

Fábio José Ayres

**Segmentação e estimação da
composição histológica da massa
tumoral em imagens de CT de
neuroblastomas**

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para a obtenção do título de Mestre em Engenharia.

São Paulo
2001

Resumo

Neuroblastoma é o tumor maligno sólido extracranial mais comum em crianças; é a terceira malignidade mais comum da infância. Avanços em radiologia tornaram possível a detecção e o preciso estadiamento da doença. Entretanto, não existem atualmente métodos de imageamento que possam transcender a detecção e análise qualitativa, em direção ao estabelecimento quantitativo da composição histológica da massa tumoral primária em neuroblastomas. Tal análise quantitativa proveria informações importantes para radiologistas e médicos, que podem levar a melhores tratamentos e práticas de acompanhamento da doença, e mesmo a evitar cirurgias de revisão em alguns casos.

O problema que investigamos neste trabalho é o aperfeiçoamento da análise da massa tumoral primária, em pacientes com neuroblastoma avançado, usando exames de tomografia computadorizada por raios-X (CT, *computed tomography*). Para atingir esta meta, propomos uma metodologia para estimar o conteúdo histológico da massa que compreende: uma técnica para segmentação semi-automática da massa tumoral primária em imagens de CT de neuroblastoma, e um método estatístico para estimar, a partir das imagens de CT segmentadas, a composição histológica do tumor primário. Os resultados do método serão posteriormente comparados com os resultados da análise histológica das correspondentes massas removidas cirurgicamente.

Abstract

Neuroblastoma is the most common extracranial solid malignant tumor in children; it is the third most common malignancy of childhood. Advances in radiology have made possible the detection and accurate staging of the disease. Nevertheless, there is no imaging method available at present that can go beyond detection and qualitative analysis, towards quantitative assessment of the histological composition of the primary tumor mass in neuroblastoma. Such quantitative analysis could provide important information for the radiologist and the physician that could lead to better treatment and follow-up, and even to the avoidance of delayed surgery.

The problem that we investigate in the present report is the improvement of the analysis of the primary tumor mass, in patients with advanced neuroblastoma, using X-ray computed tomography (CT) exams. To achieve this goal, we propose a methodology for the estimation of the histological content of the mass that comprises a technique for semi-automatic segmentation of the primary tumor mass in CT images of neuroblastoma and a statistical method to estimate, from segmented CT images, the histological composition of the primary tumor. The results of the method will later be compared with the results of histological analysis of surgically resected masses.

©1998 Universidade de São Paulo

Ayres, Fábio José

Segmentação e estimação da composição histológica da massa tumoral em imagens de CT de neuroblastomas

1. Processamento de imagens 2. Tomografia computadorizada 3. Neuroblastoma 4. Extração de características

I. Título II. Ayres, Fábio J.

1998

Composto com o sistema de documentação L^AT_EX

Conteúdo

1	Introdução	8
1.1	Objetivos	9
1.2	Motivação	10
1.3	Metodologia	10
1.4	Estrutura do texto	11
2	Análise de Neuroblastomas	12
2.1	Neuroblastoma	12
2.1.1	Incidência	13
2.1.2	Estadiamento	13
2.1.3	Análise radiológica	15
2.1.4	Diagnóstico auxiliado por computador	17
2.2	Caracterização de Tecidos usando CT	19
3	Segmentação e Estimação Histológica	22
3.1	Segmentação	23
3.1.1	Estimação do Histograma	26
3.2	Estimação da Composição Histológica	27
4	Implementação e Análise dos Resultados	33
4.1	Segmentação	34
4.2	Estimação da Composição Histológica	39
4.2.1	Teste com dados sintéticos	39
4.2.2	Teste com dados reais	52
5	Conclusão e Trabalhos Futuros	62

Lista de Figuras

2.1	Imagem de CT de um paciente com neuroblastoma Estágio III, com o contorno da massa desenhado por um radiologista (MV).	17
3.1	Representação esquemática do método proposto para a análise de neuroblastomas. EM: Expectation-maximization.	24
4.1	Segmentação manual da massa tumoral, caso ‘A’.	36
4.2	Conectividade fuzzy da região obtida por crescimento de região para o tumor do caso ‘A’ com o contorno manual superposto para comparação. Veja também a Figura 4.1.	36
4.3	Segmentação manual da massa tumoral, caso ‘B’.	37
4.4	Conectividade fuzzy da região obtida por crescimento de região para o tumor do caso ‘B’ com o contorno manual superposto para comparação. Veja também a Figura 4.3.	37
4.5	Segmentação manual da massa tumoral, caso ‘C’.	38
4.6	Conectividade fuzzy da região obtida por crescimento de região para o tumor do caso ‘C’ com o contorno manual superposto para comparação. Veja também a Figura 4.5.	38
4.7	Resultados do algoritmo EM para o modelo de mistura de Gaussianas usando dados simulados com núcleos amplamente espaçados. As funções densidade de probabilidade para os modelos com três, quatro e cinco Gaussianas não são visíveis pois se sobrepõem à função original. Veja também 4.2.	44
4.8	Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados simulados com núcleos amplamente espaçados. A linha sólida representa a função densidade de probabilidade para o dado original, e as linhas tracejadas representam as componentes estimadas. Veja também Tabela 4.2 e Figura 4.7.	45
4.9	Resultados do algoritmo EM para modelo de mistura de Gaussianas usando dados simulados com espaçamento intermediário entre núcleos (veja também Figura 4.7 e Figura 4.11). As funções densidade de probabilidade para os modelos com três, quatro e cinco Gaussianas não são visíveis pois se sobrepõem à função original. Veja também Tabela 4.3.	47

4.10	Resultados do algoritmo EM para modelo de mistura de Gaussianas usando dados simulados com espaçamento intermediário entre núcleos (veja também Figura 4.7 e Figura 4.11). A linha sólida representa a função densidade de probabilidade para o dado original, e as linhas tracejadas representam as componentes estimadas. Veja também Tabela 4.3 e Figura 4.9.	48
4.11	Resultados do algoritmo EM para modelo de mistura de Gaussianas usando dados simulados com espaçamento estreito entre núcleos (veja também Figura 4.7 e Figura 4.9). Todos os modelos se sobrepõem à função original. Veja também Tabela 4.4.	50
4.12	Resultados do algoritmo EM para modelo de mistura de Gaussianas usando dados simulados com espaçamento estreito entre núcleos. A linha sólida representa a função densidade de probabilidade para o dado original, e as linhas tracejadas representam as componentes estimadas. Veja também Tabela 4.4 e Figura 4.11.	51
4.13	Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'A'). Ver também tabela 4.5.	55
4.14	Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'A'). A linha sólida representa a função densidade de probabilidade para o dado original e as linhas tracejadas representam as componentes estimadas. veja também tabela 4.5 e figura 4.13.	56
4.15	Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'B'). Ver também tabela 4.6.	57
4.16	Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'B'). A linha sólida representa a função densidade de probabilidade para o dado original e as linhas tracejadas representam as componentes estimadas. veja também tabela 4.6 e figura 4.15.	58
4.17	Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'C'). Ver também tabela 4.7.	59
4.18	Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'C'). A linha sólida representa a função densidade de probabilidade para o dado original e as linhas tracejadas representam as componentes estimadas. veja também tabela 4.7 e figura 4.17.	60

Lista de Tabelas

2.1	International Neuroblastoma Staging System. Reproduzido de Cohen et al. [1].	14
2.2	Sobrevivência ao neuroblastoma por idade. Reproduzido de Abramson [2].	14
2.3	Sobrevivência ao neuroblastoma por estágio. Reproduzido de Abramson [2].	14
2.4	Distribuição dos pacientes com neuroblastoma segundo o estágio da doença. Reproduzido de Abramson [2].	15
2.5	Valores de CT em unidades Hounsfield (HU) para alguns tipos de tecidos abdominais. Reproduzido de Mategrano et al. [3].	20
4.1	Descrição dos casos analisados	34
4.2	Resultados do algoritmo EM, com dados simulados que consistem de núcleos amplamente espaçados.	43
4.3	Resultados do algoritmo EM para dados simulados, espaçamento mediano entre os núcleos.	46
4.4	Resultados do algoritmo EM para dados simulados, espaçamento estreito entre os núcleos.	49
4.5	Parâmetros do modelo de mistura de Gaussianas estimados para o tumor do caso 'A', manualmente segmentado. A região do tumor contém 946782 voxels, o valor médio para o tumor é 45.22 HU e o desvio padrão é 11.60 HU.	54
4.6	Parâmetros do modelo de mistura de Gaussianas estimados para o tumor do caso 'B', manualmente segmentado. A região do tumor contém 304732 voxels, o valor médio para o tumor é 54.88 HU e o desvio padrão é 20.25 HU.	54
4.7	Parâmetros do modelo de mistura de Gaussianas estimados para o tumor do caso 'C', manualmente segmentado. A região do tumor contém 24892 voxels, o valor médio para o tumor é 44.83 HU e o desvio padrão é 19.38 HU.	61

Capítulo 1

Introdução

Neuroblastoma é um tumor maligno que se origina de células indiferenciadas da crista neural, e que pode surgir em qualquer ponto ao longo dos gânglios simpáticos ou na medula adrenal [4]. É considerado uma entidade biológica fascinante e enigmática que representa, ao mesmo tempo, um dos melhores exemplos de regressão espontânea para um tumor maligno e um dos tumores com o menor potencial de cura para alguns grupos [4, 5]. O neuroblastoma é o tumor extracranial sólido maligno mais comum em crianças [2]; é a terceira malignidade mais comum da infância.

Em exames de tomografia computadorizada (*Computed Tomography* - CT), o neuroblastoma abdominal é visto como uma massa de tecido mole, usualmente suprarenal ou paravertebral, de forma irregular, lobulada, estendendo-se do flanco em direção à linha média, e sem estrutura capsular. A massa tende a ser heterogênea em função da existência de necrose tumoral misturada a tumor viável, e contém calcificações em 85% dos pacientes. Calcificações são usualmente densas, amorfas, e heterogêneas em aparência. Em algumas ocasiões, o neuroblastoma apresenta áreas centrais de necrose, exibidas como regiões de baixo coeficiente de atenuação, que são mais aparentes após melhoramento de contraste [6, 4, 2].

Apesar da utilidade demonstrada das técnicas de imageamento na detecção, delineamento, e estadiamento do tumor primário, urge melhorar o uso destas técnicas para a avaliação mais precisa da doença local que possa levar a um melhor planeja-

mento do tratamento e acompanhamento. Foglia et al. [7] argumentam que o estado do tumor primário não pode ser estabelecido definitivamente pelo diagnóstico por imagem, devido a erros de sensibilidade e especificidade da ordem de 38%, quando se procura estabelecer a viabilidade do tumor por métodos de imageamento e compara-se o resultado desta análise com os achados em cirurgia retardada. Relatam também que exames de CT não podem discernir tumor viável de tecido fibroso ou tumor não viável destruído em quimioterapia prévia.

Análise de imagens e outras técnicas baseadas em computador podem melhorar a análise radiológica oferecendo medidas mais sofisticadas, precisas e reproduzíveis da informação presente nas imagens [8, 9, 10]. Entretanto, poucos pesquisadores investigaram o potencial do diagnóstico do neuroblastoma auxiliado por computador usando imageamento médico; trabalhos correlatos publicados estão limitados à medida do volume tumoral usando fatias de CT segmentadas manualmente (planimetria) [11, 12]. Em particular, não existem métodos, com ou sem o auxílio de computadores, que possa assegurar com precisão, usando exames baseados em imagem, a quantidade de doença ativa, tecido cicatricial, fibroso, tumor benigno e outros tecidos em uma dada massa tumoral [7].

1.1 Objetivos

O objetivo do presente trabalho é o aperfeiçoamento da análise da massa tumoral primária, em pacientes com neuroblastoma avançado, usando exames de tomografia computadorizada por raio-X. Mais especificamente, objetivamos estabelecer uma metodologia quantitativa de análise de exames de CT de neuroblastomas que compreenda as seguintes etapas: segmentação, e estimação do conteúdo histológico.

1.2 Motivação

Este trabalho é motivado por sua relevância científica e social, dado que insere-se no contexto da cooperação estabelecida entre o Instituto da Criança da Faculdade de Medicina da USP (ICr-FMUSP) e o Laboratório de Sistemas Integráveis da Escola Politécnica da USP (LSI-EPUSP).

Uma metodologia de análise quantitativa dos neuroblastomas, a partir de exames de CT, traz a possibilidade de reprodução e comparação das medidas efetuadas, que pode levar a novas abordagens científicas de diagnóstico. Note-se também a maior riqueza de informação obtida em tal análise. Tais melhoramentos reduziriam a incerteza no diagnóstico por imagem da massa tumoral, após quimioterapia inicial, e eventualmente permitiriam que se evitassem algumas cirurgias de revisão. Estes argumentos evidenciam o caráter social benéfico do trabalho desenvolvido. Em particular, estima-se que aproximadamente 10% dos casos de neuroblastomas do país são tratados no ICr-FMUSP e cerca de 2 pacientes por mês são operados¹.

Cientificamente, observamos que o problema é desafiador, poucos pesquisadores investigaram o potencial do diagnóstico do neuroblastoma auxiliado por computador usando imageamento médico, no melhor de nosso conhecimento.

1.3 Metodologia

Para atingir nossa meta, propomos uma metodologia para a estimação do conteúdo histológico da massa que engloba uma técnica para a segmentação semi-automática da massa tumoral primária em imagens de CT do neuroblastoma e um método estatístico para estimar, a partir das imagens de CT segmentadas, a composição histológica do tumor primário. Os resultados do método serão posteriormente comparados com os resultados da análise histológica das massas resectadas cirurgicamente.

¹comunicação pessoal

1.4 Estrutura do texto

A dissertação é composta dos seguintes capítulos:

- Capítulo 1: Introdução.
- Capítulo 2: Análise de Neuroblastomas usando Imagens de Tomografia Computadorizada.

Expomos os conceitos referentes ao neuroblastoma e sua análise por imagens de CT, mediante revisão bibliográfica em radiologia e oncologia pediátrica. São apresentadas também considerações acerca da análise do neuroblastoma auxiliada por computador.

- Capítulo 3: Segmentação e Estimação da Composição dos Tecidos a partir de Imagens de CT de Neuroblastomas.

Justificamos e descrevemos neste capítulo a solução implementada para a segmentação e estimação da composição histológica da massa tumoral.

- Capítulo 4: Implementação e Análise dos Resultados.

Apresentamos e discutimos a sequência de testes efetuados e os resultados obtidos.

- Capítulo 5: Conclusão e Trabalhos Futuros.

Neste item apresentamos uma análise final da pesquisa, e propostas de extensão deste trabalho.

Capítulo 2

Análise de Neuroblastomas usando Imagens de Tomografia Computadorizada

2.1 Neuroblastoma

Neuroblastoma é um tumor maligno que se origina de células indiferenciadas da crista neural, e que pode surgir em qualquer ponto ao longo dos gânglios simpáticos ou na medula adrenal [4]. É considerado uma entidade biológica fascinante e enigmática que representa, ao mesmo tempo, um dos melhores exemplos de regressão espontânea para um tumor maligno e um dos tumores com o menor potencial de cura para alguns grupos [4, 5].

Existem três tipos de lesões provenientes de células dos gânglios simpáticos que formam um espectro de doenças neoplásicas. Neuroblastoma é a forma mais imatura e maligna das três e surge usualmente antes dos cinco anos de idade. Ganglioneuroblastoma é uma forma mais madura que retém algumas características de malignidade, com pico de incidência entre cinco e dez anos de idade. Ganglioneuroma é uma forma bem diferenciada e benigna, que tipicamente apresenta-se após os dez anos de idade [13].

2.1.1 Incidência

Neuroblastoma é o tumor extracranial sólido maligno mais comum em crianças [2]; é a terceira malignidade mais comum da infância. Responde por oito a 10% de todos os cânceres infantis [14] e 15% de todas as mortes relacionadas ao câncer no grupo de pacientes em idade pediátrica. A idade mediana de diagnóstico é dois anos e 90% dos casos diagnosticados estão em crianças com menos de cinco anos [2].

Sessenta e cinco por cento dos neuroblastomas estão localizados no abdômen; aproximadamente dois terços destes surgem na glândula adrenal. Quinze por cento dos neuroblastomas são torácicos, usualmente localizados nos gânglios simpáticos do mediastino posterior. Dez a doze por cento dos neuroblastomas são disseminados sem que se possa localizar seu ponto de origem [6].

2.1.2 Estadiamento

O mais recente sistema de estadiamento para neuroblastoma é o *International Neuroblastoma Staging System* (INSS), que leva em consideração achados radiológicos, resectabilidade cirúrgica, envolvimento dos nódulos linfáticos e envolvimento da medula óssea [1]. A Tabela 2.1 apresenta a definição de cada estágio deste sistema.

Os principais fatores determinantes para o prognóstico são a idade do paciente (ver Tabela 2.2) e o estágio da doença (ver Tabela 2.3) [2, 6]. A Tabela 2.4 mostra a distribuição dos pacientes com diagnóstico de neuroblastoma em relação ao estágio da doença ao diagnóstico [2].

O local do tumor primário também é dito relevante no prognóstico geral. Tumores que surgem no abdômen e pélvis tem o pior prognóstico, e tumores adrenais tem a pior taxa de mortalidade. Neuroblastomas torácicos apresentam, em geral, uma melhor taxa de sobrevivência (61%) que tumores abdominais (20%) [6].

Ressecção cirúrgica do tumor primário de um paciente com neuroblastoma avançado é recomendável quando possível. Contudo, o tumor primário de um paciente com neuroblastoma avançado (Estágios III e IV) pode não ser resectável, se há risco de

Estágio	Descrição
I	Tumor localizado com excisão completa com ou sem doença residual; análise de nódulos linfáticos ipsilaterais representativos negativa para o tumor microscopicamente (nódulos atrelados e removidos com o tumor primário podem apresentar resultado positivo).
II A	Tumor localizado com excisão incompleta; nódulos linfáticos ipsilaterais não-aderentes representativos negativos para tumor microscopicamente.
II B	Tumor localizado com ou sem excisão completa com nódulos linfáticos ipsilaterais não-aderentes positivos para tumor; análise dos nódulos linfáticos contralaterais aumentados deve ser negativa microscopicamente.
III	Tumor unilateral não resectável que infiltra além da linha média com ou sem envolvimento dos nódulos linfáticos regionais; tumor unilateral localizado com envolvimento dos nódulos linfáticos regionais contralaterais, ou tumor em linha média com extensão bilateral por meio de infiltração (não resectável) ou envolvimento dos nódulos linfáticos.
IV	Qualquer tumor primário com disseminação a nódulos linfáticos distantes, estrutura óssea, medula óssea, fígado, pele, ou outros órgãos (exceto como definido para o Estágio IV S).
IV S	Tumor primário localizado (como definido para o Estágio I, IIA, ou IIB) com disseminação limitada a pele, fígado, ou medula óssea (mínima), limitado a crianças com menos de um ano de idade.

Tabela 2.1: International Neuroblastoma Staging System. Reproduzido de Cohen et al. [1].

Idade	Percentual
< 1 ano	74
1-2 anos	43
2-3 anos	18
> 3 anos	14
Geral	36

Tabela 2.2: Sobrevivência ao neuroblastoma por idade. Reproduzido de Abramson [2].

Estágio	Sobrevivência(%)
I	95 – 100
II	82 – 83
III	42 – 81
IV	10 – 30
IV S	60 – 90

Tabela 2.3: Sobrevivência ao neuroblastoma por estágio. Reproduzido de Abramson [2].

Estágio	Pacientes (%)
I	9 – 13
II	11 – 20
III	8 – 16
IV	43 – 60
IV S	9 – 10

Tabela 2.4: Distribuição dos pacientes com neuroblastoma segundo o estágio da doença. Reproduzido de Abramson [2].

danificar estruturas vitais durante o procedimento (notavelmente quando a massa circunda a artéria aorta). O tratamento nestes casos requer quimioterapia ou radioterapia para uma diminuição inicial da massa, após o qual a cirurgia (retardada) de ressecção pode ser realizada [4].

2.1.3 Análise radiológica

A radiologia do neuroblastoma tem sido estudada expressivamente nas últimas três décadas, e diversos artigos de revisão acerca dos aspectos radiológicos da doença foram publicados [15, 16, 6, 2, 1]. Exames radiológicos podem ser úteis no diagnóstico inicial, avaliação da extensão da doença, estadiamento, avaliação pré-cirúrgica, tratamento, e acompanhamento da doença.

Diversas modalidades de imagem foram investigadas no contexto do neuroblastoma, incluindo ultra-som, CT, imageamento por ressonância magnética (*Magnetic Resonance Imaging - MRI*), cintilografia óssea e de medula, urografia excretora e radiografia convencional. CT e MRI são consideradas as melhores modalidades para a avaliação do estágio do tumor [2], resectabilidade [17], prognóstico, e seguimento [2]. Exames de CT e MRI são mandatários para a análise do tumor primário, e alguns autores relatam que MRI pode ser mais útil que CT na avaliação de doença metastática [18, 19, 20].

Comparando CT e MRI, a última possui maior precisão no diagnóstico (a maior dentre todas as modalidades de imageamento) e é tida como a mais adequada para demonstrar o espalhamento do tumor e visualização deste em relação aos vasos

sangüíneos vizinhos e vasos dentro do próprio tumor [2], que são fatores importantes na terapia e no estabelecimento da resectabilidade. MRI também provê melhor contraste de imagem para tecidos moles do que CT, e é mais promissor para a estimacão da composicão histológica [21, 22]. Outra característica vantajosa é o fato de que MRI usa radiação não-ionizante.

Entretanto, CT é considerado a modalidade de escolha em diversas investigacões, porque é comparável em utilidade com MRI na avaliacaão de doenca local [2], e possui um custo menor. Sabe-se que CT é efetivo na detecção de calcificacões (um achado a favor de neuroblastoma [2] em comparacão ao tumor de Wilms), detecção e avaliacaão de extensão da doenca local, avaliacaão do envolvimento dos nódulos linfáticos, doenca recorrente, e metástase não esquelética [6].

Em exames de CT, o neuroblastoma abdominal é visto como uma massa de tecido mole, usualmente suprarenal ou paravertebral, de forma irregular, lobulada, estendendo-se do flanco em direcção à linha média, e sem estrutura capsular. A massa tende a ser heterogênea em funçao da existênciade necrose tumoral misturada a tumor viável, e contém calcificacões em 85% dos pacientes. Calcificacões são usualmente densas, amorfas, e heterogêneas em aparência. Em algumas ocasiões, o neuroblastoma apresenta áreas de necrose centrais, exibidas como regiões de baixo coeficiente de atenuaçao, que são mais aparentes após melhoramento de contraste [6, 4, 2]. A Figura 2.1 mostra uma imagem de CT de um paciente com neuroblastoma em Estágio III, antes do tratamento, com o contorno da massa desenhado por um radiologista. Note que a massa envolve a aorta, o que a torna não resectável.

Apesar da utilidade demonstrada das técnicas de imageamento na detecção, delineamento, e estadiamento do tumor primário, urge melhorar o uso destas técnicas para a avaliacaão mais precisa da doenca local que possa levar a um melhor planejamento do tratamento e acompanhamento. Foglia et al. [7] argumentam que o estado do tumor primário não pode ser estabelecido definitivamente pelo diagnóstico por imagem, devido a erros de sensibilidade e especificidade da ordem de 38%, quan-



Figura 2.1: Imagem de CT de um paciente com neuroblastoma Estágio III, com o contorno da massa desenhado por um radiologista (MV).

do se procura estabelecer a viabilidade do tumor por métodos de imageamento e compara-se o resultado desta análise com os achados em cirurgia retardada. Relatam também que exames de CT não podem discernir tumor viável de tecido fibroso ou tumor não viável destruído em quimioterapia prévia.

2.1.4 Diagnóstico auxiliado por computador

Análise de imagens e outras técnicas baseadas em computador podem melhorar a análise radiológica oferecendo medidas mais sofisticadas, precisas e reproduzíveis da informação presente nas imagens [8, 9, 10]. Entretanto, poucos pesquisadores investigaram o potencial do diagnóstico do neuroblastoma auxiliado por computador usando imageamento médico; trabalhos correlatos publicados estão limitados à medida do volume tumoral usando fatias de CT segmentadas manualmente (planimetria) [11, 12]. Em particular, não existem métodos, com ou sem o auxílio de computadores, que possa assegurar com precisão, usando exames baseados em imagem, a quantidade de doença ativa, tecido cicatricial, fibroso, tumor benigno e outros

tecidos em uma dada massa tumoral [7].

O problema que investigamos no presente trabalho é o aperfeiçoamento da análise da massa tumoral primária, em pacientes com neuroblastoma avançado, usando exames de tomografia computadorizada por raio-X. Para atingir esta meta, propomos uma metodologia para a estimação do conteúdo histológico da massa que engloba uma técnica para a segmentação semi-automática da massa tumoral primária em imagens de CT do neuroblastoma e um método estatístico para estimar, a partir das imagens de CT segmentadas, a composição histológica do tumor primário. Os resultados do método serão posteriormente comparados com os resultados da análise histológica das massas resectadas cirurgicamente.

Os seguintes itens resumem os pontos mais relevantes discutidos acima e que enfatizam a importância do trabalho presente:

- A idade do paciente e o estágio de sua doença são os fatores de prognóstico mais relevantes. Pacientes com neuroblastoma em estágio III ou IV tem a taxa de mortalidade mais alta. A localização do tumor também é um fator de prognóstico: neuroblastoma abdominal tem o pior prognóstico.
- A maioria dos pacientes entram em tratamento nos estágios III e IV. Nestes casos, o tumor primário não pode ser removido cirurgicamente em um primeiro momento: a massa deve ser diminuída por quimioterapia ou radioterapia e então resectada por cirurgia retardada. Portanto, o acompanhamento por métodos de imagem é crítico.
- Embora métodos de imagem possam facilitar a detecção e estadiamento do tumor, observamos que uma melhor quantificação e análises mais precisas da massa tumoral não podem ser obtidas através de metodologias correntes. Em particular, não existe método que possa avaliar com segurança se a massa tumoral primária contém tumor viável, tecido fibroso, ou tumor não viável. É desejável estimar a fração de cada tipo de tecido na massa tumoral, mas não

existem recursos disponíveis para derivar esta informação a partir de imagens radiográficas.

- Técnicas de processamento de imagens médicas não foram empregadas até o presente momento para melhor caracterização da doença, exceto para planimetria.

2.2 Caracterização de Tecidos usando CT

A descoberta dos raios-X por Röntgen em 1895 e o desenvolvimento de CT por raios-X na década de 1970 foram grandes avanços no diagnóstico por imagem (ambos foram laureados com o prêmio Nobel, em Física e Medicina, respectivamente) [23]. O aparelho de tomografia computadorizada (tomógrafo) é construído com base na técnica de reconstruir uma seção tomográfica do corpo do paciente dado um certo número de projeções [23, 24, 25]. Cada projeção é composta por integrais de linha da distribuição espacial da densidade no corpo, estimadas a partir de medições com raio-X. Para um certo feixe de raios-X seguindo um caminho \mathcal{P} com uma dada posição e orientação, o relacionamento entre a intensidade I_0 do raio na fonte e a intensidade detectada I , após a passagem do raio através do corpo sendo imageado, é dado por

$$I = I_0 \exp \left[- \int_{\mathcal{P}} \rho(x) dx \right] ,$$

onde x é a variável de distância ao longo do caminho \mathcal{P} do raio.

A quantidade $\rho(x)$ sendo integrada é chamada *coeficiente de atenuação linear* do material (tecido) no ponto x , é a quantidade física que é exibida na imagem de CT reconstruída. O coeficiente de atenuação linear varia com duas propriedades do material: densidade e composição elementar [26]. O valor de ρ foi medido e tabulado para diversos materiais [27, 28], incluindo tecidos humanos e animais, com raios-X de energias diferentes (incluindo medidas com energias múltiplas, como *dual-energy*

Tecido	Valor de Atenuação (HU)	
	Média	Desvio Padrão
Bolo alimentar	-166	24
Gordura	-90	18
Bile	+16	8
Rim	+32	10
Pâncreas	+40	14
Sangue(Aorta)	+42	18
Músculo	+44	14
Baço	+46	12
Fígado	+60	14
Medula	+142	48

Tabela 2.5: Valores de CT em unidades Hounsfield (HU) para alguns tipos de tecidos abdominais. Reproduzido de Mategrano et al. [3].

imaging [26, 29, 30]).

Não é comum exibir imagens de CT em termos do coeficiente de atenuação linear (que é dependente da energia utilizada [27, 30]). Ao invés disso são usadas ‘unidades CT’ normalizadas, que são mais convenientes e independentes, até certo ponto, da energia do raio-X. A unidade CT é definida como:

$$\text{número CT} = k \frac{\rho - \rho_w}{\rho_w},$$

onde ρ é o coeficiente de atenuação linear do tecido e ρ_w é o da água. Em tomógrafos antigos, o parâmetro k valia 500. Hoje é comum usar $k = 1000$ para que se obtenha o número CT em unidades Hounsfield (*Hounsfield Units* - HU) [26]. A Tabela 2.5 mostra o valor de média e de desvio padrão dos valores CT em HU para diversos tipos de tecido no abdômen.

Logo após a invenção do tomógrafo, diversos pesquisadores focalizaram sua atenção para o problema de estimar a composição de tecidos do corpo humano a partir de imagens de CT. Alter [31] avaliou a utilidade clínica do tomógrafo e relatou a aparência de diversos órgãos, tecidos, e doenças em imagens de CT, e usou o valor destes primeiros, em HU, para a caracterização de tecidos. Phelps et al. [27],

e Rao e Gregg [28] descrevem o coeficiente de atenuação linear de diversos tecidos, e os relaciona à unidade CT ideal tal como relatada por Wilson [29] e Brooks [30]. Mategrano et al. [3] apresentam medidas do coeficiente de atenuação para tecidos na região abdominal; veja tabela 2.5. Uma discussão das fontes de erro potenciais na medida do coeficiente de atenuação é encontrada no trabalho de Williams et al. [32]. Duerinckx e Macovski [33] discutem a natureza do ruído em imagens de CT. Pullan et al. [34] estudaram a caracterização de regiões em uma imagem de CT usando momentos estatísticos centrais (tais como média, desvio padrão e skewness) obtidos a partir de histogramas do valor dos elementos volumétricos (voxels) no exame de CT e do gradiente espacial dos voxels (tomado como uma medida simplificada de textura), dentro de regiões delimitadas, com aplicação a tumores cerebrais e lesões do baço e fígado. Kramer et al. [35] relatam um trabalho semelhante ao de Pullan et al. Latchaw et al. [36] opinam que CT não é específico na separação de tumores sólidos e lesões císticas no cérebro.

Os seguintes pontos resumem a discussão prévia sobre caracterização de tecidos por CT:

- Alguns autores relataram um sucesso relativo no uso de unidades CT para diferenciar entre tecidos que são visualmente similares em imagens de CT [35, 34, 3, 32], enquanto outros relatam fracasso na mesma atividade [36, 7].
- Enquanto a medida do coeficiente de atenuação linear dos tecidos *in vitro* pode ser realizada com boa precisão [28, 27], diversas fontes de erro podem degradar o desempenho da sensitometria por CT [25].
- Modelos e técnicas estatísticas mais sofisticadas não foram empregados nos trabalhos citados acima.

Capítulo 3

Segmentação e Estimação da Composição dos Tecidos a partir de Imagens de CT de Neuroblastomas

O problema que nos propomos a investigar é a análise quantitativa da massa tumoral primária, em pacientes com neuroblastoma avançado, usando exames de CT. Para atingir esta meta, propomos uma metodologia para a estimação do conteúdo histológico da massa. Baseados em uma revisão dos capítulos anteriores, nossas premissas básicas no presente trabalho são:

- Estimação da composição histológica a partir de imagens de CT é uma tarefa importante que pode levar a melhoramentos significativos no tratamento de pacientes com neuroblastoma avançado, especialmente aqueles com a doença em estágios III e IV. Tais pacientes representam a maioria dos pacientes com neuroblastoma, e tem o pior prognóstico.
- Há certa evidência de que tecidos na cavidade abdominal podem ser distinguidos através de seus valores de CT. Embora tal evidência não seja muito

forte, atualmente estão disponíveis ferramentas estatísticas mais poderosas que aquelas empregadas nos trabalhos revistos acima.

- Ocorreram diversos avanços encorajadores no campo do processamento e análise de imagens médicas durante a década passada. Contudo, não conhecemos a aplicação de tais técnicas à análise do neuroblastoma.

Para que alcancemos as metas estabelecidas, nossa metodologia compreende uma técnica para segmentação semi-automática da massa tumoral, um modelo estatístico parametrizado para a composição histológica do tumor, e um método para estimar os parâmetros do modelo. O algoritmo de segmentação implementado é baseado no conceito de conectividade fuzzy tal como definido por Udupa e Samarasekera [37]. O modelo estatístico empregado é o modelo de mistura de Gaussianas [38], e o algoritmo para a estimação dos parâmetros do modelo é o algoritmo Expectation-maximization [39].

A metodologia completa é mostrada na Figura 3.1. O caminho superior (chamado “o caminho da imagem”) mostra a seqüência de operações baseadas em computador que levam à estimação do modelo estatístico, que será descrito nas seções seguintes. O caminho inferior (“o caminho da faca e da agulha”) consiste em obter a informação histológica a respeito do tumor a partir da biópsia e cirurgia retardada [7]. Correlação entre as informações obtidas das duas abordagens permite verificar se o método é válido.

3.1 Segmentação

Segmentação de imagens é o processo de separar uma dada imagem em regiões, tal que algumas dessas regiões possam ser identificadas como objetos significativos para uma certa tarefa. Segmentação é o primeiro passo em diversas metodologias para análise e interpretação de imagens.

Embora diversos trabalhos tenham sido relatados no campo da segmentação de

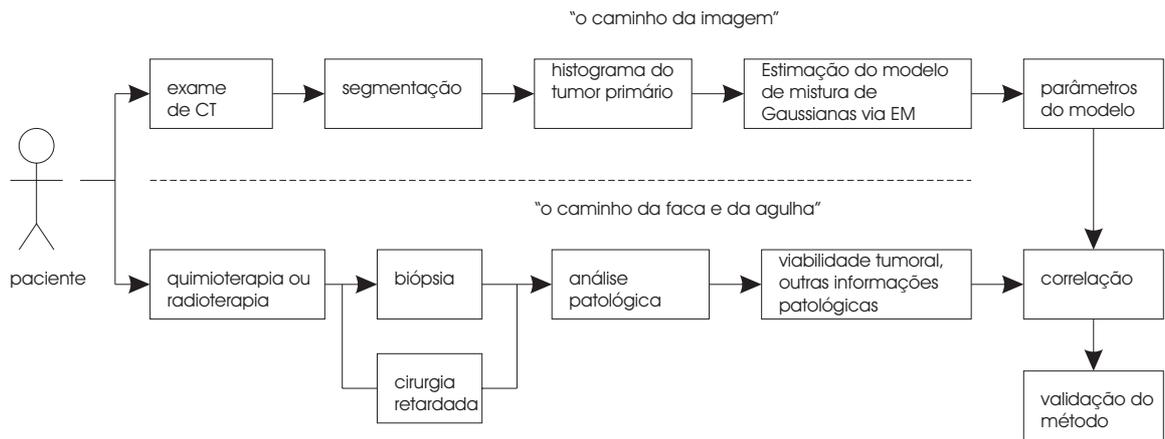


Figura 3.1: Representação esquemática do método proposto para a análise de neuroblastomas. EM: Expectation-maximization.

imagens médicas, estes parecem estar concentrados em problemas particulares tais como segmentação de estruturas cerebrais, extração dos principais vasos sanguíneos, ou análise do movimento cardíaco [8, 9]. Ressonância magnética parece ser a modalidade de escolha em muitos destes trabalhos devido a sua habilidade intrínseca em separar tecidos, e à possibilidade de se projetar diferentes seqüências de pulso que possibilitam imagear diferentes quantidades físicas do material em estudo. Os seguintes aspectos tornam difícil o problema de segmentação de massas tumorais em imagens de CT do abdômen pediátrico:

- **Baixo contraste:** Como mostrado na Tabela 2.5, a separação entre os valores de CT médios dos tecidos no abdômen são da mesma magnitude de seus desvios-padrões; este fato torna os tecidos indistinguíveis em alguns casos.
- **Morfologia variável:** A forma das estruturas encontradas no abdômen é bem menos regular que aquelas encontradas em outras partes do corpo, tais como o cérebro, o coração, ossos, ou vasos. Em particular, descrever a estrutura do neuroblastoma avançado é ainda mais difícil, pois este cresce entre os órgãos da cavidade abdominal, e não dentro de um órgão em particular; o tumor empurra as estruturas adjacentes ou amolda-se ao espaço disponível.
- **Artefatos de movimento:** embora o movimento do paciente não seja um pro-

blema em imageamento do cérebro, trata-se de um problema crucial em relação ao abdômen. Mesmo o efeito de artefatos de movimento em imageamento cardíaco pode ser controlado devido à periodicidade deste movimento - sincroniza-se o *scanner* com o eletrocardiograma. Contudo, movimentos respiratórios e peristálticos são intensos e irregulares, dificultando o imageamento apropriado de estruturas na cavidade abdominal. O imageamento de tumores pediátricos abdominais apresenta um desafio extra em manter a criança imóvel; isto é conseguido em alguns casos sedando o paciente.

O problema da segmentação automática ou semi-automática do neuroblastoma usando técnicas de processamento de imagens médicas não foi estudado, até onde foi de nosso conhecimento. A inspiração para o projeto das técnicas no trabalho presente vem de outros problemas de segmentação na área de processamento de imagens médicas. A técnica que propomos para a segmentação está baseada no conceito de *conectividade fuzzy* tal como desenvolvido no trabalho de Udupa e Samarasekera [37]. O algoritmo de conectividade fuzzy é um procedimento de crescimento de região cuja saída, para cada elemento de volume (voxel) ou elemento de imagem (pixel), é um valor em $[0, 1]$ que representa o grau fuzzy de conectividade entre o voxel considerado e o voxel semente (ou região semente) que é usado para iniciar o procedimento.

Seja C o conjunto de características para cada voxel, que compreende sua localização espacial e algumas propriedades locais (tais como textura local, valor de CT, gradiente). Vamos definir uma função $\eta: C \times C \mapsto [0, 1]$ (uma relação fuzzy) que representa a afinidade fuzzy entre dois voxels. A função η tem as propriedades: $\eta(a, a) = 1$ (a relação é reflexiva) e $\eta(a, b) = \eta(b, a)$ (a relação é simétrica). Qualitativamente, a afinidade entre dois voxels deve ser maior se estão próximos no espaço e se suas propriedades são similares, ou seja, se estão próximos no espaço de características.

Um caminho de tamanho n entre dois voxels a e b é uma seqüência $\langle c_1 =$

$a, c_2, \dots, c_n = b$) de voxels (c_i). A força do caminho é $\min[\eta(c_i, c_{i+1})]$. A conectividade entre a e b é a força do caminho mais forte entre eles, para todos os caminhos possíveis. Um objeto fuzzy pode então ser definido, para um dado voxel semente e uma certa função η , avaliando-se para cada voxel a conectividade fuzzy entre este voxel e o voxel semente. Udupa e Samarasekera [37] mostram como efetuar este cálculo usando programação dinâmica. Finalmente temos a segmentação fuzzy, considerando este valor de conectividade fuzzy entre cada voxel x e o voxel semente como o valor fuzzy de pertinência do voxel x à região do tumor.

O método descrito acima foi implementado em nosso trabalho, com η definida como:

$$\eta(a, b) = \begin{cases} 0 & \text{se } a \text{ e } b \text{ não são vizinhos,} \\ \exp \left[-\frac{1}{2} \left(\frac{f(a)+f(b)-\mu}{\sigma} \right)^2 \right] & \text{se } a \text{ e } b \text{ são vizinhos,} \end{cases}$$

onde $f(a)$ é o valor de CT do voxel a , e μ e σ são a média e o desvio padrão da região. Esta definição segue o trabalho de Udupa e Samarasekera [37]. A vizinhança considerada em nossos testes é a vizinhança 6, onde dois voxels no reticulado Z^3 são vizinhos se suas coordenadas diferem de apenas uma unidade, em apenas uma dimensão.

3.1.1 Estimação do Histograma

Tradicionalmente, o subproduto do processo de segmentação de imagens é uma estrutura de dados que informa, para cada voxel, se este pertence ou não ao objeto de estudo. Assim ocorre com a segmentação manual da massa tumoral: para cada fatia do exame de CT, o tumor é manualmente delineado, e os voxels que se encontrem dentro do contorno são declarados como pertencentes à região do tumor.

Para que possamos obter o histograma dos voxels da massa tumoral dentro deste paradigma binário de identificação de objetos, usamos o algoritmo clássico de obtenção do histograma. Inicializa-se o histograma armazenando zero em cada uma de

suas posições. Então, para cada voxel pertencente à região da massa tumoral, toma-se o valor do voxel e incrementa-se de um valor unitário a posição do histograma que corresponde a este valor.

O processo de segmentação semi-automática do exame de CT, conforme descrito acima, produz um resultado de natureza diferente daquele obtido em processos tradicionais de segmentação de imagens. No processo de segmentação *fuzzy* que descrevemos anteriormente, a cada voxel atribui-se um valor *fuzzy* que indica o grau de pertinência deste voxel à categoria dos voxels da massa tumoral. Não obstante, podemos obter o histograma da massa tumoral a partir desta segmentação difusa, sem a necessidade de tornar binária nossa segmentação *fuzzy* do exame de CT.

Segue o procedimento para a obtenção do histograma da massa tumoral a partir da segmentação *fuzzy*. Inicializa-se o histograma armazenando o valor zero em cada uma de suas posições. Para cada voxel do exame de CT, toma-se o valor deste voxel e incrementa-se, *do valor de pertinência do voxel em questão*, a posição do histograma que corresponde a seu valor. Por exemplo, se um voxel de valor 20 HU pertence ao tumor com valor *fuzzy* de pertinência 0,7, devemos acrescentar o valor 0,7 à posição do histograma correspondente ao valor de 20 HU.

Por fim, devemos normalizar o histograma, seja este obtido através da segmentação manual binária ou do processo semi-automático *fuzzy*: divide-se o valor armazenado em cada posição do histograma pela soma de todos os valores do histograma. Segue imediatamente que, após o processo de normalização, a soma dos valores do histograma vale um.

3.2 Estimação da Composição Histológica

Estando a massa tumoral segmentada e o histograma normalizado obtido, procedemos à estimação de sua composição histológica. Como discutido na Seção 2.1, a massa tumoral é heterogênea, pois necrose e tecido tumoral viável estão misturados, e em certos casos observam-se regiões centrais de necrose, mostradas como áreas de

baixa atenuação dentro da massa. Portanto, devemos desenvolver uma descrição global da massa que leve à estimação das frações de volume que correspondem a cada tipo de tecido, ao invés de procurar separar a massa em regiões distintas.

Assumimos que o valor de CT para um voxel que provém de um dado tecido interno à massa (massa benigna, necrose, tumor maligno, tecido fibroso, etc) é uma variável aleatória Gaussiana. Assim sendo, a massa como um todo é modelada estatisticamente como uma mistura de variáveis Gaussianas, um modelo conhecido como modelo de mistura de Gaussianas [40].

Seja x o valor de atenuação de CT para um dado voxel, e $\theta_i = (\mu_i, \sigma_i)$ o conjunto de parâmetros (média μ_i e desvio padrão σ_i) que descrevem a função densidade de probabilidade Gaussiana dos valores de CT do i -ésimo tipo de tecido. Denotamos a função densidade de probabilidade para x , dado que x provém do i -ésimo tipo de tecido, por $p_i(x|\theta_i)$, que é representada como

$$p_i(x|\theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[-\frac{(x - \mu_i)^2}{2\sigma_i^2} \right]. \quad (3.1)$$

Seja M o número verdadeiro de tipos diferentes de tecido em uma dada massa tumoral, que assumimos ser conhecido para o momento. (A estimação do valor de M será discutida posteriormente.) Seja α_i a probabilidade de que um dado voxel provém do i -ésimo tipo de tecido na massa tumoral. Por definição, $\sum_{i=1}^M \alpha_i = 1$. O valor de α_i pode ser visto como a fração do volume tumoral que é composta do i -ésimo tipo de tecido. Então, a função densidade de probabilidade para a massa toda é uma mistura de Gaussianas, especificada pelo vetor de parâmetros $\Theta = (\alpha_1, \dots, \alpha_M, \theta_1, \dots, \theta_M)$, e descrita por $p(x|\Theta) = \sum_{i=1}^M \alpha_i p_i(x|\theta_i)$.

Seja N o número de voxels na massa tumoral e seja $\mathbf{x} = (x_1, x_2, \dots, x_N)$ um vetor formado por todos os voxels da massa, o qual chamaremos de dado observado. Sob a condição do dado observado, a *probabilidade posterior* dos parâmetros é obtida pela regra de Bayes [38] como

$$p(\Theta|\mathbf{x}) = \frac{p(\Theta) p(\mathbf{x}|\Theta)}{p(\mathbf{x})}. \quad (3.2)$$

Aqui, $p(\mathbf{x})$ é a probabilidade do dado, independente dos parâmetros; considerando que o problema de estimação é condicionado ao dado observado, $p(\mathbf{x})$ é um termo constante. O termo $p(\Theta)$ é a probabilidade *a priori* do vetor de parâmetros Θ . O termo $p(\mathbf{x}|\Theta)$ é chamado de *verossimilhança* (*likelihood*) de Θ , denotado por $L(\Theta|\mathbf{x})$, e dado por

$$L(\Theta|\mathbf{x}) \triangleq p(\mathbf{x}|\Theta) = \prod_{i=1}^N p(x_i|\Theta). \quad (3.3)$$

Em uma abordagem Bayesiana, se conhecemos $p(\Theta)$, é desejável que tomemos como a melhor estimativa dos parâmetros do modelo de mistura de Gaussianas o valor de Θ que maximiza $p(\Theta|\mathbf{x})$: este é o princípio do *máximo a posteriori*.

No caso em que não temos nenhum conhecimento prévio sobre o valor de Θ , ou seja, se não conhecemos nada sobre sua probabilidade *a priori*, temos a situação conhecida como *flat prior* [40]. Neste caso, a Equação 3.2 torna-se $p(\Theta|\mathbf{x}) = c p(\mathbf{x}|\Theta)$, onde c é uma constante de normalização. Portanto, encontrar o valor mais provável de Θ fornecido o dado observado, sem nenhum conhecimento prévio sobre a função densidade de probabilidade dos parâmetros, é o mesmo que encontrar o valor de Θ que maximiza a verossimilhança, ou o logaritmo desta: este é o *princípio da máxima verossimilhança* (*maximum likelihood* - ML) [40, 38]. A adoção deste princípio leva a cálculos simplificados com resultados razoáveis [41]. Abordagens completamente Bayesianas para classificação e estimação de parâmetros podem prover um desempenho melhor, ao custo de maiores requerimentos de poder computacional e complexidade de implementação [42, 43].

Para maximizar a verossimilhança, usamos o algoritmo *Expectation-maximization* (EM) [40, 44, 41, 39, 38]. O algoritmo EM é um procedimento iterativo que parte de uma estimativa inicial dos parâmetros e melhora iterativamente a estimativa em direção ao máximo local da verossimilhança. O algoritmo EM genérico é composto

de dois passos: *expectativa* (expectation) e *maximização* (maximization). No passo de expectativa, calcula-se o modelo probabilístico parametrizado dado a estimativa corrente do vetor de parâmetros. No passo de maximização, encontra-se o vetor de parâmetros que maximiza o recém computado modelo paramétrico, este vetor é então considerado como a nova melhor estimativa dos parâmetros. O procedimento iterativo continua até que algum critério de parada seja satisfeito, por exemplo, até que a diferença $\log[L(\Theta_{i+1}|\mathbf{x})] - \log[L(\Theta_i|\mathbf{x})]$ ou o módulo $|\Theta_{i+1} - \Theta_i|$ do vetor diferença entre sucessivas iterações seja menor que um valor pré-definido.

Para cada tipo de tecido j definimos $p(j|x_i, \Theta)$ como a probabilidade de que o i -ésimo voxel, cujo valor é x_i , pertence ao j -ésimo tipo de tecido. Isto pode ser calculado usando a regra de Bayes.

$$p(j|x_i, \Theta) = \frac{p(j|\Theta)p(x_i|j, \Theta)}{p(x_i|\Theta)} = \frac{\alpha_j p_i(x_i|\theta_i)}{p(x_i|\Theta)}.$$

A derivação do algoritmo EM para o modelo de mistura de Gaussianas leva a um conjunto de equações iterativas que realizam os passos de expectativa e de maximização simultaneamente. Para o j -ésimo tipo de tecido, as equações de atualização da melhor estimativa são:

$$\begin{aligned}\alpha_j^{\text{nov}} &= \frac{1}{N} \sum_{i=1}^N p(j|x_i, \Theta^{\text{velho}}), \\ \mu_j^{\text{nov}} &= \frac{\sum_{i=1}^N x_i p(j|x_i, \Theta^{\text{velho}})}{\sum_{i=1}^N p(j|x_i, \Theta^{\text{velho}})}, \\ \sigma_j^{\text{nov}} &= \frac{\sum_{i=1}^N (x_i - \mu_j^{\text{nov}})^2 p(j|x_i, \Theta^{\text{velho}})}{\sum_{i=1}^N p(j|x_i, \Theta^{\text{velho}})}.\end{aligned}$$

Para estimar o valor de M , ou seja, o número de tipos de tecidos na massa tumoral, não se pode modelar M como uma variável aleatória e aplicar diretamente o princípio da máxima verossimilhança. Isto ocorre porque a máxima verossimilhança de Θ é uma função não-decrescente de M [39]. O valor estimado de M deve ser o valor que minimize uma função de custo que valorize o aumento da verossimilhança,

mas penalize valores altos de M . (Em outras palavras, deve-se construir uma probabilidade *a priori* para a variável aleatória M .) A escolha padrão para tal função de custo é a que segue o critério de Menor Comprimento de Descrição (*Minimum Description Length* - MDL); contudo, existem outros critérios para encontrar o valor de M [39]. Ferrari et al. [45, 46, 47] aplicaram com sucesso o critério MDL para encontrar o número de núcleos Gaussianos em um modelo de mistura de Gaussianas, no contexto de detecção do disco fibro-glandular em mamogramas. Contudo, nossos experimentos mostram que não é apropriado usar o critério MDL no trabalho presente, pois há grande sobreposição, no domínio da escala Hounsfield, entre os núcleos Gaussianos que desejamos identificar. Esta conclusão deriva de nossa implementação da função de custo MDL, definida como

$$\kappa = \log L(\Theta|\mathbf{x}) - \frac{\nu_M}{2} \log(M) ,$$

onde $\nu_M = (M - 1) + Md + M \frac{d(d+1)}{2}$ é o número de parâmetros livres do modelo de mistura e $d = 1$ é a dimensão do espaço de características [46].

Desta forma, como identificar o melhor número de componentes histológicas da massa tumoral primária? Devemos investigar o conhecimento dos *experts* médicos, pois inexistem modelos de comprovada eficiência para o problema em questão. Para tanto, devemos realizar a estimação dos parâmetros do modelo de mistura de Gaussianas para diversos valores de M . Os resultados obtidos são apresentados a um médico experiente, na forma de gráficos. Este médico pode então decidir qual o modelo que melhor se ajusta a sua experiência e aos valores obtidos em cirurgia retardada. Um possível trabalho futuro, extendendo esta pesquisa, é modelar este processo manual de decisão por inspeção, e buscar um padrão emergente que permita a construção de um classificador automático para decidir o valor de M .

Modelos de misturas finitas são considerados ferramentas poderosas em tarefas de classificação não-supervisionadas [39]. Modelos de mistura de Gaussianas são o tipo mais comum de modelos de mistura [39], e o algoritmo EM é o método

padrão de estimação de parâmetros em um modelo de mistura de Gaussianas [39, 48]. Modelos de mistura foram empregados com sucesso em processamento de imagens para classificação não-supervisionada [39, 48], segmentação automática do cérebro em imagens de MRI [48] e de estruturas em mamogramas [46, 47], reconhecimento automático de alvo [42], correção de não-uniformidade de intensidade em MRI [41], caracterização de tecidos [41, 49], e segmentação de volume parcial em MRI [49].

Jain et al. [39] fazem notar que tópicos atuais em pesquisa do uso do algoritmo EM são: tratamento de sua natureza local, que faz com que o algoritmo seja criticamente dependente do valor inicial de Θ ; e a falta de limitação dos parâmetros (pois, no princípio de máxima verossimilhança, nenhuma probabilidade *a priori* é atribuída a Θ) que pode levar Θ a convergir para pontos indesejáveis do espaço de características, como ter α_i e σ_i aproximando-se simultaneamente de zero para o i -ésimo núcleo Gaussiano. Enquanto este último problema não foi encontrado em nosso trabalho, o primeiro foi enfrentado: o problema de encontrar uma boa estimativa inicial de Θ . A melhor abordagem para se obter um bom ajuste do modelo estimado à dada função densidade de probabilidade (ou seja, o histograma normalizado dos voxels na massa tumoral) foi selecionar manualmente o valor inicial de média para cada núcleo Gaussiano e atribuir a σ_i um valor grande (comparado com o desvio padrão do histograma), tal que cada núcleo Gaussiano pudesse ‘sentir’ o histograma completo e que sua média convergisse para valores apropriados.

Capítulo 4

Implementação e Análise dos Resultados

Para que avaliássemos a viabilidade e o potencial das técnicas anteriormente descritas, obtivemos imagens de tomografia computadorizada de pacientes com neuroblastoma em estágio avançado, em cooperação com o Instituto da Criança da Faculdade de Medicina da Universidade de São Paulo e o Hospital Sírio-Libanês. Em nosso trabalho a coleta de dados é prospectiva, pois trata-se de um desenvolvimento novo nas instituições envolvidas. Desta forma, o volume de dados disponíveis para nosso estudo foi restrito. Não obstante, tal montante de dados permitiu que desenvolvêssemos idéias e conclusões.

Dispusemos de três conjuntos de exames de tomografia computadorizada, correspondentes a dois pacientes com neuroblastoma estágio III. Cada paciente foi examinado ao diagnóstico da doença, atividade que produziu dois conjuntos de dados pré-tratamento. Um dos pacientes faleceu durante o tratamento, o outro foi tratado com sucesso. Deste ultimo adquirimos um novo exame pré-cirurgia de resecção. Extirpada a massa, produziu-se um relatório patológico da mesma, que pôde ser confrontado com os resultados das técnicas aqui desenvolvidas. A tabela 4.1 apresenta os conjuntos de dados analisados.

Caso	Descrição
A	Paciente do sexo feminino, dois anos de idade, com neuroblastoma estágio III. No diagnóstico inicial, realizou-se exame de tomografia computadorizada da região abdominal com 49 fatias, que demonstrou a existência de grande massa abdominal. A paciente faleceu durante o tratamento.
B	Novamente uma paciente do sexo feminino, dois anos de idade, com diagnóstico de neuroblastoma estágio III. O exame de CT da região abdominal, torácica e pelvis com 128 fatias, realizado no diagnóstico inicial (portanto antes do tratamento) revelou a existência de massa tumoral abdominal.
C	Novo exame da paciente do caso B, que foi tratada com sucesso. Foi realizado um exame de CT da região abdominal, com 33 fatias, que precedeu a cirurgia de ressecção completa da massa tumoral primária. Trata-se de uma massa pequena, posto que foi reduzida no processo de tratamento quimioterápico. Dispomos de análise patológica para este caso.

Tabela 4.1: Descrição dos casos analisados

4.1 Segmentação

Vemos na figura 4.1 uma fatia do exame de CT do tumor do caso ‘A’, com o contorno da massa delineado por um radiologista experiente. O exame foi realizado no diagnóstico inicial, logo a massa observada não havia sido tratada naquele momento. Desta forma, observamos um tumor de grande volume, com densidade alta (valor médio de aproximadamente 45 HU), que reflete, qualitativamente, a grande fração de doença ativa nesta massa.

Aplicamos a este exame de CT o processo de segmentação tridimensional fuzzy, descrito anteriormente. A figura 4.2 mostra a fatia do volume segmentado que corresponde àquela exibida na figura 4.1. Nesta imagem, os pixels mais claros possuem valores maiores de pertinência à massa tumoral. Nesta mesma figura observamos o contorno da massa tumoral, tal como exibido na figura 4.1, para comparação.

Podemos observar na figura 4.2 que a região da imagem que está dentro do contorno da massa tumoral apresenta, em sua maior parte, um alto valor de pertinência à massa tumoral. Observamos também que a maioria dos pixels externos ao contorno da massa possuem baixo valor de pertinência à massa tumoral, com exceção de

uma pequena fração de pixels acima da região delimitada. Desta forma, podemos concluir que nesta fatia a segmentação fuzzy concorda com a delimitação manual do tumor. Por inspeção, tal conclusão aplica-se também às outras fatias deste mesmo exame (não exibidas no texto): a extensão do tumor, tal como obtida pelo processo de segmentação fuzzy tridimensional, concorda com a delimitação do volume do tumor.

Na figura 4.3 apresentamos uma fatia do exame de CT do tumor do caso ‘B’. Neste caso temos novamente um tumor pré-tratamento, para o qual espera-se predominância de massa tumoral viável. O resultado da segmentação fuzzy desta massa é ilustrado na figura 4.4, que exhibe a fatia do volume de pertinência que corresponde à fatia mostrada na figura 4.3. Assim como na análise do caso ‘A’, podemos observar que a extensão do tumor, identificada pelo algoritmo de segmentação fuzzy, concorda com a delimitação manual da massa.

Tal concordância não foi observada no caso ‘C’, como podemos observar nas figuras 4.5 e 4.6. Neste caso a massa tumoral foi reduzida por tratamento quimioterápico. A composição histológica desta massa é diferente daquela obtida para o tumor do caso ‘B’ (mesmo paciente, porém antes do tratamento): há menos tumor viável presente.

A figura 4.5 apresenta uma fatia do exame de CT do caso ‘C’, sobre a qual está desenhado o contorno da massa tumoral. Observe que o nível de cinza dos pixels da região interna à imagem é semelhante ao dos pixels externos ao contorno da massa: de fato, os tecidos vizinhos à massa tumoral são vistos, no exame de CT, com valores em escala Hounsfield semelhantes aos encontrados na própria massa. Em consequência, vemos na figura 4.6 que diversos voxels que estão fora do contorno do tumor recebem alto valor de pertinência à região do tumor, semelhantes à pertinência atribuída aos voxels da região do tumor.



Figura 4.1: Segmentação manual da massa tumoral, caso 'A'.

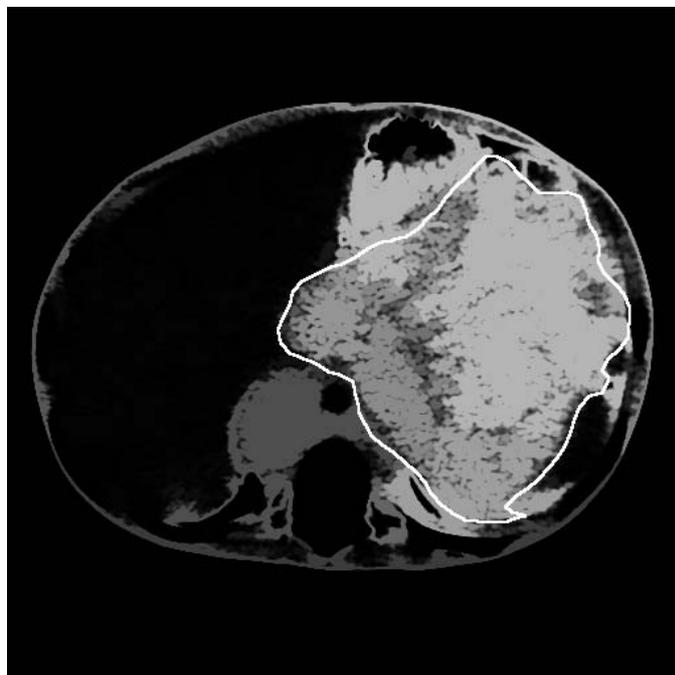


Figura 4.2: Conectividade fuzzy da região obtida por crescimento de região para o tumor do caso 'A' com o contorno manual superposto para comparação. Veja também a Figura 4.1.



Figura 4.3: Segmentação manual da massa tumoral, caso 'B'.

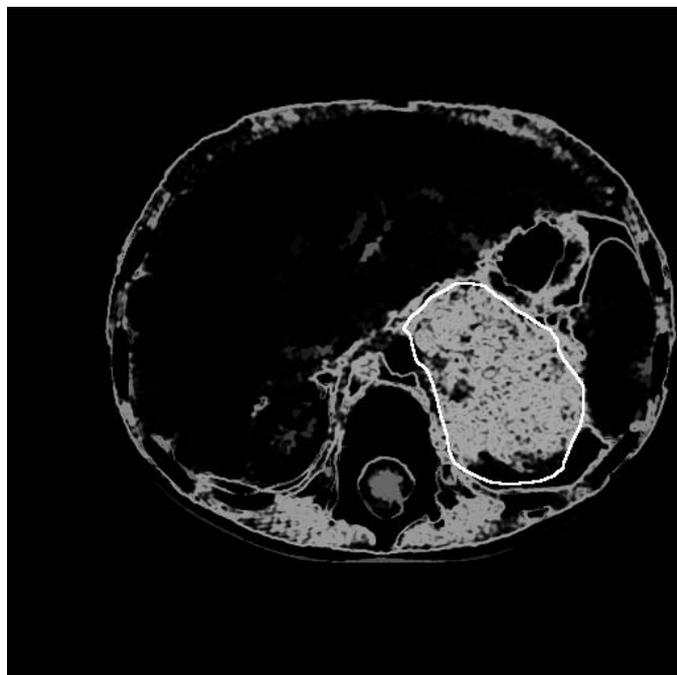


Figura 4.4: Conectividade fuzzy da região obtida por crescimento de região para o tumor do caso 'B' com o contorno manual superposto para comparação. Veja também a Figura 4.3.

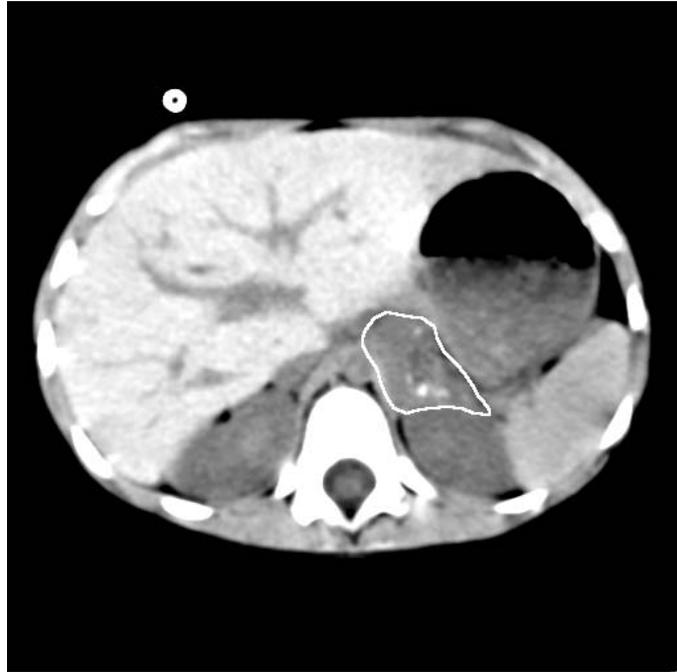


Figura 4.5: Segmentação manual da massa tumoral, caso 'C'.

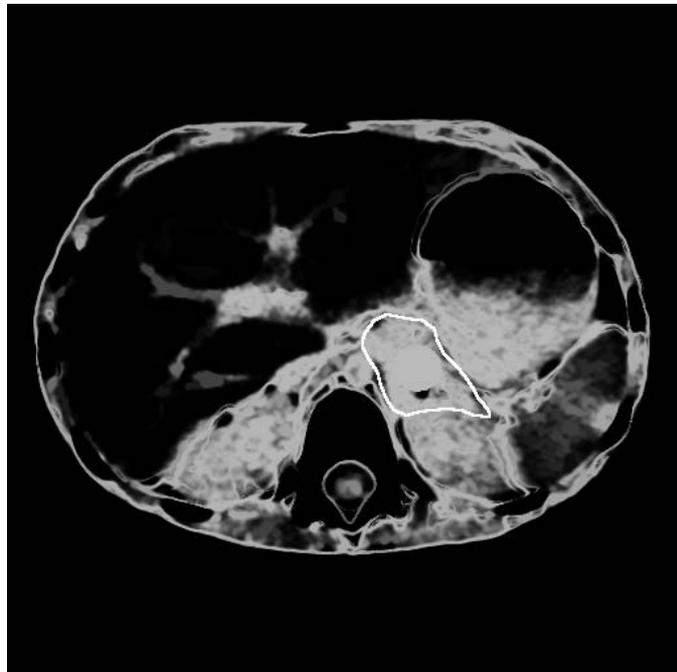


Figura 4.6: Conectividade fuzzy da região obtida por crescimento de região para o tumor do caso 'C' com o contorno manual superposto para comparação. Veja também a Figura 4.5.

4.2 Estimação da Composição Histológica

Para investigar o desempenho do algoritmo Expectation-Maximization para modelos de mistura de Gaussianas, no contexto do problema presente, implementamos o algoritmo e realizamos alguns testes com dados sintéticos e com o histograma real do tumor.

4.2.1 Teste com dados sintéticos

As funções densidade de probabilidade sintéticas são construídas em torno da faixa de valores de CT para tecidos moles, para que o algoritmo EM seja avaliado sob condições mais realistas.

No primeiro teste sintetizamos uma função densidade de probabilidade com três núcleos Gaussianos bem espaçados entre si (ou seja, há uma grande distância entre as médias se comparadas com o desvio padrão de cada Gaussiana). A tabela 4.2 apresenta os parâmetros para a síntese da função densidade de probabilidade de teste, bem como os resultados dos modelos estimados. A figura 4.7 mostra simultaneamente o gráfico da função densidade de probabilidade sintetizada, e os gráficos correspondentes aos modelos estimados. A figura 4.8 apresenta quatro gráficos, cada um destes apresenta os núcleos Gaussianos estimados correspondentes a cada modelo.

Neste caso, podemos observar na figura 4.7 que a estimação de um modelo de mistura de Gaussianas com apenas dois núcleos Gaussianos produz um resultado inadequado, que não se ajusta à função sintetizada. Tal fato é ilustrado também no primeiro gráfico da figura 4.8 e na tabela 4.2, em que podemos observar que um dos núcleos Gaussianos estimados ajusta-se a um dos núcleos da curva sintetizada, enquanto o outro núcleo estimado cobre a área de dois núcleos da curva sintetizada.

Porém, a estimação de modelos com três, quatro, e cinco núcleos Gaussianos produz resultados que se ajustam bem à curva original, conforme visto na figura 4.7. Podemos observar também que os valores de *log-likelihood* para cada caso (tabela 4.2)

concordam, com alta precisão, evidenciando que a qualidade do ajuste produzido em todos os casos é a mesma.

Notável é o ajuste do modelo estimado com três núcleos Gaussianos à curva sintetizada, que possui o mesmo número de núcleos. Observamos na tabela 4.2 que neste caso os parâmetros do modelo estimado ajustam-se aos do modelo sintetizado com precisão superior a duas casas decimais. Tal fato pode ser visualizado também na figura 4.8.

Como distinguir o melhor modelo estimado? Conforme mencionado na seção 3.2, a função de custo MDL não pode ser utilizada: tal conclusão deriva dos experimentos que aqui apresentamos. A tabela 4.2 mostra que o modelo estimado com dois núcleos Gaussianos apresenta o menor custo MDL, porém é evidente que este modelo é inadequado, como vimos acima.

A figura 4.8 mostra um padrão interessante: podemos observar que nos modelos estimados com quatro e cinco núcleos Gaussianos, há a superposição de alguns destes núcleos (mesma média e desvio padrão), que podem ser combinados. No modelo com quatro Gaussianas, os núcleos dois e três podem ser combinados em apenas uma Gaussiana. No modelo com cinco Gaussianas, podemos combinar os núcleos dois e três em uma Gaussiana, e os núcleos quatro e cinco em outra Gaussiana. Se estimamos um modelo de mistura de Gaussianas e tal comportamento se manifesta, temos a evidência de que a redução do número de Gaussianas, no modelo estimado, pode levar a um modelo mais simples para a função densidade de probabilidade sendo analisada. Esta heurística é uma primeira estratégia de seleção de modelos que pode ser aplicada ao problema em questão, de estimar o modelo de mistura de Gaussianas que melhor represente a composição histológica da massa tumoral em um paciente com neuroblastoma.

Em nosso segundo experimento com dados artificiais, novamente sintetizamos uma função densidade de probabilidade com três núcleos Gaussianos, mas os espaçamentos entre eles é menor que no caso anterior. No terceiro experimento o

espaçamento é ainda menor, de modo que a diferença entre os núcleos Gaussianos não se manifesta visualmente na figura 4.11. Portanto, vamos nos referir ao segundo e terceiro experimentos como teste com espaçamento intermediário e com espaçamento estreito, respectivamente.

O gráfico da figura 4.9 apresenta a função densidade de probabilidade sintética, com espaçamento intermediário entre núcleos, e os modelos estimados com dois, três, quatro, e cinco núcleos Gaussianos a ela superpostos. Observamos que apenas o modelo estimado com dois núcleos Gaussianos não provê uma curva que bem se ajusta à função original: nos demais casos os modelos estimados ajustam-se adequadamente. A figura 4.10 apresenta quatro gráficos, correspondentes aos quatro modelos estimados, em que podemos observar os diversos núcleos Gaussianos estimados. Ambas as figuras são derivadas dos valores da tabela 4.3.

Neste caso, observamos fenômenos similares àqueles descritos acima para o caso de núcleos Gaussianos bem espaçados. Novamente, quando o número de Gaussianas no modelo sendo estimado é igual ao real número de núcleos Gaussianos (três núcleos), os parâmetros do modelo estimado aproximam-se com alta precisão dos valores reais. Observamos também que, se estimamos um modelo cujo número de núcleos supera o número verdadeiro de Gaussianas na função densidade de probabilidade sendo modelada, então alguns núcleos do modelo estimado podem ser combinados em um único núcleo, reduzindo portanto nossa estimativa do número correto de Gaussianas. Deste modo, quando estimamos um modelo com quatro núcleos Gaussianos, os núcleos dois e três podem ser combinados em apenas um. Analogamente, quando estimamos um modelo com cinco Gaussianas, os núcleos dois e três novamente podem ser combinados, vale o mesmo para os núcleos quatro e cinco.

Entretanto, neste caso, a similaridade dos núcleos Gaussianos a serem combinados é muito menos evidente do que no teste anterior, onde a função densidade de probabilidade original apresentava um maior espaçamento entre seus núcleos

Gaussianos. Desta forma, para que possamos combinar pares de núcleos Gaussianos, devemos possuir alguma informação *a priori* sobre o número de Gaussianas na função densidade de probabilidade sendo estimada. Por exemplo, no teste de estimação de quatro ou cinco núcleos Gaussianos, podemos argumentar a favor do agrupamento das curvas estimadas (conforme apresentado no parágrafo anterior) mesmo se não conhecemos o número exato de Gaussianas na função densidade de probabilidade original: sabemos que este número deve ser pequeno, e os modelos estimados apresentam curvas bastante similares.

Em nosso último teste desta seção, construímos uma função densidade de probabilidade com três núcleos Gaussianos, com espaçamento reduzido entre eles, se comparado aos dois testes anteriores. Os parâmetros que definem este teste podem ser vistos na tabela 4.4, bem como os parâmetros dos modelos estimados com dois, três, quatro, e cinco núcleos Gaussianos. Podemos observar na figura 4.11 que a função densidade de probabilidade construída é unimodal, e sua forma não permite perceber o número de núcleos Gaussianos que a compõe. Observamos na mesma figura que os modelos estimados ajustam-se bem à curva original, mesmo aquele composto de dois núcleos apenas - fenômeno que não ocorre nos testes anteriores.

Considerando a forma da função densidade de probabilidade, e a dificuldade que esta acarreta na sua decomposição em núcleos Gaussianos, é notável que observemos um bom grau de concordância entre os parâmetros reais da curva simulada e aqueles obtidos pela estimação de um modelo de mistura de três Gaussianas, como pode ser visto na tabela 4.4. Entretanto, não podemos aplicar a heurística de estimação do número correto de Gaussianas, vista acima: note que as curvas obtidas em modelos com quatro e cinco Gaussianas dificilmente podem ser agrupadas. Mais ainda, o modelo estimado com dois núcleos Gaussianos adapta-se bem à curva simulada original, aumentando a indecisão na escolha do número ideal de núcleos Gaussianos.

Parâmetros dos dados simulados para o primeiro teste	$\alpha_1 = 0.3, \mu_1 = 0.0, \sigma_1 = 10.0$ $\alpha_2 = 0.3, \mu_2 = 40.0, \sigma_2 = 10.0$ $\alpha_3 = 0.4, \mu_3 = 80.0, \sigma_3 = 10.0$
Parâmetros estimados, 2 núcleos Gaussianos	$\alpha_1 = 0.65, \mu_1 = 23.77, \sigma_1 = 25.52$ $\alpha_2 = 0.35, \mu_2 = 81.19, \sigma_2 = 9.23$ log-likelihood = -4.819195 custo MDL = -6.552063
Parâmetros estimados, 3 núcleos Gaussianos	$\alpha_1 = 0.30, \mu_1 = 0.0, \sigma_1 = 10.00$ $\alpha_2 = 0.30, \mu_2 = 40.00, \sigma_2 = 10.00$ $\alpha_3 = 0.40, \mu_3 = 80.00, \sigma_3 = 10.00$ log-likelihood = -4.732351 custo MDL = -9.126801
Parâmetros estimados, 4 núcleos Gaussianos	$\alpha_1 = 0.30, \mu_1 = -0.01, \sigma_1 = 9.99$ $\alpha_2 = 0.13, \mu_2 = 40.02, \sigma_2 = 10.47$ $\alpha_3 = 0.17, \mu_3 = 39.98, \sigma_3 = 9.67$ $\alpha_4 = 0.40, \mu_4 = 80.01, \sigma_4 = 9.99$ log-likelihood = -4.732351 custo MDL = -12.356970
Parâmetros estimados, 5 núcleos Gaussianos	$\alpha_1 = 0.30, \mu_1 = -0.01, \sigma_1 = 10.00$ $\alpha_2 = 0.11, \mu_2 = 40.51, \sigma_2 = 10.75$ $\alpha_3 = 0.19, \mu_3 = 39.77, \sigma_3 = 9.64$ $\alpha_4 = 0.19, \mu_4 = 82.82, \sigma_4 = 9.65$ $\alpha_5 = 0.21, \mu_5 = 77.47, \sigma_5 = 9.56$ log-likelihood = -4.732351 custo MDL = -15.998417

Tabela 4.2: Resultados do algoritmo EM, com dados simulados que consistem de núcleos amplamente espaçados.

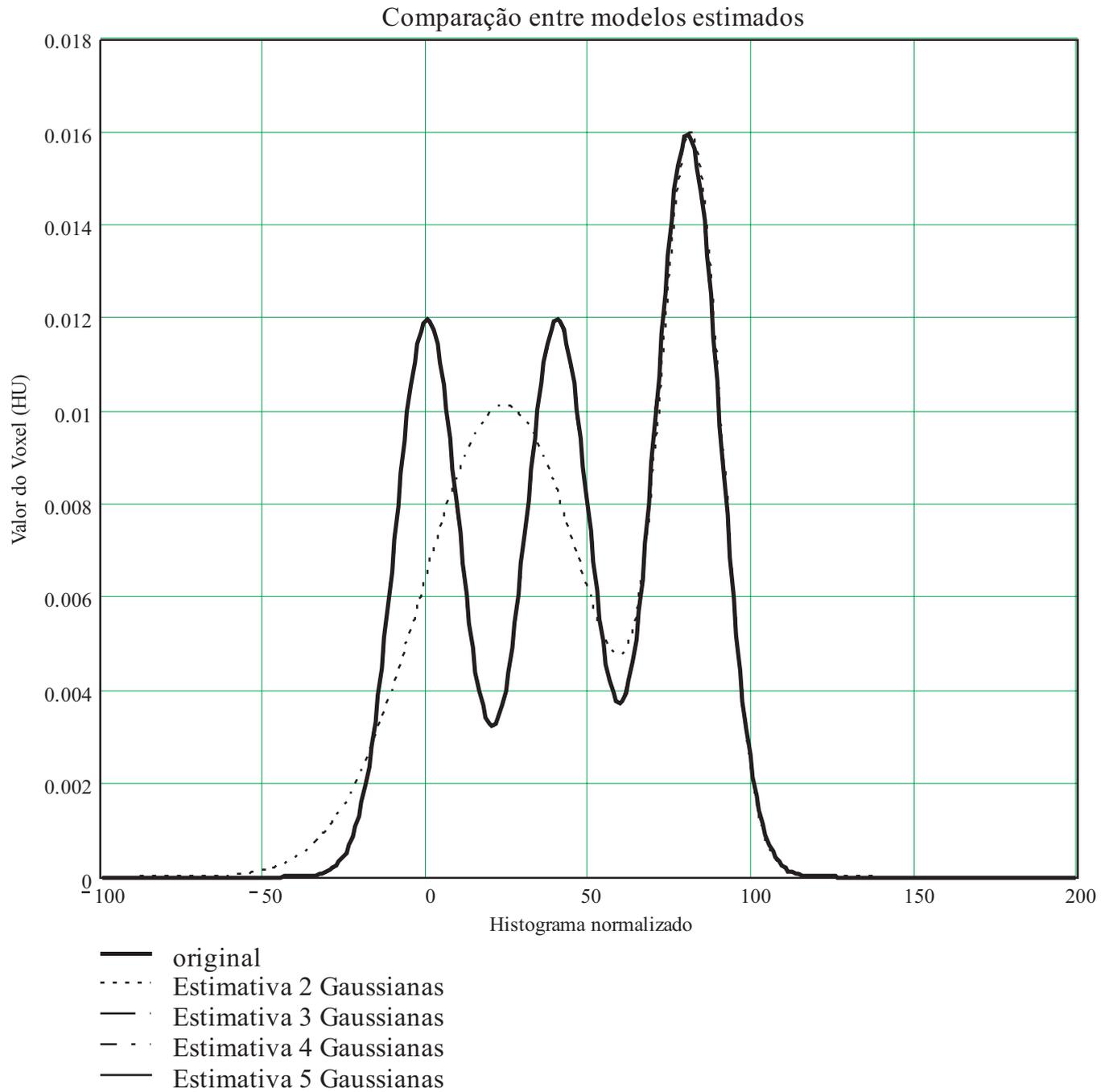


Figura 4.7: Resultados do algoritmo EM para o modelo de mistura de Gaussianas usando dados simulados com núcleos amplamente espaçados. As funções densidade de probabilidade para os modelos com três, quatro e cinco Gaussianas não são visíveis pois se sobrepõem à função original. Veja também 4.2.

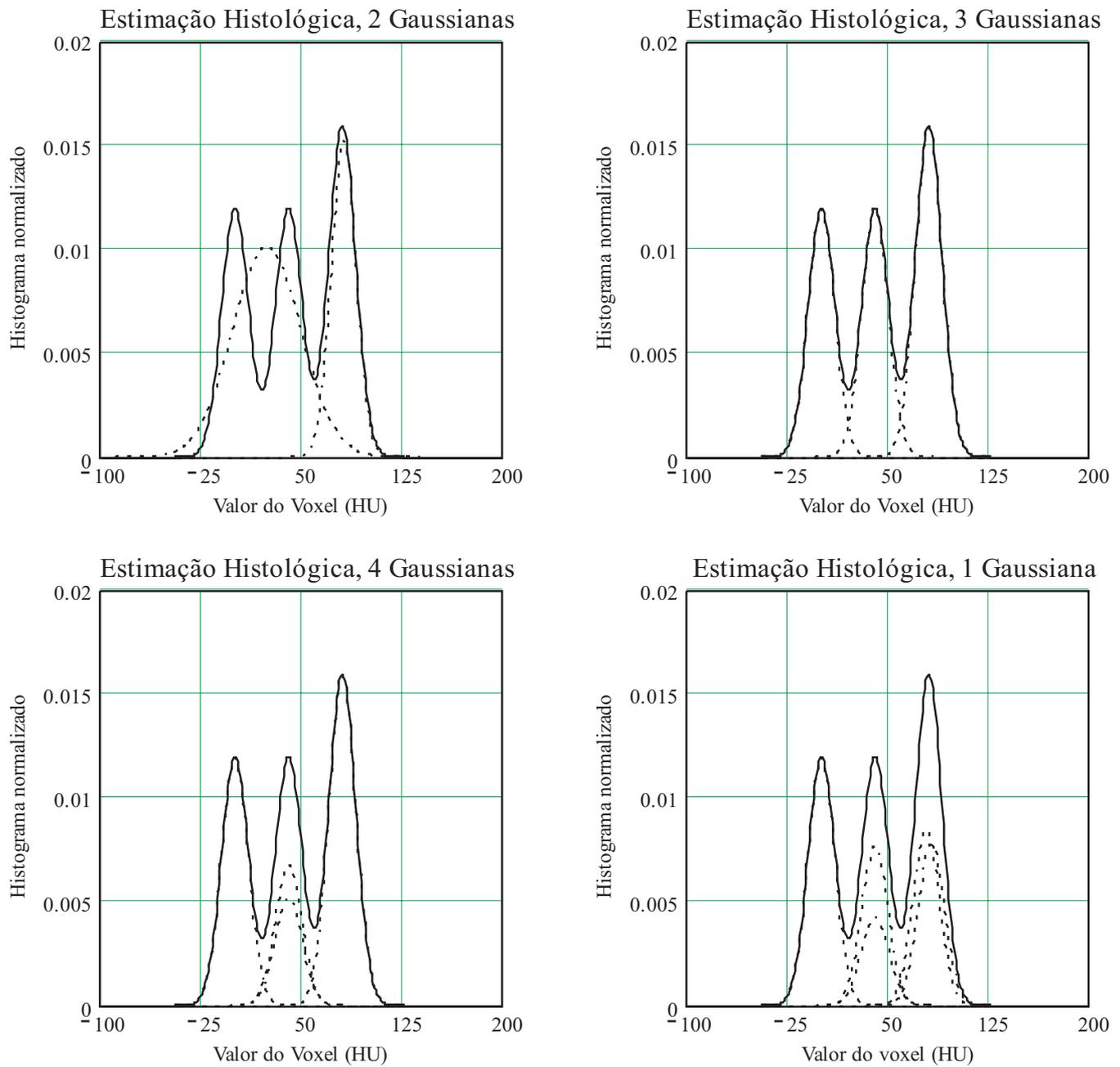


Figura 4.8: Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados simulados com núcleos amplamente espaçados. A linha sólida representa a função densidade de probabilidade para o dado original, e as linhas tracejadas representam as componentes estimadas. Veja também Tabela 4.2 e Figura 4.7.

Parâmetros dos dados simulados para o segundo teste	$\alpha_1 = 0.3, \mu_1 = 40.0, \sigma_1 = 5.0$ $\alpha_2 = 0.2, \mu_2 = 50.0, \sigma_2 = 5.0$ $\alpha_3 = 0.5, \mu_3 = 60.0, \sigma_3 = 5.0$
Parâmetros estimados, 2 núcleos Gaussianos	$\alpha_1 = 0.41, \mu_1 = 41.99, \sigma_1 = 5.86$ $\alpha_2 = 0.59, \mu_2 = 58.89, \sigma_2 = 5.51$ log-likelihood = -3.658276 custo MDL = -5.391144
Parâmetros estimados, 3 núcleos Gaussianos	$\alpha_1 = 0.30, \mu_1 = 39.99, \sigma_1 = 5.00$ $\alpha_2 = 0.20, \mu_2 = 50.01, \sigma_2 = 5.02$ $\alpha_3 = 0.50, \mu_3 = 60.01, \sigma_3 = 5.00$ log-likelihood = -3.656832 custo MDL = -8.051281
Parâmetros estimados, 4 núcleos Gaussianos	$\alpha_1 = 0.28, \mu_1 = 39.86, \sigma_1 = 4.97$ $\alpha_2 = 0.10, \mu_2 = 48.70, \sigma_2 = 6.41$ $\alpha_3 = 0.12, \mu_3 = 50.23, \sigma_3 = 4.73$ $\alpha_4 = 0.49, \mu_4 = 60.02, \sigma_4 = 5.00$ log-likelihood = -3.656835 custo MDL = -11.281454
Parâmetros estimados, 5 núcleos Gaussianos	$\alpha_1 = 0.28, \mu_1 = 39.82, \sigma_1 = 4.96$ $\alpha_2 = 0.10, \mu_2 = 48.00, \sigma_2 = 6.40$ $\alpha_3 = 0.14, \mu_3 = 50.64, \sigma_3 = 4.84$ $\alpha_4 = 0.29, \mu_4 = 61.17, \sigma_4 = 4.84$ $\alpha_5 = 0.19, \mu_5 = 58.62, \sigma_5 = 4.64$ log-likelihood = -3.656832 custo MDL = -14.922897

Tabela 4.3: Resultados do algoritmo EM para dados simulados, espaçamento mediano entre os núcleos.

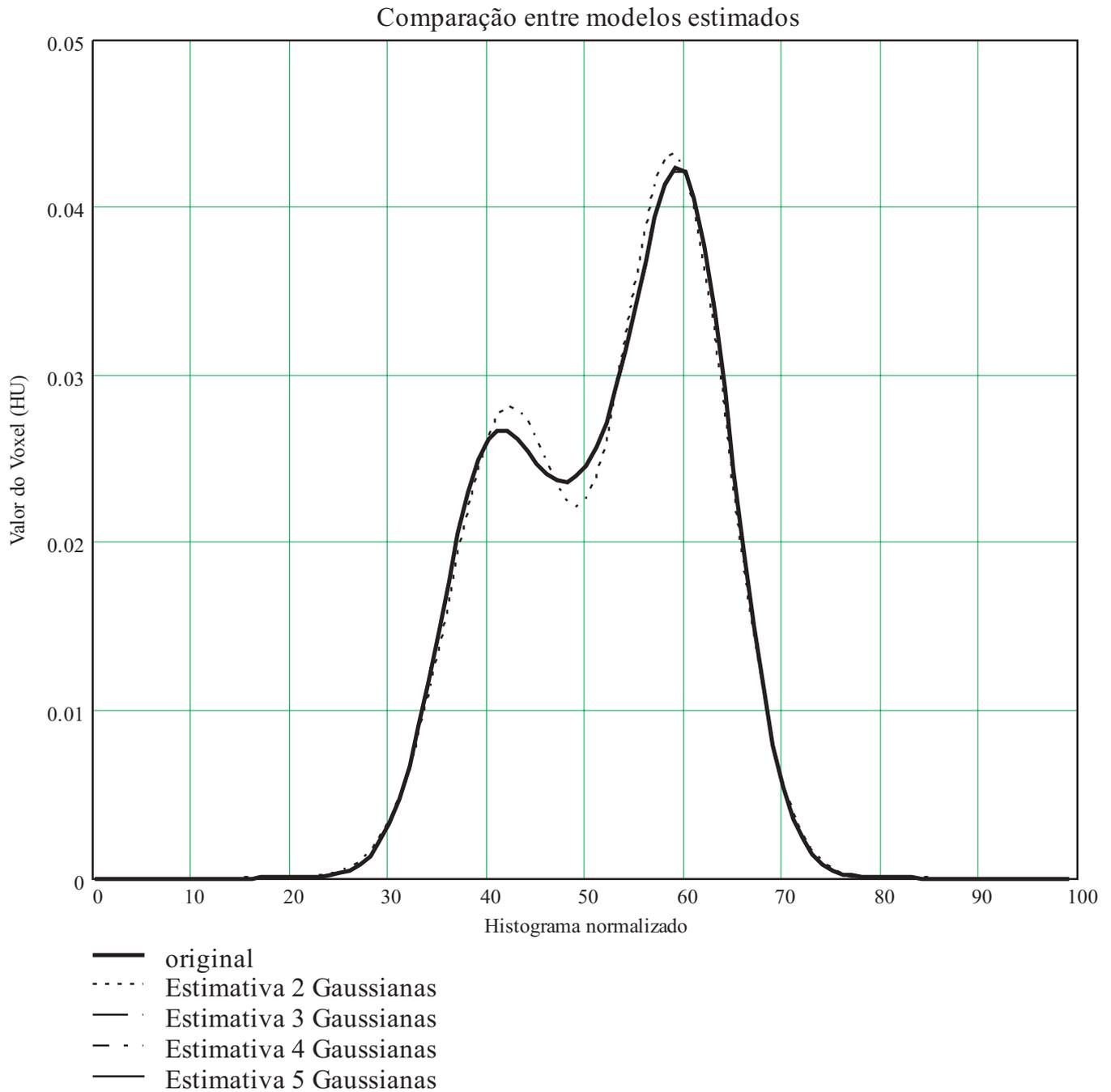


Figura 4.9: Resultados do algoritmo EM para modelo de mistura de Gaussianas usando dados simulados com espaçamento intermediário entre núcleos (veja também Figura 4.7 e Figura 4.11). As funções densidade de probabilidade para os modelos com três, quatro e cinco Gaussianas não são visíveis pois se sobrepõem à função original. Veja também Tabela 4.3.

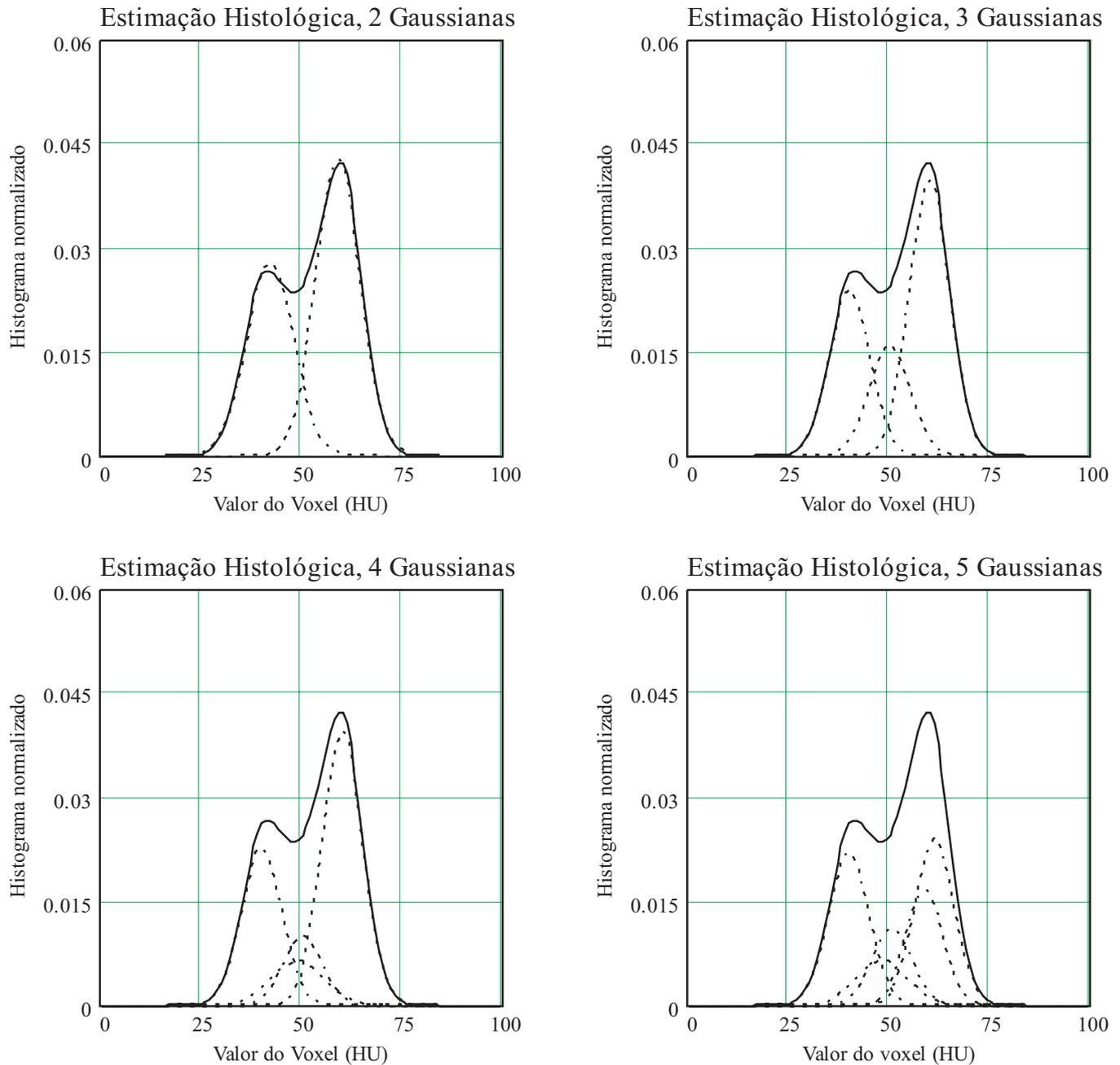


Figura 4.10: Resultados do algoritmo EM para modelo de mistura de Gaussianas usando dados simulados com espaçamento intermediário entre núcleos (veja também Figura 4.7 e Figura 4.11). A linha sólida representa a função densidade de probabilidade para o dado original, e as linhas tracejadas representam as componentes estimadas. Veja também Tabela 4.3 e Figura 4.9.

Parâmetros dos dados simulados para o terceiro teste	$\alpha_1 = 0.2, \mu_1 = 32.0, \sigma_1 = 10.0$ $\alpha_2 = 0.4, \mu_2 = 44.0, \sigma_2 = 14.0$ $\alpha_3 = 0.4, \mu_3 = 60.0, \sigma_3 = 14.0$
Parâmetros estimados, 2 núcleos Gaussianos	$\alpha_1 = 0.33, \mu_1 = 33.58, \sigma_1 = 10.88$ $\alpha_2 = 0.67, \mu_2 = 55.01, \sigma_2 = 15.04$ log-likelihood = -4.249527 custo MDL = -5.982595
Parâmetros estimados, 3 núcleos Gaussianos	$\alpha_1 = 0.19, \mu_1 = 31.88, \sigma_1 = 9.94$ $\alpha_2 = 0.37, \mu_2 = 42.75, \sigma_2 = 13.73$ $\alpha_3 = 0.44, \mu_3 = 59.29, \sigma_3 = 14.10$ log-likelihood = -4.249510 custo MDL = -8.643959
Parâmetros estimados, 4 núcleos Gaussianos	$\alpha_1 = 0.15, \mu_1 = 31.50, \sigma_1 = 9.66$ $\alpha_2 = 0.35, \mu_2 = 39.41, \sigma_2 = 12.89$ $\alpha_3 = 0.29, \mu_3 = 54.59, \sigma_3 = 12.87$ $\alpha_4 = 0.22, \mu_4 = 63.89, \sigma_4 = 13.47$ log-likelihood = -4.249509 MDL cost = -11.874129
Parâmetros estimados, 5 núcleos Gaussianos	$\alpha_1 = 0.17, \mu_1 = 31.82, \sigma_1 = 9.80$ $\alpha_2 = 0.17, \mu_2 = 35.93, \sigma_2 = 12.55$ $\alpha_3 = 0.30, \mu_3 = 47.04, \sigma_3 = 12.29$ $\alpha_4 = 0.26, \mu_4 = 59.54, \sigma_4 = 12.29$ $\alpha_5 = 0.10, \mu_5 = 68.32, \sigma_5 = 12.90$ log-likelihood = -4.249509 custo MDL = -15.515575

Tabela 4.4: Resultados do algoritmo EM para dados simulados, espaçamento estreito entre os núcleos.

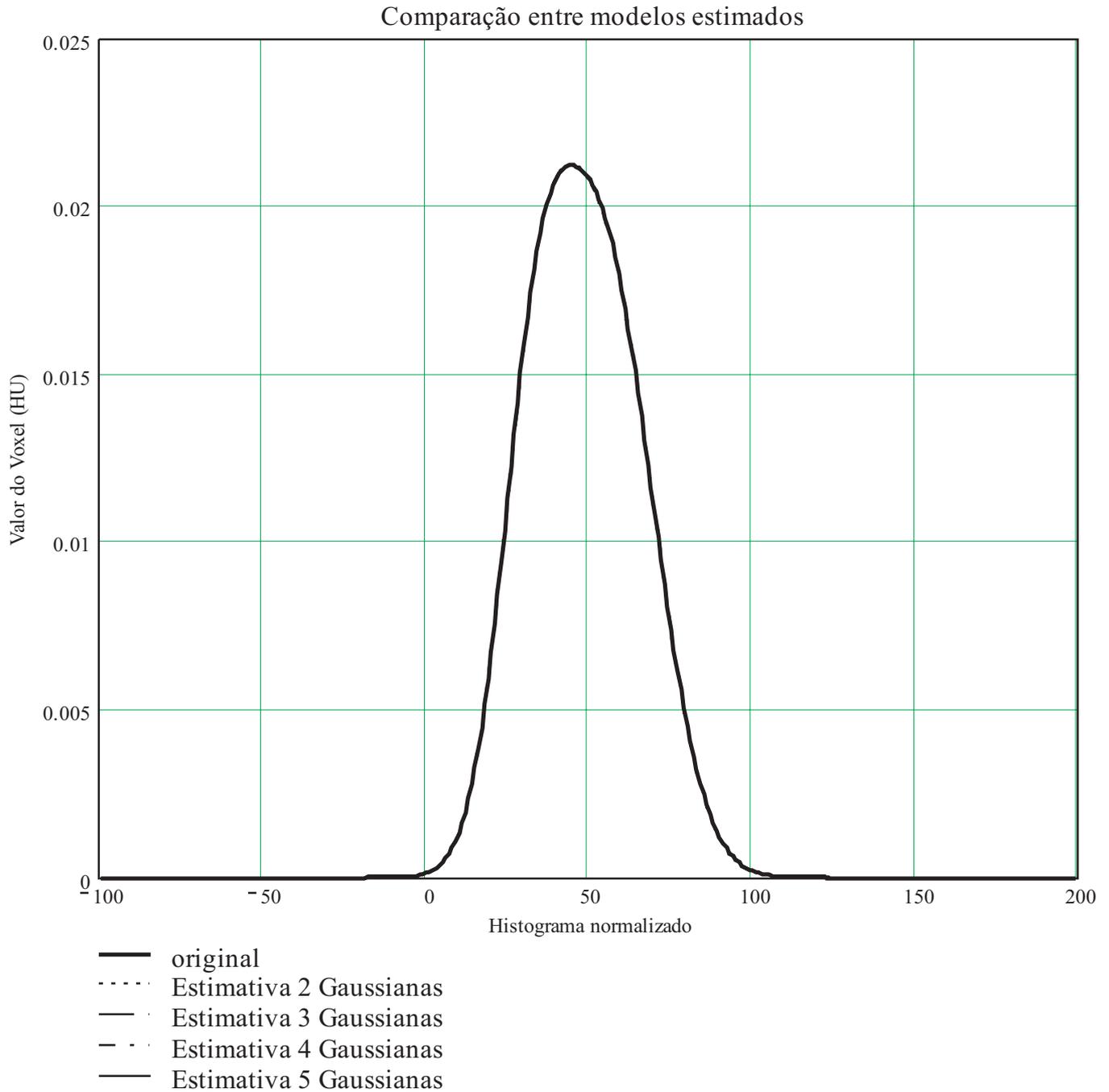


Figura 4.11: Resultados do algoritmo EM para modelo de mistura de Gaussianas usando dados simulados com espaçamento estreito entre núcleos (veja também Figura 4.7 e Figura 4.9). Todos os modelos se sobrepõem à função original. Veja também Tabela 4.4.

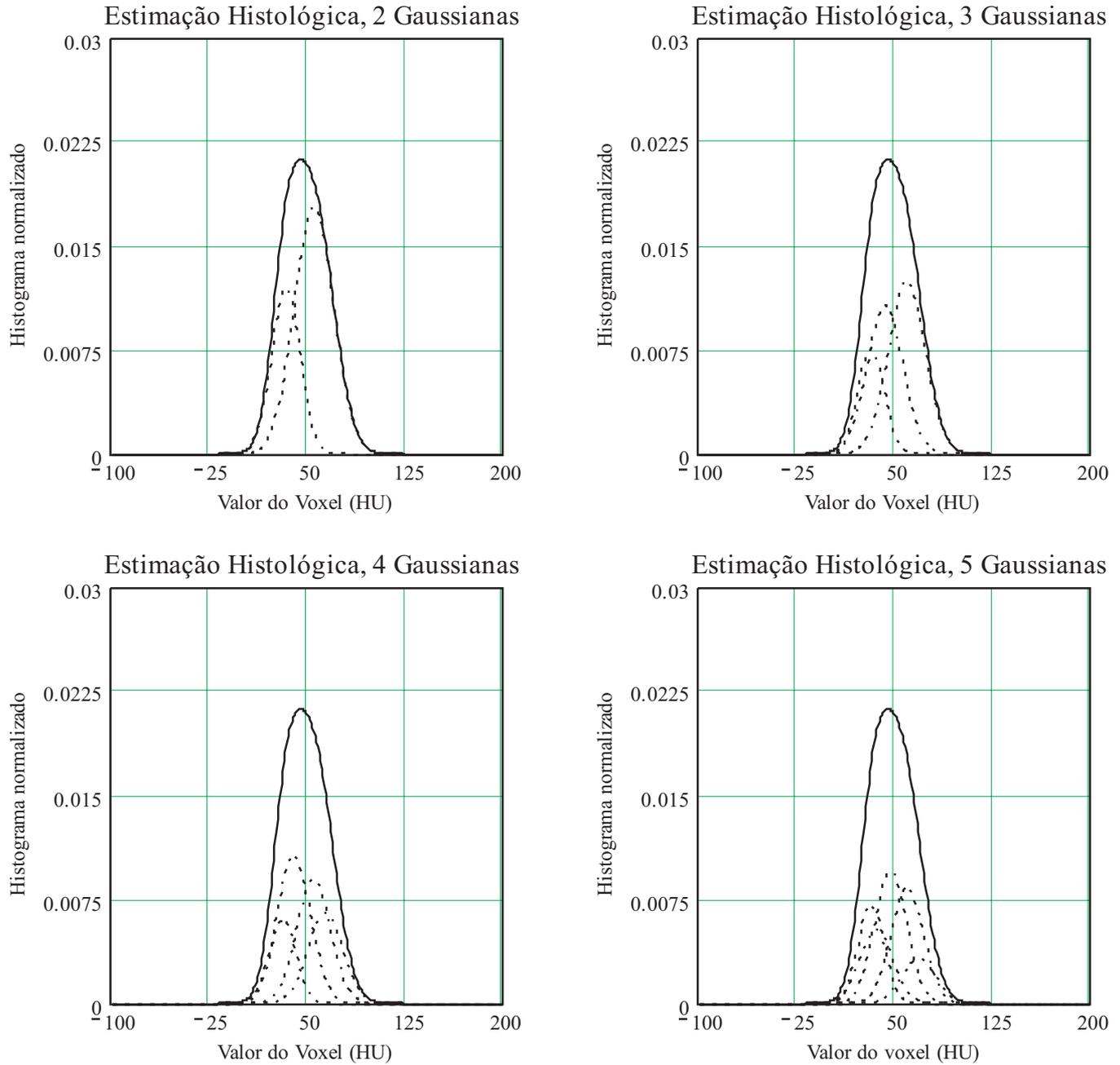


Figura 4.12: Resultados do algoritmo EM para modelo de mistura de Gaussianas usando dados simulados com espaçamento estreito entre núcleos. A linha sólida representa a função densidade de probabilidade para o dado original, e as linhas tracejadas representam as componentes estimadas. Veja também Tabela 4.4 e Figura 4.11.

4.2.2 Teste com dados reais

Nesta seção vamos analisar o comportamento do algoritmo EM para a estimação de modelos de mistura de Gaussianas, a partir de dados reais. Os dados foram obtidos a partir dos casos descritos no início do capítulo, usando segmentação manual.

Primeiro teste: Os resultados produzidos pelo algoritmo para o conjunto de dados do caso 'A' são mostrados na tabela 4.5, e nas figuras 4.13 e 4.14. Observamos que não ocorrem grupos de curvas Gaussianas que possam ser combinados, para nenhum modelo estimado. Porém observamos que os modelos com três e quatro núcleos Gaussianos são similares: as duas componentes Gaussianas de maior área (maior valor de α) são comparáveis entre os modelos. Poderíamos inferir que o número correto de núcleos Gaussianos a ser estimado dos dados reais é dois. Surpreendentemente, o modelo estimado com dois núcleos Gaussianos difere consideravelmente do resultado obtido a partir dos modelos com três e quatro núcleos Gaussianos.

Qualitativamente, o modelo que melhor representa o tumor em questão é aquele obtido a partir da estimativa com três núcleos Gaussianos, suprimindo-se a componente Gaussiana de menor valor. Tal conclusão deriva dos argumentos do parágrafo anterior, e da análise da figura 4.14 feita por um médico expert. Podemos observar no modelo estimado com três núcleos Gaussianos que, dentre as duas componentes Gaussianas de maior amplitude, a maior delas apresenta a maior média. Este comportamento pode ser explicado em termos da patologia esperada para o tumor deste caso: trata-se de um tumor antes do tratamento, logo apresenta elevado percentual de tumor viável, que é mais denso (portanto apresenta maior valor em HU) que os demais materiais que podem compor a massa tumoral.

Segundo teste: O segundo teste foi realizado a partir dos dados reais, do caso 'B', e são exibidos na tabela 4.6, e figuras 4.15 e 4.16. Temos novamente um tumor antes do tratamento quimioterápico. Porém existem duas diferenças fundamentais entre este caso e o caso 'A': no caso 'B' o tumor é menor (possui menos voxels), e

seu histograma assemelha-se mais a uma curva Gaussiana única, sem ressaltos - no caso 'A' podemos observar (figura 4.13) um ligeiro ressalto à esquerda do histograma normalizado. O tamanho reduzido do tumor implica em menor precisão nos modelos de mistura de Gaussianas estimados a partir do histograma deste tumor.

Podemos observar na tabela 4.6 que componentes Gaussianas de maior valor médio apresentam maior amplitude. Entretanto, tal comportamento não é tão evidente como no caso anterior. Isto decorre da forma quase Gaussiana do histograma do tumor, que reduz a capacidade de estimação do algoritmo, conforme observamos também na seção 4.2.1 nos testes com espaçamento estreito entre núcleos Gaussianos.

Terceiro teste: O último teste com dados reais foi realizado com os dados do tumor do caso 'C', para o qual dispomos de análise patológica. Os resultados obtidos são mostrados na tabela 4.7 e nas figuras 4.17 e 4.18. Neste caso, o tumor foi imageado após tratamento quimioterápico. Após o imageamento, a massa foi resectada e submetida à análise patológica. O relatório desta análise informa que a massa não possui mais que 30% de tumor viável.

Observando a figura 4.18, podemos verificar qualitativamente que o modelo estimado com três núcleos Gaussianos é o que melhor estima a composição histológica da massa tumoral. De fato, dentre as duas curvas Gaussianas mais significativas deste modelo, aquela com menor valor médio apresenta maior amplitude. Isto concorda com a expectativa advinda do conhecimento médico do tumor: a massa após tratamento possui um percentual reduzido de tumor viável. Mais ainda, observamos na tabela 4.7 que a curva de maior amplitude corresponde a cerca de 77% da massa tumoral, este fato concorda com o resultado da análise patológica.

Parâmetros estimados, 1 núcleo Gaussiano	$\alpha_1 = 1.00, \mu_1 = 45.22, \sigma_1 = 11.60$
Parâmetros estimados, 2 núcleos Gaussianos	$\alpha_1 = 0.27, \mu_1 = 49.88, \sigma_1 = 6.09$ $\alpha_2 = 0.73, \mu_2 = 43.47, \sigma_2 = 12.65$
Parâmetros estimados, 3 núcleos Gaussianos	$\alpha_1 = 0.14, \mu_1 = 28.75, \sigma_1 = 5.91$ $\alpha_2 = 0.84, \mu_2 = 47.82, \sigma_2 = 8.99$ $\alpha_3 = 0.02, \mu_3 = 46.60, \sigma_3 = 29.92$
Parâmetros estimados, 4 núcleos Gaussianos	$\alpha_1 = 0.14, \mu_1 = 29.28, \sigma_1 = 5.787$ $\alpha_2 = 0.08, \mu_2 = 45.45, \sigma_2 = 18.41$ $\alpha_3 = 0.78, \mu_3 = 47.96, \sigma_3 = 8.57$ $\alpha_4 = 0.001, \mu_4 = 69.80, \sigma_4 = 86.27$

Tabela 4.5: Parâmetros do modelo de mistura de Gaussianas estimados para o tumor do caso 'A', manualmente segmentado. A região do tumor contém 946782 voxels, o valor médio para o tumor é 45.22 HU e o desvio padrão é 11.60 HU.

Parâmetros estimados, 1 núcleo Gaussiano	$\alpha_1 = 1.00, \mu_1 = 54.88, \sigma_1 = 20.25$
Parâmetros estimados, 2 núcleos Gaussianos	$\alpha_1 = 0.21, \mu_1 = 50.17, \sigma_1 = 36.72$ $\alpha_2 = 0.79, \mu_2 = 56.15, \sigma_2 = 12.21$
Parâmetros estimados, 3 núcleos Gaussianos	$\alpha_1 = 0.14, \mu_1 = 44.95, \sigma_1 = 39.83$ $\alpha_2 = 0.60, \mu_2 = 55.46, \sigma_2 = 10.79$ $\alpha_3 = 0.26, \mu_3 = 58.92, \sigma_3 = 19.39$
Parâmetros estimados, 4 núcleos Gaussianos	$\alpha_1 = 0.14, \mu_1 = 44.79, \sigma_1 = 40.08$ $\alpha_2 = 0.55, \mu_2 = 54.50, \sigma_2 = 10.38$ $\alpha_3 = 0.06, \mu_3 = 65.42, \sigma_3 = 9.64$ $\alpha_4 = 0.25, \mu_4 = 58.48, \sigma_4 = 19.95$

Tabela 4.6: Parâmetros do modelo de mistura de Gaussianas estimados para o tumor do caso 'B', manualmente segmentado. A região do tumor contém 304732 voxels, o valor médio para o tumor é 54.88 HU e o desvio padrão é 20.25 HU.

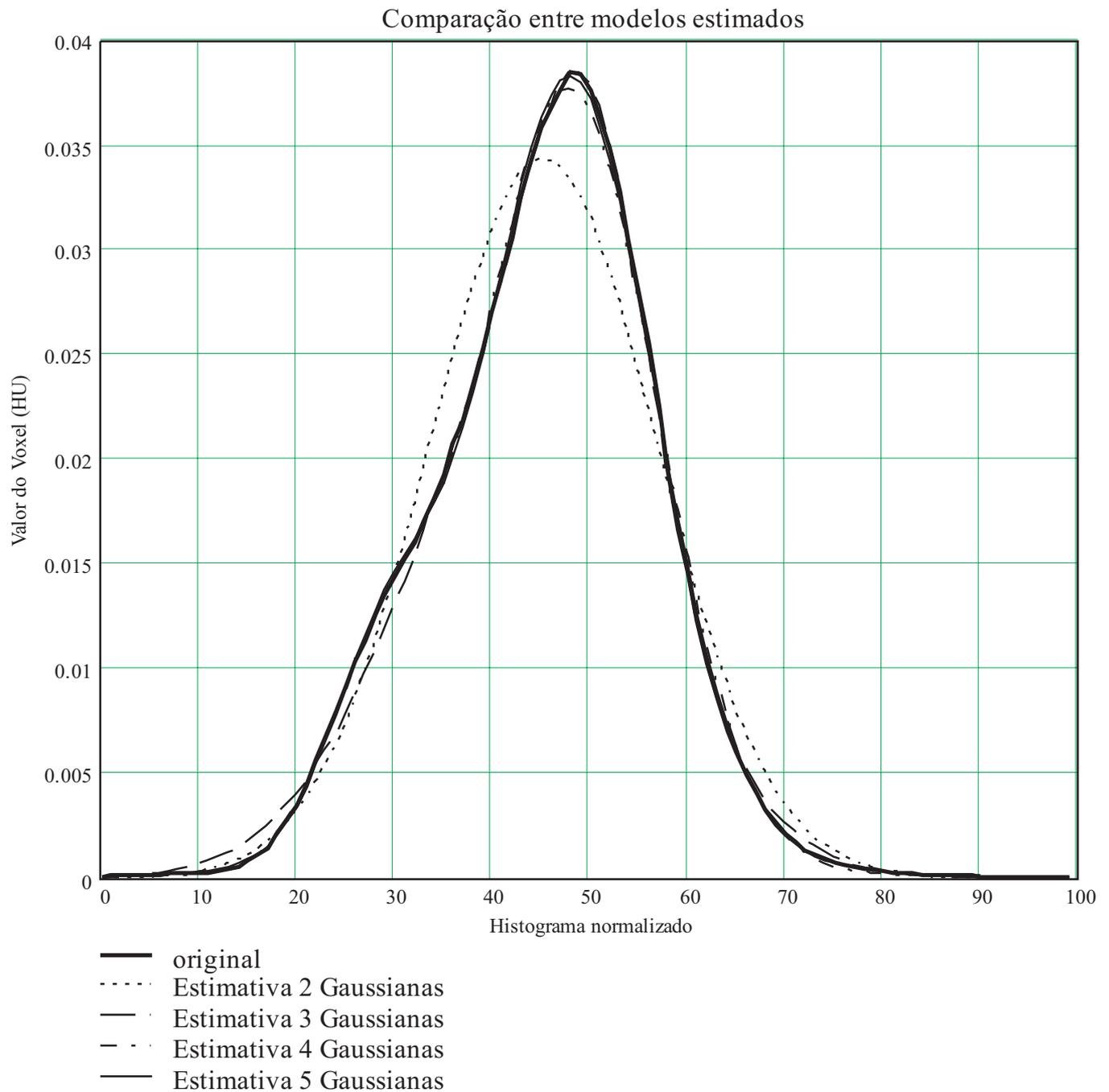


Figura 4.13: Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'A'). Ver também tabela 4.5.

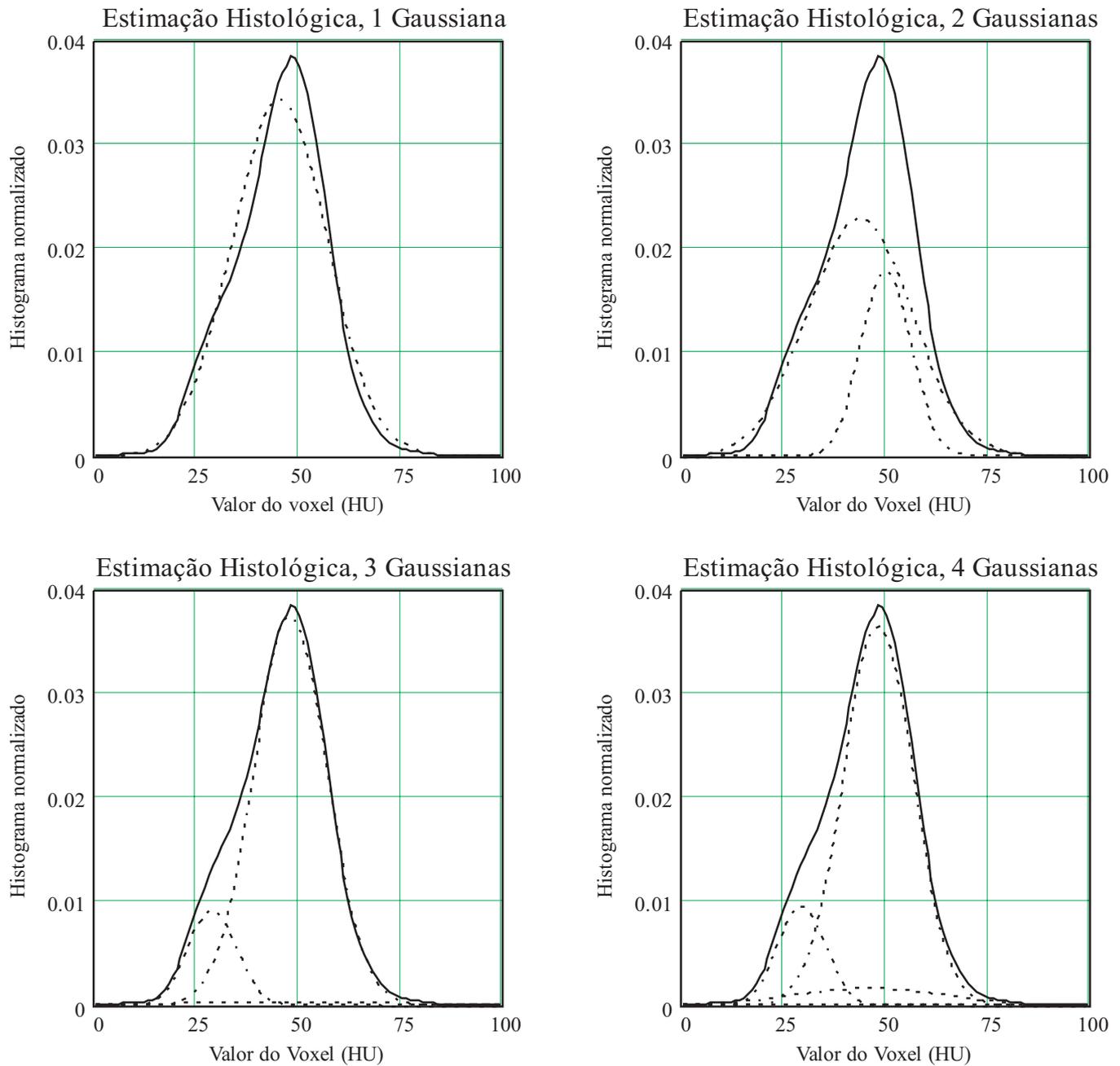


Figura 4.14: Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'A'). A linha sólida representa a função densidade de probabilidade para o dado original e as linhas tracejadas representam as componentes estimadas. veja também tabela 4.5 e figura 4.13.

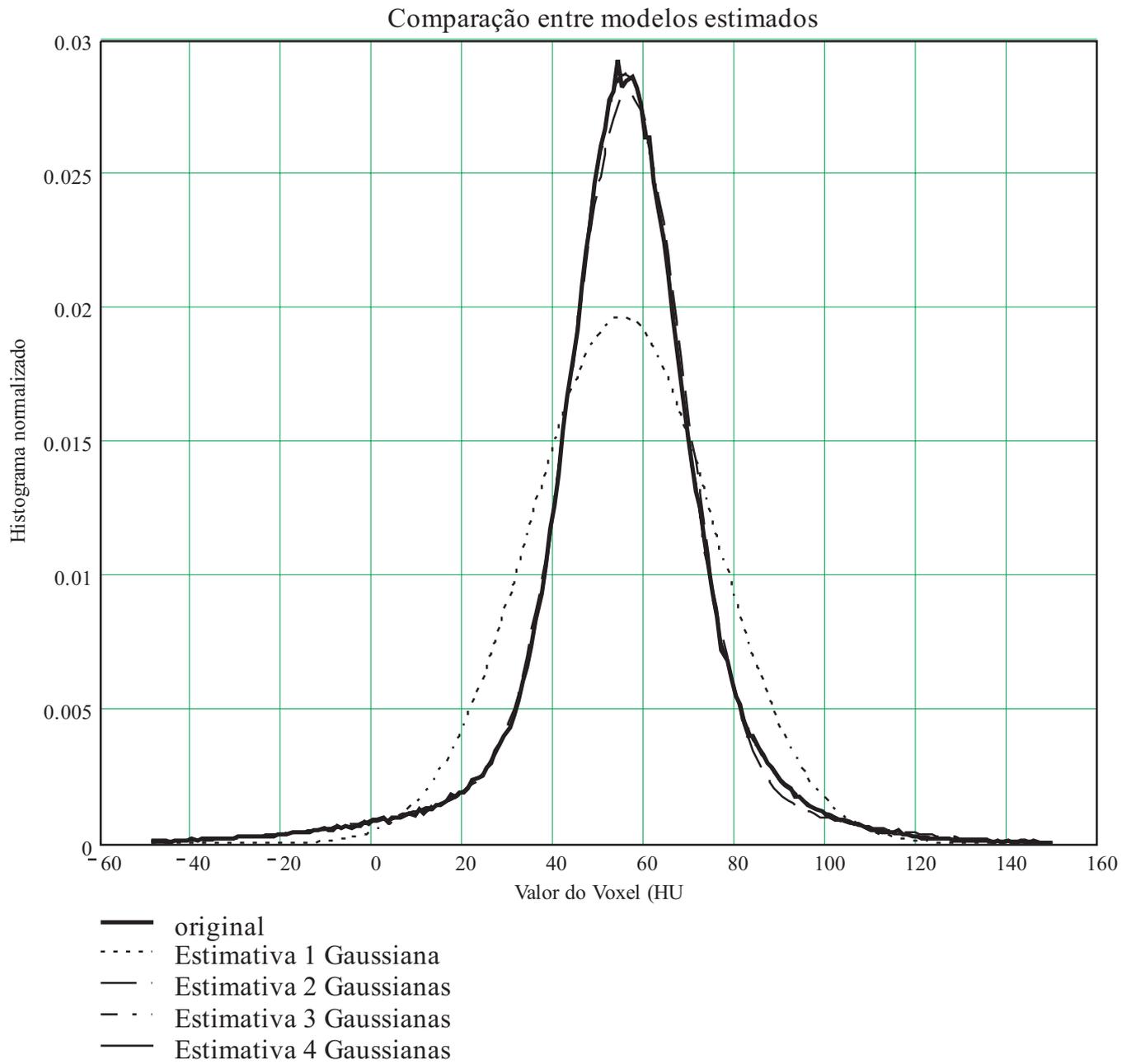


Figura 4.15: Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'B'). Ver também tabela 4.6.

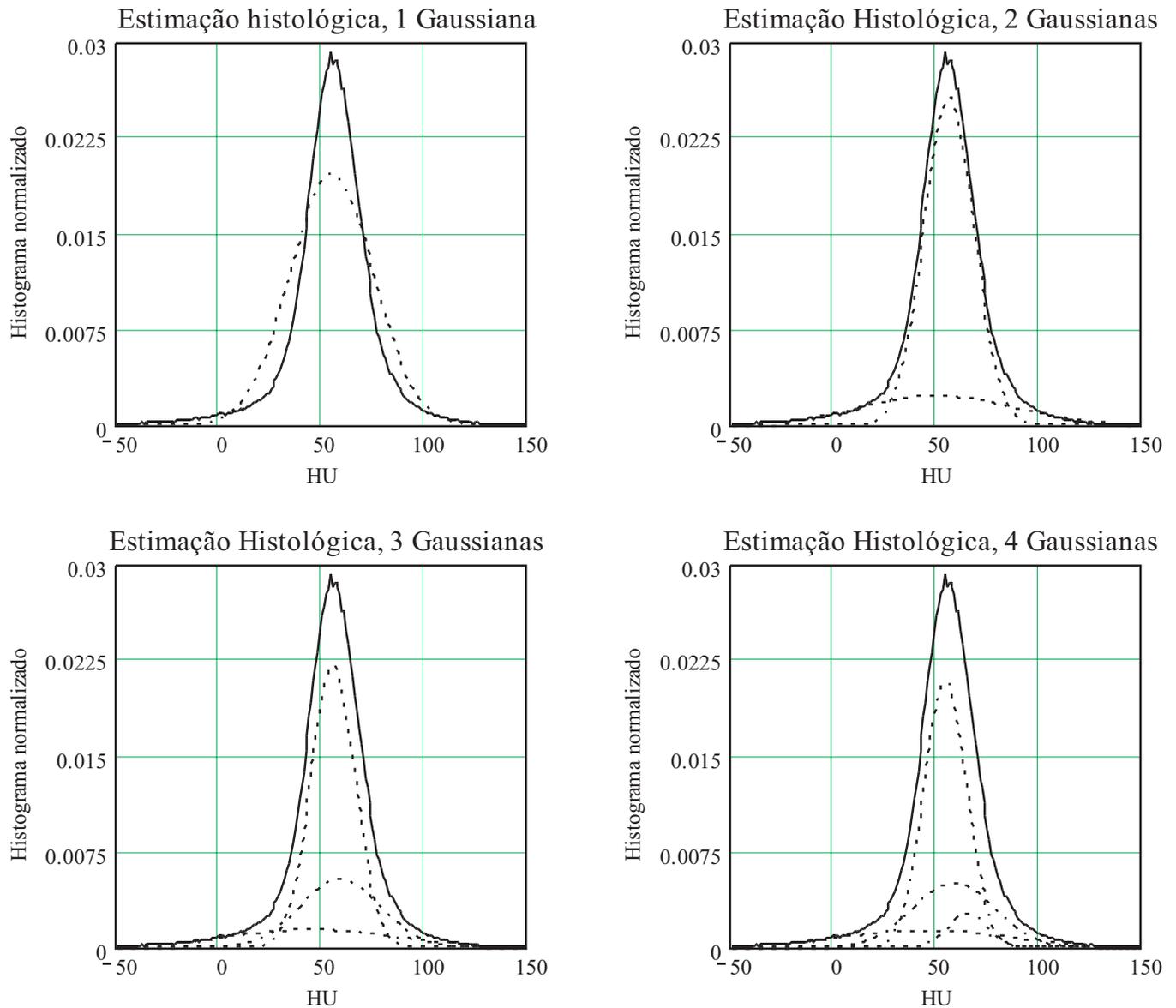


Figura 4.16: Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'B'). A linha sólida representa a função densidade de probabilidade para o dado original e as linhas tracejadas representam as componentes estimadas. veja também tabela 4.6 e figura 4.15.

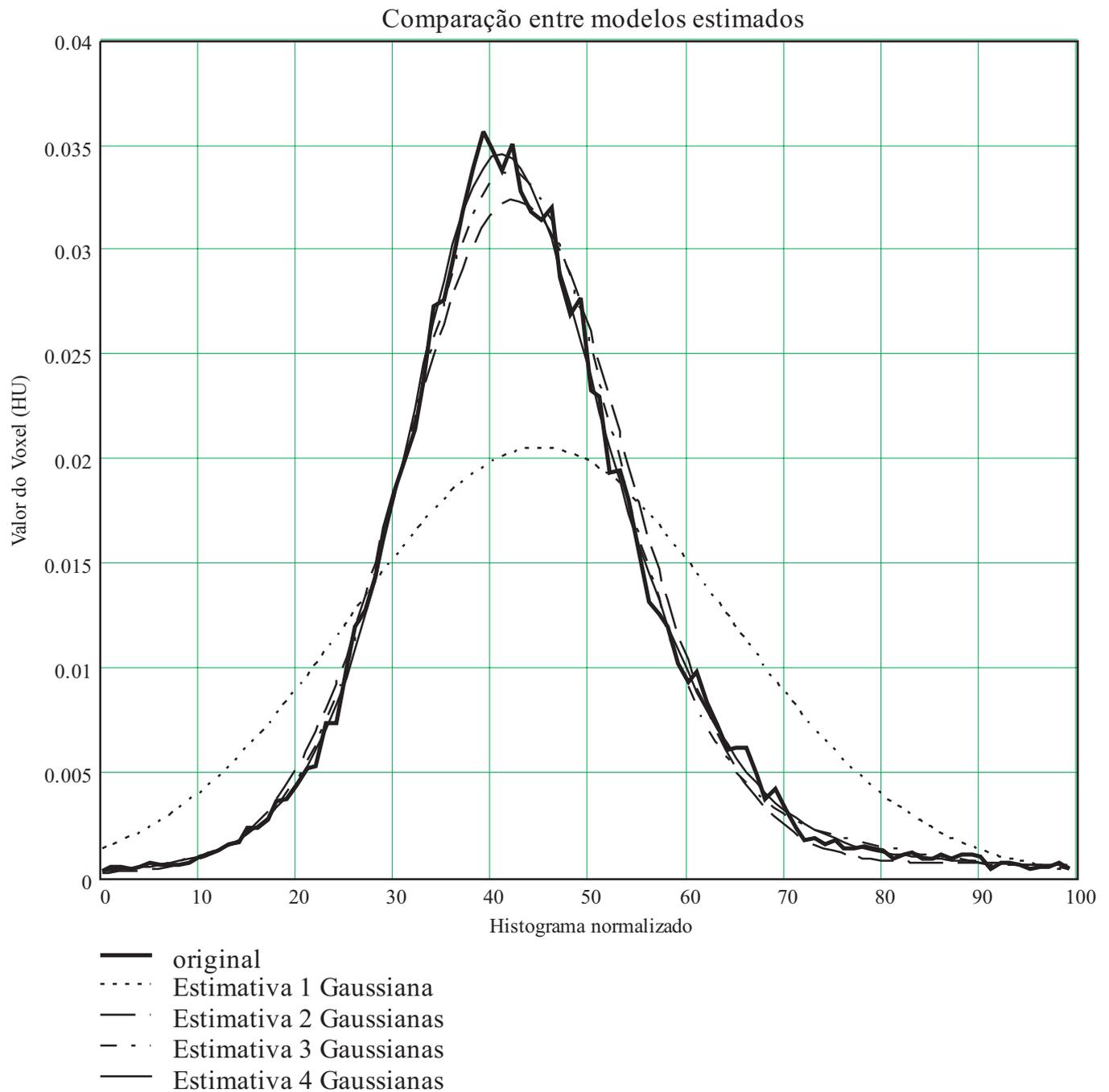


Figura 4.17: Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'C'). Ver também tabela 4.7.

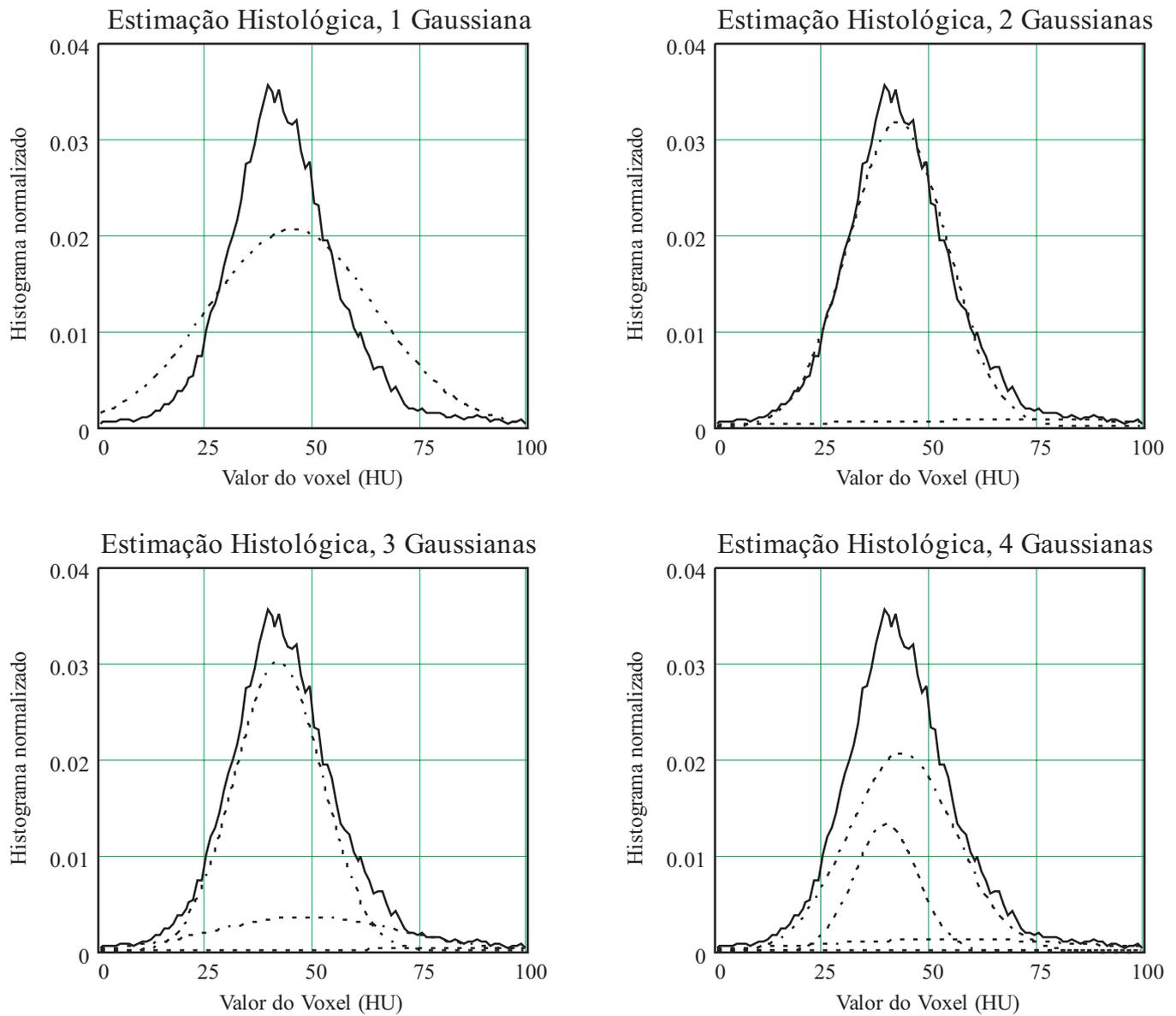


Figura 4.18: Resultados do algoritmo EM para modelos de mistura de Gaussianas usando dados reais (caso 'C'). A linha sólida representa a função densidade de probabilidade para o dado original e as linhas tracejadas representam as componentes estimadas. veja também tabela 4.7 e figura 4.17.

Parâmetros estimados, 1 núcleo Gaussiano	$\alpha_1 = 1.00, \mu_1 = 44.83, \sigma_1 = 19.38$
Parâmetros estimados, 2 núcleos Gaussianos	$\alpha_1 = 0.91, \mu_1 = 42.29, \sigma_1 = 11.47$ $\alpha_2 = 0.09, \mu_2 = 71.82, \sigma_2 = 46.60$
Parâmetros estimados, 3 núcleos Gaussianos	$\alpha_1 = 0.77, \mu_1 = 41.69, \sigma_1 = 10.17$ $\alpha_2 = 0.19, \mu_2 = 48.10, \sigma_2 = 21.54$ $\alpha_3 = 0.04, \mu_3 = 93.60, \sigma_3 = 54.77$
Parâmetros estimados, 4 núcleos Gaussianos	$\alpha_1 = 0.64, \mu_1 = 43.31, \sigma_1 = 12.35$ $\alpha_2 = 0.24, \mu_2 = 39.58, \sigma_2 = 7.30$ $\alpha_3 = 0.10, \mu_3 = 54.09, \sigma_3 = 31.13$ $\alpha_4 = 0.02, \mu_4 = 117.88, \sigma_4 = 54.61$

Tabela 4.7: Parâmetros do modelo de mistura de Gaussianas estimados para o tumor do caso 'C', manualmente segmentado. A região do tumor contém 24892 voxels, o valor médio para o tumor é 44.83 HU e o desvio padrão é 19.38 HU.

Capítulo 5

Conclusão e Trabalhos Futuros

Neste trabalho apresentamos um estudo e desenvolvimento de técnicas de processamento de imagens médicas aplicadas ao melhoramento da análise da massa tumoral primária, em pacientes com neuroblastoma avançado. Mais especificamente, estudamos técnicas para a segmentação e estimação da composição histológica da massa tumoral, a partir de exames tridimensionais de tomografia computadorizada, desde seu *rationale* até a implementação e análise destas técnicas.

Dissertamos sobre os aspectos da doença que são relevantes no projeto das técnicas de segmentação e estimação histológica, com a devida revisão da literatura científica pertinente (tanto em Medicina, como em processamento tridimensional de imagens). Seguiu-se a apresentação das técnicas implementadas, em seus fundamentos matemáticos, e uma revisão do estado da arte da pesquisa em diagnóstico do neuroblastoma auxiliado por computador. Finalmente, apresentamos os testes efetuados para um maior entendimento do problema, e dos algoritmos implementados, e discutimos a validade destes algoritmos em relação ao conjunto de dados disponíveis.

Pudemos constatar a importância social do desenvolvimento deste trabalho, e das propostas de trabalhos futuros que extendam o presente, ao revisarmos a literatura médica a respeito do neuroblastoma, e durante conversas com médicos especialistas do Instituto da Criança da Faculdade de Medicina da USP. O problema que

abordamos, o emprego de técnicas de processamento de imagens no diagnóstico do neuroblastoma, encontra-se pouco explorado, e novos desenvolvimentos serão bem recebidos pela comunidade médica.

Estando o problema pouco explorado, no melhor de nosso conhecimento, não nos surpreende que as técnicas aqui desenvolvidas não apresentem um desempenho superior, à semelhança do sucesso do processamento de imagens médicas em outros problemas, tais como análise de mamogramas ou angiografia digital tridimensional. Não obstante, em nosso trabalho coletamos elementos fundamentais para o entendimento do problema e desenvolvimento de soluções adequadas, através de nossa revisão do estado da arte e de nossos experimentos.

Podemos sugerir diversos trabalhos de extensão aos desenvolvimentos apresentados nesta dissertação, entre eles:

- Aperfeiçoamento da técnica de segmentação: como visto na seção 4.1, o desempenho da técnica de segmentação implementada ainda não é satisfatório. Isto decorre da natureza complexa da forma e da textura do neuroblastoma, visto em exame de CT. Trabalhos futuros devem incorporar novos elementos descritores da massa tumoral, tais como formato da massa ou posição relativa a outros órgãos.
- Aperfeiçoamento da técnica de estimação da composição histológica: desenvolvimento de metodologias para a seleção automática de valores iniciais para o algoritmo Expectation-maximization; seleção automática do número correto de núcleos Gaussianos.
- Construção de uma base de dados para pesquisa: o desenvolvimento da pesquisa em aplicações do processamento de imagens médicas para a análise de neuroblastomas seria beneficiado com a criação de uma base de imagens e dados patológicos, à imagem do que ocorre em alguns outros problemas de processamento de imagens, como em análise de mamogramas - os estudiosos

deste problema contam com a base de dados da Mammogram Image Analysis Society (MIAS).

- Uso de MRI: o imageamento por ressonância magnética pode ser mais adequado à investigação do neuroblastoma, em função de sua habilidade intrínseca de distinguir mais claramente tecidos moles, como aqueles presentes na cavidade abdominal.
- Desenvolvimento de técnicas semi-automáticas para auxílio ao diagnóstico: observamos que o desenvolvimento de um conjunto de programas que configure uma estação de diagnóstico, por computador, do neuroblastoma, seria benéfico à comunidade médica. Tal estação englobaria os diversos algoritmos de análise possivelmente desenvolvidos, e outras facilidades de exploração de imagens tomográficas, permitindo que o radiologista amplie sua compreensão do tumor em estudo. Poderia também estar associada a outros sistemas para exploração da massa tumoral, como sistemas de realidade virtual ou realidade aumentada, aumentando ainda mais o potencial do emprego do computador no estudo do neuroblastoma.

Bibliografia

- [1] Cohen M D, Bugaieski E M, Haliloglu M, Faught P, and Siddiqui A R. Visual presentation of the staging of pediatric solid tumors. *RadioGraphics*, 16(3):523–545, 1996.
- [2] Abramson S J. Adrenal neoplasm in children. *Radiologic Clinics of North America*, 35(6):1415–1453, 1997.
- [3] Mategrano V C, Petasnick J, Clark J, Bin A C, and Weinstein R. Attenuation values in computed tomography of the abdomen. *Radiology*, 125:135–140, October 1977.
- [4] Alexander F. Neuroblastoma. *Urologic Clinics of North America*, 27(3):383–392, 2000.
- [5] Cotterill S J, Pearson A D J, Pritchard J, Foot A B M, Roald B, Kohler J A, and Imeson J. Clinical prognostic factors in 1277 patients with neuroblastoma: results of The European Neuroblastoma Study Group ‘Survey’ 1982-1992. *European Journal of Cancer*, 36:901–908, 2000.
- [6] Bousvaros A, Kirks D R, and Grossman H. Imaging of neuroblastoma: an overview. *Pediatric Radiology*, 16:89–106, 1986.
- [7] Foglia R P, Fonkalsrud E W, Feig S A, and Moss T J. Accuracy of diagnostic imaging as determined by delayed operative intervention for advanced neuroblastoma. *Journal of Pediatric Surgery*, 24(7):708–711, 1989.

- [8] Bankman I N, editor. *Handbook of Medical Imaging: Processing and Analysis*. Academic Press, London, UK, 2000.
- [9] Sonka M and Fitzpatrick J M, editors. *Handbook of Medical Imaging, Volume 2: Medical Image Processing and Analysis*. SPIE Press, Bellingham, WA, 2000.
- [10] Duncan J S and Ayache N. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):85–106, January 2000.
- [11] Wheatley J M, Rosenfield N S, Heller G, Feldstein D, and LaQuaglia M P. Validation of a technique of computer-aided tumor volume determination. *Journal of Surgical Research*, 59(6):621–626, 1995.
- [12] Hopper K D, Singapuri K, and Finkel A. Body CT and oncologic imaging. *Radiology*, 215(1):27–40, 2000.
- [13] Meza M P, Benson M, and Slovis T L. Imaging of mediastinal masses in children. *Radiologic Clinics of North America*, 31(3):583–604, 1993.
- [14] Castleberry R P. Neuroblastoma. *European Journal of Cancer*, 33(9):1430–1438, 1997.
- [15] Stark D D, Moss A A, Brasch R C, deLorimier A A, Albin A R, London D A, and Gooding C A. Neuroblastoma: Diagnostic imaging and staging. *Radiology*, 148:101–105, July 1983.
- [16] Kirks D R, Merten D F, Grossman H, and Bowie J D. Diagnostic imaging of pediatric abdominal masses: an overview. *Radiologic Clinics of North America*, 19(3):527–545, 1981.
- [17] Boechat M I, Ortega J, Hoffman A D, Cleveland R H, Kangarloo H, and Gilsanz V. Computed tomography in Stage III neuroblastoma. *American Journal of Radiology*, 145:1456–1283, December 1985.

- [18] Corbett R, Olliff J, Fairley N, Moyes J, Husband J, Pinkerton R, Carter R, Treleaven J, McElwain T, and Meller S. A prospective comparison between magnetic resonance imaging, meta-iodobenzylguanidine scintigraphy and marrow histology/cytology in neuroblastoma. *European Journal of Cancer*, 27(12):1560–1564, 1991.
- [19] Fletcher B D, Kopywoda S Y, Strandjord S E, Nelson A D, and Pickering S P. Abdominal neuroblastoma: Magnetic resonance imaging and tissue characterization. *Radiology*, 155(3):699–703, 1985.
- [20] Sofka C M, Semelka R C, Kelekis N L, Worawattanakul S, Chung C J, Gold S, and Fordham L A. Magnetic resonance imaging of neuroblastoma using current techniques. *Magnetic Resonance Imaging*, 17(2):193–198, 1999.
- [21] Chezmar J L, Robbins S M, Nelson R C, Steinberg H V, Torres W E, and Bernardino M E. Adrenal masses: Characterization with T1-weighted MR imaging. *Radiology*, 166(2):357–359, 1988.
- [22] Kornreich L, Horev G, Kaplinsky N Z, and Grunebaum M. Neuroblastoma: evaluation with contrast enhanced MR imaging. *Pediatric Radiology*, 21:566–569, 1991.
- [23] Robb R A. X-ray computed tomography: an engineering synthesis of multidisciplinary principles. *CRC Critical Reviews in Biomedical Engineering*, 7(4):265–333, 1982.
- [24] Herman G T. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. Academic Press, Orlando, FL, 1980.
- [25] Goodenough D J. Tomographic imaging. In Beutel J, Kundel H L, and Van Metter R L, editors, *Handbook of Medical Imaging, Volume 1: Physics and Psychophysics*, chapter 8, pages 511–554. SPIE Press, Bellingham, WA, 2000.

- [26] Fullerton G D. Fundamentals of ct tissue characterization. In Fullerton G D and Zagzebski J A, editors, *Medical Physics of CT and Ultrasound: Tissue Imaging and Characterization*, pages 125–162. American Association of Physicists in Medicine, New york, NY, 1980.
- [27] Phelps M E, Hoffman E J, and Ter-Pogossian M M. Attenuation coefficients of various body tissues, fluids and lesions at photon energies of 18 to 136 keV. *Radiology*, 117:573–583, December 1975.
- [28] Rao P S and Gregg E C. Attenuation of monoenergetic gamma rays in tissues. *American Journal of Roentgenology*, 123(3):631–637, 1975.
- [29] Wilson C R. Quantitative computer tomography. In Fullerton G D and Zagzebski J A, editors, *Medical Physics of CT and Ultrasound: Tissue Imaging and Characterization*, pages 163–175. American Association of Physicists in Medicine, New York, NY, 1980.
- [30] Brooks R A. A quantitative theory of the Hounsfield unit and its application to dual energy scanning. *Journal of Computer Assisted Tomography*, 1(4):487–493, 1977.
- [31] Alter A J. Computerized tomography: A clinical perspective. In Fullerton G D and Zagzebski J A, editors, *Medical Physics of CT and Ultrasound: Tissue Imaging and Characterization*, pages 125–162. American Association of Physicists in Medicine, New York, NY, 1980.
- [32] Williams G, Bydder G M, and Kreel L. The validity and use of computed tomography attenuation values. *British Medical Bulletin*, 36(3):279–287, 1980.
- [33] Duerinckx A J and Macovski A. Information and artifact in computed tomography image statistics. *Medical Physics*, 7(2):127–134, March-April 1980.
- [34] Pullan B R, Fawcitt R A, and Isherwood I. Tissue characterization by an analysis of the distribution of attenuation values in computed tomography scans: A

- preliminary report. *Journal of Computer Assisted Tomography*, 2(1):49–54, 1978.
- [35] Kramer R A, Yoshikawa B M, Scheibe P O, and Janetos G P. Statistical profiles in computed tomography. *Radiology*, 125:145–147, October 1977.
- [36] Latchaw R E, Gold L H A, Moore J S, and Payne J T. The nonspecificity of absorption coefficients in the differentiation of solid tumors and cystic lesions. *Radiology*, 125:141–144, October 1977.
- [37] Udupa J K and Samarasekera S. Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation. *Graphical Models and Image Processing*, 58(3):246–261, May 1996.
- [38] Bishop C M. *Neural Networks for Pattern Recognition*. Oxford Press, New York, NY, 1995.
- [39] Jain A K, Duin R P W, and Mao J. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, January 2000.
- [40] Tanner M A. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer-Verlag, New York, NY, 3rd edition, 1996.
- [41] Dawant B M and Zijdenbos A P. Image segmentation. In Sonka M and Fitzpatrick J M, editors, *Handbook of Medical Imaging, Volume 2: Medical Image Processing and Analysis*, chapter 2, pages 71–127. SPIE Press, Bellingham, WA, 2000.
- [42] Copsey K and Webb A. Bayesian approach to mixture models for discrimination. In Ferri F J, Iñesta J M, Amin A, and Pudil P, editors, *Advances in Pattern Recognition, Joint IAPR International Workshops SSPR 2000 and*

- SPR 2000, [8th International Workshop on Structural and Syntactic Pattern Recognition, 3rd International Workshop on Statistical Techniques in Pattern Recognition]*, pages 491–500, Alicante, Spain, August 30 - September 1 2000. Springer. Lecture Notes in Computer Science, Vol. 1876.
- [43] Richardson S and Green P J. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society B*, 59(4):731–792, 1997.
- [44] Meng X L and van Dyk D. The em algorithm - and old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society*, 59(3):511–567, 1997.
- [45] Rangayyan R M, Ferrari R J, Desautels J E L, and Frère A F. Directional analysis of images with gabor wavelets. In *Proceedings of SIBGRAPI 200: XIII Brazilian Symposium on Computer Graphics and Image Processing*, pages 170–177, Gramado, Rio Grande do Sul, Brazil, 17-20 October 2000. IEEE Computer Society Press.
- [46] Ferrari R J, Rangayyan R M, Desautels J E L, and Frère A F. Automatic identification of the skin-air boundary, pectoral muscle, and fibro-glandular disc in mammograms. *Submitted to the IEEE Transactions on Medical Imaging*, August 2000.
- [47] Ferrari R J, Rangayyan R M, Desautels J E L, and Frère A F. Segmentation of mammograms: Identification of the skin-air boundary, pectoral muscle, and fibro-glandular disc. In *Proceedings of the 5th International Workshop on Digital Mammography*, page in press, Toronto, ON, June 2000.
- [48] Goldszal A F and Pham D L. Volumetric segmentation. In Bankman I N, editor, *Handbook of Medical Imaging: Processing and Analysis*, chapter 12, pages 185–194. Academic Press, London, UK, 2000.

- [49] Laidlaw D H, Fleischer K W, and Barr A H. Partial volume segmentation with voxel histograms. In Bankman I N, editor, *Handbook of Medical Imaging Processing and Analysis*, chapter 13, pages 185–194. Academic Press, London, UK, 2000.