

Segmentação e Aquisição de Partes Genéricas para Reconhecimento

DÍBIO L. BORGES
ROBERT B. FISHER

Department of Artificial Intelligence
University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL, Scotland, UK
dibio, rbf@aifh.ed.ac.uk

Abstract. The ability to acquire generic parts of 3D objects is an important step for recognition of objects in 3D space. Most natural articulated objects can be described as having a small number of salient 3D parts. Range data provides ready-to-use information to derive a reasonably stable and accurate differential structure of the surfaces, which allow us to track discontinuities between these parts. In this paper we present a mixed and comprehensive approach to segmenting and extracting prototypical parts from 3D articulated objects. This approach enables the acquisition of generic volumetric parts of a 3D object for the purpose of recognition.

1 Introdução

Uma das principais questões sendo pesquisadas em Visão por Computador é a de como segmentar uma cena em regiões significativas que tenham uma interpretação visual estável e podem ser extraídas com segurança. Esse processo de segmentação é de especial importância como o primeiro estágio de um sistema que almeja o reconhecimento de objetos 3D articulados utilizando a noção de partes. Esse artigo descreve um processo de segmentação amplo que visa a identificação de primitivas volumétricas adequadas para posterior reconhecimento.

Para atingir a segmentação desejada é apresentada aqui uma abordagem mista e completa para identificar partes de um objeto articulado. Essa abordagem compreende estimativas estáveis e seguras da estrutura diferencial das superfícies do objeto, seguidas de um agrupamento dinâmico das discontinuidades detectadas na superfície e de um ajuste através de regressão não-linear de modelos deformáveis de superquádricas para as regiões segmentadas.

O objetivo final da segmentação é conseguir um conjunto de modelos deformáveis de superquádricas onde cada modelo do conjunto está relacionado a uma única parte do objeto. Parte-se do pressuposto que três tipos de discontinuidades (profundidade, orientação e pontos onde ocorrem mínimos das Direções Principais) detectadas localmente na estrutura diferencial da superfície do objeto podem ser agrupadas de uma maneira a se tornarem os contornos de partes salientes de um objeto 3D articulado. Este agrupamento é formulado como um processo de minimização de energia de curvas deformáveis, as quais são inicializadas em pontos de discontinuidade e en-

tão atraídas para caminhos de mudanças mínimas nas Direções Principais (com vizinhança fixa) até atingirem outras discontinuidades. Após este estágio o conjunto final das curvas indicará as regiões na superfície do objeto a serem representadas como partes separadas. Propõe-se também que superquádricas com deformações globais são capazes de capturar a noção de partes de um objeto, e nesse sistema implementa-se um ajuste por meio de regressão não-linear de superquádricas deformáveis para cada uma das regiões segmentadas com o intuito de extrair um modelo final do objeto para uso na identificação de sua forma e posição.

São utilizadas imagens codificadas por distâncias, as quais são adquiridas por um sensor 3D do tipo triangulação (a raio laser) construído em nosso laboratório. A partir da imagem CD (Codificada por Distancias) se faz necessário um procedimento estável e preciso para estimar as Curvaturas Gaussiana, Média e Principais assim como as Direções Principais para identificar e agrupar as discontinuidades.

A estrutura global da abordagem aqui descrita possui semelhanças a trabalho recente realizado por [Ferrie et al. (1990)], e difere do mesmo pela inclusão de condições para preservação de discontinuidades no processo de suavização e cálculos de consistência de curvatura, agrupamento dinâmico de todas as discontinuidades detectadas e pelo uso de modelos deformáveis de superquádricas para representar as partes do objeto. Uma das motivações desse trabalho é estabelecer um procedimento direto para extrair um conjunto significativo de partes deformáveis para objetos 3D articulados com a meta de possibilitar a identificação da forma e a localização espacial de uma

grande gama de objetos naturais e manufaturados.

O restante desse artigo divide-se em cinco (5) seções. Na Seção 2 descreve-se o cálculo das estimativas estáveis da estrutura diferencial das superfícies. A Seção 3 descreve o procedimento de agrupamento das discontinuidades guiado pelas Direções Principais na superfície. A representação em Superquádricas e o procedimento de ajuste por regressão não-linear dos modelos globalmente deformáveis são apresentados na Seção 4. As Seções 5 e 6 incluem respectivamente exemplos do processo de segmentação e conclusões e extensões do trabalho.

2 Estrutura Diferencial de Objetos 3D

O uso de técnicas de geometria diferencial para modelar e segmentar imagens CD tem sido tópico de importante investigação em Visão por Computador [Besl-Jain (1985)]. Uma abordagem comum é a de derivar *descrições qualitativas das superfícies* utilizando-se dos sinais das Curvaturas Gaussiana e Média estimadas das porções da superfície. Existem oito (8) formatos diferentes de superfícies que podem ser classificadas com os sinais das Curvaturas Gaussiana e Média [do Carmo (1979)]. Este tipo de segmentação tem sido aplicado a objetos 3D com superfícies planas e não-planas com relativo sucesso [Besl-Jain (1985)], [Trucco-Fisher (1992)].

A estrutura local de uma superfície 3D porém é mais rica, e ao invés de utilizar somente as Curvaturas Gaussiana (K) e Média (H) é possível também de se obter a partir de um ajuste local as Curvaturas Principais (κ_1, κ_2) e as Direções Principais (\vec{d}_1, \vec{d}_2). O problema com os métodos tradicionais para se calcular tais estimativas é que eles são instáveis e não precisos o bastante para o uso em processos de identificação. Para superar este problema combina-se aqui um procedimento de suavização que preserva as discontinuidades, juntamente com um algoritmo de consistência de dois estágios para se conseguir estimativas confiáveis para a estrutura local da superfície.

A eliminação de ruído em imagens CD anterior ao cálculo das curvaturas da superfície constitui-se em um módulo essencial no tratamento deste tipo de imagens devido aos efeitos da quantização. O uso de suavização gaussiana obscurece os pontos de discontinuidade na superfície, os quais devem ser preservados pois eles se constituem em características essenciais da superfície do objeto. [Cai (1989)] sugere o uso de suavização por difusão com prevenção contra a suavização sobre os pontos de discontinuidade ao invés da suavização gaussiana que é padrão para a remoção de ruído. [Trucco-Fisher (1992)] aprimoraram este procedimento de suavização impondo uma condição de contorno adaptativa na equação de

difusão e nesse trabalho este algoritmo de suavização é utilizado como pré-filtragem nas imagens CD.

As estimativas iniciais da estrutura local da superfície, $D(\mathbf{P}) = (\mathbf{P}, \kappa_{1P}, \kappa_{2P}, \mathbf{N}_P, \vec{d}_{1P}, \vec{d}_{2P})$, são conseguidas de maneira similar a outras abordagens onde para cada ponto $\mathbf{P} \in S$ ajusta-se localmente um parabolóide da forma $g(\mathbf{u}, \mathbf{v}) = a\mathbf{u}^2 + b\mathbf{u}\mathbf{v} + c\mathbf{v}^2$ sobre a superfície. A estrutura local $D(\mathbf{P})$ pode então ser calculada a partir de $g(\mathbf{u}, \mathbf{v})$. Duas particularidades do ajuste aqui adotado são: 1) A origem do sistema de coordenadas paramétrico $\langle \mathbf{u}, \mathbf{v}, g \rangle$ é colocada no ponto \mathbf{P} alinhando o eixo g com a normal, e 2) pontos com gradientes de orientação e profundidade muito altos são preservados.

As estimativas obtidas a partir do primeiro estágio são confiáveis para o propósito de classificação das superfícies pelos sinais de (H) e (K); contudo, elas não são confiáveis e precisas o suficiente para Direções Principais consistentes. Para se conseguir estimativas estáveis e precisas utiliza-se aqui um método desenvolvido inicialmente para aplicações em imagens tomográficas por Sander e Zucker [Sander-Zucker (1990)]. O método consiste em uma minimização iterativa onde para cada ponto \mathbf{P} na superfície os Conjuntos de Darboux $D()$ dos pontos na vizinhança são extrapolados a partir de $D(\mathbf{P})$. A partir deste novo conjunto de $D()$ na vizinhança de \mathbf{P} , uma função custo é montada para minimizar a variação em $D(\mathbf{P})$ utilizando-se as seguintes medidas de erro:

$$E_1 = \min \sum_{\alpha=1}^n \|\mathbf{N}_P - \mathbf{N}_{()P\alpha}\|^2 + (\kappa_1 - \kappa_{1()P\alpha})^2 + (\kappa_2 - \kappa_{2()P\alpha})^2 + \lambda((\mathbf{N}_P \cdot \mathbf{N}_{()P\alpha}) - 1). \quad (1)$$

$$E_2 = \min_{\theta} \sum_{\alpha=1}^n [1 - (\vec{d}_{1P}(\theta) \cdot \vec{d}_{1P\alpha})^2]. \quad (2)$$

onde α indexa toda a vizinhança do ponto.

A minimização está sujeita às seguintes condições sobre as Direções Principais ($\vec{d}_{1P}, \vec{d}_{2P}$) e a normal (\mathbf{N}_P):

$$\begin{aligned} (\vec{d}_{1P} \cdot \mathbf{N}_P) &= 0; & (\mathbf{N}_P \cdot \mathbf{N}_{()}) &= 1; \\ (\vec{d}_{1P} \cdot \vec{d}_{1P}) &= 1 \end{aligned} \quad (3)$$

Para maiores detalhes sobre o algoritmo de minimização ver [Sander-Zucker (1990)].

Qualquer uma das Direções Principais pode ser usada em E_2 . A convergência é controlada por uma medida composta $\mathcal{U}_S^{(i)}$ de E_1 e E_2 , onde procura-se atingir uma variação mínima da medida ao longo das

iterações.

$$\mathcal{U}_S^{(i)} = \sum_j E_{j_1}^{(i)} + E_{j_2}^{(i)}; \quad \mathbf{P}_j \in S. \quad (4)$$

3 Partes dos Objetos

O problema de se definir o que constitui uma parte de um objeto é de grande importância e ao mesmo tempo de difícil solução. Regularidades são percebidas por nós a partir do mundo visual e estas regularidades servem para ajudar na organização de porções da imagem em partes significativas. Um conjunto destas partes se constitui em um modelo aproximado do objeto.

Padrões de discontinuidades nas superfícies se constituem em evidência importante para se definir a extensão de uma parte do objeto. Dois tipos de discontinuidades úteis aqui são as discontinuidades de profundidade e discontinuidades de orientação. Hoffman e Richards [Hoffman-Richards (1985)] propõem que um outro princípio importante para decomposição em partes é a "regularidade de transversalidade", a qual sugere que os pontos de mínimo de cada Curvatura Principal são pontos do contorno entre as partes.

As discontinuidades de profundidade são encontradas de maneira simples estabelecendo-se um limiar na imagem CD buscando identificar assim mudanças abruptas ($> \text{limiar}$) entre os valores dos *pixels* naquela vizinhança. As discontinuidades de orientação são por sua vez detectadas como sendo o local onde o ângulo entre as normais à superfície, dentro de uma pequena vizinhança, é maior que um valor pré-determinado. O ângulo nessa implementação é igual a $\pi/6$. As discontinuidades nas Curvaturas Principais são definidas como os pontos onde:

$$\kappa'_1(x, y) |_{\vec{d}_1} = 0 \quad E \quad \kappa_1(x, y) < 0. \quad (5)$$

ou,

$$\kappa'_2(x, y) |_{\vec{d}_2} = 0 \quad E \quad \kappa_2(x, y) < 0. \quad (6)$$

Todos os três tipos de discontinuidades, profundidade, orientação e os pontos de mínimo de curvatura principal são importantes na determinação da extensão de uma parte do objeto. O problema a ser resolvido a partir de obtidas as discontinuidades é o de como combinar todas elas de modo a isolar as partes mais salientes de objetos articulados [Marr-Nishihara (1978)].

Sugere-se aqui como uma estrutura adequada para derivar estes contornos trabalhar com a noção de ajuste de curvas por minimização de energia, ou contornos ativos [Kass et al. (1987)]. Nesse trabalho,

contornos ativos de comprimento fixo são colocados inicialmente em pontos de discontinuidades detectados como descrito anteriormente e o crescimento destes contornos seguem a direção de um campo potencial local (na direção do plano tangente) [Zucker et al. (1988)], o qual é derivado a partir da estrutura diferencial da superfície em 3D (ver Seção anterior).

A formulação do problema como ajuste de curva por minimização de energia proporciona um procedimento dinâmico de combinar os pontos segmentados do objeto. A estrutura local da superfície entra como uma condição necessária e importante na minimização. O processo pode ser visto como uma minimização da seguinte curva deformável:

$$\ell(s, t) = (x(s, t); y(s, t)). \quad 0 \leq s \leq 1 \quad (7)$$

com energia cinética funcional $T(\ell)$:

$$T(\ell) = \frac{1}{2} \int_0^1 \mu |\ell_t|^2 ds. \quad (8)$$

onde μ é uma constante. A energia potencial funcional $U(\ell)$ é:

$$U(\ell) = \frac{1}{2} \int_0^1 (\omega_1(s) |\ell_s|^2 + \omega_2(s) |\ell_{ss}|^2 + I(\ell) + S(\ell)) ds. \quad (9)$$

onde:

- $\omega_1(s)x_s$ é o fator de tensão da curva;
- $\omega_2(s)x_{ss}$ é o fator de rigidez da curva;
- $I(\ell)$ é o campo potencial da estrutura local da superfície;
- $S(\ell)$ é a força entre curvas muito próximas.

A solução procurada do sistema acima diz respeito às curvas $x(s, t)$ e $y(s, t)$ para as quais

$$\int_{t_1}^{t_2} T(\ell) - U(\ell) dt. \quad (10)$$

é um mínimo.

4 Representação por Superquádricas

Primitivas volumétricas capturam muito bem a noção de partes em objetos naturais [Marr-Nishihara (1978)], e em objetos manufaturados [Biederman (1987)]. Estas primitivas são também adequadas para indexação desde que elas possuam um número relativamente pequeno de parâmetros quando comparadas a outras primitivas.

Uma superfície superquádrica pode ser definida pelo seguinte vetor no espaço 3D:

$$x(\eta, \omega) = \begin{bmatrix} a_1 \cos^{\epsilon_1}(\eta) \cos^{\epsilon_2}(\omega) \\ a_2 \cos^{\epsilon_1}(\eta) \sin^{\epsilon_2}(\omega) \\ a_3 \sin^{\epsilon_1}(\eta) \end{bmatrix} \quad (11)$$

onde,

$$-\pi/2 \leq \eta \leq \pi/2. \quad (12)$$

$$-\pi \leq \omega < \pi. \quad (13)$$

Os parâmetros a_1 , a_2 e a_3 definem a extensão de uma superquádrica nas coordenadas x , y e z respectivamente. ϵ_1 e ϵ_2 são os parâmetros de quadratura. A capacidade de modelagem das superquádricas pode ser expandida ainda mais com a introdução de deformações do tipo *encurvar* e *adelgaçar*. Outras deformações por exemplo que criam cavidades [Solina-Bajcsy (1990)], e deformações locais [Terzopoulos-Metaxas (1990)] também são possíveis, embora elas precisam de diferentes métodos para a recuperação dos parâmetros a partir de imagens CD.

Para se construir uma superfície superquádrica a partir dos dados de entrada, a equação (11) pode ser manipulada para se conseguir a seguinte equação implícita:

$$\left(\left(\frac{x}{a_1} \right)^{2/\epsilon_2} + \left(\frac{y}{a_2} \right)^{2/\epsilon_2} \right)^{\epsilon_2/\epsilon_1} + \left(\frac{z}{a_3} \right)^{2/\epsilon_1} = 1. \quad (14)$$

Referindo a esta equação pode-se determinar a função de posição [Solina-Bajcsy (1990)]:

$$\mathbf{F}(x, y, z) = \left(\left(\left(\frac{x}{a_1} \right)^{2/\epsilon_2} + \left(\frac{y}{a_2} \right)^{2/\epsilon_2} \right)^{\epsilon_2/\epsilon_1} + \left(\frac{z}{a_3} \right)^{2/\epsilon_1} \right)^{\epsilon_1} \quad (15)$$

A função (15) determina onde um dado ponto $[x, y, z]$ se localiza relativamente a superfície da superquádrica

- $\mathbf{F}(x, y, z) = 1$, ponto (x, y, z) está na superfície.
- $\mathbf{F}(x, y, z) > 1$, ponto (x, y, z) está fora da superfície.
- $\mathbf{F}(x, y, z) < 1$, ponto (x, y, z) está dentro da superfície.

A equação (15) define a superfície superquádrica em um sistema de coordenadas centrado no objeto (x_s, y_s, z_s) . Utilizando-se um sistema de coordenadas genérico (com origem em (p_x, p_y, p_z) e ângulos de Euler

(ϕ, θ, ψ) pode-se expressar a função (15) em posição genérica por:

$$\mathbf{F}(x_w, y_w, z_w) = \mathbf{F}(x_w, y_w, z_w; a_1, a_2, a_3, \epsilon_1, \epsilon_2, \phi, \theta, \psi, p_x, p_y, p_z). \quad (16)$$

A função \mathbf{F} expandida, equação (16) possui 11 parâmetros sem a inclusão das deformações globais. Ao incluir as deformações (encurvar e adelgaçar) o número total de parâmetros se torna igual a 15,

$$\mathbf{F}(x_w, y_w, z_w) = \mathbf{F}(x_w, y_w, z_w; a_1, \dots, a_{11}, b, \beta, T_x, T_y) \quad (17)$$

onde,

- (x_w, y_w, z_w) é um ponto na superfície da superquádrica
- a_1, \dots, a_{11} são os 11 parâmetros da superquádrica sem deformação
- b, β são os parâmetros relativos à deformação de encurvar ao longo do eixo z
- T_x, T_y são os parâmetros relativos à deformação de adelgaçar ao longo do eixo z

Utilizando um método de minimização não-linear por mínimos-quadrados, o algoritmo de Levenberg-Marquadt, também usado por [Solina-Bajcsy (1990)], os 15 parâmetros da superfície superquádrica podem ser estimados através da minimização da seguinte expressão:

$$\sum_{i=1}^N [\mathbf{R}(x_i, y_i, z_i; a_1, \dots, a_{15})]^2. \quad (18)$$

onde,

$$\mathbf{R} = \sqrt{a_1 a_2 a_3} (\mathbf{F} - 1). \quad (19)$$

e N é o número de pontos observados. Na implementação aqui descrita utilizam-se todos os pontos adquiridos da imagem CD em cada porção da superfície do objeto, os quais são isolados pelos contornos encontrados anteriormente.

Para incluir as deformações deve haver uma prioridade na ordem de recuperação de cada uma delas. De acordo com Leyton [Leyton (1987)] a ordenação a seguir preserva melhor a estrutura do objeto.

$$\text{Transl.}(\text{Rot.}(\text{Enc.}(\text{Adelg.}(x)))) \quad (20)$$

5 Exemplos

Esta Seção inclui exemplos da segmentação de um objeto articulado 3D. A Figura 1 mostra uma imagem CD de um objeto com articulações, a Figura 2 mostra as discontinuidades detectadas inicialmente pelos métodos descritos na Seção 3. A Figura 3 mostra a extensão das discontinuidades (usando os contornos ativos) que resultam nos contornos das partes do objeto. A Figura 4 apresenta o modelo final recuperado ajustando uma superfície superquádrica para cada região (parte) segmentada do objeto. O sistema de modelagem utilizado para se construir as superfícies superquádricas após a recuperação dos parâmetros na minimização é o "Thingworld" [Pentland-Sclaroff (1991)].

6 Conclusões

A segmentação de objetos 3D em partes (como protótipos) proporciona a um sistema artificial de visão a possibilidade de raciocínio sobre a localização, a forma e a função de objetos complexos, naturais e manufaturados. A estrutura diferencial da superfície do objeto constitui-se em informação essencial sobre onde se deve segmentar um objeto com articulações. Até o momento pouquíssimas tentativas em se resolver este problema têm sido relatadas na literatura, nenhuma porém com grande sucesso. Propõe-se aqui uma solução completa à segmentação em partes de objetos com articulações, a qual é adequada para uma abordagem de reconhecimento baseada em protótipos de partes.

Testes iniciais baseados neste processo de segmentação aqui apresentado mostra resultados promissores. Uma das utilizações imediatas para esse processo de segmentação é como primeiro estágio de um sistema de reconhecimento de objetos em 3D para objetos articulados naturais e manufaturados. Esta classe de objetos representa um conjunto numeroso de objetos encontrados em muitas situações práticas, e a segmentação dos mesmos em um conjunto pequeno de partes 3D é um passo importante na direção de se localizar e identificar as suas formas completas através de um sistema de visão artificial.

Agradecimentos. O primeiro autor gostaria de agradecer ao CNPq/MCT pelo auxílio financeiro ao projeto, recebido através do programa de bolsas no exterior. Dúbio L. Borges agradece também a S. Sclaroff e A. Pentland (MIT, Media Lab) pelas discussões e pela ajuda na modelagem de superquádricas.

7 Referências

P. Besl and R. Jain, Three-dimensional object recog-

- inition, *ACM Comp. Surveys*, **17** (1985) 75–145.
- I. Biederman, Recognition-by-components: A theory of human image understanding, *Psychological Review*, **94** (1987) 115–147.
- L. Chi, A diffusion smoothing approach to sculptured surfaces, *The Mathematics of Surfaces III*, In D. Handscomb, Clarendon Press, 1989.
- M. P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, 1976.
- F. Ferrie, J. Lagarde and, P. Whaite, *Proc. First European Conf. Computer Vision ECCV-90, Antibes, France*, (1990) 387–396.
- D. Hoffman and W. Richards, Parts of recognition, *Visual Cognition*, In S. Pinker, MIT Press, (1985) 65–96.
- M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active contour models, *Int. J. Computer Vision*, **1** (1987) 321–331.
- M. Leyton, Nested structures of control: An intuitive view, *Computer Vision, Graphics, and Image Processing*, **37** (1987) 20–53.
- D. Marr and K. Nishihara, Representation and recognition of the spatial organization of three-dimensional shapes, *Proc. Royal Society of London B*, **200** (1978) 269–294.
- A. Pentland, Perceptual organization and the representation of natural form, *Artificial Intelligence*, **28** (1986) 293–331.
- A. Pentland and S. Sclaroff, Closed-form solutions for physically based shape modeling and recognition, *IEEE Trans. P.A.M.I.*, **13** (1991) 715–729.
- P. Sander and S. Zucker, Inferring surface trace and differential structure from 3-D images, *IEEE Trans. P.A.M.I.*, **12** (1990) 833–854.
- F. Solina and R. Bajcsy, Recovery of parametric models from range images: The case for superquadrics with global deformations, *IEEE Trans. P.A.M.I.*, **12** (1990) 133–147.
- D. Terzopoulos and D. Metaxas, Dynamic 3D models with local and global deformations: Deformable superquadrics, *Proc. Third Int. Conf. Computer Vision ICCV-90, Osaka, Japan*, (1990) 606–615.
- E. Trucco and R. Fisher, Computing surface-based representations from range images, *Proc. IEEE Int. Symp. Intelligent Control ISIC-92, Glasgow*, (1992) 275–280.
- S. Zucker, C. David, A. Dobbins and L. Iverson, The organization of curve detection: Coarse tangent fields and fine spline coverings, *Proc. Second Int. Conf. Computer Vision ICCV-88, Florida*, (1988) 568–577.

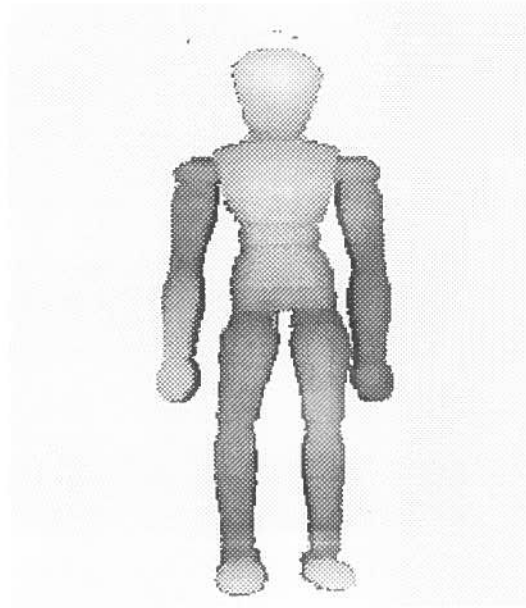


Figura 1: Imagem CD de um objeto articulado

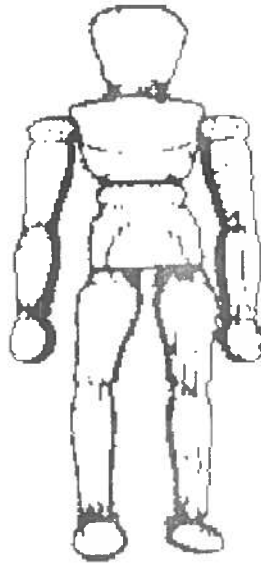


Figura 2: Discontinuidades detectadas

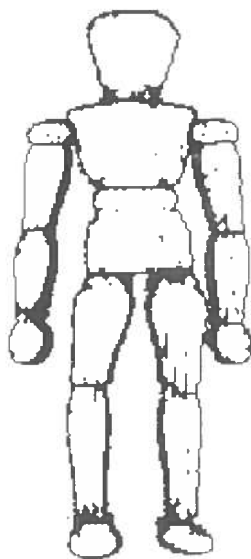


Figura 3: Contornos após extensão das regiões segmentadas

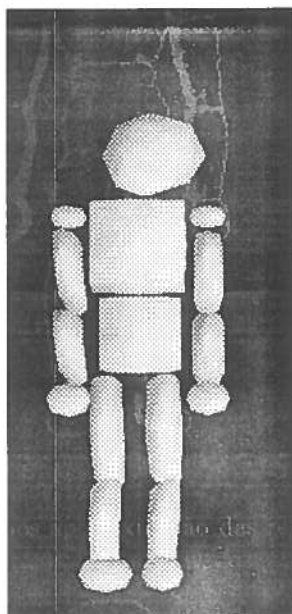


Figura 4: Modelos recuperados de superquádricas deformáveis