

Universidade de São Paulo
Instituto de Física de São Carlos

Análise e Classificação de Formas
Biológicas

USP/IFSC/SBI



Renata Antonia Tadeu Arantes

Tese de Doutorado apresentada ao Instituto de Física de São Carlos, da Universidade de São Paulo, para obtenção do título de Doutora em Ciências – Física Aplicada – Opção Computacional.

Orientador: Prof. Dr. Luciano da Fontoura Costa

São Carlos - 2004

IFSC-USP SERVIÇO DE BIBLIOTECA
INFORMAÇÃO

Arantes, Renata Antonia Tadeu.

“Análise e Classificação de Formas Biológicas”
Renata Antonia Tadeu Arantes — São Carlos, 2004.

Tese (Doutorado) – Área de Física da Universidade de
São Paulo, 2004. Páginas:129

Orientador: Prof. Dr. Luciano da Fontoura Costa

1. Morfologia. 2.Análise de Forma 3. Curvatura Digital.

I - Título



MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE RENATA ANTÔNIA
TADEU ARANTES APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS,
UNIVERSIDADE DE SÃO PAULO, EM 13-01-2004.

COMISSÃO JULGADORA:

Prof. Dr. Luciano da Fontoura Costa (Orientador e Presidente) – IFSC/USP

Profa. Dra. Agma Juci Machado Traina – ICMC/USP

Prof. Dr. Adilson Gonzaga – EESC/USP

Prof. Dr. Roberto Marcondes César Júnior – IME/USP

Prof. Dr. Marcelo Emílio Beletti – UFU

Dedico este trabalho

A minha mãe Lourayde pelo amor, dedicação, carinho, oportunidade e sacrifícios.

Ao meu querido filho Rhenan pelo amor e compreensão pela distância nestes anos.

Agradecimentos

Ao meu orientador Luciano da. F. Costa pela oportunidade oferecida para execução dessa tese, pela excelente e competente orientação, pela confiança depositada em mim, sem os quais esse trabalho não teria sido realizado.

Ao Professor Sérgio Furtado dos Reis pela grande colaboração com as imagens, pela disponibilidade, ajuda e amizade.

Ao Professor Roberto, Celina e Erika que muito contribuíram para o desenvolvimento do trabalho.

Aos amigos do laboratório CVRG: Bruno, Ricardo, Cesar, Abel, Edson, Rafael, Mateus, Marcelo, Erbe, Marconi, Leandro, Linder, Célia, Dilson, pela convivência amigável durante todo o período de doutoramento, pelo deliciosos churrascos e em especial ao Pedro e Giancarlo pelas discussões do trabalho e ajuda significativa na execução da tese.

Aos técnicos Marquinho e Leandro que sempre auxiliaram no dia a dia, por tudo e pela grande amizade.

A ex-secretária do grupo Sueli e a mais nova Sonia, Ivanilda pela amizade e assistência sempre que necessária.

As minhas amigas Silvia, Daniela, Andrea e Katia o meu muito obrigado pela ajuda, contribuição e pelos momentos felizes, e momentos de ansiedade que compartilhamos.

Aos bons e velhos amigos Katarine, Ana, Sandra, Edmarcio, Ana Rita, Cesar, Mauricio, aos amigos do Ladi, valeu cada momento que passamos juntos.

A todos os funcionários do IFSC, as bibliotecárias que sempre se dedicaram em me dispor toda a bibliografia de que precisei para compor minha tese, as secretárias da Pós

Wladerez e Cris pela enorme paciência, aos professores que contribuíram para minha formação, meu muito obrigado.

Aos meus Pais pela estrutura familiar proporcionando a serenidade e calma mesmo nos momentos mais difíceis, em especial a minha querida mãe que incondicionalmente compartilha comigo momentos de ansiedade, incerteza e esperança, meu muito obrigado pelo apoio emocional, carinho e anos de dedicação exclusiva ao meu filho.

Ao meu filho Rhenan, pela paciência de esperar todo este tempo, pelo amor, carinho, incentivo, a razão de tudo, deixo aqui registrado “Eu que te amo mais”.

A minha avó Alayde pelo carinho, amor e ajuda.

Aos meus irmãos Leticia e Junior, cunhados (Lucia e Raul) e sobrinhos (Matheus e Vinnicius) por compreender minha distância mesmo sem muitas vezes entender o motivo. Vocês sempre me proporcionaram uma alegria imensa.

A Deus por mais um sonho realizado e pela perfeita inserção de pessoas na minha vida.

A Cnpq pelo apoio financeiro.

A todos vocês pessoas maravilhosas que fizeram parte da realização de um sonho, meu
MUITO OBRIGADO

Sumário

LISTA DE FIGURAS.....	IX
LISTA DE TABELAS.....	XVI
LISTA DE ABREVIATURAS.....	XIX
RESUMO.....	XX
ABSTRACT.....	XXI

CAPÍTULO 1 – INTRODUÇÃO

1.1 – Considerações Iniciais.....	1
1.2 – Motivação do Trabalho.....	3
1.3 – Objetivo do trabalho.....	4
1.4 – Organização do texto.....	4

CAPÍTULO 2 – FORMAS BIOLÓGICAS

2.1 – Formas na Natureza.....	6
2.2 – Morfometria.....	8
2.3 – Dados em morfometria.....	10
2.4 – Taxonomia.....	12
2.5 – Princípios de formas.....	13
2.6 – Análise de formas por “landmarks”.....	14
2.7 – Conceitos de formas.....	16
2.8 – Caracterização de formas.....	17

CAPÍTULO 3 – RECONHECIMENTO DE PADRÕES

3.1 – Relevância.....	18
3.2 – Extração de características.....	19
3.3 – Extração de características baseadas nos “landmarks”.....	19

3.4 – Seleção de atributos.....	21
3.5 – Estado da Arte de Reconhecimento de Padrões.....	23
3.6 – Conceitos de classificação.....	22
3.7 – Tipos de classificação.....	24
3.8 – Fases do processo de Reconhecimento de Padrões.....	25

CAPÍTULO 4 – MATERIAIS E MÉTODOS PROPOSTOS

4.1 – Fases envolvidas na pesquisa.....	28
4.2 – Base de dados utilizada.....	30
4.3 – Softwares utilizados.....	30
4.4 – Métodos utilizados para a segmentação.....	30
4.4.1 – Suavização.....	31
4.4.2 – Binarização.....	34
4.4.3 – Dilatação e Erosão.....	35
4.4.4 – Detecção de contorno.....	36
4.5 – Métodos utilizados para extração de atributos.....	37
4.5.1 – Métodos de Distância.....	37
4.5.2 – Análise de Deformações.....	46
4.5.2.1 – Thin-Plate Spline.....	47
4.5.2.2 – Warps.....	50
4.5.3 – Curvatura.....	54
4.5.4 – Análise de Componente Principal	63
4.6 – Métodos de Classificação implementados.....	65

CAPÍTULO 5 – RESULTADOS

5.1 – Resultados extraídos por curvatura contínua.....	69
--	----

5.1.1- Curvatura contínua obtida de crânios de roedores <i>T. apereoides</i>	69
5.1.2 – Curvatura contínua obtida de amostras de mandíbulas de roedores <i>T. apereoides</i>	73
5.1.3- Curvatura contínua obtida de amostras de crânios de canídeos.....	74
5.2 – Resultados extraídos por curvatura digital.....	78
5.2.1 – Resultados de estudos da variação geográfica da espécie de roedores (<i>T.apereoides</i>) usando métodos de classificação não supervisionados (clustering hierárquicos.....	78
5.2.2 – Validação dos resultados extraídos das medidas de curvatura digital usando métodos de feature selection FSS para crânios.....	84
5.2.3 – Resultados obtidos de curvatura digital para mandíbulas.....	86
5.2.4 – Validação dos resultados extraídos das medidas de curvatura digital usando métodos de feature selection FSS para mandíbulas.....	89
5.2.5 – Resultados obtidos de curvatura digital para amostras de crânio de canídeos.....	90
5.3 – Resultados extraídos de Thin-plate Spline e Warps.....	93
5.3.1 - Resultados plotados de warps principal, parcial e relativos para crânio da espécie <i>T. apereoides</i>	95

5.4 – Classificação usando métodos supervisionados e não-supervisionados.....	96
5.4.1- Supervisionados.....	96
5.4.1.1 – Roedores.....	97
5.4.1.2 – Canideos.....	108
5.4.2 - Não supervisionados.....	112
5.4.2.1 – Clustering hierárquico.....	112
5.4.2.2 – K-médias.....	112

CAPÍTULO 6: RESULTADOS DO TRABALHO

6.1 – Contribuições.....	116
6.2 – Desenvolvimentos Futuros.....	119
6.3 – Produção Científica, Tecnológica.....	119
6.3.1 – Produção Bibliográfica.....	119
6.3.1.1 – Textos em Revistas.....	119
6.3.2.1 – Participação em eventos.....	120

CAPÍTULO 7: REFERÊNCIAS BIBLIOGRÁFICAS

Lista de Figuras

- Figura 1:** A natureza com suas formas e cores..... 7
- Figura 2:** Vista da mandíbula da espécie *T. apereoides* com seus respectivos landmarks..... 11
- Figura 3:** As atividades normalmente implicadas em análise de formas biológicas. Resultados parciais, incluindo as imagens originais, medidas e resultados, são armazenados e organizados em bases de dados..... 13
- Figura 4:** Processo de segmentação realizado até obtenção do contorno e o passo seguinte inserção de pontos de controle. (A) imagem original ; (b) imagem limiarizada; (c) extração do contorno..... 14
- Figura 5:** Exemplo da espécie *T. apereoides* (crânio lateral) com landmark.....15
- Figura 6:** Exemplos de formas..... 17
- Figura 7:** Imagens de canídeos (a) Imagem original canídeo sem etiquetas; (b) inserção de marcos na forma usando o programa TpsDig; (c) contorno da forma ; (d) Inserção de landmarks no contorno da forma..... 20
- Figura 8:** As três fases do reconhecimento de padrão. [K.Castleman, (1995)]... 26
- Figura 9:** Diagrama de blocos geral do sistema..... 29
- Figura 10:** Exemplo genérico de uma máscara 3 x 3..... 31
- Figura 11:** Amostras pertencentes ao grupo canídeos. (original).....32
- Figura 12:** Amostras pertencentes ao grupo canídeos. (original, sem etiquetas).....32

Figura 13: Amostra canideos (mesma figura 12) mas agora no matlab.....33

Figura 14 : Amostra canideos, imagem passou por filtro de mediana (2 vezes).....33

Figura 15 : Amostras pertencentes ao grupo canideos. (Binarizadas).....34

Figura 16: (a) Imagem original , (b) Imagem dilatada.....35

Figura 17: Amostras pertencentes ao grupo canideos (fechamento).....36

Figura 18 : Contorno da amostras do grupo canideos. (Utilizamos o programa ceguinho para fazer a detecção de contorno).....37

Figura 19 : Amostras de espécies de *T. apereoides* a serem comparadas, mostrando as diferenças de tamanho, posição e orientação.....39

Figura 20: Amostras de espécies de *T. apereoides* centrados na origem após a translação..... 40

Figura 21: Espécie de *T. apereoides*. Definição dos segmentos de linhas utilizados para o cálculo do tamanho centróide. Distância de cada *landmark* ao centro de massa..... 41

Figura 22: Espécie de *T. apereoides* configuradas após a translação e proporcionalização, continuam centradas na origem..... 42

Figura 23: Thin-plate Spline obtido pela adição de pontos de controle. Observe que a deformação na placa é obtida pela energia de dobra cada vez maior..... 50

Figura 24: Uma forma genérica representada em termos de N pontos *landmarks*, vetores e ângulos, definido para cada landmark v_i 58

Figura 25: Landmarks morfológicos definidos pela vista dorsal , lateral e ventral do esqueleto de *Thrichomys apereoides*..... 61

Figura 26: (a) conjunto anterior, (b) conjunto novo de variáveis..... 63

Figura 27: Organograma dos Resultados.....68

Figura 28: Gráfico das amostras de crânios das espécies *T. apereoides*, (a) landmarks plotados para ponto adicional, (b) landmarks interpolados linearmente, (c) funções gaussianas (d) Funções das coordenadas x e y, cálculo das derivadas e *fourier*, (e) curvaturas.....69

Figura 29: Plot do crânio lateral para curvatura contínua com ponto adicional.....70

Figura 30: Dendrograma do método Ward para curvatura contínua usando ponto adicional para amostras de crânio *T. apereoides*.....71

Figura 31: Plot do crânio lateral para curvatura contínua sem ponto adicional.....72

Figura 32: Plot da mandíbula para curvatura contínua sem ponto adicional.....73

Figura 33: Amostras de crânios de canideos), (a) landmarks plotados para ponto adicional, (b) landmarks interpolados linearmente, (c) funções gaussianas (d) Funções das coordenadas x e y, cálculo das derivadas e *fourier* (e) curvaturas.....75

Figura 34: Plot do crânio para curvatura contínua com ponto adicional.....75

Figura 35: Dendrograma do método Ward para curvatura contínua usando ponto adicional para amostras de canideos. (curvatura).....76

Figura 36: Plot do crânio (canideo) para curvatura contínua sem ponto adicional.....77

Figura 37: Dendrograma do método Ward para curvatura contínua sem ponto adicional para amostras de canideos. (somente *feature* de curvatura).....78

Figura 38: Vista lateral do crânio da espécie *T. apereoides* contendo 14 *landmarks*.....80

Figura 39: Dendrograma obtido do método ward usando curvatura digital, considerando (energia e retirando o primeiro e último ângulo) para obtenção das 5 classes.....81

Figura 40: Amostras de crânio (*T. apereoides*) (a) Programa desenvolvido no matlab (Aglomerativo).....83

Figura 41: Amostras de crânio (*T. apereoides*); Método Aglomerativo (S-Plus)...83

Figura 42: Amostras de crânio (*T. apereoides*); Método Divisivo (S-Plus).....84

Figura 43: Vista lateral da mandíbula da espécie *T. apereoides* contendo 11 *landmarks*.....87

Figura 44: Dendrograma obtido pelo método Ward usando curvatura digital e energia, para mandíbula.....88

Figura 45: Método hierárquico Aglomerativo gerado no Splus com amostras de mandíbulas.....88

Figura 46: Método hierárquico divisivo gerado no Splus com amostras de mandíbulas.....89

Figura 47: Amostras de canideos *Atelocynos* (a) TpsDig com seus respectivos *landmarks*; (b) Matlab, *landmarks* interpolados.....91

Figura 48: Dendrograma obtido pelo método Ward das amostras de canideos usando curvatura digital e energia.....92

Figura 49: (a) Método Aglomerativo.....92

Figura 50: (b) Método Divisivo.....93

Figura 51 : Amostra da vista lateral do crânio *T. apereoides* e plot da deformação da mesma vista no programa Thin-plate.....94

Figura 52: Plot dos principal warps para espécie de referência.....95

Figura 53: Plot do parcial warps para espécie de referência.....95

Figura 54: (a) Configuração de referência, (b) configuração de referência conectada.....95

Figura 55: Relative Warp 1 + deformação.....96

Figura 56: Plot da matriz de peso do relative warps.....96

Figura 57: Plote bivariado dos centróides e 95% região confidente para os primeiro dois eixos derivados da análise canônica para a vista lateral do esqueleto de 20 populações *Thrichomys apereoides* do nordeste e sudeste do Brasil..... 98

Figura 58: Análise Canônica Variada considerando razão e curvatura digital para dois grupos de população: um grupo do nordeste e outro do sudeste..... 98

Figura 59 : Forma original de 5 amostras (a)-(c), escolhida randomicamente dos 428 casos, considerando a versão distorcida $\eta=0.1$ (d)-(f), $\eta=0.2$ (g)-(i), $\eta=0.3$ (j)-(l) and $\eta=0.4$ (m)-(o).....	100
Figura 60 : Resultados da Análise Canônica Variada com as versões distorcidas mostradas na figura 59.....	101
Figura 61 : Análise Canônica Variada (a) só energia (b) energia e curvatura, para 1 combinação de ângulo.....	101
Figura 62 : Análise Canônica Variada (a) só energia (b) energia e curvatura, para 2 combinações de ângulos.....	102
Figura 63 : Configuração original de duas amostras coletadas aleatoriamente dos 428 casos de <i>T apereoides</i>	102
Figura 64 : Configuração transladada de duas amostras coletadas aleatoriamente dos 428 casos de <i>T apereoides</i>	103
Figura 65 : Configuração escalonada de duas amostras coletadas aleatoriamente dos 428 casos de <i>T apereoides</i>	103
Figura 66 : Amostra de 428 espécies de <i>T. apereoides</i> , da coleção do Museu do Rio de Janeiro. As amostras usadas foram coletadas das regiões central, ocidental e nordeste do Brasil. Amostras sem o cálculo das distâncias.....	104
Figura 67 : Resultados da análise procrustes para uma amostra de 428 espécies de <i>T. apereoides</i>	104
Figura 68 : Distâncias ajustadas, depois de transladar, escalonar e rotacionar.....	105

Figura 69 : Análise discriminante das amostras de crânio de roedores, as *features* extraídas são curvatura digital.....105

Figura 70 : Análise discriminante das amostras de crânio de roedores, as *features* extraídas são curvatura contínua.....106

Figura 71: Análise discriminante das amostras de mandíbulas de roedores, as *features* extraídas são curvatura contínua. (20 *features*).....107

Figura 72: Análise discriminante das amostras de mandíbulas de roedores, *features* extraídas de curvatura digital.....107

Figura 73: Análise discriminante da espécie do gênero *Atelocynus*: (1) feminino , (2) masculino.....109

Figura 74: Análise discriminante de todas as espécies de canídeos *Atelocynus* e *Cerdocyon* (1) feminino , (2) masculino. (curvatura digital).....110

Figura 75: Análise discriminante de todas as espécies de canídeos *Atelocynus* e *Cerdocyon* (1) feminino , (2) masculino. (20 *features* de curvatura contínua).....110

Figura 76: Análise discriminante de todas as espécies de canídeos mostradas na tabela 24: (1) feminino , (2) masculino. Usando *features* extraídas de curvatura digital.....111

Figura 77: Análise discriminante de todas as espécies de canídeos mostradas na tabela 24: (1) feminino , (2) masculino. Usando *features* extraídas de curvatura contínua. (20 *features*).....111

LISTA DE TABELAS

Tabela 1: Espécies de canideos utilizados na segmentação.....	31
Tabela 2: Comparação dos métodos Complete, Ward e Averege para curvatura contínua, usando o ponto adicional.....	70
Tabela 3: Comparação dos métodos usado para variação morfológica dos crânios das espécimes <i>T. apereoides</i> . (sem a primeira e última curvatura).....	71
Tabela 4: Quantidade de amostras N1 e N2 e sigma.....	72
Tabela 5: Combinação dos métodos usado para variação morfológica dos crânios das espécimes <i>T. apereoides</i> . (sem a primeira e última curvatura).....	73
Tabela 6: Quantidade de amostras N1, N2 e sigma.....	73
Tabela 7: Método Ward usado para variação morfológica dos crânios das espécies <i>T. apereoides</i> . (sem a primeira e última curvatura).....	74
Tabela 8: Quantidade de amostras N1, N2 e sigma.....	76
Tabela 9: Método Ward usado para variação morfológica dos crânios das espécimes <i>T. apereoides</i> . (sem a primeira e última curvatura).....	76
Tabela 10: Quantidade de amostras N1 , N2 e sigma.....	77
Tabela 11: Método Ward usado para variação morfológica dos crânios das espécies <i>T. apereoides</i> . (sem a primeira e última curvatura).....	77

Tabela 12: Localidades do nordeste, centro-oeste e sudeste do Brasil utilizadas nos estudos da variação.....79

Tabela 13: Método Ward usado para variação morfológica dos crânios das espécies *T. apereoides*. (sem a primeira e última curvatura).....80

Tabela 14: Dados das populações mostradas no dendrograma da figura 39.....82

Tabela 15: Resultados de validação encontrados utilizando o algoritmo FSS para curvatura digital em amostras de crânio (*T. apereoides*).....85

Tabela 16: Resultados encontrados para crânios (*T. apereoides*) utilizando método FSS considerando apenas um ângulo.....85

Tabela 17: Resultados de validação encontrados utilizando FSS para curvatura digital e energia.....86

Tabela 18: Método Ward usado para variação morfológica das mandíbulas das espécimes *T. apereoides*. (sem a primeira e última curvatura).....87

Tabela 19: Resultados de validação encontrados utilizando o algoritmo FSS para curvatura digital em amostras de mandíbulas (*T. apereoides*).....90

Tabela 20: Resultados de validação encontrados utilizando FSS para curvatura digital e energia.....90

Tabela 21: Método Ward usado para variação morfológica das mandíbulas das espécies *T. apereoides*. (sem a primeira e última curvatura).....91

Tabela 22: Espécies de canídeos estudadas.....108

Tabela 23: Resultados k-médias para crânios de *T.apereoides*, 1:1 (80.59%) , 2:2 (57.0%) para *features* de energia e curvatura; 1:1 (33.0%) e 2:2 (89.0%) para *features* de curvatura e razão.....113

Tabela 24: Resultados k-médias para crânios de *T.apereoides*, 1:1 (88.0%) , 2:2 (62.58%) para *features* de curvatura contínua.....114

Tabela 25: Resultados k-médias para mandíbulas de *T.apereoides*, 1:1 (92.30%) , 2:2 (70.0%) para *features* de energia e curvatura.....114

Tabela 26: Resultados k-médias para mandíbulas de *T.apereoides*, 1:1 (0.0%) , 2:2 (97.89%) para *features* de curvatura contínua.....114

Tabela 27: Resultados k-médias para canídeos, para *features* de (a) energia e curvatura; (b) curvatura + razão.....114

Tabela 28: Resultados k-médias para crânio de canídeos, 1:1 (22.85%) , 2:2 (85.71%) para *features* de curvatura contínua.....115

Lista de Abreviaturas

CVA = Canonical Variate Analysis

CVRG = Cybernetic Vision Research Group

DC = Digital Curvature

FSS = Forward Sequential Selection

IFSC = Instituto de Física de São Carlos

PCA = Principal Component Analysis

RESUMO

Este trabalho representa um estudo pertencente a área de visão computacional, mais especificamente à morfologia biológica e evolução. A análise de formas é fundamental para a solução de muitos problemas relacionados a visão, cujas técnicas de visão computacional podem ser aplicadas em estudos sobre evolução relacionados a biologia. Neste trabalho prestamos uma atenção especial a complexidade do método de análise de formas, introduzindo uma nova e simples característica (a curvatura digital) para ser representadas por marcos anatômicos (“*landmarks*”). É importante observar que este tipo de representação é amplamente aplicado em problemas relacionados à área de morfologia biológica. O trabalho também está direcionado à extração e seleção de características (*features*) mais informativas obtidas de dados de *landmarks*, analisando vantagens e desvantagens de sua aplicação. As melhores *features* morfológicas extraídas serão usadas como ferramentas para classificar as amostras de roedores *Thrichomys apereoides* da família *Echimydae* e obter esquemas hierárquicos (taxonomias) das espécimes e comparar com taxonomias tradicionais. A distribuição geográfica das amostras é também levada em consideração e é observado que existe um bom acordo entre tal distribuição e os grupos obtidos pela análise discriminante considerando as propriedades morfológicas das amostras. Os principais resultados deste trabalho é que a verificação ao menos para os tipos de dados e problemas considerados, um simples método como a curvatura digital pode conduzir para melhores resultados do que aqueles obtidos pelos tradicionais.

ABSTRACT

This work addresses the application of computational vision to the area of biological morphology and evolution. Shape analysis plays an important role in the solution of many related problems. In this work we pay special attention to the complexity of shape analysis methods, introducing a new and simple feature (the digital curvature) to be estimated from anatomical landmarks. It is important to observe that this type of representation is widely applied in related problems in biological morphology. The current work also addresses the extraction and selection of the most informative features obtained from landmark data, analyzing the advantages and disadvantages of each approach. The best morphological features extracted are used to classify the samples (*Thrichomys apereoides* rodents from the Echimydae family) in order to get hierarchical structures (taxonomy) of the species, which is compared to the traditional taxonomy. The geographical distribution of the samples is also taken into account and it is observed that there is a good agreement between such a distribution and the groups obtained by discriminant analysis considering the morphological properties of the samples. The main results of this work is the verification that, at least for the considered type of data and problem, a simple method such as the digital curvature can lead to better results than those obtained by the traditional, more complex, approaches.

INTRODUÇÃO

Neste capítulo são apresentadas a justificativas do trabalho e a organização da presente tese.

1.1 Considerações Iniciais

O projeto de pesquisa trata da análise de formas de imagens anatômicas para taxonomias das espécimes, em especial 428 amostras de crânios, 223 amostras de mandíbulas de espécimes de roedores *T. apereoides* e 522 amostras o que

equivale a mais 4 gêneros de dimorfismo sexual canídeos, sendo 174 apenas da vista dorsal.

As coleções de mamíferos *T. apereoides* estão no Museu Nacional do Rio de Janeiro (MNRJ), Brasil, e conta com a colaboração do Professor Sérgio Furtado dos Reis (do Departamento de Parasitologia, Instituto de Biologia, Universidade Estadual de Campinas (Unicamp)). E as coleções de crânios de canídeos estão abrigadas nos Museus: Museum of Vertebrate Zoology, Berkeley , CA, (MVZ); Museo Argentino de Ciencias Naturales “Bernardino Rivadavia”, B. Aires, (MACN); National Museum, Smithsonian Institution, Washington, DC, (USNM); Museu Paraense Emilio Goeldi, Belém, (MPEG); Museu de Zoologia da Universidade de São Paulo, (MZUSP); Field Museum of Natural History, Chicago, (FMNH); American Museum of Natural History – New York , NY, (AMNH); Muséum National d’Histoire Naturelle – Paris, (MNHN); e British Museum of Natural History, Londres, (BMNH); Museu da PUC de Minas Gerais, (PUCMG) e conta com a colaboração de Erika Hingst-Zaher do Museu de Zoologia da Universidade de São Paulo e do professor Roberto Marcondes Cesar Junior do Departamento de Ciência da Computação, Instituto de Matemática e Estatística (IME).

A proposta é desenvolver métodos eficientes de morfometria geométrica e comparar com abordagens já existentes e aquelas desenvolvidas por nosso grupo no IFSC .

Considerando que a estrutura morfológica num organismo tem duas variáveis (tamanho e forma) e que a forma pode fornecer maior informação sobre a definição da entidade biológica, os dados em estudo para descrição e análise da variação da forma serão estudados e analisados utilizando “*landmarks*” morfológicos (pontos estes de correspondência em cada estrutura individual que são homólogos entre e dentro das populações), ou seja, são dominantes na forma.

A escolha dos *landmarks* morfológicos se realiza após um processo de obtenção da forma onde a imagem é fotografada em três vistas: lateral, dorsal e ventral, e digitalizada com seus respectivos *landmarks*. Neste processo de digitalização os *landmarks* são inseridos manualmente por biólogos experientes, capazes de selecionar os melhores marcos antômicos da forma biológica. Após definidos e ligados ou

interpolados os “*landmarks*”, é possível realizar a extração de um conjunto de medidas usando a abordagem clássica de Bookstein, baseada na técnica 2D-THIN PLATE spline [Monteiro, R. Leandro; Reis, Sérgio. F (1999)], distância procrustes e warps assim como as novas abordagens desenvolvidas neste trabalho incluindo energia, curvatura digital, ângulos e razões e também as abordagens desenvolvidas por nosso grupo, incluindo Energia Multiescala, Esqueletos Multiescala, Curvatura Multiescala e outras [Bookstein, F. L (1990)].

Por fim, essas medidas serão usadas como entradas para classificar as amostras e obter esquemas hierárquicos (taxonomias) das espécimes a serem comparadas com taxonomias tradicionais.

1.2 - Motivação do trabalho

O grupo de pesquisa em Visão Cibernética (CVRG¹) do IFSC na Universidade de São Paulo tem mais de 9 anos de existência, e representa uma interface entre visão de computador e uma série de áreas relacionadas, especialmente visão biológica e neurociência computacional, mas também incluindo matemática, física, psicologia e filosofia.

O grupo CVRG deve esse sucesso aos seus objetivos de organização em sempre manter esta multidisciplinaridade. Até agora já foram realizados vários trabalhos de relevância em visão biológica e neurociência computacional o que permitiu a formação de diversos estudantes em mestrado e doutorado.

O grupo CVRG conta com uma interação eficaz com vários grupos de pesquisa internacionais e colaborações com pesquisadores das áreas de ciências biológicas que tem contribuído ativamente para o desenvolvimento dessa área em assuntos de grande interesse e pesquisa de ponta.

¹ CVRG- Cybernetic Vision Research Group

Desse modo, a presente proposta se insere nesse contexto, como parte do projeto maior, visando interagir os resultados que serão obtidos com total ênfase em análise e classificação de formas biológicas em especial mandíbulas de ratos (do Departamento de Parasitologia, Instituto de Biologia, Universidade Estadual de Campinas Professor Sérgio Furtado dos Reis) e crânios de canídeos (do Museu MVZ, MACN, USNM, MPEG, MZUSP, FMNH, AMNH e BM) com as abordagens já existentes para análise e classificação de imagens.

Portanto, uma importante finalidade do presente projeto está em comparar abordagens existentes no grupo CVRG com as novas que serão desenvolvidas, usando mais de um classificador, assim como com métodos tradicionalmente usados em morfologia biológica.

1.3 – Objetivo do trabalho

Parte deste trabalho foi investigar um conjunto extenso de técnicas de segmentação, análise de formas, classificação, extração e seleção de características bem como seu desempenho e aplicações em imagens de Anatomia.

O objetivo também foi selecionar um conjunto de medidas, que fossem suficientes para obter informações sobre a forma, especialmente taxonomias a serem obtidas através do uso de técnicas de “*clustering*” hierárquicos.

Finalmente, foi querer obter o melhor desempenho possível com o conjunto de imagens, combinando os melhores efeitos das técnicas avaliadas. Pretende-se com isso aumentar o detalhamento de informações.

1.4 – Organização do texto

Partindo-se das considerações iniciais expostas, este trabalho está dividido em 7 capítulos, os capítulos contêm uma revisão bibliográfica a respeito da

base teórica envolvida no sistema computacional e explicações sobre as técnicas desenvolvidas, bem como os resultados obtidos.

Assim, conceitos sobre formas biológicas, formas em geral, extração e seleção de atributos e reconhecimentos de padrões estão apresentados no primeiros capítulos e a seguir as técnicas computacionais desenvolvidas. Um breve resumo dos capítulos encontrados nesta tese:

- ⇒ Capítulo 2 : Conceitos sobre formas biológicas importantes para a compreensão das técnicas propostas e a análise de formas representadas por “*landmarks*”.
- ⇒ Capítulo 3: Conceitos de Reconhecimento de padrões envolvendo extração de características baseadas nos *landmarks*, seleção dos melhores atributos, conceitos e tipos de classificação.
- ⇒ Capítulo 4: Detalhes dos métodos desenvolvidos durante esta pesquisa, quais métodos utilizados para segmentação e extração das *features*, métodos da distância, análise de deformações e curvaturas.
- ⇒ Capítulo 5: Resultados dos estudos da variação geográfica da espécie *Thrichomys apereoides* usando *clustering* hierárquico e análise discriminante.
- ⇒ Capítulo 6: Conclusão do trabalho.
- ⇒ Capítulo 7: Referências bibliográficas do trabalho.

Formas Biológicas

Capítulo 2

Este capítulo tem o objetivo de apresentar conceitos básicos sobre formas biológicas, importantes para a compreensão das técnicas propostas neste trabalho.

2.1 – Formas na Natureza

A natureza nos proporciona um espetáculo maravilhoso. Espetáculo este que é a sua diversidade de formas e cores que podem ser vistas na figura 1. A natureza é muito generosa com tanta elegância e beleza.



Figura 1 - A natureza com suas formas e cores.

Uma das razões para tal diversidade e beleza são mudanças que ocorrem no clima e também a evolução das espécies. As espécies com características propícias são as que combinam e se adaptam melhor com o novo ambiente devido a alguma mudança no clima e também são as espécies que mais tem chances de sobreviver e competir com outras espécies [L. F. da Costa, Andréa G. Campos and E.T.M Manoel (2001)].

Como já identificado há tempos, os principais fenômenos no mundo biológico relacionam-se com o processo evolutivo das espécies, que envolve intensa cooperação e competição. Fica logo claro que as formas da vida não são acidentais e que um estreito relacionamento entre a forma e a função é verificado qualquer que seja a escala espacial. Por exemplo, um pássaro que depende do voo para capturar tipos específicos de insetos dos quais se alimenta, precisa possuir uma aerodinâmica, determinada pela forma de seu corpo e asa, particularmente adequada à altura e velocidade necessárias para a sua alimentação. E o coração de um beija-flor possui uma conformação que lhe permite bater na frequência necessária para manter o alto metabolismo desta pequena ave.

Enquanto o ambiente e o conteúdo genético são fundamentais na determinação das formas biológicas, os seres vivos também atuam no sentido de modificar o seu próprio ambiente. Desta maneira, as formas dos seres vivos, na plenitude das escalas espaciais, representa importante fator a ser considerado e estudado no sentido de melhor compreendermos a evolução e ecologia dos seres vivos.

De acordo com [Raff, R. A. (1996)] o crescimento da população humana e a conseqüente degradação ambiental aumenta disparadamente a extinção, de forma que ecólogos e biólogos estão tentando estimar apenas quantas espécies ainda têm realmente na terra.

Alguns ecossistemas como florestas tropicais e recifes de corais são extremamente ricos em espécies. Esta grande diversidade que existe apresenta dois grandes atributos. O primeiro é simplesmente o grande número de diferentes espécies e o segundo é a disparidade, ou seja, o número ainda não relacionado de espécies. A disparidade é uma característica forte que nos mostra quão diferentes os organismos podem ser. O grau de disparidade é aproximadamente indicado pelo mais alto nível taxonômico.

Tendo em vista os diversos princípios relacionando forma em biologia, o estudo das formas biológicas, especialmente em combinação com outros tipos de pesquisa tais como taxonomia, é de grande importância e se insere em colaboração para futuras pesquisas na extração de medidas de grande potencial na evolução morfológica que possam comparar ou classificar estas espécies.

O uso de medidas e modelos matemáticos computacionais no estudo das formas biológicas caracteriza-se como uma das principais oportunidades científicas para os próximos anos.

Tais perspectivas têm sido catalizadas por desenvolvimentos na teoria de análise de imagens (veja, por exemplo, [Costa, L. F & Cesar, R. M (2001)] para uma introdução ao assunto) e pelo sempre crescente potencial de computação.

2.2 - Morfometria

A morfometria está em um período de pleno desenvolvimento, com seus métodos sendo desenvolvidos e aprimorados. Este desenvolvimento só foi possível quando no final da década de 80 a metodologia sintetizou descrições geométricas com análise estatísticas.

Um dos maiores responsáveis pela revolução morfométrica foi Fred L. Bookstein em [Bookstein, F. L (1991)] que define morfometria como “o estudo estatístico da covariância entre mudanças de forma e fatores causais”, ou seja, a morfometria estuda as causas das diferenças de forma entre organismos, sejam elas ecológicas ou filogenéticas.

Com a incorporação de conceitos geométricos na análise morfométrica, forma foi definida como “todas as propriedades de uma configuração de pontos que não se alteram por efeitos de tamanho, posição e orientação”, [Bookstein, F.L (1989a)].

Para [Blackit (1965)] morfometria designa métodos que servem para medir a distância de forma entre espécimes e a partir daí construir fenogramas.

Segundo [Monteiro, R. Leandro; Reis, Sérgio. F (1999)] as vantagens destas novas técnicas morfométricas ficarão evidentes com as diferentes maneiras de medir variações no tamanho e na forma dos organismos. Considerando o enorme poder estatístico dessas técnicas e seu amplo potencial de aplicação, estas técnicas serão de grande utilidade para medir a variação do tamanho e forma entre os organismos.

Essas variações podem ser consideradas como a essência da biologia, já que estão associadas aos padrões de evolução fenotípica em grandes escalas do tempo e espaço, às bases genéticas da variação inter e intrapopulacional.

D’Arcy Thompson em [Thompson, W. D’arcy (1952)] não teve sucesso na elaboração de um método estatístico que pudesse mostrar seus conceitos de transformações. O método das coordenadas proposto como técnica de estudo e comparação de formas relacionadas chama a atenção pela beleza gráfica dos diagramas, mas não permite um estudo estatístico da variação da forma.

Vários pesquisadores em morfologia quantitativa desenvolveram métodos que, em geral, conseguiam extrair informações relacionadas à geometria dos organismos, mas que não permitiam o estudo estatístico da variação da forma. [Rohlf, F. J. & Marcus L. F. (1993)], [Tucker, A. S. & Sharpe, P. T. (1999)], [Walker, J. A. (1997)], [Badyaev, A. V. & Foresman, K. R. (2000)].

2.3 – Dados em Morfometria

A confusão até a década de 80 foi porque até então não sabiam o que seriam “dados morfométricos” e isso só foi superado por [Bookstein (1984)] que propôs o método de coordenada da forma, método este que estudava as diferenças de formas em amostras.

Existem vários caminhos para conduzir à análise morfométrica devido à grande diversidade de tipos de dados tipicamente utilizados em estudo. Temos os dados convencionais, que utilizam por exemplo distâncias, ângulos, área, dados que usam da homologia, ou seja, da posição de marcos de forma para forma, dados de contorno, onde a forma pode ser descrita em termos de coordenadas cartesianas de uma sequência de pontos sem correspondência biológica (não possui homologia) e dados que utilizam marcos anatômicos sendo que a homologia é considerada um mapeamento de pontos onde o importante é a correspondência relacionando pontos a pontos e não partes a partes.

Os métodos morfométricos geométricos recentemente desenvolvidos representam uma importante contribuição estatística e geométrica para estudos da forma biológica, ou seja, a forma de um organismo e sua estrutura biológica é um aspecto muito importante para o estudo científico e a forma fornece uma ligação entre o genótipo e o ambiente, elo este que pode ser útil para o processo de seleção se existirem alterações.

A forma biológica pode ser caracterizada por um número de comprimentos de medidas e também pelo arranjo de um conjunto de pontos de *landmarks* (*landmarks* são pontos de correspondência em cada estrutura individual que são homólogos entre e dentro das populações) que pode ser localizado em cada espécie ou por um esboço do objeto. Maiores detalhes sobre *landmarks* podem ser esclarecidos nas próximas seções. Cada um desses tipos de dados morfométricos são associados com métodos particulares para análise estatística e cada método tem sua própria vantagem.

Visto que a escolha do método mais adequado dependerá tanto do organismo estudado, quanto da natureza do problema proposto. Esta escolha vai

dependem de um conhecimento sobre as propriedades de cada método e o tipo de resultado que cada um oferece.

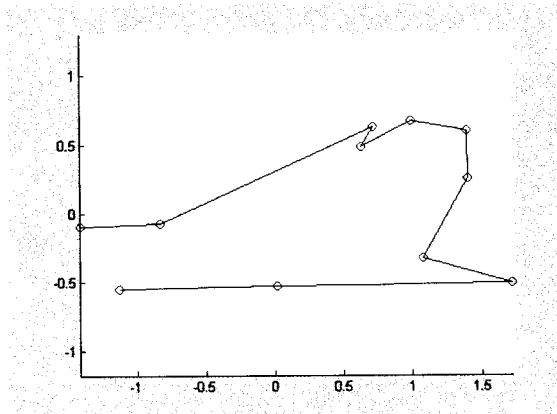


Figura 2: Vista da mandíbula da espécie *T. apereoides* com seus respectivos 11 landmarks.

Na literatura, o estudo da variação da forma tem usado distâncias entre *landmarks* na estrutura biológica, bem como ângulos e razões, que chamamos de morfometria tradicional. Estes estudos tradicionais usam uma combinação linear de distâncias inter-landmarks, normalmente PCA (Análise de Componente Principal), e outras medidas de tamanho como: área, volume e tamanho centróide. Mas estes métodos tem sido substituídos por métodos morfométricos geométricos modernos, com uma clara definição matemática e com uma maior quantidade de informação presente em diferentes tipos de dados estatísticos que construíram a forma.

As vantagens fundamentais da morfometria geométrica sobre a tradicional incluem medidas de diferenças entre as formas, por exemplo, distância Procrustes, ou seja o desenvolvimento de métodos estatísticos avançados para estudo da forma e o desenvolvimento de novas técnicas para representação gráfica.

A morfometria tradicional baseia-se nas aplicações de análise multivariada de padrão de coleções de distâncias como ângulos e razões. Estas variáveis representam apenas parte da informação que pode ser obtida de posições relativas de

landmarks na qual estas medidas são baseadas. Estes métodos não levam em conta a quantidade de informação sobre a relação espacial entre as variáveis medidas.

Esperamos extrair informação biológica de dados morfométricos e que os métodos sejam capazes de fornecer informação adicional com poder estatístico maior.

Existem algumas razões teóricas para usar morfometria geométrica:

⇒ As distâncias inter-*landmarks* usadas pelos métodos tradicionais são limitadas a fornecer informações considerando diferenças entre espécimes que são alinhadas com os eixos. Já as coordenadas associadas com *landmarks* captura variação em todas as possíveis direções [Bookstein, F. L (1986)].

⇒ Para a mesma distância inter-*landmarks* ter a mesma quantidade de informação que a distância de coordenadas *landmarks* têm, considerando variação da forma e tamanho, teria que incluir todas as direções inter-*landmarks* possíveis na análise, mas isto pode gerar um problema de dimensionalidade.

2.4 - Taxonomia

A taxonomia numérica, ao contrário da tradicional, tem um método matemático para explicá-la. Os matemáticos construíram um complexo campo de investigação chamado de análise multidimensional (ou multivariada). Uma parte dele é a análise de agrupamento. Uma parte desta, aplicada aos organismos, recebeu o nome de taxonomia numérica, que é a simples transferência para a biologia de um pequeno número de técnicas matemáticas que têm larga aplicação fora da biologia e que serve para juntar coisas comparáveis de qualquer tipo, como pedras e borboletas. Em suma, é uma taxonomia sem teoria biológica e com métodos importados.

2.5 - Princípios de Formas

O processo inicia-se com a aquisição das formas, por exemplo utilizando-se uma câmera ou microscópio para se obter imagens dos objetos cujas formas desejamos investigar. Após separarmos cada uma das formas das imagens, o que tecnicamente se denomina segmentação, utilizamos métodos computacionais para se obter uma série de medidas relevantes das formas sob análise, medidas estas que são normalmente organizadas em termos estatísticos, tais como histogramas, a fim de se permitir uma visualização mais clara e completa de tendências exibidas pelas medidas.

A forma do objeto é invariante sob a transformação de translação, escala e rotação, ou seja, a forma é toda informação geométrica que permanece quando estes efeitos são filtrados fora de um objeto [Dryden I. L. & K. V. Mardia (1988)].

Neste ponto, podemos realizar diversas abordagens de modelagem, classificação ou puramente interpretação dos resultados. Os diversos dados obtidos no processo como um todo, incluindo as imagens originais, as medidas e os resultados obtidos, são organizadas em bases de dados de modo a melhor se estruturar e facilitar o processamento e acesso às informações.

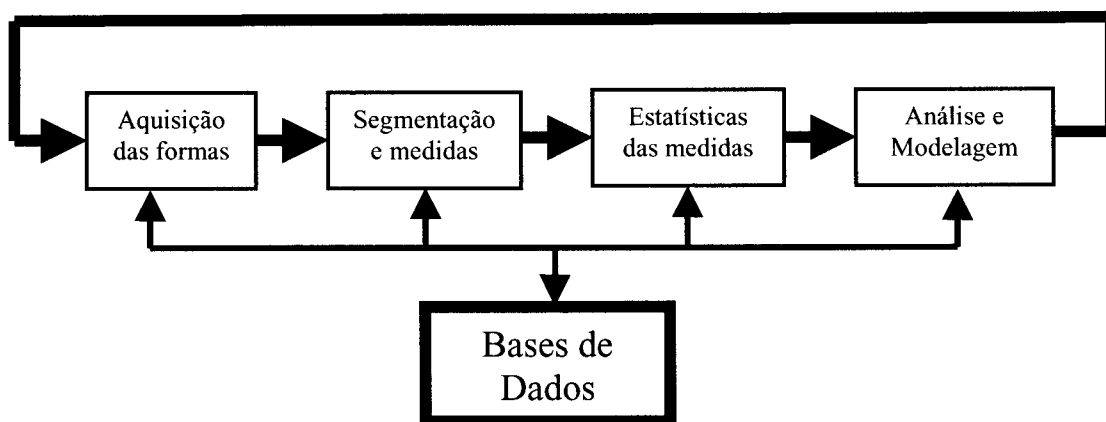


Figura 3 – As atividades normalmente implicadas em análise de formas biológicas. Resultados parciais, incluindo as imagens originais, medidas e resultados, são armazenados e organizados em bases de dados.

Na prática queremos comparar objetos com diferentes formas e medidas, quantificar a distância entre duas formas e métodos para análise estatística da forma.

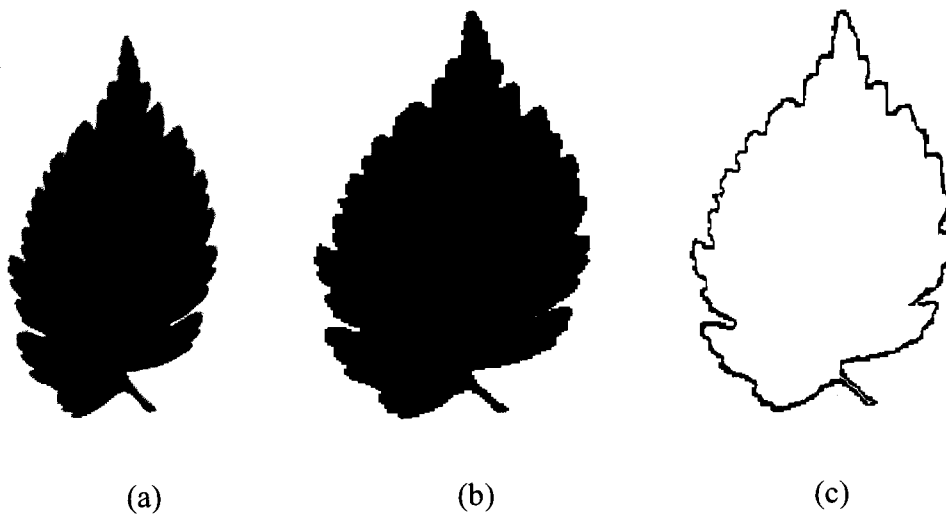


Figura 4 - Processo de segmentação realizado até obtenção do contorno e o passo seguinte inserção de pontos de controle. (A) imagem original ; (b) imagem limiarizada ; (c) extração do contorno.

2.6 - Análise de Formas por *landmarks* ("marcos anatômicos")

Existe uma necessidade muito grande de um método morfométrico que faça a análise de forma independente de seu tamanho pois os resultados dos métodos tradicionais não eram satisfatórios já que a morfometria multivariada continha apenas informações sobre a forma e tamanho das espécimes.

Por isso a análise de formas dentro da área de morfometria é baseada em marcos e todo o trabalho de pesquisa utilizou marcos anatômicos, também conhecidos como "*landmarks*". E no contexto de visão computacional a forma se baseia em contornos, ou seja, envolve a descrição da forma em termos de suas bordas.

A forma do objeto é descrita nomeando a forma que lhe parece mais familiar. Um caminho prático é localizar os pontos de cada objeto. Descrevemos formas localizando um número finito de pontos em cada espécie as quais chamamos *landmarks*.

Um landmark é um ponto de correspondência em cada objeto que adiciona entre e dentro das populações, usualmente pontos de curvatura máxima ou junções entre estruturas, ou pontos localizados em regiões de interesse. Observe na figura 5 abaixo alguns *landmarks* obtidos do crânio (vista lateral) dos roedores da família *Tapereoides*.

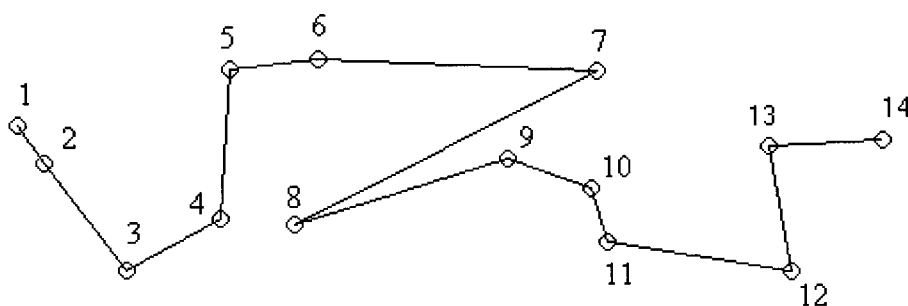


Figura 5 - Exemplo da espécie *T. apereoides* (Crânio lateral) com *landmarks*.

Existem três tipos de *landmarks* em nossas aplicações: anatômicos, matemáticos e pseudo *landmarks*.

⇒ **Landmark Anatômico:** São alguns pontos em organismos estritamente biológicos, por exemplo suturas. *Landmarks* anatômicos designam partes de um organismo que correspondem em termos de derivação biológica e estas partes são chamadas homólogas. São pontos que fazem parte da anatomia biológica do esqueleto sendo observado.

⇒ **Landmarks Matemáticos:** São pontos localizados num objeto de acordo com algumas propriedades geométricas ou matemáticas da figura. Ex: Num ponto de alta curvatura ou num ponto extremo. O uso de *landmarks* matemáticos é particularmente útil em reconhecimento automático e análises.

⇒ **Pseudo-landmarks:** São pontos construídos num organismo, localizados tanto ao redor do esboço ou entre *landmarks* anatômicos ou matemáticos. Os pseudo-landmarks podem também ser úteis quando adicionados em superfícies, onde esses pontos podem ser localizados numa grade regular por cima de cada superfície.

2.7 – Conceitos de Formas

Segundo [Dryden I. L. & K. V. Mardia (1998)] forma é toda informação geométrica que permanece quando efeitos de rotação, translação e escala são filtrados fora de um objeto. Já [Costa, L. F. & Cesar, R. M. (2001)] dizem que forma é um conjunto conectado de pontos.

Na prática, estamos interessados em comparar formas e observar se ambas possuem o mesmo tamanho, qual a distância entre elas e aprender métodos estatísticos para a análise de forma. Neste capítulo sobre formas biológicas vimos que os biólogos desejam analisar formas para estudar como elas mudam durante o crescimento, a evolução, como está relacionada ao tamanho, como pode ser alterada por uma doença e como pode também ser relacionada com o sexo, idade ou condições ambientais.

2.8 – Caracterização de formas

A forma é caracterizada em termos de área, perímetro, diâmetro, número de arestas. Exemplos de algumas formas.

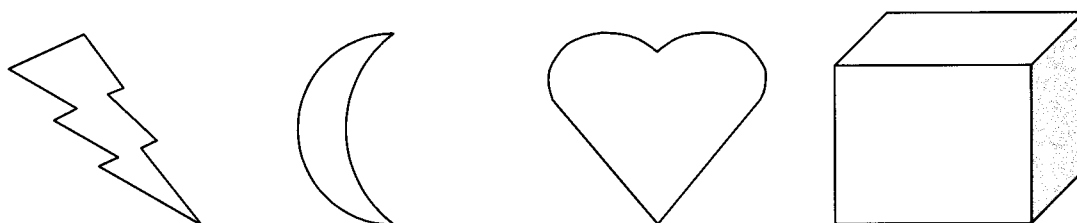


Figura 6 - Exemplos de formas.

O processo de seleção de qual atributo escolher para definir uma forma não é fácil e deve-se levar em conta toda a informação detalhada da forma.

No trabalho de pesquisa vamos analisar formas biológicas e para isso vamos aprender e usar métodos específicos já mencionados, que utilizam “*landmarks*”. No entanto é interessante conhecer os conceitos de formas e quais métodos são utilizados para caracterizar uma forma comum.

Reconhecimento de Padrões

Neste capítulo iremos salientar a importância de extrair e selecionar características para mais adiante efetuarmos a classificação. Também serão apresentados os métodos e bases teóricas de Reconhecimento de Padrões.

3.1 - Relevância

A área de análise de formas e classificação está evoluindo continuamente. Com o intuito de auxiliar na árdua tarefa da avaliação e análise de estruturas, trabalhos vêm sendo desenvolvidos utilizando sistemas computacionais que

empregam técnicas de análise de imagens. É importante salientar a importância da extração e seleção de atributos para efetuar o reconhecimento do objeto.

É importante salientar também algumas dificuldades iniciais quanto a extração, pois existe algumas questões como organizar e visualizar “*features*”, qual *feature* extrair e como medir *features* selecionadas dos objetos, estes são passos iniciais de extrema importância.

3.2 – Extração de Características

Raramente se usa apenas uma medida de cada padrão para que o classificador decida a qual classe pertence. O mais comum é extrair várias medidas de cada padrão² e representá-las através de um vetor, o qual servirá de entrada para o classificador. Por isso é importante organizar os dados de uma maneira correta antes de serem apresentados ou analisados, os dados devem ser armazenados em um formato adequado. Em linhas gerais, os algoritmos de extração criam novas características a partir de transformações ou combinações do conjunto de características originais. As características são extraídas através do processamento da imagem.

3.3 – Extração de Características baseadas nos *landmarks*.

Para este tipo de extração por *landmarks* é necessário primeiramente a transformação da imagem original em representações por contorno. A inserção de *landmarks* deve ser ordenada de maneira igual para que quando duas formas forem comparadas não ocorra problemas de homologias. A inserção destes *landmarks* também

2- Padrão – Um conjunto de características que caracteriza um objeto. Uma família de padrões que possuem propriedades comuns são chamadas classes.

pode ser feita diretamente na forma usando o programa TpsDig [Rohlf, F.J. (1999)] ou também podemos usar o programa PhotoModeler Lite [S. Flach, J. Hargrave, S. Moore, G. Rempel, A. Valji and N. Walford (1999)] no caso de algumas amostras, onde os pares de coordenadas x,y de cada *landmarks* da forma são armazenados. Figura 7.

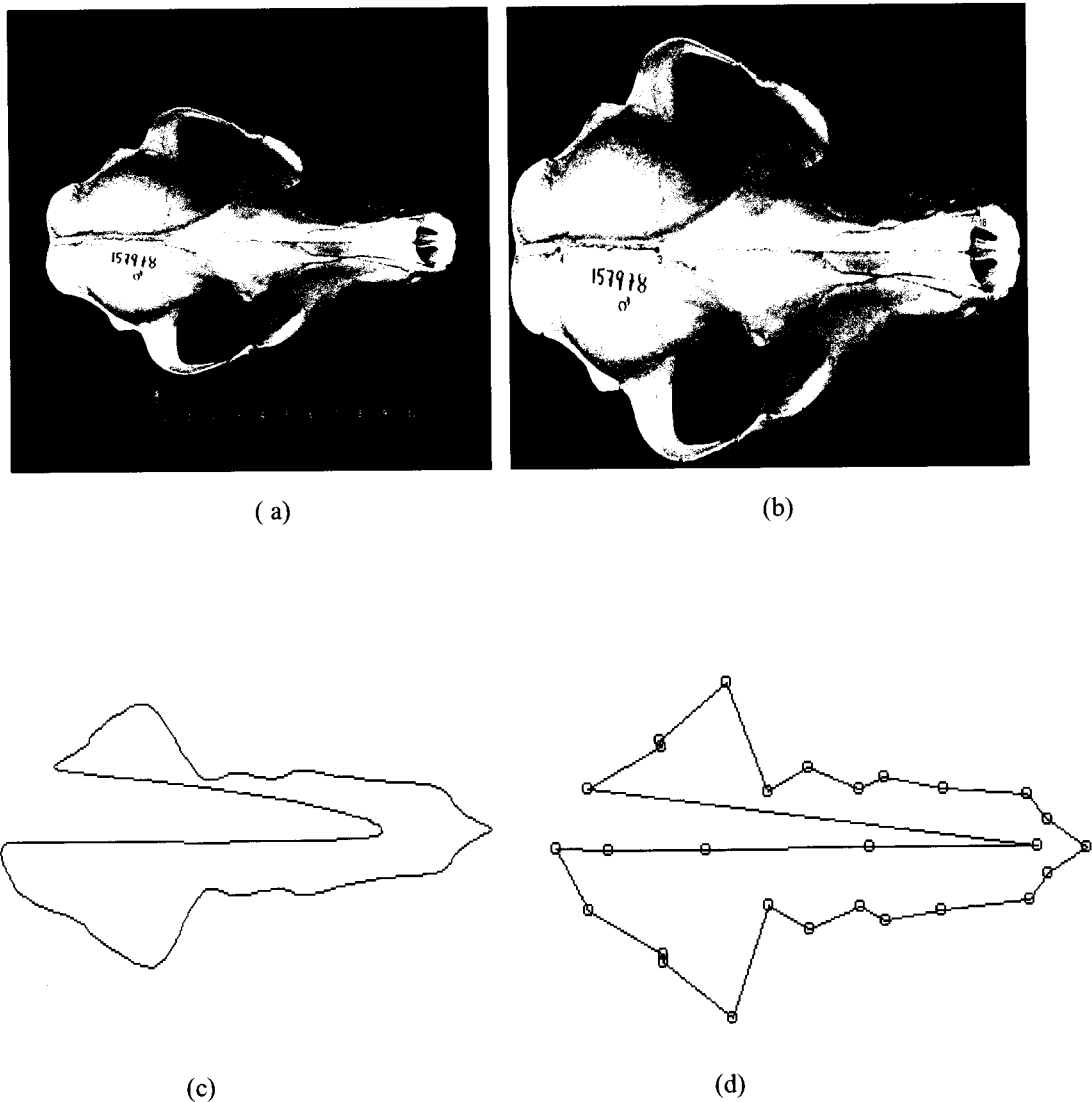


Figura 7: Imagens de canideos (a) Imagem original canideo sem etiquetas; (b) inserção de marcos na forma usando o programa TpsDig; (c) contorno da forma ; (d) Inserção de *landmarks* no contorno da forma.

A sequência de marcos na forma pode ser interpolada construindo uma matriz das coordenadas x e y , assim podemos armazenar uma representação significativa do contorno da forma. Essa matriz vai ser utilizada como entrada de vários algoritmos de extração de características.

3.4 – Seleção de atributos

A seleção de atributos é feita para construir um sistema de classificação o mais exato possível. Um sistema de classificação contém um modelo de como estes atributos descrevem as diferentes classes e espera-se que este modelo permita novos exemplos. Se por alguma razão não desejamos usar todas os atributos avaliados, então devemos realizar a seleção de atributos escolhendo um subconjunto de atributos daqueles disponíveis.

Um dos pontos críticos da análise de formas é a escolha das medidas a serem consideradas. Mais especificamente, o problema é que existe um número virtualmente infinito de possíveis medidas, ao mesmo tempo em que não dispomos de metodologias matemáticas que claramente nos indiquem quais são as mais apropriadas. Assim, a escolha precisa ser feita considerando-se não apenas a experiência anterior dos pesquisadores, mas especialmente os aspectos específicos de cada problema que na maioria das vezes não são tão evidentes quanto parecem. Algumas regras gerais são:

- ⇒ a escolha da característica mais discriminativa;
- ⇒ evitar características com alta correlação;
- ⇒ manter o número de características o menor possível;
- ⇒ escolher características invariantes a transformações geométricas específicas.

De fato, o estudo das formas é frequentemente realizado no sentido de se compreender melhor esses próprios aspectos. Uma maneira de se resolver esse

impasse é partirmos de medidas de forma implicadas por um conjunto de hipóteses iniciais e utilizar os resultados da análise morfométrica assim conduzida de maneira a eventualmente definir-se novas hipóteses e medidas, motivando um ciclo de investigação.

Uma alternativa particularmente interessante para se escolher as medidas envolve o uso de sistemas computacionais onde medidas mais relevantes, no sentido de melhor caracterizarem os fenômenos estudados, são identificadas dentre um vasto número de medidas iniciais, um processo conhecido como seleção de atributos (“*feature selection*”) [W. Siedleki and J. Sklansky, (1989)], [P. Pudil, J. Novovicová, and J. Kittler (1994)], [P. Somol, P. Pudil, J. Novovicová, and P. Paclíck (1999)].

Embora a escolha das medidas seja dependente de cada problema em questão, observa-se que algumas medidas geométricas apresentam melhor potencial para uso geral do que outras. Exemplos de medidas morfológicas particularmente poderosas incluem medidas de comprimento e área, dimensão fractal (normalmente utilizada na caracterização da complexidade e cobertura espacial de formas), energia armazenada na forma, número e tipo dos pontos de saliência (por exemplo vértices e pontos de ramificação ao longo da forma), e distribuição de curvaturas, dentre muitas outras. Algumas dessas medidas são ilustradas ao longo do texto.

Os algoritmos de seleção identificam, segundo determinado critério, o melhor subconjunto do conjunto de características original. Em geral, a seleção de características reduz o custo de medição de dados, e as características selecionadas mantêm sua interpretação física original, permanecendo as propriedades que possuíam quando foram criadas. Já as características transformadas geradas por extração podem prover uma habilidade de discriminação ainda melhor que o subconjunto das características originais.

O processo de seleção de atributos tem como função transformar as medidas extraídas dos padrões (imagens) em características significativas no sentido de maximizar-se a separação das classes. Intuitivamente, o maior conjunto de características deixaria nossos modelos de classificação mais ricos e capazes de classificar exemplos desconhecidos.

Teoricamente, a classificação exata aumentaria com o número de *features* ou dimensões nos dados a serem modelados. Na situação do mundo real este não é o caso, pois a classificação exata pode diminuir em consequência do resultado de se ter também muitas *features*, isto é conhecido como dimensionalidade.

Apesar da importância de seleção de features, não há regras ou procedimentos definitivos para essa tarefa em cada aplicação particular, principalmente quando o número de características disponíveis for grande. Em tais casos, a escolha automática das medidas mais representativas pode se fazer considerando-se o poder de discriminação das mesmas quando aplicadas em formas diferentes.

3.5 - Estado da Arte de Reconhecimento de Padrões

O Reconhecimento de Padrões é de grande importância dentro do processo de análise de formas tendo como objetivo a classificação dos objetos em classes. O trabalho está especificamente concentrado na busca de novas técnicas de análise e classificação de formas de imagens biológicas, mas uma importante ênfase será utilizar a combinação de vários classificadores. Difícil tarefa, pois classificadores diferentes podem dar resultados diferentes. Portanto vamos considerar esquemas de combinação de classificadores.

São componentes de um processo para fazer reconhecimento de padrões: análise discriminante [Estes, (1965)], métodos de correlação, extração de características, pré-processamento, clustering [Fortier, J.J (1964)], [Bonner, R.E. (1964)] e aprendizado não-supervisionado, seguidos de algoritmos de classificação [Nagy, G (1968)]. A razão é porque talvez exista uma vasta coleção de problemas altamente variados e algumas destas técnicas contribuem para a solução de alguns destes problemas.

Um dos passos mais importantes no reconhecimento de padrões é a seleção de um conjunto apropriado de características com bom poder discriminativo. [Allais, D.C (1964)]. Não existe uma solução geral para a escolha de uma ótima característica e obtenção de um ótimo algoritmo de classificação. Na literatura dois tipos são identificados: algoritmos de classificação supervisionada e algoritmos de classificação não-supervisionada.

Os algoritmos de classificação supervisionada associam um padrão a uma classe pré-definida enquanto que na não-supervisionada a definição das classes não está pré-definida.

3.6 – Conceitos de Classificação

A classificação da forma é mais um caso específico da classificação em geral, o que significa que os resultados obtidos pela classificação podem ser transferidos para outras áreas específicas. No geral, classificar significa definir classes ou categorias para objetos de acordo com as suas propriedades [MacQueen, J,(1967)].

3.7 – Tipos de Classificação

⇒ **Classificação Supervisionada** – Amostras conhecidas, como conjunto de treinamento de cada classe de objetos previamente conhecidas, são fornecidas como protótipos para classificação de objetos adicionais. Neste caso as características não são evidentes e também não são especificadas, a classificação é feita por um critério imposto. Este tipo de classificação envolve dois estágios; (1) aprender: estágio onde os critérios e métodos são treinados no protótipo e (2) reconhecimento: o sistema treinado é usado para classificar novos objetos.

⇒ **Classificação Não-Supervisionada** – As amostras de um conjunto de objetos não tem protótipo especificado e característica sugerida, é uma situação encontrada por taxonomistas que adoram classificar seres vivos existentes ao redor do mundo. Quando o esquema de classificação adotado consiste no treinamento de classes obtidas através da maximização de similaridade entre os objetos em cada classe e a minimização de similaridade em diferentes classes, a classificação não-supervisionada é chamada de “*clustering*” e cada grupo de objeto obtido de um “*cluster*”. Ou seja, este tipo de classificação é muito mais difícil do que classificação supervisionada [Patrick, E.A. (1966)].

A classificação é sempre realizada por algumas propriedades tais como: medidas, características, parâmetros, critério de classificação porém não existe uma regra indicando como fazer a melhor escolha.

3.8 - Fases do processo de Reconhecimento de Padrões

A primeira fase é segmentação da imagem ou isolamento na qual cada objeto é encontrado e a imagem é isolada do resto da cena. [Kenneth R. Castleman (1995)]. A segunda fase é chamada extração de características, onde os objetos são medidos, gerando atributos.

A característica é uma função de uma ou mais medidas computadas mas que quantificam alguma característica significativa do objeto. Este processo de extração de características produz um conjunto de características que, tiradas juntas, compreende o vetor de características. Este processo reduz a quantidade de informação comparada com o objeto original.

A terceira fase é a classificação: a saída compreende a classe à qual cada objeto pertence. Cada objeto é reconhecido como sendo um tipo particular e o

reconhecimento é implementado como um processo de classificação. A classificação é baseada unicamente no vetor de características. As três fases são:

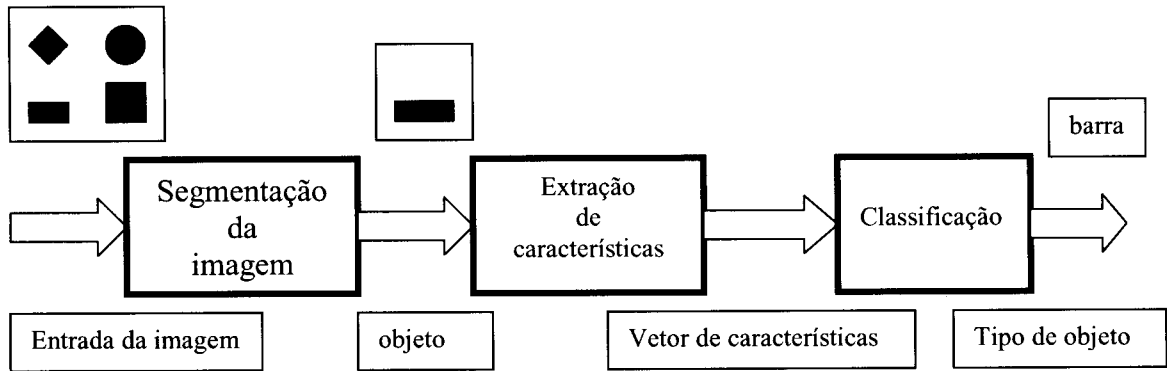


Figura 8 - As três fases do reconhecimento de padrões. [Kenneth, K. Castleman, (1995)].

Capítulo 4

Materiais e Métodos Propostos

Neste capítulo é apresentada a parte prática e são descritos os principais métodos empregados para a realização deste trabalho. Nos itens abaixo serão descritos como foram realizadas as etapas de segmentação, extração, seleção dos principais atributos e classificação.

Para o desenvolvimento do trabalho, foi necessário realizar uma ampla pesquisa sobre as diversas técnicas. Testes extensivos nas imagens foram realizados e pôde-se assim avaliar o melhor desempenho e melhor efeito.

4.1- Fases envolvidas na pesquisa

1. A parte inicial da pesquisa foi fazer a captura das imagens que foram usadas no desenvolvimento do trabalho. As imagens usadas foram de crânios e mandíbulas de roedores (Unicamp) e crânios de canídeos (Museu Smithsonian de Washington). Nesta fase cada crânio foi colocado paralelo ao plano focal sob uma Pixera (Pixera Corporation, Los Gatos, California) um sistema de câmera digital.
2. Após a aquisição, o passo seguinte foi a realização do pré-processamento na imagem, objetivando a eliminação de ruídos, suavização e utilização de técnicas para isolamento de regiões. Segundo [Kenneth R. Castleman (1995)] o propósito de realizar este pré-processamento é separar os objetos de interesse de outras estruturas não muito importantes usando técnicas como filtragem por mediana, média, filtros passa alta (Sobel, Roberts).
3. Após a aquisição e eliminação do ruído, nesta fase vamos determinar "*landmarks*". "*Landmarks*" são pontos dominantes que ocorrem na forma como por exemplo picos de curvatura e fissuras ósseas [Duarte, L.C., Monteiro, L.R., Von Zuben, F.J. & Reis, S.F. (2000)]. Estes marcos anatômicos como também podem ser chamados, serão escolhidos dependendo do problema em estudo e da forma (imagem) a ser marcada. Esta escolha é muito importante pois existe a necessidade de identificar regiões na imagem que representam partes relevantes do objeto. Estes *landmarks* morfológicos foram primeiramente definidos com base nas características topológicas do crânio. Detectando alguns pontos dominantes, muitos dos problemas em análise e reconhecimento de objetos podem ser resolvidos. As coordenadas x, y e z de cada *landmark* foram obtidas usando um software PhotoModeler Lite [S. Flach, J. Hargrave, S. Moore, G. Rempel, A. Valji and N. Walford (1999)].
4. Após determinar os "*landmarks*", é importante extrair características que resultem em alguma informação quantitativa de interesse ou que sejam básicas para discriminação entre classes de objetos e que também sejam suficientes para obter

informações sobre a forma. Várias medidas de cada objeto serão feitas usando a abordagem clássica de Bookstein, baseada na técnica 2D-THIN PLATES [Bookstein. F.L (1991)] assim como as abordagens do CVRG (Cybernetic Vision Research Group), incluindo Curvatura Multiescala [Cesar Jr., R.M. & Costa, L. da F., (1995)], Energia Multiescala, Esqueletos Multiescala, Dilatações Exatas, Dimensão Fractal Multiescala. Realizando-se uma extensa extração de atributos, poderemos escolher quais as melhores medidas para dar entrada no classificador.

Finalmente, clusterizar pelas medidas, ou seja definir ou identificar a classe envolvida na população através do uso de vários classificadores.

Veja o diagrama abaixo.

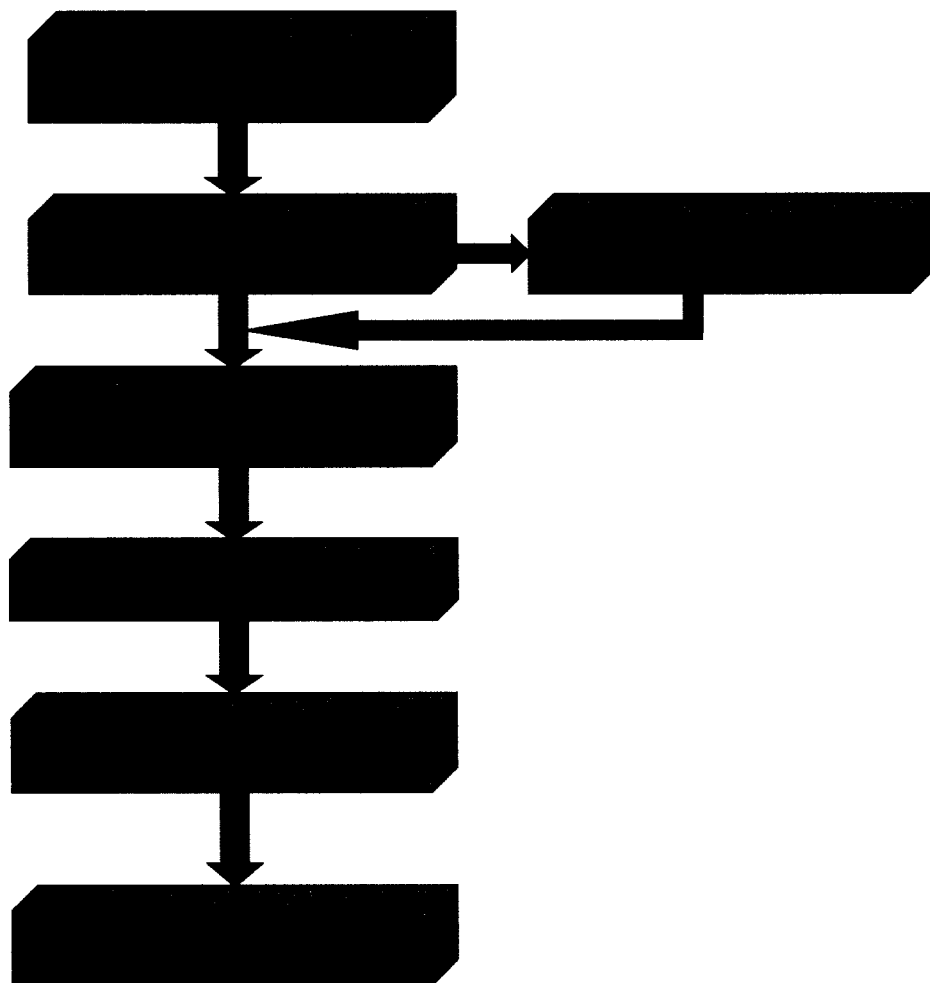


Figura 9 : Diagrama de blocos geral do sistema.

4.2 – Base de dados utilizada

A base de dados utilizada é composta de 428 amostras de crânios e 223 amostras de mandíbulas de espécimes de roedores *T. apereoides* abrigadas nas coleções do Museu Nacional, Rio de Janeiro, Brasil e conta com a colaboração (do Prof. Sérgio Furtado do Reis do Departamento de Parasitologia, Instituto de Biologia, Unicamp) na aquisição das imagens e no acompanhamento dos resultados obtidos na análise.

A base de dados possui também mais 4 gêneros de dimorfismo sexual canídeos, entre eles 33 imagens de *Atelocynos*, 217 de *Cerdocyon*, 6 imagens de *Chrysocyon* e mais 12 de *Lycalopex* com uma resolução em torno de 1016 x 768 pixels e estão abrigadas na coleção do Museu AMNH e conta com a colaboração do professor Roberto Marcondes Cesar Junior do departamento de Ciência da Computação, Instituto de Matemática e Estatística (IME) também no acompanhamento dos resultados obtidos na análise.

4.3 – Softwares utilizados

Para a implementação e testes analisados neste trabalho, usamos uma máquina Pentium II, 450 Mhz, 256 Mb, HD 40Gb e a plataforma usada foi o ambiente operacional “Windows-98” e o software “Matlab 6.0 – Release 12”.

4.4 – Métodos utilizados para a Segmentação

Os métodos que fazem parte do processo de segmentação utilizados foram suavização, binarização, dilatação, erosão e detecção de contorno. Estes métodos foram usados nas amostras de canídeos. Ver tabela abaixo.

Gênero	Espécie	Museu	Quantidade
Atelocynus	Microtis	AMNH	33
Cerdocyon	Thous	USNM	217
Chrysocyon	Brachiurus	FMNH	15
Dusicyon	australis	BM	6
Lycalopex	vetulus	MZUSP	12
		Total	283

Tabela 1. Espécies de canídeos utilizados na segmentação.

Estas espécies estão também divididas em vistas dorsal, lateral, ventral, occipital e mandíbula.

4.4.1 – Suavização

Operações de suavização são usadas para diminuir efeitos espúrios resultantes do processo de aquisição da imagem (ruídos por exemplo). As técnicas descritas neste item foram citadas por Gonzalez e Wintz em [Gonzalez C. Rafael., Woods E. Richard (1987)] e [Gonzalez C. Rafael., Woods E. Richard (2000)].

⇒ **Filtragem Mediana:** É uma técnica de suavização onde cada pixel da imagem final é substituído pelo nível de cinza mediano em uma vizinhança do pixel. O nível mediano m de um conjunto de valores é tal que metade dos valores no conjunto são menores que m e outra metade é constituída de valores maiores que m . Genericamente, uma máscara de tamanho (3x3), pode ser representada da forma indicada na figura abaixo.

W_1	W_2	W_3
W_4	W_5	W_6
W_7	W_8	W_9

Figura 10 – Exemplo genérico de uma máscara 3 x 3.

Considerando $W_i = 1, \dots, 9$, os coeficientes da máscara em questão.

Exemplo de imagens pré-processadas.

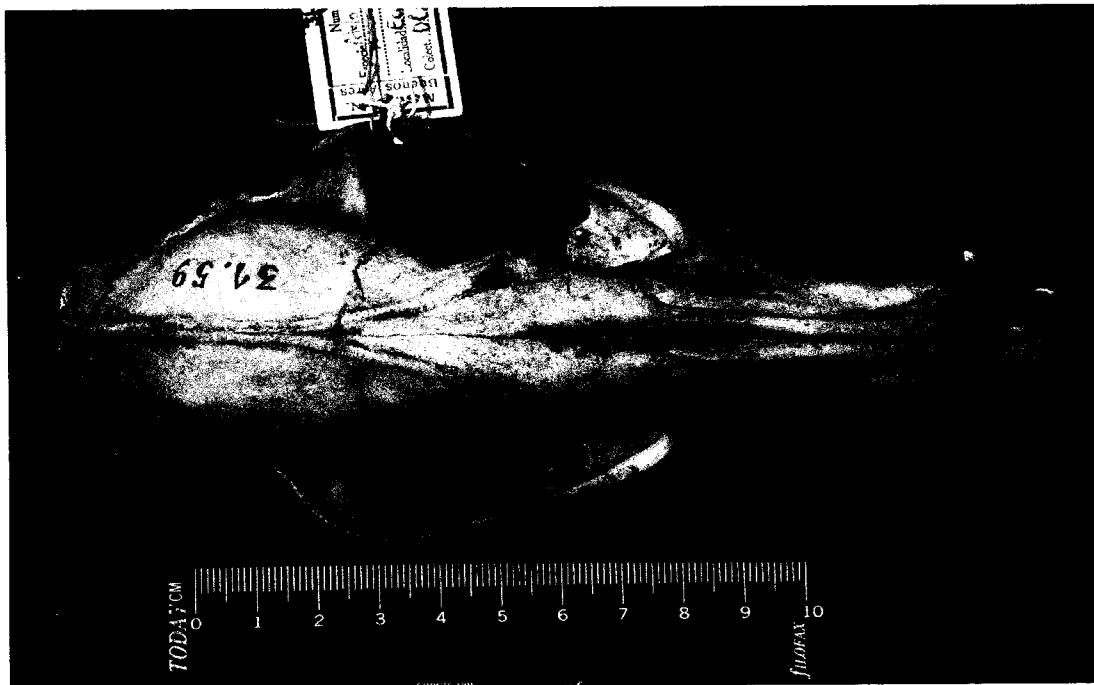


Figura 11- Amostras pertencentes ao grupo canideos. (original)

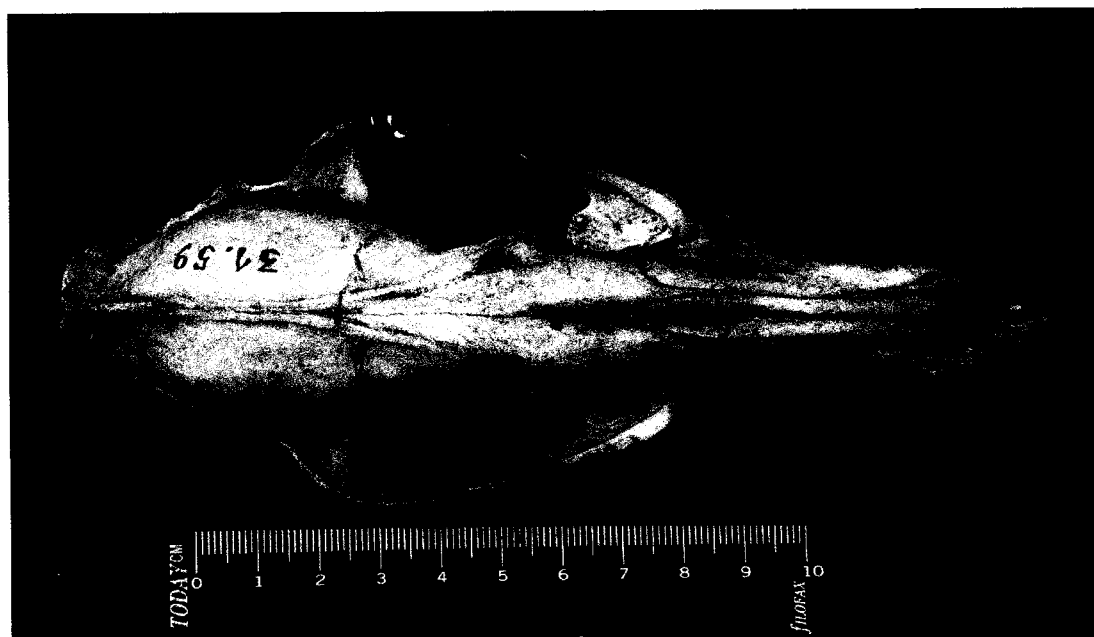


Figura 12 - Amostras pertencentes ao grupo canideos. (original, sem etiquetas)

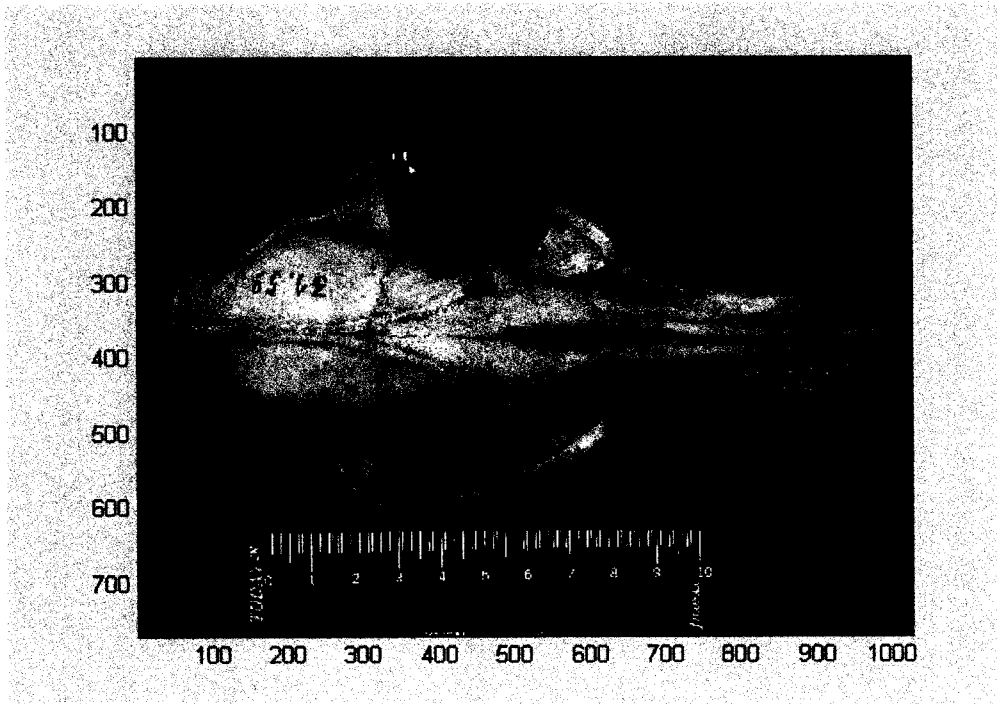


Figura 13 - Amostra canideos (mesma figura 12) mas agora no matlab.

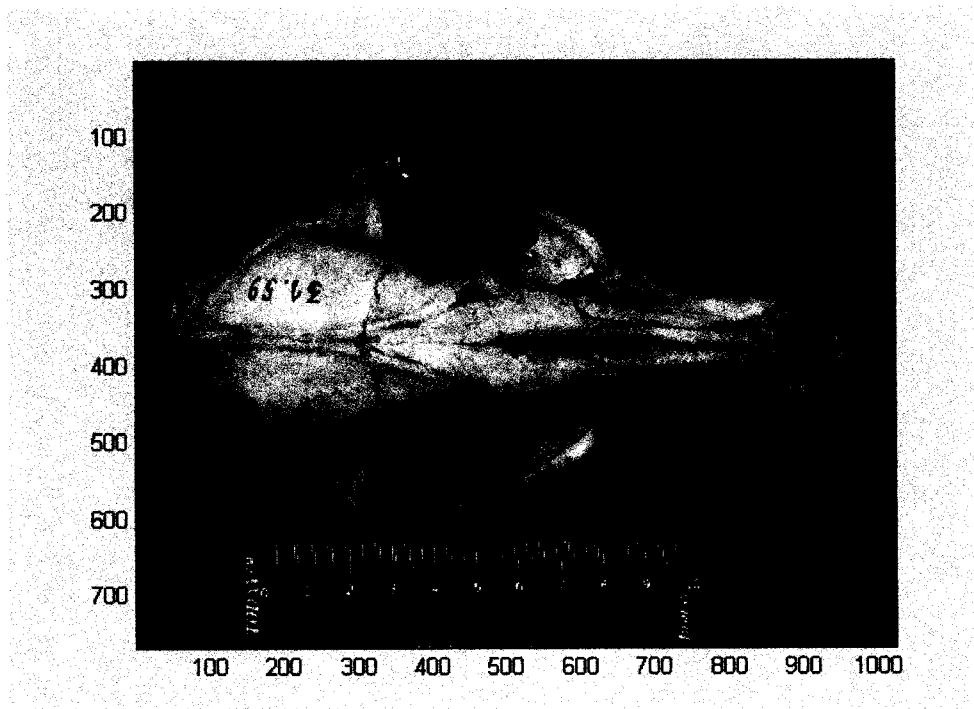


Figura 14 - Amostra canideos, imagem passou por filtro de mediana (2 vezes).

A figura 11 é a imagem original, ainda com suas etiquetas. E a figura 12 é a imagem já pronta para ser binarizada, pois as etiquetas já foram retiradas manualmente, mas antes aplicou-se o filtro de mediana 2 vezes e o tempo de processamento gasto em média em cada imagem foi de 4 mim.

4.4.2 - Binarização

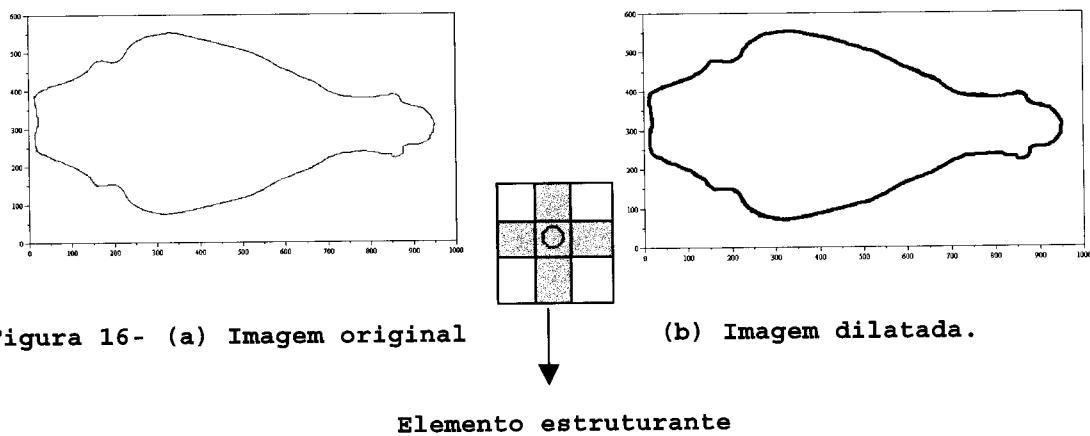
É um tipo de “thresholding”, com o objetivo de transformar a imagem em uma matriz com somente dois níveis de cinza: zero ou intensidade máxima. É determinado um valor de limiar que será utilizado para definir os novos valores dos pixels nas imagens. Os pixels com valores abaixo do limiar são transformados em zero e aqueles com valores iguais ou superiores ao limiar é atribuído o máximo valor de intensidade. A limiar das imagens ficaram entre $L= 30$ e $L=60$ e o tempo de processamento de cada imagem foi 1 mim. Veja figura 15.



. Figura 15 - Amostras pertencentes ao grupo canideos.
(Binarizadas)

4.4.3 - Dilatação e Erosão

A dilatação e erosão fazem parte dos algoritmos de morfologia matemática citadas em [Dougherty (1992)] capazes de resolver problemas de processamento complexos na forma. Esta operação envolve uma imagem binária e um elemento estruturante. Elemento estruturante é uma matriz que percorre toda a imagem binária, este elemento estruturante pode passar pela imagem quantas vezes você achar necessário. Para cada tamanho de elemento estruturante, uma saída diferente para a mesma imagem é processada. A figura 16 mostra um exemplo de uma imagem de dimorfismo para canídeos do gênero *Atelocynus* da espécie *microtis*, onde o elemento estruturante passou três vezes pela imagem.



A dilatação morfológica é aplicada em muitos problemas no processamento de imagens e principalmente na análise de formas biológicas o que permite a implementação de outras técnicas interessantes, pode ser usada também para redução de ruídos [Haralick, R.M., Stenberg, S.R., Zhuang, X. (1987)].

⇒ Dilatação : a imagem binária é dilatada pelo elemento estruturante. Ou seja, o elemento estruturante é deslocado de modo que sua origem passe por cada pixel do objeto quantas vezes forem necessários para eliminar o ruído. Em termos visuais a dilatação regulariza os contornos, elimina estruturas claras.

⇒ A erosão da imagem segue os mesmos processos da dilatação, necessitando também de uma imagem binária e um elemento estruturante, só que ela faz o processo inverso, ou seja, suprime pequenos canais estreitos e atua sobre as estruturas escuras. Tudo isto em função da forma e dimensão do elemento estruturante considerado.

Veja a figura 17, a imagem foi dilatada e erodida 3 vezes e o tempo de processamento foi de 10 s para cada imagem.

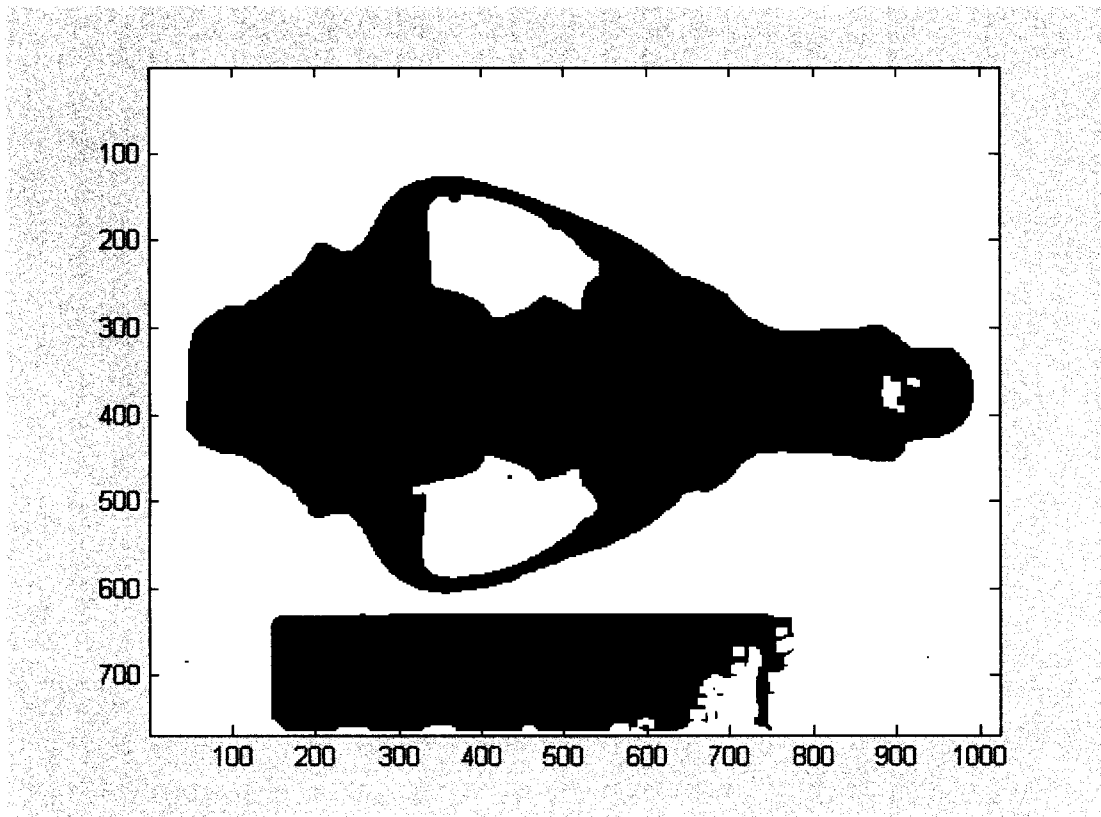


Figura 17 - Amostras pertencentes ao grupo canídeos (fechamento).

4.4.4 - Detecção de contorno

Partindo da definição de borda como uma fronteira entre duas regiões com níveis de cinza relativamente distintos, os algoritmos utilizados para a detecção de bordas são estruturados de forma a detectar as discontinuidades existentes nas transições. O tempo de processamento gasto nesta operação foi 10 s para cada imagem.

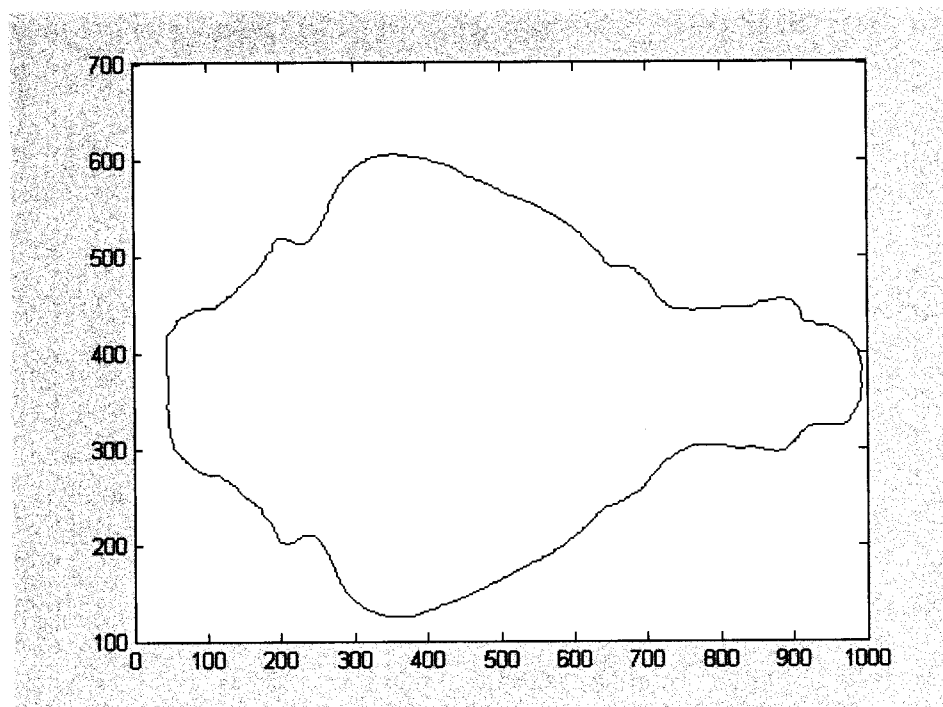


Figura 18 - Contorno da amostras do grupo canideos. (Utilizamos o programa ceguinho para fazer a detecção de contorno).

No geral, o tempo gasto total de processamento para cada imagem foi em média 5 min e 20 s.

4.5 - Métodos utilizados para extração de atributos

Os métodos utilizados neste trabalho para a extração dos atributos serão usados futuramente na classificação.

4.5.1 - Métodos de Distância

Os métodos de distância que usam sobreposição são utilizados para uma comparação de configuração de *landmarks* em espécies. Estes métodos mostram a diferença na forma entre as espécies, problema este importante na morfometria.

O problema de comparação de configurações de *landmarks* em duas ou mais espécies nos leva a usar várias técnicas que caracterizam esta diferença na forma. Uma alternativa é usar métodos de ajuste de sobreposição na qual os *landmarks* homólogos de uma espécie são sobrepostos em outra espécie para otimizar alguma medida de ajuste. Assim, a própria sobreposição é útil tendo uma medida quantitativa do grau de ajuste entre duas ou mais espécies.

⇒ ***Distância Procrustes (Análise Procrustes generalizada - ajuste dos mínimos quadrados)***

A distância Procrustes é a menor soma dos quadrados da distância entre pontos homólogos num par de configuração após a aplicação do processo de superposição [Rohlf, F.J. (1990)].

Para tal processo se realizar, vai ser necessário transladar, proporcionalizar e rotacionar os objetos. Este é um ajuste muito simples que considera parâmetros globais como as diferenças na rotação, translação e escala.

Geometricamente, podemos dizer que estes parâmetros correspondem à superposição de uma espécie na outra colocando-se exatamente um *landmark* sobreposto ao outro *landmark* correspondente da outra espécie. Esta diferença na posição dos *landmarks* corresponde à diferença na forma.

O problema de encontrar a melhor rotação sempre existe, mas para isto foi usado o critério dos mínimos quadrados [Bookstein, F.L (1996a)] e [Bookstein, F.L. (1996b)]. É claro que nem todos estes métodos podem ser usados para todos os tipos de dados. Para espécies que possuem poucos *landmarks* homólogos, recomendam-se outros métodos.

⇒ Passos da implementação do método

Como os procedimentos abaixo relacionados podem apresentar alguma variação dependendo do autor, foi implementado a proposta de [Rohlf, F.J & D.E. Slice (1990)]. A primeira coisa a ser feita foi armazenar as coordenadas x e y dos objetos em matrizes, depois transladar estas coordenadas para um lugar comum, ou seja, colocando na origem do sistema de eixos de maneira que os centróides (pontos médios das configurações) estejam sobrepostos.

Para tal, esta transformação é realizada diretamente pela multiplicação da matriz de configuração original pela matriz identidade I e uma matriz com todos os elementos iguais a $1/p$. O resultado mostra configurações centradas na origem dos eixos e o efeito de posição eliminado.

$$X_i' = (I - P) * X_I \quad (1)$$

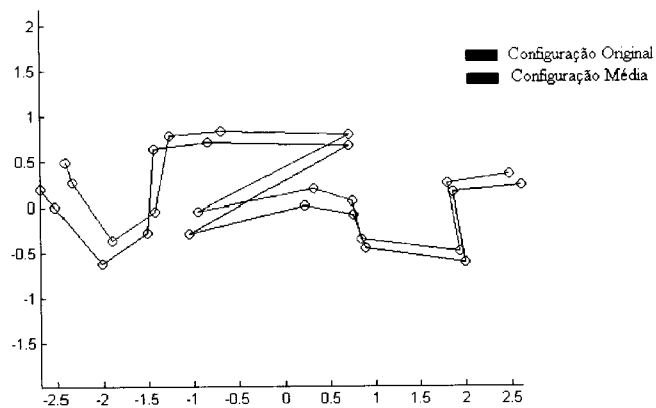


Figura 19: Amostras de espécies de *T. apereoides* a serem comparadas, mostrando as diferenças de tamanho, posição e orientação.

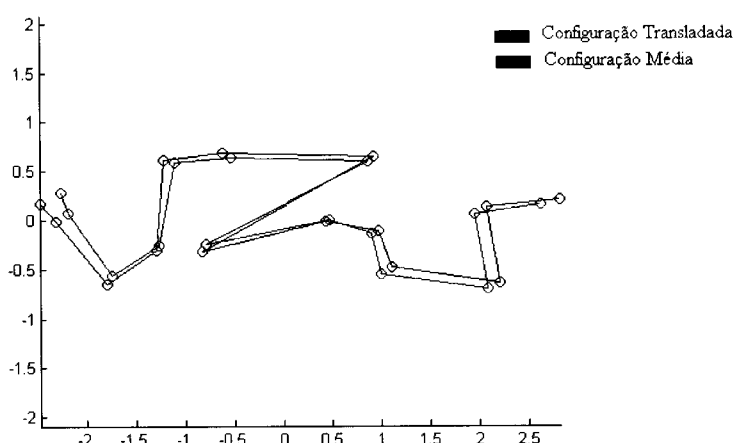


Figura 20: Amostras de espécies de *T. apereoides* centrados na origem após a translação.

O segundo passo é tirar o efeito de tamanho de modo que todas as configurações tenham o mesmo tamanho e por último rotacionar segundo um critério de otimização. Critério este denominado distância Procrustes.

Segundo [Monteiro, R. Leandro., Reis, Sérgio. F. (1999)] a padronização do tamanho requer, antes de tudo, uma definição de tamanho que envolve, por sua vez, a definição de uma medida. O centróide, que representa o centro de massa de uma configuração, é fundamental para a definição do tamanho. Na morfometria geométrica, [Dryden, I.L & K.V. Mardia (1998)] diz que o tamanho é definido em termos de centróide da configuração.

Para isso precisamos calcular as distâncias de cada *landmark* à origem. Com os valores das distâncias de cada *landmark* ao centro de massa, podemos calcular o tamanho centróide.

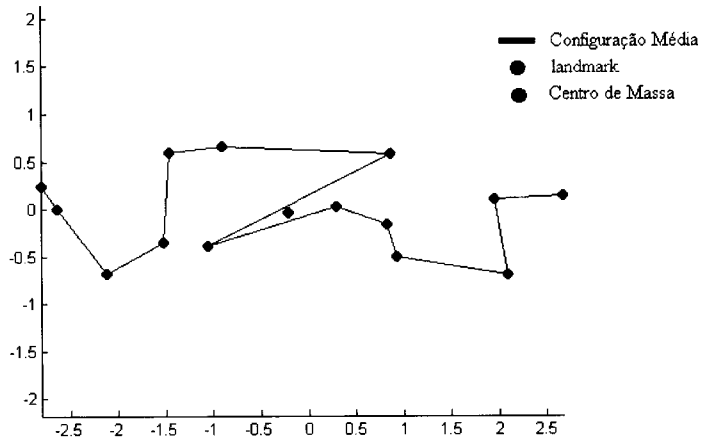


Figura 21: Espécie de *T. apereoides*. Definição dos segmentos de linhas utilizados para o cálculo do tamanho centróide. Distância de cada landmark ao centro de massa.

⇒ Cálculo do tamanho Centróide:

$$S_i^i = \sqrt{\text{Tr}((I - P) X_i X_i^t (I - P))} \quad (2)$$

Uma vez definida e calculada a variável de tamanho a ser utilizada, voltamos ao problema de padronização pelo tamanho. A escala é feita igualando o tamanho do centróide de cada configuração igual a 1. Isto pode ser feito facilmente dividindo o valor de cada coordenada pelo tamanho centróide. Estas novas coordenadas agora definem configurações centradas na origem (centra as coordenadas na média) e cujo tamanho do centróide é igual a 1.

⇒ Transladar e Escalonar as coordenadas para cada um dos N objetos pela equação:

$$X'_i = (I - P) X_i / S_i' \quad (3)$$

Onde I é uma matriz identidade (p x p) e P é uma matriz (p x p) com todos os elementos iguais a 1/p.

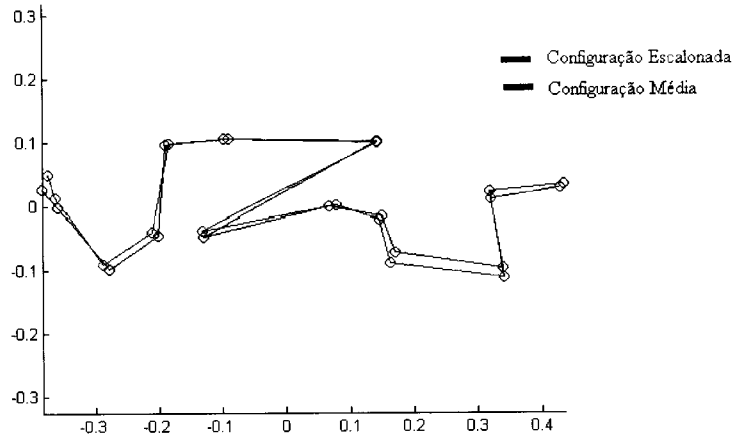


Figura 22: Espécie de *T. apereoides* configuradas após a translação e proporcionalização, continuam centradas na origem.

⇒ O passo seguinte é fazer uma aproximação inicial, fixando uma coordenada já transladada e escalonada e rotacionar todas as outras em relação a esta fixa.

$$Y = X'_i \quad (4)$$

para $i = 2, 3, 4, \dots, n$, usando o ajuste dos mínimos quadrados. Y é a matriz com a média das coordenadas já rotacionadas, transladadas e proporcionalizadas.

Quando rotacionamos, removemos o efeito de orientação nas coordenadas. O objetivo é fazer com que os *landmarks* correspondentes nas configurações estejam o mais próximo possível entre si (respectivamente),

querendo com isso obter a menor soma de distâncias entre *landmarks* correspondentes. Para computar a matriz de rotação antes precisamos calcular H que corresponde a melhor rotação.

$$H = VSU^t \quad (5)$$

A princípio devemos calcular os autovalores e autovetores de uma matriz A qualquer:

$$A = X' i * Y \quad (6)$$

Para calcular U é só fazer $A \cdot A^t$ e para calcular V é só fazer $A^t \cdot A$. E S é uma matriz diagonal com $s_{ii} = \pm 1$. Feito o cálculo do H é só multiplicarmos em relação ao objeto de configuração já transladado e escalonado. Com este resultados já temos as configurações transladadas, escalonadas e rotacionadas em relação à coordenada fixa.

⇒ Para um novo ajuste computar um novo Y com a média das coordenadas transladadas, escalonadas e rotacionadas. Calcular também o resíduo inicial Sr da soma dos quadrados e um fator de escalas p_i para $i = 1, \dots, n$.

$$Sr = n(1 - \text{Tr}(Y \cdot Y^t)) \quad (7)$$

⇒ Repetir o processo dos mínimos quadrados novamente só que agora rotacionar em relação ao objeto médio e multiplicar pelo fator de escala.

$$X_i^* = p_i \cdot X_i \cdot H^T \quad (8)$$

A matriz X_i^* é transladada, escalonada e rotacionada em relação ao objeto médio. E $H = VSU^t$ (*singular value decomposition*).

⇒ Agora é só calcular uma nova média Y^* da configuração resultante X_i^* então faça:

$$X_i^{**} = \rho_i^* / \rho_i \cdot X_i^* \quad (9)$$

Onde :

$$\frac{\rho_i^*}{\rho_i} = \sqrt{\frac{\text{Tr}(x_i^* \cdot y^{*t})}{\text{Tr}(x_i^* \cdot x_i^{*t}) \text{Tr}(y \cdot y^t)}} \quad (10)$$

O propósito é assegurar que:

$$\sum_i \text{Tr}(X_i^{**} \cdot X_i^{**t}) = n \quad (11)$$

⇒ É necessário agora calcular um novo Y^{**} da configuração resultante X_i^{**} , este novo Y^{**} é a média balanceada de X_i^* pela relação ρ_i^* / ρ_i . E também calcular um novo resíduo da soma dos quadrados.

$$S_{r^{**}} = n(1 - \text{Tr}(Y^{**} \cdot Y^{**t})) \quad (12)$$

⇒ E por último verificar se o resíduo está dentro da tolerância esperada ($S_{r^{**}} \leq 10^{-3}$ ou 10^{-4}). Caso contrário $Y = Y^{**}$, $\rho_i = \rho_i^*$, $X_i^* = X_i^{**}$ e volta a rotacionar usando o objeto médio como referência.

Percebe-se que a avaliação da quantidade de variação é um passo importante em qualquer análise morfométrica. Existe também um programa escrito por Rohlf para este fim, chamado TPSSMALL [Rohlf, F.J (1998)]. Observa-se que uma vantagem de usar este método de sobreposição é que os resultados são limitados exclusivamente à própria localização dos *landmarks*.

Este critério permite o desenvolvimento de uma solução direta para duas espécies ou para várias espécies. A desvantagem deste método é que se existe diferença local entre as formas então estas diferenças são distribuídas sobre a configuração inteira dos marcos e em muitos casos tais diferenças locais são prováveis ser de maior interesse.

⇒ **Tamanho centróide**

Um dos atributos na forma é seu tamanho e existem algumas normas que nos permite o seu cálculo. O tamanho centróide é a raiz quadrada da soma das distâncias euclidianas ao quadrado entre cada ponto de landmark e o centro de massa da forma.

$$\|S\| = \sqrt{\sum_{i=1}^n [(S_{x,i} - S_{x,0})^2 + (S_{y,i} - S_{y,0})^2]} \quad (13)$$

Considera-se a média aritmética do centro de massa como sendo:

$$S_{x,0} = \frac{\sum_{i=1}^n S_{x,i}}{n} \quad S_{y,0} = \frac{\sum_{i=1}^n S_{y,i}}{n} \quad (14)$$

⇒ **Média das distâncias**

Uma alternativa de medir o tamanho da forma é tirar a média das distâncias de cada ponto de landmark ao centro de massa:

$$\|S\| = \frac{\sum_{i=1}^n \sqrt{(S_{x,i} - S_{x,0})^2 + (S_{y,i} - S_{y,0})^2}}{n} \quad (15)$$

4.5.2 – Análise de Deformações

Análise de deformações são outras maneiras de decompor as diferenças de forma criando subespaços tangentes cujas características são lineares. Uma deformação ou transformação é um mapeamento de localizações de pontos em uma imagem para novas localizações em outra imagem. As transformações utilizadas para alinhar duas imagens podem ser globais ou locais [Monteiro, R. Leandro., Reis, Sérgio. F.(1999)].

Uma transformação uniforme ou global é aquela que transforma a configuração dos *landmarks* como um todo, multiplicando as coordenadas de cada landmark por uma matriz constante e as transformações localizadas estão restritas a determinadas regiões da configuração, onde cada landmark é multiplicado por diferentes valores, que dependem exclusivamente da proximidade relativa entre os *landmarks* [Bookstein, F.L. (1991)].

Isso significa que deformações ou variações na forma em componentes estatísticos e geométricos, juntos, podem ter interpretações úteis.

As variações na forma de uma amostra são analisadas pela projeção dos indivíduos nos eixos pré-determinados pela configuração de referência. A configuração de referência é qualquer configuração escolhida, mas podemos escolher a configuração de referência como sendo a média das configurações. Vamos expor alguns métodos que fazem estas variações na forma e análise das deformações.

4.5.2.1- Thin -Plate Spline

Thin-plate Spline pode ser visto como uma folha fina de algum material duro com extensão infinita [Bookstein, F.L (1989b)] e [Costa, L.F & Cesar, R.M (2001)]. Para criar o thin-plate spline é necessário ter pontos de

controle, quando isto ocorre e estes pontos de controle específicos ao longo da placa são deslocados, a placa é submetida a uma deformação que minimiza a energia de dobra total E implicada pela transformação, ou seja, os pontos de controle dão forma ao spline. O método permite análises estatísticas da variação em uma ou mais amostras [Bookstein, F.L (1991)].

⇒ Pontos de Controle na Forma

A determinação dos pontos de controle é uma tarefa muito importante. Esta seleção dos pontos de controle é feita a partir das características da amostra em estudo. É necessário que a quantidade de pontos de controle seja suficiente para caracterizar adequadamente as distorções de interesse (veja sobre tipos de pontos de controle na seção 2.6).

⇒ Interpolação Thin-Plate Spline

Para calcularmos os pontos do spline utilizamos as matrizes T , S , M e H . As funções thin-plate spline são obtidas pela distância de coordenadas do ponto em que a função deve ser avaliada para as coordenadas de pontos de referência. À medida que se afasta dos pontos de referência, a função apresenta um comportamento quase linear. Considere uma função básica (g) ρ onde ρ é um valor real não negativo.

$$\rho = \sqrt{x^2 + y^2} \quad (16)$$

A interpolação thin-plate spline é uma função:

$$\varphi(x, y) = a + b_x x + b_y y + \sum_{k=1}^n w_k g(\sqrt{(x - x_k)^2 + (y - y_k)^2}) \quad (17)$$

A matriz T , ($n \times n$) avalia a distância entre cada par de pontos de controle.

$$T = \left[g(\rho_{i,j}) \right]_{n \times n} = \begin{bmatrix} 0 & g(\rho_{1,2}) & g(\rho_{1,n-1}) & g(\rho_{1,n}) \\ g(\rho_{2,1}) & 0 & & \\ & & 0 & \\ g(\rho_{n-1,1}) & g(\rho_{n-1,2}) & 0 & g(\rho_{n-1,n}) \\ g(\rho_{n,1}) & g(\rho_{n,2}) & g(\rho_{n,n-1}) & 0 \end{bmatrix}$$

(18)

Onde:

$$\rho_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

(19)

A matriz (nx3) S_e contém coordenadas x e y dos pontos de controle, e uma coluna inicial de 1. E a matriz M [(n+3)x(n+3)] é feita da composição das matrizes T, S_e .

$$S_e = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_n & y_n \end{bmatrix} \quad M = \begin{bmatrix} T & S_e \\ S_e^t & 0 \end{bmatrix}$$

(20)

A matriz $[(n+3) \times 1]$ H contém coordenadas z (as alturas dos pontos de controle).

$$H = [h(x_1, y_1) h(x_2, y_2) h(x_n, y_n) 000]^T \quad (21)$$

A matriz $C [(n+3) \times 1]$ contém todos coeficientes envolvidos na equação da função do thin-plate spline.

$$C = [W_1 W_2 W_n a b_x b_y] \quad (22)$$

Agora os coeficientes requeridos para interpolar através dos pontos de controle n podem ser facilmente obtidos resolvendo-se a seguinte equação:

$$C = M^{-1} H \quad (23)$$

E a energia de dobramento pode também ser obtida com a equação:

$$E = W^T T W \quad (24)$$

Os coeficientes das equações acima podem solucionar o sistema de equações lineares. Estas equações são aplicadas para análise do thin-plate tradicional (forma plana). Para ilustrar a sequência de deformações do thin-plate, abaixo está um exemplo do crânio da espécie *T. apereoides*, vista lateral.

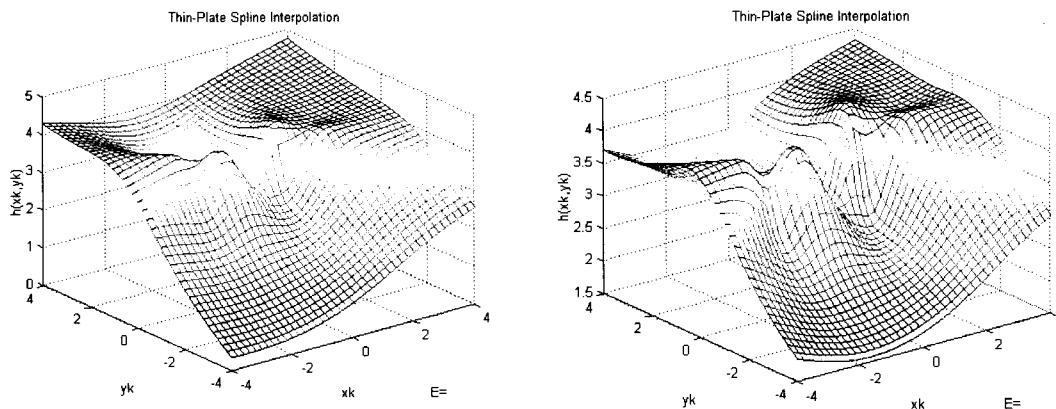


Figura 23: Thin-plate Spline obtido pela adição de pontos de controle. Observe que a deformação na placa é obtida pela energia de dobra cada vez maior.

Podemos também usar um par de Thin-Plate Spline para representar pontos de *landmarks*. Para maiores detalhes sobre a computação dos coeficientes e da energia de deformação consultar [Costa, L.F & Cesar, R.M (2001)].

4.5.2.2 – Warps

Para [Rohlf, F.J (1993)] warps são alguns tipos de transformações que ocorrem na forma que influencia a configuração de marcos como um todo, multiplicando as coordenadas de cada landmark por uma matriz constante. Este método foi desenvolvido por [Bookstein, F.L (1989)] e detalhado na parte de matemática em [Bookstein, F.L (1991)], esta técnica *relative warps* é uma alternativa para análise da variação morfométrica da população em dados de *landmarks*. Esta técnica pode ser aplicada para análise do crescimento do crânio em ratos.

A implementação do método consiste do ajuste de uma função de interpolação do thin-plate spline vista por [Bookstein, F.L (1989b)]. Cada amostra de espécie possui coordenadas x,y de *landmarks* que foram marcadas. A variação ocorrida de uma amostra para outra é pela variância nos parâmetros da função ajustada e isto pode ser mostrado por uma matriz de energia que toma por base

coordenadas dos *landmarks* da amostra de referência. A amostra de referência é uma configuração média dos *landmarks* depois de alguns alinhamentos apropriados da espécie.

O método *warps* é uma técnica desenvolvida por [Bookstein, F.L (1989b)] para análise de variação morfométrica dentro de uma amostra de população baseado em dados de *landmarks*. Passos seguidos para implementação do método:

⇒ O primeiro passo é pegar a matriz de coordenadas x,y de p *landmarks* das espécies. Usaremos então a matriz X para definir a matriz $(2n \times p)$ de coordenadas para todas as espécies na amostra. Então colocar as coordenadas X em uma matriz X_x e uma outra matriz X_y com as coordenadas y .

⇒ Continuando, é necessário arrumar uma configuração de referência, isto pode ser feito usando o método procrustes, que permite construir uma espécie de referência. Este método já foi explicado na seção 4.5.1 sobre distância procrustes. O resultado é colocado em uma matriz que chamamos matriz de referência X_c .

⇒ A partir da matriz X_c , vamos computar a matriz de energia de deformação L_p^{-1} , é importante reforçar que as deformações principais apenas descrevem mudanças na configuração de referência. É conveniente definir a matriz L e sua inversa. Esta matriz é construída a partir da matriz:

$$L = \begin{bmatrix} P_p - Q \\ Q_t - 0 \end{bmatrix}, (p+3) \times (p+3) \quad (25)$$

A matriz P_p é definida pelos valores de $U(r_{i,j})$ para as distâncias entre os *landmarks* na configuração de referência e a matriz Q contém as coordenadas de configuração de referência mais uma coluna de números 1. Esta matriz L segue os mesmos

procedimentos do cálculo do thin-plate spline na seção 4.5.2.1. Após o cálculo de L podemos calcular a sua inversa.

A matriz de energia de deformação, L_p^{-1} está localizada na parte superior esquerda da matriz inversa de L , ou seja, a matriz de energia corresponde apenas à porção $(p \times p)$ da inversa de L .

⇒ A decomposição espectral da matriz de energia L_p^{-1} produz as deformações principais normalizadas.

$$L_p^{-1} = E \Lambda E^t \quad (26)$$

Onde Λ é uma matriz diagonal $(p \times p)$ de autovalores e E é uma matriz $(p \times p)$ de autovetores (linhas= *landmarks*, colunas = autovetores normalizados) da matriz de energia de deformação. Os autovetores desta matriz são chamados deformações principais (“principal warps”), maiores informações podem ser encontradas em [Bookstein, F.L (1989b)].

As magnitudes destes autovalores são inversamente relacionadas à escala. Os autovalores que possuem maior energia estão associados com deformações principais que descrevem mudanças em pequena escala, ou seja, localizadas. E os menores autovalores estão associados com escalas maiores. Pelo menos três dos autovetores vão ter autovalores iguais a zero pois eles correspondem a parte do componente afim que são de escala infinita e não pode ser localizada. Estes autovetores correspondem à redução das dimensões que retiram os efeitos de tamanho, posição e orientação. Estes autovalores iguais a zero e seus autovetores correspondentes podem ser deletados para reduzir Λ para uma matriz $(p-3 \times p-3)$ e E para uma matriz $(p \times p-3)$.

⇒ A matriz Λ é formada por escores que indicam quanto de cada deformação principal é necessário para realizar a deformação desejada, a estes escores chamamos deformações parciais, [Bookstein, F.L (1991)]. Vamos computar o peso da matriz

W , como uma projeção das coordenadas x, y dos n objetos no principal warps. Como as configurações são compostas de coordenadas x, y vamos projetar separadamente as dimensões das coordenadas, produzindo um conjunto de $(2p-6)$ dimensões.

$$W = \left[W_x | W_y \right] \quad (27)$$

Onde:

$$V_x = X_x - 1_n \otimes X_c \quad (28)$$

$$V_y = X_y - 1_n \otimes X_c \quad (29)$$

A matriz V_x é uma matriz ($n \times p$) de coordenadas x das diferenças entre os n objetos e o objeto de referência. 1_n é um vetor coluna de n números 1, e V_y é a matriz correspondente de coordenadas y , finalmente:

$$V = \left[V_x | V_y \right] \quad (30)$$

$$W = \frac{1}{\sqrt{n}} V (I_2 \otimes X_c) \quad (31)$$

Se temos uma amostra, teremos uma matriz W com as projeções dos espécimes nas deformações principais expressando a contribuição relativa de cada deformação principal para a deformação total, [Rohlf, F.J. (1993)].

⇒ Após ter calculado a matriz W com os escores das deformações parciais, cada linha de W corresponde a uma combinação linear do principal warp normalizado. As deformações relativas nada mais são que uma análise de componentes principais da matriz W . O método de deformações relativas também conhecido por “relative warps” foi sugerido por [Bookstein, F.L (1989b)], com uma variação para população intrapopulacional. Agora é só pegar a matriz W e calcular uma matriz de variância-covariância dos escores das deformações parciais e calcular em seguida os autovalores e autovetores. Estes componentes principais são as deformações relativas. Agora temos a deformação relativa de cada combinação de espécimes, e a influência de cada par de coordenada x , y na amostra.

4.5.3 – Curvatura

Segundo [Costa, L.F & Cesar, R.M (2001)] a curvatura é uma das características mais importantes que podem ser extraídas dos contornos porque esta importante medida pode ser estimada ao longo do esboço do objeto. Porém uma vez retirado todos os valores de curvatura, somente os picos de curvatura mais altos são considerados para representação do objeto.

⇒ Contínua

A curvatura contínua caracteriza-se por ser obtida de maneira pontual ao longo do desenvolvimento da forma, descrevendo uma “função” contínua que representa a curva original. A curvatura contínua foi baseada em Fourier. É também utilizada a função Gaussiana para a filtragem mostrada:

$$G_{r,s} = \exp\left(-\frac{(r^2 + s^2)}{2\sigma^2}\right) \quad (32)$$

Sendo $c(t) = (x(t), y(t))$ a representação paramétrica do contorno, a curvatura $k(t)$ de $c(t)$ é definida como:

$$k(t) = \frac{x'(t)y''(t) - x''(t)y'(t)}{(x'(t)^2 + y'(t)^2)^{3/2}} \quad (33)$$

A energia armazenada na forma (E) pode ser obtida como sendo a integral do quadrado dos valores da curvatura ($k(t)$), como:

$$E = \int (k(t))^2 dt \quad (34)$$

O método de curvatura contínua tem uma boa precisão, pois trabalha com todas as informações relativas à geometria que a forma contém.

Os algoritmos a serem implementados para os programas principais, aqui em linguagem natural:

⇒ Análise por Curvatura Contínua

Abre o arquivo;

Monta a Matriz;

Retira o Ruído;

Interpola linearmente;

Cria um Vetor Complexo com coordenadas x, y;

Monta Fourier;

Interpola a função;

Monta Gaussiana para filtragem;

Calcula as derivadas;

Calcula a curvatura $k(i)$;

Calcula a energia;

Gera o histograma;
Gera Dendrograma;
Guarda Dados;

⇒ Curvatura Digital

Curvatura digital é uma nova aplicação e tem como característica a caracterização morfológica e a classificação da forma de *landmarks*. Depois de herdar diversas características originais da curvatura contínua, a curvatura digital fornece a invariância às translações, rotações, deformações locais da forma e é tolerante a escala .

A energia de dobra, uma característica global da forma, pode ser diretamente estimada dos valores da curvatura [Cesar, R.M & Costa, L.F (1997)]. A aplicação destas características para análise morfológica e geográfica para diferenciação de roedores da espécie *Thrichomys apereoides* conduz a resultados encorajadores, indicando uma correspondência entre distribuição geográfica e morfológica [Duarte, L.C., Monteiro, L.R., Von Zuben, F.J. & Reis, S.F. (2000)].

Foi usada a curvatura digital e a energia de dobra como uma descrição dos testes padrões da variação geográfica na forma cranial em *Thrichomys apereoides* (família Echimyidae), um roedor que possui distribuição no nordeste e sudeste do Brasil, Paraguai e Bolívia [Mares, M.A., Willig, M.R & Lacher, T.E (1985)]. Em toda sua distribuição, os *Thrichomys. apereoides* habitam ambientes rochosos em domínios da caatinga e do cerrado em Brasil e o chaco no Paraguai [Moojen, J. (1952)]; [Streilein, K.E. (1982)]; [Éguas et al. (1985)].

A curvatura digital foi usada como um procedimento morfométrico para gerar as características que podem ser usadas como métodos estatísticos para análise multivariada a variação da forma. Basicamente, a curvatura digital pode ser compreendida como a diferença entre os ângulos dos pares dos vetores subseqüentes definidos pela seqüência de landmark. Esta medida associa uma diferença do ângulo a cada ponto do landmark.

A motivação principal para adotar a curvatura digital como uma característica da forma foi que ela possui diversas das características originais da

curvatura contínua como um descritor da forma, por exemplo [Attneave, F. (1954)]; [Costa, L.F & Cesar, R.M (2001)].

Primeiramente, a curvatura digital é invariante à translação e à rotação e pode facilmente ser modificada a fim tornar-se invariante a escala, por exemplo normalizando a forma pela energia ou pelo diâmetro [Costa, L.F & Cesar, R.M (2001)], ou outra normalização conveniente como indicado recentemente por [Llenonart, J., Salat, J & Torres, G.J. (2000)].

Segundo, é uma medida local, no sentido que uma pequena perturbação ou uma variação pequena da forma em um ponto limita o efeito espacial sobre a descrição da forma. Além disso, e ao contrário da curvatura contínua, a curvatura digital é também invariante a escala. Esta medida foi relacionada também aos pontos salientes ao longo da forma [Attneave, F. (1954)], e a soma de seus valores quadrados ao longo da forma é sabida para ser proporcional à energia de dobra armazenada na forma, que pode fornecer uma característica global útil da forma [Young, I.T., Calvert, T.W. & Bowie, J.E. (1974)]; [Cesar, R.M. & Costa, L.F (1997)].

Finalmente, a curvatura digital é uma medida simples que pode ser fácil e obtida economicamente usando operações elementares do vetor. Vejamos a seguir.

⇒ Medidas baseadas na curvatura digital:

Vamos representar os N *landmarks* pelo vetor \vec{v}_i , como se vê na figura abaixo:

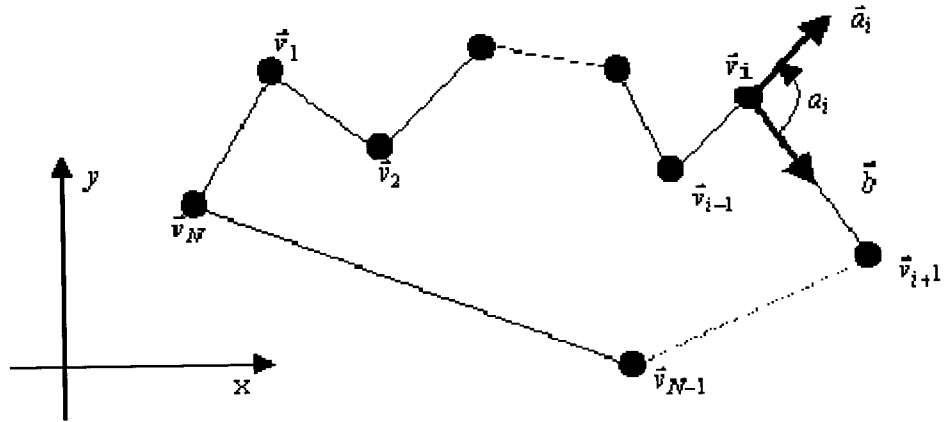


Figura 24: Uma forma genérica representada em termos de N pontos landmarks, vetores e ângulos, definido para cada landmark v_i .

Em cada landmark $v_i = 2, 3, \dots, N-1$, é possível definir dois vetores \vec{a}_i e \vec{b}_i como indicado nas equações (1) e (2), respectivamente.

$$\vec{a}_i = \frac{\vec{v}_i - \vec{v}_{i-1}}{\|\vec{v}_i - \vec{v}_{i-1}\|} \quad (35)$$

$$\vec{b}_i = \frac{\vec{v}_{i+1} - \vec{v}_i}{\|\vec{v}_{i+1} - \vec{v}_i\|} \quad (36)$$

Em particular:

$$\vec{a}_1 = \frac{\vec{v}_1 - \vec{v}_N}{\|\vec{v}_1 - \vec{v}_N\|} \quad (37)$$

$$\vec{b}_1 = \frac{\vec{v}_2 - \vec{v}_1}{\|\vec{v}_2 - \vec{v}_1\|} \quad (38)$$

$$\vec{a}_N = \frac{\vec{v}_N - \vec{v}_{N-1}}{\|\vec{v}_N - \vec{v}_{N-1}\|} \quad (39)$$

$$\vec{b}_N = \frac{\vec{v}_1 - \vec{v}_N}{\|\vec{v}_1 - \vec{v}_N\|} \quad (40)$$

O menor ângulo α_i , entre os vetores \vec{a}_i e \vec{b}_i , é dado conseqüentemente através do produto escalar.

$$\alpha_i = \cos^{-1} \left(\frac{\vec{a}_i \cdot \vec{b}_i}{\|\vec{a}_i\| \|\vec{b}_i\|} \right) \quad (41)$$

Estes ângulos expressam a agudeza da forma em cada ponto do landmark, quanto maior o valor de α_i , mais afiado o vértice. Quando esta medida não indicar a concavidade em cada vértice, esta pode imediatamente ser conseguida usando o produto do vetor entre \vec{a}_i e \vec{b}_i para orientar o ângulo.

$$k_i = \text{sgn}(\vec{a}_i \times \vec{b}_i) \alpha_i \quad (42)$$

onde sgn é a função sigma, retornando 1 para argumentos positivos e -1 para argumentos negativos. Agora o valor k_i definido em cada ponto do landmark v_i pode ser compreendido como curvatura digital da forma nesse ponto [Desbrun, M., Meyer, M., Schroder, P. & Barr, A.H (2000)]. Caso os *landmarks* sejam marcados no sentido horário ao longo da forma, os vértices convexos e côncavos implicarão valores de k_i negativos e positivos, (oposto é verificado para o sentido anti-horário).

Enquanto a integral dos valores quadrados da curvatura ao longo de um contorno contínuo for proporcional à energia de dobra armazenada na forma, uma medida diretamente similar, a energia de dobramento digital ε , pode ser obtida para uma forma do landmark, sendo definido pela soma dos valores quadrados da curvatura digital.

$$\varepsilon = \sum_{i=1}^N k_i^2 \quad (43)$$

⇒ Medida de distância

Vamos definir agora duas distâncias para cada landmark (X_i, Y_i).

$$a = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (44)$$

$$b = \sqrt{(x_{i-1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (45)$$

Agora vamos definir a razão entre as distâncias:

$$C = \frac{a}{b} \quad (46)$$

Observe que a forma original pode ser completamente recuperada a partir dos ângulos e razões.

⇒ Método implementado

Para usar esta técnica foram usados 428 amostras das espécimes de roedores *T. apereoides*, cada espécime possui 14 *landmarks* morfológicos e todos os *landmarks* homólogos nas espécies. Estes landmarks foram definidos pela vista lateral do esqueleto e podem ser vistos na figura abaixo.

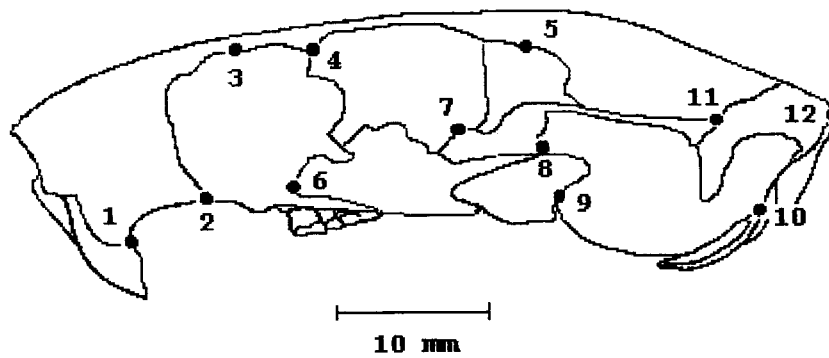


Figura 25 . Landmarks morfológicos definidos pela vista dorsal , lateral e ventral do esqueleto de *Thrichomys apereoides*.

Cada esqueleto foi obtido pela câmera digital (Pixera Corporation, Califórnia) e suas coordenadas x e y de cada landmark foram obtidas usando o software TPSDIG [Rohlf, F.J. (1999)]. Para mais detalhes na definição de *landmarks* [Reis, S.F., Duarte, L.C., Monteiro, L.R & Von Zuben, F.J. (2002a)]; [Reis, S.F., Duarte, L.C., Monteiro, L.R & Von Zuben, F.J. (2002b)].

Também foram considerados neste trabalho duas análises de dados multivariados, clustering hierárquico e análise canônica (CVA). As aproximações hierárquicas para análise de cluster [Anderberg, M.R. (1973)]; [Romesburg, H.C. (1990)] são baseadas na aglomeração progressiva dos indivíduos e sub-cluster considerando distâncias ou medidas de dispersão. Neste trabalho usamos aproximação hierárquica baseada no método Ward, que executa a aglomeração progressivamente, minimizando a dispersão intra-cluster.

O resultado é uma estrutura de árvore chamada dendrograma, expressando o relacionamento entre diversos sub-clusters e observações. As classes definidas para todo o número possível das classes podem convenientemente ser obtidas cortando o dendrograma em pontos específicos da altura. No caso atual, duas classes são obtidas dos dendrogramas, correspondendo à população do norte e do sul. O desempenho conseguido pode apropriadamente ser caracterizado nos termos de classificação matrizes.

A outra análise de dados multivariados é a análise canônica, método que incorpora a informação externa sobre os grupos definidos a priori, onde a estrutura do grupo se levanta naturalmente dos dados em consequência das populações diferentes da amostragem [Krzanowski, W.J (1989)]. Essencialmente, dado p variáveis e g grupos, o i -th grupo ($i=1, \dots, p$) é uma amostra do tamanho n_i ($\sum_{i=1}^g n_i = n$) de uma distribuição multivariada normal $N_p(\mu, \Sigma)$, é possível calcular o grupo médio \bar{x}_i (conhecido como o centróide da amostra), a média total da amostra (\bar{x}) e entre os grupos matrizes de covariância W e B [Ringrose, T.J (1996)].

Então $s = \min(g-1, p)$ eixos canônicos da amostra a_k para $K=1, \dots, s$ são escolhidos sucessivamente para maximizar $a_k^T B a_k / a_k^T W a_k$. A posição do centróide [i -th] da amostra centróide \bar{x}_i projetada na amostra k -th do eixo canônico a_k da amostra é dado por $\bar{z}_{ik} = a_k^T (\bar{x}_i - \bar{x})$ e esta informação é usada para distinguir a natureza das diferenças entre grupos.

As regiões da confiança em torno dos centróide da população para eixos canônicos foram construídas usando o método Bootstrap [Ringrose, T.J (1996)]. O Bootstrap é baseado na $N(\bar{x}_i, \hat{\Sigma})$ distribuição onde \bar{x}_i é o vetor médio para a amostra i -th. A análise canônica foi executada para render autovalores e autovetores de Bootstrap [Von Zuben, F.J., Duarte, L.C., Stangenhuis, G., Pessoa, L.M & dos Reis, S.F. (1998)].

A curvatura digital é uma medida global da forma, a energia de dobra, foram usadas com os procedimentos clustering e análise canônica aglomerando a a variação geográfica entre amostras das localidades diferentes dos roedores *T.apereoides*.

A descrição dos parâmetros da forma revelados pela curvatura digital e pela distribuição espacial das populações é uma primeira etapa na definição de unidades evolutivas independentes na natureza ao nível de espécie, a qual é unidade fundamental da biodiversidade.

A simplicidade das características da forma derivadas dos formalismos da curvatura digital associado com os métodos estatísticos multivariados lineares fornece um estudo promissor da variação morfológica e de suas implicações sistemáticas e evolutivas.

4.5.4 – Análise de Componentes Principais (PCA)

O objetivo da análise é sempre verificar se alguns componentes podem “explicar” a maior parte da variação nos dados originais. Se isso acontece, [Monteiro, R. Leandro., Reis, Sérgio. F. (1999)] diz que a dimensionalidade efetiva do problema é menor do que p , o número original de variáveis. Neste caso, espera-se que os primeiros componentes tenham significado e nos ajudem a compreender a estrutura e o padrão de variação expressa nos dados, permitindo a visualização do problema a partir de um menor número de variáveis.

A análise dos componentes principais transforma um conjunto de variáveis correlacionadas em um novo conjunto de variáveis não-correlacionadas. É uma técnica matemática que não requer a especificação de um modelo estatístico para explicar a estrutura do “erro” [H. Holtelling, (1993)].

⇒ Dedução dos componentes principais

Vamos deduzir os métodos algébricos necessários. Suponha que $X^t = (X_1, \dots, X_p)$ seja um conjunto de variáveis de dimensão p com média μ e matriz de covariância Σ . O nosso problema é encontrar um novo conjunto de variáveis, digamos Y_1, Y_2, \dots, Y_p , que não sejam correlacionadas e cujas variâncias decresçam da primeira para última.



Figura 26: (a) conjunto anterior, (b) conjunto novo de variáveis.

Para [Kenneth R. Castleman (1995)] o vetor médio de X pode ser estimado da seguinte maneira:

$$m_x = \frac{1}{L} \sum_{i=1}^L x_i \quad (47)$$

E a matriz de covariância por:

$$C_x = \varepsilon \left\{ (X - m_x)(X - m_x)^t \right\} = \frac{1}{L} \sum_{i=1}^L x_i x_i^t - m_x m_x^t \quad (48)$$

Agora vamos definir a matriz de transformação linear que gera um novo vetor Y a partir de algum vetor X. ($X \rightarrow Y$).

$$Y = A(X - m_x) \quad (49)$$

Onde A é construído de modo que suas fileiras sejam os autovetores de C_x . O vetor transformado Y, é um vetor randômico com zeros. E a matriz de covariância é relacionada a X por:

$$C_y = A C_x A^t \quad (50)$$

As linhas de A são autovetores de C_x , C_y é a matriz diagonal contendo os autovalores de C_x ao longo desta diagonal. Então λ_k são autovalores de C_y :

$$C_y = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_n \end{bmatrix} \quad (51)$$

A transformação linear A remove a correlação entre as variáveis. Além disso, cada λ_k é a variância de Y_k . Nós podemos reconstruir o vetor X de uma transformada do vetor Y.

$$X = A^{-1}y + m = A^t y + m$$

(52)

Nós podemos reduzir a dimensão do vetor Y ignorando um ou mais dos autovetores que tem autovalores menores. A técnica para tal redução K-L também pode ser chamada de transformada de Karhunen-Loève que pode ser usada para este conjunto de vetores Y [K. Karhunen, 1947)].

4.6 – Métodos de Classificação implementados

Para [Costa, L.F. & Cesar, R.M (2001)] classificar significa atribuir classes ou categorias para itens de acordo com as suas propriedades. A propriedade pode ser explorada através de esquemas de classificação hierárquica, como os considerados para taxonomia das espécies. Classificar envolve a escolha entre vários atributos, distância, critério de classificação e parâmetros, cada escolha conduz possivelmente a diferentes classificações. Não existe uma regra exata indicando como fazer a melhor escolha. Uma das razões porque classificar não é fácil é o número de possibilidades na qual P objetos podem ser classificados dentro de q classes.

⇒ Clustering Hierárquico

Significa que um grupo de M objetos dentro de K classes é classificada progressivamente de acordo com alguns parâmetros, tipicamente a distância ou similaridade entre os vetores de características representando os objetos. Ou seja, para os objetos que forem mais similares um com os outros, a distância entre eles será menor, e são agrupados dentro de sub-classes antes do que os objetos que são menos similares e o processo finaliza uma vez que todos os objetos tenham sido juntados dentro de um cluster. A classificação hierárquica obtida pode ser

representada como uma árvore chamada dendrograma. Um esquema interessante desta técnica é que a estrutura do clustering é definida pela similaridade entre os objetos, deixando claro o relacionamento entre os vários clusters.

⇒ Análise Canônica

O princípio básico em uma análise canônica é desenvolver uma combinação linear em cada um dos conjuntos de variáveis tal que a correlação entre os dois conjuntos seja maximizada. Na análise canônica não existe a distinção entre variável independente e dependente, existem somente dois conjuntos de variáveis em que se busca a máxima correlação entre ambos.

Para [Monteiro, R. Leandro., Reis, Sérgio. F (1999)] a análise das variáveis canônicas fornece uma descrição das diferenças entre grupos especificados a priori em um conjunto de dados multivariados. A especificação dos grupos depende essencialmente do delineamento do problema biológico, sendo que os grupos podem representar indivíduos de populações de uma espécie ou espécies diferentes.

Resultados

Capítulo 5

O objetivo deste capítulo é mostrar os resultados obtidos usando as técnicas propostas, comparando ou combinando estas técnicas afim de alcançar o objetivo do projeto de pesquisa, que era investigar um conjunto extenso de técnicas de segmentação, análise de formas, extração e seleção de características, classificação bem como seu desempenho e aplicações em imagens de anatomia.

Neste capítulo também serão apresentados resultados de comparação ou combinação de medidas, cujos resultados podem ser visualizados através de gráficos ou tabelas. Podemos verificar os resultados à partir da seção 5.1. Abaixo está um organograma dos resultados.

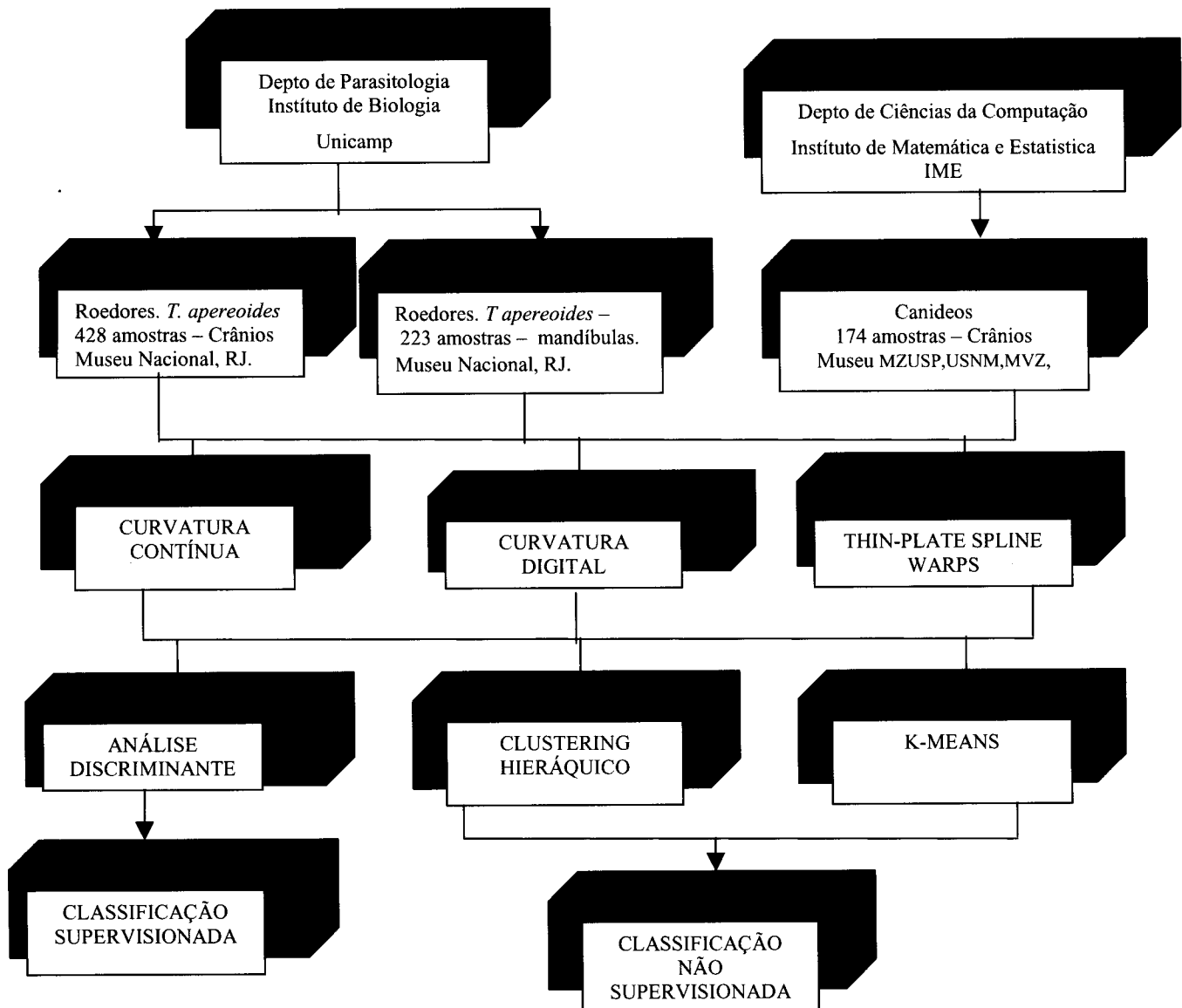


Figura 27 – Organograma dos Resultados.

5.1 – Resultados extraídos por curvatura contínua

5.1 1- Curvatura contínua obtida de crânios de roedores. (*T. apereoides*)

Os gráficos a seguir foram obtidos através do programa que calcula curvatura contínua.

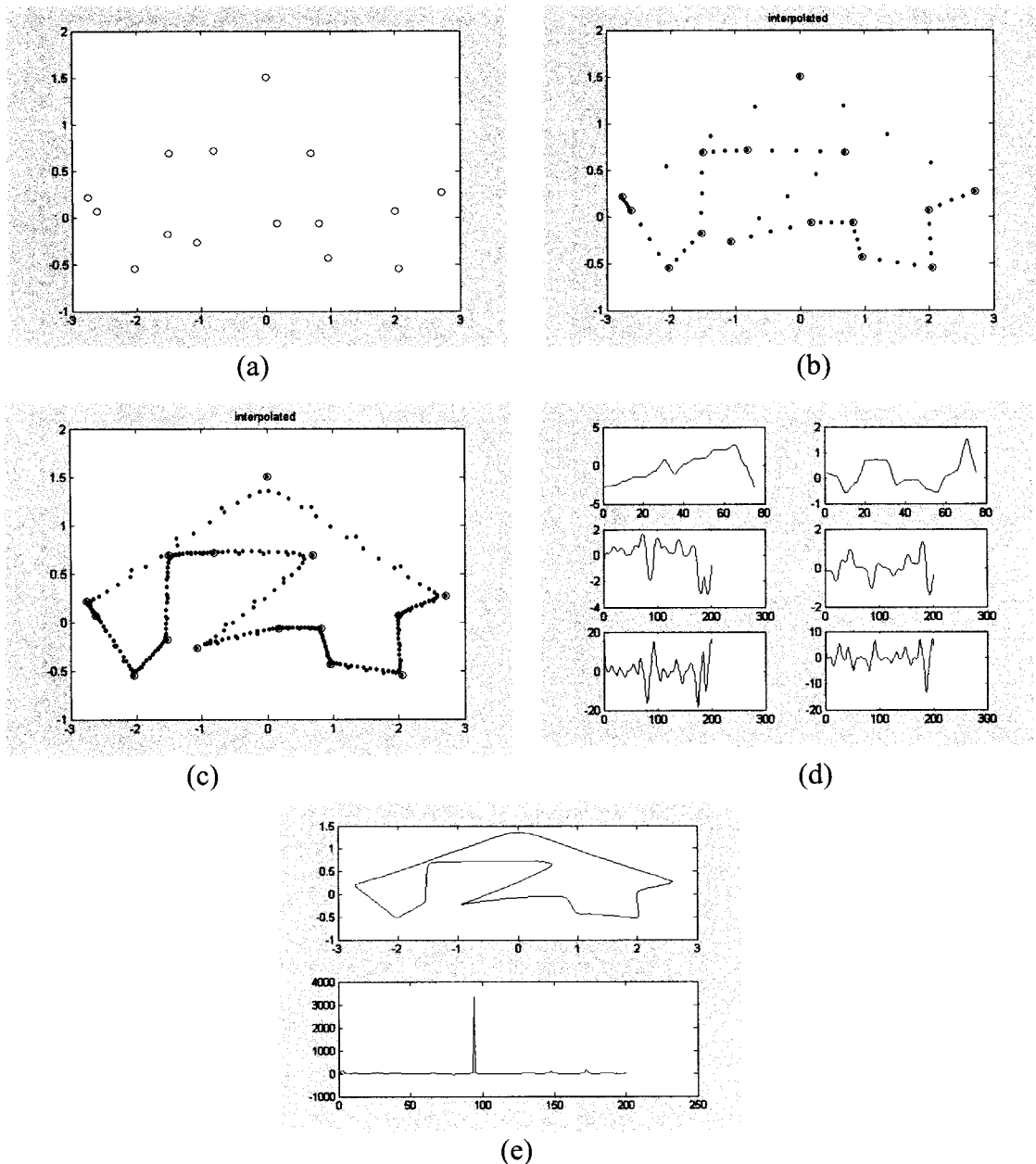


Figura 28:Gráfico das amostras de crânios das espécies *T. apereoides*, (a) landmarks plotados para ponto adicional, (b) landmarks interpolados linearmente, (c) funções gaussianas (d) Funções das coordenadas x e y, cálculo das derivadas e fourier, (e) curvaturas.

- Resultados obtidos das amostras de crânios (*T. apereoides*) para curvatura contínua utilizando um ponto adicional.

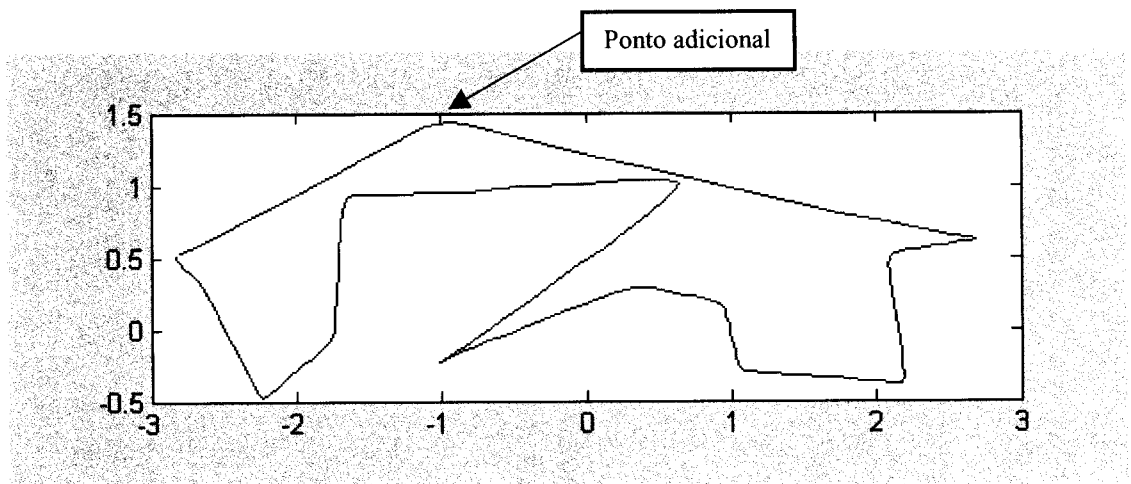


Figura 29: Plot do crânio lateral para curvatura contínua com ponto adicional.

Tabela 2 – Comparação dos métodos Complete, Ward e Averege para curvatura contínua, usando o ponto adicional.		
Quantidade de sigma	N1	N2
12	134	294

Tabela 2: Comparação dos métodos Complete, Ward e Averege para curvatura contínua, usando o ponto adicional.

Nesta tabela temos a quantidade de amostras de cada classe pertencentes a N1 e N2 e o valor de sigma que foram variados até achar a melhor curva de suavização. Neste exemplo usamos sigma 12 para amostra de crânios *T. apereoides*. O uso de um ponto adicional foi para fechar a imagem, por isso foi criado um ponto adicional $x = -1$ e $y = 1.5$ para não passar por cima das outras linhas.

• Resultados da comparação:

Método Complete

E + Curv	1	2
1	7,46%	92,54%
2	13,61%	86,39%

Curv	1	2
1	5,22%	94,78%
2	1,71%	98,29%

Energia	1	2
1	41,79%	58,21%
2	34,36%	65,64%

Método Ward

E + Curv	1	2
1	60,44%	39,56%
2	10,07%	79,93%

Curv	1	2
1	60,44%	39,56%
2	10,41%	79,59%

Energia	1	2
1	60,44%	39,56%
2	47,96%	52,04%

Método Average

E + Curv	1	2
1	99,65%	0,35%
2	100%	0%

Curv	1	2
1	99,65%	0,35%
2	100%	0%

Energia	1	2
1	4,47%	95,53%
2	2,05%	97,95%

Tabela 3 : Comparação dos métodos usados para variação morfológica dos crânios das espécimes *T. apereoides*. (sem a primeira e última curvatura).

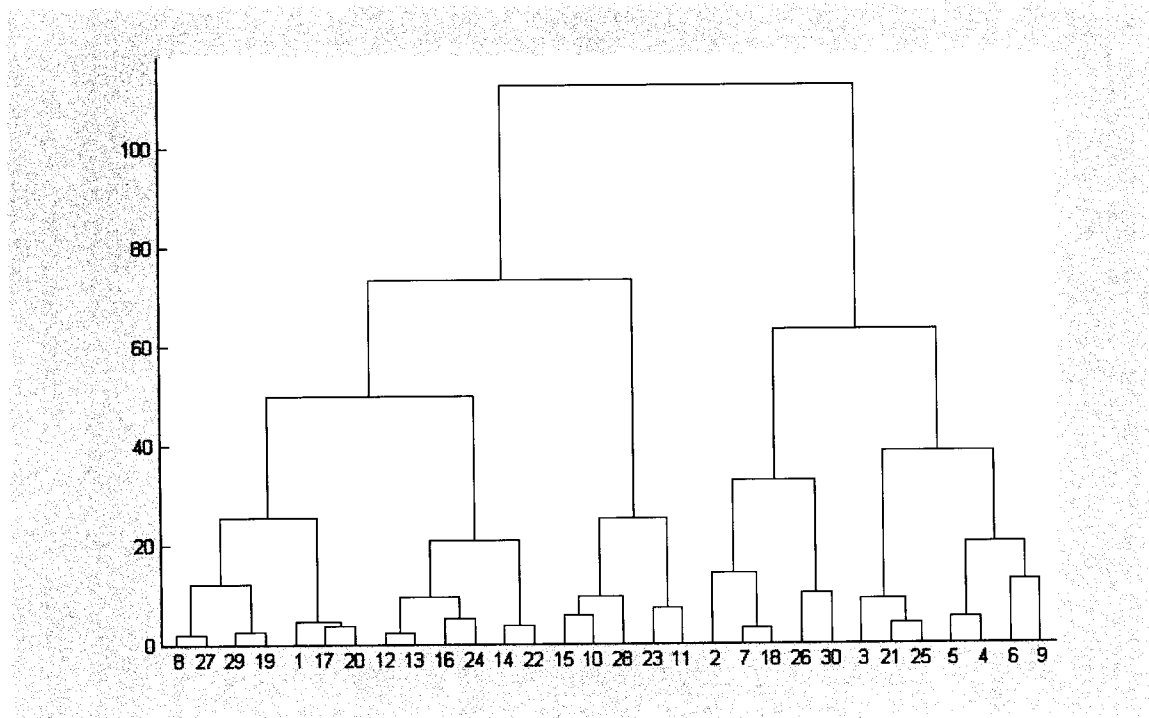


Figura 30: Dendrograma do método ward para curvatura contínua usando ponto adicional para amostras de crânio *T. apereoides*.

As amostras foram testadas nestes três métodos, obtendo melhor desempenho no método Ward.

- Resultado obtido das amostras de crânios (*T. apereoides*) para curvatura contínua retirando o ponto adicional.

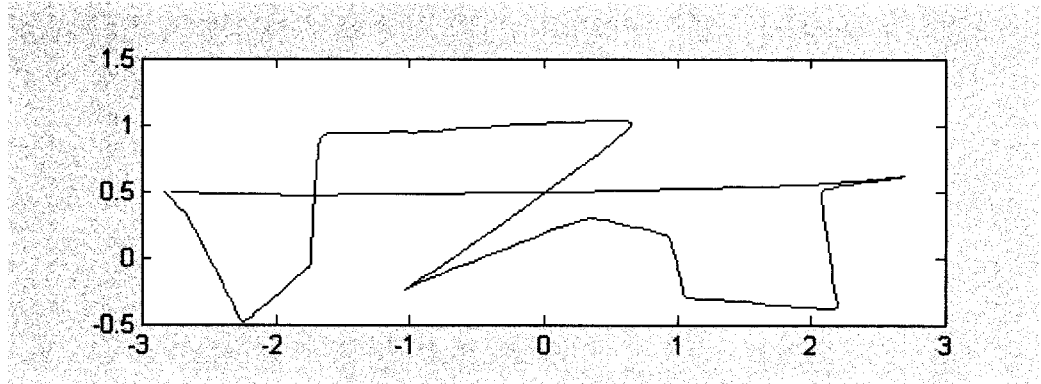


Figura 31: Plot do crânio lateral para curvatura contínua sem ponto adicional.

Tabela 4 – Comparação dos métodos Complete, Ward e Averege para curvatura contínua retirando o ponto adicional		
Quantidade de sigma	N1	N2
15	134	294

Tabela 4: Quantidade de amostras N1 e N2 e sigma.

Nesta tabela variamos o sigma novamente pois agora estamos testando a imagem sem o ponto adicional, apenas fechamos a imagem, então foi necessário escolher outro sigma que melhor suavizasse a imagem.

- Resultados da comparação:

Método Complete

E + Curv	1	2
1	94,77%	5,23%
2	56,47%	43,53%

Curv	1	2
1	2,98%	97,02%
2	25,17%	74,83%

Energia	1	2
1	64,18%	35,82%
2	28,57%	71,43%

Método Ward

E + Curv	1	2	Curv	1	2	Energia	1	2
1	88,06%	11,94%	1	88,06%	11,94%	1	46,27%	53,73%
2	35,03%	64,97%	2	35,03%	64,97%	2	13,95%	86,05%

Método Average

E + Curv	1	2	Curv	1	2	Energia	1	2
1	0%	100%	1	0%	100%	1	54,48%	45,52%
2	9,53%	90,47%	2	3,06%	96,94%	2	18,03%	81,97%

Tabela 5 : Combinação dos métodos usado para variação morfológica dos crânios das espécimes *T. apereoides*. (sem a primeira e última curvatura).

5.1.2 - Curvatura contínua obtida de amostras de mandíbulas de roedores (*T. apereoides*)

Resultado obtido das amostras de mandíbulas (*T. apereoides*) para curvatura contínua.

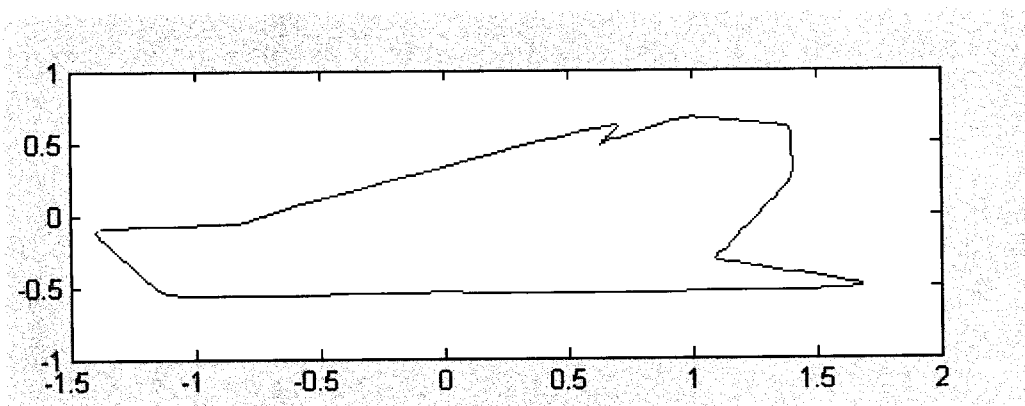


Figura 32: Plot da mandíbula para curvatura contínua sem ponto adicional.

- Resultados da comparação referentes a tabela 8 mostrada abaixo.

Tabela 6 – Comparação dos métodos Complete, Ward e Averege para curvatura contínua retirando o ponto adicional		
Quantidade de sigma	N1	N2
15	13	190

Tabela 6: Quantidade de amostras N1, N2 e sigma.

Como o método complete e average não mostraram bons resultados, nem colocamos na comparação. Usamos somente o método ward que apresenta melhores resultados entre os três testados.

Método Ward

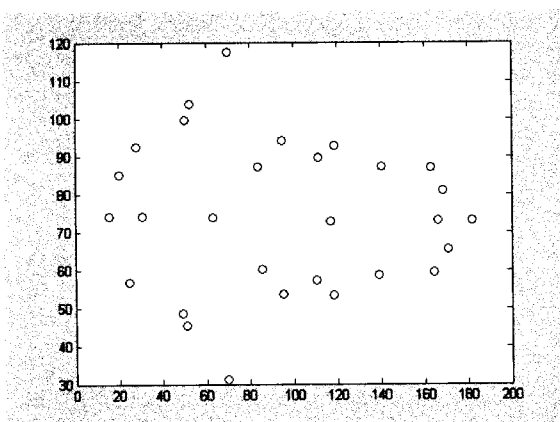
E + Curv	1	2
1	84,61%	15,39%
2	30,53%	69,47%

Curv	1	2
1	84,61%	15,39%
2	30,53%	69,47%

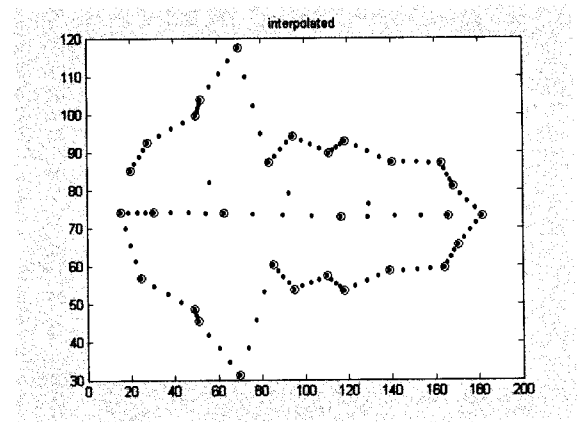
Energia	1	2
1	30,76%	69,24%
2	2,11%	97,89%

Tabela 7 : Método Ward usado para variação morfológica dos crânios das espécimes *T. apereoides*. (sem a primeira e última curvatura).

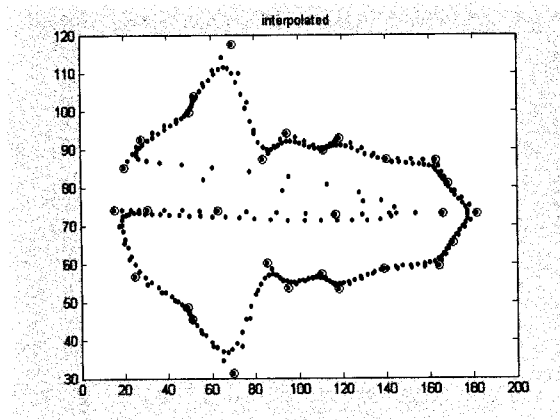
5.1.3 - Curvatura contínua obtida de amostras de crânios. (*canideos*)



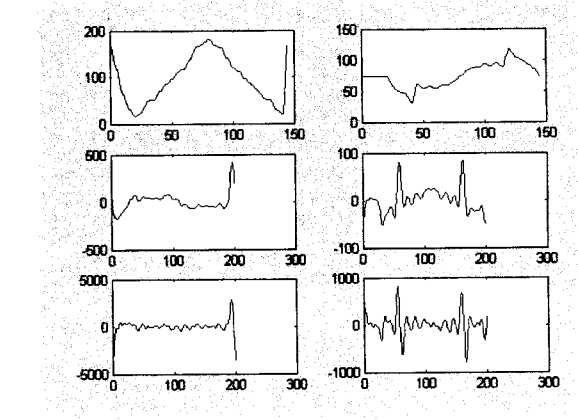
(a)



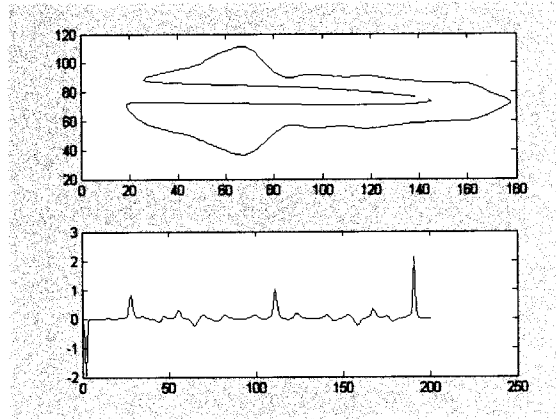
(b)



(c)



(d)



(e)

Figura 33: Amostras de crânios de canídeos), (a) landmarks plotados para ponto adicional, (b) landmarks interpolados linearmente, (c) funções gaussianas (d) Funções das coordenadas x e y, cálculo das derivadas e fourier (e) curvaturas.

- Resultados obtidos das amostras de crânios (*Canídeos*) para curvatura contínua utilizando um ponto adicional.

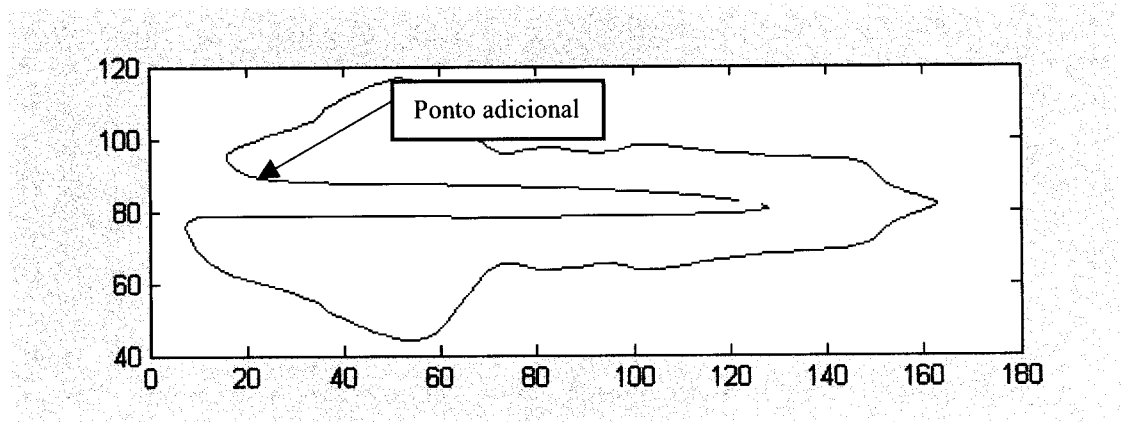


Figura 34: Plot do crânio para curvatura contínua com ponto adicional.

- Resultados de Comparação usando os mesmos dados da tabela abaixo.

Tabela 8 – Comparação dos métodos Complete, Ward e Averege para curvatura contínua colocando o ponto adicional		
Quantidade de sigma	N1	N2
12	85	88

Tabela 8: Quantidade de amostras N1 , N2 e sigma

Método Ward

E + Curv	1	2
1	77.77%	22.23%
2	57.9%	42.10%

Curv	1	2
1	77.77%	22.23%
2	57.9%	42.10%

Energia	1	2
1	94.44%	5.56%
2	89.48%	10.52%

Tabela 9 : Método Ward usado para variação morfológica dos crânios das espécies *T. apereoides*. (sem a primeira e última curvatura).

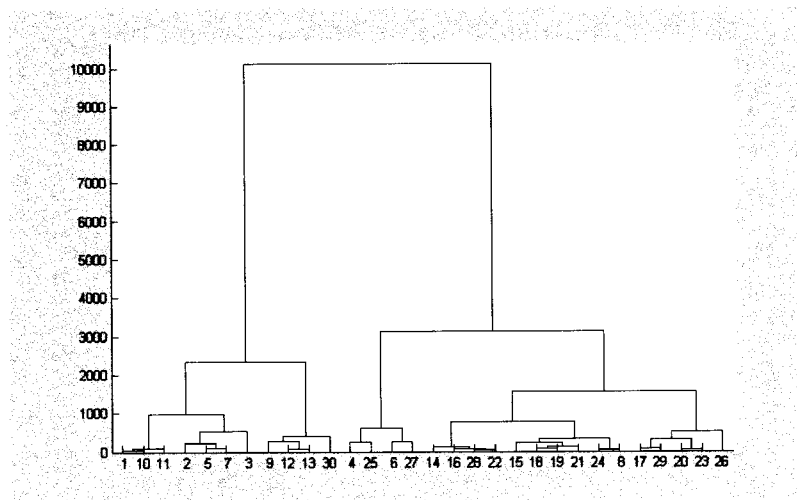


Figura 35: Dendrograma do método ward para curvatura contínua usando ponto adicional para amostras de canideos. (curvatura)

- Resultado obtido das amostras de crânios (Canídeos) para curvatura contínua.

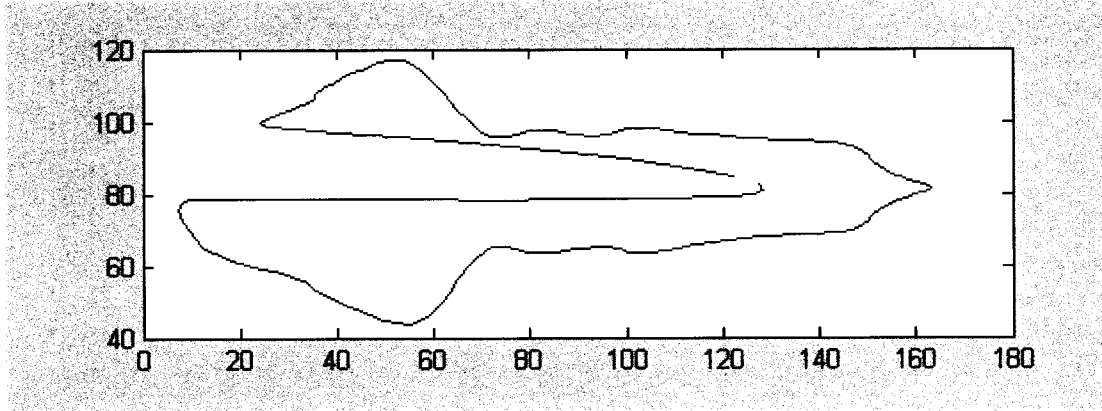


Figura 36: Plot do crânio (canídeo) para curvatura contínua sem ponto adicional.

- Resultados da comparação referentes a tabela abaixo.

Tabela 10- Comparação dos métodos Complete, Ward e Average para curvatura contínua retirando o ponto adicional		
Quantidade de sigma	N1	N2
15	85	88

Tabela 10: Quantidade de amostras N1, N2 e sigma.

Método Ward

E + Curv	1	2	Curv	1	2	Energia	1	2
1	72.22%	27.78%	1	72.22%	27.78%	1	11.11%	88.89%
2	57.9%	42.10%	2	57.9%	42.10%	2	0%	100%

Tabela 11 : Método Ward usado para variação morfológica dos crânios das espécimes *T. apereoides*. (sem a primeira e última curvatura).

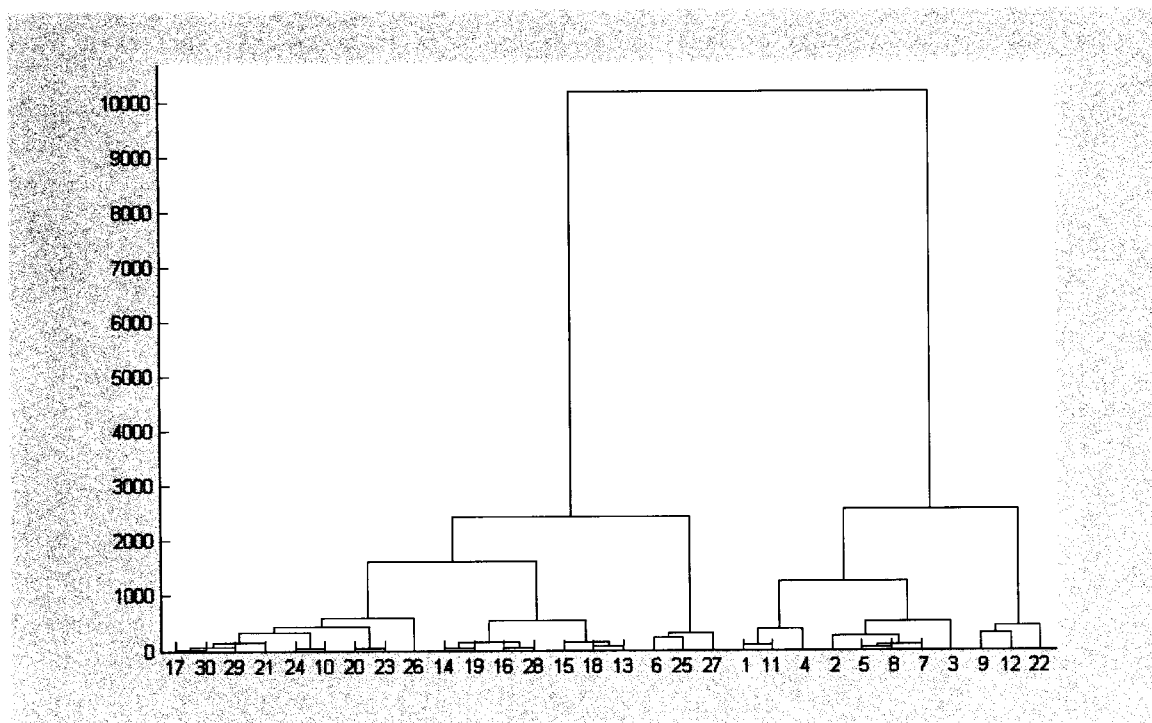


Figura 37: Dendrograma do método ward para curvatura contínua sem ponto adicional para amostras de canideos. (somente feature de curvatura)

5.2 - Resultados extraídos por curvatura digital

5.2.1 - Resultados de estudos da variação geográfica da espécie de roedores (*Thrichomys apereoides*) usando métodos de classificação não supervisionados. (clustering hierárquicos)

No estudo da variação geográfica utilizamos critérios para fazer o levantamento dos marcos anatômicos e a definição de coordenadas (x,y) de 14 landmarks da vista lateral do crânio de roedores da espécie *Thrichomys apereoides*. Os

critérios utilizados neste levantamento foram estritamente biológicos, isso significa que os pontos anatômicos marcados fazem parte da própria anatomia biológica do esqueleto e a definição desses *landmarks* foram feitas manualmente por biólogos experientes. Esta inserção de *landmarks* na anatomia do esqueleto foi feita cuidadosamente em um programa onde o resultado desejado era obter pares de coordenadas x,y de cada espécie analisada. Neste estudo foram considerados (20) vinte populações do nordeste, centro-oeste e sudeste do Brasil, ver tabela 12. Os nomes das localidades foram abreviados e fixados nos arquivos utilizados juntamente com os números de séries da espécie na coleção do Museu como forma de identificação.

Localidades	
Sal= Salinas – MG	Moc = Jaiba – MG
Fei = Feira de Santana – BA	Palm = Palmeiras – BA
Bomj = Bom Jesus da Lapa - BA	Barr = Barreiras – BA
Caru = Caruaru – PE	Camp = Campos Sales- CE
Bod = Bodoco – PE	Pri = Princesa Isabel – PB
Ala = Alagoinha – PE	Flo = Floresta – PE
Pind = Palmeira dos Indios - AL	Vic = Vicososa – AL
Ipan = Santana do Ipanema - AL	Ita = Itapage – CE
Cra = Crato – Ceara	Tri = Triunfo – PE
Serr = Serra da Mesa- Goias	

Tabela 12 - Localidades do nordeste, centro-oeste e sudeste do Brasil utilizadas nos estudos da variação.

Na primeira análise foi usado curvatura digital e energia de dobramento para crânios dos roedores. Para tal avaliação geográfica da variação morfológica foi usado o método Ward e clustering hierárquico. Os resultados indicaram uma correspondência bem definida entre as distribuições geográficas e morfológicas colocando a curvatura digital como uma característica apropriada da forma. Os dois grupos descobertos podem ser caracterizados pela forma da vista lateral do esqueleto e continuidade geográfica. A

amostra do crânio analisado pode ser visto na figura 38 com os 14 marcos anatômicos já marcados.

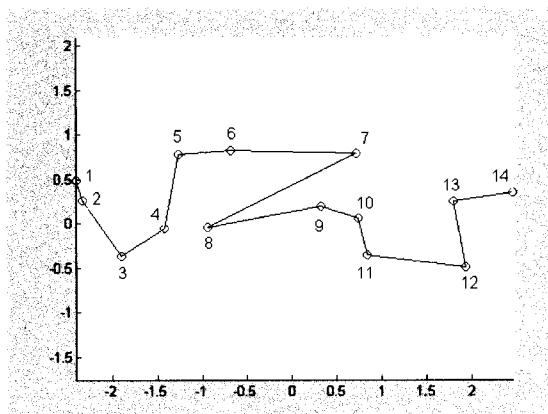


Figura 38: Vista lateral do crânio da espécie *T. apereoides* contendo 14 landmarks.

Continuando com os testes, após ter sido realizado a extração de todas as curvaturas da forma, observamos que quando retiramos a primeira e última curvatura de cada espécie, os resultados foram mais promissores como podemos observar na tabela 13 onde usamos o método Ward.

Método Ward

E + Curv	1	2	Curv	1	2	Energia	1	2
1	70,89%	29,11%	1	70,89%	29,11%	1	91,04%	8,9%
2	26,34%	73,66%	2	26,34%	73,66%	2	68%	32%

Tabela 13 : Método Ward usado para variação morfológica dos crânios das espécies *T. apereoides*. (sem a primeira e última curvatura).

Podemos observar no primeiro quadro da tabela onde esta mostrando energia + curvatura que as diagonais indicam a porcentagem de acertos das amostras que eram 1 e

permaneceram 1 (70,89%) e das amostras que eram 2 e permaneceram 2 (73,66%) e assim sucessivamente nos quadros somente com curvatura e energia.

Com as curvaturas digitais obtivemos dendrogramas que puderam expressar o relacionamento entre diversos sub-clusters. As classes as quais queríamos foram obtidas cortando o dendrograma em pontos específicos da altura. (ver figura 39). As duas classes obtidas do dendrograma, correspondem à população do norte e do sul.

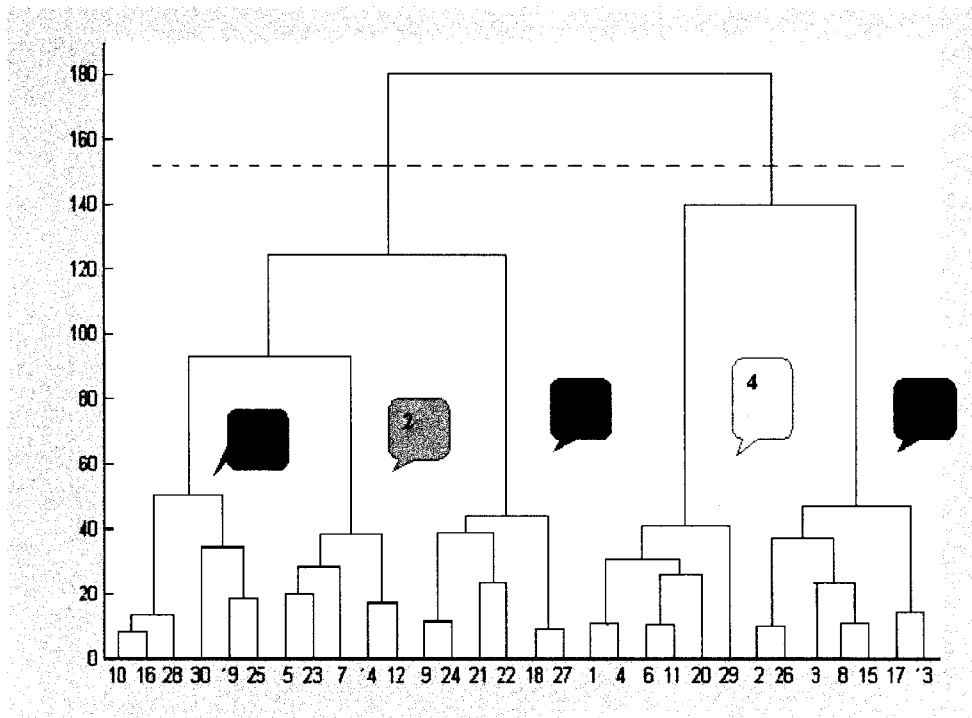


Figura 39: Dendrograma obtido do método ward usando curvatura digital, considerando (energia e retirando o primeiro e último ângulo) para obtenção das 5 classes.

As classes 1, 2 e 3 obtidas foram consideradas região sul e as classes 4 e 5 foram consideradas região norte.

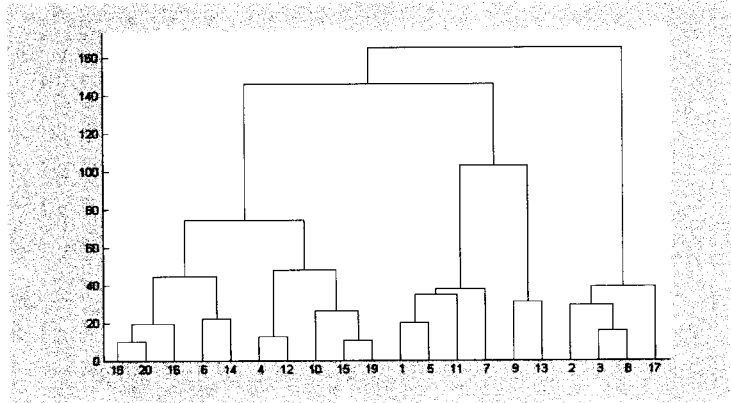
- A tabela abaixo representa resultados detalhados encontrados no dendrograma da figura 39.

Tabela 14 -Crânio - Dados das populações						
Localidade	Região 1	2	3	4	5	Total indivíduos por região
BA	11	26	42	4	8	91
	12,1%	28,5%	46,2%	4,4%	8,8%	100%
MG	9	10	9			28
	32,15%	35,7%	32,15%			100%
PE		22	14	49	57	142
		15,49%	9,89%	34,5%	40,12%	100%
CE		22	7	21	31	81
		27,16%	8,64%	25,93%	38,27%	100%
AL		16	2	20	21	59
		27,12%	3,4%	33,89%	35,59%	100%
PB			2	9	4	15
			13,34%	60%	26,66%	100%

Tabela 14 : Dados das populações mostradas no dendrograma da figura 39.

Nesta tabela podemos observar que as localidades da Bahia e Minas Gerais foram consideradas região Sul e as demais localidades região Norte.

Algumas outras análises foram feitas nas amostras utilizando os dados em um programa aplicativo chamado SPLUS, um software estatístico. Fizemos a análise de clusters usando o método *Hierarchical Clustering* onde este método funciona com um algoritmo que vai combinando ou dividindo grupos existentes produzindo uma estrutura hierárquica mostrando a ordem na qual os grupos são agrupados ou divididos. O Agglomerative Hierarchical Clustering inicia com cada observação em grupo separado e procede até todas observações serem um único grupo (Fig 41). E o método Divisive Hierarchical Clustering inicia com toda a observação num único grupo e procede até cada observação em grupos separados (Fig 42). Os dendrogramas podem ser vistos nos gráficos abaixo.



(a)

Figura 40: Amostras de crânio (*T. apereoides*) (a) Programa desenvolvido no matlab (Aglomerativo).

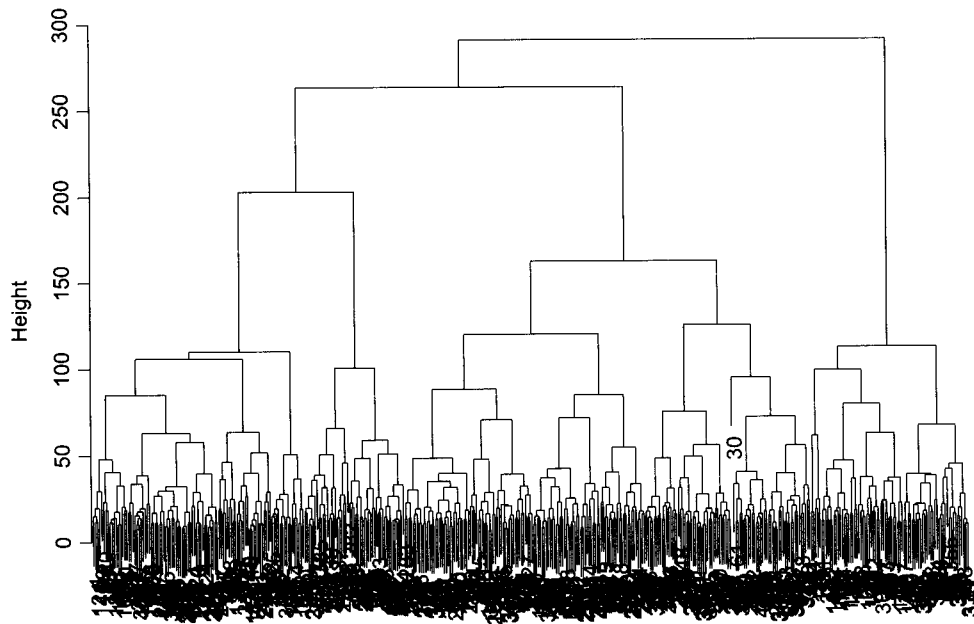


Figura 41: Amostras de crânio (*T. apereoides*); Método Aglomerativo (S-Plus)

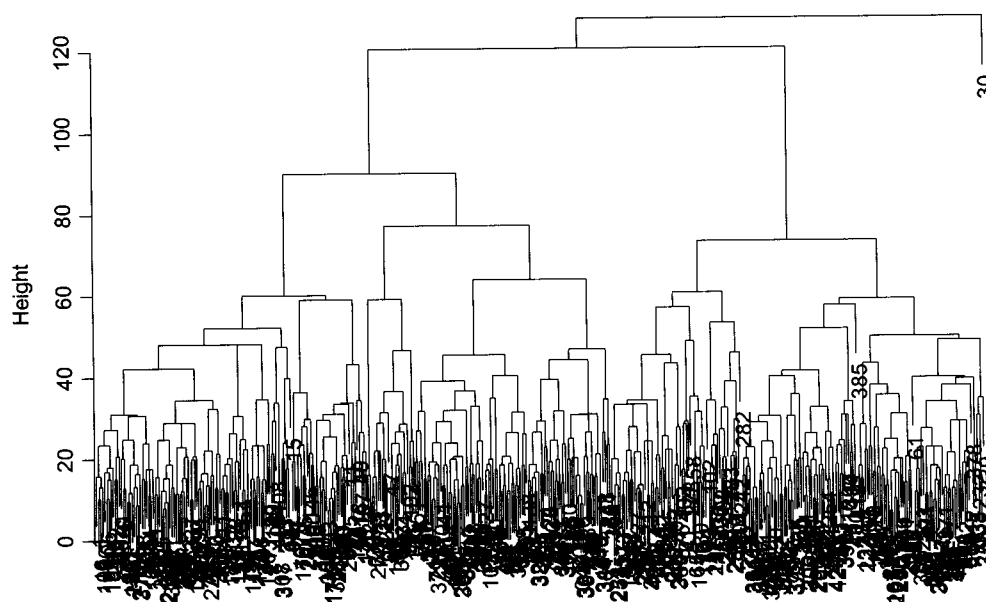


Figura 42: Amostras de crânio (*T. apereoides*): Método Divisivo (S-Plus).

5.2.2 - Validação dos resultados extraídos das medidas de curvatura digital usando métodos de feature selection (FSS) para crânios.

Com o uso de '*feature selection*' tentamos reduzir o número de dimensões e melhorar o desenvolvimento de algumas medidas independentes. O algoritmo de pesquisa sequencial mais comum para seleção de features é o algoritmo FSS (*forward sequential selection*). O FSS inicia sem atributos, avalia todos os subconjuntos de features com exatamente uma feature e seleciona uma com o melhor desempenho. E então adiciona para este sub-conjunto que tem o melhor rendimento pra

sub-conjuntos de tamanho maior. Este ciclo se repete até nenhuma melhoria seja obtida do sub-conjunto atual. Para a validação dos resultados obtidos da curvatura digital usamos combinações de ângulos, considerando duas regiões (norte e sul).

- Busca realizada utilizando apenas combinação de ângulo.

Tabela 15: Resultados de validação encontrados utilizando o algoritmo FSS para curvatura digital em amostras de crânio (*T. apereoides*).

Combinação/ângulo	N1	N2	Busca/ ângulo	Acertos 1/N1	Acertos 2/N2
1	134	294	ALFA {4}	80,60%	50,68%
			ALFA {5}	80,60%	63,26%
2	134	294	ALFA {2,5}	70,9%	76,19%
			ALFA {5,7}	77,61%	64,62%
3	134	294	ALFA {2,4,6}	61,0%	88,0%
			ALFA {2,5,6}	69,0%	78,0%
			ALFA {2,5,8}	68,0%	70,0%
			ALFA {3,5,6}	66,0%	79,0%
4	134	294	ALFA {2,5,6,7}	82,0%	67,0%

Tabela 15: Resultados de validação encontrados utilizando o algoritmo FSS para curvatura digital em amostras de crânio (*T. apereoides*).

- Exemplo: Resultado encontrado para 1 combinação :

Ângulos	1	2	3	4	5	6	7	8	9	10	11	12
N1	141	148	256	257	294	290	274	241	233	198	179	131
N2	287	280	172	171	134	138	154	187	195	230	249	297
Acertos1	72	79	77	108	108	72	57	66	97	16	25	58
Acertos2	215	201	179	149	186	218	217	175	136	214	224	239

Tabela 16: Resultados encontrados para crânios (*T. apereoides*) utilizando método FSS considerando apenas um ângulo.

- Busca realizada utilizando combinação de ângulo e energia, considerando duas regiões (norte e sul).

Combinação/ângulo	N1	N2	Busca/ ângulo	Acertos 1/N1	Acertos 2/N2
1	134	294	ALFA {4}	80,60%	50,68%
			ALFA {5}	80,60%	63,26%

Tabela 17: Resultados de validação encontrados utilizando FSS para curvatura digital e energia.

- As medidas para as amostras 1,2,3,4 não mudaram com as energias.

5.2.3 - Resultados obtidos de curvatura digital para mandíbulas.

Os critérios utilizados neste levantamento foram estritamente biológicos, isso significa que os pontos anatômicos marcados fazem parte da própria anatomia biológica do esqueleto. Esta inserção de *landmarks* na anatomia do esqueleto foi feita cuidadosamente em um programa onde o resultado desejado era obter 11 pares *landmarks* de coordenadas (x,y) da mandíbula de roedores da espécie *Thrichomys apereoides*.

Na primeira análise foi usado curvatura digital e energia de dobramento para mandíbulas dos roedores. Para tal avaliação geográfica da variação morfológica foi usado o método Ward e clustering hierárquico.

O estudo das mandíbulas possuem coordenadas (x,y) de 11 marcos anatômicos. (ver figura 43).

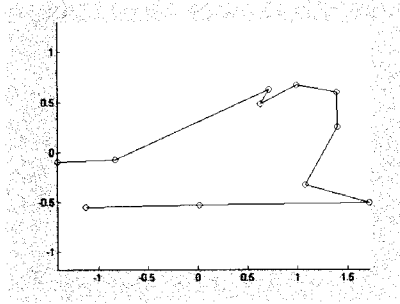


Figura 43: Vista lateral da mandíbula da espécie *T. apereoides* contendo 11 landmarks.

Continuando com os testes, após ter sido realizada a extração de todas as curvaturas da forma, observamos que quando retiramos a primeira e última curvatura de cada espécie, os resultados foram mais promissores como podemos observar na tabela 18 onde usamos o método Ward.

Método Ward

E + Curv	1	2	Curv	1	2	Energia	1	2
1	76,92%	23,08%	1	76,92%	23,08%	1	100%	0%
2	30,0%	70,0%	2	30,0%	70,0%	2	17,37%	82,63%

Tabela 18 : Método Ward usado para variação morfológica das mandíbulas das espécimes *T. apereoides*. (sem a primeira e última curvatura).

Podemos observar no primeiro quadro da tabela onde esta mostrando energia + curvatura que as diagonais indicam a porcentagem de acertos das amostras que eram 1 e permaneceram 1 (76,92%) e das amostras que eram 2 e permaneceram 2 (70%) e assim sucessivamente nos quadros somente com curvatura e energia.

- Obtenção do dendrograma para mandíbulas usando o método Ward.

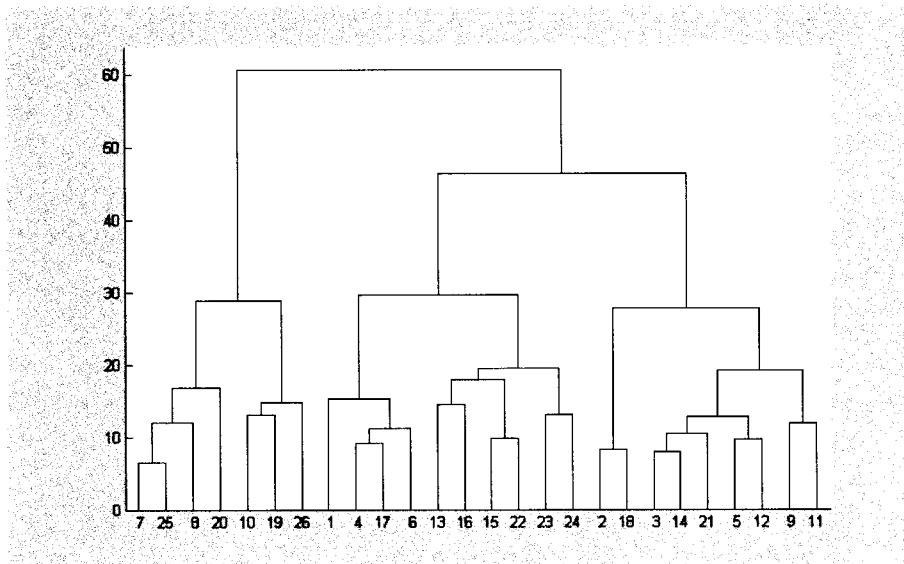


Figura 44: Dendrograma obtido pelo método Ward usando curvatura digital e energia, para mandíbula.

Os mesmos resultados foram analisados no programa Splus utilizando o método aglomerativo e divisivo. Veja Figura 45 e 46.

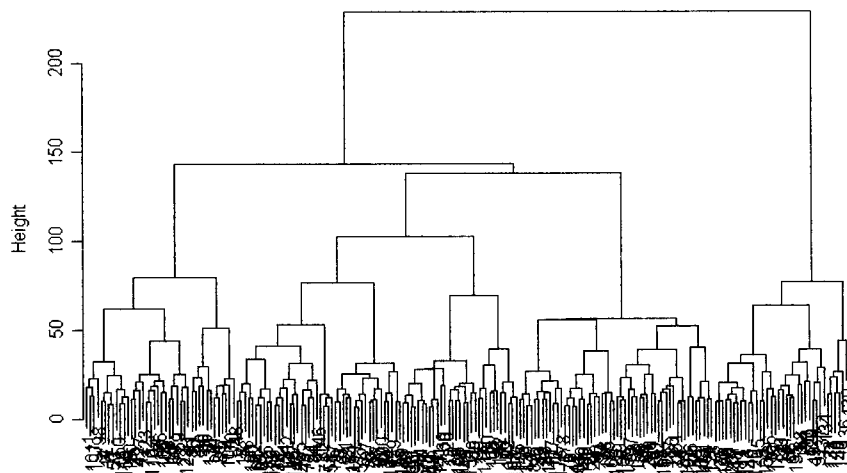


Figura 45 - Método hierárquico Aglomerativo gerado no Splus com amostras de mandíbulas.

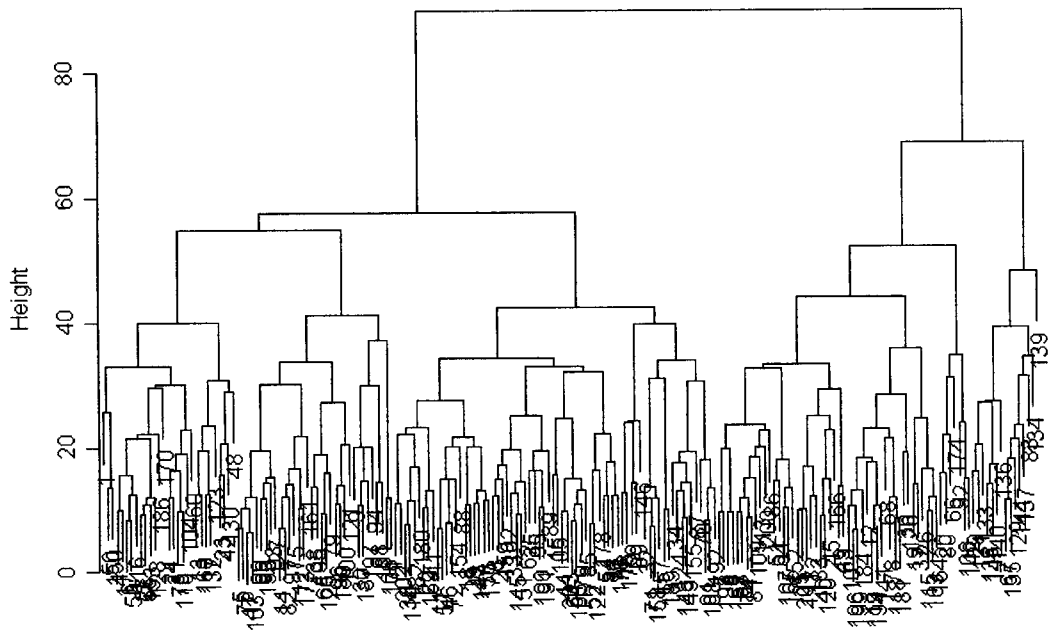


Figura 46 - Método hierárquico divisivo gerado no Splus com amostras de mandíbulas

5.2.4 - Validação dos resultados extraídos das medidas de curvatura digital usando métodos de feature selection (FSS) para mandíbulas.

A validação dos resultados das mandíbulas usando métodos de feature selection (FSS) usam combinações de ângulos para as curvaturas digitais considerando duas regiões (norte e sul).

- Busca realizada utilizando apenas combinação de ângulo.

Tabela 19: Resultados de validação encontrados utilizando o algoritmo FSS para curvatura digital em amostras de mandíbulas (<i>T. apereoides</i>).					
Combinação/ângulo	N1	N2	Busca/ ângulo	Acertos 1/N1	Acertos 2/N2
1	13	190	ALFA {4}	84,61%	75,78%
2	13	190	ALFA {4,11}	84,61%	85,78%
3	13	190	ALFA {3,7,10}	84,61%	83,68%
4	13	190	ALFA {3,5,7,11}	76,92%	85,78%
	13	190	ALFA {1,3,6,10}	84,61%	85,26%

Tabela 19 : Resultados de validação encontrados utilizando o algoritmo FSS para curvatura digital em amostras de mandíbulas (*T. apereoides*).

Busca realizada utilizando combinação de ângulo e energia, também considerando duas regiões (norte e sul) para mandíbulas.

Tabela 20: Resultados de validação encontrados utilizando FSS para curvatura digital e energia					
Combinação/ângulo	N1	N2	Busca/ ângulo	Acertos 1/N1	Acertos 2/N2
1	13	190	ALFA {10}	100 %	82,63%
2	13	190	ALFA {2,4}	84,61%	77,89%
3	13	190	ALFA {2,6,9}	84,61%	83,68%
4	13	190	ALFA {2,5,7,8}	100%	82,10%

Tabela 20: Resultados de validação encontrados utilizando FSS para curvatura digital e energia.

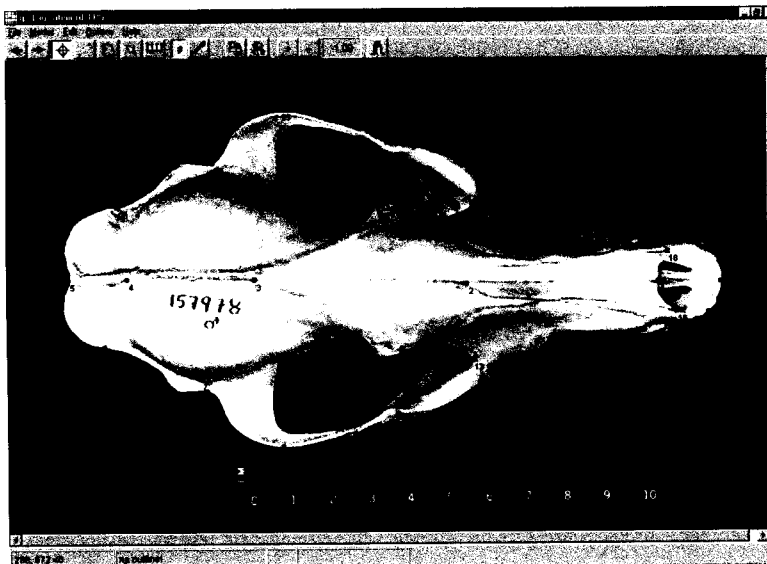
5.2.5 - Resultados obtidos de curvatura digital para crânios de canídeos.

Os critérios utilizados neste levantamento foram estritamente biológicos, isso significa que os pontos anatômicos marcados fazem parte da própria anatomia biológica do esqueleto. Esta inserção de *landmarks* na anatomia do esqueleto

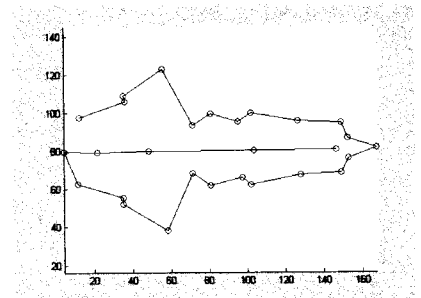
foi feita cuidadosamente em um programa onde o resultado desejado era obter 28 pares marcos de coordenadas (x,y) de crânios de canídeos.

Na primeira análise foi usado curvatura digital e energia de dobramento para mandíbulas dos roedores. Para tal avaliação geográfica da variação morfológica foi usado o método Ward e clustering hierárquico.

O estudo dos crânios possuem coordenadas (x,y) de 28 marcos anatômicos. (ver figura abaixo). Estes marcos podem ser marcados no programa TpsDig ou Photomodeler.



(a)



(b)

Figura 47 - Amostras de canídeos *Atelocynos* (a) TpsDig com seus respectivos landmarks; (b) Matlab, landmarks interpolados.

Método Ward

E + Curv	1	2
1	44,44%	55,56%
2	36,85%	63,15%

Curv	1	2
1	44,44%	55,56%
2	36,85%	63,15%

Energia	1	2
1	38,88%	61,12%
2	21,06%	78,94%

Tabela 21 : Método Ward usado para variação morfológica das mandíbulas das espécies *T. apereoides*. (sem a primeira e última curvatura).

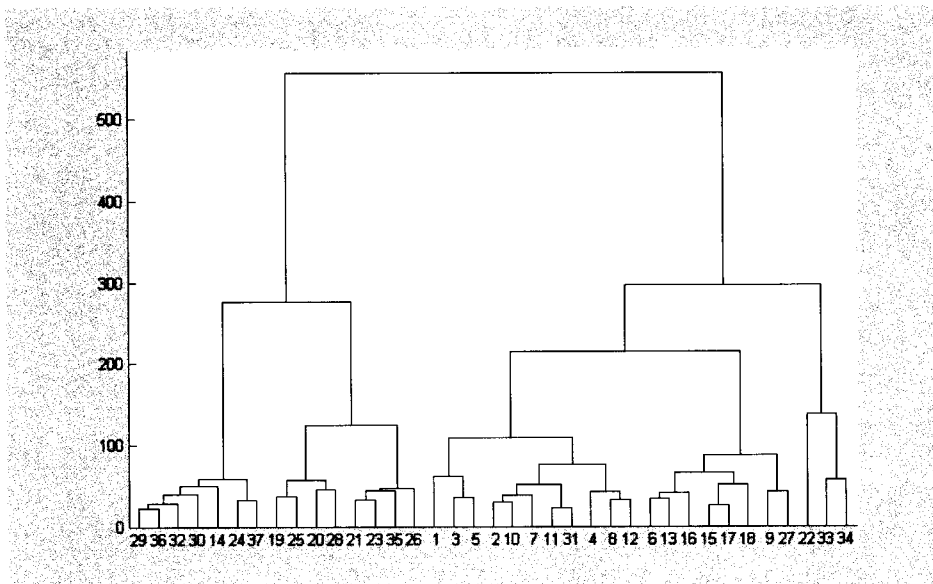


Figura 48 - Dendrograma obtido pelo método Ward das amostras de canideos usando curvatura digital e energia.

Os mesmos resultados foram analisados no programa Splus utilizando o método aglomerativo e divisivo. Veja Figura 49 e 50.

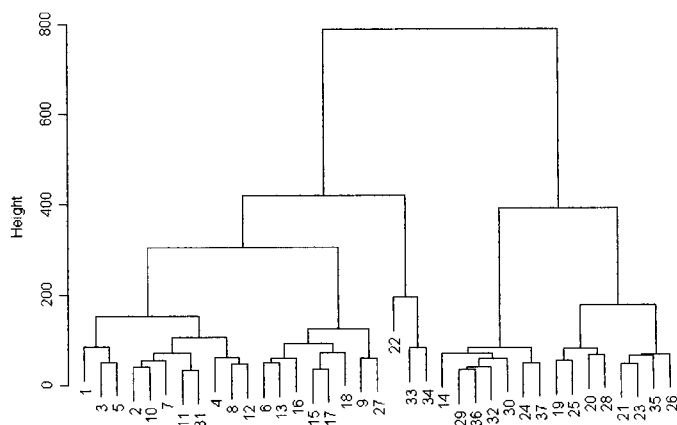


Figura 49 - (a) Método Aglomerativo

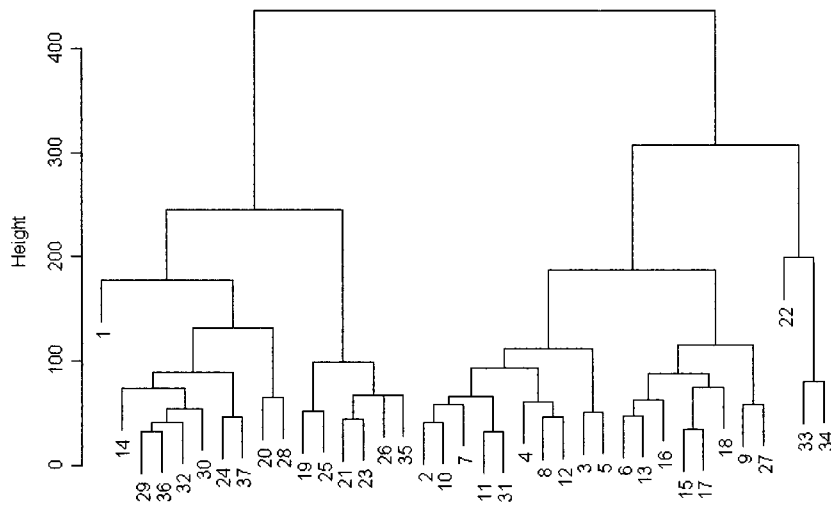


Figura 50 - (b) Método Divisivo.

5.3 - Resultados extraídos de Thin-Plate Spline e Warps

Estes métodos fazem parte da metodologia tradicional e usamos nossas amostras nestes métodos só como efeito de comparação.

Os resultados obtidos com os métodos thin-plate e warps são absolutamente gráficos.

Usamos os programas da literatura desenvolvidos por [Rohlf, F.J. (1998)] como TpsSpln e TpsRelw para gerar os gráficos com os dados das amostras.

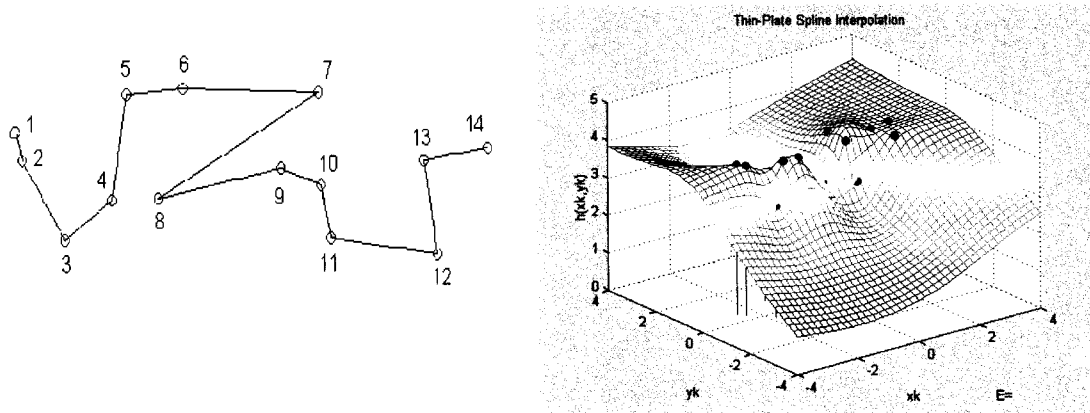


Figura 51 - Amostra da vista lateral do crânio *T. apereoides* e plot da deformação da mesma vista no programa thin-plate.

Usamos o programa TpsSpln para nossas amostras de crânios *T. apereoides*. O TpsSpln computa as transformações de thin-plate e os pesos de partial warps como variáveis de forma e expressa os resultados de uma análise morfométrica como uma deformação da forma. Também usamos o TpsRelw, este programa corresponde a uma análise do componente principal da variação na amostra, ou seja o programa facilita a análise de variação na forma para dados de landmark entre mais do que duas espécimes. O programa lê um conjunto de coordenadas x,y para uma amostra e expressa a variação delas em termos de *relative warps*. Relative warps são eixos de componentes principal de um espaço multivariado na qual cada ponto corresponde para uma espécime e os eixos são inversamente os pesos dos principais warps de uma matriz de energia de dobra definida por uma configuração de referência de *landmarks*.

5.3.1 - Resultados plotados de warps principal, parcial e relativos para crânio da espécie *T. apereoides*.

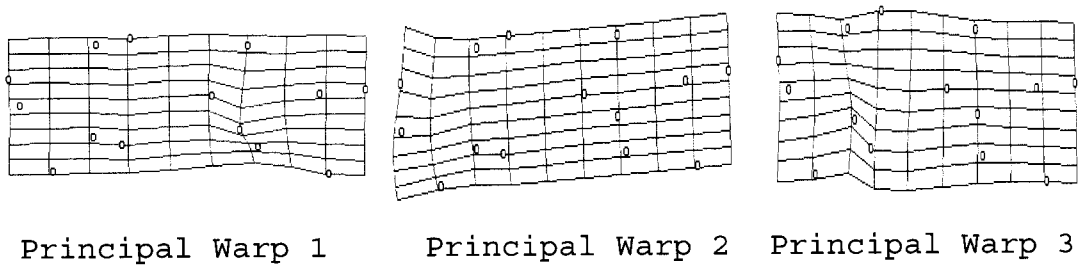


Figura 52: Plot dos principal warps para espécie de referência.

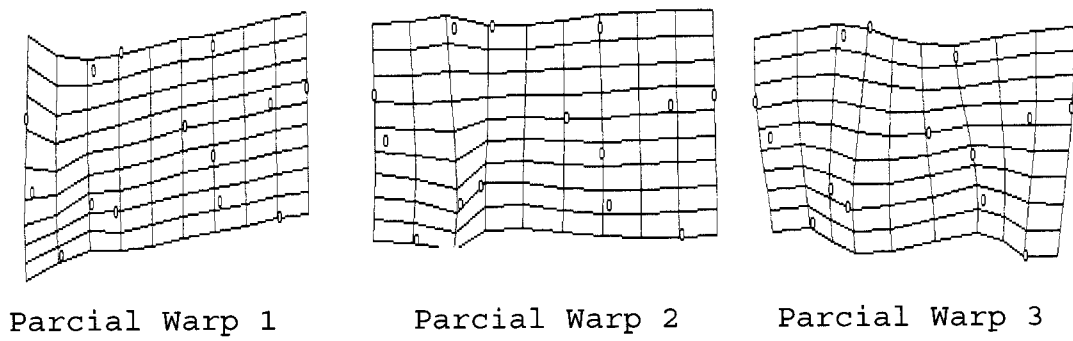


Figura 53: Plot do parcial warps para espécie de referência.

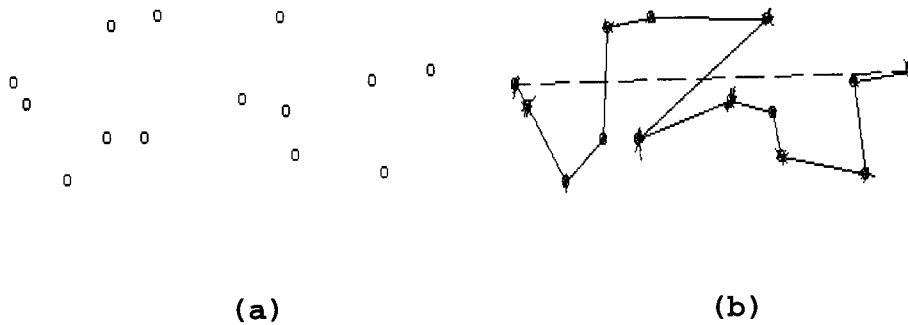


Figura 54: (a)Configuração de referência, (b) configuração de referência conectada.

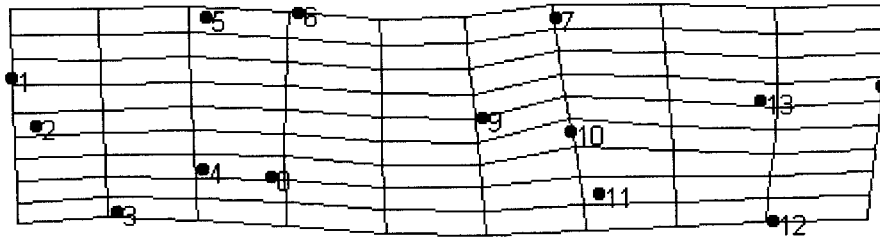


Figura 55 : Relative Warp 1 + deformação

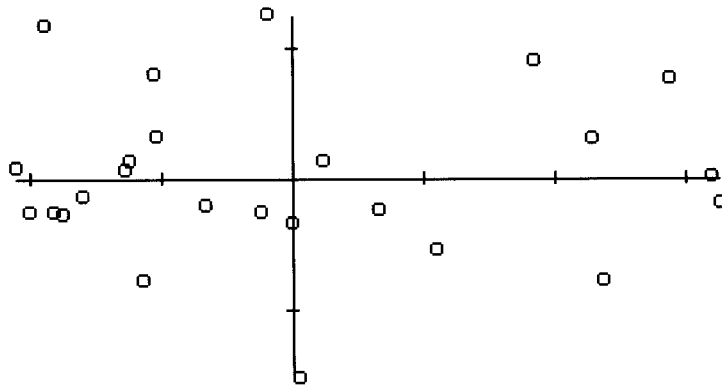


Figura 56:Plot da matriz de peso do relative warps.

5.4 - Classificação usando métodos supervisionados e não-supervisionados.

Nesta seção apresento resultados de clustering hierárquicos e análise discriminante usando amostras biológicas de crânios e mandíbulas de roedores e de crânios de canídeos.

5.4.1 - Supervisionados

5.4.1.1 – Roedores (*T. apereoides*)

Os resultados de estudos da variação geográfica da espécie de roedores (*Thrichomys apereoides*) usando métodos de classificação supervisionados podem ser vistos nesta seção. Foram testadas 428 amostras a fim de extrair curvaturas digitais (DC) e energias de dobra (E). E os seguintes vetores das características foram considerados:

- (i) todas as curvaturas digitais e energias de dobra;
- (ii) todas as curvaturas digitais; e energias de dobra somente.
- (iii) toda curvatura digital e razão.

Os resultados obtidos pelo método ward, na classificação das matrizes indicam uma correspondência geográfica bem definida entre os indivíduos do conjunto original e os dois conjuntos obtidos considerando as medidas morfológicas adotadas. O acordo é particularmente forte quando as curvaturas digitais são feitas como características morfológicas e a inclusão da energia de dobramento não melhorou as correspondências.

Dos métodos supervisionados usamos a análise canônica de 12 curvaturas digitais. As análises das curvaturas digitais e da energia de dobramento mostrou que os eixos canônicos 1 e 2 explicaram (63%) da variação, com o primeiro eixo contendo uma parcela principal da variação (46%). Este eixo discrimina dois grupos das populações: um grupo do norte composto das populações (a–m) e outro do sul incluindo populações (n– t). A única exceção é a população h que ocupa uma posição intermediária entre estes dois clusters principais das populações.

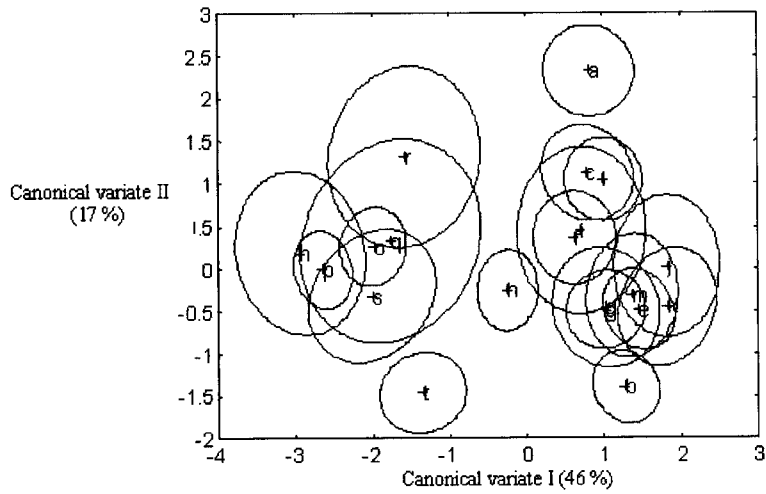


Figura 57 - Plote bivariado dos centróides e 95% região confiante para os primeiro dois eixos derivados da análise canônica para a vista lateral do esqueleto de 20 populações *Thrichomys apereoides* do nordeste e sudeste do Brasil.

Uma ordenação idêntica foi obtida em uma análise excluindo a medida da energia de dobra. Uma melhor separação foi obtida considerando razão e curvatura digital. Veja a figura abaixo.

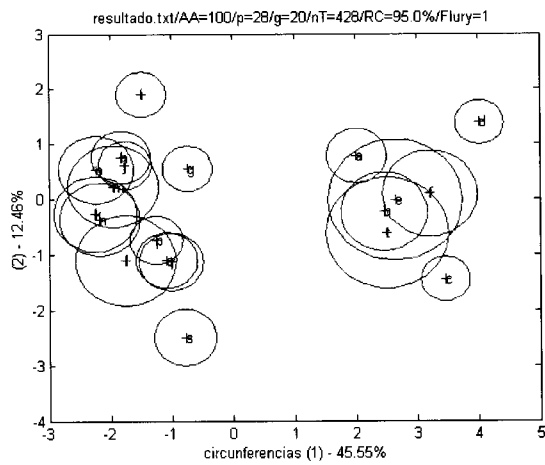


Figura 58 : Análise Canônica Variada considerando razão e curvatura digital para dois grupos de população: um grupo do nordeste e outro do sudeste.

O trabalho com curvatura digital relatou uma característica geométrica para a análise de forma biológica. Mais especificamente, esta característica simples foi usada para a caracterização morfológica das espécimes *T. apereoides* coletadas nas diversas localidades do nordeste e sudeste do Brasil.

Também acrescentamos na metodologia ruído, ou seja, fomos aumentando o grau de perturbação nos pontos de *landmarks*. Mais especificamente, os *landmarks* de cada objeto foram deslocados de suas posições originais por um vetor de perturbação $\vec{p} = (\delta x, \delta y)$, onde δx e δy são números randômicos no intervalo de $-0.5\eta \leq \delta x, \delta y < 0.5\eta$, $\eta > 0$, de modo que a escolha de η determina a intensidade de perturbação. A Figura 59 mostra a forma original de 5 amostras (a)-(c), escolhido randomicamente dos 428 casos, com várias versões distorcidas considerando $\eta=0.1$ (d)-(f), $\eta=0.2$ (g)-(i), $\eta=0.3$ (j)-(l) and $\eta=0.4$ (m)-(o). E a análise canônica destes resultados são mostradas na figura 60. E fica claro no conjunto de plots que a metodologia mostrou, pelo menos nos dados considerados na análise, uma robusta perturbação η até 0.2, com uma sobreposição substancial entre duas classes sendo observado para grandes valores de sobreposição.

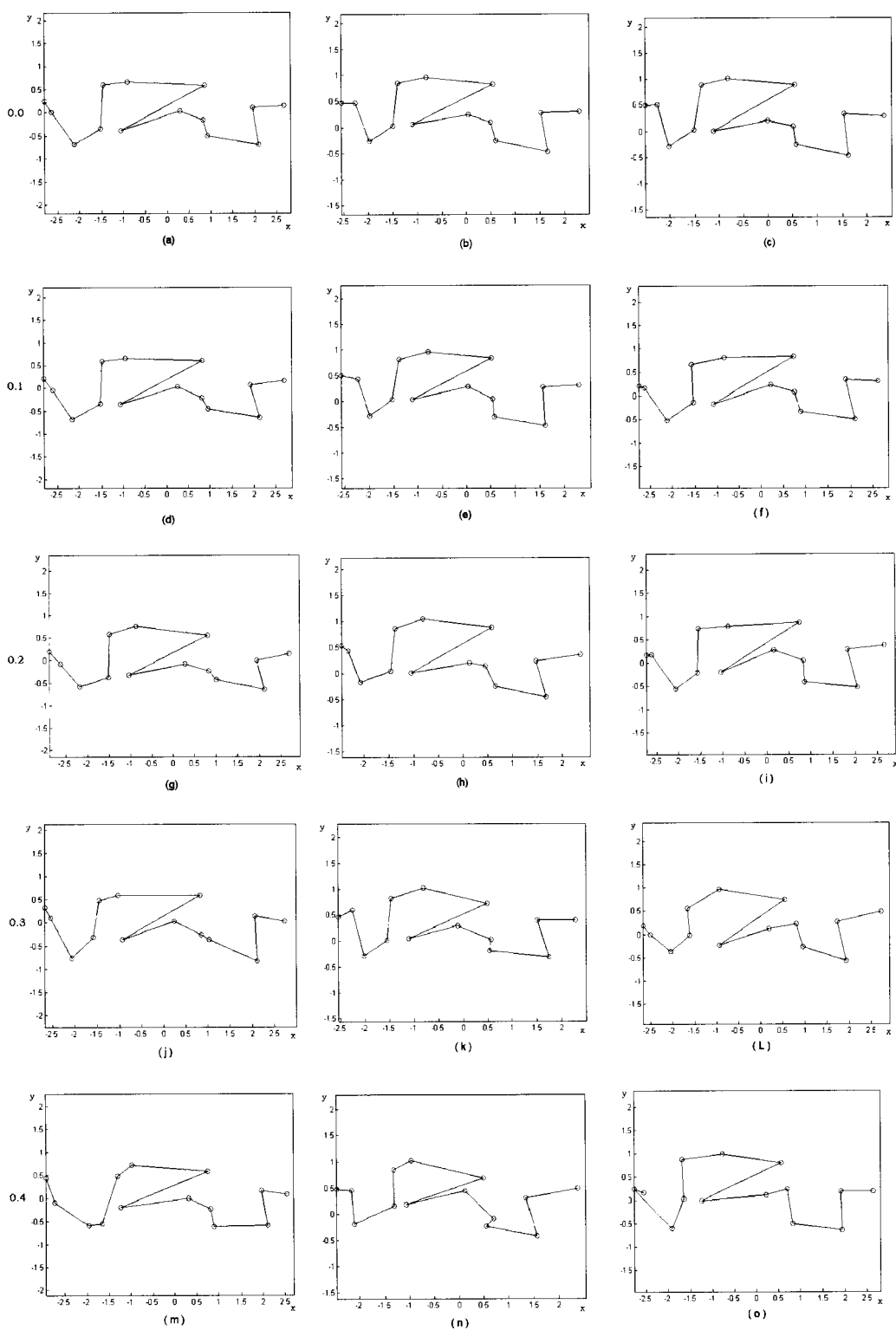


Figura 59 –Forma original de 5 amostras (a)-(c), escolhida randomicamente dos 428 casos, considerando a versão distorcida $\eta=0.1$ (d)-(f), $\eta=0.2$ (g)-(i), $\eta=0.3$ (j)-(l) and $\eta=0.4$ (m)-(o).

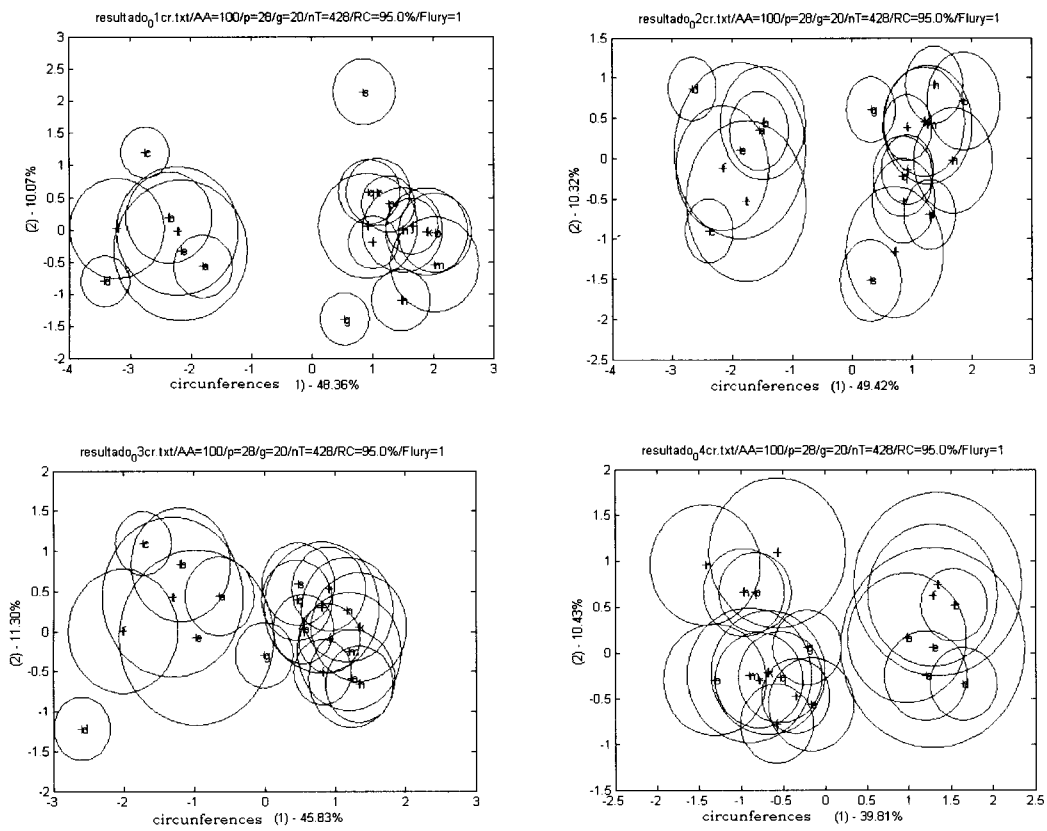


Figura 60 - Resultados da Análise Canônica Variada com as versões distorcidas mostradas na figura 59.

Fizemos também algumas combinações com os ângulos das amostras e analisamos quais dos ângulos tinham mais influência. E dos resultados de validação junto com curvatura digital aplicamos análise canônica.

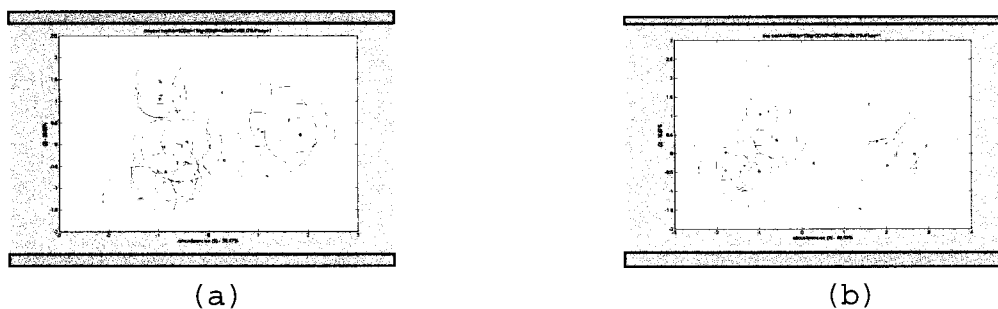
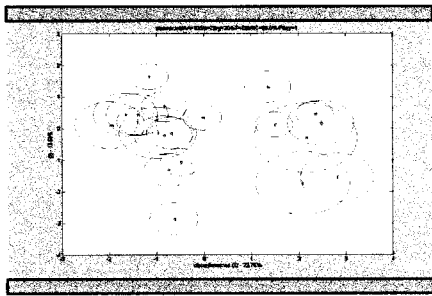
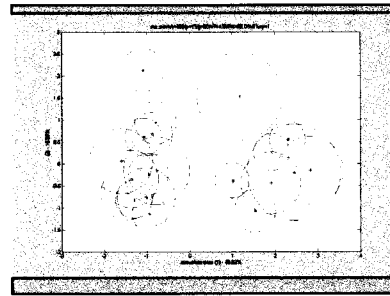


Figura 61: Análise Canônica Variada (a) só energia (b) energia e curvatura, para 1 combinação de ângulo.



(a)



(b)

Figura 62: Análise Canônica Variada (a) só energia (b) energia e curvatura, para 2 combinações de ângulos.

Também aproveitando as amostras de canídeos implementamos o método Procruste já citado anteriormente. Este método consiste no ajuste das distâncias após o processo de sobreposição. A menor soma dos quadrados da distância entre pontos homólogos num par de configuração chamamos de distância Procruste. Veja os resultados obtidos após cada passo da implementação. As amostras em vermelho determinam as espécies e a azul determina o objeto de configuração média de todas as espécies.

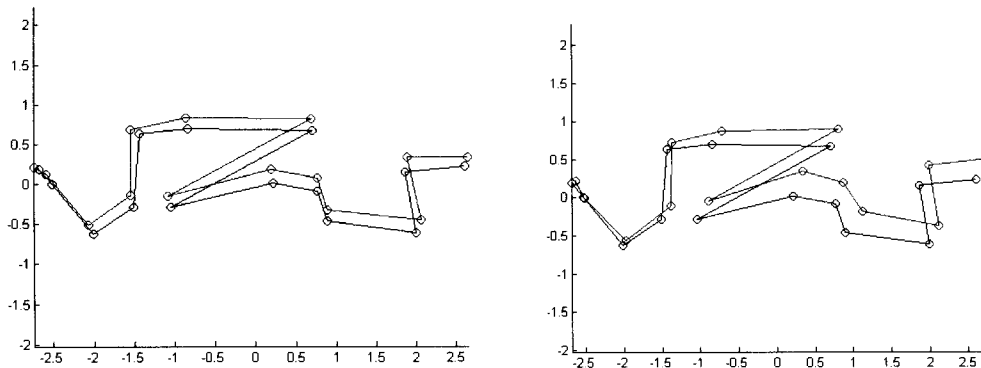


Figura 63 - Configuração original de duas amostras coletadas aleatoriamente dos 428 casos de *T. apereoides*.

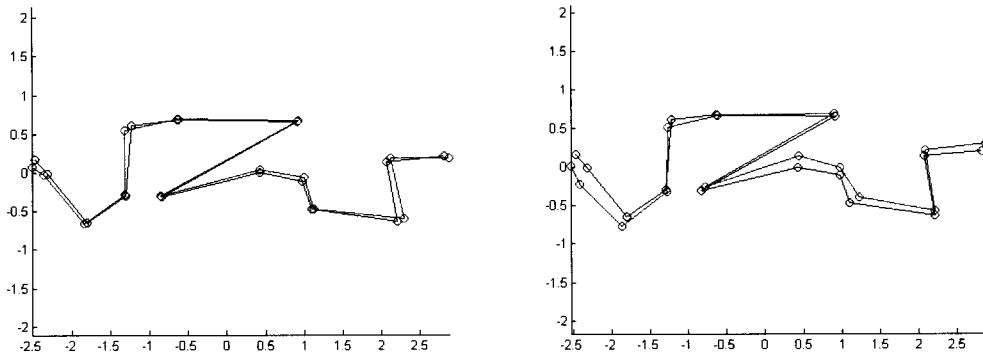


Figura 64 - Configuração transladada de duas amostras coletadas aleatoriamente dos 428 casos de *T* apereoides.

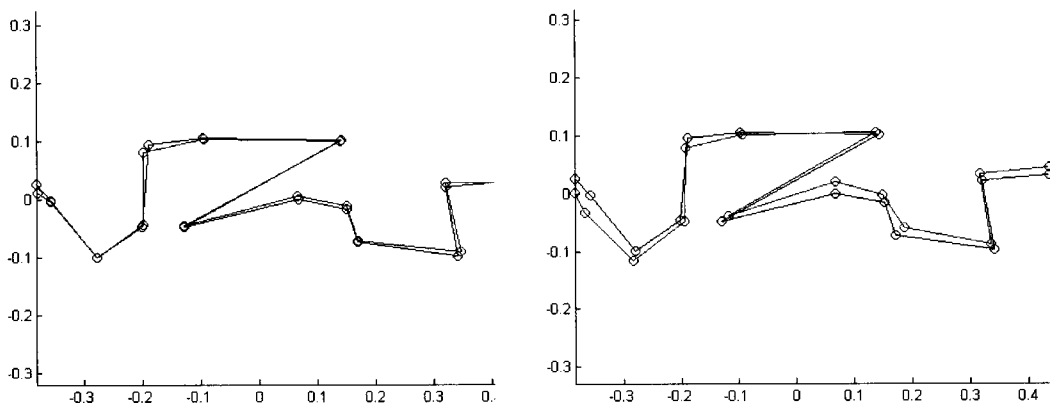


Figura 65 - Configuração escalonada de duas amostras coletadas aleatoriamente dos 428 casos de *T* apereoides.

Após os passos de translação, rotação, o programa fez a análise ortogonal procruste usando o ajuste dos mínimos quadrados e o resultado pode ser visto na figura 67.

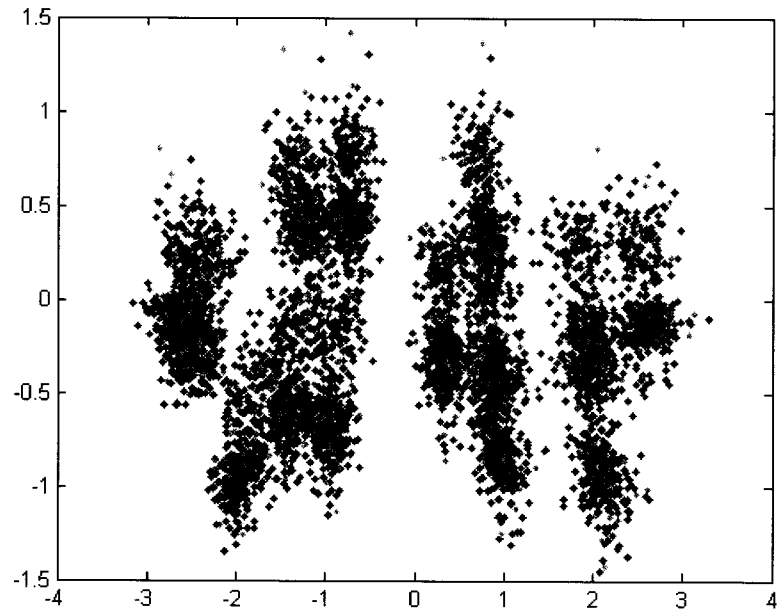


Figura 66: Amostra de 428 espécies de *T. apereoides*, da coleção do Museu do Rio de Janeiro. As amostras usadas foram coletadas das regiões central, ocidental e nordeste do Brasil. Amostras sem o cálculo das distâncias.

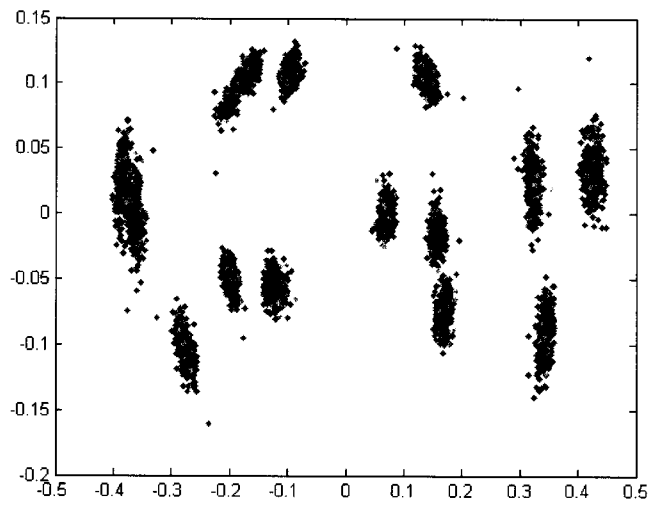


Figura 67: Resultados da análise procrustes para uma amostra de 428 espécies de *T. apereoides*.

Ajuste das distâncias após a translação, escalonamento, e rotação. Veja figura 68.

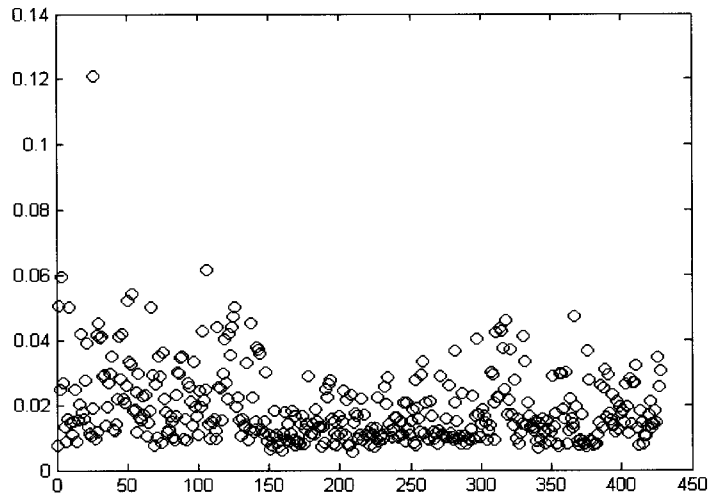


Figura 68: Distâncias ajustadas, depois de transladar, escalonar e rotacionar.

Analisando as 428 amostras, extraímos features de curvatura contínua e curvatura digital. O objetivo era detectar espécies do Sul e do Norte. Estes dados serviram como entrada no cálculo da análise discriminante. Veja figura 69.

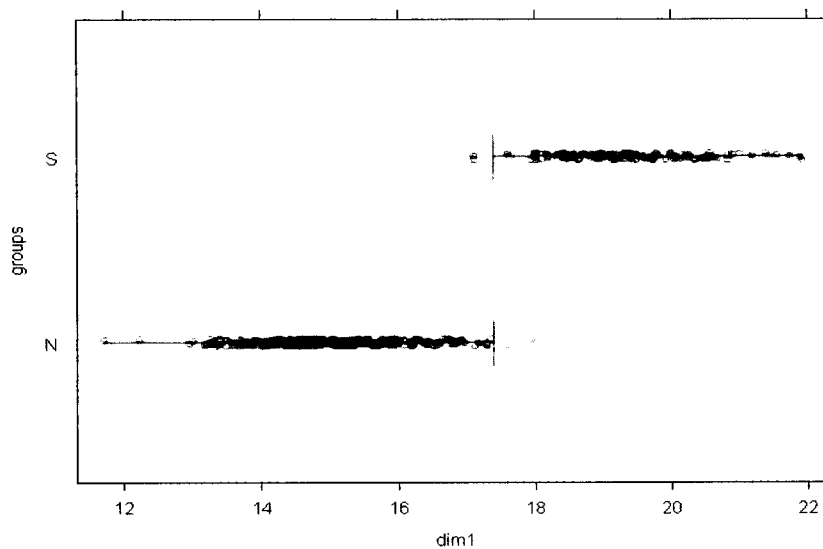


Figura 69 - Análise discriminante das amostras de crânio de roedores, as features extraídas são curvatura digital.

Podemos observar no gráfico acima que houve uma boa separação em amostras pertencentes ao sul e amostras pertencentes ao norte usando dados de curvatura digital.

Usamos dados da curvatura contínua para efeitos de comparação e os resultados mostraram que quando realizamos a análise discriminante destas medidas de curvatura contínua os resultados foram menos eficientes do que quando usamos medidas de curvatura digital.

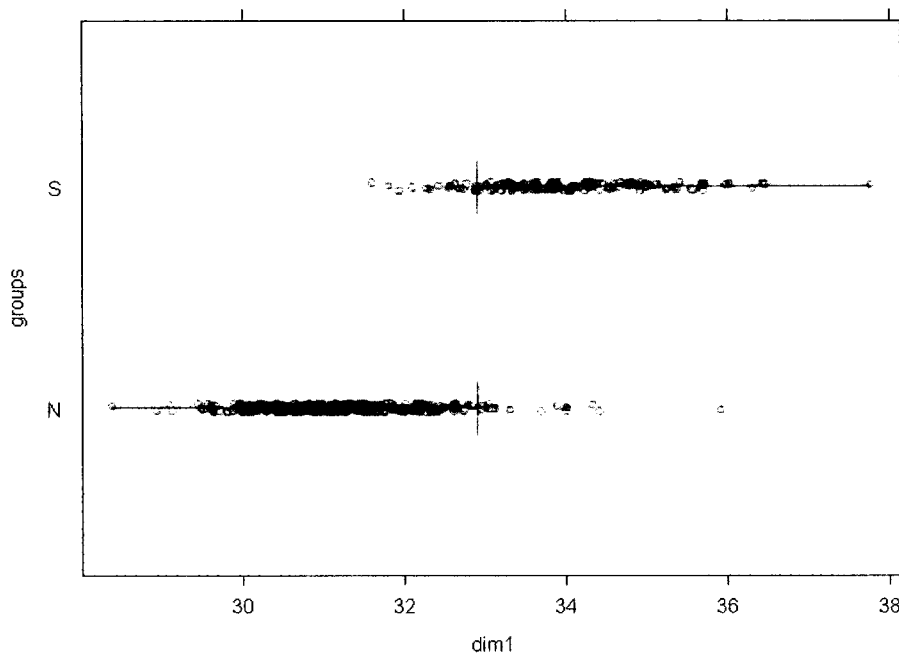


Figura 70 - Análise discriminante das amostras de crânio de roedores, as features extraídas são curvatura contínua.

As mesmas medidas de curvatura contínua e curvatura digital foram agora feitas para amostras de mandíbulas e podemos observar na figura 71 e 72 que o mesmo acontece quando usamos curvatura digital. Os resultados apresentam de uma maneira melhor obtendo um melhor desempenho na separação em amostras pertencentes a região sul e amostras pertencentes a região norte.

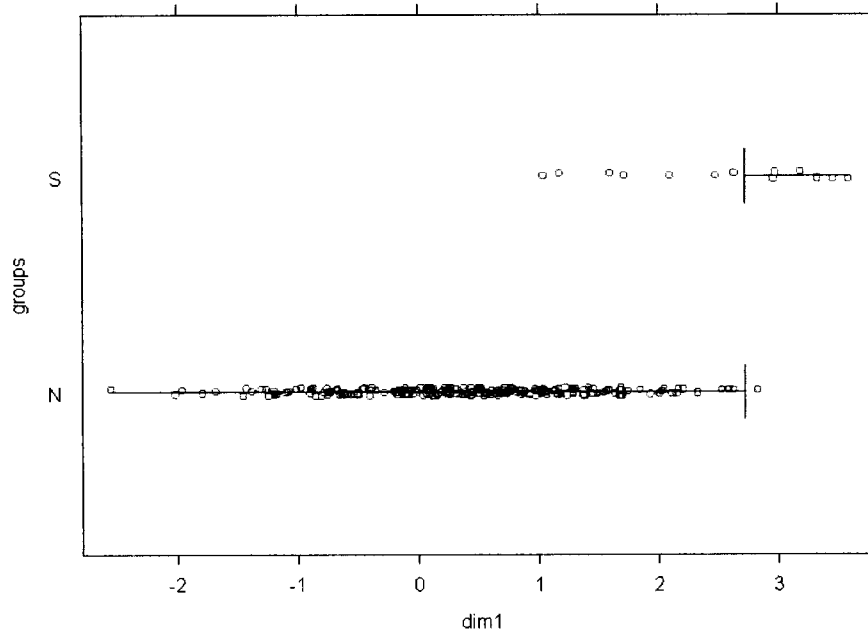


Figura 71 - Análise discriminante das amostras de mandíbulas de roedores, as *features* extraídas são curvatura continua. (20 *features*).

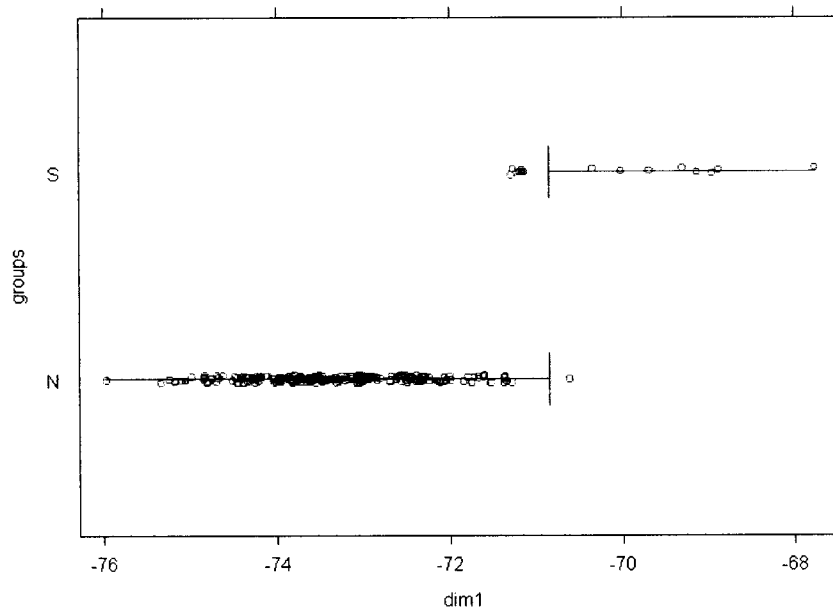


Figura 72 - Análise discriminante das amostras de mandíbulas de roedores, *features* extraídas de curvatura digital.

5.4.1.2 - Canídeos

Obtivemos também resultados das diferenças entre sexo masculino e feminino nas amostras de crânios de canídeos. As amostras de canídeos estudadas estão na tabela 22, foram feitos testes usando análise discriminante em algumas espécies e os resultados podem ser vistos na figura 73.

Gênero	Espécie	Museu	Quantidade
Atelocynus	Microtis	AMNH	13
Cerdocyon	Thous	USNM	24
Chrysocyon	Brachiurus	FMNH	13
Dusicyon	australis	BM	2
Lycalopex	vetulus	MZUSP	18
		Total	70

Tabela 22 - Espécies de canídeos estudadas

A figura 73 abaixo indica a análise discriminante da espécie de canídeos do gênero *Atelocynus*. Para aplicar a análise discriminante retiramos dados de curvatura. A espécie *Atelocynus* possui 13 amostras sendo elas 7 do sexo masculino e 6 do sexo feminino. No gráfico o grupo 1 determina feminino e 2 masculino.

Podemos observar que o gráfico abaixo separou todas as amostras corretamente usando as features extraídas da curvatura digital.

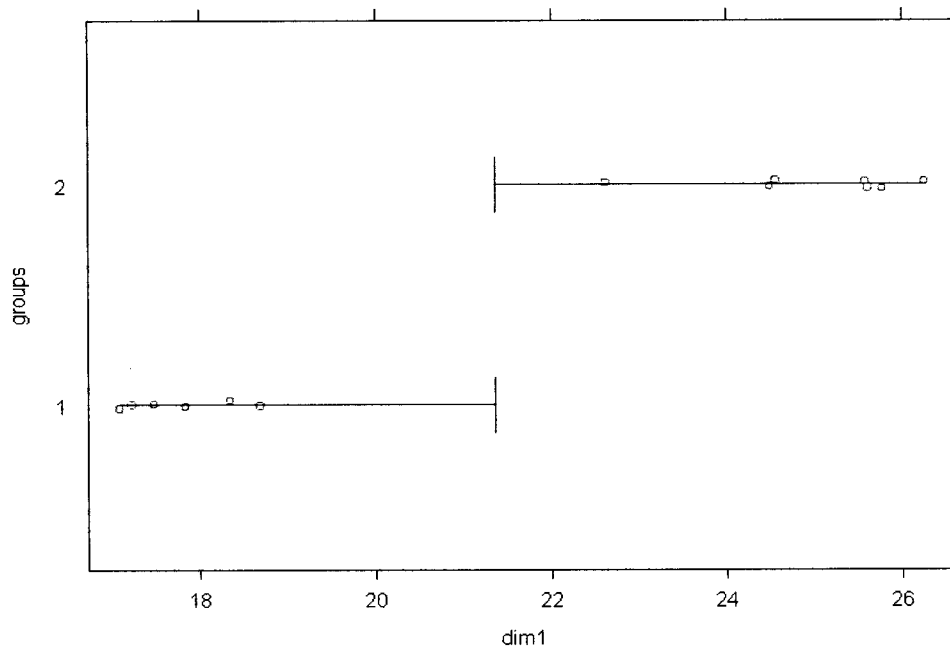


Figura 73 - Análise discriminante da espécie do gênero Atelocynus: (1) feminino , (2) masculino.

A mesma análise foi feita para todas as amostras da tabela 22 .Veja figura 74. Observe no gráfico a divisão de grupo f (feminino) e m (masculino). Podemos observar o quão importante esta “feature” extraída da curvatura significa, pois conseguimos deixar bem separado os sexos das amostras.

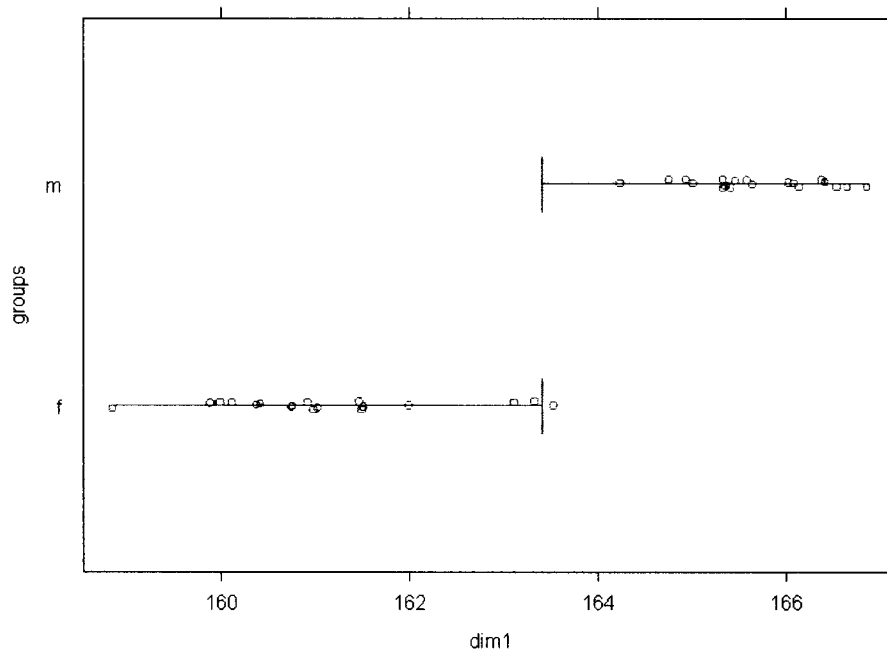


Figura 74 - Análise discriminante de todas as espécies de canídeos *Atelocynus* e *Cerdocyon* (1) feminino , (2) masculino. (curvatura digital)

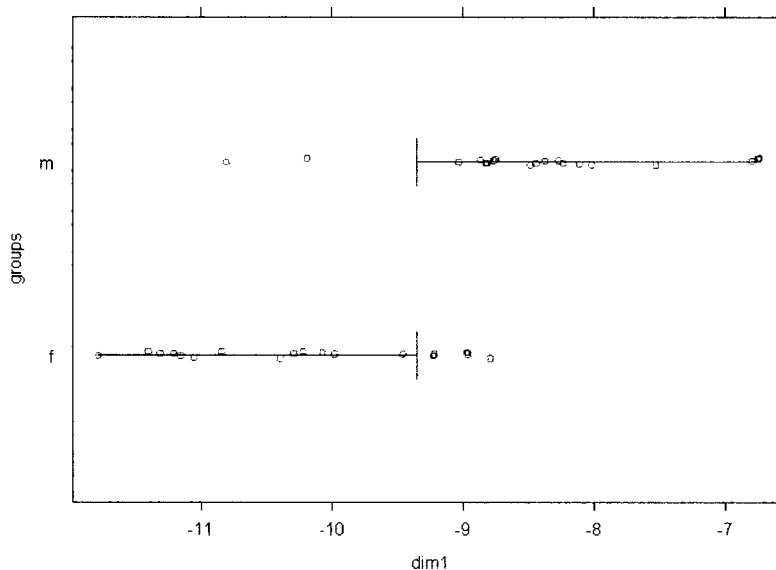


Figura 75 - Análise discriminante de todas as espécies de canídeos *Atelocynus* e *Cerdocyon* (1) feminino , (2) masculino. (20 features de curvatura continua).

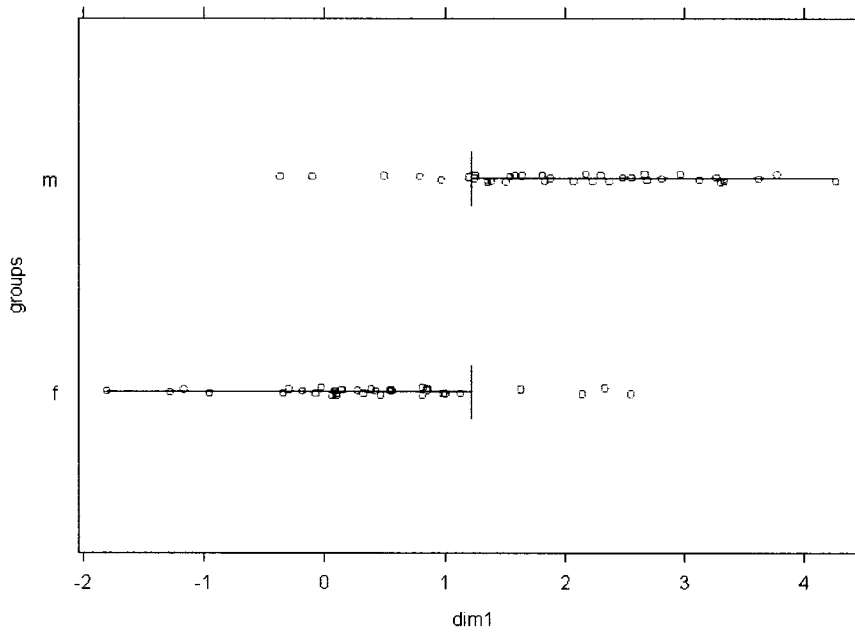


Figura 76 - Análise discriminante de todas as espécies de canídeos mostradas na tabela 24: (1) feminino , (2) masculino. Usando features extraídas de curvatura digital.

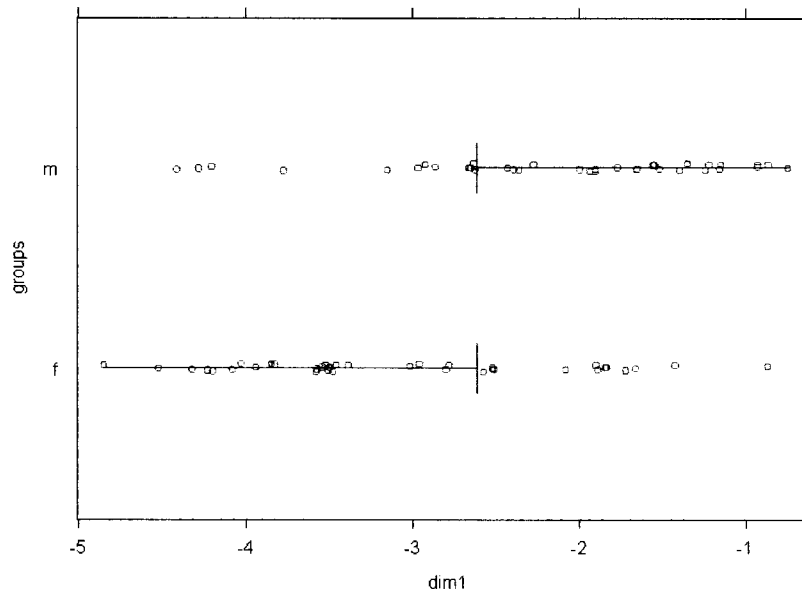


Figura 77 - Análise discriminante de todas as espécies de canídeos mostradas na tabela 24: (1) feminino , (2) masculino. Usando features extraídas de curvatura continua. (20 features).

5.4.2 - Não-Supervisionados

5.4.2.1 - Clustering Hierárquico

Na classificação não supervisionada sabe-se que o conhecimento sobre a geração do padrão é disponível. Tudo que é fornecido vem a ser um conjunto de observações representado em termos de seus respectivos vetores de características. É difícil determinar se os pontos pertencem a um ou mais clusters definidos ou se eles correspondem simplesmente a cluster adicionais.

A verdade é que existe infinitas escolhas, envolvendo várias combinações de proximidades e medidas de dispersão. Na prática clustering é de nenhuma maneira uma tarefa fácil. Ou seja, o critério de clustering impõe uma estrutura nos vetores de características que podem ou não corresponder aquele atualmente subjacente as observações originais. Um critério usado é o critério de similaridade. Agrupa coisas de modo que objetos na mesma classe sejam tão similar quanto possível e objetos de todos os dois conjuntos distintos sejam tão diferente quanto possível. Maximizar similaridade com os clusters automaticamente implica minimizar dissimilaridade entre objetos de clusters distintos. Outro importante fator é como definir um número correto de cluster, na qual afeta substancialmente o resultado alcançado.

Os resultados de estudos da variação geográfica da espécie de roedores (*Thrichomys. apereoides*) usando métodos de classificação não supervisionados foram colocados no decorrer do capítulo 5 a partir da seção 5.2. Porém nesta seção usamos as amostras de *T. apereoides* e também extraímos resultados dos métodos k-means.

5.4.2.2 - K-means

Este é o mais simples e mais fundamental dos métodos de clusterização. Pode-se especificar o número de clusters a serem localizados nos dados e os pontos (de fato, vetores) médios iniciais para cada cluster ou classe. Se não

especificado, o algoritmo arbitrariamente selecionará uma semente, ou seja, uma posição inicial para os K clusters.

Cada ponto do conjunto de dados é atribuído a uma classe cujo ponto médio é o mais próximo deste ponto. O novo conjunto de cada ponto médio é então calculada a partir de resultados de classificações prévias e os pontos são reatribuídos aos novos clusters. O procedimento continua até que não haja mudanças significativas nas atribuições dos pontos de uma iteração e a próxima ou o número de iterações tenha atingido o número máximo de iterações que o usuário especificou. O algoritmo Isodata é basicamente uma extensão do K-médias, pois o usuário também especifica o número desejado de clusters, contudo após cada iteração, o Isodata examina estatisticamente cada cluster e aplica divisão, união ou descarte do cluster de acordo com algum critério. Isto pode levar ao número de categorias resultantes que não concorde com o número especificado pelo usuário inicialmente. Nas nossas amostras usamos o software S-Plus para fazer a análise de clusters onde o usuário também determina o número de cluster e o máximo de interações.

Temos alguns resultados de *features* de curvatura digital para amostras de crânios de *T. apereoides* usando o K-médias.

E + Curv	1	2	Curv + razão	1	2
1	80.59%	19.41%	1	33.0%	67.0%
2	43.0%	57.0%	2	11.0%	89.0%

Tabela 23: Resultados k-médias para crânios de *T.apereoides*, 1:1 (80.59%) , 2:2 (57.0%) para *features* de energia e curvatura; 1:1 (33.0%) e 2:2 (89.0%) para *features* de curvatura e razão.

Temos também alguns resultados de *features* de curvatura contínua para amostras de crânios de *T. apereoides* usando o K-médias.

Curvatura Contínua	1	2
1	88.0%	12.0%
2	37.42%	62.58%

Tabela 24: Resultados k-médias para crânios de *T.apereoides*, 1:1 (88.0%) , 2:2 (62.58%) para *features* de curvatura contínua.

Temos alguns resultados de *features* de curvatura digital para amostras de mandíbulas de *T. apereoides* usando o K-médias.

E + Curv	1	2	Curv + razão	1	2
1	92.30%	7.7%	1	92.30%	7.7%
2	30.0%	70.0%	2	28.95%	71.05%

Tabela 25: Resultados k-médias para mandíbulas de *T.apereoides*, 1:1 (92.30%) , 2:2 (70.0%) para *features* de energia e curvatura.

Temos também alguns resultados de *features* de curvatura contínua para amostras de mandíbulas de *T. apereoides* usando o K-médias.

Curvatura Contínua	1	2
1	0%	100%
2	2.11%	97.89%

Tabela 26: Resultados k-médias para mandíbulas de *T.apereoides*, 1:1 (0.0%) , 2:2 (97.89%) para *features* de curvatura contínua.

Temos alguns resultados de *features* de curvatura digital para amostras de crânio de canídeos usando o K-médias.

E + Curv	1	2	Curv + razão	1	2
1	42.85%	57.15%	1	42.85%	57.15%
2	31.43%	68.57%	2	34.29%	65.71%

(a)

(b)

Tabela 27: Resultados k-médias para canídeos, para *features* de (a) energia e curvatura; (b) curvatura + razão.

Temos também alguns resultados de *features* de curvatura contínua para amostras de crânio de canídeos usando o K-médias.

Curvatura Contínua	1	2
1	22.85%	77.15%
2	14.29%	85.71%

Tabela 28: Resultados k-médias para crânio de canídeos, 1:1 (22.85%) , 2:2 (85.71%) para *features* de curvatura contínua.

Capítulo 6

Conclusão

Esse trabalho tem sido uma importante contribuição para os trabalhos desenvolvidos na área de morfologia, no aspecto de apresentar técnicas de caracterização morfológica, em especial, para caracterização na variação na forma craniana dentro e entre as populações do roedor equimídeo *Thrichomys apereoides* amostradas nas localidades do Brasil.

6.1 - Contribuições

Um das primeiras conclusões do trabalho é que a escolha da base de dados (amostras) foi extremamente importante para o sucesso e validação da pesquisa. A diversidade de amostras usadas no desenvolvimento do projeto ajudou substancialmente no processo de escolha da espécie e a quantidade de imagem disponibilizada pelos professores foi fundamental neste processo de seleção.

O desempenho do sistema, em todas as suas etapas (segmentação, extração, seleção e classificação) está diretamente relacionada ao conjunto de dados utilizados tanto para implementação quanto para sua validação. O uso de uma amostra comum em alguns testes para todos os métodos foi de extrema importância. Além da escolha das imagens a serem digitalizadas, o processo de digitalização incluindo a inserção dos marcos anatômicos foi também outro fator importante para o bom desempenho do sistema.

Com relação aos atributos utilizados, foi possível observar que alguns foram tão eficazes o quanto se esperava mas outros não foram tão bons na diferenciação das formas. Este problema pode estar relacionado na escolha do melhor atributo, uma tarefa difícil a ser realizada.

É importante ressaltar que o uso das técnicas morfométricas, teve uma grande vantagem por fornecer muita informação, mas houve também a possibilidade de desenvolvimento e implementação de um novo método como a Curvatura Digital capaz de garantir o melhor desempenho do trabalho.

No que se refere a proposta inicial, um método de morfometria geométrica foi desenvolvido e as comparações com os métodos da literatura também foram testados.

O trabalho realizado com morfometria baseado em landmarks morfológicos associou informações geométricas e estatísticas às formas biológicas.

Com relação as medidas, todas foram usadas como entradas nos classificadores e também obteve-se esquemas hierárquicos (taxonomias) das espécies de algumas medidas escolhidas.

Desse modo, a descrição dos testes padrões da variação na forma morfológica dentro de uma população ou entre populações foi fundamental para definir os limites individuais evolucionários independentes na natureza.

Portanto, a curvatura digital como uma característica geométrica, nos surpreendeu quanto aos resultados por mostrar ser eficiente para a análise de formas biológicas.

Os resultados com esta característica simples obteve acertos no método ward de classificação hierárquica de 70,89% para região norte e de 73,66% para região sul com amostras de crânio de roedores, estes foram bem satisfatórios em relação a curvatura contínua, lembrando que a curvatura contínua é lenta e exige um alto poder computacional.

Nos métodos de seleção de atributos como parte da validação dos resultados extraídos das medidas de curvatura digital observamos que na combinação dos ângulos, o ângulo de número 5 (sutura entre o pré-maxilar frontal e uma ponte antorbital do maxilar) apareceu quase que em todas as combinações, levando a concluir que este marco morfológico na imagem, exerce uma influência bem significativa podendo ser uma característica bem relevante no processo de separação das classes. Os resultados ficaram na faixa de 70,0 % a 80,0 % de acerto.

Portanto, quanto aos métodos conclui-se que a curvatura digital apresentou melhores resultados, uma vez que além de ter apresentado resultados tão bons ou melhores que alguns outros métodos, foram aqueles que exigiram menor poder computacional, tempo de processamento e complexidade de implementação.

Sobre os métodos de classificação supervisionados e não supervisionados, observamos no método supervisionado que a análise canônica para a vista lateral do esqueleto de *T. apereoides*, considerando razão e curvatura, obteve uma melhor separação deixando bem definido duas regiões geográficas norte e sul. E com as amostras de canídeos a separação foi entre os sexo feminino e masculino.

Pode se dizer que as análises foram razoavelmente boas para ambos tipos de curvatura, mas a curvatura digital permitiu uma melhor separação usando análise discriminante. Este resultado foi um tanto surpreendente por que a curvatura contínua é conhecida por apresentar uma representação mais completa da forma original. As explicações para tais resultados talvez seja porque o método de curvatura contínua implique numa representação dos esqueletos com formas de contornos fechados. Isso

nos leva a entender que a curvatura digital obteve uma representação mais significativa e mais robusta, facilitando a separação entre os indivíduos das duas regiões geográficas.

Quanto a classificação não supervisionada, o uso do clustering hierárquico junto com a técnica ward e com o uso do dendrograma demonstrou ser bastante eficiente.

Por fim, pode-se afirmar que o desempenho dos métodos implementados e técnicas estudadas foram bastante satisfatórios, uma vez que atingiram as expectativas de acerto quanto a localização geográfica, indicando uma correspondência geográfica bem definida entre os indivíduos.

6.2 - Desenvolvimentos Futuros

Uma medida a ser tomada na continuação do desenvolvimento do trabalho é o uso do método para outras aplicações em morfologia biológica ou mesmo para outras áreas.

Podemos também procurar aperfeiçoar o método e aplicar em imagens 3D.

6.3 - PRODUÇÃO CIENTÍFICA, TECNOLÓGICA

6.3.1 - PRODUÇÃO BIBLIOGRÁFICA

6.3.1.1 Textos em revistas

1 Luciano da F. Costa; Sérgio F. dos Reis; **Renata A. T. Arantes**, Ana C. R. Alves and Giancarlo Mutinari. Biological shape analysis by digital curvature. Pattern Recognition, Volume 37, Issue 3, Pages 515-524.

Referências adicionais: Estados Unidos/Inglês; Meio de divulgação: Impresso;

Homepage: <http://www.elsevier.nl/inca/publications/store/3/2/8/>.

2 C. M Takemura, R.M. Cesar-Jr, **R. A. T. Arantes**, L. da F. Costa, E. Hingst-Zaher, S.F dos Reis, V. Bonato. Morphometrical Data Analysis using Wavelets. Real Time Imaging, (Submetido).

6.3.2.1 - PARTICIPAÇÃO EM EVENTOS

1 *Application of wavelets in the analysis of sexual dimorphism in the skull of South American canids (Mammalia, Canidae)*. Erika Hingst-Zaher, Celina Maki Takemura, Roberto Marcondes Cesar-Jr, **Renata A. T. Arantes**, Luciano da Fontoura Costa; 1 st International Conference on Bioinformatics and Computational Biology. 2003.; USP; Centro de Convenções; Ribeirão Preto, Brasil, 14 a 16 de maio de 2003.

2 *Biological shape analysis by techniques based on digital curvature: A comparison*. **Renata A. T. Arantes**, Luciano da F. Costa, Sérgio F. dos Reis; 1 st International Conference on Bioinformatics and Computational Biology. 2003. USP, Centro de Convenções, Ribeirão Preto, Brasil, 14 a 16 de maio de 2003.

3 VII Workshop da Pós Graduação do IFSC. 2003.
Referências adicionais: Brasil; Nome do evento: Política Científica e Pós-Graduação;
Nome da instituição promotora: Instituto de Física de São Carlos; Local: USP; Cidade: São Carlos. Participação: Comissão Organizadora

4 *Análise e Classificação de Formas Biológicas*, **Renata Antonia Tadeu Arantes**, Luciano da Fontoura Costa, I Workshop Regional de Engenharia Biomédica. 2002. Departamento de Engenharia Elétrica da EESC – USP, São Carlos, SP, Brasil, 17 e 18 de outubro de 2002.

5 *Análise e Classificação de Formas Biológicas*, **Renata Antonia Tadeu Arantes**, Luciano da Fontoura Costa, VI Workshop de Pós-Graduação em Física do Instituto de Física de São Carlos – USP, Brasil, 25 a 27 de setembro de 2002.

6 *Geographic Differentiation in the Rodent *Thrichomys Apereoides* using a Digital and Continuous Curvature*, **Renata A. T. Arantes**, G. Mutinari, Luciano da F. Costa, Sérgio F. dos Reis, Biomat 2003, III Brazilian Symposium of Mathematical and Computational Biology, novembro 22 a 26 de 2003, Instituto FIOCRUZ, Rio de Janeiro, RJ, Brasil

Capítulo 7

Referências Bibliográficas

[Allais,D.C (1964)]. “The selection of measurements for prediction,” D.C Allais, Tech. Rept. 6103-9, Stanford Electronics Lab., Stanford, Calif., November 1964.

[Anderberg, M. R. (1973)]. *Cluster Analysis for Applications*. New York: Academic Press.

[Atchley, W. R & Hall, B. K. (1991)]. A model for development and evolution of complex morphological structures. *Biol. Rev.* **66**,101-157.

- [Atchley, W. R., Cowley, D. E., Vogl, D. & McLellan, T. (1992)]. Evolutionary divergence, shape change, and genetic correlation structure in the rodent mandible. *Syst. Biol.* **41**, 196-221.
- [Attneave, F. (1954)]. Some informational aspects of human visual perception. *Psych. Rev.* **61**, 183-193.
- [Badyaev, A. V. & Foresman, K. R. (2000)]. Extreme environmental change and evolution: stress-induced morphological variation concordant with patterns of evolutionary divergence in shrew mandibles. *Proc. R. Soc. Lond. B* **267**, 371-437.
- [Bookstein, F.L (1986)]Size and shape spaces for landmark data in two dimensions, *Statistical Science*, 1:18 1-242.
- [Bookstein, F.L (1989a)]"Size and shape": a comment on semantics. *Syst. Zool.* 38:173-180.
- [Bookstein, F.L (1989b)] Principal warps: thin-plate splines and the decomposition of deformations. *IEEE. Transactions on Pattern Analysis and Machine Intelligence*, 11: 567-585.
- [Bookstein, F.L (1990)] Four metrics for image variation. In proceedings of the XI International conference on information processing in medical imaging (Ortendahl, D. & J. Llacer, eds) Alan R. Liss. Inc. New York.
- [Bookstein, F.L (1991)] Morphometric tools for landmark data. Cambridge University Press. 435 pp.
- [Bookstein, F.L (1996a)] Combining the tools of geometric morphometrics, p. 131-151. In: Marcus,L.; Corti; A.Loy; G.Naylor; D.E.Slice (eds), *Advances in morphometrics*. NATO ASI Series A: Life Sciences, vol.284. Plenum Publishing Corp, New York.

- [Bookstein, F.L (1996b)] Standard formula for the uniform shape component in landmark data, p. 153-168. In : Marcus, L;M. Corti; A.Loy; G.Naylor; D.E. Slice (eds), *Advances in morphometrics*. NATO ASI Series A: Life Sciences, vol 284. Plenum Publishing Corp, New York.
- [Blackit(1965)]. *Morphometrics*, p. 225-249. In: Waterman, T.H. & J. Morowitz (eds); *Theoretical and mathematical biology*. Blaisdell, New York.
- [Bonner, R.E.(1964)]. “On some clustering techniques,” R.E. Bonner, *IBM J.Research and Develop.*,vol 8, pp.22-32, January 1964.
- [Cesar, R. M. & Costa, L. F. (1996)]. Towards effective planar shape representation with multiscale digital curvature analysis based on signal processing techniques. *Pat. Recog.* 29, 1559-1569.
- [Cesar Jr., R. M. & Costa, L. da F., (1995)] Piecewise Linear Segmentation of Digital Contours in $O(N \cdot \log(N))$ through a Technique Based on Effective Digital Curvature Estimation, *Real-Time Imaging*, vol 1, pp 409-417, 1995.
- [Cesar, R.M. & Costa, L.F. (1997)]. Application and assessment of multiscale bending energy for morphometric characterization of neural cells. *Rev. Sci. Instr.* 68, 2177-2186.
- [Costa, L.F & Cesar, R.M (2001)]. *Shape Analysis and Classification: Theory and Practice*. Boca Raton, FL: CRC Press.
- [Costa, L. F.; Campos, A. G. & Manoel, E. T. M. (2001)]. An integrated approach to shape analysis: results and perspectives. In: *International Conference on Quality Control by Artificial Vision*, Le Creusot, France, 23-34.

[Duarte, Luiza C.; Monteiro, R. Leandro; Zuben, Fernando José Von; Reis, Sergio Furtado (2000)] “ *Variation in Mandible Shape in *Thrichomys apereoides* (Mammalia: Rodentia): Geometric Analysis of a Complex Morphological Structure,*”

[Desbrun, M.; Meyer, M.; Schröder, P. & Barr, A. H. (2000)]. Discrete differential-geometry operators in nD , Caltech, *Internal Report*.

[Dryden, I.L. & K.V. Mardia. (1998)] Statistical Shape Analysis. Wiley, New York.

[Duarte, L. C., Monteiro, L. R., Von Zuben, F. J. & Reis, S. F. (2000)]. Variation in mandible shape in *Thrichomys apereoides* (Mammalia: Rodentia): geometric analysis of a complex morphological structure. *Sys. Biol.* **49**, 563-578.

[Estes, (1965)]. “Measurement selection for linear discriminants used in pattern classification, “ S. E Estes Ph.D. dissertation, Stanford University, Stanford, Calif., 1964; also Research Rept. RJ-331, IBM Corp., San Jose, Calif., April 1.

[Fortier, J.J (1964)]. “Clustering Procedures,” J.J Fortier and H. Solomon; presented at the Internat. Symp. on Multivariate Analysis, Dayton, Ohio, June 1965; also Rept. 7, Dept. of Statistics, Stanford University, Stanford, Calif., March 20, 1964.

[Gonzalez C. Rafael., Woods E. Richard (2000)], “*Processamento de Imagens Digitais,*” Editora Edgard Blucher Ltda, 2000.

[Gonzalez C. Rafael., Woods E. Richard (1987)], “*Digital Image Processing,*” Addison-Wesley.

[Haralick, R.M, Stenberg, S.R., Zhuang, X. (1987)]. Image Analysis Using Mathematical Morphology. *IEEE Trans Pattern Analysis and Machine Intelligence*, **9** (4), p. 532-550. Julho 1987.

- [H. Holtelling. (1933)]. “Analysis of a Complex of Statistical Variables into Principal Components,” *J. Educ. Psychol.*, 24, 417-441, 498-520,1993.
- [K. Karhunen, (1947)]. *Über Lineare Methoden in der Wahrscheinlichkeitsrechnung*,” *Ann. Acad. Sci Fennicae. Ser. A.I.37* (English translation by I. Selin. “On linear Methods in Probability Theory,” Doc. T-131, The RAND Corp., Santa Monica, CA, 1960),1947.
- [Krzanowski, W. J. (1989)]. *Principles of Multivariate Analysis: a User's Perspective*. Oxford: Clarendon Press.
- [Kenneth R. Castleman (1995)]. *Digital Image Processing*; Prentice-Hall signal Processing Series.
- [L.F da Costa, Andréa G. Campos and E.T. M. Manoel (2001)] . *An Integrated Approach to Shape Analysis: Results and Perspectives*. International Conference on Quality Control by Artificial Vision.. Vol 1, 23-34.
- [Llenonart, J., Salat, J. & Torres, G. J. (2000)]. Removing allometric effects of body size in morphological analysis. *J. Theor. Biol.* 205:85-93.
- [Mares, M. A., Willig, M. R. & Lacher, T. E. (1985)]. The Brazilian caatinga in South American zoogeography: tropical mammals in dry regions. *J. Biogeogr.* **12**, 57-69.
- [MacQueen. J, (1967)]. “Some methods for classification and analysis of multivariate observations.” J. MacQueen, *Proc. 5th Berkeley Symp. on Statistics and Probability*. Berkeley, Calif: University of California Press, 1967, pp. 281-297.
- [Moojen, J.(1952)]. *Os Roedores do Brasil*. Rio de Janeiro, Instituto Nacional do Livro.

[Monteiro, R. Leandro; Reis, Sergio. F (1999)].; *Princípios de Morfometria Geométrica. Holos, Editora.*

[Patrick, E.A. (1966)]. “ Nonsupervised sequential classification and recognition of patterns,” E.A Patrick and J.C. Hancock, IEEE Trans. Information Theory, vol IT-12,pp 362-372, July 1966.

[P. Pudil, J. Novovicová, and J. Kittler (1994)]. Floating search methods in feature selection. *Pattern Recognition Letters*, 15:1119-1125, November.

[P. Somol, P. Pudil, J. Novovicová, and P. Paclík (1999)]. Adaptive floating search methods in feature selection. *Pattern Recognition Letters*, 20:1157-1163, 1999.

[Raff, R. A. (1996)]. *The Shape of Life: Genes, Development, and the Evolution of Animal Form.* Chicago: University Press.

[Reis, S. F., Durte, L. C., Monteiro, L. R. & Von Zuben, F. J. (2002a)]. 2002. Geographic variation in cranial morphology in *Thrichomys apereoides* (Rodentia: Echimyidae). I. Geometric descriptors and patterns of variation in shape. *J. Mamm. Journal of Mammalogy*. **83**(2): 333-344, May 2002

[Reis, S. F., Durte, L. C., Monteiro, L. R. & Von Zuben, F. J. (2002b)]. 2002. Geographic variation in cranial morphology in *Thrichomys apereoides* (Rodentia: Echimyidae). II Geographic units, morphological discontinuities, and sampling gaps. *Journal of Mammalogy*. **83**(2): 345-353, May 2002.

[Ringrose, T. J. (1996)]. Alternative confidence regions for canonical variate analysis. *Biometrika* **83**, 575-587.

[Rohlf,F.J. (1990)]. Fitting curves to outlines, p.167-178. In:Rohlf.F.J. & F.L Bookstein (eds), Proceedings of the Michigan morphometric workshop. Special publication number 2, The University of Michigan Museum of Zoology, Ann Arbor.

- [Rohlf, F. J. & D. E. Slice (1990)]. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst Zool.* 39:40-59.
- [Rohlf, F. J. (1999)]. 'tpsDig', Version 1.18, Stony Brook, NY: Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook.
- [Rohlf, F. J. & Marcus L. F. (1993)]. A revolution in morphometrics. *Trends Ecol. Evol.* 8, 129-132.
- [Rohlf, F. J. (1993)]. Relative Warp Analysis and an example of its application to mosquito wings, p.131-159. In: Marcus, L.F.; Morphometrics. Monografias, Museu Nacional de Ciencias Naturales, Madrid.
- [Rohlf, F. J. (1998)]. tpsSmall: is shape variation small? Version 1.11. Department of Ecology and Evolution, State University of New York at Stony Brook.
- [Romesburg, H. C. (1990)]. *Cluster Analysis for Researchers*. Florida: Robert E. Krieger.
- [S. Flach, J. Hargrave, S. Moore, G. Rempel, A. Valji and N. Walford. (1999)]. Photomodeler lite (eos systems inc.).
- [Streilein, K. E. (1982)]. The ecology of small mammals in the semiarid Brazilian caatinga. IV. Habitat selection. *Ann. Carnegie Mus.* 51, 31-343.
- [Thompson, W. D'arcy (1952)]. On Growth and Form. Volume I, Cambridge at the University Press.
- [Tucker, A. S. & Sharpe, P. T. (1999)]. Molecular genetics of tooth morphogenesis and patterning: the right shape in the right place. *J. Dent. Res.* 78, 826-834.
- [Walker, J. A. (1997)]. Ecological morphology of lacustrine threespine stickleback *Gasterosteus aculeatus* L. (Gasterosteidae) body shape. *Biol. J. Linn. Soc.* 61, 3-50.

[W. Siedleki and J. Sklansky, (1989)]. A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, 10:335-347, November.

[Young, I. T; Calvert, T. W. & Bowie, J. E. (1974)]. An analysis technique for biological shape I. *Info. Contr.* **25**, 357-370.

[Von Zuben, F. J., Duarte, L. C., Stangenhans, G., Pessoa, L. M. & dos Reis, S. F. (1998)]. Bootstrap confidence regions for canonical variates: Application to studies of evolutionary differentiation. *Biometrical Journal*. **40**(3), 327-339.