância entre o centróide da Espécie 1 e as demais espécies. A Figura 6.38b a distância entre o centróide da Espécie 2 e as demais espécies. E assim até a Figura 6.38t que representa a distância entre o centróide da Espécie 20 e as demais espécies. Os critérios para avaliação do método de classificação são os mesmos citados anteriormente: (i) clusters separados (verde) – para distâncias superiores a 1,0; (ii) clusters parcialmente sobrepostos (amarelo) – para distâncias entre 0,6 e 1,0 e (iii) clusters sobrepostos (vermelho) – para distâncias inferiores a 0,6.

(a)

(b)

(c)

Figura 6.38: Análise de clusters das 20 espécies. (a) distância entre o centróide da espécie 1 e as demais espécies; assim até (t) distância entre o centróide da espécie 20 e as demais espécies.

O conjunto de dados foi classificado através da técnica de redes neurais (RNA). O algoritmo escolhido para treinamento da rede foi o *back-propagation*. A Tabela 6.19 apresenta os resultados do processo de classificação por RNA. Nesse experimento o

conjunto de dados (noventa vetores de características de vinte classes distintas) foi dividido em 75% para treinamento e 25% para teste da rede. A rede foi capaz de classificar corretamente 55% das espécies com uma taxa de erro (MSE) nos testes de 0,1956. A coluna %A representa as porcentagens de acertos (classificações corretas) das espécies e, a coluna %E representa as porcentagens de erros (classificações incorretas).

TABELA 6.19: Resultado da classificação por redes neurais dos dados de biometria dos esqueletos – todas as espécies (75% treinamento/25% testes).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	%A	%E	
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0%	100%	
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
9	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
10	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	100%	0%	
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0%	100%	
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0%	100%	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	100%	0%	
15	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0%	100%	
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	100%	0%	
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	100%	0%	
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	100%	0%	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0%	100%	
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0%	100%	
																					%	55%	45%

6.5 Experimento III: Comparação dos Métodos

Nos experimentos apresentados inicialmente (Experimento I e Experimento II), o objetivo principal foi verificar a potencialidade tanto de métodos de complexidade quanto de biometria, na classificação das espécies. Contrário a esse paradigma o objetivo do Experimento III é avaliar a potencialidade do local de onde as características são extraídas, ou seja, avaliar se as características extraídas das nervuras são melhores ou piores que as características extraídas dos esqueletos. Dessa forma, os conjuntos de dados foram reunidos em dois grupos:

- (a) complexidade e biometria das nervuras: esse conjunto contém todas as características extraídas das nervuras, de todas as espécies utilizadas nos

experimentos. Assim, esse conjunto de dados contém: noventa amostras de espécies arbóreas (quarenta de *Passiflora* e cinquenta da Mata Atlântica e do Cerrado, divididas em vinte espécies diferentes), sendo que para cada amostra vinte e cinco características foram extraídas (seis características de complexidade e dezenove características de biometria).

- (b) Complexidade e biometria dos esqueletos: da mesma forma do conjunto anterior, esse grupo contém noventa vetores de características com vinte e cinco características por vetor (seis características de complexidade e dezenove de biometria).

No último experimento realizado, todas as características extraídas de todas as espécies são reunidas em um único conjunto. Dessa forma, foram agrupados noventa vetores de características (das vinte espécies arbóreas distintas) contendo cinquenta dimensões: doze dimensões referentes aos dados de complexidade, tanto de nervura quanto esqueleto e, trinta e oito dimensões relativas aos dados de biometria, tanto de nervura quanto de esqueleto. Os dados foram classificados através da técnica de análise de agrupamentos (clusters) e redes neurais artificiais. Os resultados desses métodos são apresentados nas próximas seções.

6.5.1 Comparação: Complexidade e Biometria das Nervuras

A Figura 6.39 apresenta os resultados da análise de agrupamento (clusters) das características de complexidade e biometria das nervuras. Esse conjunto contém noventa vetores de características (com vinte e cinco dimensões) com vinte classes (espécies arbóreas) diferentes. Para avaliação dos resultados foram adotados os seguintes critérios: (i) clusters separados (verde) – para distâncias entre os centróides superiores a 1,0; (ii) clusters parcialmente sobrepostos (amarelo) – para distância entre os centróides no intervalo de 0,6 a 1,0 e (iii) clusters sobrepostos (vermelho) – para distância entre os centróides inferiores a 0,6. Na Figura 6.39a são representadas as distâncias entre o centróide da Espécie 1 e as demais espécies; e assim até a Figura 6.39t que representa a distância a Espécie 20 e as demais espécies.

Figura 6.39: Análise dos clusters por distância mínima. (a) distância da espécie 1 aos demais clusters; (b) distância da espécie 2 aos demais clusters; e assim até (t) distância da espécie 20 aos demais clusters.

Para classificação por redes neurais foi utilizado o algoritmo de treinamento *back-propagation*. A topologia adota para rede foi 25x10x5 (vinte e cinco neurônios na camada de entrada; 10 neurônios na camada oculta e cinco neurônios na camada de saída). A Tabela 6.20 apresenta os resultados da classificação por redes neurais no conjunto de complexidade e biometria das nervuras. Nesse experimento 75% do conjunto foi necessário para treinamento da rede e 25% para testes. A rede classificou corretamente 95% das espécies e a taxa de erro (MSE) foi de 0,0143.

TABELA 6.20: Resultado da classificação por redes neurais dos dados de complexidade e biometria das nervuras – todas as espécies (75% treinamento/25% testes).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	%A	%E
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	100%	0%
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	100%	0%
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	100%	0%
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	100%	0%
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	100%	0%
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	100%	0%
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	100%	0%
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	100%	0%
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	100%	0%
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0%	100%
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	100%	0%
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	100%	0%
																				%	95%	5%

6.5.2 Comparação: Complexidade e Biometria dos Esqueletos

Na Figura 6.40 são ilustrados os resultados da classificação por agrupamentos realizada no conjunto de dados. Esse conjunto contém características obtidas através dos métodos

computacionais de complexidade e biometria dos esqueletos das espécies. Para avaliação do método o critério adotado foi o mesmo citado anteriormente: clusters separados; sobrepostos e parcialmente sobrepostos. A Figura 6.40a apresenta a distância entre o centróide da Espécie 1 e as demais espécies. E assim até a Figura 6.40t que ilustra a distância entre o centróide de Espécie 20 e as demais espécies.

Figura 6.40: Análise de clusters. (a) distância entre a espécie 1 e as demais espécies; assim até (t) distância entre a espécie 20 e as demais espécies.

A Tabela 6.21 apresenta os resultados do processo de classificação por redes neurais artificiais. Para esse experimento foram utilizados 75% do conjunto para treinamento e 25% para testes. A rede classificou corretamente 65% das espécies, com taxa de erro (MSE) de 0,18982.

TABELA 6.21: Resultado da classificação por redes neurais dos dados de complexidade e biometria dos esqueletos – todas as espécies (75% treinamento/25% testes).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	%A	%E
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%
9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0%	100%
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	100%	0%
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	100%	0%
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0%	100%
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0%	100%
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0%	100%
15	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	100%	0%
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	100%	0%
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	100%	0%
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0%	100%
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	100%	0%
																				%	65%	35%

6.5.3 Comparação: Todas as características extraídas

A Figura 6.41 apresenta os resultados da análise de agrupamentos (clusters) utilizando todas as características extraídas (cinquenta) de todas as espécies utilizadas (vinte). Novamente, os critérios adotados são os mesmos determinados anteriormente: (i) clusters separados (verde) – distância superior a 1,0; (ii) clusters parcialmente sobrepostos (amarelo) – distâncias entre 0,6 e 1,0 e (iii) clusters sobrepostos (vermelho) – distâncias inferiores a 0,6. Na Figura 6.41a é representada a distância entre o centróide da Espécie 1 e as demais dezenove espécies. A Figura 6.41b ilustra a distância entre o centróide da Espécie 2 e as demais espécies. E assim até a Figura 6.41t que representa a distância do centróide da Espécie 20 até as demais espécies.

Figura 6.41: Análise das distância entre os centróides dos agrupamentos. Distância entre os centróides das espécies e o centróide da: (a) Espécie 1; (b) Espécie 2; e assim até (t) Espécie 20.

A Tabela 6.22 apresenta os resultados da classificação por redes neurais. O algoritmo de treinamento utilizado foi o *back-propagation*, com uma topologia de 50x25x5. A rede classificou corretamente 85% das espécies, com erro (MSE) = 0,1254.

TABELA 6.22: Resultado da classificação por redes neurais de todos os dados coletados de todas as espécies (75% treinamento/25% testes).

todas as espécies (75% treinamento/25% testes):																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	%A	%E	
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	100%	0%	
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	100%	0%	
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	100%	0%	
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	100%	0%	
13	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	100%	0%	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	100%	0%	
15	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%	100%	
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	100%	0%	
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	100%	0%	
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	100%	0%	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0%	100%	
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	100%	0%	
																					%	85%	15%

6.6 Análise e Discussão dos Resultados

Nessa seção os resultados experimentais apresentados ao longo do capítulo 6 são analisados e discutidos. Para facilitar o acompanhamento das linhas de discussões seguidas, tomaremos como base o diagrama de execução dos experimentos apresentado anteriormente e, novamente ilustrado na Figura 6.42.

Figura 6.42: Diagrama dos experimentos executados para identificação das espécies vegetais.

Como ilustrado na Figura 6.42, os experimentos para identificação de espécies arbóreas através da análise de forma interna de seus órgãos foliares, foram divididos em três módulos:

- Complexidade (Experimento I): cujo objetivo foi classificar as espécies através de informações obtidas através da dimensão fractal.
- Biometria (Experimento II): no qual a classificação das espécies utiliza como base informações biométricas dos órgãos foliares
- Comparação (Experimento III): nesse módulo as informações coletadas sobre complexidade e biometria foram correlacionadas, tentando alcançar melhores resultados.

Todos os experimentos foram realizados com casos reais de espécies, nos quais dois grupos de imagens foram utilizados: (a) imagens de espécies de maracujás silvestres (*Passiflora*) e (b) imagens de espécies da Mata Atlântica e do Cerrado. Os experimentos com esses grupos foram sempre executados de maneira separada e, por fim os grupos de imagens foram reunidos. Essa separação aconteceu por causa da

importância da classificação de espécies dentro de um mesmo gênero, principalmente no caso dos maracujás silvestres. Devido a semelhanças morfológicas e moleculares, a correta classificação das espécies de *Passiflora* (cerca de 450) é altamente complexa. Os resultados obtidos com a classificação das espécies de maracujá foram de fundamental importância para pesquisas correlatas na ESALq/USP.

No Experimento I o objetivo principal foi identificar espécies arbóreas através de informações da sua complexidade. Para isso, características foram coletadas a partir de pontos de interesse da curva resultante da análise multiescala, tanto para as nervuras quanto para os esqueletos dos órgãos foliares. Assim, o experimento foi dividido em dois subgrupos: Complexidade das Nervuras e Complexidade dos Esqueletos.

Avaliando os resultados obtidos com experimentos que envolveram a complexidade das nervuras, podemos concluir:

- A. Resultados das Espécies de *Passiflora*: em uma escala de: péssimo, ruim, regular, bom e ótimo, os resultados alcançados podem ser avaliados entre bom e ótimo. As redes neurais treinadas classificaram corretamente 90% das espécies, porém a análise de clusters não foi capaz de separar todas as espécies.
- B. Resultados das Espécies da Mata Atlântica e do Cerrado: nesse caso os resultados podem ser determinados como bom. A análise de clusters não alcançou resultados muito satisfatórios e, as redes classificaram (em média) 80% das espécies.
- C. Correlação entre as Espécies: as características dos resultados nesse experimento foram inversa ao caso B. As redes não foram muito bem na classificação das espécies, apenas 60%. Por outro lado, a análise de clusters obteve resultados satisfatórios. Em geral, esses resultados podem ser avaliados como bom.

Com resultados obtidos pelos experimentos que utilizaram características de complexidade extraídas dos esqueletos, podemos concluir:

- A. Resultados das Espécies de *Passiflora*: os resultados para esse grupo podem ser avaliados como bons. As redes classificaram, em média, 75% das espécies corretamente e análise de clusters foi satisfatória.
- B. Resultados das Espécies da Mata Atlântica e do Cerrado: nesse caso os resultados podem ser avaliados como regular. Na análise de clusters, grande parte das espécies não foi separada corretamente (diversos clusters ficaram sobrepostos). E, as redes neurais classificaram apenas 50% das espécies.
- C. Correlação entre as Espécies: os resultados também não foram satisfatórios sendo classificadas como regular. As redes novamente classificaram apenas 50% das espécies, e análise de clusters não foi capaz de separar corretamente a maioria das espécies.

Analizando os resultados obtidos com a Complexidade dos Órgãos foliares, podemos concluir que: (i) os resultados obtidos com classificação das características extraídas das nervuras são consideravelmente melhores do que os resultados obtidos com as características dos esqueletos; (ii) em geral a classificação por análise de clusters foi melhor que a classificação por redes neurais, isso pode ser justificado pela quantidade de amostras utilizadas nos experimentos e (iii) o resultado da classificação das espécies de *Passiflora* foi melhor que as espécies da Mata Atlântica e do Cerrado.

O objetivo principal do Experimento II foi caracterizar as espécies arbóreas através de informações obtidas por biometria. Da mesma forma do que o experimento de Complexidade o experimento de Biometria foi dividido em dois subgrupos: Biometria das Nervuras e Biometria dos Esqueletos.

Na Biometria das Nervuras, as conclusões obtidas com os resultados da classificação dos vetores de características são:

- A. Resultados das Espécies de *Passiflora*: nesse caso os resultados podem ser avaliados entre bom e ótimo. As redes neurais classificaram corretamente uma média de 85% das espécies e, os clusters (das espécies) na grande maioria dos casos ficaram separados.

- B. Resultados das Espécies da Mata Atlântica e do Cerrado: os resultados também podem ser determinados entre bom e ótimo. As redes novamente acertaram 85% dos casos em média. O desempenho da classificação por clusters ficou um pouco abaixo dos resultados das espécies de *Passiflora*.
- C. Correlação entre as Espécies: As redes classificaram corretamente 90% das espécies, que foram reunidas em um único grupo. A classificação por análise de clusters também alcançou desempenho satisfatório. No geral, podemos avaliar os resultados entre bom e ótimo.

Avaliando os resultados alcançados com os experimentos de Biometria dos Esqueletos, podemos concluir:

- A. Resultados das Espécies de *Passiflora*: o desempenho da classificação dos vetores de características extraídos pode ser avaliado como bom. As redes classificaram em média 75% das espécies e no caso da análise dos clusters os resultados foram satisfatórios (com algumas sobreposições de clusters).
- B. Resultados das Espécies da Mata Atlântica e do Cerrado: nesse experimento os resultados não foram tão positivos sendo avaliados como regular. O resultado da análise de clusters foi razoável e, as redes neurais acertaram em média 56,5% das espécies.
- C. Correlação entre as Espécies: os resultados podem ser avaliados como regular, principalmente pelo desempenho das redes neurais (apenas 55% das espécies). Na análise de clusters muitas espécies foram separadas, porém alguns clusters ficaram parcialmente sobrepostos.

Com base nos resultados obtidos pelos experimentos que envolveram Biometria dos órgãos foliares, podemos concluir: (i) novamente a classificação baseada em informações coletadas das nervuras foi superior do que a classificação dos esqueletos e, (ii) o desempenho das redes neurais foi bem melhor do que nos experimentos com Complexidade, acompanhando os resultados obtidos com análise de clusters.

No Experimento III, todos os vetores de características coletados foram reunidos em três grupos: (i) complexidade e biometria das nervuras; (ii) complexidade e biometria dos esqueletos e (iii) todos os vetores.

Os resultados da classificação do primeiro grupo, que reuniu todos os vetores de características de todas as espécies extraídos a partir de nervura, foram altamente satisfatórios. Nesse caso podem avaliar os resultados obtidos como ótimos. Na classificação por análise de clusters a vasta maioria das espécies foi classificada corretamente (clusters separados). No caso das redes neurais o desempenho foi mais surpreendente ainda, pois 90% das espécies foram corretamente classificadas. Vale lembrar que nesse experimento todas as espécies foram reunidas em um único grupo de imagens, sendo assim, foram utilizadas 90 amostras de órgãos foliares (40 de *Passiflora* e 50 da Mata Atlântica e do Cerrado). No segundo grupo (complexidade e biometria dos esqueletos), como era de esperar os resultados foram inferiores aos do primeiro grupo. As redes classificaram corretamente 65% das vinte espécies utilizadas nos experimentos e, na análise de clusters algumas espécies não foram exatamente separadas (clusters parcialmente sobrepostos).

Por fim, o terceiro grupo de comparação reuniu todos os vetores de características (tanto de nervura quanto de esqueleto) de todas as espécies. Os resultados foram muito satisfatórios, também podendo ser avaliados como ótimo. Os resultados da classificação por análise de clusters foram excelentes, não contendo nenhum cluster sobreposto. A rede neural classificou corretamente 85% das vinte espécies utilizadas.

6.7 Considerações Finais

Nesse capítulo foram apresentados os resultados experimentais da classificação de espécies arbóreas através das características extraídas de complexidade e biometria. Para validação dos métodos desenvolvidos, foram utilizados dois conjuntos reais de imagens de espécies arbóreas: (i) imagens de espécies de *Passiflora* e (ii) imagens de espécies da Mata Atlântica e do Cerrado. Com os vetores de características extraídos com métodos computacionais, foi possível realizar vários experimentos. O Experimento I apresenta os resultados da classificação das espécies utilizando os dados referentes as medidas de complexidade. No Experimento II são discutidos os resultados da classificação das espécies através das características extraídas a partir de biometria. E por fim, no Experimento III as informações coletadas são reunidas em grupos específicos.

Conclusões

A identificação de espécies vegetais é uma tarefa muito complexa, principalmente devido a biodiversidade das espécies. Os taxonomistas, e outros profissionais de áreas correlatas, além enfrentar esse verdadeiro desafio, precisam ser altamente qualificados. Se não bastasse isso, problemas sociais como a falta de conscientização ecológica relacionada a devastação continuada dos recursos vegetais, dificultam mais ainda essa tarefa.

Nesse trabalho foram apresentados os como métodos computacionais de análise de imagens e extração de características, podem contribuir para identificação de espécies vegetais. O sistema TreeVis, associado aos bons resultados obtidos pelos métodos apresentados, pode complementar as metodologias empregadas pelos taxonomistas, facilitando e agilizando a tarefa de identificação de espécies.

O objetivo principal do trabalho foi identificar espécies vegetais através de informações (características) extraídas, essencialmente, da parte interna dos órgãos foliares. Como objetos internos de interesse foram considerados o conjunto de nervuras e o esqueleto. Para coleta das informações, as imagens dos órgãos foliares foram submetidas a dois grupos de técnicas computacionais: (a) complexidade – no qual as características foram extraídas a partir dos resultados da análise multiescala da dimensão fractal e (b) biometria – em que diversos métodos para extração de medidas biométricas foram desenvolvidos, com base nos métodos manuais utilizados por taxonomistas. Nos dois grupos de técnicas, as características foram extraídas tanto das nervuras quanto dos esqueletos dos órgãos foliares.

Na etapa seguinte do projeto, foram aplicadas técnicas de reconhecimento de padrões nas informações coletadas dos órgãos foliares, visando a classificação das espécies. Diversos experimentos foram efetuados, envolvendo as características extraídas de complexidade e biometria. Com base nos resultados alcançados, foi possível avaliar a potencialidade das características extraídas, bem como dos métodos desenvolvidos.

Analisando os resultados obtidos, podemos concluir que na grande maioria dos casos, os experimentos envolvendo nervuras alcançaram melhores resultados do que os que utilizavam características do esqueleto. Isso pode ser justificado pelo fato do esqueleto estar intimamente associado a área foliar das espécies. Dessa maneira, para as amostras utilizadas, as características essencialmente internas, como é o caso das nervuras, são potencialmente mais discriminantes do que outros tipos de características. Por outro lado, vale lembrar que os esqueletos utilizados nesse experimento foram obtidos através da esqueletonização multiescala, que gera esqueletos a partir de um determinado limiar. Com isso, experimentos utilizando outros valores de limiar poderão encontrar resultados mais ou menos significativos.

Um ponto importante deve ser destacado sobre os resultados obtidos com as espécies de maracujás silvestres. A alta taxa de acerto na identificação das espécies, encontrada através dos métodos computacionais, foi comparada a resultados obtidos por meio de reconstruções filogenéticas. Surpreendentemente, os resultados alcançados com as duas abordagens demonstraram muita semelhança (PÁDUA, ROSSIN *et al.*, 2004; PLOTZE, FALVO *et al.*, 2004). Nesse caso, a associação entre métodos computacionais e biológicos, comprovadamente, contribuiu para avanços das pesquisas nas duas áreas. Além disso, esses resultados sinalizam a viabilidade da utilização de técnicas de análise de imagens, para identificação de espécies vegetais através de características extraídas das nervuras e do esqueleto.

O desenvolvimento de projetos multidisciplinares, como nesse trabalho sobre identificação de espécies arbóreas, demanda além dos conhecimentos naturais sobre computação, o estudo e entendimento do problema biológico que está sendo abordado. Para esse projeto, foi necessário realizar uma revisão bibliográfica sobre assuntos relacionados a taxonomia de espécies vegetais, bem como estudar os métodos para classificação de órgãos foliares. Com base no material pesquisado e, em reuniões com

biólogos e botânicos, foi possível determinar quais características deveriam ser exploradas. Esse fator foi imprescindível principalmente para o desenvolvimento dos algoritmos de biometria. Se por um lado, os projetos multidisciplinares podem parecer um desafio, por outro os esforços são recompensadores, pois conseguem demonstrar a real aplicabilidade da computação.

7.1 Contribuições do Trabalho

As principais contribuições deste trabalho foram o desenvolvimento e implementação de métodos computacionais para análise, e extração de características da forma interna dos órgãos foliares. Além disso, este trabalho avaliou a potencialidade dos métodos desenvolvidos na identificação das espécies. Outras contribuições alcançadas ao longo do desenvolvimento do projeto são apresentadas a seguir:

- estudo e implementação dos métodos para estimação da dimensão fractal;
- comparação e avaliação dos métodos de dimensão fractal e sua aplicabilidade na identificação de órgãos foliares (PLOTZE, FALVO *et al.*, 2003; PLOTZE, FALVO *et al.*, 2004);
- estudo e implementação de algoritmos de esqueletonização para imagens binárias;
- análise comparativa da eficiência dos algoritmos de esqueletonização (PLOTZE e BRUNO, 2004);
- desenvolvimento e implementação de algoritmos para biometria de órgãos foliares e estruturas tubulares (PLOTZE, CARVALHO *et al.*, 2003; BRUNO, PLOTZE *et al.*, 2004);
- comparação e complementação dos resultados pesquisas envolvendo reconstrução filogenética e DNA (PLOTZE, PÁDUA *et al.*, 2004; PLOTZE, PÁDUA *et al.*, 2004).

7.2 Trabalhos Futuros

A complexidade do projeto associada ao curto espaço de tempo para desenvolvimento, abre espaço para que diversos outros trabalhos possam ser desenvolvidos:

- utilização dos métodos desenvolvidos em outras espécies vegetais;
- desenvolvimento de novos experimentos com um número maior de amostras, tanto das espécies de maracujás silvestres, quanto das espécies da Mata Atlântica e do Cerrado;
- aumento do número de vetores de características coletados através de técnicas de reamostragem como *bootstrapping*.
- desenvolvimento de um método automático para segmentação das nervuras de órgãos foliares;
- análise das nervuras dos órgãos foliares com base nas informações de textura;
- utilização de outros limiares para esqueletonizar as imagens de órgãos foliares através do algoritmo multiescala;
- classificação das informações coletadas de complexidade e biometria através de outras técnicas de reconhecimento de padrões, como: algoritmos genéticos, *support vector machines*, etc.

Anexo

A Tabela A.1 apresenta os vetores de características obtidos através da dimensão fractal multiescala, aplicada em imagens de nervuras de *Passiflora*. Os dados foram normalizados entre $[-1,1]$. Na tabela as linhas de 1-4 representam os valores obtidos para as curvas da espécie 1 (*Passiflora coriacea*); as linhas de 5-8 espécie 2 (*P. foetida*); linhas de 9-12 espécie 3 (*P. miersii*); linhas de 13-16 espécie 4 (*P. organensis*); e assim até as linhas de 36-40 da espécie 10 (*P. alata*).

TABELA A.1: Vetores de características extraídos a partir dos pontos de interesse da dimensão fractal multiescala de espécies de *Passiflora*.

#	P1x	P1y	P2x	P2y	P3x	P3y
1	-0,01512	0,56278	0,25992	0,56006	0,73689	0,38741
2	-0,02914	0,66264	0,33108	0,62648	0,86364	0,40195
3	-0,03123	0,70600	0,27678	0,55577	0,61723	0,39422
4	0,01676	0,54716	0,35992	0,48171	0,67796	0,36423
5	-0,04545	0,86364	0,78302	0,86364	0,46171	0,86002
6	0,01694	0,82593	0,79325	0,82857	0,44232	0,86364
7	0,01628	0,83600	0,70334	0,78088	0,42387	0,81213
8	0,01913	0,83503	0,71510	0,75354	0,43986	0,79543
9	0,10440	0,37349	0,55546	0,75764	0,69266	0,48480
10	0,08569	0,36795	0,51629	0,73152	0,57065	0,49773
11	0,11559	0,53908	0,44045	0,70948	0,60374	0,46794
12	0,08496	0,38113	0,45538	0,70605	0,64514	0,46787
13	0,15662	0,43538	0,43386	0,62276	0,68317	0,43631
14	0,21078	0,44376	0,41247	0,53304	0,70311	0,41079
15	0,25808	0,43263	0,34804	0,52750	0,49157	0,40872

16	0,12335	0,42332	0,31388	0,55929	0,59246	0,40524
17	0,11819	0,37952	0,83027	0,79067	0,71628	0,59340
18	0,43729	0,78791	0,73551	0,75089	0,67019	0,55273
19	0,07541	0,42002	0,65195	0,77384	0,79173	0,52122
20	0,38418	0,59600	0,86364	0,73971	0,50146	0,61506
21	-0,02884	0,73203	0,60234	0,59757	0,33521	0,52786
22	-0,01277	0,72624	0,61234	0,67528	0,50081	0,53360
23	0,08585	0,65588	0,65265	0,68961	0,36503	0,64290
24	-0,02734	0,76118	0,74285	0,67800	0,37222	0,72996
25	0,03491	0,85583	0,18246	0,27615	0,11181	0,34611
26	0,22207	0,62342	0,50016	0,27988	0,29906	0,29387
27	0,09480	0,37571	0,25436	0,20642	0,10373	0,34050
28	0,09686	0,39909	0,24787	0,15610	0,46515	0,03846
29	0,67570	0,73556	0,17157	0,14473	0,19353	0,14808
30	0,64302	0,77631	0,41043	0,23408	0,25474	0,27152
31	0,50837	0,70228	0,53752	0,23585	0,47386	0,09781
32	0,86364	0,64445	0,31931	0,03687	0,65120	-0,04545
33	0,05608	0,71277	-0,04545	0,19567	0,15274	0,28084
34	0,10907	0,44087	0,08523	0,17771	0,15038	0,27952
35	0,12385	0,63867	0,02826	0,13448	-0,01404	0,25036
36	0,12079	0,21104	0,05111	0,07178	-0,04545	0,26925
37	0,11483	0,06458	0,08908	0,21117	0,02748	0,46987
38	0,14628	-0,04545	0,35070	-0,04545	0,42082	0,18466
39	0,14192	0,10218	0,08706	0,20577	-0,02787	0,46184
40	0,11122	-0,01759	0,09665	0,12425	0,05641	0,42758

Na Tabela A.2 são apresentados os dados relativos a dimensão fractal multiescala aplicada em imagens de nervuras, de espécies da Mata Atlântica e do Cerrado. esses dados estão normalizados entre $[-1, 1]$. As linhas de 1 a 5 representam os dados obtidos da espécie 1; as linhas de 6-10 da espécie 2; as linhas 11-15 da espécie 3; e assim até as linhas de 46-50 da espécie 10.

TABELA A.2: Vetores de características extraídos a partir dos pontos de interesse da dimensão fractal multiescala de espécies da Mata Atlântica e do Cerrado brasileiro.

#	P1x	P1y	P2x	P2y	P3x	P3y
1	-0,04545	0,74945	0,86364	0,86364	0,67794	0,78056
2	0,05995	0,74050	0,14656	0,42025	0,86364	0,29600
3	0,01512	0,68540	0,29801	0,46457	0,81715	0,33661
4	0,01531	0,72104	0,26213	0,45626	0,68148	0,34937
5	-0,04290	0,53564	0,32264	0,51586	0,67208	0,41147

6	0,08272	0,84312	0,39084	0,72432	0,42522	0,83462
7	0,77778	0,86364	-0,04545	0,32476	0,59587	0,34036
8	0,32749	0,83070	-0,02733	0,37038	0,59933	0,32818
9	0,48552	0,85630	0,45470	0,55826	0,58213	0,42567
10	0,10518	0,85191	0,04757	0,45139	0,80906	0,35626
11	0,16904	0,12467	0,70621	0,61837	0,31229	0,70661
12	0,07413	0,05560	0,48910	0,60012	0,14578	0,86364
13	0,09907	0,06531	0,51516	0,48013	0,32558	0,51512
14	0,26751	0,04037	0,50125	0,37551	0,39237	0,38138
15	0,18399	0,03033	0,37715	0,39929	0,31354	0,47404
16	0,00905	0,07300	0,67925	0,58860	0,36157	0,65492
17	0,27314	0,23934	0,27337	0,37289	0,63490	0,32018
18	0,27333	0,19489	-0,03375	0,19913	0,35834	0,27902
19	0,28178	0,20220	0,08568	0,24322	0,52350	0,28515
20	0,25080	0,26760	0,39498	0,47308	0,50854	0,42945
21	0,52595	0,73822	0,44868	0,51435	0,59265	0,48452
22	0,73364	0,69724	0,36376	0,49545	0,56589	0,49577
23	0,43420	0,59252	0,47439	0,54980	0,61895	0,44477
24	0,23775	0,57431	0,27402	0,48357	0,51954	0,42801
25	0,70681	0,75807	0,60262	0,59217	0,35686	0,60169
26	0,25069	0,23425	0,08734	0,27212	0,41292	0,25368
27	0,28589	0,22030	-0,00136	0,17848	0,57631	0,24986
28	0,21331	0,22957	0,05025	0,20880	0,65468	0,22921
29	0,28649	0,27465	0,19976	0,26735	0,66750	0,19461
30	0,26048	0,28837	0,05450	0,25752	0,45326	0,22550
31	0,28212	0,07520	0,00502	0,10607	0,36771	0,15046
32	0,38444	0,04350	0,06625	0,08867	0,44047	0,10638
33	0,49173	0,01673	0,08856	0,09097	0,21324	0,18686
34	0,47763	0,09193	0,19777	0,19979	0,41823	0,19580
35	0,49184	0,02703	0,19485	0,13990	0,22655	0,16809
36	0,65130	-0,04545	0,22300	-0,04508	0,14440	-0,02509
37	0,47342	0,05543	0,28854	0,19368	0,11108	0,29888
38	0,82503	0,08182	0,30534	0,10849	-0,04545	0,21580
39	0,62410	0,00754	0,37977	0,08201	0,19962	0,11886
40	0,55852	0,14249	0,32032	0,22230	0,19100	0,25156
41	0,45366	0,10542	0,51019	0,35476	0,25508	0,39248
42	0,53562	-0,04391	0,11811	-0,04545	0,22454	-0,04545
43	0,78000	0,16143	0,47479	0,26844	0,11234	0,34616
44	0,85912	0,12228	0,36784	0,11845	0,09851	0,18348
45	0,86364	0,00720	0,36233	0,08833	0,16615	0,10566
46	0,83798	0,03243	0,60354	0,25591	0,01498	0,51411
47	0,76837	0,23651	0,41600	0,28734	0,07983	0,39936

48	0,82688	0,24206	0,43205	0,20812	0,10770	0,29293
49	0,81384	0,21297	0,54577	0,25913	0,09976	0,33524
50	0,81989	0,19398	0,43975	0,21930	0,17754	0,21721

A Tabela A.3 apresenta os vetores de características referentes extraídas a partir da dimensão fractal multiescala em imagens esqueletonizadas de espécies de *Passiflora*. Os dados estão organizados por espécies como na Tabela A.1

TABELA A.3: Vetores de características extraídos a partir dos pontos de interesse da dimensão fractal multiescala de imagens esqueletonizadas das espécies de *Passiflora*.

#	P1x	P1y	P2x	P2y	P3x	P3y
1	-0,04205	0,84814	0,08484	0,86128	0,86364	0,53834
2	-0,03829	0,86364	0,12222	0,80729	0,71047	0,57823
3	-0,03905	0,82968	0,09487	0,71973	0,53253	0,54843
4	-0,03700	0,74799	-0,04545	0,69818	0,19590	0,54825
5	0,75851	0,14585	0,35322	0,11389	-0,00507	0,71338
6	0,71414	0,41470	0,26582	0,35253	0,07279	0,86364
7	0,73432	0,28360	0,26868	0,25412	-0,04545	0,79655
8	0,70221	0,19887	0,28441	0,24721	-0,01857	0,80000
9	-0,02824	0,18253	0,12894	0,59209	0,40660	0,46848
10	-0,03432	0,34717	0,18295	0,72263	0,53785	0,48801
11	-0,03448	0,10093	0,12069	0,59610	0,34160	0,50868
12	-0,03829	0,38922	0,09395	0,78125	0,48076	0,56819
13	-0,04545	0,61080	0,04599	0,75659	0,48279	0,59966
14	-0,03858	0,65328	0,13143	0,84046	0,53861	0,63711
15	-0,03729	0,41570	0,31435	0,83829	0,65470	0,60770
16	-0,03284	0,71141	0,20208	0,86364	0,65909	0,60449
17	0,03370	0,36886	0,42400	0,61664	0,26477	0,22886
18	0,04169	0,30672	0,33944	0,61041	0,43120	0,49191
19	0,04442	-0,04545	0,86364	0,37891	0,40523	-0,04545
20	0,10175	0,13419	0,53985	0,20943	0,12578	0,06367
21	0,09918	0,71146	0,34732	0,34659	0,20740	0,05221
22	-0,02384	0,75530	0,18618	0,43670	0,61585	-0,02009
23	0,03676	0,75570	0,31075	0,37896	0,55756	0,00026
24	-0,01907	0,79344	0,30238	0,39452	0,44827	0,09216
25	-0,01410	0,66447	0,19006	0,38011	0,45285	0,06001
26	-0,02808	0,79020	0,21344	0,36735	0,25902	0,25712
27	0,86364	0,69728	0,24286	0,27037	0,25130	0,09951
28	-0,03566	0,76857	0,18012	0,36677	0,36655	0,12026
29	0,63075	0,59480	0,14680	0,17895	0,25627	0,04070

30	0,38449	0,64899	0,47903	0,18307	0,21234	0,11261
31	0,65359	0,53790	0,42943	-0,04545	0,24412	-0,01030
32	0,55088	0,47924	0,59034	0,06097	0,27761	0,08619
33	-0,03526	0,83830	0,16815	0,44126	0,46198	0,11054
34	-0,04053	0,83702	0,14897	0,44838	0,47852	0,09538
35	-0,03727	0,83099	0,16014	0,38168	0,37123	0,16315
36	-0,03146	0,83043	0,23231	0,41118	0,60912	0,00693
37	-0,04472	0,29589	0,04045	0,77107	0,62456	0,63137
38	-0,04145	0,49762	0,10224	0,73883	0,61927	0,47790
39	-0,04206	0,39092	-0,00834	0,84149	0,66688	0,59828
40	-0,04092	0,30527	0,04588	0,70880	0,61070	0,49858

A Tabela A.4 apresenta os pontos de interesse extraídos a partir da dimensão fractal multiescala em imagens esquelotonizadas de órgãos foliares da Mata Atlântica e do Cerrado. A organização das espécies na Tabela é idêntica a Tabela A.2.

TABELA A.4: Vetores de características extraídos através da dimensão fractal multiescala aplicada em imagens esquelotonizadas das espécies da Mata Atlântica e do Cerrado

#	P1x	P1y	P2x	P2y	P3x	P3y
1	0,13313	0,88616	0,40651	0,57432	0,96024	0,16661
2	0,19621	0,84698	0,29401	0,42940	1,00000	0,10517
3	0,16566	0,85549	0,40555	0,49711	0,84709	0,18266
4	0,16897	0,84841	0,37732	0,48468	0,77687	0,19557
5	0,12568	0,87862	0,46597	0,58747	0,79355	0,14760
6	0,11802	0,80634	0,57918	0,64328	0,75208	0,14576
7	0,24418	0,76827	0,04727	0,27109	0,81530	0,11624
8	0,21767	0,73277	0,06376	0,22055	0,67331	0,08487
9	0,11787	0,87102	0,33577	0,55468	0,88528	0,18266
10	0,19580	0,81289	0,15308	0,36313	0,76865	0,15314
11	0,02988	1,00000	1,00000	1,00000	0,75934	0,34133
12	0,00000	0,96005	0,38573	0,81623	0,53901	0,34133
13	0,07594	0,86372	0,51020	0,68174	0,69639	0,22878
14	0,08290	0,70087	0,14295	0,45062	0,46566	0,22878
15	0,24404	0,58759	0,55400	0,50996	0,57294	0,17343
16	0,09593	0,91353	0,61469	0,71196	0,80156	0,22325
17	0,17916	0,81762	0,35860	0,44335	0,74427	0,15314
18	0,22448	0,73101	0,07199	0,26748	0,68091	0,11255
19	0,22013	0,79089	0,18324	0,31250	0,67852	0,11439
20	0,12538	0,86453	0,48683	0,59008	0,85585	0,18450
21	0,11204	0,90332	0,56938	0,67450	0,81039	0,22878
22	0,10785	0,90430	0,51301	0,64990	0,84896	0,23063

23	0,10075	0,90538	0,62287	0,70892	0,84992	0,21956
24	0,08096	0,85614	0,39695	0,61237	0,68429	0,22878
25	0,06364	0,88768	0,73240	0,78776	0,46699	0,33026
26	0,18222	0,79628	0,34425	0,42735	0,80828	0,14945
27	0,22852	0,82648	0,09965	0,32733	0,96904	0,12177
28	0,21937	0,81588	0,13325	0,33659	0,93193	0,11808
29	0,20158	0,82150	0,18345	0,38100	0,98854	0,10886
30	0,18894	0,83395	0,18926	0,39184	0,85087	0,10886
31	0,22714	0,78759	0,14262	0,27883	0,64340	0,13838
32	0,21168	0,76312	0,05716	0,24323	0,62376	0,14391
33	0,25484	0,63983	0,00000	0,25438	0,49959	0,14576
34	0,21294	0,69097	0,10510	0,28963	0,90496	0,09410
35	0,31075	0,53785	0,12760	0,25929	0,33396	0,12177
36	0,41909	0,09934	0,12020	0,07028	0,21204	0,11808
37	0,22838	0,32811	0,33800	0,37908	0,33000	0,23616
38	0,22355	0,02851	0,16607	0,21243	0,00000	0,24354
39	1,00000	0,00000	0,41298	0,18048	0,25662	0,00000
40	0,28464	0,19124	0,23186	0,33100	0,26108	0,23063
41	0,23132	0,65605	0,58493	0,53995	0,47582	0,23432
42	0,93468	0,19544	0,11287	0,00000	0,32547	0,07565
43	0,49970	0,48051	0,20166	0,27900	0,26589	0,16605
44	0,56513	0,19663	0,29126	0,17597	0,47814	0,14760
45	0,43445	0,10211	0,44060	0,22886	0,26589	0,16790
46	0,49771	0,12770	0,64117	0,39282	0,09869	0,22694
47	0,08463	0,59396	0,52385	0,59781	0,36861	0,29151
48	0,23790	0,18586	0,46952	0,34575	0,03096	0,35609
49	0,10821	0,58370	0,61323	0,53487	0,27738	1,00000
50	0,12523	0,54187	0,25148	0,38223	0,40339	0,20111

As próximas tabelas apresentam os vetores de características coletados através de biometria dos órgãos foliares, tanto em imagens de nervuras quanto de esqueleto. A Tabela A.5 apresenta os dados referentes as nervuras das espécies de *Passiflora*. Na Tabela A.6 são ilustrados os dados relativos as nervuras das espécies Mata Atlântica e do Cerrado. A Tabela A.7 apresenta os dados coletados dos esqueletos das espécies de *Passiflora* e, na Tabela A.8 os dados das espécies da Mata Atlântica e do Cerrado. Todos os vetores estão normalizados entre [-1,1]

TABELA A.5: Vetores de características extraídos através de Biometria das nervuras dos órgãos foliares de *Passiflora*

	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19
1	0,056	0,076	0,060	0,096	0,000	0,014	0,000	0,301	0,914	0,076	0,330	0,019	0,037	0,512	0,450	0,143	0,000	0,551	0,680
2	0,111	0,034	0,025	0,024	0,015	0,000	0,000	0,247	0,883	0,000	0,272	0,000	0,037	0,462	0,429	0,286	0,034	0,636	0,460
3	0,111	0,039	0,026	0,024	0,169	0,217	0,000	0,339	0,713	0,167	0,361	0,065	0,037	0,630	0,548	0,286	0,165	0,857	0,094
4	0,056	0,111	0,085	0,085	0,082	0,198	0,000	0,527	1,000	0,478	0,580	0,060	0,000	0,853	0,822	0,071	0,080	0,800	1,000
5	1,000	0,618	0,764	0,846	0,794	0,628	1,000	0,482	0,000	0,379	0,308	0,885	0,852	0,153	0,050	0,714	0,785	0,803	0,439
6	0,778	0,497	0,619	0,722	0,814	0,667	0,875	0,299	0,091	0,142	0,230	0,646	0,889	0,000	0,166	0,571	0,809	0,493	0,483
7	0,944	0,710	0,770	0,820	0,859	0,657	0,938	0,264	0,125	0,453	0,170	0,776	1,000	0,011	0,072	0,786	0,852	0,377	0,476
8	0,778	0,767	0,857	0,869	0,911	0,686	0,750	0,431	0,094	0,777	0,264	0,889	0,778	0,210	0,337	0,786	0,915	0,719	0,366
9	0,389	0,726	0,831	0,842	1,000	0,884	0,313	0,166	0,388	0,507	0,641	0,251	0,370	0,111	0,123	0,500	0,995	0,621	0,511
10	0,278	0,854	0,893	0,898	0,983	0,667	0,188	0,300	0,567	0,855	0,637	0,225	0,259	0,244	0,239	0,500	0,969	0,655	0,526
11	0,222	0,658	0,750	0,766	1,000	0,884	0,125	0,223	0,531	0,860	0,730	0,247	0,222	0,349	0,484	0,429	0,981	0,409	0,489
12	0,278	0,979	1,000	1,000	1,000	0,749	0,188	0,226	0,671	0,810	0,761	0,126	0,259	0,115	0,000	0,500	1,000	0,582	0,635
13	0,556	0,585	0,564	0,807	0,186	0,145	0,500	0,417	0,420	0,303	0,052	0,401	0,519	0,093	0,298	0,286	0,189	0,727	0,078
14	0,611	0,601	0,569	0,786	0,233	0,473	0,563	0,301	0,425	0,288	0,148	0,416	0,593	0,042	0,157	0,286	0,237	0,746	0,097
15	0,556	0,595	0,568	0,795	0,335	0,473	0,563	0,295	0,152	0,261	0,067	0,394	0,481	0,123	0,419	0,429	0,332	0,674	0,412
16	0,389	0,516	0,601	0,798	0,176	0,425	0,375	0,422	0,270	0,368	0,088	0,402	0,444	0,171	0,347	0,286	0,182	0,585	0,422
17	0,833	0,634	0,608	0,768	0,318	0,580	0,813	0,362	0,385	0,435	0,260	0,757	0,704	0,188	0,354	0,500	0,312	0,642	0,477
18	0,778	0,672	0,590	0,757	0,390	0,705	0,750	0,363	0,439	0,593	0,018	0,783	0,815	0,120	0,277	0,500	0,383	0,508	0,512
19	0,944	0,882	0,712	0,876	0,732	0,870	1,000	0,402	0,489	0,750	0,516	1,000	0,852	0,217	0,058	0,571	0,717	0,466	0,450
20	0,833	0,929	0,736	0,927	0,707	0,952	0,875	0,455	0,400	0,830	0,631	0,898	0,889	0,135	0,133	0,429	0,709	0,685	0,396
21	0,556	0,743	0,742	0,775	0,866	1,000	0,625	0,208	0,507	0,495	0,298	0,588	0,556	0,206	0,331	0,857	0,862	0,479	0,471
22	0,667	0,544	0,528	0,568	0,561	0,546	0,625	0,000	0,429	0,113	0,000	0,393	0,556	0,053	0,206	0,857	0,562	0,650	0,375
23	0,611	0,683	0,716	0,737	0,782	0,744	0,625	0,148	0,398	0,349	0,248	0,478	0,556	0,120	0,079	0,714	0,763	0,613	0,358
24	0,556	0,594	0,735	0,710	0,809	0,816	0,563	0,124	0,391	0,454	0,278	0,491	0,556	0,131	0,245	0,714	0,811	0,458	0,412
25	0,278	0,441	0,637	0,591	0,628	0,652	0,188	0,264	0,473	0,429	0,371	0,246	0,259	0,271	0,477	0,429	0,613	0,497	0,629
26	0,278	0,454	0,687	0,641	0,655	0,676	0,250	0,226	0,506	0,474	0,290	0,323	0,259	0,372	0,469	0,429	0,642	0,511	0,421
27	0,222	0,297	0,534	0,486	0,777	0,715	0,125	0,483	0,475	0,691	0,556	0,347	0,296	0,330	0,574	0,143	0,751	0,639	0,021
28	0,222	0,404	0,562	0,543	0,653	0,734	0,250	0,221	0,533	0,440	0,281	0,352	0,296	0,336	0,430	0,500	0,661	0,347	0,459
29	0,056	0,000	0,000	0,000	0,588	0,382	0,000	0,786	0,582	0,564	0,581	0,189	0,111	0,577	0,790	0,214	0,538	0,268	0,449
30	0,167	0,072	0,126	0,122	0,648	0,546	0,000	0,840	0,497	0,670	0,677	0,225	0,185	0,405	0,858	0,000	0,598	0,000	0,000
31	0,056	0,000	0,002	0,002	0,846	0,700	0,000	1,000	0,511	1,000	1,000	0,133	0,074	0,613	1,000	0,143	0,823	0,449	0,527
32	0,056	0,003	0,002	0,002	0,692	0,512	0,000	0,962	0,671	0,891	0,740	0,306	0,074	1,000	1,000	0,286	0,879	0,160	0,239
33	0,111	0,019	0,013	0,012	0,546	0,271	0,000	0,525	0,861	0,347	0,616	0,107	0,037	0,742	0,652	0,286	0,542	0,653	0,459
34	0,000	0,033	0,045	0,066	0,519	0,237	0,000	0,483	0,898	0,294	0,656	0,090	0,037	0,699	0,679	0,143	0,511	1,000	0,354
35	0,111	0,013	0,006	0,020	0,596	0,430	0,000	0,643	0,540	0,522	0,611	0,155	0,148	0,391	0,678	0,143	0,591	0,425	0,498
36	0,056	0,020	0,016	0,023	0,620	0,343	0,000	0,624	0,749	0,502	0,672	0,148	0,074	0,648	0,712	0,143	0,608	0,864	0,074
37	0,556	0,944	0,873	0,833	0,931	0,797	0,500	0,104	0,282	0,702	0,414	0,393	0,481	0,122	0,020	0,714	0,903	0,606	0,492
38	0,722	1,000	0,864	0,803	0,896	0,618	0,688	0,179	0,231	0,700	0,275	0,493	0,630	0,069	0,169	1,000	0,874	0,419	0,476
39	0,500	0,771	0,818	0,787	0,876	0,729	0,438	0,011	0,307	0,377	0,484	0,247	0,407	0,064	0,043	0,714	0,867	0,764	0,368
40	0,500	0,966	0,940	0,905	0,973	0,932	0,438	0,188	0,192	0,961	0,477	0,512	0,333	0,449	0,233	0,786	0,978	0,716	0,457

TABELA A.6: Vetores de características extraídos através de Biometria das nervuras dos órgãos foliares da Mata Atlântica e do Cerrado																			
	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19
1	0,000	0,000	0,000	0,000	0,374	0,510	0,000	0,833	0,851	0,778	0,833	0,000	0,000	0,822	0,835	0,000	0,000	0,000	0,000
2	0,000	0,000	0,000	0,000	0,646	0,647	0,000	1,000	1,000	1,000	1,000	0,042	0,000	1,000	1,000	0,000	0,000	0,000	0,000
3	0,000	0,000	0,000	0,000	0,535	0,529	0,000	0,932	0,937	0,910	0,932	0,037	0,000	0,978	0,980	0,000	0,000	0,000	0,000
4	0,000	0,000	0,000	0,000	0,495	0,490	0,000	0,907	0,914	0,877	0,907	0,035	0,000	0,973	0,975	0,000	0,000	0,000	0,000
5	0,000	0,000	0,000	0,000	0,384	0,451	0,000	0,846	0,858	0,786	0,839	0,021	0,000	0,913	0,920	0,000	0,000	0,000	0,000
6	0,167	0,387	0,326	0,329	0,162	0,196	0,160	0,102	0,096	0,060	0,097	0,115	0,222	0,102	0,121	0,160	0,654	0,657	0,146
7	0,167	0,458	0,411	0,454	0,788	0,843	0,160	0,283	0,157	0,208	0,233	0,222	0,222	0,179	0,175	0,160	0,906	0,644	0,152
8	0,167	0,427	0,381	0,417	0,697	0,686	0,160	0,252	0,144	0,147	0,212	0,186	0,259	0,112	0,179	0,160	0,859	0,431	0,408
9	0,167	0,376	0,348	0,389	0,364	0,294	0,160	0,167	0,112	0,072	0,145	0,164	0,222	0,137	0,125	0,160	0,726	0,664	0,122
10	0,167	0,449	0,389	0,408	0,606	0,627	0,160	0,222	0,141	0,172	0,191	0,203	0,222	0,165	0,163	0,160	0,838	0,702	0,171
11	0,667	0,554	0,465	0,461	0,000	0,039	0,640	0,000	0,005	0,022	0,000	0,398	0,704	0,000	0,000	0,520	0,560	0,635	0,170
12	0,750	0,509	0,457	0,448	0,030	0,000	0,720	0,007	0,000	0,000	0,001	0,542	0,778	0,016	0,007	0,720	0,581	0,681	0,414
13	0,792	0,631	0,525	0,513	0,273	0,196	0,760	0,045	0,023	0,131	0,038	0,695	0,815	0,039	0,022	0,760	0,679	0,655	0,473
14	0,917	0,699	0,584	0,567	0,525	0,392	0,880	0,073	0,041	0,210	0,059	0,966	0,963	0,056	0,026	0,880	0,791	0,766	0,723
15	1,000	0,680	0,675	0,654	0,273	0,314	1,000	0,092	0,021	0,178	0,080	0,941	1,000	0,044	0,026	1,000	0,684	0,646	0,528
16	0,083	0,098	0,119	0,159	0,111	0,157	0,080	0,311	0,089	0,504	0,274	0,326	0,148	0,451	0,332	0,040	0,615	0,379	0,679
17	0,083	0,067	0,076	0,091	0,535	0,471	0,080	0,494	0,150	0,836	0,418	0,505	0,148	0,643	0,484	0,080	0,795	0,284	0,424
18	0,083	0,053	0,067	0,091	0,657	0,608	0,080	0,550	0,168	0,948	0,466	0,597	0,148	0,742	0,562	0,080	0,842	0,556	0,204
19	0,083	0,107	0,125	0,159	0,545	0,588	0,080	0,517	0,156	0,899	0,444	0,544	0,185	0,521	0,491	0,080	0,799	1,000	1,000
20	0,083	0,053	0,067	0,091	0,323	0,275	0,080	0,420	0,118	0,682	0,351	0,446	0,185	0,436	0,435	0,080	0,705	0,250	0,385
21	0,292	0,587	0,562	0,603	0,232	0,137	0,280	0,060	0,081	0,084	0,084	0,125	0,333	0,026	0,018	0,200	0,662	0,654	0,157
22	0,208	0,649	0,572	0,625	0,343	0,118	0,200	0,089	0,121	0,172	0,124	0,103	0,259	0,060	0,060	0,200	0,718	0,496	0,435
23	0,292	0,597	0,586	0,629	0,253	0,118	0,280	0,089	0,060	0,102	0,096	0,159	0,333	0,042	0,025	0,280	0,671	0,679	0,151
24	0,333	0,591	0,543	0,576	0,303	0,176	0,320	0,067	0,071	0,121	0,079	0,243	0,370	0,060	0,029	0,320	0,697	0,588	0,114
25	0,417	0,590	0,503	0,517	0,162	0,078	0,400	0,037	0,034	0,057	0,040	0,227	0,481	0,009	0,015	0,360	0,637	0,744	0,220
26	0,458	0,734	0,581	0,588	0,626	0,529	0,480	0,095	0,081	0,263	0,103	0,442	0,519	0,064	0,025	0,480	0,880	0,618	0,623
27	0,417	1,000	0,804	0,821	0,798	0,745	0,400	0,151	0,127	0,450	0,174	0,320	0,481	0,039	0,030	0,320	0,889	0,746	0,169
28	0,583	0,871	0,703	0,701	0,768	0,608	0,560	0,108	0,099	0,318	0,128	0,422	0,667	0,014	0,025	0,520	0,897	0,744	0,098
29	0,667	0,816	0,661	0,650	0,768	0,667	0,760	0,092	0,069	0,307	0,100	0,674	0,815	0,035	0,011	0,760	0,906	0,633	0,567
30	0,625	0,844	0,678	0,677	0,788	0,627	0,640	0,102	0,092	0,271	0,116	0,486	0,778	0,005	0,037	0,640	0,962	0,645	0,532
31	0,583	0,845	0,891	0,895	0,838	0,902	0,560	0,191	0,095	0,454	0,187	0,884	0,630	0,139	0,087	0,560	0,944	0,711	0,446
32	0,583	0,998	1,000	1,000	0,879	1,000	0,600	0,223	0,081	0,538	0,205	0,908	0,630	0,145	0,087	0,600	0,953	0,611	0,421
33	0,625	0,929	0,907	0,906	0,758	0,882	0,640	0,230	0,055	0,463	0,203	1,000	0,667	0,151	0,099	0,640	0,889	0,721	0,431
34	0,542	0,769	0,837	0,845	0,818	0,863	0,520	0,202	0,096	0,352	0,195	0,734	0,630	0,102	0,087	0,520	0,953	0,854	0,483
35	0,625	0,906	0,859	0,857	0,869	0,961	0,640	0,184	0,070	0,430	0,165	0,883	0,667	0,123	0,054	0,640	0,932	0,724	0,393
36	0,417	0,886	0,901	0,932	0,909	0,784	0,400	0,263	0,067	0,537	0,181	0,712	0,481	0,169	0,128	0,400	0,957	0,684	0,411
37	0,458	0,587	0,634	0,646	0,485	0,412	0,440	0,165	0,028	0,210	0,114	0,673	0,481	0,155	0,065	0,440	0,774	0,673	0,384
38	0,458	0,761	0,732	0,752	0,687	0,627	0,440	0,207	0,043	0,409	0,136	0,827	0,481	0,207	0,078	0,440	0,863	0,661	0,344
39	0,208	0,696	0,783	0,890	0,667	0,588	0,360	0,201	0,069	0,409	0,145	0,514	0,333	0,213	0,102	0,280	0,846	0,558	0,589
40	0,583	0,669	0,668	0,670	0,535	0,431	0,560	0,149	0,024	0,236	0,095	0,775	0,630	0,112	0,061	0,560	0,799	0,828	0,345
41	0,417	0,609	0,621	0,643	0,455	0,275	0,400	0,122	0,051	0,187	0,094	0,451	0,444	0,101	0,074	0,400	0,765	0,651	0,125
42	0,458	0,815	0,877	0,899	1,000	0,843	0,440	0,223	0,104	0,493	0,181	0,909	0,519	0,207	0,134	0,440	1,000	0,664	0,383
43	0,542	0,673	0,734	0,739	0,707	0,549	0,520	0,183	0,057	0,314	0,131	0,912	0,556	0,184	0,084	0,440	0,868	0,650	0,417
44	0,333	0,633	0,736	0,779	0,798	0,667	0,320	0,275	0,052	0,437	0,191	0,778	0,407	0,251	0,166	0,320	0,927	0,646	0,454
45	0,625	0,794	0,843	0,840	0,778	0,627	0,640	0,178	0,067	0,349	0,131	0,951	0,667	0,140	0,060	0,640	0,893	0,656	0,460
46	0,500	0,671	0,634	0,644	0,465	0,275	0,520	0,126	0,039	0,219	0,080	0,699	0,556	0,124	0,064	0,480	0,765	0,723	0,316
47	0,417	0,671	0,646	0,668	0,475	0,333	0,400	0,112	0,056	0,220	0,094	0,417	0,481	0,071	0,077	0,400	0,791	0,816	0,579
48	0,500	0,685	0,691	0,703	0,586	0,353	0,480	0,154	0,043	0,280	0,096	0,736	0,519	0,154	0,075	0,480	0,825	0,676	0,378
49	0,458	0,794	0,747	0,760	0,535	0,353	0,480	0,138	0,045	0,335	0,087	0,631	0,556	0,105	0,094	0,480	0,799	0,684	0,407
50	0,417	0,751	0,720	0,744	0,697	0,431	0,400	0,173	0,069	0,388	0,116	0,592	0,444	0,151	0,112	0,400	0,863	0,677	0,218

TABELA A.7: Vetores de características extraídos através de Biometria dos esqueletos dos órgãos foliares de *Passiflora*

	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19
1	0,000	0,000	0,000	0,000	0,000	0,765	0,000	0,816	0,852	0,853	0,874	0,151	0,083	0,718	0,714	0,000	0,000	0,000	0,000
2	0,143	0,000	0,000	0,000	0,136	0,662	0,143	0,394	0,191	0,264	0,294	0,330	0,250	0,511	0,382	0,667	0,136	0,394	0,383
3	0,143	0,000	0,000	0,000	0,225	0,412	0,143	0,375	0,177	0,331	0,283	0,355	0,250	0,551	0,372	0,667	0,225	0,418	0,394
4	0,000	0,000	0,000	0,000	0,000	0,544	0,000	0,988	0,984	1,000	0,992	0,211	0,083	1,000	1,000	0,000	0,000	0,000	0,000
5	1,000	0,873	0,548	0,688	0,607	0,456	1,000	0,000	0,000	0,040	0,000	0,604	1,000	0,223	0,213	1,000	0,607	0,313	0,416
6	0,429	0,537	0,493	0,514	0,675	0,500	0,429	0,113	0,092	0,138	0,075	0,519	0,417	0,478	0,273	0,667	0,675	0,460	0,442
7	0,429	0,543	0,497	0,516	0,681	0,676	0,429	0,160	0,117	0,153	0,072	0,520	0,417	0,479	0,299	0,667	0,681	0,430	0,411
8	0,571	0,616	0,472	0,575	0,702	0,544	0,571	0,147	0,063	0,097	0,067	0,550	0,500	0,420	0,268	1,000	0,702	0,511	0,362
9	0,286	0,130	0,159	0,130	0,754	0,588	0,286	0,170	0,161	0,036	0,232	0,336	0,500	0,253	0,320	0,667	0,754	0,503	0,465
10	0,143	0,000	0,000	0,000	0,743	0,912	0,143	0,457	0,355	0,116	0,351	0,274	0,333	0,314	0,374	0,333	0,743	1,000	1,000
11	0,143	0,000	0,000	0,000	0,770	0,382	0,143	0,256	0,276	0,137	0,369	0,281	0,333	0,322	0,388	0,667	0,770	0,488	0,466
12	0,143	0,000	0,000	0,000	0,832	0,765	0,143	0,472	0,355	0,060	0,460	0,242	0,250	0,372	0,495	0,333	0,832	0,958	0,958
13	0,000	0,000	0,000	0,000	0,000	1,000	0,000	1,000	1,000	0,841	0,868	0,169	0,083	0,801	0,799	0,000	0,000	0,000	0,000
14	0,000	0,000	0,000	0,000	0,000	0,265	0,000	0,643	0,712	0,776	0,815	0,308	0,167	0,724	0,513	0,000	0,000	0,000	0,000
15	0,000	0,000	0,000	0,000	0,000	0,412	0,000	0,463	0,573	0,552	0,636	0,112	0,083	0,532	0,526	0,000	0,000	0,000	0,000
16	0,000	0,000	0,000	0,000	0,000	0,191	0,000	0,632	0,695	0,811	0,843	0,145	0,083	0,686	0,682	0,000	0,000	0,000	0,000
17	0,143	0,704	0,704	0,704	0,000	0,206	0,143	0,350	0,178	0,317	0,289	0,372	0,250	0,577	0,534	0,000	0,000	0,000	0,000
18	0,143	0,648	0,648	0,648	0,000	0,265	0,143	0,290	0,211	0,320	0,295	0,378	0,250	0,588	0,549	0,000	0,000	0,000	0,000
19	0,429	0,370	0,214	0,140	0,105	0,294	0,429	0,165	0,077	0,000	0,345	0,493	0,667	0,279	0,391	0,333	0,110	0,332	0,332
20	0,143	1,000	1,000	1,000	0,283	0,000	0,143	0,446	0,331	0,614	0,685	0,507	0,250	0,791	0,557	0,667	0,288	0,374	0,366
21	0,571	0,398	0,335	0,374	0,764	0,015	0,571	0,035	0,014	0,100	0,178	0,568	0,750	0,285	0,296	0,667	0,764	0,450	0,428
22	0,143	0,000	0,000	0,000	0,576	0,162	0,143	0,392	0,198	0,514	0,280	0,453	0,250	0,705	0,403	0,667	0,576	0,440	0,420
23	0,143	0,000	0,000	0,000	0,717	0,147	0,143	0,415	0,231	0,591	0,327	0,493	0,250	0,769	0,446	0,667	0,717	0,441	0,400
24	0,286	0,120	0,147	0,120	0,738	0,029	0,286	0,226	0,091	0,273	0,265	0,520	0,333	0,604	0,454	0,667	0,738	0,438	0,422
25	0,143	0,000	0,000	0,000	0,576	0,162	0,143	0,396	0,204	0,529	0,287	0,505	0,250	0,788	0,457	0,667	0,576	0,414	0,398
26	0,143	0,000	0,000	0,000	0,592	0,074	0,143	0,405	0,200	0,541	0,294	0,505	0,250	0,788	0,458	0,667	0,592	0,430	0,415
27	0,143	0,000	0,000	0,000	0,717	0,015	0,143	0,436	0,238	0,608	0,341	0,572	0,250	0,893	0,526	0,667	0,717	0,418	0,394
28	0,143	0,000	0,000	0,000	0,592	0,059	0,143	0,338	0,176	0,535	0,290	0,503	0,250	0,784	0,453	0,667	0,592	0,432	0,413
29	0,429	0,173	0,159	0,173	0,775	0,147	0,429	0,189	0,082	0,365	0,117	0,742	0,417	0,690	0,337	0,667	0,743	0,396	0,386
30	0,429	0,278	0,255	0,278	0,817	0,162	0,429	0,185	0,113	0,396	0,137	0,789	0,583	0,521	0,372	0,667	0,801	0,374	0,366
31	0,429	0,228	0,209	0,228	1,000	0,118	0,429	0,292	0,150	0,551	0,222	1,000	0,500	0,777	0,527	0,667	1,000	0,372	0,368
32	0,429	0,142	0,146	0,142	0,859	0,132	0,429	0,302	0,088	0,504	0,121	0,939	0,417	0,877	0,429	0,667	0,843	0,475	0,468
33	0,143	0,000	0,000	0,000	0,639	0,574	0,143	0,626	0,286	0,567	0,306	0,477	0,250	0,744	0,426	0,667	0,639	0,434	0,410
34	0,143	0,000	0,000	0,000	0,602	0,662	0,143	0,611	0,300	0,552	0,298	0,491	0,250	0,765	0,440	0,667	0,602	0,432	0,412
35	0,143	0,000	0,000	0,000	0,654	0,412	0,143	0,530	0,281	0,591	0,317	0,518	0,250	0,808	0,468	0,667	0,654	0,431	0,413
36	0,143	0,000	0,000	0,000	0,649	0,412	0,143	0,584	0,263	0,579	0,312	0,507	0,333	0,590	0,448	0,333	0,649	0,800	0,800
37	0,143	0,000	0,000	0,000	0,791	0,515	0,143	0,319	0,280	0,113	0,397	0,272	0,250	0,419	0,446	0,667	0,791	0,512	0,484
38	0,143	0,000	0,000	0,000	0,754	0,956	0,143	0,417	0,347	0,240	0,328	0,338	0,333	0,389	0,377	0,667	0,754	0,523	0,477
39	0,000	0,000	0,000	0,000	0,921	0,118	0,000	0,666	0,831	0,582	1,000	0,000	0,000	0,000	0,000	0,000	0,921	0,000	0,000
40	0,143	0,000	0,000	0,000	0,812	0,191	0,143	0,187	0,268	0,161	0,401	0,296	0,333	0,340	0,410	0,667	0,812	0,516	0,471

TABELA A.8: Vetores de características extraídos através de Biometria dos esqueletos dos órgãos foliares da Mata Atlântica e do Cerrado																			
	F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19
1	0,067	1,000	1,000	1,000	0,557	0,654	0,000	0,784	0,792	0,735	0,779	0,333	0,167	0,291	0,367	0,000	0,788	0,000	0,000
2	0,000	0,000	0,000	0,000	0,774	0,788	0,000	0,908	0,910	0,887	0,906	0,371	0,000	1,000	1,000	0,000	0,000	0,000	0,000
3	0,000	0,000	0,000	0,000	0,670	0,731	0,000	0,849	0,853	0,814	0,845	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
4	0,000	0,000	0,000	0,000	0,670	0,731	0,000	0,849	0,853	0,814	0,845	0,350	0,000	0,944	0,944	0,000	0,000	0,000	0,000
5	0,000	0,000	0,000	0,000	0,509	0,615	0,000	0,763	0,772	0,702	0,752	0,322	0,000	0,867	0,867	0,000	0,000	0,000	0,000
6	0,133	0,464	0,330	0,252	0,425	0,327	0,129	0,098	0,119	0,117	0,115	0,343	0,278	0,174	0,173	0,083	0,725	0,892	0,892
7	0,133	0,603	0,435	0,344	1,000	0,885	0,129	0,183	0,221	0,467	0,240	0,449	0,222	0,292	0,250	0,083	1,000	0,892	0,892
8	0,200	0,751	0,497	0,371	0,962	0,846	0,097	0,277	0,224	0,503	0,306	0,485	0,278	0,250	0,181	0,167	0,982	0,948	0,000
9	0,067	0,258	0,258	0,258	0,538	0,404	0,032	0,423	0,251	0,278	0,368	0,331	0,167	0,289	0,320	0,000	0,779	0,000	0,000
10	0,067	0,596	0,596	0,596	0,830	0,750	0,032	0,428	0,374	0,553	0,468	0,405	0,167	0,355	0,298	0,000	0,919	0,000	0,000
11	0,000	0,000	0,000	0,000	0,000	0,019	0,000	0,466	0,482	0,344	0,454	0,216	0,000	0,582	0,582	0,000	0,000	0,000	0,000
12	0,000	0,000	0,000	0,000	0,142	0,000	0,065	0,142	0,110	0,216	0,149	0,244	0,000	0,658	0,658	0,000	0,000	0,000	0,000
13	0,133	0,725	0,564	0,480	0,368	0,192	0,032	0,290	0,254	0,334	0,313	0,318	0,222	0,204	0,203	0,083	0,698	1,000	1,000
14	0,400	0,430	0,225	0,150	0,519	0,288	0,226	0,089	0,084	0,059	0,088	0,426	0,444	0,131	0,110	0,167	0,770	0,892	0,057
15	0,467	0,563	0,285	0,186	0,509	0,288	0,290	0,064	0,080	0,175	0,075	0,479	0,556	0,116	0,061	0,333	0,766	0,646	0,373
16	0,000	0,000	0,000	0,000	0,377	0,192	0,000	0,682	0,693	0,609	0,674	0,292	0,000	0,786	0,786	0,000	0,000	0,000	0,000
17	0,067	0,497	0,497	0,497	0,698	0,538	0,032	0,412	0,323	0,460	0,421	0,394	0,167	0,345	0,276	0,083	0,856	0,000	0,000
18	0,200	0,724	0,424	0,278	0,953	0,712	0,161	0,138	0,198	0,460	0,183	0,464	0,278	0,239	0,217	0,083	0,977	0,892	0,892
19	0,067	0,603	0,603	0,603	0,887	0,615	0,032	0,466	0,381	0,579	0,487	0,438	0,167	0,384	0,298	0,083	0,946	0,892	0,892
20	0,067	0,649	0,649	0,649	0,509	0,385	0,097	0,139	0,165	0,345	0,177	0,326	0,167	0,284	0,256	0,000	0,766	0,000	0,000
21	0,000	0,000	0,000	0,000	0,443	0,250	0,000	0,719	0,730	0,655	0,713	0,305	0,000	0,821	0,821	0,000	0,000	0,000	0,000
22	0,000	0,000	0,000	0,000	0,453	0,250	0,000	0,725	0,736	0,662	0,719	0,307	0,000	0,827	0,827	0,000	0,000	0,000	0,000
23	0,000	0,000	0,000	0,000	0,387	0,231	0,000	0,687	0,697	0,616	0,680	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
24	0,133	0,308	0,218	0,162	0,415	0,154	0,065	0,217	0,161	0,136	0,213	0,333	0,222	0,214	0,228	0,083	0,721	0,948	0,948
25	0,133	0,632	0,515	0,444	0,217	0,096	0,032	0,258	0,203	0,231	0,259	0,290	0,222	0,185	0,200	0,083	0,626	0,948	0,948
26	0,133	0,421	0,319	0,268	0,726	0,577	0,129	0,229	0,131	0,386	0,228	0,392	0,222	0,254	0,224	0,000	0,869	0,000	0,000
27	0,067	0,013	0,013	0,013	0,934	0,769	0,000	1,000	1,000	1,000	1,000	0,409	0,167	0,359	0,540	0,000	0,968	0,000	0,000
28	0,133	0,967	0,778	0,669	0,906	0,712	0,032	0,439	0,407	0,576	0,492	0,413	0,222	0,268	0,278	0,000	0,955	0,000	0,000
29	0,133	0,636	0,467	0,381	0,830	0,673	0,065	0,275	0,270	0,450	0,327	0,402	0,222	0,260	0,204	0,000	0,919	0,000	0,000
30	0,067	0,603	0,603	0,603	0,792	0,635	0,000	0,924	0,927	0,901	0,917	0,383	0,167	0,335	0,303	0,000	0,901	0,000	0,000
31	0,067	0,762	0,762	0,762	0,915	0,904	0,032	0,482	0,375	0,672	0,511	0,447	0,167	0,393	0,297	0,083	0,959	0,948	0,948
32	0,067	0,821	0,821	0,821	0,962	0,981	0,097	0,235	0,236	0,440	0,277	0,460	0,167	0,405	0,307	0,083	0,982	0,948	0,948
33	0,400	0,826	0,371	0,223	0,840	0,942	0,323	0,088	0,094	0,419	0,103	0,504	0,500	0,138	0,125	0,167	0,923	0,800	0,149
34	0,200	0,868	0,516	0,355	0,906	0,865	0,161	0,184	0,167	0,579	0,196	0,460	0,278	0,237	0,203	0,083	0,959	0,797	0,797
35	0,400	0,850	0,380	0,229	0,868	1,000	0,323	0,108	0,101	0,465	0,119	0,519	0,444	0,162	0,168	0,333	0,937	0,942	0,080
36	0,867	0,532	0,268	0,168	0,745	0,442	0,677	0,134	0,064	0,093	0,086	0,943	0,833	0,155	0,085	0,750	0,883	0,845	0,323
37	0,533	0,510	0,239	0,153	0,491	0,327	0,484	0,044	0,024	0,133	0,016	0,633	0,611	0,142	0,070	0,583	0,757	0,779	0,330
38	0,800	0,538	0,207	0,122	0,651	0,500	0,742	0,079	0,000	0,179	0,018	0,932	0,889	0,144	0,076	1,000	0,838	0,796	0,364
39	0,933	0,544	0,242	0,153	0,623	0,462	0,581	0,119	0,013	0,198	0,049	1,000	0,944	0,145	0,085	0,917	0,820	0,853	0,326
40	0,867	0,452	0,253	0,160	0,509	0,250	0,581	0,088	0,015	0,090	0,060	0,759	0,889	0,114	0,055	0,833	0,766	0,844	0,314
41	0,267	0,450	0,248	0,168	0,462	0,250	0,194	0,079	0,083	0,091	0,084	0,428	0,333	0,180	0,118	0,250	0,743	0,841	0,042
42	0,867	0,794	0,342	0,210	0,981	0,808	1,000	0,081	0,060	0,389	0,050	0,936	0,944	0,135	0,079	0,917	0,991	0,838	0,307
43	0,533	0,681	0,281	0,166	0,670	0,500	0,516	0,057	0,046	0,306	0,029	0,585	0,556	0,144	0,089	0,333	0,842	0,893	0,065
44	0,533	0,585	0,259	0,162	0,717	0,635	0,613	0,077	0,027	0,181	0,041	0,824	0,667	0,172	0,088	0,750	0,869	0,734	0,413
45	1,000	0,627	0,298	0,184	0,689	0,538	0,774	0,114	0,011	0,245	0,077	0,911	1,000	0,122	0,058	0,833	0,851	0,847	0,313
46	0,667	0,462	0,185	0,112	0,340	0,058	0,355	0,081	0,022	0,118	0,020	0,862	1,000	0,116	0,088	0,750	0,685	0,751	0,432
47	0,533	0,450	0,201	0,121	0,358	0,192	0,419	0,000	0,032	0,050	0,000	0,443	0,556	0,106	0,079	0,167	0,694	0,998	0,157
48	0,667	0,399	0,197	0,127	0,557	0,269	0,355	0,107	0,057	0,059	0,073	0,816	0,778	0,144	0,091	0,667	0,797	0,743	0,495
49	0,467	0,423	0,229	0,142	0,434	0,288	0,290	0,046	0,086	0,123	0,057	0,492	0,611	0,108	0,117	0,417	0,730	0,721	0,377
50	0,600	0,397	0,277	0,165	0,575	0,192	0,452	0,049	0,061	0,000	0,040	0,540	0,667	0,108	0,122	0,500	0,797	0,965	0,113

Referências

- (BAJA e THIEL, 1996) BAJA, G. S. e E. THIEL. Skeletonization algorithm running on path-based distance maps. *Image and Vision Computing*, v.14, p.47-57. 1996.
- (BARBARÁ e CHEN, 2000) BARBARÁ, D. e P. CHEN. Using the fractal Dimension to Cluster Datasets. *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. Boston: ACM Press, 1-10 p., 2000
- (BLUM, 1967) BLUM, H. A transformation for Extracting New Descriptors of Shape. In *Models for the Perception of Speech and Visual Form*. Cambridge: MIT Press, 1967.
- (BOCKHOFF, 2001) BOCKHOFF, R. C. Plant Morphology: The Historic Concepts of Wilhelm Troll, Walter Zimmermann and Agnes Arber. *Annals of Botany*, v.88, p.1153-1172. 2001.
- (BRAGA, CARVALHO *et al.*, 2000) BRAGA, A. P., A. P. L. F. CARVALHO e T. B. LUDERMIR. *Redes Neurais Artificiais: Teoria e Aplicações*. Rio de Janeiro: LTC. 262 p., 2000.
- (BRUNO, 2000) BRUNO, O. M. *Paralelismo em Visão Natural e Artificial*. (Doutorado). Instituto de Física de São Carlos, Universidade de S. Paulo, São Carlos, 2000.
- (BRUNO e COSTA, 2001) BRUNO, O. M. e L. F. COSTA. A Parallel Approach for a Plant Taxonomy System. *Iber-American Symposium on Pattern Recognition*. Florianópolis, 159-165 p., 2001
- (BRUNO, PLOTZE *et al.*, 2004) BRUNO, O. M., R. O. PLOTZE e L. A. V. D. CARVALHO. Algoritmo para Processamento de Imagens da Retina in vivo com Aplicações na Biometria. *Scientia* (to appear). 2004.
- (CARDOSO, 2000) CARDOSO, C. M. V. *Vascularização Foliar em Espécies de Eugenia L.(Myrtaceae) da Bacia Hidrográfica do Rio Tibagi, Paraná*. (Mestrado). Instituto de Biociências, Universidade Estadual Paulista, Rio Claro, 2000.
- (CARLIN, 2000) CARLIN, M. Measuring the Complexity of Non-Fractal Shapes by a Fractal Method. *Pattern Recognition Letters*, v.21, p.1013-1017. 2000.

- (CASTLEMAN, 1979) CASTLEMAN, K. R. Digital Image Processing. New Jersey: Prentice-Hall. 429 p., 1979.
- (CHAMARTHY, STANLEY *et al.*, 2004) CHAMARTHY, P., R. J. STANLEY, G. CIZEK, R. LONG, S. ANTANI e G. THOMA. Image Analysis Techniques for Characterizing Disc Space Narrowing in Cervical Vertebrae Interfaces. Computerized Medical Imaging and Graphics, v.28, n.1-2, p.39-50. 2004.
- (CONSULARO, 2000) CONSULARO, L. A. Técnica de Mineração de dados para Análise de Imagens. (Doutorado). Instituto de Física, Universidade de São Paulo, São Carlos, 2000.
- (COSTA, 1999) COSTA, L. F. Multidimensional Scale-Space Shape Analysis. Proceedings International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging. Santorini - Greece, 241-247 p., 1999
- (COSTA, 2000) COSTA, L. F. Robust Skeletonization Through Exact Euclidean Distance Transform and Its Application to Neuromorphometry. Real-Time Imaging, v.6, p.415-431. 2000.
- (COSTA, 2003) COSTA, L. F. Characterizing Spatial Gene Expression Profiles by Using the Local Multiscale Fractal Dimension. Pharmatech, p., 2003
- (COSTA e BIANCHI, 2002) COSTA, L. F. e A. G. C. BIANCHI. A Outra Dimensão da Dimensão Fractal. Ciência Hoje, v.31, n.183, p.40-47. 2002.
- (COSTA e CESAR, 2000) COSTA, L. F. e R. M. CESAR. Shape Analysis and Classification: Theory and Practice. Pennsylvania: CRC Press. 659 p., 2000.
- (COSTA, MANOEL *et al.*, 2002) COSTA, L. F., E. T. M. MANOEL, F. FAUCEREAU, J. CHELLY, J. VAN PELT e G. RAMAKERS. A Shape Analysis Framework for Neuromorphometry. Network, v.13, n.3, p.283-310. 2002.
- (DOUGHERTY e HENEBRY, 2001) DOUGHERTY, G. e G. M. HENEBRY. Fractal Signature and Lacunarity in the Measurement of the Texture of Trabecular Bone in Clinical CT Images. Medical Engineering & Physics, n.23, p.369-380. 2001.
- (DUDA e HART, 1973) DUDA, R. O. e P. E. HART. Pattern Classification and Scene Analysis. New York: Wiley-Interscience, 1973.
- (FALCÃO, COSTA *et al.*, 2002) FALCÃO, A. X., L. F. COSTA e G. S. CUNHA. Multiscale Skeletons by Image Foresting Transform and its Application to Neuromorphometry. Pattern Recognition, v.35, p.1571-1582. 2002.
- (FALVO, PLOTZE *et al.*, 2004) FALVO, M., R. O. PLOTZE e O. M. BRUNO. TreeVis - Um Laboratório Biométrico para Análise e Identificação de Estruturas Foliares. 2004.
- (FISHER, 1936) FISHER, R. A. The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, v.7, p.179-188. 1936.
- (GONZALEZ e WOODS, 2002) GONZALEZ, R. C. e R. E. WOODS. Digital Image Processing: Addison-Wesley, 2002.

- (HAYKIN, 2001) HAYKIN, S. Neural Networks - A Comprehensive Foundation: Prentice Hall. 900 p., 2001.
- (HE e YAN, 2000) HE, R. e H. YAN. Stroke Extraction as Pre-Processing Step to Improve Thinning Results to Chinese Characters. Pattern Recognition Letters, v.21, p.817-825. 2000.
- (HOFER, GOURLAY *et al.*, 2001) HOFER, J. M. I., C. W. GOURLAY e T. H. N. ELIS. Genetic Control of Leaf Morphology: A Partial View. Annals of Botany, v.88, p.1129-1139. 2001.
- (HOLT, STEWART *et al.*, 1987) HOLT, C. M., A. STEWART, M. CLINT e R. H. PERROT. An Improved Parallel Thinning Algorithm. Communications of the ACM, v.30, p.156-160. 1987.
- (INSELBERT e DIMSDALE, 1990) INSELBERT, A. e B. DIMSDALE. Parallel Coordinates: A Tool for Visualizing MultiDimensional Geometry. IEEE Visualization 1990: IEEE Computer Society, 361-378 p., 1990
- (JAIN, 1989) JAIN, A. K. Fundamentals of Digital Image Processing. London: Prentice Hall, 1989.
- (JAIN, DUIN *et al.*, 2000) JAIN, A. K., R. P. W. DUIN e J. MAO. Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, v.22, n.1, p.4-37. 2000.
- (JOHNSON e WICHERN, 1988) JOHNSON, R. A. e D. W. WICHERN. Applied Multivariate Statistical Analysis: Englewood Cliffs: Prentice Hall. 607 p., 1988.
- (LAM, LEE *et al.*, 1992) LAM, L., S.-W. LEE e C. Y. SUEN. Thinning Methodologies - A Comprehensive Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, v.14, n.9, p.869-885. 1992.
- (LÓPEZ e VÍLCHEZ-CABRERIZO, 2001) LÓPEZ, F. M. e M. A. VÍLCHEZ-CABRERIZO. An Improved Method to Estimate the Fractal Dimension of Physical Fractals Based on the Hausdorff Definition. Physica A, v.298, p.387-399. 2001.
- (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. Oxford: W.H. Freeman, 1983.
- (MANOEL, COSTA *et al.*, 2002) MANOEL, E. T. M., L. F. COSTA, J. STREICHER e G. B. MÜLLER. Multiscale Fractal Characterization of Three-Dimensional Gene Expression Data. Proceedings of XV Brazilian Symposium of Computer Graphics and Image Processing (SIBGRAPI'02). Fortaleza, 269-274 p., 2002
- (MCCULLOCH e PITTS, 1943) MCCULLOCH, W. S. e W. PITTS. A Logical Calculus of Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, v.5, p.115-133. 1943.

- (MELHI, IPSON *et al.*, 2001) MELHI, S. S., S. S. IPSON e W. BOOTH. A Novel Triangulation Procedure for Thinning Hand-Written Text. *Pattern Recognition Letters*, v.22, p.1059-1071. 2001.
- (MINSKY e PAPERT, 1969) MINSKY, M. e S. PAPERT. *Perceptrons*. Cambridge MA: MIT Press, 1969.
- (NEUSIUS e OLSZEWSKI, 1994) NEUSIUS, C. e J. OLSZEWSKI. A Noniterative Thinning Algorithm. *ACM Transactions on Mathematical Software*, v.20, n.1, p.5-20. 1994.
- (NILCEU, COSTA *et al.*, 1999) NILCEU, A. M., L. F. COSTA, R. A. LOTUFO e S. A. VELASTIN. Estimating Crowd Density With Minkowsky Fractal Dimension. *International Conference on Acoustics, Speech and Signal Processing*. Phoenix - Arizona, p., 1999
- (OLIVEIRA e LEVKOWITZ, 2003) OLIVEIRA, M. C. F. e H. LEVKOWITZ. From Visual Data Exploration to Visual Data Mining: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, v.9, n.3, p.378-394. 2003.
- (OLIVEIRA, 1997) OLIVEIRA, P. R. *Redes Neurais Artificiais para Extração de Características*. (Dissertação de Mestrado). Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 1997.
- (PÁDUA, ROSSIN *et al.*, 2004) PÁDUA, J. G., R. P. ROSSIN, R. O. PLOTZE, M. FALVO, L. C. BERNACCI, O. M. BRUNO e G. C. X. OLIVEIRA. Utilização de Morfometria, Filogenia e Marcadores SSR para Quantificação da Diversidade Genética do Gênero *Passiflora* e suas Implicações Evolutivas e Ecológicas. *Congresso Nacional de Genética*. Florianópolis, p., 2004
- (PÁDUA, VIEIRA *et al.*, 2002) PÁDUA, J. G., M. L. C. VIEIRA e G. C. X. OLIVEIRA. Phylogenetics Relationships Among Some Species of *Passiflora* (Passifloraceae). 48o. *Congresso Nacional de Genética*. Águas de Lindóia, p., 2002
- (PARKER, 1997) PARKER, J. R. *Algorithms for Image Processing and Computer Vision*. New York: John Wiley & Sons, 1997.
- (PAVLIDIS, 1982) PAVLIDIS, T. *Algorithms for Graphics and Image Processing*. Maryland: Computer Science Press, 1982.
- (PEITGEN e SAUPE, 1988) PEITGEN, H. O. e D. SAUPE. *The Science of Fractal Images*. New York: Springer, 1988.
- (PEIXOTO e MORIM, 2003) PEIXOTO, A. L. e M. P. MORIM. *Coleções Botânicas: Documentação da Biodiversidade Brasileira*. *Ciência e Cultura*, v.55, n.3, p.21-24. 2003.
- (PINTO, 2001) PINTO, S. C. D. *Estimação da Dimensão Fractal de Imagens de SPM*. (Dissertação de Mestrado). Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2001.
- (PLOTZE e BRUNO, 2004) PLOTZE, R. O. e O. M. BRUNO. Estudo e Comparação de Algoritmos de Esqueletonização para Imagens binárias. *IV Congresso Brasileiro de Computação - CBComp*. Itajaí-SC, p., 2004

- (PLOTZE, CARVALHO *et al.*, 2003) PLOTZE, R. O., L. A. V. D. CARVALHO e O. M. BRUNO. An Algorithm to Perform Geometric Analysis of Eye Fundus Blood Vessels. International Conference on BioInformatics and Computacional Biology - ICoBiCoBi. Ribeirão Preto, p., 2003
- (PLOTZE, FALVO *et al.*, 2003) PLOTZE, R. O., M. FALVO e O. M. BRUNO. Identificação de Espécies Arbóreas Através da Dimensão Fractal. I Workshop de Informática aplicada ao Meio Ambiente - Congresso Brasileiro de Computação. Itajai-SC, 1-10 p., 2003
- (PLOTZE, FALVO *et al.*, 2004) PLOTZE, R. O., M. FALVO, J. G. PÁDUA, G. C. X. OLIVEIRA, M. L. C. VIEIRA e O. M. BRUNO. Análise Biométrica de Órgãos Foliares de Passiflora. 49a. Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria. Uberlândia, 1-5 p., 2004
- (PLOTZE, PÁDUA *et al.*, 2004) PLOTZE, R. O., J. G. PÁDUA, M. FALVO, O. M. BRUNO, M. L. C. VIEIRA e G. C. X. OLIVEIRA. Leaf Shape Analysis by the Multiscale Minkowsky Fractal Dimension, a New Morphometric Method: A Study in Passiflora L. (Passifloraceae). Canadian Journal of Botany (submetido). 2004.
- (PLOTZE, PÁDUA *et al.*, 2004) PLOTZE, R. O., J. G. PÁDUA, M. FALVO, M. L. C. VIEIRA, G. C. X. OLIVEIRA e O. M. BRUNO. Passiflora and Heliconius: Phylogenetic and Morphometric Approaches to Understand Leaf Shape Variation as an Escape Strategy. Taxon (submetido). 2004.
- (RAUBER, 1998) RAUBER, T. W. Redes Neurais Artificiais. ERI - Encontro Regional de Informática. Nova Friburgo: Sociedade Brasileira de Computação, 201-228 p., 1998
- (RICH e KNIGHT, 1983) RICH, E. e K. KNIGHT. Artificial Intelligence. New York: McGraw Hill. 448 p., 1983.
- (ROSENBLATT, 1958) ROSENBLATT, F. The perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review, v.65, n.386-408. 1958.
- (ROSENFELD, 2001) ROSENFELD, A. From Image Analysis to Computer Vision: An Annotated Bibliography, 1955-1979. Computer Vision and Image Understanding, v.84, p.298-324. 2001.
- (RUMELHART, HINTON *et al.*, 1986) RUMELHART, D. E., G. E. HINTON e R. J. WILLIAMS. Learning Internal Representations by Error Propagation. In: D. E. RUMELHART e J. L. MCCLELLAND (Ed.). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge: MIT Press. Learning Internal Representations by Error Propagation, p.318-362 1986
- (RUTOVITZ, 1966) RUTOVITZ, D. Pattern recognition. Journal of the Royal Statistical Society, v.129, p.504-530. 1966.
- (SARKAR e CHAUDHURI, 1992) SARKAR, N. e B. B. CHAUDHURI. An Efficient Approach to Estimate Fractal Dimension of Textural Images. Pattern Recognition, v.25, p.1035-1042. 1992.

- (SARKAR e CHAUDHURI, 1994) SARKAR, N. e B. B. CHAUDHURI. An Efficient Differential Box Counting Approach to Compute Fractal Dimension of Image. IEEE Transactions on Systems, Man and Cybernetics, v.24, p.115-120. 1994.
- (SHROEDER, 1996) SHROEDER, M. Fractals, Chaos, Power Laws - Minutes from a Infinite Paradise. New York: W.H. Freeman and Company, 1996.
- (SMITH, 1999) SMITH, S. W. The scientist and engineer's guide to digital signal processing: California Technical Publishing, 1999.
- (SOILLE, 2000) SOILLE, P. Morphological Image Analysis Applied to Crop Field Mapping. Image and Vision Computing, v.18, p.1025-1032. 2000.
- (SONKA, HLAVAC *et al.*, 1999) SONKA, M., V. HLAVAC e R. BOYLE. Image Processing Analysis, and Machine Vision: Pacific Grove: Brooks Cole. 770 p., 1999.
- (STENTIFORD, 1983) STENTIFORD, F. Some new heuristics for thinning binary handprinted caracteres for OCR. IEEE Transactions on Systems, Man and Cybernetics, v.13, n.1, p.81-84. 1983.
- (SUN, 2004) SUN, D. W. Computer Vision - An Objective, Rapid and Non-Contact Quality Evaluation Tool for the Food Industry. Journal of Food Engineering, v.61, n.1, p.1-2. 2004.
- (TRICOT, 1995) TRICOT, C. Curves and Fractal Dimension. New York: Springer-Verlag, 1995.
- (TURNER, BLACKLEDGE *et al.*, 1998) TURNER, M. J., J. M. BLACKLEDGE e P. R. ANDREWS. Fractal Geometry in Digital Imaging. Cambridge: Academic Press. 328 p., 1998.
- (VAJNA, ROVATTI *et al.*, 2000) VAJNA, Z. M. K., R. ROVATTI e R. FRAZZONI. Fingerprint Ridge Distance Computation Methodologies. Patter Recognition, v.33, p.69-80. 2000.
- (VANDERPLANK, 2000) VANDERPLANK, J. Passion Flowers. Cambridge: MIT Press. 224 p., 2000.
- (VOSS, 1988) VOSS, R. F. Fractal in Nature: from Characterization to Simulation. In: H. O. PEITGEN e D. SAUPE (Ed.). The Science of Fractal Images. New York: Springer. Fractal in Nature: from Characterization to Simulation, p.21-70. 1988.
- (WARD, 2004) WARD, M. O. XmdvTool home page. (acessado em:06/2004) 2004.
- (WING, WILF *et al.*, 1999) WING, S., P. WILF, K. JOHNSON, L. J. HICKEY, B. ELLIS e A. ASH. Manual of Leaf Architecture - Morphological Description and Categorization of Dicotyledonous and Net-veined Monocotyledonous Angiosperms by Leaf Architecture. Washington: Leaf Architecture Working Group. 65 p., 1999.
- (WINTERS e SANTOS-VICTOR, 2002) WINTERS, N. e J. SANTOS-VICTOR. Information Sampling for Vision-Based Robot Navigation. Robotics and Autonomous Systems, v.41, n.2-3, p.145-159. 2002.

- (YOUNG, 1994) YOUNG, T., Y. Handbook of Pattern Recognition and Image Processing. London: Academic Press, v.2. 584 p., 1994.
- (ZHANG e SUEN, 1984) ZHANG, T. Y. e C. Y. SUEN. A Fast Parallel Algorithm for Thinning Digital Patterns. Communications of ACM, v.27, n.3. 1984.