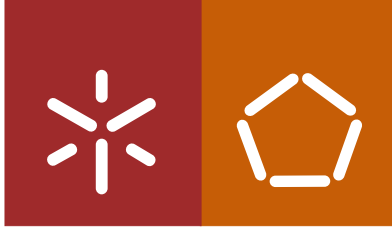


Universidade do Minho
Escola de Engenharia

Pedro Miguel Pereira Vieira

**Segmentação de imagens da Cápsula
Endoscópica com aplicações na deteção
de tumores e na reconstrução 3-D da
mucosa intestinal**



Universidade do Minho

Escola de Engenharia

Pedro Miguel Pereira Vieira

**Segmentação de imagens da Cápsula
Endoscópica com aplicações na deteção
de tumores e na reconstrução 3-D da
mucosa intestinal**

Dissertação de Mestrado
Mestrado Integrado em Engenharia Biomédica
Ramo Eletrónica Médica

Trabalho efetuado sob a orientação do
Professor Doutor Carlos Lima

outubro de 2013

DECLARAÇÃO

Nome: Pedro Miguel Pereira Vieira

Endereço eletrónico: a55551@alunos.uminho.pt

Número do cartão de cidadão: 13781465

Título da Dissertação: Segmentação de imagens da Cápsula Endoscópica com aplicações na deteção de tumores e na reconstrução 3-D da mucosa intestinal

Orientador: Professor Doutor Carlos Manuel Gregório Santos Lima

Ano de conclusão: 2013

Designação do Mestrado: Mestrado Integrado em Engenharia Biomédica – Ramo Eletrónica Médica

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, ___ / ___ / _____

Assinatura: _____

Agradecimentos

Os meus primeiros agradecimentos vão para o Professor Carlos Lima, o meu orientador neste trabalho, primeiro pela oportunidade que me deu para trabalhar com ele, mas também por todos os conselhos dados, conhecimentos que transmitiu e a confiança que depositou em mim.

Aos orientadores e pessoas responsáveis pelo meu trabalho na *KULeuven*, pela possibilidade de ter trabalhado durante um semestre num dos maiores centros de investigação em imagem médica do mundo.

Também as pessoas que me acompanharam durante a experiência de Erasmus merecem um agradecimento especial, devido a todas as aventuras, viagens e passeios que vivemos, além do convívio do dia-a-dia durante aqueles seis meses especiais.

A todos os meus amigos de Braga, que me acompanham há tanto tempo, agradeço toda a companhia e insistência, tal como toda a compreensão pelas minhas ausências constantes.

À Cristiana e à Rita, grandes amigas e futuras médicas, agradeço o tempo despendido pela leitura, análise e correção dos textos clínicos presentes nesta dissertação.

A todos os colegas que me acompanharam nestes últimos meses no laboratório, Carla, João, Luís, Sara e Susana, o meu muito obrigado por toda a companhia e conselhos que me deram, tal como a alegria que havia para ir trabalhar todos os dias.

Um obrigado muito especial à Joana, ao José e ao Sérgio, porque além de terem sido as pessoas que desde o quarto ano me acompanham todos os dias, estiveram sempre prontos a ajudar, e que tiveram a árdua tarefa de me aturar.

E porque é impossível agradecer a toda a gente individualmente, um grande obrigado a todos os grandes amigos que fiz no curso, principalmente à segunda família que ganhei, e que espero nunca mais perder. Estes não tinham sido a mesma coisa sem vocês.

Por último, um agradecimento especial à minha família, aos meus pais, ao meu irmão e principalmente à minha irmã, pela compreensão, ajuda e carinho que me deram durante toda esta etapa.

Resumo

A cápsula endoscópica é um dispositivo que contém uma pequena câmara e que é usado para capturar imagens do trato gastrointestinal de forma não invasiva. Como o resultado é um vídeo bastante longo, a taxa de detecção de patologias é baixa, logo existe a necessidade de melhorar a qualidade dos vídeos ao adicionar informação às *frames* ou então ao fazer uma detecção automática de patologias ou estruturas de relevo.

Os objetivos do trabalho descrito nesta dissertação eram especialmente dois. O primeiro consistia na reconstrução tridimensional de *frames* individuais da cápsula endoscópica, tal como encontrar uma forma de relacionar imagens consecutivas para assim poder, no futuro, juntar informações da terceira dimensão ao longo do vídeo. O segundo objetivo é a detecção automática de tecido tumoral em cada *frame* retirada da cápsula endoscópica.

Para ambos os objetivos foi usado o algoritmo de segmentação *Expectation-Maximization*, seguindo uma abordagem *Maximum a Posteriori*, usando o espaço de cor Lab para retirar o efeito da luminosidade. No primeiro caso, é realizada a segmentação com diferentes números de regiões, para aferir o número ótimo em cada *frame*. Ao calcular os coeficientes de correlação entre vetores de *features* de cor e textura das várias regiões, é possível detetar regiões correspondentes entre as várias *frames*. No final a reconstrução das *frames* é realizada com o uso do método *Shape from Shading*. Na detecção automática de tumores, depois de dividir a imagem em duas regiões diferentes, o método proposto envolve o cálculo de descritores estatísticos dos histogramas em cada região, e a posterior classificação usando diferentes classificadores supervisionados.

Chegou-se à conclusão que a segmentação para encontrar os diferentes tecidos, apesar de apresentar resultados quando analisada individualmente, tem de ser refinada para que as regiões em *frames* diferentes tenham uma maior semelhança. A reconstrução individual apresenta bons resultados, aumentando a informação em cada *frame*, aumentando assim a capacidade de diagnóstico. O método de detecção automática de tumores proposto apresenta resultados bastante promissores quando comparados com estado da arte, tanto o método de segmentação, tal como o conjunto de *features* usado. A precisão do sistema concebido teve uma melhoria de pelo menos 6% em relação a métodos do estado da arte implementados para a mesma base de dados.

Abstract

Wireless Capsule Endoscopy is a non-invasive device that contains a small camera used for capturing images from the gastrointestinal tract. The result of this exam is very long video, which leads to a low detection rate of several pathologies. This fact is the reason why it's needed for the increase of quality of the videos, either adding new information to every frame, either implementing algorithms for automatic detection of pathologies or important structures.

The objectives of the work described in this text were specially two. The first one consisted in the construction of an algorithm for the 3-D reconstruction of individual frames and also to find similar regions in consecutive regions, so in the future this information could lead to a video of combined 3-D images. The second objective is the automatic detection of tumor tissue in capsule endoscopy frames.

The segmentation in both objectives used the Expectation-Maximization algorithm, with a Maximum a Posteriori approach, using the Lab color space so the light effect could be removed. In the first case, the segmentation is performed with a different number of regions, so it could be possible to find the best number of regions in each frame. Correlation coefficients were then calculated between vectors containing color and texture features of the different regions, so corresponding regions could be found. Reconstruction of frames was performed using the Shape from Shading method. In the automatic detection of tumors, after dividing the image in two different regions, the proposed method includes the computation of statistical features from the histograms of each region, and the classification with supervised classifiers.

As conclusion, it is possible to say that the segmentation to find different kinds of tissues need to be refined so the correct corresponding regions can be found. The individual reconstruction presents good results, increasing the visual information in each frame, increasing in that way the diagnosis capability. The method for tumor detection presents promising results when compared to state of the art methods. The final system precision had an increase of at least 6% in relation with state of the art algorithm implemented with the same database.

Índice

| | |
|--|-------------|
| Agradecimentos | iii |
| Resumo | v |
| Abstract | vii |
| Índice de Figuras | xiii |
| Índice de Tabelas | xv |
| Lista de Acrónimos..... | xvii |
| Capítulo 1 Introdução | 1 |
| 1.1. Motivação..... | 1 |
| 1.2. Objetivos | 2 |
| 1.3. Organização da dissertação | 4 |
| Capítulo 2 Estado da Arte | 5 |
| 2.1. Detecção de tumores | 6 |
| 2.2. Reconstrução tridimensional..... | 6 |
| 2.2.1. Endoscopia | 7 |
| 2.2.2. Cápsula endoscópica..... | 8 |
| Capítulo 3 Contexto clínico | 11 |
| 3.1. Sistema gastrointestinal..... | 11 |
| 3.1.1. Intestino Delgado..... | 12 |
| 3.2. Técnicas de imagiologia do trato gastrointestinal | 15 |
| 3.2.1. Endoscopia | 15 |
| 3.2.2. Cápsula Endoscópica..... | 18 |
| Capítulo 4 Segmentação | 23 |
| 4.1. Modelo estatístico..... | 24 |

| | |
|---|-----------|
| 4.2. Regra de Bayes | 25 |
| 4.3. A Estimativa de Máxima Verosimilhança..... | 26 |
| 4.4. O algoritmo EM | 28 |
| 4.5. Inicialização..... | 29 |
| 4.6. Critério de Informação Bayesiano..... | 30 |
| Capítulo 5 Reconstrução 3-D..... | 33 |
| 5.1. Iluminação de superfícies | 33 |
| 5.2. Shape from Shading..... | 35 |
| Capítulo 6 Detecção de tumores | 45 |
| 6.1. Correção luminosa | 46 |
| 6.1.1. Espaços de cor..... | 47 |
| 6.1.2. Remoção da luminosidade..... | 49 |
| 6.2. Extração de <i>features</i> | 50 |
| 6.2.1. Informação de cor | 51 |
| 6.2.2. Informação de textura..... | 53 |
| 6.3. Relação entre regiões | 59 |
| Capítulo 7 Implementação | 61 |
| 7.1. Base de dados..... | 61 |
| 7.2. Software..... | 62 |
| 7.3. Implementação da segmentação | 62 |
| 7.3.1. Implementação do Algoritmo EM | 62 |
| 7.3.2. Pós-processamento | 64 |
| 7.4. Implementação da reconstrução 3-D..... | 65 |
| 7.5. Implementação da detecção de tumores | 66 |
| Capítulo 8 Resultados | 69 |
| 8.1. Segmentação | 69 |
| 8.1.1. Inicialização..... | 69 |
| 8.1.2. Segmentação de tecidos..... | 71 |
| 8.1.3. Relação entre regiões | 77 |

| | |
|---|------------|
| 8.1.4. Tempo de execução..... | 80 |
| 8.2. Reconstrução tridimensional | 81 |
| 8.2.1. Tempo de execução..... | 85 |
| 8.3. Detecção automática de tumores | 85 |
| 8.3.1. Resultados com segmentação..... | 86 |
| 8.3.2. Resultados sem segmentação..... | 91 |
| 8.3.3. Comparação com estado da arte | 95 |
| 8.3.4. Tempo de execução..... | 97 |
| Capítulo 9 Conclusão e trabalho futuro | 101 |
| 9.1. Conclusões..... | 101 |
| 9.2. Contribuições originais..... | 103 |
| 9.3. Trabalho futuro..... | 104 |
| Bibliografia | 105 |

Índice de Figuras

| | |
|--|----|
| Figura 3.1. Esquema do intestino delgado e suas regiões. Adaptado de [3]..... | 12 |
| Figura 3.2. Exemplo de uma úlcera duodenal (a – Adaptado de (Fusco 2013a)), um pólipó intestinal (b – Adaptado de (Umeoka et al. n.d.)) e uma hemorragia intestinal (c – Adaptado de (Clancy 2008)). | 14 |
| Figura 3.3. Réplica do <i>Lichtleiter</i> original. Adaptado de (Natalin & Landman 2009). | 16 |
| Figura 3.4. Endoscópio semi-flexível de Wolf-Schindler. Adaptado de (Anthony J. DiMarino & Benjamin 2002). | 17 |
| Figura 3.5. Exemplo de um endoscópio atual (a – Adaptado de (Olympus America 2011)), e alguns exemplos de ferramentas que podem ser colocadas na ponta do endoscópio: pinça de biópsia (b), escova citológica (c) e base para fórceps (d) – Adaptado de (Cotton et al. 2009). .. | 18 |
| Figura 3.6. Cápsula Endoscópica M2A™ (em cima) e os seus constituintes (em baixo). Adaptado de (Qureshi 2004). | 19 |
| Figura 4.1. <i>Frame</i> da cápsula endoscópica e histogramas de cana um dos canais do espaço de cor Lab..... | 25 |
| Figura 4.2. Exemplo do funcionamento do algoritmo k-means. a) pontos verdes é o grupo de dados que se pretende dividir, b) inicialização do algoritmo com as cruces como centros, do c) ao h), iterações sucessivas do algoritmo com solução final em i). Adaptado de (Alpaydin 2010). .. | 30 |
| Figura 5.1. Definição de ângulo incidente (i) e ângulo de reflexão (e). | 34 |
| Figura 5.2. Um modelo da projeção de uma imagem usando uma cápsula endoscópica..... | 37 |
| Figura 5.3. Representação dos ângulos τ (centro) e σ (direita) tendo em conta uma fonte luminosa arbitrária no espaço (esquerda)..... | 38 |
| Figura 6.1. Exemplo de <i>frame</i> que tem uma zona de hiperintensidade no seu centro..... | 46 |
| Figura 6.2. Esquema que revela como os atributos de cada um dos espaços de cor estão relacionados entre si. Adaptado de (Anon 2010). | 49 |
| Figura 6.3. <i>Frame</i> da cápsula endoscópica que apresenta hiperintensidade (esquerda). Mesma imagem com canal L menos 20 unidades (centro) e canal de luminância mais 20 unidades (direita). | 50 |
| Figura 6.4. Histogramas dos canais de cor <i>a</i> e <i>b</i> de uma região de uma <i>frame</i> da cápsula endoscópica. | 51 |

| | |
|--|----|
| Figura 6.5. Histogramas dos canais de cor <i>a</i> e <i>b</i> de uma região de uma <i>frame</i> da cápsula endoscópica. | 52 |
| Figura 6.6. Exemplo de duas decomposições da DWT de uma imagem. | 57 |
| Figura 6.7. Divisão no domínio das frequências da Transformada Discreta <i>Curvelet</i> | 58 |
| Figura 7.1. Segmentação de imagem da CE (a) em duas regiões, usando a matriz de covariâncias completa (b) e aproximando a uma matriz diagonal (c)..... | 63 |
| Figura 7.2. Algoritmo de pós-processamento do resultado da segmentação. | 64 |
| Figura 8.1. Exemplo de segmentação de <i>frame</i> da CE (a) usando o espaço de cor RGB (b), o espaço de cor Lab (c) e o espaço de cor Lab sem o canal da luminosidade (d). | 71 |
| Figura 8.2. Exemplo de segmentação de <i>frame</i> da CE (a) usando o espaço de cor RGB (b), o espaço de cor Lab (c) e o espaço de cor Lab sem o canal da luminosidade (d). | 73 |
| Figura 8.3. Segmentação da <i>frame</i> da CE apresentada na Figura 6.1 em diferentes números de regiões: duas a sete (a-f). | 73 |
| Figura 8.4. Valor do BIC das segmentações com vários números de classes representadas na Figura 8.3. | 74 |
| Figura 8.5. Exemplo de <i>frame</i> retirada de exame da CE. | 75 |
| Figura 8.6. Segmentação da <i>frame</i> da CE apresentada na Figura 8.5 em diferentes números de regiões: duas a sete (a-f). | 76 |
| Figura 8.7. Valor do BIC das segmentações com vários números de classes representadas na Figura 8.6. | 76 |
| Figura 8.8. Três <i>frames</i> consecutivas de um exame da CE e as respectivas segmentações. | 78 |
| Figura 8.9. Resultado da reconstrução tridimensional da <i>frame</i> da Figura 6.1, da iteração número 1 até a iteração número 7 (a-g) e iteração número 30 (h). | 82 |
| Figura 8.10. Comparação de reconstrução tridimensional usando apenas a imagem original (a) ou o algoritmo SfS (b). | 83 |
| Figura 8.11. Exemplo de reconstrução usando o SfS de três <i>frames</i> diferentes - à esquerda as <i>frames</i> da CE (a1, b1 e c1) e à direita as respectivas reconstruções (a2, b2 e c2). | 84 |

Índice de Tabelas

| | |
|--|----|
| Tabela 5.1. Coordenadas de cada ponto do esquema da Figura 5.2. | 37 |
| Tabela 7.1. Características da câmara e lente da Pillcam SB. | 65 |
| Tabela 8.1. Média, mediana, máximo e mínimo dos valores do desvio padrão de vários parâmetros das distribuições das 496 imagens usadas. | 70 |
| Tabela 8.2. Valor do grau de confiança da correspondência entre as várias regiões da <i>frame 1</i> e da <i>frame 2</i> | 79 |
| Tabela 8.3. Valor do grau de confiança da correspondência entre as várias regiões da <i>frame 2</i> e da <i>frame 3</i> | 79 |
| Tabela 8.4. Resultados da correspondência de regiões entre <i>frames</i> | 80 |
| Tabela 8.5. Medidas estatísticas do tempo (em segundos) da execução do algoritmo de segmentação nas várias imagens. | 81 |
| Tabela 8.6. Medidas estatísticas do tempo (em segundos) da execução do algoritmo de segmentação nas várias imagens. | 85 |
| Tabela 8.7. Organização das <i>features</i> de cor utilizadas neste trabalho. | 86 |
| Tabela 8.8. Resultados usando o classificador MLP e correlação entre <i>features</i> de diferentes regiões retiradas dos histogramas dos canais <i>a</i> e <i>b</i> do Lab. | 87 |
| Tabela 8.9. Resultados usando o classificador SVM e correlação entre <i>features</i> de diferentes regiões retiradas dos histogramas dos canais <i>a</i> e <i>b</i> do Lab. | 87 |
| Tabela 8.10. Resultados da classificação com os classificadores MLP e SVM usando o set C sem a <i>feature</i> sem significância. | 87 |
| Tabela 8.11. Resultados usando o classificador MLP e correlação entre <i>features</i> de textura retiradas de diferentes regiões dos canais <i>a</i> e <i>b</i> do Lab. | 88 |
| Tabela 8.12. Resultados usando o classificador SVM e correlação entre <i>features</i> de textura retiradas de diferentes regiões dos canais <i>a</i> e <i>b</i> do Lab. | 88 |
| Tabela 8.13. Resultados usando correlação de <i>features</i> de textura entre as duas regiões, sem a <i>feature</i> com maior valor-p acima do nível de significância. | 89 |
| Tabela 8.14. Resultados usando correlação de <i>features</i> de textura da DCT entre as duas regiões, sem as duas <i>features</i> com maior valor-p acima do nível de significância. | 90 |

| | |
|--|----|
| Tabela 8.15. Resultados usando o classificador MLP e correlação entre <i>features</i> de textura retiradas de diferentes regiões dos canais do espaço HSV..... | 90 |
| Tabela 8.16. Resultados usando o classificador SVM e correlação entre <i>features</i> de textura retiradas de diferentes regiões dos canais do espaço HSV..... | 91 |
| Tabela 8.17. Resultados usando o classificador MLP e <i>features</i> retiradas dos histogramas dos canais <i>a</i> e <i>b</i> do Lab em toda a imagem..... | 92 |
| Tabela 8.18. Resultados usando o classificador SVM e <i>features</i> retiradas dos histogramas dos canais <i>a</i> e <i>b</i> do Lab em toda a imagem..... | 92 |
| Tabela 8.19. Resultados usando o classificador MLP e <i>features</i> com significância retiradas dos histogramas dos canais <i>a</i> e <i>b</i> do Lab em toda a imagem. | 93 |
| Tabela 8.20. Resultados usando o classificador SVM e <i>features</i> com significância retiradas dos histogramas dos canais <i>a</i> e <i>b</i> do Lab em toda a imagem. | 93 |
| Tabela 8.21. Resultados usando o classificador MLP e <i>features</i> de textura retiradas dos canais <i>a</i> e <i>b</i> do Lab em toda a imagem. | 94 |
| Tabela 8.22. Resultados usando o classificador SVM e <i>features</i> de textura retiradas dos canais <i>a</i> e <i>b</i> do Lab em toda a imagem. | 94 |
| Tabela 8.23. Resultados usando o classificador MLP e <i>features</i> de textura retiradas do espaço HSV em toda a imagem..... | 95 |
| Tabela 8.24. Resultados usando o classificador SVM e <i>features</i> de textura retiradas do espaço HSV em toda a imagem..... | 95 |
| Tabela 8.25. Comparação entre <i>features</i> propostas neste trabalho e estado da arte. | 96 |
| Tabela 8.26. Medidas estatísticas do tempo (em segundos) da execução do algoritmo de segmentação com vários parâmetros nas várias imagens..... | 98 |
| Tabela 8.27. Tempo de processamento de um exame completo e precisão de cada uma das versões do algoritmo testadas..... | 98 |

Lista de Acrónimos

- ASIC – *Application-Specific Integrated Circuit*
- BIC – *Bayesian Information Criterion*
- CCD – *Charged-Coupled Device*
- CMOS – *Complementary Metal Oxide Semiconductor*
- CE – Cápsula Endoscópica
- CTC – Colonoscopia por Tomografia Computadorizada
- DCT – *Discrete Curvelet Transform*
- DWT – *Discrete Wavelet Transform*
- EM – *Expectation-Maximization*
- FDA – *Food and Drug Administration*
- fdp – Função densidade de probabilidade
- HIA – Hemorragia Intestinal Alta
- HIB – Hemorragia Intestinal Baixa
- ICP – *Iterative Closest Point*
- LED – *Light Emitting Diode*
- MAP – *Maximum a Posteriori*
- MLP – *Multilayer Perceptron*
- MRF – *Markov Random Fields*
- RBF – *Radial Basis Function*
- RM – Ressonância Magnética
- ROI – *Regions of Interest*
- SfM – *Shape from Motion*
- SfS – *Shape from Shading*
- SIFT – *Scale-invariant feature transform*
- SNS – Sistema Nacional de Saúde
- SVM – *Support Vector Machine*

Lista de Acrónimos

TAC – Tomografía Axial Computarizada

TF – Transformada de Fourier

TGI – Trato gastrointestinal

Capítulo 1 Introdução

Este capítulo apresenta as principais motivações e objetivos que guiaram o trabalho desenvolvido ao longo desta dissertação, apresentando-se ainda um breve resumo e explicação da organização da dissertação.

1.1. Motivação

Atualmente são ainda poucos os sistemas que permitem fazer a visualização direta do sistema gastrointestinal, principalmente do intestino delgado. Isto deve-se à dificuldade de se conseguir chegar a este órgão de uma maneira minimamente invasiva. Um dos poucos métodos existentes é a cápsula endoscópica (CE), um dispositivo que consiste numa microcâmara, ingerido pelo paciente, e que tira continuamente fotografias durante mais de 8 horas ao trato gastrointestinal (TGI). Existem contudo algumas limitações de diagnóstico inerentes à análise destes exames, já que num estudo de 2012 concluiu-se que mesmo gastroenterologistas experientes na análise de vídeos de CE apresentam uma deteção de patologias inferior a 50% (Zheng et al. 2012). Para melhorar estes resultados, duas abordagens podem ser seguidas: a construção de métodos de deteção automática de patologias recorrentes no TGI, ou o aumento da qualidade dos vídeos da cápsula endoscópica, incluindo novas informações que um vídeo comum não possui, como por exemplo uma terceira dimensão ou uma maior fluidez entre *frames*.

Nesta dissertação foram feitos dois trabalhos, de certa forma complementares, porque abordam duas técnicas diferentes para a melhoria do diagnóstico dos médicos quando visualizam vídeos da CE. O primeiro tem como objetivo o aumento da qualidade dos vídeos da CE, ao adicionar uma terceira dimensão às imagens, incluindo assim informação de forma das estruturas que é muito importante para a detecção de diversas patologias. Esta não está visível em imagens 2-D, o que faz com que a sua inclusão nos vídeos da cápsula endoscópica seja muito importante para a análise dos vídeos por parte dos gastroenterologistas. O segundo tem como objetivo a detecção automática de tumores em imagens isoladas da CE, o que facilita o diagnóstico destes exames, já que os médicos podem analisar mais profundamente as imagens assinaladas pelo sistema, diminuindo significativamente o número de imagens a analisar pelo médico e consequentemente os denominados erros de natureza subjetiva, especialmente aqueles que tiverem o “stress” como origem.

Existem atualmente alguns exames feitos ao sistema gastrointestinal que incluem informação tridimensional, como por exemplo a endoscopia virtual usando a Tomografia Axial Computadorizada (TAC). Basicamente esta técnica usa um conjunto de *slices* de um exame de TAC para reconstruir uma zona do sistema gastrointestinal. Devido ao reduzido detalhe desta tecnologia, não é possível utilizá-la no intestino delgado, sendo apenas usada na parte final do intestino (intestino grosso), sendo conhecida por Colonoscopia por Tomografia Computadorizada (CTC). As estruturas 3-D podem ainda ser obtidas a partir de exames de Ressonância Magnética (RM), onde o detalhe é bastante mais elevado que no caso da tecnologia TAC, e não existe qualquer radiação envolvida, tendo como principal desvantagem o elevado custo associado.

Devido às várias desvantagens das alternativas existentes, o uso de vídeos da cápsula endoscópica para a reconstrução 3-D da mucosa intestinal do intestino delgado seria um avanço de grande importância para a melhoria das condições de diagnóstico quer este seja efetuado exclusivamente por gastroenterologistas, quer sejam usados sistemas automáticos de apoio ao diagnóstico.

1.2. Objetivos

O principal objetivo do trabalho aqui proposto era construir uma ferramenta que aumentasse a qualidade de um vídeo de cápsula endoscópica, incorporando uma terceira dimensão às imagens e assim adicionar informação de forma, à mucosa do intestino delgado.

Esta informação adicionada conjuntamente com os respetivos métodos de a obter deve ser aproveitada para melhorar o desempenho de algoritmos de classificação para deteção de tecido tumoral.

A principal referência no desenvolvimento deste trabalho está descrita em (Karargyris & Bourbakis 2011b). Aqui é usado um algoritmo de segmentação para dividir as várias *frames* em diferentes regiões, para poder calcular uma medida de similaridade entre elas. Usando apenas os pares de *frames* com um certo valor de similaridade, é feita a interpolação de imagens intermédias para aumentar virtualmente o *frame rate* destas sequências. Uma das desvantagens deste trabalho é a impossibilidade de utilizar a maioria das *frames* devido ao elevado *threshold* de similaridade usado, logo só é usado com pares de *frames* muito semelhantes entre si (algumas delas quase iguais).

Os objetivos desta dissertação são então os seguintes:

- Desenvolvimento de um algoritmo de segmentação de *frames* de cápsula endoscópica baseado em diferenças de cor e removendo a influência da luminosidade, usando uma abordagem *Maximum a Posteriori* (MAP).
- Extração de *features* de cor e textura para comparar regiões de *frames* consecutivas.
- Classificação das regiões segmentadas em tecido normal e tecido tumoral. Análise do desempenho desta abordagem relativamente à abordagem mais comum de processamento da *frame* completa.
- Aplicação de um método de *Shape from Shading* (SfS) para a reconstrução 3-D de imagens 2-D, fazendo as devidas considerações da câmara da CE e do modelo do sistema.

As três primeiras partes são baseadas em dois trabalhos já desenvolvidos (Vieira et al. 2012; Vieira et al. 2013). Estas envolvem o uso do espaço de cor Lab para a remoção da influência da saturação luminosa no processo de segmentação que usa uma abordagem MAP multidimensional. Posteriormente é usado o critério de informação bayesiana (*Bayesian Information Criterion*) para determinar o número de regiões de natureza diferente em cada uma das *frames*. A extração de *features* envolve a computação de medidas estatísticas de transformadas complexas como *wavelets* e *curvelets* para a caracterização da textura, e a computação de medidas estatísticas de canais de cor representativos de diferenças entre cores opostas para a caracterização da informação de cor. A reconstrução tridimensional das *frames* da

CE é feita usando o algoritmo de SfS proposto por Tsai (Tsai & Shah 1994), com certas modificações relacionadas com características das câmaras usadas na tecnologia de CE.

Relativamente à deteção automática de *frames* contendo tecido tumoral (Vieira et al. 2013), foi essencialmente conseguido um aumento de desempenho relativamente aos sistemas já existentes usando uma abordagem baseada em Regiões de interesse (ROI) não rígidas e fornecidas pelo segmentador quando divide a imagem em duas regiões. Com base nestas ROI é efetuado o cálculo de diversas medidas estatísticas dos histogramas de canais do espaço de cor Lab e a posterior classificação com o uso de classificadores supervisionados.

1.3. Organização da dissertação

O texto desta dissertação foi dividido em nove capítulos. O Capítulo 2 é composto por um estado da arte sobre a generalidade de algoritmos de processamento de imagem usados na CE, mais especificamente os que fazem a deteção de tumores e os de reconstrução tridimensional. O Capítulo 3 consiste no contexto clínico deste trabalho, mais concretamente o sistema gastrointestinal, o intestino delgado e as técnicas imagiológicas para o estudo destes órgãos (endoscopia e cápsula endoscópica). O Capítulo 4 apresenta os métodos utilizados neste trabalho em relação à segmentação de tecidos. No Capítulo 5 são explicados métodos de reconstrução 3-D a partir de imagens 2-D, principalmente o algoritmo *Shape from Shading*. No Capítulo 6 são explicadas as especificidades da deteção automática de tumores, tal como a remoção da luminosidade da CE e as *features* utilizadas. No Capítulo 7 são apresentadas certas especificações na implementação dos algoritmos, tal como bases de dados e *softwares* utilizados durante a realização do trabalho. Os resultados obtidos nas várias fases do trabalho aqui descrito são apresentados no Capítulo 8. Para finalizar, no Capítulo 9 são apresentadas as principais conclusões, bem como diversas propostas para trabalho futuro.

Capítulo 2 Estado da Arte

O processamento de imagens e vídeos da CE tem sido um tema bastante estudado dentro da comunidade científica. Em relação a aplicações de visão por computador e processamento de imagem na CE salientam-se alguns temas mais comuns: a segmentação de vídeo, a classificação de *frames* em relação a patologias presentes nos tecidos e a reconstrução tridimensional.

Em (Kodogiannis et al. 2007) são propostos dois esquemas diferentes para a extração de *features* a partir de informação de textura, tanto no domínio cromático como no domínio acromático, para depois serem usadas na classificação das *frames* como normais ou anormais. A primeira analisa a imagem à procura de elementos de textura caracterizados anteriormente (*textons*), definindo assim a imagem a partir dos elementos presentes. A segunda é uma abordagem estatística, onde descritores são calculados a partir dos histogramas dos vários canais dos espaços de cor RGB e HSV. Com um objetivo um pouco diferente, em (Cunha et al. 2008) são extraídas *features* a partir de descritores visuais de MPEG-7, para fazer a segmentação topográfica dos vídeos da cápsula endoscópica e estimar o tempo de trânsito gástrico e intestinal. A segmentação topográfica do vídeo da CE foi também desenvolvida em (Mackiewicz et al. 2008), onde cada *frame* era dividida em 12 subimagens iguais com formas quadradas e sempre na mesma posição, para descartar imagens com pouca informação dos tecidos intestinais. Era depois retirada informação de cor, textura e movimento das regiões de interesse para usar em

classificadores supervisionados e assim definir a região do TGI de onde o tecido existente na *frame* provém.

2.1. Deteção de tumores

Para deteção de tecido tumoral em imagens da endoscopia, em (Maroulis et al. 2003) são usadas estatísticas de segunda ordem calculadas no domínio da transformada de *wavelet* (bandas 1, 2 e 3) de cada uma das imagens; e em (Karkanis et al. 2003) são usadas também estatísticas de segunda ordem no domínio da transformada de *wavelet*, mas aqui nas bandas 4, 5 e 6. A deteção de tumores em *frames* de CE, e conseqüentemente em tecido do intestino delgado, usa medidas estatísticas de descritores de textura de Haralick retirados de matrizes de co-ocorrência, usando diferentes bandas da transformada de *wavelet* para selecionar a informação relevante (Lima et al. 2008; Barbosa et al. 2008a; Barbosa et al. 2008b). Foram também já usadas *features* estatísticas provenientes das escalas de maior ordem da transformada de *curvelet* para fazer a classificação de *frames* tumorais (Barbosa et al. 2009; Martins et al. 2010); e mais recentemente, a elaboração de esquemas mais complexos que incluem o uso de momentos de segunda e terceira ordem calculados entre medidas estatísticas de descritores de Haralick de matrizes de co-ocorrência em diferentes canais de cor e diferentes escalas da transformada *wavelet* (Barbosa et al. 2012).

Em todas estas abordagens são usados descritores de textura baseados em transformadas complexas, o que eleva o tempo e complexidade computacional. Além disso, com o uso sempre de toda a imagem para codificar a informação de tumor, se o tecido lesionado tiver uma área pequena em relação à área total da imagem, esta poderá perder-se dentro da informação codificada pelo tecido normal. A abordagem da segmentação das imagens em duas regiões e a posterior extração de informação de cor fazem com que se consiga ultrapassar alguns dos problemas referidos anteriormente.

2.2. Reconstrução tridimensional

A reconstrução tridimensional de imagens e vídeos genéricos tem sido bastante estudada já desde os anos 70. Foi no início desta década que um intenso estudo por parte de Horn possibilitou o desenvolvimento do primeiro algoritmo de *Shape from Shading* (Horn 1970),

viabilizando a reconstrução 3-D a partir de imagens isoladas. Mais tarde um método fotométrico foi desenvolvido para determinar a orientação de superfícies usando várias imagens de um mesmo objeto (Woodham 1980). Estes dois princípios foram melhorados ao longo dos anos, alterando os métodos e algoritmos usados, sendo neste momento ainda um tema em estudo.

No resto deste capítulo serão explicados mais em pormenor alguns dos estudos usados em imagens e vídeos retirados de exames ao TGI, mais especificamente a endoscopia e CE.

2.2.1. Endoscopia

O primeiro método de reconstrução 3-D com o objetivo de aplicar a sequências de imagens endoscópicas (Forster & Tozzi 2000) usava o algoritmo SfS com remoção das componentes refletivas especulares. Foi implementado ainda um método de correção radial de distorção devido à qualidade das imagens vindas da câmara do endoscópio. Apesar de em imagens de teste os resultados terem sido considerados bastante razoáveis, quando aplicados em imagens reais do TGI, o resultado não foi o esperado. Isto deveu-se principalmente às grandes irregularidades das formas das superfícies e à falta de qualidade das imagens do endoscópio. Mais tarde foi desenvolvido um método baseado também em SfS, mas usando múltiplas imagens (Wu et al. 2009). O sistema multi-imagem usado envolve o conhecimento da posição do endoscópio ao longo do tempo, usando um algoritmo *Iterative Closest Point* (ICP) para melhorar as correspondências das estruturas entre *frames*. Além disso, a câmara presente no endoscópio tem que ser calibrada com um novo método para este algoritmo funcionar corretamente, o que impossibilita o uso de um método destes em vídeos já adquiridos e em que o instrumento não foi assim calibrado.

Em (Kaufman & Wang 2008) é usado uma combinação do algoritmo SfS e do algoritmo *Shape from Motion* (SfM). Este último método permite a descoberta dos parâmetros de movimento da câmara do endoscópio, tal como a localização 3-D de certos pontos característicos das superfícies reconstruídas pelo SfS. No final, zonas parciais das superfícies são integradas através de métodos de registo. Este método, apesar de apresentar boas perspetivas para o uso em imagens reais de exames de endoscopia, apenas foi testado em imagens simuladas e sem alterações abruptas.

Mais tarde surgiu um método diferente que permitia juntar duas imagens endoscópicas, fazendo uma reconstrução em simultâneo das duas (Thormaehlen et al. 2002). Este estudo faz a deteção de pontos característicos em cada *frame*, correspondendo depois os conjuntos de pontos de duas *frames* consecutivas para encontrar os melhores pares. Assim é possível fazer a estimação

do movimento da câmara do endoscópio. Um estudo posterior desenvolveu já um algoritmo que segue o mesmo princípio, fazendo a correspondência entre três *frames* consecutivas em vez de duas (Faltin & Behrens 2010). Apesar dos bons resultados deste último estudo em imagens de teste, não foram usadas imagens de endoscopia até ao momento para provar a fiabilidade do algoritmo para esta aplicação.

Num estudo de 2004 (Wu et al. 2004) foi também desenvolvido um modelo da câmara presente nos endoscópios, baseado no modelo de uma câmara *pinhole* (ou estenopeica), com a adição de um método de factorização entre pontos de *frames* consecutivas para determinar o movimento da câmara e a forma do objeto. Um estudo posterior (Wang et al. 2008) usa um vídeo completo de um exame de endoscopia para fazer a reconstrução, adicionando ao modelo construído com todas as *frames* já processadas, a informação da nova imagem, aumentando assim o modelo enquanto se vai avançando no vídeo. Este modelo, como resultado final não conseguiu gerar uma verdadeira reconstrução tridimensional, mas apenas conjuntos de pontos correspondentes nas várias *frames*, constituindo por isso ainda um estudo bastante preliminar.

Outra abordagem para o problema da reconstrução é usar certas características do endoscópio para ajudar no correto funcionamento do algoritmo. A forma cilíndrica do endoscópio, e a posição dele em relação ao tubo digestivo (sempre paralelo ao eixo do tubo), faz com que possa ser usado um modelo cilíndrico do tubo digestivo para fazer a reconstrução tridimensional da superfície deste (Zhou et al. 2008). É usada uma solução MAP com uma formulação baseada em *Markov Random Fields* (MRF) para fazer a reconstrução a partir dos dados das imagens. Outra abordagem relacionada com uma especificidade do endoscópio é usando o conhecimento da posição da ponta da ferramenta do endoscópio (González et al. 2009). A partir desta informação e desta posição, propaga-se a informação da profundidade da imagem, fazendo assim a reconstrução.

2.2.2. Cápsula endoscópica

A maioria algoritmos usados para a reconstrução tridimensional em exames de endoscopia convencional não podem ser diretamente migrados para aplicar em imagens da CE. Isto deve-se em parte às seguintes razões:

- *Frame rate* reduzido da CE em relação à endoscopia convencional.

- Movimentos abruptos da CE devido aos movimentos peristálticos no TGI e ao não controlo do dispositivo por parte do pessoal médico.
- Falta de conhecimento com precisão adequada das coordenadas espaciais do dispositivo em cada momento do exame

Mesmo com as dificuldades supramencionadas, têm sido desenvolvidos alguns estudos que abordaram a reconstrução tridimensional em imagens da CE. No entanto, ainda nenhuma aplicação obteve resultados suficientemente completos e consistentes que permitam a sua integração num sistema utilizável pela comunidade médica.

Em (Prasath et al. 2012) é feita a segmentação da mucosa intestinal recorrendo ao algoritmo de contornos ativos, usando métodos de correção luminosa. Posteriormente usa-se a região de interesse resultante do processo de segmentação para fazer a reconstrução usando o SfS tradicional. Este método tem como resultado final um modelo tridimensional para cada uma das *frames* da cápsula, não endereçando o problema da junção das reconstruções das *frames* consecutivas.

A reconstrução 3D usando múltiplas *frames* da cápsula endoscópica foi feita em (Sun et al. 2010), usando apenas pares de *frames* consecutivas. Neste trabalho é necessário fazer a calibração da câmara presente na cápsula, sendo que a junção das *frames* é feita encontrando pontos de interesse nas duas imagens (usando o detetor de Harris e Stephens), com o objetivo de encontrar pontos correspondentes nas duas *frames*. Usando depois princípios de estereoscopia, é possível calcular o movimento e posição da câmara em relação às imagens. Apesar de ter apresentado resultados promissores, não chegou a ser testado em imagens reais da CE. Mais tarde, um método semelhante foi desenvolvido (Fan & Meng 2011). Aqui também pontos de interesse nas várias imagens são encontrados, mas desta vez usando o método *Scale Invariant Features Transform* (SIFT), que deteta os extremos do laplaciano de gaussianas; e depois a partir de descritores de cor e gradiente comparam-se os pontos entre *frames* consecutivas. Posteriormente, partindo de métodos de geometria epipolar é possível determinar as relações geométricas entre os pontos. Numa fase posterior, são geradas *meshes* a partir da triangulação de Delaunay e mapeamento de textura para construir o modelo resultante.

Em (Karargyris & Bourbakis 2011b) está descrito o trabalho mais completo que aborda a reconstrução 3D de imagens da cápsula endoscópica. Aqui é endereçado um importante problema que é o baixo *frame rate* desta tecnologia. Assim, cada *frame* é dividida em várias regiões, usando um algoritmo de segmentação de crescimento de regiões baseado na abordagem fuzzy. Através

de características de cor, textura, forma e tamanho; regiões de *frames* consecutivas são comparadas para encontrar regiões correspondentes entre *frames*. Através de um sistema de pontuação desenvolvido pelos autores, uma medida de similaridade é calculada, descartando pares de *frames* que tenham um valor abaixo de um certo *threshold* (neste caso, menor que 80%). Com os pares de imagens de interesse é efetuada interpolação, para aumentar virtualmente o *frame rate* nessas sequências de imagens. Este algoritmo de interpolação é aplicado tendo em conta o conhecimento do movimento das regiões entre *frames*, tanto entre regiões com informação mútua como regiões sem informação mútua. No final, é possível construir modelos 3-D de cada uma das *frames* (originais e interpoladas) através do método SfS. Como principais desvantagens do método apresentado, pode-se mencionar a eliminação de grande parte das *frames* devido ao *threshold* de similaridade aplicado, e a não junção da informação de várias *frames* nos modelos tridimensionais. Estas desvantagens levam à obtenção de um vídeo altamente descontínuo e lento com demasiadas *frames* eliminadas por insuficiência de similaridade com as suas vizinhas. Deste modo, nem se consegue naturalidade suficiente de movimento no vídeo gerado, nem porventura a informação 3-D gerada será suficiente para atenuar os efeitos da informação 2-D perdida por eliminação das *frames* que tornariam o movimento no vídeo mais natural.

Capítulo 3 Contexto clínico

Neste capítulo será primeiro abordada a estrutura, funções e principais características do TGI. Numa segunda parte, será abordada a história das técnicas de visualização e imagiologia desta zona do organismo, com uma explicação mais exaustiva da cápsula endoscópica, as suas principais características, vantagens e desvantagens.

3.1. Sistema gastrointestinal

O TGI consiste num conjunto de órgãos ocios, desde a boca até ao ânus, incluindo glândulas acessórias que segregam várias substâncias importantes para o correto funcionamento deste sistema (Boron & Boulpaep 2003). O tubo digestivo pode ser dividido em duas partes, o TGI superior, que inclui a boca, faringe, esófago e estômago, e o TGI inferior, constituído pelo intestino delgado (duodeno, jejuno e íleon) e o intestino grosso (cego, cólon, sigmóide, recto e ânus) (Brunicardi et al. 2006). A cápsula endoscópica é usada principalmente para a visualização do intestino delgado, órgão difícil de visualizar com outras técnicas de imagiologia usadas hoje em dia (como a endoscopia digestiva alta e baixa).

3.1.1. Intestino Delgado

O intestino delgado é o órgão que se estende desde o final do estômago (piloro) até ao começo do intestino grosso (válvula ileocecal), e é responsável pela maior parte da digestão e absorção da comida e nutrientes. O comprimento médio deste órgão é de 6,9 metros para o sexo masculino e de 7,1 metros para o sexo feminino, apesar de apresentar grandes variações: sendo que os valores extremos são o mínimo de 4,7 metros e um máximo de 9,7 metros (Gray 1918).

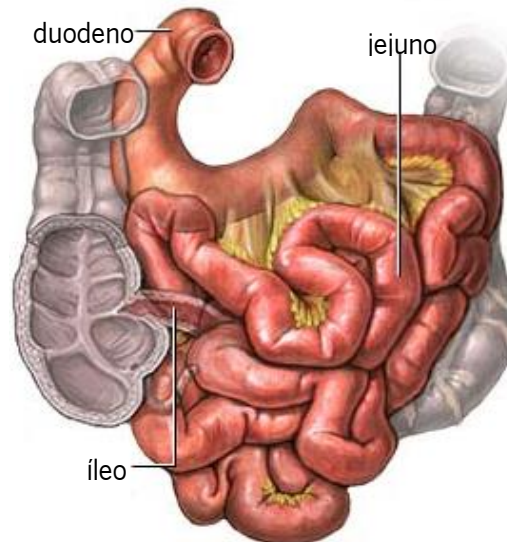


Figura 3.1. Esquema do intestino delgado e suas regiões. Adaptado de [3].

O intestino pode ser dividido em três regiões distintas: o duodeno, o jejunum e o íleo (Treves 1785) (Figura 3.1). O diâmetro difere em cada uma das zonas, sendo que vai diminuindo enquanto se chega ao final desta porção de intestino. O duodeno pode chegar aos 5 cm de diâmetro, o jejunum aos 4 cm e o íleo aos 3 cm (Gray 1918; Davis 1935).

Os alimentos ingeridos, depois de passarem pelas duas primeiras fases de digestão e absorção (na boca e no estômago), encontram-se num estado denominado de quimo, juntamente com água, ácido clorídrico e enzimas digestivas (Thomson et al. 2001). Este semifluido, que tem um pH com um valor de aproximadamente 2, ativa a secreção de enzimas libertadas assim que entra no duodeno, possibilitando a continuação da digestão dos alimentos. Quando o quimo ultrapassa o piloro, são segregadas enzimas pancreáticas e biliar, que irão ter um papel de destaque na digestão de lípidos, proteínas e hidratos de carbono (Boron & Boulpaep 2003). Com a junção destas novas enzimas alcalinas, o pH do quimo aumenta, levando à ativação de outras enzimas que facilitam o catabolismo de algumas moléculas, como é o caso de partículas de gordura (Gray 1918). É também durante o transporte dos alimentos por este órgão que se dá a

maior parte da absorção dos nutrientes pelo organismo, sendo este processo facilitado pela grande área da superfície do intestino, aumentado pela existência de vilosidades e microvilosidades intestinais.

Face à enorme extensão do intestino delgado, este necessita de um mecanismo que permita a propulsão dos alimentos em digestão ao longo de todo o seu trato, denominando-se de movimentos peristálticos das paredes intestinais. Este processo é composto por ondas de relaxamento e contração que se propagam ao longo do intestino, fazendo com que quimo avance até ao fim do íleo. Estes movimentos são possíveis devido à existência de duas camadas de tecido muscular – uma camada interior circular e uma camada exterior longitudinal. Quando a primeira camada contrai, a segunda relaxa, levando à diminuição do lúmen, o que propulsiona os alimentos, transportando-os para a frente. Quando o músculo longitudinal contrai, existe relaxamento da camada interior, aumentando o lúmen intestinal, facilitando a entrada de novos alimentos naquela zona (Boron & Boulpaep 2003).

No intestino delgado existem diversas lesões que podem ocorrer, sendo de destacar as úlceras, os pólipos e as hemorragias resultantes de outras patologias subjacentes. Estas ocorrem em grande número aqui, por ser o órgão visceral mais sujeito a lesões, apesar de ser o mais preparado para lidar com problemas desta natureza. A sua grande elasticidade e a facilidade com que os seus anéis se sobrepõem em cima dos seguintes, faz com que as contusões e situações em que altas pressões são colocadas sobre este órgão, não tenham consequências muito problemáticas (Treves 1785).

As úlceras pépticas são erosões na mucosa do intestino, sendo que no intestino delgado ocorrem principalmente no duodeno, e quando benignas, geralmente na primeira porção deste. Estas podem ser simples ou múltiplas, podem ter uma forma circular ou oval (pode ser visto Figura 3.2a) e podem ser superficiais ou profundas. São geradas principalmente por aumento na produção do suco gástrico, fluido ácido que em grandes quantidades pode superar os mecanismos de defesa da mucosa duodenal, outro mecanismo pode dever-se à presença da bactéria *Helicobacter pylori*, que interfere nas defesas naturais contra o ácido gástrico (Berkow & Beers 2003).

Os pólipos devem-se a um crescimento de tecido anormal da parede intestinal (Figura 3.2b). Ao contrário das úlceras, estas lesões geralmente não apresentam formas constantes e regulares (The Gale Group Inc. 2008). A maioria dos pólipos existentes no intestino são benignos, apesar de o risco de se tornarem cancerosos aumente quanto maior for o seu tamanho e o seu

número (Berkow & Beers 2003). Estas lesões podem ser encontradas no íleo, jejuno e também no duodeno (em ordem de diminuição da sua frequência), sendo que estes pólipos no intestino têm um crescimento considerado como bastante lento, o que faz com que a sua sintomatologia seja insidiosa, explicando a importância do diagnóstico precoce e correto (The Gale Group Inc. 2008).

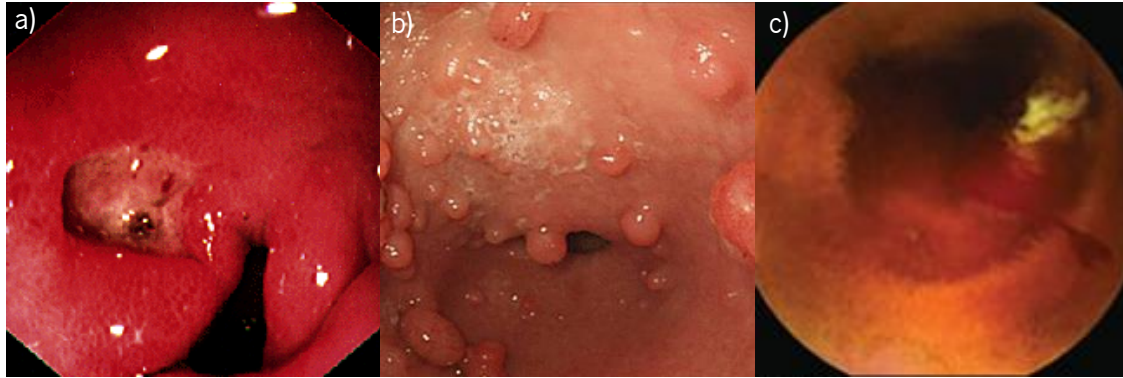


Figura 3.2. Exemplo de uma úlcera duodenal (a – Adaptado de (Fusco 2013a)), um pólipio intestinal (b – Adaptado de (Umeoka et al. n.d.)) e uma hemorragia intestinal (c – Adaptado de (Clancy 2008)).

Uma hemorragia digestiva corresponde a qualquer perda hemática com origem no TGI e ocorre quando a parede intestinal começa a sangrar anormalmente (Figura 3.2c), sendo muito relevante na prática médica devido à sua alta prevalência. Podem distinguir-se dois tipos de hemorragia digestivas, consoante a sua origem que são a hemorragia digestiva alta (HIA) e baixa (HIB); sendo que a primeira se refere a sangramentos do esófago até ao ângulo de Treitz (ângulo formado ao nível da junção entre o duodeno e o jejuno) e a segunda aos sangramentos que ocorrem entre o ângulo de Treitz e o ânus (Longo et al. 2011). Recentemente, já se começa a utilizar o termo hemorragia intestinal média para os sangramentos neste órgão específico (Prakash & Zuckerman 2003). As hemorragias digestivas podem também ser caracterizadas consoante se tratem de quadros agudos, geralmente maciços e de diagnóstico mais fácil, e quadros crónicos, que demoram meses a ser diagnosticados podendo mesmo ser indetetáveis (apenas detetáveis por pesquisa de sangue nas fezes ou no decurso do estudo de uma anemia crónica). Isto surge devido à dificuldade de visualização de hemorragias no intestino delgado através dos métodos de imagem utilizados atualmente na prática clínica (Longmore et al. 2011). Apesar desta dificuldade, apenas 5% das hemorragias no TGI ocorrem no intestino delgado (Weragama 2011), sendo que os métodos complementares de diagnóstico atualmente mais usados são, frequentemente, suficientes para detetar as hemorragias existentes nas restantes porções do TGI. No que diz

respeito às HDB, mais especificamente às localizadas no intestino delgado, estas podem ter diversas etiologias, como algumas das lesões mencionadas anteriormente (úlceras e pólipos), mas podem dever-se também à doença de Crohn (doença inflamatória intestinal), a ectasias vasculares (malformações arteriais ou venosas) e a neoplasias (Lewis 2000). Em indivíduos jovens a hemorragia por divertículo de Meckel (divertículo congénito do intestino delgado) é frequente (Longmore et al. 2011).

3.2. Técnicas de imagiologia do trato gastrointestinal

Hoje em dia existem várias técnicas que permitem a visualização do sistema gastrointestinal. A mais usada é a endoscopia (alta ou baixa), uma técnica invasiva que não permite observar a maior porção do intestino delgado. Existe uma nova técnica, a cápsula endoscópica, que permite ver todo o sistema gastrointestinal de uma forma não-invasiva para o paciente.

3.2.1. Endoscopia

A palavra endoscopia tem origem nos termos gregos *endo*, que significa dentro e *skopéin*, que significa observar (Kramme et al. 2011). Pode-se então concluir que a junção destes dois termos significa “observar o que está dentro”. Realmente, o exame de endoscopia tem como objetivo visualizar o interior de órgãos ocultos ou cavidades do organismo, usando um instrumento chamado de endoscópio (Berkow & Beers 2003).

Os primórdios do endoscópio surgiram no século XIX, mais propriamente em 1806, pelas mãos do médico alemão Philipp Bozzini, com um aparelho inventado pelo mesmo, denominado de *Lichtleiter* (cuja tradução significa condutor de luz) - Figura 3.3. Este foi construído usando espelhos posicionados de forma a projetar as imagens das estruturas do organismo para o olho humano (Natalin & Landman 2009). A fonte de luz era apenas uma vela, devido à falta de alternativas na altura em que foi construído, e eram usados tubos duplos de alumínio para inserir nas cavidades (Vilardell 2005). Apesar de ter sido uma revolução médica quando surgiu, esta invenção foi ridicularizada entre os especialistas. Além disso as aplicações eram limitadas, devido apenas à possibilidade de visualização de uma pequena área da mucosa da bexiga, e da falta de iluminação do aparelho (Natalin & Landman 2009; Zajackowski & Zamann 2004).



Figura 3.3. Réplica do *Lichtleiter* original. Adaptado de (Natalin & Landman 2009).

Os sucessores de Bozzini tentaram melhorar ao máximo as condições de iluminação do aparelho. Em 1826, Pierre Salomon Segalas adicionou uma segunda vela e um cateter para drenar os conteúdos da bexiga e assim facilitar a sua visualização (Natalin & Landman 2009). Foi em 1853 que o termo “endoscópio” foi usado pela primeira vez, devido a uma invenção do urologista francês Antoin Jean Désormeaux. Este baseou-se no *Lichtleiter* para construir um aparelho, onde substituiu as velas por uma chama resultante da combustão de uma solução alcoólica a 96% com terebintina, que poderá ser focada, aumentando a capacidade do dispositivo (Natalin & Landman 2009).

Foi em 1873 que Gustave Trouvé implementou iluminação elétrica pela primeira vez num endoscópio. Este foi um passo importante na evolução deste dispositivo, já que possibilitou uma iluminação de qualidade, e que não usasse fontes de luz com perigos para os pacientes, como era o caso das chamas (Zajaczkowski & Zamann 2004).

Outro dos grandes problemas no desenvolvimento do endoscópio não era só o sistema de iluminação, mas também a possibilidade de este ser flexível e assim permitir a visualização de zonas mais longínquas no organismo. Em 1882, Johann von Mikulicz-Radecki desenvolveu um primeiro instrumento cuja parte final podia ser dobrada até um ângulo de 30° sem que a capacidade de visualização fosse perdida. Apesar de não ter sido implementado na área médica pelo seu criador, esta tecnologia permitiu um grande avanço para que fosse possível essa aplicação específica no futuro (Zajaczkowski & Zamann 2004). Foi então apenas em 1932 que Georg Wolf, juntamente com o médico Rudolf Schindler, produziu um endoscópio semi-flexível

(Figura 3.4). Este pode ser considerado o primeiro endoscópio moderno, tendo 78 cm de comprimento, com um segmento rígido de 34 cm e um diâmetro de 8.5 mm e um segmento flexível com 12 mm de diâmetro (Vilardell 2005; Anthony J. DiMarino & Benjamin 2002).



Figura 3.4. Endoscópio semi-flexível de Wolf-Schindler. Adaptado de (Anthony J. DiMarino & Benjamin 2002).

Um dos mais importantes desenvolvimentos nesta tecnologia foi a inclusão da fibra ótica na sua construção. Isto foi conseguido no ano de 1957, por Basil Hirschowitz e Larry Curtis, apesar da fonte de iluminação utilizada (lâmpada de filamento de tungsténio) ser de pouca intensidade originando imagens escuras, e ainda alguma distorção de cor (Shah 2002). Este, além de ter sido o primeiro endoscópio que usava fibra ótica, foi também o primeiro instrumento de endoscopia completamente flexível. Pode-se considerar que o principal precursor do atual endoscópio foi inventado em 1966, por Karl Storz. Este reaproveitou o mecanismo anterior, mas a partir de princípios físicos enunciados por Harold Hopkins, substituiu os espaços de ar entre as várias fibras do tubo do endoscópio por pequenas lentes de vídeo, aumentando em grande escala a qualidade de imagem do mesmo (Zajaczkowski & Zamann 2004; Knyrim et al. 1990).

Os desenvolvimentos futuros deste instrumento e que resultaram no dispositivo que se usa nos dias de hoje foi incluir *Charged-Coupled Device* (CCD), e inúmeros acessórios com fins diagnósticos e terapêuticos ou cirúrgicos (incluindo bisturis, tesouras, pinças, etc) (Knyrim et al. 1990; Natalin & Landman 2009) - Figura 3.5.

Foi em 1969 que a tecnologia CCD foi criada pelos físicos Willard Boyle e George Smith. Este é um *chip* cujo objetivo inicial era armazenar dados em forma de cargas elétricas (Boyle & Smith 1970). O CCD, com a aplicação de armazenamento de memória teve muito pouco sucesso, mas mais tarde Tompsett, usando as diversas vantagens desta tecnologia como linearidade da resposta e alta sensibilidade para aplicar o CCD na formação de imagens (Tompsett et al. 1971). Esta tecnologia foi implementada em endoscópios apenas em 1983, depois do chip de CCD ser miniaturizado e as suas funções serem otimizadas por Welch Allyn (Anthony J. DiMarino & Benjamin 2002). Uma das alterações que teve de ser feita para que esta tecnologia fosse aplicada num endoscópio é que um chip CCD apenas recolhe intensidades em níveis de cinzento, e sendo

que existem certos problemas que são diagnosticados devido a variações na cor dos tecidos, é preciso inserir três CCD's para que cada um detete a contribuição de cada uma das cores primárias, vermelho, verde e azul.

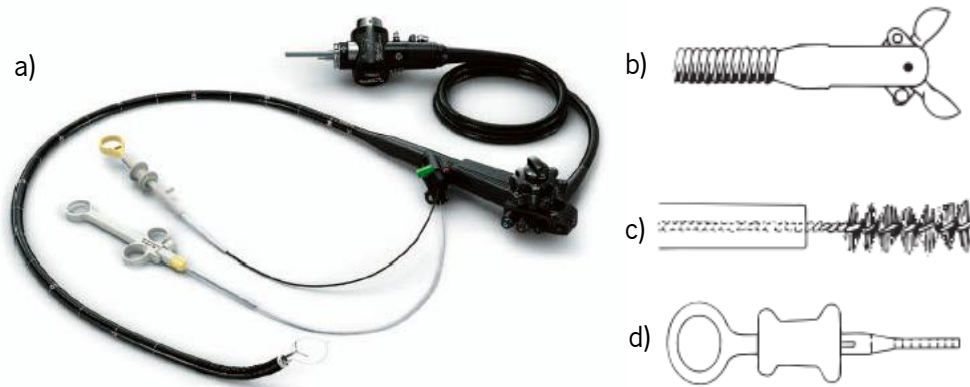


Figura 3.5. Exemplo de um endoscópio atual (a – Adaptado de (Olympus America 2011)), e alguns exemplos de ferramentas que podem ser colocadas na ponta do endoscópio: pinça de biópsia (b), escova citológica (c) e base para fórceps (d) – Adaptado de (Cotton et al. 2009).

Apesar de todos os avanços tecnológicos no desenvolvimento do endoscópio, este continua a ter limites na visualização do TGI. Sendo que através da endoscopia digestiva alta é possível ver da boca até ao duodeno, e através da endoscopia digestiva baixa (ou colonoscopia), é possível ver do ânus até ao íleo. Assim, a maior parte do intestino delgado não pode ser visualizado usando a endoscopia convencional (Natalin & Landman 2009). Foi devido a estas limitações que a cápsula endoscópica foi pensada e desenvolvida, tendo especial indicação no caso de pesquisa da doença de Crohn e de hemorragias digestivas ocultas (Adler & Gostout 2003).

3.2.2. Cápsula Endoscópica

A cápsula endoscópica é um dispositivo ingerível, com uma câmara miniaturizada, cujo intuito é de adquirir imagens da mucosa do TGI (Nakamura & Terano 2008). Foi em 1981 que Gavriel Iddan, um engenheiro mecânico israelita que trabalhava no desenvolvimento de dispositivos de imagem para mísseis, juntamente com o gastroenterologista Eitan Scapa, teve a ideia de desenvolver um “míssil” miniaturizado que pudesse passar pelo TGI e gravar imagens deste ao longo de todo o percurso (Perman 2011). O principal problema era a falta de tecnologia na altura para poder miniaturizar os vários componentes necessários para o desenvolvimento deste dispositivo. Além disso, o dinheiro para trabalhar nesse projeto era escasso e dentro do meio, não se pensava que fosse atingido o objetivo de Iddan (Fusco 2013b). O desenvolvimento

do dispositivo só foi possível depois de certos avanços na tecnologia da altura, das quais se podem destacar: o aparecimento da tecnologia *Complementary Metal Oxide Semiconductor* (CMOS), que possibilitou que todos os componentes da câmara fossem colocados num único chip; avanços na tecnologia *Application-Specific Integrated Circuit* (ASIC) que permitiram a integração de um transmissor de vídeo com eficiência e largura de banda suficiente que coubesse na cápsula e ainda *Light Emitting Diodes* (LEDs) que possibilitassem iluminação com qualidade suficiente para eliminar reflexões internas no tecido (Perman 2011).

Em 1998, Iddan e Gavriel Meron fundaram a *Given Imaging*, empresa que desenvolveu a primeira cápsula endoscópica, a M2A^{TM 1} (Nakamura & Terano 2008). Em 2000 começaram os ensaios clínicos desta cápsula, onde foi comprovado um diagnóstico mais eficiente de anomalias no intestino delgado do que as técnicas concorrentes (enteroscopia e técnicas cirúrgicas) (Perman 2011; Yu 2001). Um ano depois, em Maio de 2001 este produto recebeu a marcação da Conformidade Europeia (necessária para a comercialização no Espaço Económico Europeu), tendo sido aprovado pela *Food and Drug Administration* (FDA) (necessário para a comercialização nos Estados Unidos da América) em Agosto de 2001 (Yu 2001; YOQNEAM 2003).

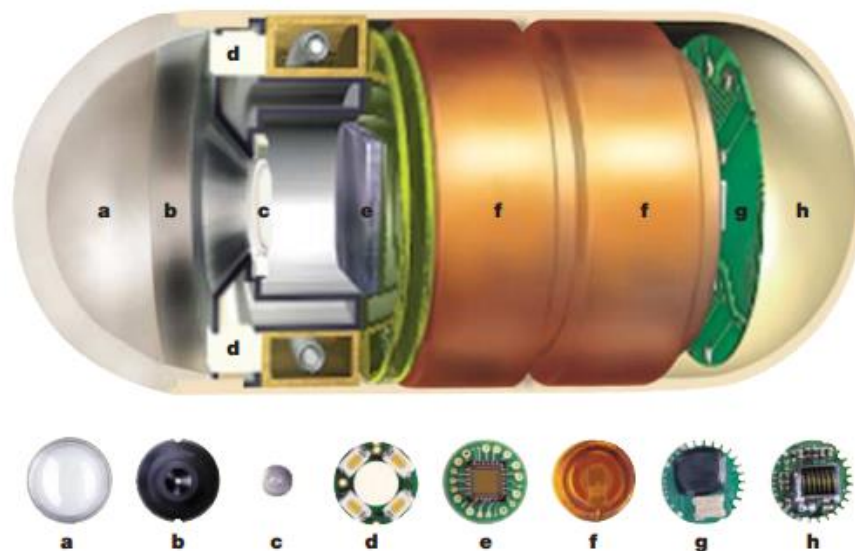


Figura 3.6. Cápsula Endoscópica M2ATM (em cima) e os seus constituintes (em baixo). Adaptado de (Qureshi 2004).

¹ *Mouth to Anus*. Traduzindo significa "Da boca ao ânus"

Esta primeira cápsula tinha dimensões de 11 mm x 26 mm e um peso de aproximadamente 3,7 gramas. A Figura 3.6 mostra esta cápsula e os seus componentes: uma cúpula ótica (a), um apoio para a lente (b), uma lente (c), quatro LEDs (d), um chip com uma câmara a cores (e), duas baterias (f), um transmissor de radiofrequência (g) e uma antena (h). Este dispositivo usa um chip CMOS, que necessita menos potência do que chips de CCD (usado na maioria dos endoscópios e câmaras digitais), podendo operar a níveis de luz bastante baixos. O campo de visão da câmara é de 140° e a capacidade de ampliação é de 1:8. O limite de tamanho de objetos que é possível detetar é de 0,1 mm e são transmitidas em média 50.000 imagens durante as 8 horas de filme (Yu 2001). Todas as imagens são guardadas e transmitidas para um gravador de dados portátil que é usado na cintura do paciente (Perman 2011).

Em Setembro de 2004, a mesma companhia renomeou a M2A para *Pillcam SB*² para que pudesse continuar uma gama de cápsulas endoscópicas. Em Novembro do mesmo ano, a *Pillcam ESO* é aprovada pela FDA, como uma cápsula endoscópica para a visualização do esófago (Gillson 2004). Novos desenvolvimentos desta tecnologia pela *Given Imaging* foi o *Agile Patency System* em 2006, cápsula biodegradável para percorrer o TGI, com o objetivo de confirmar se é possível a ingestão de uma cápsula endoscópica e que não há obstrução intestinal (Hillcrest 2006), (Herrerias et al. 2008). No mesmo ano foi também desenvolvida uma cápsula com o principal objetivo de observar o intestino grosso (*Pillcam COLON*). Esta teve a necessidade de sofrer algumas alterações em relação ao dispositivo original devido a certas particularidades do intestino grosso. Foi adicionada uma terceira bateria para garantir que a cápsula funcionasse durante todo o trajeto do intestino grosso, e foi adicionada uma segunda câmara, já que devido ao maior diâmetro deste órgão (em relação ao intestino delgado), havia a possibilidade de rotação deste e de haver várias imagens das mesmas zonas sem haver imagens de outras (Adler & Metzger 2011).

Em 2007 uma segunda geração da cápsula endoscópica convencional foi lançada para o mercado com o nome de *Pillcam SB 2*. A grande inovação em relação à sua antecessora era o campo de visão mais alargado, já que aumentou de 140° para 156°. Uma nova versão (*Pillcam SB 3*) está agora a ser preparada para entrar no mercado com novas funcionalidades e desenvolvimentos, como um aumento na resolução das imagens, de 256x256 para 320x320 e um aumento do *frame rate* de 2 fps para 6 fps (Given Imaging 2013; Johnson et al. 2008).

² SB – *Small Bowel*. Traduzindo significa “Intestino Delgado”

Outras empresas desenvolveram cápsulas endoscópicas depois da *Given Imaging*. A primeira foi a *Olympus*, ao criar a *EndoCapsule*, uma cápsula com características muito semelhantes à *Pillcam SB*, exceto que usa um sensor de imagem CCD (Cave et al. 2008). Dois anos mais tarde, em 2007, a *IntroMedic* introduziu no mercado a *MiroCam*, que é um dispositivo cuja maior diferença com os seus concorrentes é a tecnologia de transmissão das imagens para o módulo externo (Kim et al. 2010).

A CE, apesar de ser a única forma de visualizar diretamente o intestino delgado e de ser a única técnica não invasiva para ver todo o TGI, ainda não tem uma grande utilização no Sistema Nacional de Saúde (SNS), já que é um produto não participado pela Segurança Social bem como pelos principais subsistemas de saúde, incluindo as principais seguradoras de saúde.

Capítulo 4 Segmentação

A segmentação de imagens é o processo de as dividir em múltiplas regiões (segmentos), e é usado geralmente para identificar objetos ou outras informações relevantes de uma imagem. No contexto deste trabalho a segmentação é usada para ajudar a identificar regiões caracterizadas por tecido da mesma natureza em *frames* consecutivas, estando também incluída num sistema que tem o objetivo de fazer a deteção automática de tumores em imagens da CE.

Os algoritmos de segmentação dividem-se principalmente em três grandes grupos: os manuais, os automáticos e os semiautomáticos (Suetens 2009). Os primeiros têm muitas desvantagens, incluindo a quantidade de tempo despendido pelos profissionais que a irão fazer, e ainda a variabilidade intrasujeito e intersujeito. Assim sendo, o nível de confiança é baixo em relação a outros métodos de segmentação (Suetens 2009; Withey & Koles 2007). Os métodos automáticos não têm qualquer influência humana no seu funcionamento, podendo ser utilizadas informações sobre modelos estatísticos das intensidades das imagens que serão usadas, formas das regiões de interesse, ou o encontro de fronteiras das regiões nas imagens (Suetens 2009). Os métodos semiautomáticos, geralmente usam as vantagens de cada um dos grupos anteriores, já que necessitam de um *input* da parte do utilizador, o que poderá melhorar a performance do segmentador nos casos em que as imagens apresentem algumas diferenças em relação ao esperado.

A segmentação de regiões em imagens de cápsula endoscópica tem sido uma área em grande desenvolvimento nos últimos tempos. A maioria destes algoritmos usam informações de cor (Vieira et al. 2012; Prasath et al. 2012; Bourbakis 2005) ou de textura (Klepaczko et al. 2010; Karargyris & Bourbakis 2011a). Num trabalho anterior (Vieira et al. 2012), a informação de cor foi suficiente para fazer a segmentação de tecido intestinal (mais especificamente, tecido tumoral), por isso neste trabalho será também usada apenas informação de cor para a segmentação. A segmentação será feita usando a abordagem MAP, sendo a estimativa dos parâmetros do modelo obtida recorrendo ao algoritmo *Expectation-Maximization* (EM), que será descrito nas secções seguintes.

4.1. Modelo estatístico

Para fazer a segmentação segundo o algoritmo EM, é necessário determinar um modelo estatístico para a distribuição dos dados que se encontram nas amostras (píxeis de cada imagem). Assim sendo, considera-se que cada píxel da imagem pertence a uma região, e o número de regiões em cada uma das *frames* será o número de diferentes tecidos que existirá nessa imagem.

Na Figura 4.1 é possível observar os histogramas de cada um dos canais de cor do espaço Lab de uma *frame* de exemplo. No caso do canal L, pode-se ver um *cluster* com grande intensidade, que se refere à zona de saturação no centro da imagem (mostrada com maior pormenor na Figura 6.1). No caso dos restantes canais de cor, pode-se ver a contribuição de vários *clusters* para a construção do histograma, apesar de ser difícil precisar visualmente o número de regiões presentes. Considerando que cada um dos *clusters* segue uma distribuição gaussiana, o modelo da imagem é considerado como uma mistura gaussiana, sendo a sua fdp dada na Equação (4.1).

$$p(x) = \sum_{i=1}^k \pi_i G(\mu_i, \Sigma_i) \quad (4.1)$$

$$\sum_{i=1}^k \pi_i = 1 \quad (4.2)$$

em que x é a intensidade do píxel, k o número de componentes da mistura (o número de regiões da imagem), π_i o coeficiente (predominância) da componente i na mistura, μ_i a média e Σ_i a matriz de covariância da componente i (Marin et al. 2005).

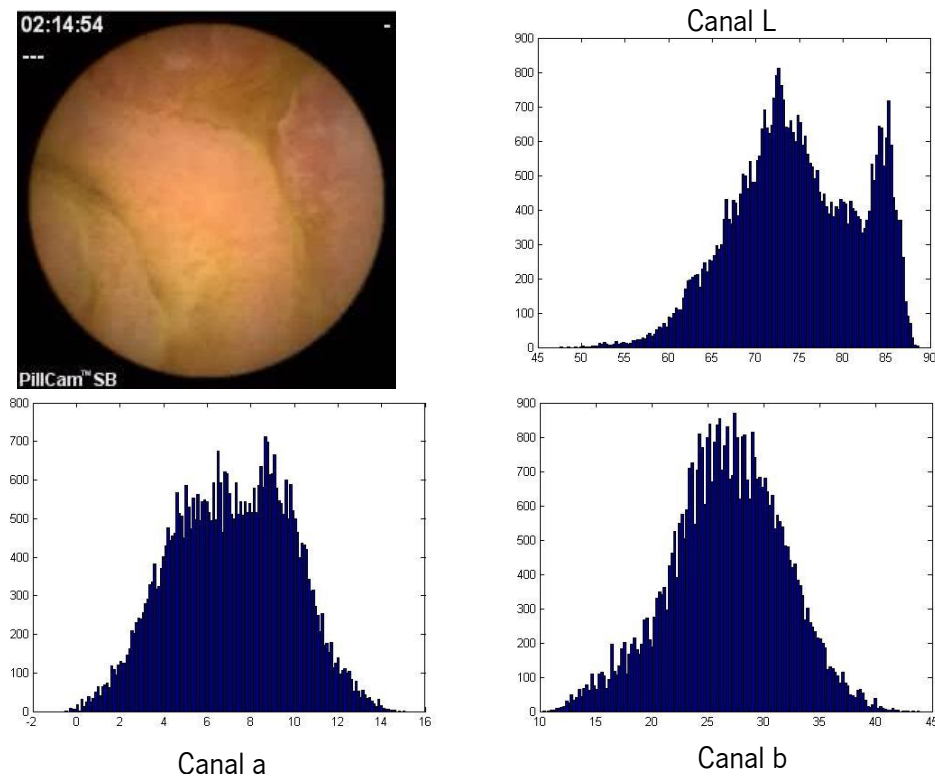


Figura 4.1. *Frame* da cápsula endoscópica e histogramas de cana um dos canais do espaço de cor Lab.

É importante referir que o modelo usado neste trabalho é uma mistura gaussiana multivariável, devido aos dois canais de cor (a e b) existentes em cada uma das imagens. Assim sendo, cada pixel será caracterizado em termos de intensidade por um vetor que contém as intensidades dos dois canais.

A abordagem MAP tem por objetivo encontrar os parâmetros que melhor caracterizam as componentes da mistura, especificamente o coeficiente, a média e a covariância, que maximizam a Equação (4.1) dada a imagem.

4.2. Regra de Bayes

A regra de Bayes, ou teoria de decisão bayesiana, é uma abordagem muito utilizada em problemas de reconhecimento de padrões. Esta tem como base a atribuição de custos às decisões de classificação corretas e incorretas e tomar decisões usando as distribuições probabilísticas dos objetos ao serem classificados.

Se houver como exemplo duas classes (C_1 e C_2), pode-se escrever a Regra de Bayes em termos de probabilidades *a posteriori* como:

se $P(C_1/x_j) > P(C_2/x_j)$, então considera-se que x_j pertence a C_1 , senão x_j pertence a C_2 .

Sendo $P(C_i/x_j)$, a probabilidade *a posteriori* de o vetor x_j pertencer à classe C_i (Soares 2006).

Matematicamente, define-se a regra de Bayes com a equação seguinte:

$$P(C_i|x_j) = \frac{p(x_j|C_i) \cdot P(C_i)}{p(x_j)} \quad (4.3)$$

em que $p(x_j/C_i)$ corresponde à função densidade de probabilidade (fdp) que é condicional à classe e ao vetor de dados x_j ; $P(C_i)$ é a probabilidade *a priori* da classe C_i e $p(x_j)$ é a fdp do vetor de dados x_j (Soares 2006).

Juntando as duas definições da Regra de Bayes, e como $p(x_j)$ é independente das classes, é possível obter uma nova definição:

Se $p(x_j/C_1) \cdot P(C_1) > p(x_j/C_2) \cdot P(C_2)$, então considera-se que x_j pertence a C_1 , senão x_j pertence a C_2 .

Será este conceito usado no algoritmo EM.

4.3. A Estimativa de Máxima Verosimilhança

A estimativa de máxima verosimilhança é por definição aquela que maximiza a probabilidade de ocorrência da observação. No contexto Bayesiano esta estimativa ocorre por maximização da Equação (4.3). Não havendo estatística suficiente para maximizar diretamente $p(C_i/x_j)$ a solução passa pela maximização iterativa de $p(x_j/C_i) \cdot P(C_i)$ usando o algoritmo EM pois $p(x_j)$ não depende da classe C_i .

A verosimilhança do pixel j dado a existência de k classes na imagem é dada pela mistura Gaussiana apresentada na Equação (4.1) e reescrita por conveniência na Equação (4.4) com terminologia mais apropriada à terminologia apresentada na secção anterior .

$$p(x_j|\varphi) = \sum_{i=1}^k P(C_i) \cdot p(x_j | C_i, \varphi_i) \quad (4.4)$$

em que φ_i é o vetor dos parâmetros que caracterizam a classe i do modelo estatístico usado (Bilmes 1998). A verosimilhança da imagem completa (n píxeis), assumindo que as intensidades

dos píxeis são estatisticamente independentes, é então dada por:

$$p(X|\varphi) = \prod_{j=1}^n p(x_j|\varphi) \quad (4.5)$$

em que X corresponde ao conjunto de todos os dados das amostras (x_0, x_1, \dots, x_n) .

Aplicando o logaritmo à equação anterior, obtém-se o logaritmo da verossimilhança, vulgarmente designado por log-verossimilhança das amostras (Soares 2006):

$$l(\varphi) = \log p(X|\varphi) = \log \prod_{j=1}^n p(x_j|\varphi) = \sum_{j=1}^n \log p(x_j|\varphi) = \sum_{j=1}^n \log \sum_{i=1}^k P(C_i) \cdot p(x_j|C_i, \varphi_i) \quad (4.6)$$

onde $l(\varphi)$ representa a função log-verossimilhança dos dados. A aplicação do logaritmo torna-se conveniente em virtude das pdf's serem Gaussianas, simplificando os cálculos e não tendo influência no processo de maximização dado que a função logaritmo é uma função crescente com o argumento.

Aplicando o operador gradiente à log-verossimilhança, obtém-se (Bilmes 1998):

$$\nabla_{\varphi_k} l(\varphi) = \sum_{j=1}^n \frac{1}{p(x_j|\varphi)} \cdot \nabla_{\varphi_k} \left[\sum_{i=1}^k P(C_i) \cdot p(x_j|C_i, \varphi_i) \right] = \sum_{j=1}^n \frac{1}{p(x_j|\varphi)} \nabla_{\varphi_k} [P(C_k) \cdot p(x_j|C_k, \varphi_k)] \quad (4.7)$$

Como:

$$\begin{aligned} P(C_i|x_j, \varphi_i) &= \frac{p(x_j|C_i, \varphi_i) \cdot P(C_i)}{\sum_{i=1}^k P(C_i) \cdot p(x_j|C_i, \varphi_i)} = \frac{p(x_j|C_i, \varphi_i) \cdot P(C_i)}{p(x_j|\varphi)} \Leftrightarrow \\ \Leftrightarrow \frac{1}{p(x_j|\varphi)} &= \frac{P(C_i|x_j, \varphi_i)}{p(x_j|C_i, \varphi_i) \cdot P(C_i)} \end{aligned} \quad (4.8)$$

Substituindo na equação do gradiente de $l(\varphi)$, chega-se à seguinte conclusão:

$$\begin{aligned} \nabla_{\varphi_k} l(\varphi) &= \sum_{j=1}^n P(C_k|x_j, \varphi_k) \cdot \frac{\nabla_{\varphi_k} [P(C_k) \cdot p(x_j|C_k, \varphi_k)]}{p(x_j|C_k, \varphi_k) \cdot P(C_k)} = \\ &= \sum_{j=1}^n P(C_k|x_j, \varphi_k) \nabla_{\varphi_k} \log [p(x_j|C_k, \varphi_k) \cdot P(C_k)] \end{aligned} \quad (4.9)$$

Para maximizar a verossimilhança, será necessário igualar a função anterior a zero.

$$\sum_{j=1}^n P(C_k|x_j, \varphi_k) \nabla_{\varphi_k} \log [p(x_j|C_k, \varphi_k) \cdot P(C_k)] = 0 \quad (4.10)$$

A fdp de uma distribuição gaussiana multivariável é dada pela seguinte equação:

$$p(x_j | C_k, \varphi_k) = \frac{1}{(2\pi_k)^{\frac{D}{2}} \cdot |\Sigma_k|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}(x_j - \mu_k)^T \Sigma_k^{-1} (x_j - \mu_k)\right) \quad (4.11)$$

onde D define a dimensão da distribuição normal. Neste caso D é igual a dois (correspondente aos dois canais de cor considerados para cada imagem) (Gut 1995).

Maximizando a Equação (4.10) em ordem a cada um dos parâmetros do modelo obtêm-se as seguintes expressões para a média, covariância e probabilidade *a priori* (Bilmes 1998):

$$\hat{\pi} = \frac{1}{n} \sum_{j=1}^n P(C_k | x_j, \varphi_k) \quad (4.12)$$

$$\hat{\mu} = \frac{\sum_{j=1}^n P(C_k | x_j, \varphi_k) x_j}{\sum_{j=1}^n P(C_k | x_j, \varphi_k)} \quad (4.13)$$

$$\hat{\Sigma} = \frac{\sum_{j=1}^n P(C_k | x_j, \varphi_k) (x_j - \mu)(x_j - \mu)^T}{\sum_{j=1}^n P(C_k | x_j, \varphi_k)} \quad (4.14)$$

4.4. O algoritmo EM

O maior problema do uso da máxima verosimilhança quando se usam misturas Gaussianas é ter o conjunto de dados incompleto. Ou seja, não sabendo a que *cluster* de dados cada pixel pertence, não é possível determinar diretamente os parâmetros de cada um dos conjuntos (Moon 1996).

Assim, é usado o algoritmo EM, que tem como objetivo preencher esta lacuna, logo vai permitir descobrir os parâmetros que definem estes modelos, mesmo não sabendo que dados pertencem a cada classe (Bilmes 1998). Este é um algoritmo iterativo que maximiza a esperança da verosimilhança dos dados completos, tendo em conta uma estimativa anterior e condicionado pelo conjunto de observações.

Numa fase inicial é necessário colocar as observações do conjunto de dados num *cluster*, que será discutido na secção 4.5. A partir da inicialização é possível avançar com o algoritmo. Este tem duas fases principais (Borman 2004):

Fase-E (expectativa): tendo a estimativa dos parâmetros obtida na iteração anterior, calcula-se a verosimilhança de cada uma das amostras para cada um dos *clusters* a partir da seguinte equação:

$$p(x_j | C_k, \varphi_k) = \pi_k \cdot \frac{1}{(2\pi_k)^{\frac{D}{2}} \cdot |\Sigma_k|^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}(x_j - \mu_k)^T \Sigma_k^{-1}(x_j - \mu_k)\right) \quad (4.15)$$

Isto dará como resultado as probabilidades *a posteriori* das amostras em cada um dos *clusters*.

Fase-M (maximização): tendo neste passo as k probabilidades *a posteriori* (número de classes) de cada uma das amostras, maximiza-se estes valores, atribuindo cada amostra ao *cluster* cuja verosimilhança é máxima. No final deste passo são calculadas novas estimativas para cada um dos parâmetros.

Depois da Fase-M, o algoritmo volta à Fase-E, até que se obtenha a convergência requerida e escolhida pelo utilizador na função verosimilhança.

4.5. Inicialização

A inicialização é um passo muito importante no algoritmo EM, já que o seu resultado é bastante dependente deste passo, especialmente no contexto de um problema multivariável, como neste caso. Como o objetivo é maximizar a função de verosimilhança do modelo, e este processo é iterativo, pode ser encontrado um máximo local que não corresponda ao requerido máximo global quando os parâmetros iniciais não são os mais corretos (Biernacki et al. 2003).

Vários métodos podem ser usados para fazer a inicialização dos parâmetros do modelo, sendo que o mais utilizado é o *k-means*. Este algoritmo tem como objetivo determinar um número k de pontos (chamados centros) que minimizam a distância euclidiana média quadrática entre cada um dos dados e o centro mais próximo. Este é inicializado ao escolher k pontos como centros iniciais, e em cada iteração calcula-se a distância aos centros atuais, atribuindo cada observação ao grupo que minimiza essa distância, calculando no final da iteração novos centros. O algoritmo é parado quando de uma iteração para a seguinte não haja alteração na classificação de nenhuma observação (Kanungo et al. 2002; Dicintio 2012). O funcionamento deste algoritmo pode ser visto no exemplo da Figura 4.2.

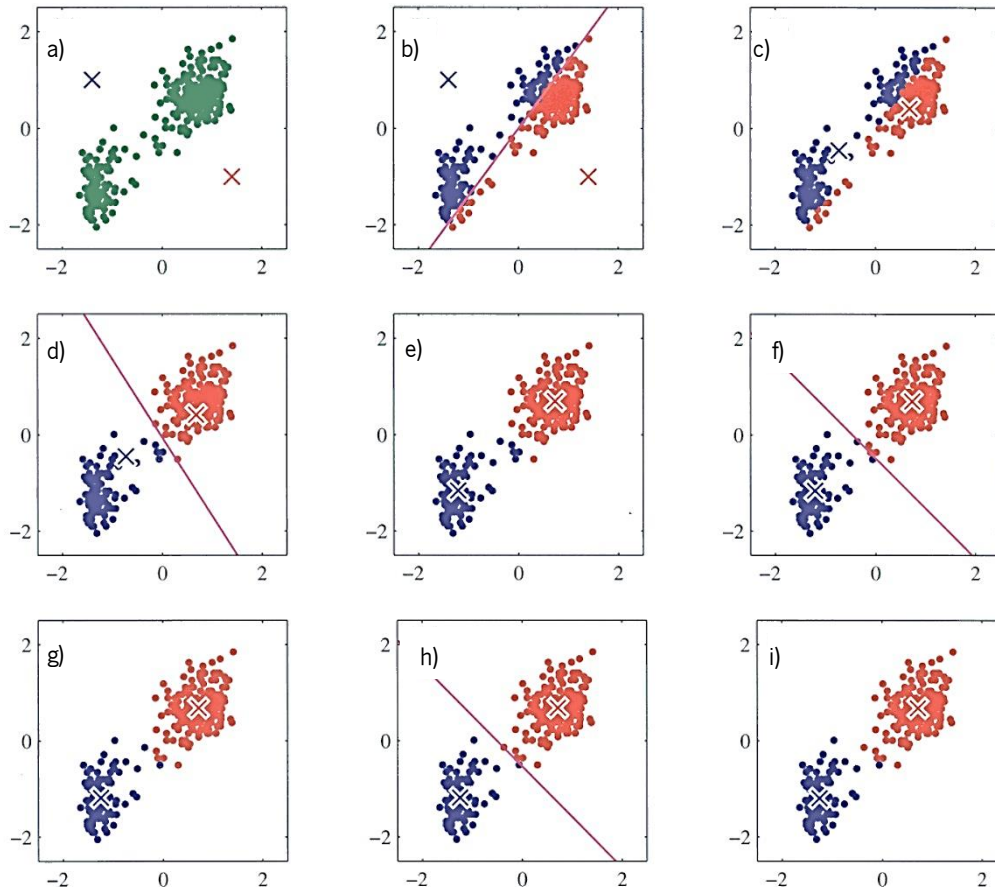


Figura 4.2. Exemplo do funcionamento do algoritmo *k-means*. a) pontos verdes é o grupo de dados que se pretende dividir, b) inicialização do algoritmo com as cruzes como centros, do c) ao h), iterações sucessivas do algoritmo com solução final em i). Adaptado de (Alpaydin 2010).

Como a inicialização do algoritmo *k-means* é aleatória, o resultado do mesmo também o será, no entanto a aleatoriedade é pouco significativa em especial porque representa apenas uma estimativa inicial a ser trabalhada pelo algoritmo EM. Deste modo a saída do EM não sofre, na maioria das aplicações práticas, alterações significativas que tenham sido propagadas desde a saída deste algoritmo.

4.6. Critério de Informação Bayesiano

Nas secções anteriores discutiu-se a forma de fazer a segmentação de uma *frame* da cápsula endoscópica em diversas regiões. Mas, para executar tanto o método EM como o método *k-means*, é necessário saber-se à *priori* o número de regiões efetivamente diferentes na imagem.

Em diagnóstico automático este número de regiões efetivamente diferentes numa imagem é normalmente desconhecido à *priori*, pelo que atribuindo-lhe um número ao acaso ou tendo por base algum critério heurístico terá sempre como resultado a obtenção de um modelo subótimo. Este processo pode ser aceitável para algumas aplicações como sejam aquelas baseadas na classificação de *frames* completas, no entanto quando o objetivo é comparar imagens consecutivas da CE, onde o número de regiões efetivamente diferentes varia de imagem para imagem, esta informação tem que ter uma precisão relativamente elevada. Deste modo será necessário implementar um método que permita saber qual o número mais provável de regiões efetivamente diferentes que existe em cada uma das imagens.

Num trabalho anterior, a comparação de modelos era feita tendo em conta apenas o valor da verosimilhança de cada um dos modelos (Vieira et al. 2012). Mas é preciso ter em conta que com o aumento do número de *clusters*, a dimensão do modelo aumenta substancialmente, causando um aumento da sua verosimilhança (Gupta et al. 2010). Assim, será necessário penalizar o aumento da dimensionalidade do problema. Um dos métodos que permite fazer isto tem o nome de *Bayesian Information Criterion* (BIC), que é, de entre os critérios bayesianos disponíveis, aquele que mais se adequa para ser usado com misturas gaussianas (Steele & Raftery 2009). É definido como:

$$BIC(G) = -2l(\varphi) + d \ln(n) \quad (4.16)$$

onde G é a mistura gaussiana respetiva e d o número de parâmetros dessa mistura. No caso de uma mistura gaussiana, o número de parâmetros é a soma do número de coeficientes de mistura, médias e elementos da matriz de covariância:

$$\#\pi = k - 1 \quad (4.17)$$

$$\#m = kD \quad (4.18)$$

$$\#\Sigma = k \frac{D(D+1)}{2} \quad (4.19)$$

$$d = k \left(\frac{D(D+1)}{2} + D - 1 \right) \quad (4.20)$$

Tendo em conta que o objetivo é minimizar o valor do BIC, então será necessário executar o algoritmo de segmentação com as possibilidades do número de componentes de misturas que

Critério de Informação Bayesiano

se considerarem, calcular o BIC para cada uma das misturas, e no final comparar os valores e encontrar o mínimo.

Capítulo 5 Reconstrução 3-D

Neste capítulo serão abordados os vários conhecimentos teóricos necessários para o desenvolvimento do algoritmo de reconstrução tridimensional de imagens 2-D, principalmente o método SfS. Para fazer a reconstrução tridimensional de objetos a partir de imagens, é necessário perceber a interação que as fontes de luz fazem com as superfícies, e assim poder determinar o mapa de profundidades de cada imagem. Assim, conhecendo a disposição espacial das luzes e da objetiva na cápsula, será possível inferir certos aspectos sobre a forma das superfícies intestinais.

5.1. Iluminação de superfícies

Quando uma superfície de um objeto tem uma luz a incidir sobre ela, uma imagem dessa superfície vai apresentar sombras (que representam variações da intensidade de luz refletida), podendo este efeito ser usado para determinar a forma do objeto (Horn 1970). Para isto é necessário saber o modelo de refletância da superfície em causa, que pode ser lambertiano, especular ou híbrido (Nayar et al. 1989).

Como pode ser visto na Figura 5.1, cada ponto reflete uma parte da luz, que será recebida pela câmara e assim usada para formar a imagem. A quantidade de energia que é refletida pelo ponto depende da orientação da superfície e a direção da fonte de luz nesse ponto, além das

propriedades de refletância (Horn 1970). Geralmente, as superfícies são descritas por um modelo híbrido, ou seja, uma componente lambertiana (I_L) e uma componente especular (I_S). Assim, a intensidade em cada ponto da imagem pode ser expresso como (Nayar et al. 1989):

$$I = I_L + I_S \quad (5.1)$$

A intensidade lambertiana representa a componente difusa da reflexão. As superfícies que apresentam características puramente lambertianas são aquelas que parecem igualmente brilhantes independentemente do ângulo de visualização do observador (Acharya & Ray 2005). Existem dois mecanismos que levam a uma reflexão lambertiana: o primeiro é quando os raios de luz que incidem na superfície são refletidos múltiplas vezes entre as ondulações da superfície antes de se espalharem para o espaço. Como estas reflexões são aleatórias, a energia incidente é distribuída em todas as direções, levando a uma reflexão difusa. O segundo mecanismo ocorre quando a luz penetra a superfície e devido a não-homogeneidades internas, os raios de luz são refletidos múltiplas vezes e refratados devidos a diferenças de índices de refração dentro da superfície. Como estes acontecimentos são aleatórios, a reflexão ocorre em todas as direções, considerando-se difusa (Nayar et al. 1989).

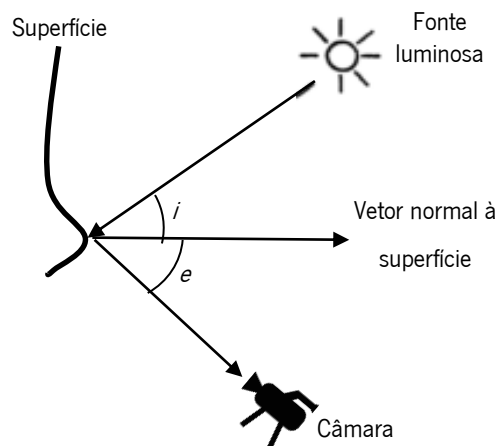


Figura 5.1. Definição de ângulo incidente (i) e ângulo de reflexão (e).

A quantidade de energia refletida será então proporcional à área de superfície (A) do elemento considerado visto da direção da câmara, e às características da superfície (albedo – ρ).

$$I_L = A\rho \cos \theta_i \quad (5.2)$$

A outra componente de reflexão que se verifica em certas superfícies é a especular. Esta ocorre quando o ângulo incidente é igual ao ângulo de reflexão, que ocorre principalmente em

superfícies espelhadas (Nayar et al. 1989). O resultado desta reflexão são duas componentes: um pico especular e um lobo especular. O pico tem intensidade zero em todas as direções exceto numa pequena área à volta da direção de reflexão especular. Enquanto isto, o lobo especular espalha-se à volta da direção de reflexão especular (Beckmann & Spizzichino 1968). A intensidade relativa destas componentes vai ser dependente tanto das propriedades da superfície, como da luz incidente (mais propriamente a rugosidade da superfície e o comprimento de onda da luz) (Nayar et al. 1989).

Existem diversos modelos para simular a reflexão especular. O mais simples usa a função impulso de Dirac:

$$I_s = B\delta(\theta_i - 2\theta_e) \quad (5.3)$$

Na Equação (5.3), B representa a força da componente especular (proporcional a σ/λ , que representam respetivamente a rugosidade da superfície e o comprimento de onda da luz incidente) e θ_i o ângulo entre a direção da fonte luminosa e a direção de visualização (Tsai & Shah 1994). Outros modelos mais complexos foram desenvolvidos mais tarde. Phong desenhou um modelo que usava a exponenciação do cosseno do ângulo de visualização para prever a reflexão especular, com a principal desvantagem de os parâmetros usados não terem significado físico (Phong 1975). Torrance e Sparrow, ao considerarem que a superfície é composta por pequenas facetas orientadas aleatoriamente com características de espelho, e que a reflexão especular é resultado do produto da energia da luz incidente, do coeficiente de Fresnel, da distribuição da orientação das facetas e um fator de atenuação geométrica (Torrance & Sparrow 1967). Um modelo derivado do Torrance-Sparrow tornava-o mais simples, considerando que a orientação das facetas podia ser modelada por uma distribuição gaussiana e o resto das variáveis como constantes (Healey & Binford 1988).

5.2. Shape from Shading

Existem quatro tipos de algoritmos de SfS: algoritmos de minimização, algoritmos de propagação, algoritmos locais e algoritmos lineares. Os algoritmos de minimização têm como objetivo minimizar uma função de energia ao longo de toda a imagem. Como existem diversas variáveis desconhecidas que tornam estas funções sistemas indeterminados, é necessário incluir certas restrições que podem ser de várias naturezas: restrições de iluminação, suavidade, gradiente de intensidades, entre outras. Os algoritmos de minimização são considerados bastante

mais lentos que os restantes, mas também mais robustos. Têm como principal desvantagem o uso de restrições que possibilitem a minimização do problema, que podem ser difíceis de determinar dependendo das aplicações (Zhang et al. 1999).

Os métodos de propagação têm como princípio a propagação da informação de forma de um conjunto de pontos para o resto da imagem. Uma das dificuldades do algoritmo é determinar os pontos iniciais que irão propagar a informação de forma. A maioria das vezes estes são os pontos que têm maior intensidade na imagem, mas nem sempre este corresponde ao melhor resultado (Zhang et al. 1999). O primeiro algoritmo de SfS, proposto por Horn (Horn 1970), é incluído neste grupo de métodos.

Os algoritmos locais, pertencentes ao terceiro grupo, derivam a forma das superfícies na imagem tendo em conta suposições sobre o tipo de superfícies presentes. São os menos utilizados, já que para cada aplicação é necessário um conhecimento grande sobre o tipo de superfícies que irão ser fotografadas (Zhang et al. 1999).

Em termos de algoritmos lineares, o objetivo será linearizar o mapa de refletância das superfícies presentes na imagem, fazendo posteriormente a computação da solução ótima a partir do novo mapa (Zhang et al. 1999). Foi escolhido um algoritmo deste grupo, nomeadamente o proposto por Tsai (Tsai & Shah 1994), já que apresenta bons resultados para superfícies esféricas (o que é mais comum no TGI), é bastante rápido em relação à maioria dos restantes algoritmos de SfS, e é bastante robusto para superfícies lambertianas e especulares (Zhang et al. 1999; Tsai & Shah 1994; Karargyris & Bourbakis 2011b).

Tendo em conta os conhecimentos da reflexão de luz nas várias superfícies descritos na secção anterior, é possível construir algoritmos com o objetivo de determinar a profundidade de cada ponto da superfície da parede intestinal.

Conhecendo a estrutura interna do dispositivo da cápsula endoscópica SB, é possível fazer um modelo da projeção de uma imagem usando este aparelho (Figura 5.2). Tendo em conta a disposição dos 4 LED's no plano da lente, é possível inferir que existe a contribuição da iluminação das quatro fontes para a reflexão de cada ponto da superfície. Assim, pode-se modelar a refletância a partir da intensidade de cada imagem segundo a seguinte expressão (Wu 2010):

$$E(x, y) = \sum_{i=1}^4 R_i(p, q) \quad (5.4)$$

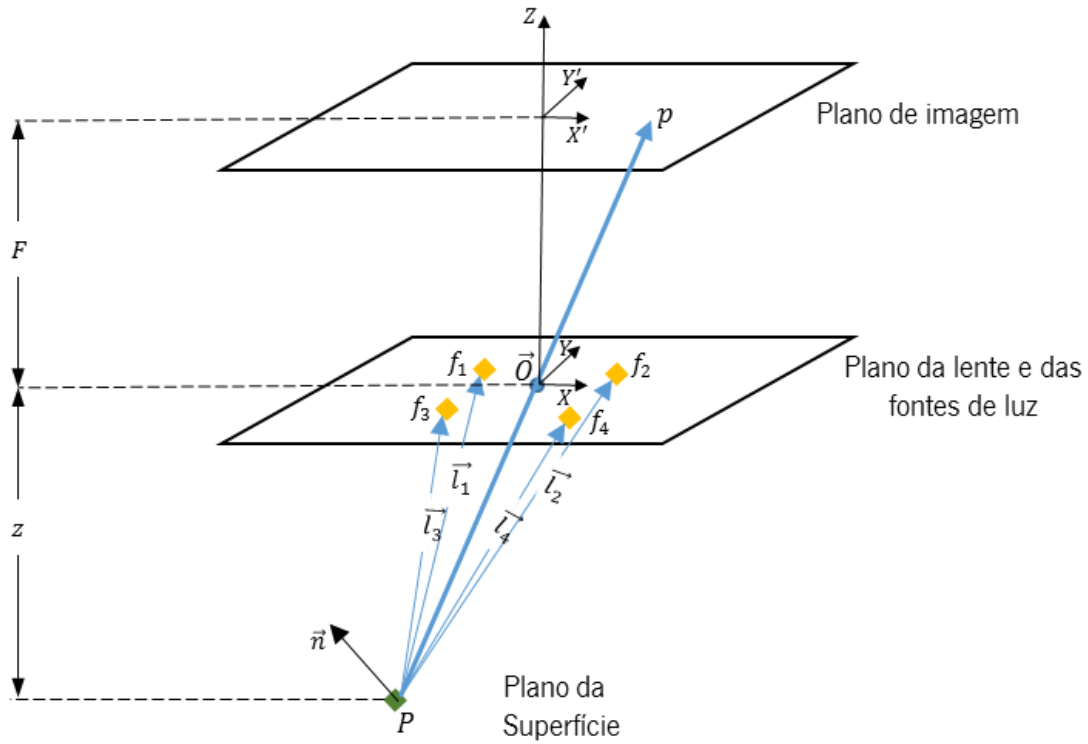


Figura 5.2. Um modelo da projeção de uma imagem usando uma cápsula endoscópica.

Tabela 5.1. Coordenadas de cada ponto do esquema da Figura 5.2.

| Variável | Designação | Coordenadas |
|-------------|--|---|
| P | Posição real na superfície intestinal | $[x, y, z] = \left[\tilde{x} \frac{z}{F}, \tilde{y} \frac{z}{F}, z \right]$ |
| P | Posição na imagem da câmara do ponto P | $[\tilde{x}, \tilde{y}, F]$ |
| f_i | Posição da fonte luminosa i | $[\pm a, \pm b, 0]$ |
| \vec{l}_i | Vetor do raio de luz proveniente da fonte de luz i | $[x \pm a, y \pm b, z]$ |
| \vec{O} | Posição da lente da cápsula endoscópica | $[0,0,0]$ |
| \vec{n} | Vetor normal à superfície no ponto P | $\frac{\left[-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1 \right]}{\sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 + 1}}$ |
| z | Profundidade do ponto P | — |
| F | Distância focal | — |

O termo $R_i(p, q)$ representa a função de refletância da fonte de luz i em cada ponto (x, y) . O modelo de refletância proposto por Tsai define apenas uma fonte de luz, sendo que o método aqui apresentado irá ser uma adaptação deste modelo para quatro fontes iguais, sendo que cada uma provoca uma refletância diferente tendo em conta os ângulos que faz com cada ponto e as derivadas parciais da profundidade – Equação (5.5) (Tsai & Shah 1994).

$$R_i(p, q) = \frac{1 + pp_{si} + qq_{si}}{\sqrt{1 + p^2 + q^2} \sqrt{1 + p_{si}^2 + q_{si}^2}} \quad (5.5)$$

As derivadas parciais podem ser aproximadas discretamente por:

$$p = \frac{\partial z}{\partial x} = z(x, y) - z(x - 1, y) \quad (5.6)$$

$$q = \frac{\partial z}{\partial y} = z(x, y) - z(x, y - 1) \quad (5.7)$$

As variáveis p_{si} e q_{si} podem ser definidas por:

$$p_{si} = \frac{\cos \tau_i \sin \sigma_i}{\cos \sigma_i} \quad (5.8)$$

$$q_{si} = \frac{\sin \tau_i \sin \sigma_i}{\cos \sigma_i} \quad (5.9)$$

Nas equações (5.8) e (5.9), os ângulos σ_i e τ_i representam os ângulo entre o ponto correspondente e a fonte de luz i (Chantler et al. 1994), representados na seguinte figura:

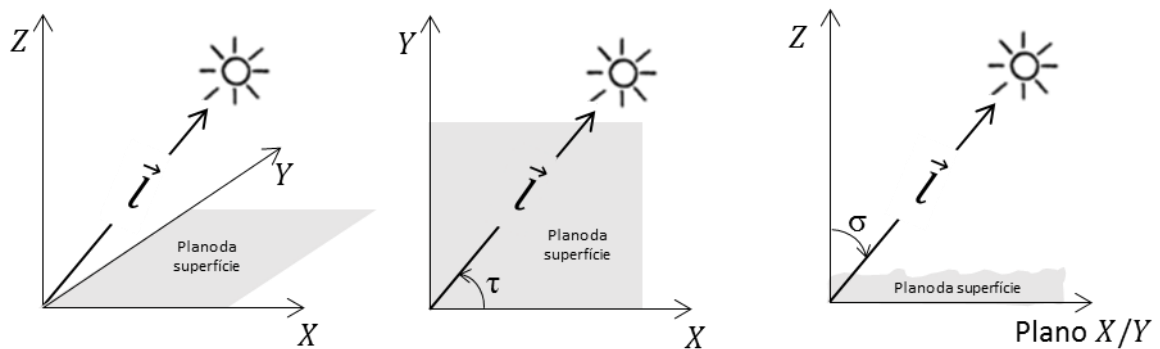


Figura 5.3. Representação dos ângulos τ (centro) e σ (direita) tendo em conta uma fonte luminosa arbitrária no espaço (esquerda).

Assim, substituindo as equações (5.8) e (5.9) na equação (5.5), fica-se com a expressão seguinte:

$$R_i(p, q) = \frac{\cos \sigma_i + p \cos \tau_i \cos \sigma_i + q \sin \tau_i \sin \sigma_i}{\sqrt{1 + p^2 + q^2}} \quad (5.10)$$

Ao incluir todas as fontes luminosas na equação geral que relaciona o modelo da refletância à intensidade da imagem, tem-se:

$$E(x, y) = \sum_{i=1}^4 \frac{\cos \sigma_i + p \cos \tau_i \cos \sigma_i + q \sin \tau_i \sin \sigma_i}{\sqrt{1 + p^2 + q^2}} \quad (5.11)$$

Pode-se reescrever a equação geral do modelo da refletância como:

$$0 = E(x, y) - \sum_{i=1}^4 R_i(p, q) \quad (5.12)$$

$$0 = f(E(x, y), z(x, y), z(x-1, y), z(x, y-1)) \quad (5.13)$$

A partir da expansão em séries de Taylor (apenas até aos termos de primeira ordem), da função f (Equação (5.13)), tendo em conta um mapa de profundidades da superfície em causa z^{n-1} tem-se:

$$\begin{aligned} 0 &= f(E(x, y), z(x, y), z(x-1, y), z(x, y-1)) \\ 0 &\approx f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\ &+ (z(x, y) - z^{n-1}(x, y)) \frac{\partial}{\partial z(x, y)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\ &+ (z(x-1, y) - z^{n-1}(x-1, y)) \frac{\partial}{\partial z(x-1, y)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\ &+ (z(x, y-1) - z^{n-1}(x, y-1)) \frac{\partial}{\partial z(x, y-1)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \end{aligned} \quad (5.14)$$

A Equação (5.14) pode ser reescrita, ao desenvolver os vários termos e mudando de membros os termos relacionados com a iteração $n-1$ e a profundidade real.

$$\begin{aligned}
 & z(x, y) \times \frac{\partial}{\partial z(x, y)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\
 & + z(x-1, y) \times \frac{\partial}{\partial z(x-1, y)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\
 & + z(x, y-1) \times \frac{\partial}{\partial z(x, y-1)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\
 & = -f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \tag{5.15} \\
 & + z^{n-1}(x, y) \times \frac{\partial}{\partial z(x, y)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\
 & + z^{n-1}(x-1, y) \times \frac{\partial}{\partial z(x-1, y)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\
 & + z^{n-1}(x, y-1) \times \frac{\partial}{\partial z(x, y-1)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1))
 \end{aligned}$$

É possível escrever em forma de vetor:

$$\begin{pmatrix} 0 & \dots & a_{x,y-1} & 0 & \dots & a_{x-1,y} & a_{x,y} & 0 & \dots \end{pmatrix} \cdot \begin{pmatrix} z_{1,1} \\ \dots \\ z_{x-1,y} \\ z_{z,y-1} \\ \dots \\ z_{N,N} \end{pmatrix} = b_{x,y} \tag{5.16}$$

Na Equação (5.16), tem-se que:

$$a_{x,y} = \frac{\partial}{\partial z(x, y)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \tag{5.17}$$

$$\begin{aligned}
 b_{x,y} & = -f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\
 & + z^{n-1}(x, y) \times \frac{\partial}{\partial z(x, y)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\
 & + z^{n-1}(x-1, y) \times \frac{\partial}{\partial z(x-1, y)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1)) \\
 & + z^{n-1}(x, y-1) \times \frac{\partial}{\partial z(x, y-1)} f(E(x, y), z^{n-1}(x, y), z^{n-1}(x-1, y), z^{n-1}(x, y-1))
 \end{aligned} \tag{5.18}$$

A Equação (5.16) é relativa a um ponto com coordenadas genéricas (x, y) . Quando estas expressões são aplicadas a uma imagem $N \times N$, tem-se uma equação da forma $A.z = B$, em que A é uma matriz $N^2 \times N^2$ e z e B vetores de tamanho $N^2 \times 1$. Basta imaginar uma imagem comum

com tamanho de 256 x 256, e para descobrir os valores de z , será necessário calcular a inversa de uma matriz de tamanho 65536 x 65536, o que é computacionalmente bastante pesado.

Assim, Tsai usa o método iterativo de Jacobi para resolver o sistema mencionado acima. Este método consiste em estimar o valor da profundidade $z(x,y)$, a partir de uma aproximação inicial $z^0(x,y)$, resolvendo iterativamente o sistema linear de equações $A.z = B$ (Saad 2003). Assim, o valor da profundidade na iteração n pode ser calculado a partir de todos os valores do mapa de profundidades da iteração $n-1$, exceto o do próprio ponto. É então possível substituir na Equação (5.15) os termos $z^{n-1}(x-1, y)$ e $z^{n-1}(x, y-1)$ por $z(x-1, y)$ e $z(x, y-1)$, respetivamente. Esta substituição leva a que o terceiro e quarto termos do membro da direita da Equação (5.15) sejam anulados. Esta equação fica assim numa forma mais reduzida:

$$\begin{aligned} 0 &= f(z(x, y)) \\ 0 &\approx f(z^{n-1}(x, y)) + (z(x, y) - z^{n-1}(x, y)) \frac{\partial}{\partial z(x, y)} f(z^{n-1}(x, y)) \end{aligned} \quad (5.19)$$

Assim, na iteração número n , tem-se $z(x, y) = z^n(x, y)$, logo pode-se reescrever a Equação (5.19):

$$z^n(x, y) = z^{n-1}(x, y) + \frac{-f(z^{n-1}(x, y))}{\frac{\partial}{\partial z(x, y)} f(z^{n-1}(x, y))} \quad (5.20)$$

onde o valor da derivada parcial de $f(z^{n-1}(x, y))$ em relação a $z(x, y)$ é dado por:

$$\frac{\partial}{\partial z(x, y)} f(z^{n-1}(x, y)) = \sum_{i=1}^N -1 \times \left(\frac{p_{si} + q_{si}}{\sqrt{p^2 + q^2 + 1} \sqrt{p_{si}^2 + q_{si}^2 + 1}} - \frac{(p+q)(pp_{si} + qq_{si} + 1)}{\sqrt{(p^2 + q^2 + 1)^3} \sqrt{p_{si}^2 + q_{si}^2 + 1}} \right) \quad (5.21)$$

Em cada iteração é então apenas necessário calcular o valor de $f(z^{n-1}(x, y))$ e o valor da primeira derivada dessa função, $\frac{\partial}{\partial z(x, y)} f(z^{n-1}(x, y))$; tendo em conta uma estimativa de $z^0(x, y)$ para todos os pixels da imagem. Dependendo da forma da superfície em causa, o valor da Equação (5.21) pode-se tornar igual a zero, o que causará a divisão por zero no cálculo da profundidade na Equação (5.20), levando a uma indeterminação no cálculo desta variável. Para

resolver este problema e impedir que esta divisão aconteça, é necessário fazer certas modificações, tornando a Equação (5.20) na seguinte forma:

$$z^n(x, y) = z^{n-1}(x, y) + K^n \left(-f \left(z^{n-1}(x, y) \right) \right) \quad (5.22)$$

Na equação acima, o valor K^n terá que satisfazer três condições para cumprir o desejado. A primeira é que terá que ser aproximadamente igual ao inverso de $\frac{\partial}{\partial z(x, y)} f \left(z^{n-1}(x, y) \right)$. A

segunda é que quando o valor de $\frac{\partial}{\partial z(x, y)} f \left(z^{n-1}(x, y) \right)$ se aproxima de zero, então K^n será igual a zero. A terceira e última é que quando $z^n(x, y)$ se aproxima do verdadeiro valor de $z(x, y)$, K^n deverá ser igual a zero. Assim, K^n será definido da seguinte forma:

$$K^n = \frac{S^n(x, y)M(x, y)}{W + S^n(x, y)M^2(x, y)} \quad (5.23)$$

onde:

$$M(x, y) = \frac{\partial}{\partial z(x, y)} f \left(z^{n-1}(x, y) \right) \quad (5.24)$$

$$S^n(x, y) = E \left\{ \left(z^n(x, y) - z(x, y) \right)^2 \right\} \quad (5.25)$$

Nas expressões acima, $E\{\}$ é o operador esperança e W é um valor pequeno, mas não nulo. Assim sendo, pode-se concluir o seguinte:

- K^n será aproximadamente igual a $\frac{1}{M(x, y)}$, que é o inverso de

$$\frac{\partial}{\partial z(x, y)} f \left(z^{n-1}(x, y) \right).$$

- Quando o valor de $M(x, y)$ se aproxima de zero, também K^n será aproximadamente igual a zero.
- Quando $z^n(x, y)$ se aproximar de $z(x, y)$, $S^n(x, y)$ irá ser igual a zero, logo K^n será também nulo e $z^n(x, y)$ igual a $z^{n-1}(x, y)$.

Assim, todas as condições são cumpridas aplicando estas expressões. E, ao analisar a Equação (5.23), pode-se concluir que a solução deste problema se assemelha à implementação de um filtro de Kalman para a resolução dos valores de profundidade. Este algoritmo é usado geralmente para obter estimativas de variáveis desconhecidas a partir de um conjunto de observações que contêm ruído (Maybeck 1982).

Algumas especificidades da implementação deste algoritmo serão discutidas no Capítulo 7.

Capítulo 6 Deteção de tumores

A par do objetivo principal deste trabalho (reconstrução tridimensional do tecido intestinal), foi desenvolvido também um algoritmo com o principal objetivo de fazer a deteção automática de tumores em *frames* da CE usando a segmentação necessária ao processo de reconstrução para seleção de regiões de interesse. Apesar de muitos estudos em relação à segmentação de vídeo, deteção automática de *frames* anormais e mais especificamente, de tumores; a maioria usa toda a imagem para classificar as *frames*. Em (Mackiewicz et al. 2008), efetivamente é realizada a segmentação de *frames*, mas esta é feita descartando sempre uma parte exterior das imagens que contém tecido (e que por vezes pode ser relevante), e além disso o resultado é sempre uma segmentação grosseira da imagem porque a imagem é dividida apenas em 12 blocos de modo rígido.

Esta abordagem de divisão da imagem por blocos é importada do processamento de imagens de colonoscópio convencional que apresenta dimensão da ordem de pelo menos 16 vezes a dimensão das imagens de CE. A partição de imagens grandes em blocos, mesmo de forma rígida, apresenta vantagens significativas ao nível do processamento e evita que características locais possam ser diluídas por toda a imagem. No entanto não é desta forma que o médico observa uma imagem, partindo-a rigidamente em partes e observando cuidadosamente cada uma delas, mesmo que nenhuma característica relevante pareça à primeira vista estar presente. O médico

parte de facto a imagem não de forma rígida mas concentrando-se em regiões localizadas mais ou menos aleatoriamente mas que apresentam padrões que diferem dos padrões conhecidos do tecido normal. É esta abordagem que é seguida no âmbito deste trabalho e para a qual é apresentada uma implementação.

6.1. Correção luminosa

Num estudo anterior (Vieira et al. 2012), chegou-se à conclusão que a segmentação de *frames* de vídeos retirados de cápsula endoscópica, sem qualquer processamento, leva a uma incorreta seleção de regiões que correspondem a diferentes tipos de tecidos. Isto acontece porque quando a cápsula endoscópica está perpendicular e muito perto da parede intestinal, então a zona central da imagem terá uma intensidade maior, devido ao aumento da luminosidade desta região. Na Figura 6.1 é possível ver o exemplo de uma *frame* em que isto ocorre.

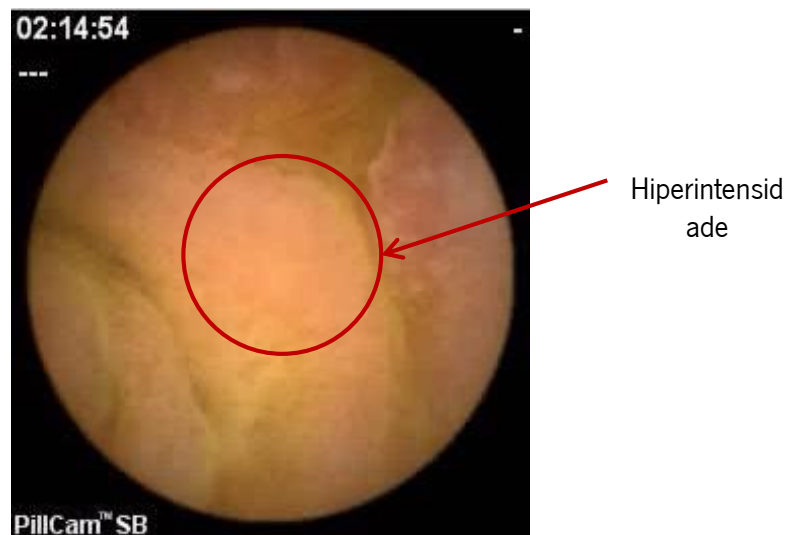


Figura 6.1. Exemplo de *frame* que tem uma zona de hiperintensidade no seu centro.

A solução desenvolvida anteriormente tinha como objetivo a seleção de tecido tumoral em *frames* em que se soubesse de partida que este tipo de tecido existia. A principal aplicação para este trabalho era a de treino médico na deteção de tecido tumoral em *frames* da CE, ao contrário da que foi desenvolvida aqui, que é a deteção automática da presença de tumor em *frames* da cápsula endoscópica. Este trabalho consistia na execução do algoritmo de segmentação em 2 e 3 regiões, comparando depois os resultados através do valor da verosimilhança. Tendo em conta que as imagens teriam, em condições normais, dois tipos de tecidos (tumoral e normal), se o melhor modelo fosse aquele com duas regiões, significava que não havia hiperintensidade; mas

se fosse o caso contrário, era sinal que havia uma terceira zona na imagem, que seria a zona de hiperintensidade.

Devido aos resultados não serem os esperados em todas as imagens com o método anterior, outra metodologia foi proposta para ultrapassar o problema do aumento de iluminação em certas zonas da imagem.

6.1.1. Espaços de cor

Um espaço de cor é uma forma de representar matematicamente uma imagem a cores (Ford & Roberts 1998). O mais utilizado e conhecido é o chamado RGB (*Red, Green, Blue*). Tal como o nome deste espaço de cor indica, cada cor irá ser representada por três números, que correspondem à quantidade de cada uma das cores primárias (vermelho, verde e azul) presentes nessa cor. Apesar de ser muito utilizado em sistemas de aquisição de imagens, não tem em conta a percepção da visão humana em relação à distinção de cores (Pascale 2003).

Esta percepção é formada por três atributos diferentes: a luminosidade, o tom (ou matiz) e a saturação (Hunt & Pointer 2011). O primeiro atributo define se uma cor parece ter mais ou menos luz, podendo definir regiões mais claras ou mais escuras. O segundo é definido como as pareências ou diferenças que a cor tem com o azul, vermelho, amarelo e verde. Por último, a saturação define o quanto uma área exibe mais ou menos tom (Ford & Roberts 1998; Hunt & Pointer 2011).

Para se melhorar a definição de um espaço de cor de acordo com a percepção que o ser humano tem das cores, foi desenvolvido o espaço HSV (*Hue, Saturation, Value*). Este espaço de cor tem três atributos (tom, saturação e valor), construídos de forma a que cada um deles tivesse as características dos atributos da percepção humana das cores (Ford & Roberts 1998). O espaço HSV é muito usado em ferramentas de processamento de imagem, quando o objetivo é automatizar um método que é possível ser feito através do olho humano. Apesar de este ser um espaço de cor com grandes potencialidades aplicacionais, e o canal V ter em conta os níveis de iluminação da região, prova-se que não existe independência suficiente entre os canais deste espaço para se poder considerar apenas os canais H e S sem se perder informação de cor (Woodland & Labrosse 2005).

O espaço CIE $L^*a^*b^*$ (Lab) foi criado com o objetivo de uniformizar os espaços de cor, encontrando um que fosse completo o suficiente para poder definir todas as cores existentes, e que cada um dos canais tivesse significado (McLaren 2008). Neste espaço de cor, o canal L

representa a luminosidade, e varia entre os valores 0 (preto) e 100 (branco difuso), enquanto os canais a e b representam as componentes de cores opostas. Mais especificamente, o canal a define a posição da cor entre o verde (valores negativos) e o magenta (valores positivos), e que o canal b define a posição da cor entre o azul (valores negativos) e o amarelo (valores positivos) (Ford & Roberts 1998). É importante referir que nenhum dos canais que contêm a informação de cor tem limites teóricos para os seus valores, ao contrário da luminância. No caso deste espaço de cor, o canal L é independente dos restantes canais de cor, logo pode-se concluir que é possível retirar o canal L sem se perder informação de cor (Woodland & Labrosse 2005).

Para se fazer a conversão do espaço de cor original das imagens (RGB) para o espaço Lab é necessário fazer duas transformações: a primeira tem como objetivo transformar o espaço RGB para o espaço XYZ, enquanto que a segunda tem como resultado final o espaço Lab (Wyszecki & Stiles 2000). As transformações são obtidas usando as seguintes equações:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (6.1)$$

$$L = 116f_y - 16 \quad (6.2)$$

$$a = 500(f_x - f_y) \quad (6.3)$$

$$b = 200(f_y - f_z) \quad (6.4)$$

$$f_x = \begin{cases} \sqrt[3]{x_r} & x_r > \varepsilon \\ \frac{\kappa x_r + 16}{116} & x_r \leq \varepsilon \end{cases} \quad (6.5)$$

$$f_y = \begin{cases} \sqrt[3]{y_r} & y_r > \varepsilon \\ \frac{\kappa y_r + 16}{116} & y_r \leq \varepsilon \end{cases} \quad (6.6)$$

$$f_z = \begin{cases} \sqrt[3]{z_r} & z_r > \varepsilon \\ \frac{\kappa z_r + 16}{116} & z_r \leq \varepsilon \end{cases} \quad (6.7)$$

$$x_r = \frac{X}{X_r} \quad y_r = \frac{Y}{Y_r} \quad z_r = \frac{Z}{Z_r} \quad (6.8)$$

Na Figura 6.2 é possível ver uma comparação entre os três espaços de cor que foram mencionados neste texto, onde é possível perceber as principais características de cada um deles. À esquerda é possível observar que o espaço RGB pode ser representado por um cubo, e que na ponta de cada um dos seus vértices que estão em cima dos eixos estão posicionadas as cores: vermelho, verde e azul. É possível perceber também a dificuldade em separar a informação de luz da informação de cor neste espaço de cor, já que não existe uma divisão aparente entre elas. A meio da imagem (relativa ao espaço HSV) pode-se ver que realmente o tom (H), equivalente ao ângulo no cone apresentado representa a cor principal apresentada e a saturação (S), que equivale à distância ao centro representa a pureza dessa cor. Em relação ao valor (V), realmente não representa a luminosidade em termos perceptuais, já que se pode ver que transporta alguma informação de cor já que com o máximo valor, existem ainda bastantes cores que podem ser visualizadas. Em relação ao espaço de cor da direita (o Lab), quando se atinge o máximo no canal L, o resultado é o branco difuso e no caso contrário, a cor preta, o que será o esperado num canal que represente perfeitamente a luminosidade de uma imagem.

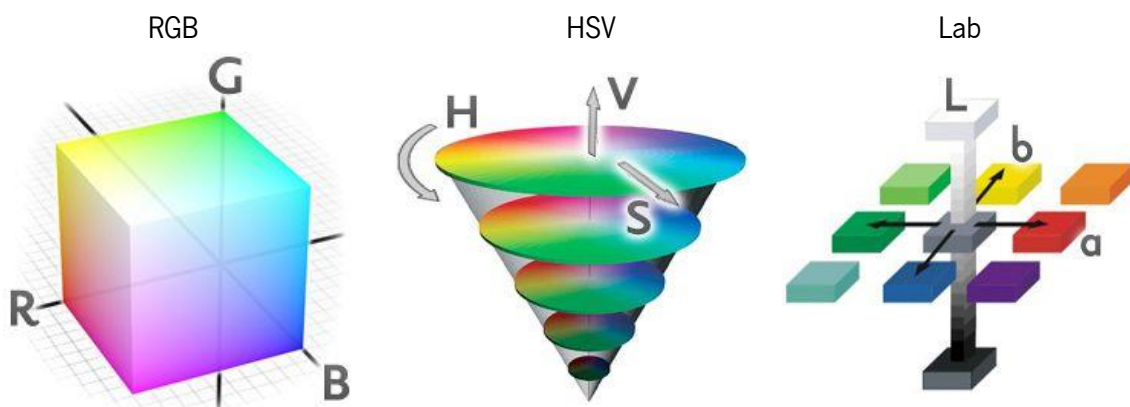


Figura 6.2. Esquema que revela como os atributos de cada um dos espaços de cor estão relacionados entre si. Adaptado de (Anon 2010).

6.1.2. Remoção da luminosidade

Como foi dito anteriormente, o espaço de cor Lab tem o seu primeiro canal que contém a informação da iluminação da região, e os dois últimos canais que contém toda a informação de

cor. Usando a *frame* da Figura 6.1 como referência foram feitas algumas alterações no domínio deste espaço de cor para mostrar a importância do canal L nos casos em que existe saturação de luz numa parte do tecido intestinal.

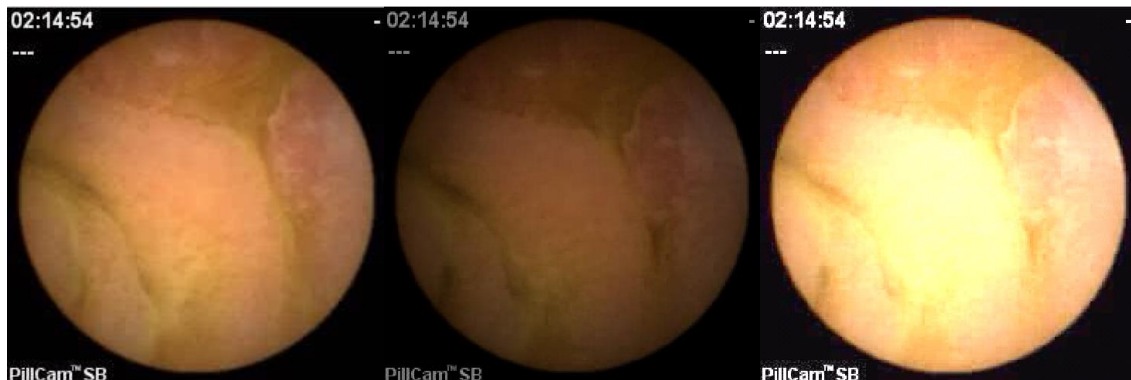


Figura 6.3. *Frame* da cápsula endoscópica que apresenta hiperintensidade (esquerda). Mesma imagem com canal L menos 20 unidades (centro) e canal de luminância mais 20 unidades (direita).

Na Figura 6.3 pode-se ver o resultado destas operações. Na imagem central foi retirado o valor de 20 a todos os píxeis no canal L, e na imagem da direita foi adicionado o valor de 20 a todos os píxeis no mesmo canal. É possível verificar que com a diminuição do valor no canal L é possível ver uma maior homogeneidade na continuação do tecido, o que é bastante piorado com o aumento do valor de L na imagem da direita. Assim sendo, tirando a informação da iluminação e usando apenas a informação de cor codificada nos canais a e b pode melhorar a capacidade de diferenciar os vários tipos de regiões numa *frame* da cápsula endoscópica.

Não é possível mostrar a imagem anterior sem o canal L porque em termos visuais é também a iluminação que nos ajuda a ter percepção das cores, mas em termos matemáticos e computacionais é possível retirar informações importantes das imagens, sem o conhecimento da luminosidade nestas.

6.2. Extração de *features*

Uma região de uma imagem de tecido intestinal pode ser caracterizada por vários atributos, sendo que estes podem ter a sua natureza na cor, na textura, na forma ou no tamanho (Karargyris & Bourbakis 2009). Apesar de poderem ser usadas *features* destas quatro naturezas para caracterizar tipos de tecido, no contexto da deteção de tumores, *features* de tamanho e forma não serão as melhores, já que não são características relevantes destes problemas.

6.2.1. Informação de cor

Na secção anterior discutiu-se como o espaço de cor Lab e a remoção do canal L pode ser importante para a caracterização de tecidos sem que a intensidade de luz dos LED's da cápsula endoscópica interfira nessa caracterização. Como no método de segmentação é usado um modelo probabilístico para a classificação de cada um dos pixels, é plausível concluir que medidas estatísticas retiradas a partir do histograma de cada um dos canais são apropriadas para caracterizar uma região do tecido intestinal. Existem muitas medidas estatísticas que podem ser usadas para caracterizar uma distribuição, como por exemplo a média, variância, entropia, curtose, energia, entre outros (Srinivasan & Shobha 2008).

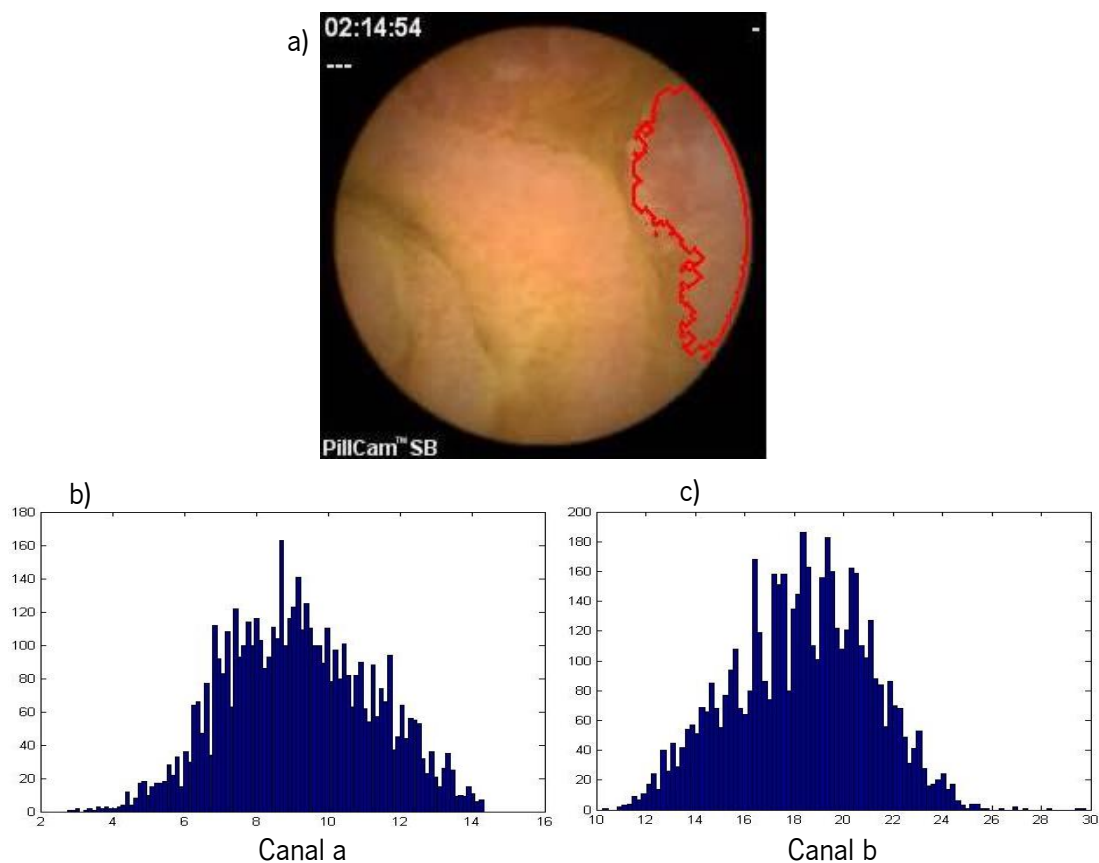


Figura 6.4. Histogramas dos canais de cor *a* e *b* de uma região de uma *frame* da cápsula endoscópica.

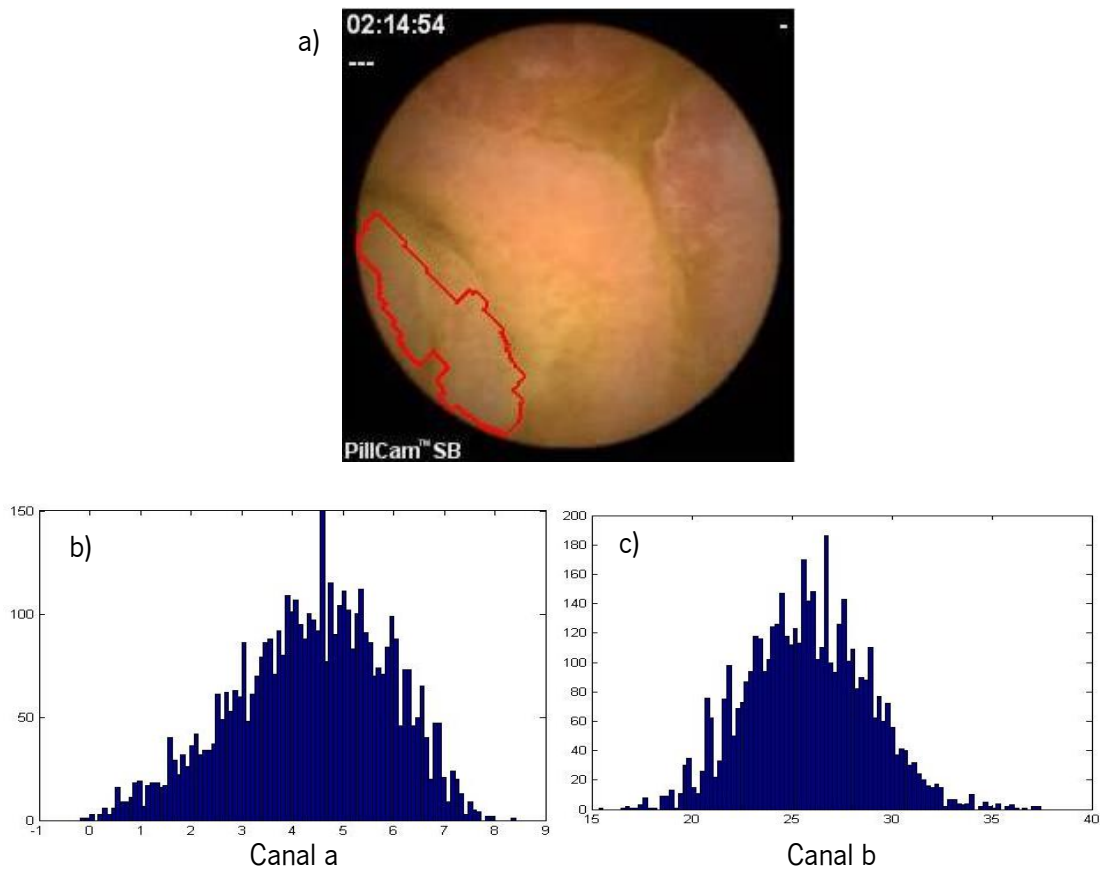


Figura 6.5. Histogramas dos canais de cor *a* e *b* de uma região de uma *frame* da cápsula endoscópica.

Na Figura 6.4 e Figura 6.5 é possível ver os histogramas dos dois canais de cor de duas regiões diferentes de uma *frame* da cápsula endoscópica. Ao comparar os histogramas entre cada uma das regiões, uma das coisas que é possível verificar é a diferença entre as médias das distribuições. Na região da Figura 6.4 vê-se que o histograma do canal *a* está mais à direita e o do canal *b* está mais à esquerda em relação à região da Figura 6.5. Outra das diferenças entre os histogramas das diferentes regiões é a largura destes. No caso da primeira região é facilmente perceptível que o histograma do canal *a* é mais “largo” e que o histograma do canal *b* é mais “estrito”. É possível verificar também a diferente forma das distribuições, principalmente na zona direita dos histogramas.

Estas observações sugerem que boas medidas estatísticas de dispersão para caracterizar a cor de regiões podem ser a média (descreve a centralidade da distribuição), variância e entropia (descrevem a largura e dispersão da distribuição) e a curtose e percentil de 95% da função cumulativa (descrevem a forma das distribuições). As equações que permitem calcular as primeiras quatro medidas são as seguintes:

$$\mu_t = E\{X_t\} = \frac{1}{N_t} \sum_i x_i \quad (6.9)$$

$$\sigma_t^2 = E\{(X_t - \mu_t)^2\} = \frac{1}{N_t} \sum_i (x_i - \mu_t)^2 \quad (6.10)$$

$$H_t = E\{-\ln p(X_t)\} = -\sum_k p_k \ln p_k \quad (6.11)$$

$$k_t = \frac{E\{(X_t - \mu_t)^4\}}{\sigma_t^4} = \frac{\frac{1}{N_t} \sum_i (x_i - \mu_t)^4}{\left(\frac{1}{N_t} \sum_i (x_i - \mu_t)^2\right)^2} \quad (6.12)$$

Para calcular o percentil de 95% primeiro é necessário calcular a posição do percentil em relação à população:

$$n_{95} = \frac{95}{100} \times N + \frac{1}{2} \quad (6.13)$$

Arredonda-se ao número inteiro mais próximo, e depois verifica-se qual o valor que corresponde à posição n_{95} na população X ordenada em forma crescente.

6.2.2. Informação de textura

A textura, apesar de ser fácil de reconhecer, é um termo difícil de definir e que na maioria das vezes é definida de forma diferente por cada autor, tendo em conta a aplicação particular onde este a irá usar. Mas um conceito que é constante em todas elas é que a textura está sempre relacionada com a variabilidade espacial nas intensidades dos pixels (Tuceryan & Jain 1998).

Dentro dos métodos usados para definir diferentes texturas numa imagem, os mais usados são aqueles que usam transformadas (wavelets e curvelets), e também descritores estatísticos (como as *features* de Haralick).

6.2.2.1. Transformada *Wavelet*

As *wavelets* são usadas para preencher um espaço no processamento de sinal em que o uso da Transformada de Fourier (TF) não permite obter os mesmos resultados. A TF ($F(\omega)$) de uma função $f(t)$ é dada pela seguinte expressão:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt \quad (6.14)$$

Com a observação da expressão (6.14), consegue-se observar que $F(\omega)$ terá como resultado coeficientes que representam a presença das frequências em todo o sinal. Assim, quando existe algum fenômeno que representa uma mudança abrupta no sinal num curto espaço de tempo, esta informação é difícil de detetar usando esta transformada. Pode-se então concluir que a TF tem uma maior utilidade quando é usada para sinais estacionários (Oppenheim et al. 1983).

É também possível fazer a truncagem dos sinais no domínio temporal e analisar cada um destes segmentos no domínio da transformada. Esta ferramenta tem o nome de *Short-Time Fourier Transform* (STFT). Este operador é caracterizado pela seguinte equação:

$$F(\tau, \omega) = \int_{-\infty}^{+\infty} f(t)w(t-\tau)e^{-j\omega t} dt \quad (6.15)$$

onde $w(t)$ é uma função janela com uma largura fixa. Esta ferramenta possibilita uma análise de frequências em vários segmentos de um sinal, mas ao necessitar da definição de uma largura fixa, torna impossível a análise do sinal sob várias resoluções (Jin et al. 2005).

O desenvolvimento das *wavelets* possibilitou ultrapassar esta dificuldade das resoluções fixas da STFT. Usar este tipo de transformadas permite fazer representações de tempo e frequência a partir de dados como sinais ou imagens. Estas são funções matemáticas que segmentam os dados nas suas diferentes componentes de frequência fazendo depois a análise de cada uma das componentes com uma resolução correspondente à sua escala (Cohen & Kovacevic 1996). O uso de *wavelets* permite fazer uso do princípio de limite de Gabor da melhor maneira (que define que não é possível ter, simultaneamente, uma boa resolução no tempo e na frequência) (Soares 2006). Isto acontece porque permite uma boa resolução no domínio das frequências para as frequências mais baixas, e uma maior resolução temporal para as frequências mais altas. Devido a estas características, esta ferramenta tem as melhores aplicações para sinais com componentes de frequência baixa de longa duração e de frequências altas de curta duração (Jin et al. 2005).

Uma *wavelet* é definida a partir de um espaço constituído por funções ortonormais entre si: a função *wavelet* mãe ($\psi(t)$) e a função de escala mãe ($\varphi(t)$). O conjunto de funções *wavelet* e de funções de escala pode ser descrito a partir das respetivas funções mãe:

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k) \quad (6.16)$$

$$\varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k) \quad (6.17)$$

onde k determina a posição da função e j a sua largura. As várias funções de escala podem ser descritas por um subespaço de cada resolução (V_j). Assim, cada subespaço de resolução (j) está contido no subespaço de maior resolução consecutivo (Cohen & Kovacevic 1996).

As funções *wavelet* de cada escala estão também contidas num subespaço W_j . Os espaços das funções de escala e das funções *wavelet* podem ser relacionados de forma que cada espaço W_j é o complemento ortogonal de V_j em relação ao subespaço V_{j+1} . Assim, como as funções *wavelet* pertencem ao conjunto constituído pelas funções de escala da resolução superior, então estas podem ser representadas pela soma ponderada de funções escala com o dobro da sua resolução:

$$\psi(x) = \sum_n h_\psi(x) \sqrt{2} \varphi(2x - n) \quad (6.18)$$

onde $h_\psi(x)$ contém os coeficientes da função *wavelet*. Para o caso da utilização de transformadas *wavelet* numa imagem, é necessário definir a Transformada Discreta *Wavelet* (DWT). Os coeficientes de aproximação e de detalhe são dados, respetivamente, por:

$$W_\varphi(j_0, k) = \frac{1}{\sqrt{M}} \sum_n f[n] \varphi_{j_0, k}[n] \quad (6.19)$$

$$W_\psi(j, k) = \frac{1}{\sqrt{M}} \sum_n f[n] \psi_{j, k}[n] \quad (6.20)$$

onde o valor de M corresponde ao número de amostras igualmente espaçadas no intervalo de suporte das funções de base. É possível ainda, a partir das equações (6.19) e (6.20) obter uma relação entre os coeficientes *wavelet* de escalas consecutivas:

$$W_\varphi(j, k) = \sum_m h_\varphi(m - 2k) W_\varphi(j + 1, m) \quad (6.21)$$

$$W_\psi(j, k) = \sum_m h_\psi(m - 2k) W_\psi(j + 1, m) \quad (6.22)$$

É então possível perceber que o cálculo dos coeficientes de uma escala é feito com a convolução dos coeficientes de uma escala posterior, com os coeficientes da função *wavelet* com

um espaçamento de $2k$. Assim, para obter os coeficientes de novas escalas, será necessário fazer a decimação das escalas anteriores, depois de passar pelos filtros (de aproximação ou detalhe).

Quando os dados que se querem analisar são imagens, é necessário usar *wavelets* com duas dimensões, e não com uma dimensão como foi explicado anteriormente. Uma *wavelet* com duas dimensões é apenas uma extensão das *wavelets* 1-D, já que os filtros usados para fazer esta transformada podem ser calculados a partir de filtros com uma dimensão. Os coeficientes de detalhe, quando se faz a análise com duas dimensões, têm agora também uma componente direcional, neste caso: horizontal, vertical e diagonal. O resultado são quatro funções, já que o resto dos produtos que podem ser feitos têm um resultado de uma só dimensão (Gonzalez & Woods 2008).

$$\varphi(x, y) = \varphi(x)\varphi(y) \quad (6.23)$$

$$\psi^H(x, y) = \psi(x)\varphi(y) \quad (6.24)$$

$$\psi^V(x, y) = \varphi(x)\psi(y) \quad (6.25)$$

$$\psi^D(x, y) = \psi(x)\psi(y) \quad (6.26)$$

Facilmente se chega, fazendo uma extensão das funções base 1-D, às equações que representam estes conjuntos de funções:

$$\psi_{j,m,n}(x, y) = 2^{j/2} \psi^i(2^j x - m, 2^j y - n) \quad (6.27)$$

$$\varphi_{j,m,n}(x, y) = 2^{j/2} \varphi(2^j x - m, 2^j y - n) \quad (6.28)$$

onde i representa a direção da função *wavelet* (H , V ou D), e m e n o nível de translação em relação à função mãe. A DWT de uma imagem será dada então pelas seguintes expressões:

$$W_\varphi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f[x, y] \varphi_{j_0, m, n}[x, y] \quad (6.29)$$

$$W_\psi^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f[x, y] \psi_{j, m, n}^i[x, y] \quad (6.30)$$

A determinação da DWT pode ser feita através de filtragens sucessivas seguidas por decimações por um fator de dois. Na Figura 6.6 é possível ver o que ocorre em duas decomposições sucessivas da DWT de uma imagem.

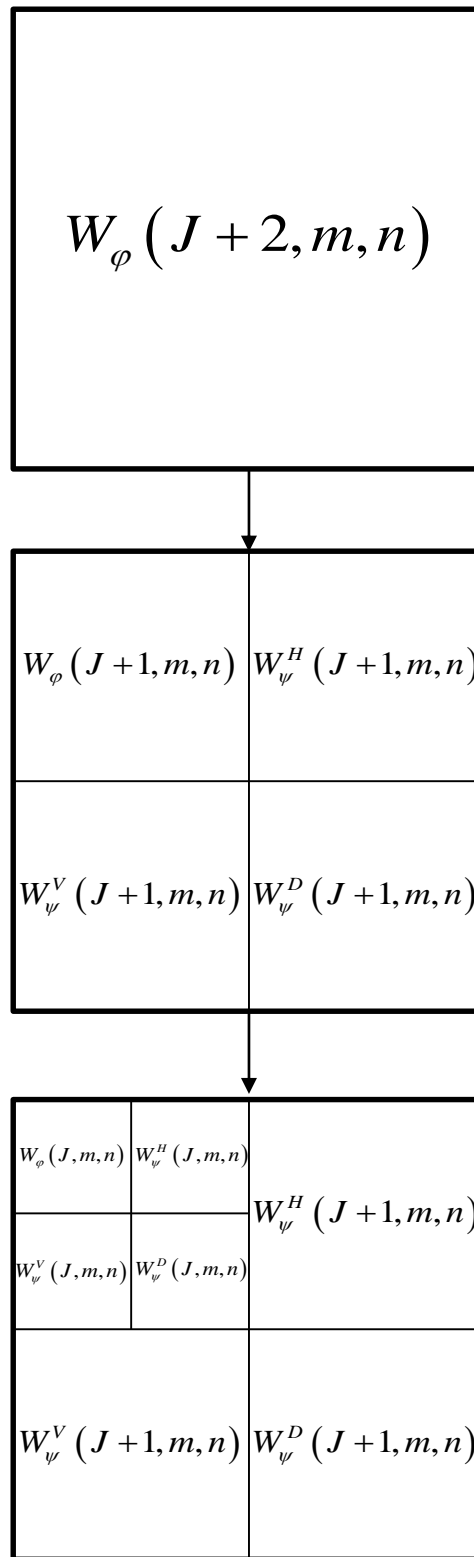


Figura 6.6. Exemplo de duas decomposições da DWT de uma imagem.

Os coeficientes de detalhe são aqueles que contêm mais informação de textura, por isso serão estes os usados para retirar *features* que representam textura. A média e variância dos coeficientes de cada uma das direções, de cada um dos canais representam boas codificações de textura de uma imagem (Martins et al. 2010).

6.2.2.2. Transformada *Curvelet*

Como foi visto anteriormente, a transformada *wavelet* de uma imagem, além de informação de frequência com resolução espacial, também dá informação direcional desta frequência. Estas direções são apenas três e podem ser poucas para analisar imagens com padrões mais complexos (Martins et al. 2010).

Assim, pode ser necessário ter mais informação direcional e é nesse contexto que foi criada a transformada *curvelet*, que é baseada numa noção anisotrópica da escala, tendo uma grande sensibilidade direcional (Ma & Plonka 2007). Usando *curvelets*, existe um parâmetro angular na função de base que será o que vai dar a sensibilidade direcional desta operação. Esta ferramenta é baseada na divisão do espaço 2D no domínio da frequência em regiões concêntricas seguindo-se a divisão destas regiões num certo número de ângulos com uma relação fixa. Assim, o resultado final será um conjunto de janelas concêntricas $W(r)$, e um conjunto de janelas angulares $V(\theta)$, tal como pode ser visto na Figura 6.7. Para cada uma destas divisões, haverá uma janela de frequências U_j , que corresponderá à TF de uma função *curvelet* baseada na função *curvelet* mãe $\varphi_j(x)$ (Candès et al. 2006).

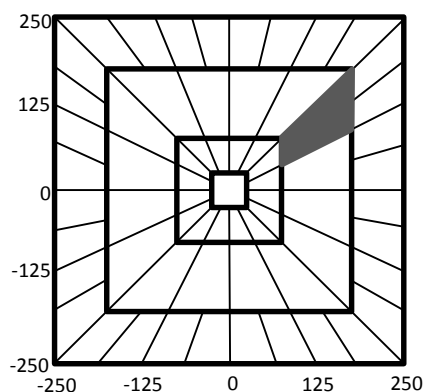


Figura 6.7. Divisão no domínio das frequências da Transformada Discreta *Curvelet*.

Os ângulos de rotação das janelas angulares em cada uma das escalas (a_j , em que $a_j = 2^{-j}$, para $j \geq 0$) serão obtidos pela seguinte equação:

$$\theta_{j,l} = \frac{\pi l 2^{-\lceil j/2 \rceil}}{2} \quad \text{com } l = 0, 1, \dots, 4 \times 2^{\lceil j/2 \rceil} - 1 \quad (6.31)$$

onde $\lceil x \rceil$ significa o menor número inteiro, maior ou igual a x . A partir dos ângulos de rotação determinados anteriormente, e a partir da função *curvelet* mãe $\varphi_j(x)$, é possível determinar as funções *curvelet* para cada escala e orientação para a posição $x_k^{(j,l)} = R_{\theta_{j,l}}^{-1}(k_1 \cdot 2^{-j}, k_2 \cdot 2^{-j/2})$. A sequência dos parâmetros de translação é dada por $k = (k_1, k_2) \in Z^2$ (Candès et al. 2006).

$$\varphi_{j,l,k}(x) = \varphi_j\left(R_{\theta_{j,l}}\left(x - x_k^{(j,l)}\right)\right) \quad (6.32)$$

Ao fazer a discretização do problema, e passando para um *input* de duas dimensões, é possível determinar os coeficientes *curvelet* a partir da equação seguinte.

$$c^D(j, l, k_1, k_2) = \sum_{\substack{0 \leq m < M \\ 0 \leq n < N}} f[m, n] \varphi_{j,l,k_1,k_2}^D[m, n] \quad (6.33)$$

6.3. Relação entre regiões

Além de um conjunto de *features* que caracterize o tecido presente em cada região é necessário obter uma medida de semelhança entre as várias regiões, usando as características propostas na secção anterior.

Em estatística, é possível calcular este valor usando dois tipos de métricas: métricas de distância ou métricas de dependência. As métricas do primeiro tipo quantificam a distância entre dois vetores, logo neste contexto o objetivo será minimizar esse valor; enquanto as do segundo tipo medem a semelhança entre dois vetores, logo o objetivo é maximizar o seu valor. Apesar de os seus valores terem significados contrários, o objetivo para o seu uso é muito semelhante (Hill & Lewicki 2006).

Dentro das métricas de distância podem-se salientar a distância euclidiana (definida como a raiz quadrada do somatório do quadrado das diferenças entre os valores de cada vetor) e a distância de Manhattan (definida como o somatório do módulo das diferenças entre os valores de cada vetor) (Hill & Lewicki 2006). Dentro das métricas de similaridade, as mais importantes e

mais utilizadas são o coeficiente de correlação de Pearson (medida que permite analisar a dependência linear entre dois conjuntos de variáveis), a similaridade do cosseno (calcula o cosseno do ângulo formado pelos vetores de características) e a informação mútua (que mede a dependência mútua entre conjuntos de variáveis aleatórias) (Wong & Yao 1987).

Para medir a relação entre duas regiões diferentes será usado o coeficiente de correlação de Pearson (ou coeficiente de correlação produto-momento). Este indicador mede o grau de relação entre duas variáveis, e o seu valor situa-se no intervalo entre -1 e 1. Este pode ter diferentes representações matemáticas, sendo que as mais comuns são as seguintes:

$$\phi_{X_1 X_2} = \frac{E\{(X_1 - E\{X_1\})(X_2 - E\{X_2\})\}}{\sigma_{X_1} \sigma_{X_2}} \quad (6.34)$$

$$\phi_{X_1 X_2} = \frac{\text{cov}(X_1, X_2)}{\text{cov}(X_1) \text{cov}(X_2)} \quad (6.35)$$

Este tipo de métrica foi usado para comparar regiões de *frames* consecutivas para encontrar regiões que representem zonas coincidentes entre imagens.

Capítulo 7 Implementação

Nos Capítulos 4, 5 e 6 foram abordados os métodos implementados durante a realização deste trabalho. No entanto essa implementação requer algumas considerações inerentes à particularidade do trabalho e que serão analisadas e discutidas neste capítulo. Serão também abordados outros temas tal como as bases de dados usadas e as plataformas de *software* utilizadas na implementação.

7.1. Base de dados

As imagens usadas neste trabalho são provenientes de duas fontes. Um conjunto de imagens foi fornecido pelo Dr. Jaime Ramos, do Departamento de Gastroenterologia do Hospital dos Capuchos, em Lisboa. Este primeiro conjunto de imagens consiste em 496 imagens da CE, onde 196 consistem em *frames* que contêm tecido tumoral e as restantes 300 contêm apenas tecido normal; sendo que não existe nenhuma continuidade entre as várias *frames*. Outro conjunto de imagens consiste num exame completo de um paciente, sem qualquer marcação manual. Estas últimas foram fornecidas como exame exemplo pela *Given Imaging*.

Foram usados estes dois conjuntos distintos de imagens para poder testar diferentes partes do algoritmo aqui proposto. Enquanto a primeira, sendo composta por imagens catalogadas em relação ao tipo de tecido presente, foi usada para validar o algoritmo de segmentação

corretamente, e mais especificamente a detecção automática de tumores. Em relação à segunda, esta é composta por uma sequência de *frames* consecutivas, o que é útil para a validação dos algoritmos de reconstrução 3-D e relação entre regiões afins em *frames* consecutivas.

7.2. Software

A implementação dos métodos e algoritmos propostos ao longo deste trabalho, ou seja os algoritmos de reconstrução, segmentação, extração de features e cálculos intermédios, foi efetuada usando a linguagem de programação MATLAB[®]. A classificação dos conjuntos de *features* foi efetuada usando o programa WEKA, um pacote *open source* para *machine learning*.

Todos os algoritmos foram processados num computador com processador quadcore i7 da Intel de 2.3 GHz com 6 GB de memória RAM.

7.3. Implementação da segmentação

O algoritmo de segmentação foi integralmente implementado no MATLAB[®], usando o método *k-means* existente na *Statistics Toolbox* para a inicialização, enquanto o restante processo foi construído de raiz.

7.3.1. Implementação do Algoritmo EM

Para um correto funcionamento, a implementação do algoritmo EM requereu alguma atenção em especial no que às características dos dados diz respeito. Por exemplo, uma das particularidades do espaço de cor usado, o espaço Lab, é possuir uma correlação baixa entre os respetivos canais (Woodland & Labrosse 2005; Reinhard et al. 2001). Devido a esta importante propriedade, as matrizes de covariância calculadas em cada uma das iterações do processo de segmentação pode ser aproximada por uma matriz diagonal (os coeficientes de covariância entre o canal i e o canal j , em que $i \neq j$ são considerados nulos). Num estudo anterior (Vieira et al. 2012) esta aproximação já foi feita em relação ao espaço de cor HSV e RGB, mostrando que no primeiro caso (baixo grau de correlação entre os canais) esta aproximação leva a resultados muito semelhantes a usar os verdadeiros valores de covariância; enquanto no segundo caso (grande correlação entre os canais) mostra uma grande diferença quando se faz esta aproximação. Ao

fazer esta consideração, a complexidade computacional do problema proposto é reduzida consideravelmente, reduzindo assim o tempo de computação do algoritmo.

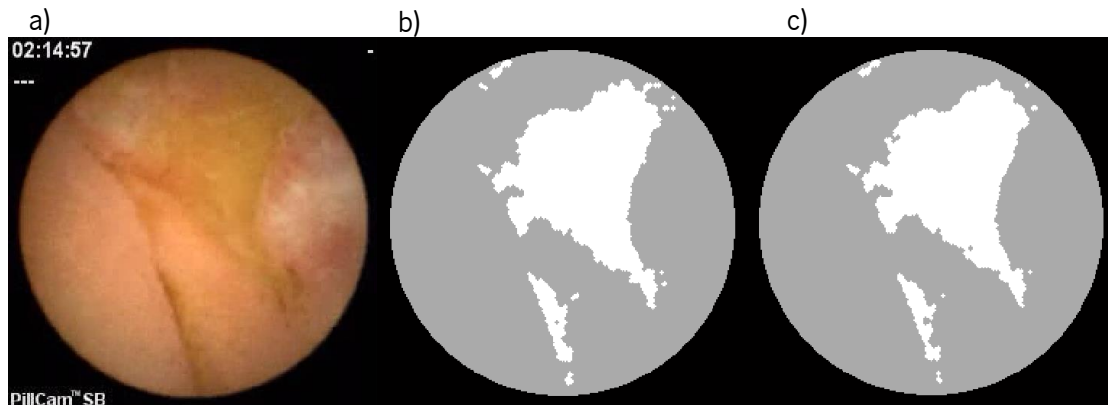


Figura 7.1. Segmentação de imagem da CE (a) em duas regiões, usando a matriz de covariâncias completa (b) e aproximando a uma matriz diagonal (c).

A Figura 7.1 mostra o desempenho do algoritmo de segmentação usando como aproximação a matriz de covariância diagonal (c), e a matriz completa (b). Atendendo a que geralmente os tumores se estendem por áreas significativamente extensas, as diferenças encontradas neste exemplo mostram que a aproximação é aceitável não comprometendo minimamente o processo de diagnóstico.

Como o algoritmo utilizado é iterativo, um ponto bastante importante é o critério de paragem a considerar. Teoricamente, o processo de estimação iterativa dos parâmetros do modelo deve ser interrompido quando de uma iteração para a seguinte a verosimilhança não seja incrementada mais que um pré-determinado *threshold*. Como o estabelecimento deste *threshold* é pelo menos difícil de justificar, um processo alternativo é frequentemente usado e consiste em parar o processo quando não existe qualquer passagem de píxeis de uma para outra classe, ou seja, quando não há qualquer alteração entre iterações. Por vezes podem existir oscilações de um ou mesmo de vários píxeis entre classes de iteração para iteração pelo que, para prever estas situações é comum fazer o algoritmo parar ao ser atingido um certo número, normalmente elevado de iterações. Como máximo de iterações considerou-se o valor de 100, uma vez que nas imagens testadas nunca houve valores superiores a 50 iterações.

Tal como foi abordado na Secção 4.6. a segmentação é efetuada n vezes, e cada uma das vezes considerando que se pretende dividir a imagem em n *clusters* diferentes, dado que não é conhecido o número de classes existente em cada *frame*. O número máximo de *clusters* em que o tecido pode ser dividido é uma questão bastante importante, principalmente em relação ao tempo de processamento dado que quanto maior for o número de *clusters*, maior será o tempo

de processamento. Por outro lado, a detecção do maior número possível de estruturas (pólipos, úlceras, etc), dado que podem existir vários com diferentes elevações e/ou profundidades, será fundamental para a qualidade do vídeo 3-D gerado. Tendo em conta a dimensão das imagens (256x256), a percentagem dos píxeis que está dentro do campo de visão e que pertencem assim a tecido intestinal (aproximadamente 67%) e a análise de várias imagens da base de dados por médicos, chegou-se à conclusão que um bom valor para n é 7. Isto porque um número superior de regiões numa imagem pequena faz com que algumas dessas regiões sejam pequenas demais para se tirar informação relevante delas.

7.3.2. Pós-processamento

O resultado da segmentação apresentava muitas vezes pequenas regiões pertencentes a uma classe isoladas no meio de outra classe maior. Como isso, dificilmente representaria um tecido diferente que existe na imagem, decidiu-se fazer um pós-processamento à segmentação para retirar estas pequenas regiões, usando métodos de morfologia matemática para isso.

A morfologia matemática é uma técnica muito usada para analisar e processar imagens em termos das suas estruturas geométricas, sendo usada mais geralmente e com melhores resultados em imagens binárias (Gonzalez & Woods 2008). Então, para retirar estas pequenas regiões isoladas decidiu-se implementar o algoritmo que pode ser visto na Figura 7.2.

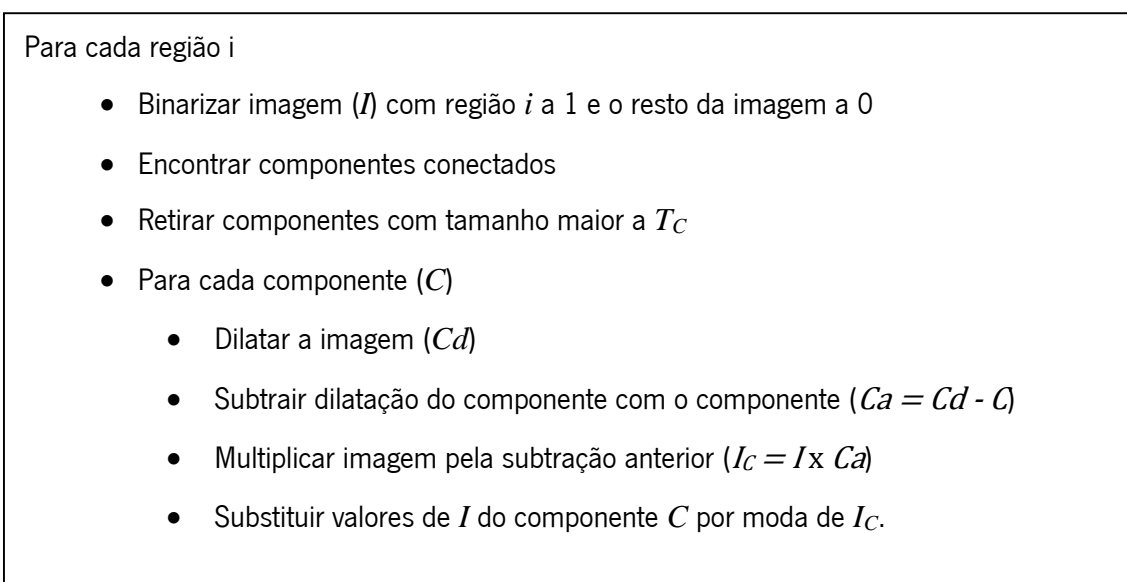


Figura 7.2. Algoritmo de pós-processamento do resultado da segmentação.

Este algoritmo pode então ser aplicado ao resultado das segmentações, independentemente do número de regiões em que a imagem foi dividida, sendo que no âmbito desta tese foi usado um valor de 50 para T_c .

7.4. Implementação da reconstrução 3-D

Na implementação do algoritmo de SfS descrito no Capítulo 5 tiveram que ser feitas certas considerações em parte devido à falta de conhecimento de algumas características técnicas da câmara da CE. Como se pode ver na Tabela 5.1, um ponto de uma superfície projetado no plano de uma câmara tem as suas coordenadas a depender da distância focal (F) dessa câmara. Na literatura consultada, esse valor nunca é dado, sendo que as características disponíveis da câmara e lente presente na *Pillcam SB* (cápsula usada para retirar as imagens das bases de dados usadas) são as que se podem ver na Tabela 7.1.

Tabela 7.1. Características da câmara e lente da Pillcam SB.

| | |
|------------------------------|---------------------|
| Dimensões | 11 mm x 26 mm |
| Frame rate | 2 fps |
| Profundidade de campo | 0 – 30 mm |
| Ângulo de campo | 140° |
| Área de cobertura | 500 mm ² |

O ângulo de campo (θ) pode ser descrito matematicamente pela Equação (7.1), onde D representa a dimensão do sensor da câmara e F a distância focal.

$$\theta = 2 \arctan\left(\frac{D}{2F}\right) \quad (7.1)$$

Tendo em conta que o sensor tem sensivelmente o tamanho da cápsula (em termos de diâmetro), é possível facilmente passar a expressão acima em ordem à distância focal (Equação (7.2)), e fazendo as substituições necessárias chega-se facilmente a um valor de F aproximadamente igual a 2 mm.

$$F = \frac{D}{2 \tan\left(\frac{\alpha}{2}\right)} \quad (7.2)$$

Para implementar a parte final do algoritmo relativa à aproximação a um filtro de Kalman, era necessário calcular vários parâmetros para a função do ganho (esta função pode ser vista na Equação (5.23). Decidiu-se usar um valor de W igual a 1×10^{-7} por ser indicado em (Tsai & Shah 1994) que este deveria ser um valor pequeno mas não nulo. Para o cálculo de S^n , como pode ser visto na Equação (5.25), é necessário ter o valor da profundidade real para calcular o “erro” em cada iteração. Como claramente esse valor não está disponível, já que é o objetivo do uso deste algoritmo, é necessário o cálculo do valor de S^n de outra forma. A partir da leitura de (Rojas 2005) percebe-se que se pode calcular da seguinte forma:

$$S^n(x, y) = (1 - K^n M(x, y)) S^{n-1}(x, y) \quad (7.3)$$

O valor se S^0 é considerado de 0,01 por ter sido considerado um bom valor nas experiências efetuadas, e o mapa de profundidades igual à unidade, para dar um bom valor inicial ao algoritmo (não afetando a capacidade de convergência).

7.5. Implementação da detecção de tumores

As questões discutidas na secção anterior relativas à segmentação foram também usadas na detecção de tumores, sendo para isso feita a divisão da imagem em duas regiões. Neste caso é considerado *a priori* o número de duas regiões, não sendo então necessário o uso do método BIC.

A classificação das *features* calculadas para as várias *frames* foi feita usando classificadores supervisionados, que são métodos que necessitam de uma fase de aprendizagem, onde o utilizador tem de apresentar um conjunto de dados corretamente classificados, e assim posteriormente o algoritmo poder fazer esta classificação corretamente. Os algoritmos utilizados foram redes neuronais, mais especificamente *Multilayer Perceptron* (MLP), e *Support Vector Machines* (SVM). O primeiro algoritmo simula uma rede de neurónios (células básicas do sistema nervoso) a um nível computacional, combinando as características de entrada linearmente e no final, através de uma regra de decisão, decidir qual a classe que mais se adequa. Este método permite a inclusão de várias camadas (*layers*), o que aumenta a complexidade da forma da divisão do espaço. O algoritmo SVM tem como objetivo definir um hiperplano, que é aquele que tem a maior distância para os dados de treino mais próximos de cada uma das classes (esta distância é chamada de margem). Este é um modelo mais flexível e menos sensível a *outliers* quando os parâmetros são corretamente ajustados, além de que é possível incluir *kernels* que definam outro

tipo de formas de planos para fazer a divisão entre classes (Mohri et al. 2012). Neste trabalho foi usado um kernel *Radial Basis Function* (RBF).

Na classificação, independentemente do classificador usado, foi utilizado um método de validação cruzada *k-fold* (k-partições), tendo sido escolhido um valor de 10 para k. Este método tem como princípio a divisão da amostra inicial (conjunto de imagens) em K subamostras iguais. Destas subamostras, 1 delas é retida para ser usada como conjunto de teste para o classificador, enquanto as outras K-1 subamostras serão usadas como conjunto de treino. Este processo é então repetido K vezes, para garantir que todas as subamostras são usadas exatamente uma vez como conjunto de teste. Os resultados da classificação são então obtidos através da média dos resultados de cada uma das partições efetuadas. A maior vantagem deste método é garantir que todos os dados são usados como treino e teste, e que todos os dados são incluídos também apenas e exatamente uma vez no conjunto de treino (Oliveira 2008).

Para poder analisar os resultados obtidos neste estudo tiveram que ser calculadas métricas, para poder comparar e quantificar cada teste. As escolhidas foram a sensibilidade (taxa de *frames* com tecido tumoral que são corretamente classificadas como tumorais – Equação (7.4)), especificidade (taxa de *frames* normais que são corretamente classificadas como normais – Equação (7.5)) e precisão (taxa de *frames* corretamente classificadas no conjunto – Equação (7.6)). Nestas equações, o valor de *TP* (*true positive*) corresponde ao número de *frames* corretamente classificadas como tumorais, *TN* (*true negative*) corresponde ao número de *frames* corretamente classificadas como normais, *FN* (*false negative*) corresponde ao número de *frames* tumorais incorretamente classificadas, *FP* (*false positive*) corresponde ao número de *frames* normais incorretamente classificadas, *P* (*positive*) corresponde ao número de *frames* tumorais, *N* (*negative*) corresponde ao número de *frames* normais.

$$Sensibilidade = \frac{TP}{P} \times 100\% = \frac{TP}{FN + TP} \times 100\% \quad (7.4)$$

$$Especificidade = \frac{TN}{N} \times 100\% = \frac{TN}{TN + FP} \times 100\% \quad (7.5)$$

$$Precisão = \frac{TP + TN}{TP + TN + FN + FP} \times 100\% \quad (7.6)$$

Para discutir e avaliar as *features* selecionadas para a classificação de cada imagem como tumoral ou não tumoral, foi feita uma análise estatística para cada uma das *features*. Testes bilaterais e emparelhados foram realizados, em que foi usado o teste de t-Student para dados com

Implementação da detecção de tumores

distribuição normal e o teste de Mann-Whitney para dados não-paramétricos. Como nível de significância, foi usado o valor de 0,001; logo quando o valor-p era inferior a este nível, a *feature* correspondente era considerada significativa para separar corretamente *frames* com tecido tumoral de *frames* só com tecido normal.

Capítulo 8 Resultados

Neste capítulo serão apresentados e discutidos os mais relevantes resultados obtidos no âmbito desta tese. O capítulo começa por apresentar e discutir alguns resultados intermédios do método de segmentação, bem como alguns resultados visuais da segmentação de *frames* da cápsula endoscópica. Seguidamente os resultados referentes à classificação de imagens como tumorais ou não tumorais serão apresentados e discutidos, no contexto de *features* baseadas quer na cor, quer na textura. Finalmente serão apresentados os resultados da reconstrução tridimensional de *frames* isoladas.

8.1. Segmentação

Os resultados da implementação do algoritmo de segmentação são divididos em três diferentes secções: o método de inicialização, a segmentação propriamente dita, e a relação entre as diferentes regiões.

8.1.1. Inicialização

Tal como foi apresentado na Secção 4.5. a inicialização do algoritmo de segmentação é feita usando o método *k-means*. Este método apresenta resultados finais com algum grau de aleatoriedade que se pode estender, embora possivelmente em grau menor, ao segmentador.

Para verificar o efeito que o método de inicialização escolhido pode ter nos resultados da segmentação, foram utilizadas as 496 imagens catalogadas como contendo tumores ou apenas tecido normal. Assim, foram feitos 100 ensaios do algoritmo de segmentação (com uma divisão de cada imagem em 2 regiões, usando o canal *a* e o canal *b*), sendo que em cada um dos ensaios o método *k-means* foi executado de forma independente. Foi calculado o desvio padrão de todos os elementos dos vetores de média das distribuições, para os vários ensaios em cada uma das imagens. Estes resultados têm como objetivo mostrar como é que a aleatoriedade do método *k-means* ao longo de 100 ensaios influencia os resultados da segmentação, apresentando-se os valores da média, do máximo e do mínimo do conjunto dos desvios padrão (Tabela 8.1).

Tabela 8.1. Média, mediana, máximo e mínimo dos valores do desvio padrão de vários parâmetros das distribuições das 496 imagens usadas.

| <u>Desvio Padrão</u> | Média | Mediana | Mínimo | Máximo |
|-----------------------------|--------------|--------------------------|---------------|---------------|
| m_{a1} | 0,0033 | $1,2434 \times 10^{-14}$ | 0 | 0,3860 |
| m_{a2} | 0,0183 | $4,2633 \times 10^{-14}$ | 0 | 1,8474 |
| m_{b1} | 0,0085 | $1,5987 \times 10^{-14}$ | 0 | 1,4236 |
| m_{b2} | 0,0160 | $5,6843 \times 10^{-14}$ | 0 | 2,3672 |

Como é possível ver na Tabela 8.1, os valores médios do desvio padrão são da ordem da grandeza das centésimas e milésimas, pode-se considerar que existe uma influência muito reduzida da aleatoriedade do método de inicialização da segmentação utilizado.

Olhando para os valores máximos, vê-se que existe valores na ordem das unidades, o que representa uma variação significativa nos resultados finais dos parâmetros das distribuições. Mas, analisando para os valores do desvio padrão de cada uma das imagens, conclui-se que este valor máximo só é atingido em 2 imagens, o que representa apenas 0,4% das imagens existentes. Comparando os valores da mediana e da média chega-se à conclusão que a maioria dos valores está bastante abaixo da média, concluindo que a maioria das vezes a aleatoriedade do *k-means* não influencia em nada os resultados.

Além do cálculo dos descritores estatísticos da tabela anterior foi feito um teste estatístico semelhante ao explicado na Secção 7.5. Considerando a hipótese nula de a média dos valores de desvio padrão dos vários parâmetros ser diferente de nula, e ao fazer um teste com uma amostra

t de Student, chegou-se a um p -value menor que 0,05 (nível de significância considerado) para todos os parâmetros. Assim, com todas estas análises, pode-se então concluir que o método de inicialização escolhido no contexto do método de segmentação MAP proposto, não influencia significativamente os resultados da segmentação, já que as soluções finais são sempre bastante próximas quando realizados vários ensaios.

8.1.2. Segmentação de tecidos

Tal como explicado no Capítulo 4, o método de segmentação seguido neste trabalho foi uma abordagem MAP. Uma grande particularidade da aplicação deste método foi a remoção da luminosidade, usando o espaço de cor Lab. Na Figura 8.1 é possível ver um exemplo de segmentação com vários espaços de cor da *frame* (Figura 8.1a) já mostrada na Figura 6.1. Esta imagem, tal como já foi dito, contém uma região central com uma saturação facilmente perceptível provocada por uma grande proximidade da cápsula nesta zona. A não consideração do canal da luminosidade no algoritmo de segmentação tem como objetivo remover este efeito.

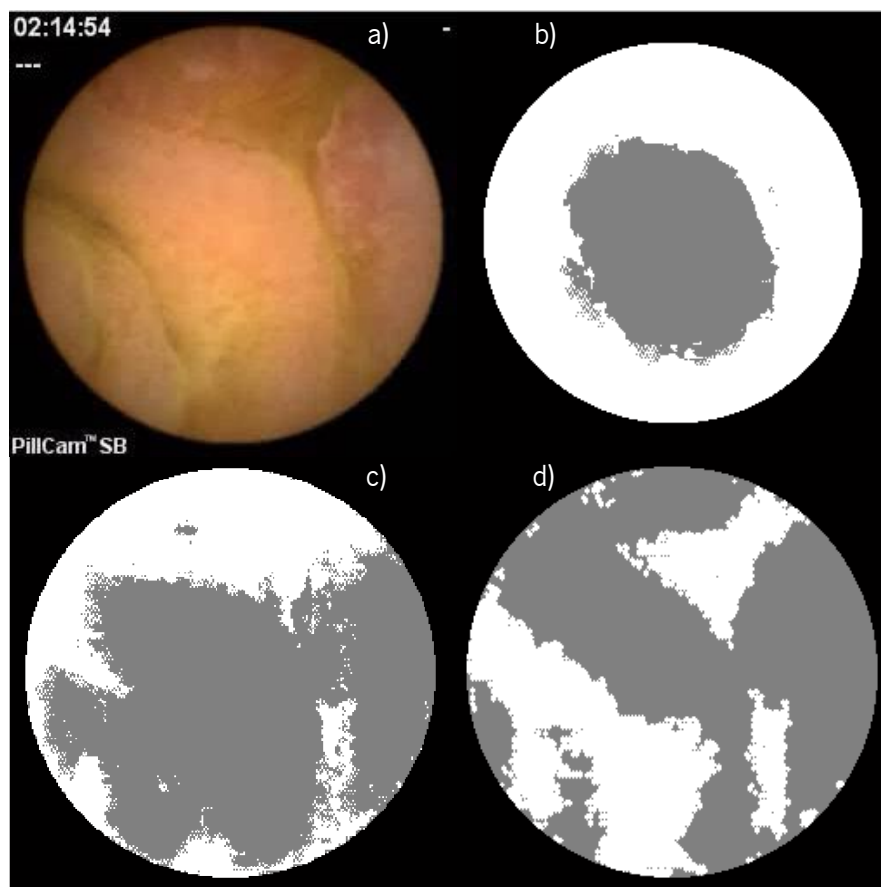


Figura 8.1. Exemplo de segmentação de frame da CE (a) usando o espaço de cor RGB (b), o espaço de cor Lab (c) e o espaço de cor Lab sem o canal da luminosidade (d).

Ao fazer a segmentação em duas classes no espaço de cor RGB (Figura 8.1b) pode-se ver claramente que a divisão não teve influência no tipo de tecido, mas somente na quantidade de luz presente em cada região. No caso em que se usam os três canais do espaço de cor Lab (Figura 8.1c), vê-se uma maior abrangência do algoritmo de segmentação, incluindo outras partes da imagem, além da região saturada. Usando apenas o canal *a* e *b* (Figura 8.1d), ou seja, descartando o canal L, o resultado é bastante melhor. Apesar de uma pequena parte central da imagem estar incluída na mesma classe que o tecido com tumor (zona à direita e zona superior), a delimitação é bastante melhor que nos casos anteriores.

Na Figura 8.2 é possível ver outro exemplo da segmentação de uma *frame* retirada da CE (Figura 8.2a), mas neste caso a saturação do tecido não é tão evidente como no caso anterior. Nesta imagem há claramente dois tipos de tecido intestinal presentes (indicados com os números 1 e 2 na Figura 8.2a). Na Figura 8.2b vê-se o resultado da segmentação do tecido considerando o espaço de cor RGB. Aqui é possível ver claramente que foi incluído uma zona do tecido 2 na classe do tecido 1, mais especificamente a zona central da imagem onde existe uma maior luminosidade devido à luz da câmara da CE. Quando é usado o espaço de cor Lab para fazer a segmentação (Figura 8.2c), consegue-se ver que a zona central da imagem (onde existe alguma saturação), não está incluída na classe do tecido 1, apesar da delimitação deste tecido não ser a melhor, muito provavelmente pela diferença de luz nas bordas deste tecido que o canal L reconhece como sendo diferentes do resto. Na última imagem (Figura 8.2d), está presente o resultado da segmentação da *frame* original usando apenas o canal *a* e o canal *b* do espaço de cor Lab. Aqui, além de não estar incluída a zona de saturação na classe do tecido 1, as bordas deste tecido têm uma melhor delimitação, incluindo uma maior parte deste tecido que não estava presente na segmentação com o canal da luminosidade.

Nos exemplos mostrados anteriormente foi considerada apenas a questão da remoção da luminosidade das imagens ao usar o espaço de cor Lab. Outro problema endereçado neste trabalho foi a forma de descobrir o número de classes ótimo existente em cada *frame*.

Tal como apresentado na Secção 4.6. usou-se o método BIC para descobrir qual o número de regiões existentes em cada imagem. Tal como foi dito anteriormente, não é possível quantificar os resultados nesta secção, já que não existe nenhuma segmentação de referência feita por especialista, nem sequer a definição do número de regiões que cada imagem possui. Assim, não é possível fazer uma análise além da descritiva.

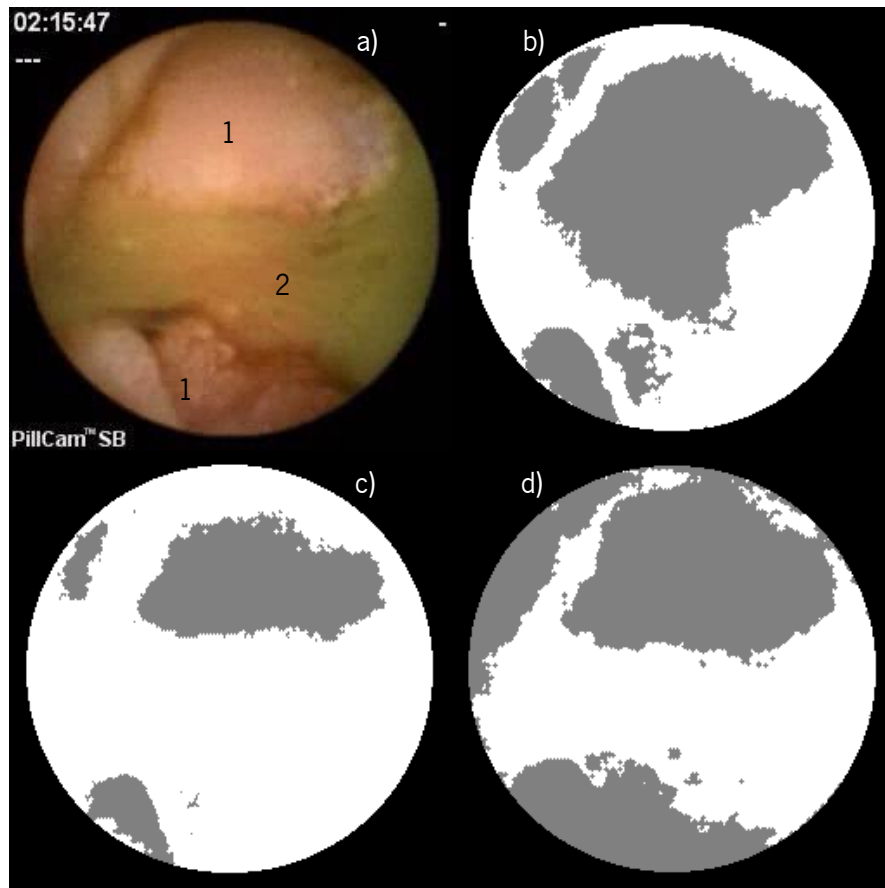


Figura 8.2. Exemplo de segmentação de frame da CE (a) usando o espaço de cor RGB (b), o espaço de cor Lab (c) e o espaço de cor Lab sem o canal da luminosidade (d).

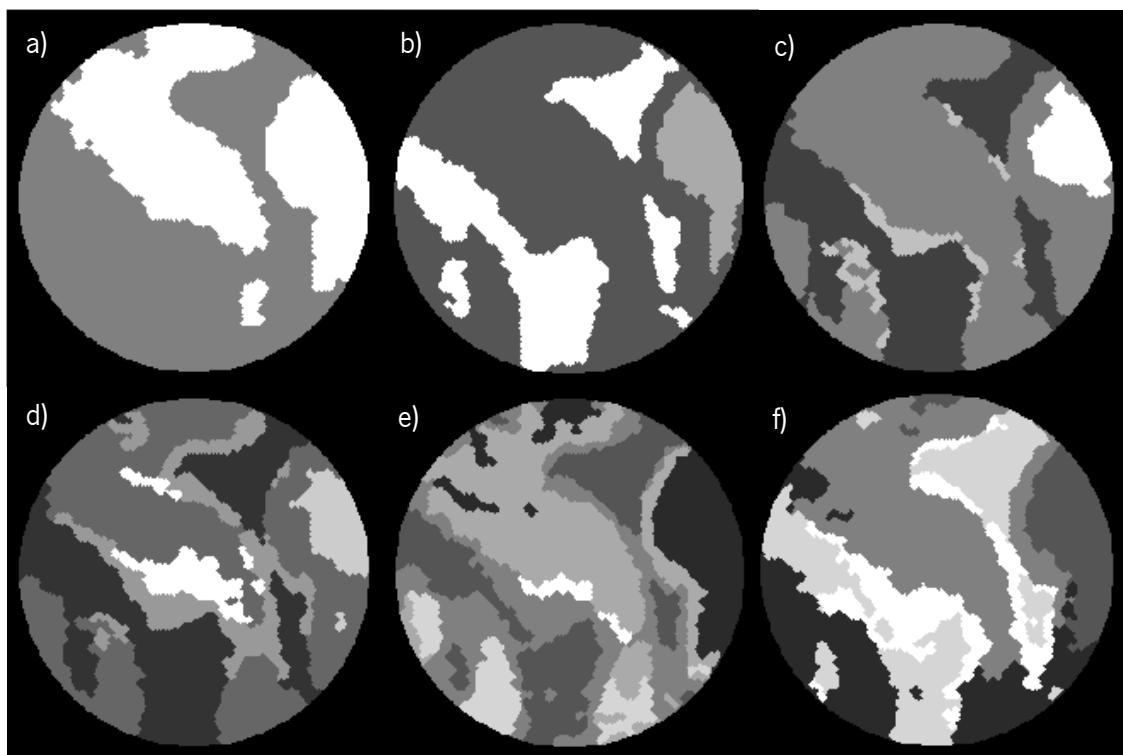


Figura 8.3. Segmentação da *frame* da CE apresentada na Figura 6.1 em diferentes números de regiões: duas a sete (a-f).

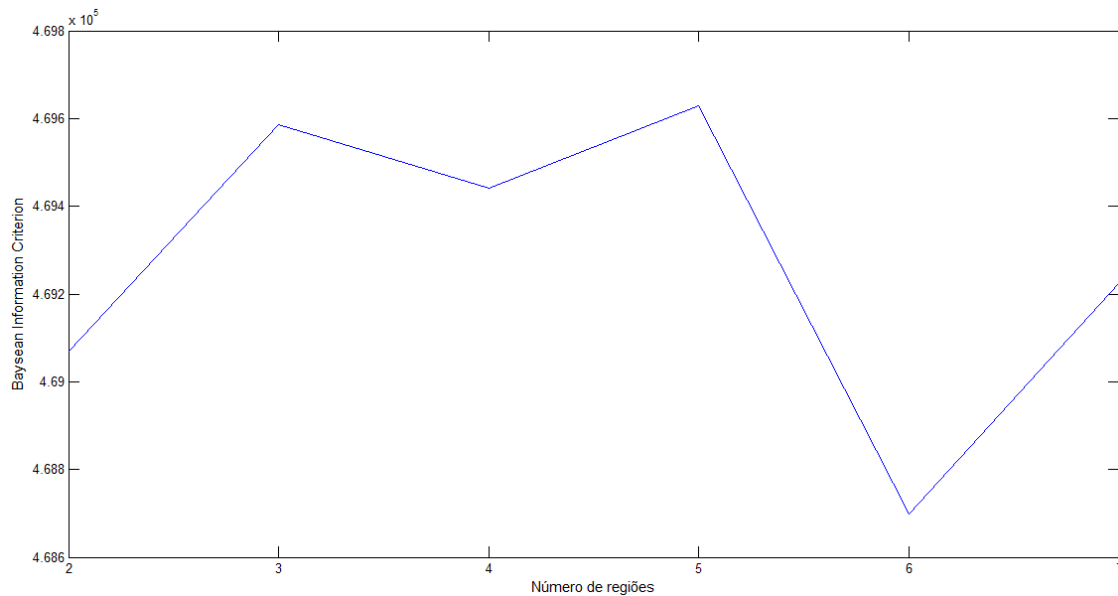


Figura 8.4. Valor do BIC das segmentações com vários números de classes representadas na Figura 8.3.

Pegando novamente no exemplo da *frame* presente na Figura 6.1, executou-se o algoritmo de segmentação completo, ou seja, separou-se o tecido de duas classes até sete classes. Os resultados destas segmentações podem ser vistos na Figura 8.3. Foi feito também o cálculo do valor de BIC para cada uma das segmentações, tendo sido construído um gráfico com todos estes valores, que pode ser visto na Figura 8.4.

Ao analisar a Figura 8.4 é possível verificar que o menor valor de BIC ocorre na segmentação da imagem em 6 regiões (resultado na Figura 8.3e). Apesar de ser complicado verificar visualmente a melhoria do resultado entre as várias segmentações, é possível fazer várias considerações:

- Entre a imagem a e b houve um aumento no valor de BIC (logo, uma piora na segmentação). A maior diferença entre estas duas imagens é a separação em duas classes do tecido tumoral à direita da imagem (representado a branco na imagem da Figura 8.3a e a cinzento escuro e claro na Figura 8.3b). Esta zona, visualmente, parece pertencer à mesma região, o que leva a concluir que realmente a segmentação piorou.
- O valor do BIC entre a segmentação em três, quatro e cinco regiões não tem grandes alterações, podendo-se concluir que os dados da imagem se adequam de forma semelhante a estas três divisões. Uma característica muito significativa no resultado da segmentação em quatro e cinco regiões é a divisão da região já

mencionada anteriormente, situada na zona direita da imagem. Esta divisão pode ser uma das causas do elevado valor do BIC, tal como tinha também acontecido com a segmentação em três regiões.

- O menor valor do BIC representa a divisão em seis regiões, onde a zona de referência já mencionada anteriormente pertence quase toda a uma só região, ao contrário do que acontecia nas restantes segmentações.

Para continuar com a análise descritiva deste método de segmentação, decidiu-se aplicá-lo a uma outra *frame* da CE (Figura 8.5), onde os resultados das várias segmentações podem ser vistos na Figura 8.6 e cada valor de BIC na Figura 8.7.

Observando a Figura 8.7 é possível ver que o menor valor de BIC ocorre quando é realizada a segmentação em cinco regiões (correspondente à Figura 8.7d). Outro dado importante a analisar é a grande diferença entre o valor de BIC para a segmentação em duas regiões em relação aos restantes valores, já que este é bastante superior. Isto significa que os dados não se adaptam com tanta precisão a uma distribuição Gaussiana bimodal do que a uma mistura com mais que 2 componentes, sendo que a mistura com 5 componentes é aquela que probabilisticamente melhor modela as intensidades da *frame*. É por isso mais verosímil que a *frame* contenha 5 regiões efetivamente distintas.



Figura 8.5. Exemplo de *frame* retirada de exame da CE.

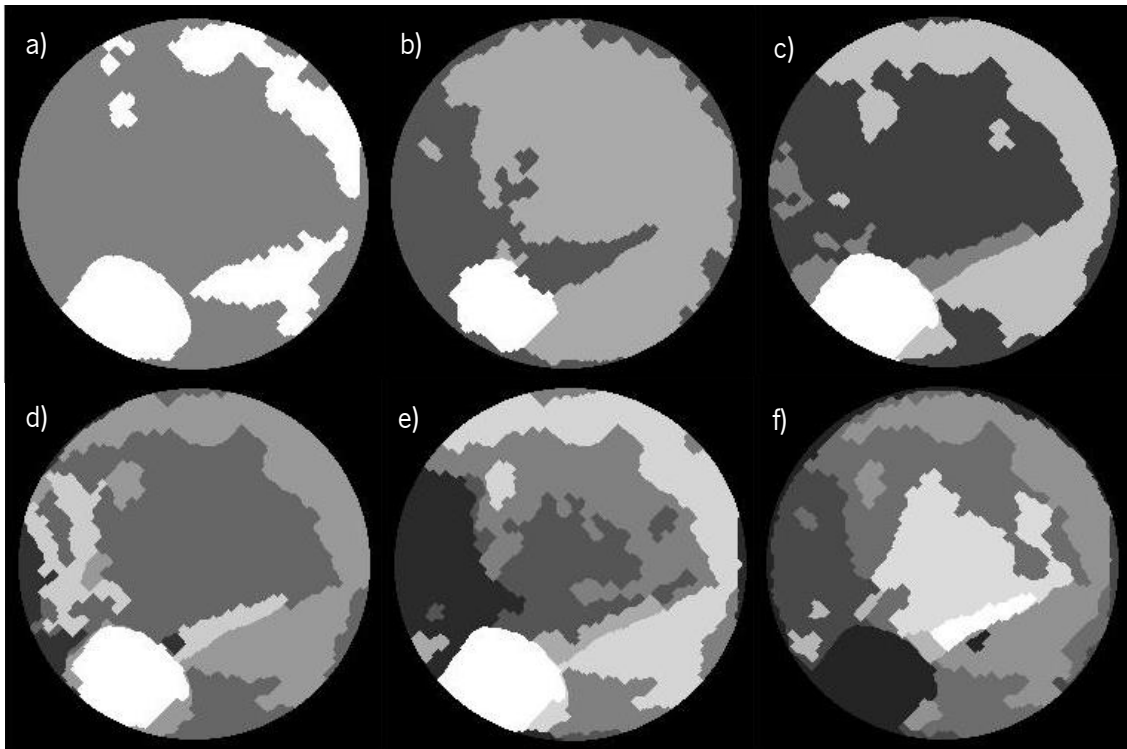


Figura 8.6. Segmentação da *frame* da CE apresentada na Figura 8.5 em diferentes números de regiões: duas a sete (a-f).

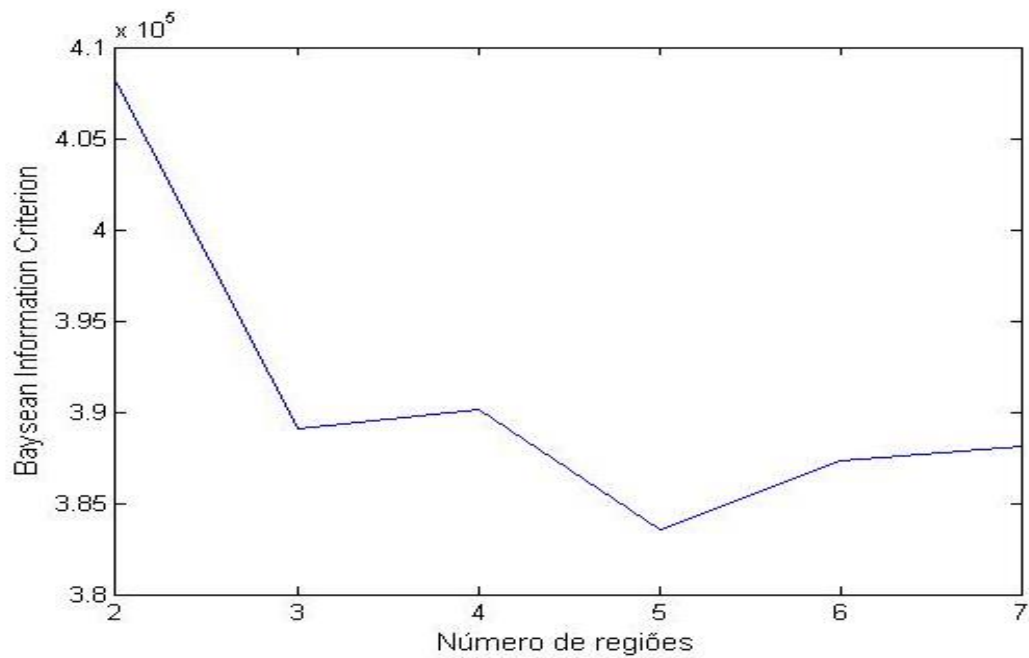


Figura 8.7. Valor do BIC das segmentações com vários números de classes representadas na Figura 8.6.

8.1.3. Relação entre regiões

Tendo 2 *frames* consecutivas segmentadas, cada uma contendo eventualmente um número diferente de regiões efetivas, importa fazer uma correspondência entre regiões de ambas as *frames*, para através da estimação do movimento gerar *frames* intermédias que permitam obter um vídeo exibindo variações suaves. Esta correspondência terá que ser feita com base nas características dos tecidos, já que no contexto da CE, quer a fonte (CE), quer o alvo (mucosa intestinal) são ambos móveis com velocidade instantânea dependente dos movimentos peristálticos e por isso imprevisíveis. Para efetuar esta análise, usaram-se três *frames* consecutivas de um exame da CE para fazer a segmentação, calcularam-se *features* de cor (médias e variâncias dos histogramas dos canais *a* e *b* do espaço de cor Lab) e textura (médias e variâncias das várias direções e escalas dos coeficientes da DCT). Como medida de semelhança entre regiões foram usados os coeficientes de correlação entre as *features* de regiões das *frames* consecutivas (que podem ser vistas na Tabela 8.2 e Tabela 8.3).

Cada região é caracterizada por várias *features* diferentes: de cor são descritores estatísticos de cada um dos canais, e de textura descritores estatísticos de cada uma das direções para cada um dos canais. Consequentemente cada vetor de características é composto por uma observação de cada variável. Modelando o conjunto destas variáveis por um processo estocástico, e aplicando a propriedade da ergodicidade dos processos estocásticos, aproxima-se a distribuição do conjunto das variáveis pela distribuição de cada uma delas individualmente (Yates & Goodman 2005). Deste modo, o cálculo dos coeficientes de correlação usado neste trabalho caracteriza corretamente a relação que o conjunto de variáveis de cada uma das regiões tem entre si.

O cálculo de coeficientes de correlação entre vetores de características e a consequente determinação de regiões correspondentes em *frames* consecutivas necessita de um threshold para poder concluir sobre o real significado destes valores. No caso dos valores aqui calculados é difícil inferir o melhor valor para se poder considerar que a correlação é boa. Assim, decidiu-se calcular uma probabilidade a partir dos valores de correlação usando a Equação (8.1). Os valores resultantes terão como significado o grau de confiança na correspondência entre as duas regiões consideradas.

$$P_{12} = \frac{\phi_{12}}{\max \phi} \quad (8.1)$$

Segmentação

Considerando que existe, em todos os casos, pelo menos uma região em cada uma das duas imagens que pertencem ao mesmo tecido, esta será aquela que terá o coeficiente de correlação superior, e conseqüentemente um grau de confiança de 100%. Considera-se um limite de confiança de 95% para aceitar a correspondência entre duas regiões.

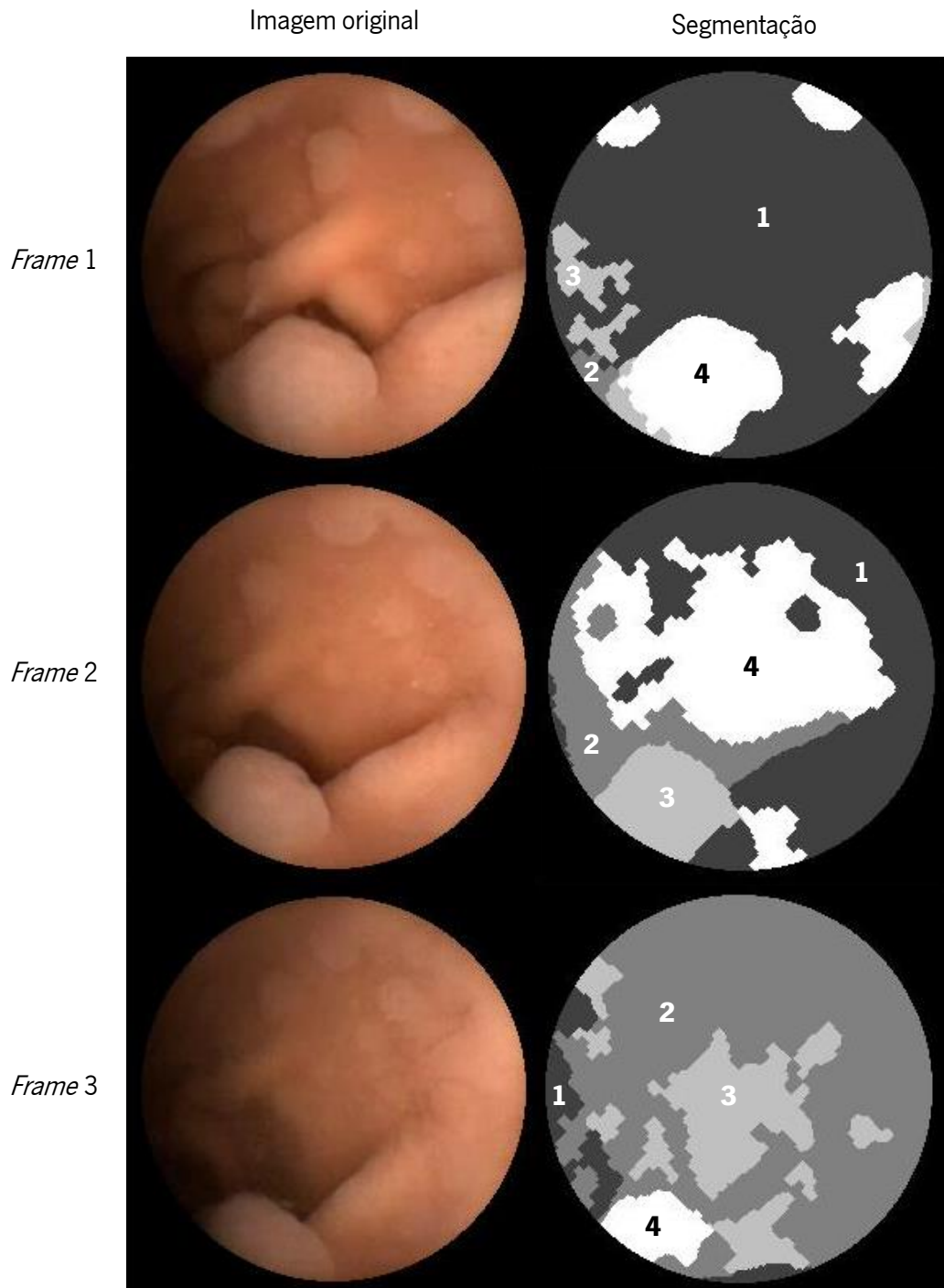


Figura 8.8. Três *frames* consecutivas de um exame da CE e as respectivas segmentações.

Para analisar os resultados desta secção é necessário olhar para as tabelas relativas aos graus de confiança, procurando o máximo, e se for superior ao limite considerado (0,95), as regiões são consideradas correspondentes nas diferentes *frames*. As regiões consideradas que possuíam correspondências nas diferentes *frames* estão indicadas na Tabela 8.4.

Tabela 8.2. Valor do grau de confiança da correspondência entre as várias regiões da *frame* 1 e da *frame* 2.

| <i>Frame</i> 2 <i>Frame</i> 1 | Região 1 | Região 2 | Região 3 | Região 4 |
|----------------------------------|---------------|---------------|---------------|----------|
| Região 1 | 0,9998 | 0,9108 | 0,9784 | 0,8814 |
| Região 2 | 0,5285 | 0,8027 | 0,3190 | 0,0389 |
| Região 3 | 0,7656 | 0,9613 | 0,6073 | 0,3742 |
| Região 4 | 0,9626 | 0,7764 | 1,0000 | 0,9723 |

Chega-se assim à conclusão que na *frame* 1, três das quatro regiões têm correspondência na *frame* 2. O maior grau de confiança é na correspondência entre a região 4 da primeira *frame* e a região 3 da segunda *frame*, que visualmente correspondem às regiões com maior relevância nas imagens. O segundo maior grau de confiança corresponde à comparação das regiões 1 das duas imagens, apesar de estas serem regiões que visualmente não correspondem exatamente aos mesmos tecidos. O terceiro maior valor corresponde à comparação entre as regiões 3 e 2, da primeira e segunda imagem respetivamente, onde visualmente se apercebe que não corresponde exatamente ao mesmo tecido, já que uma grande parte da região da segunda imagem não está incluída na região da primeira imagem.

Tabela 8.3. Valor do grau de confiança da correspondência entre as várias regiões da *frame* 2 e da *frame* 3.

| <i>Frame</i> 3 <i>Frame</i> 2 | Região 1 | Região 2 | Região 3 | Região 4 |
|----------------------------------|----------|---------------|---------------|---------------|
| Região 1 | 0,6516 | 0,9824 | 0,8865 | 0,9923 |
| Região 2 | 0,9024 | 0,8308 | 0,6505 | 0,8691 |
| Região 3 | 0,4672 | 0,9919 | 0,9686 | 0,9992 |
| Região 4 | 0,2105 | 0,9981 | 1,0000 | 0,9206 |

Ao analisar a Tabela 8.3, pode-se observar que o maior grau de confiança corresponde à correspondência entre a região 4 e região 3, da segunda e terceira *frame* respetivamente. Estas

são regiões que apesar de não serem exatamente as mesmas, estão situadas aproximadamente na mesma localização. O segundo maior valor corresponde à comparação entre a região 3 e região 4 da segunda e terceira *frame* respetivamente, o que corresponde às zonas de tecido com maior correlação na comparação anterior (Tabela 8.2). Em relação ao terceiro maior valor, este corresponde à comparação entre a região 1 e a região 2 da segunda e terceira *frame* respetivamente. Neste caso, as duas regiões correspondem aproximadamente ao mesmo tecido.

Tabela 8.4. Resultados da correspondência de regiões entre *frames*.

| Frame 1 | Frame 2 | Frame 3 |
|----------------|----------------|----------------|
| Região 1 | Região 1 | Região 2 |
| Região 2 | – | – |
| Região 3 | Região 2 | – |
| Região 4 | Região 3 | Região 4 |
| – | Região 4 | Região 3 |
| – | – | Região 1 |

Fazendo uma análise geral a estes resultados, conclui-se que apesar do cálculo dos coeficientes de correlação entre *features* das diferentes regiões darem resultados aceitáveis, a segmentação precisa de ser melhorada para que estas regiões estejam melhor definidas nas diferentes *frames*.

8.1.4. Tempo de execução

Outra questão importante em relação ao algoritmo de segmentação utilizado é o seu tempo de execução. Ao executar o algoritmo de segmentação proposto em cada uma das 496 imagens da base de dados foram medidos os tempos de computação, mostrando na Tabela 8.5 a média, desvio padrão, máximo e mínimo dos valores destes tempos. Ao analisar estes valores, pode-se concluir que o tempo médio pouco inferior a 10 segundos pode-se considerar bom, tendo em conta o número de vezes que é necessário executar o algoritmo de segmentação (6 vezes, entre a 2 regiões a 7 regiões). Além disso, observa-se um desvio padrão bastante pequeno, concluindo-se assim que não existem grandes diferenças entre o número de iterações necessárias para o algoritmo de segmentação parar nas várias imagens.

Tabela 8.5. Medidas estatísticas do tempo (em segundos) da execução do algoritmo de segmentação nas várias imagens.

| | Média | Desvio padrão | Mínimo | Máximo |
|------------------|--------------|----------------------|---------------|---------------|
| Tempo (s) | 9,3229 | 2,3038 | 2,3076 | 17,2952 |

8.2. Reconstrução tridimensional

Tal como anteriormente mencionado, o método de reconstrução 3D usado foi o SfS, com a abordagem feita por Tsai. Como o algoritmo é iterativo, analisar-se-á a evolução dos resultados entre cada iteração.

Como se mostra na Figura 8.9, resultado da reconstrução no final das várias iterações do algoritmo aplicado à *frame* da Figura 6.1, existe uma clara diferença nas várias imagens. Podem-se tirar várias conclusões a partir da análise das várias imagens desta figura:

- Na primeira imagem a reconstrução é bastante fraca, devido principalmente a uma inicialização sem qualquer relação com a imagem usada, o que torna difícil ao algoritmo calcular com grande exatidão a profundidade de cada ponto da imagem
- Ao longo das iterações verifica-se uma suavização da reconstrução resultante, principalmente a partir das 3^a e 4^a iterações, sendo que a partir dessa iteração os resultados são muito semelhantes entre si, apresentando apenas uma pequena suavização em certas estruturas.
- Apresentou-se na última imagem a iteração número trinta para se verificar que este é um algoritmo de rápida convergência, já que mesmo quando aplicadas muitas iterações, os resultados não apresentam melhorias significativas.

É possível fazer a consideração que o nível de intensidade de cada imagem pode ser proporcional à profundidade do píxel correspondente. Assim sendo, fazendo a transformação de uma *frame* do espaço RGB para níveis de cinzentos, é possível fazer também uma reconstrução tridimensional considerando que a intensidade corresponde exatamente à altura do tecido, tendo como resultado a Figura 8.10a. Pode-se comparar este resultado com o obtido após 30 iterações com o algoritmo SfS (Figura 8.10b), para perceber as vantagens do uso de um algoritmo destes.

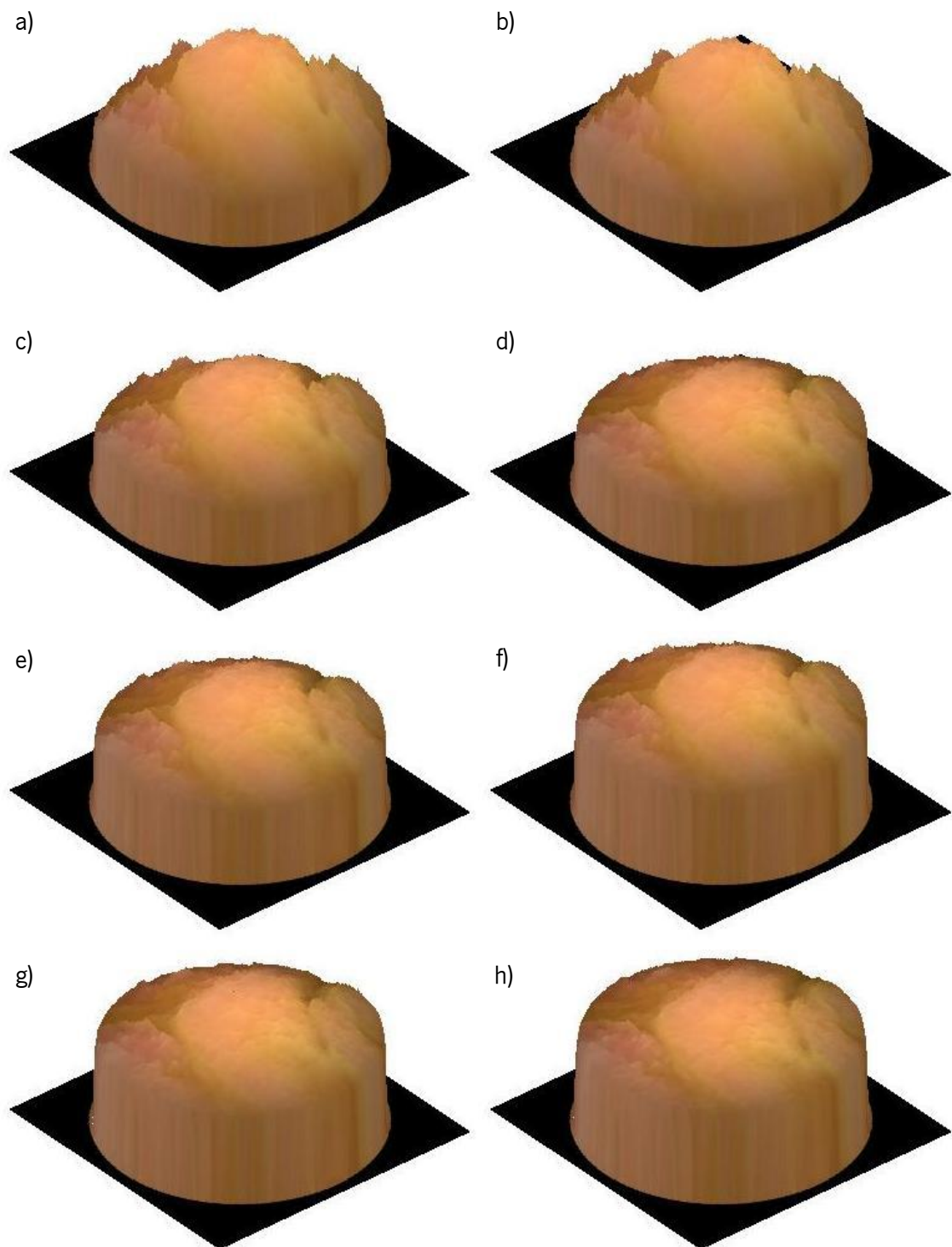


Figura 8.9. Resultado da reconstrução tridimensional da *frame* da Figura 6.1, da iteração número 1 até a iteração número 7 (a-g) e iteração número 30 (h).

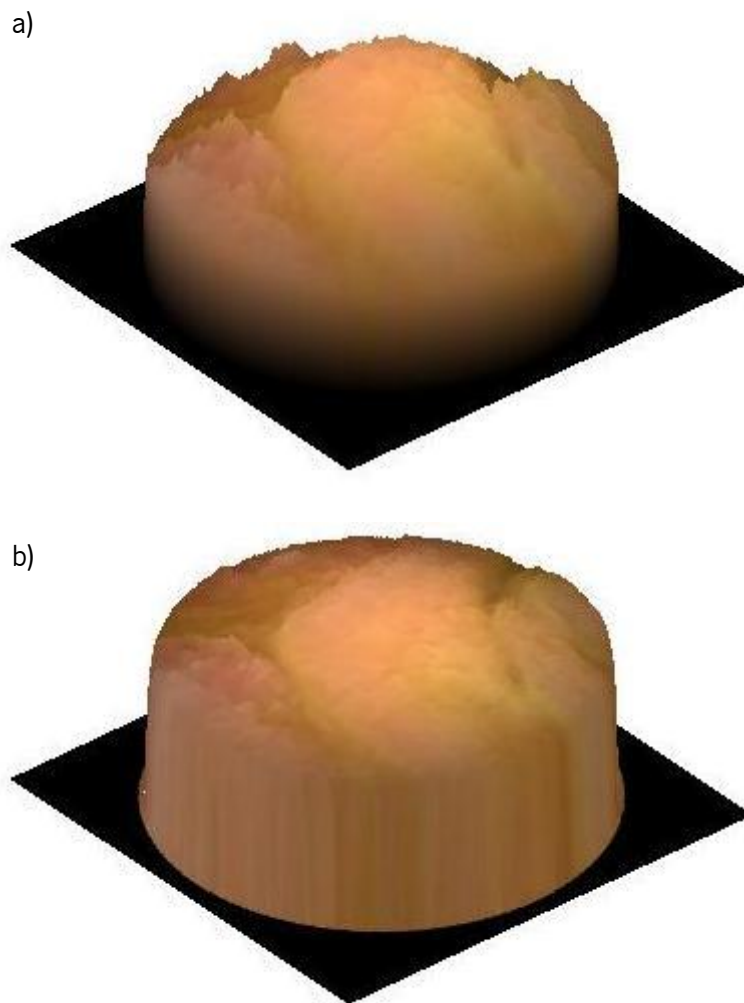


Figura 8.10. Comparação de reconstrução tridimensional usando apenas a imagem original (a) ou o algoritmo SfS (b).

Analisando a Figura 8.10 é possível verificar a vantagem do algoritmo SfS. Com o uso apenas da intensidade como profundidade do tecido vê-se uma menor suavidade o que torna o tecido bastante irrealista em relação ao resultado do algoritmo SfS. É de salientar também a melhoria desta reconstrução partindo apenas da intensidade com a reconstrução resultante das iterações números um e dois (Figura 8.9a e Figura 8.9b), o que considera que esta podia ser uma boa melhoria da inicialização deste algoritmo num trabalho futuro.

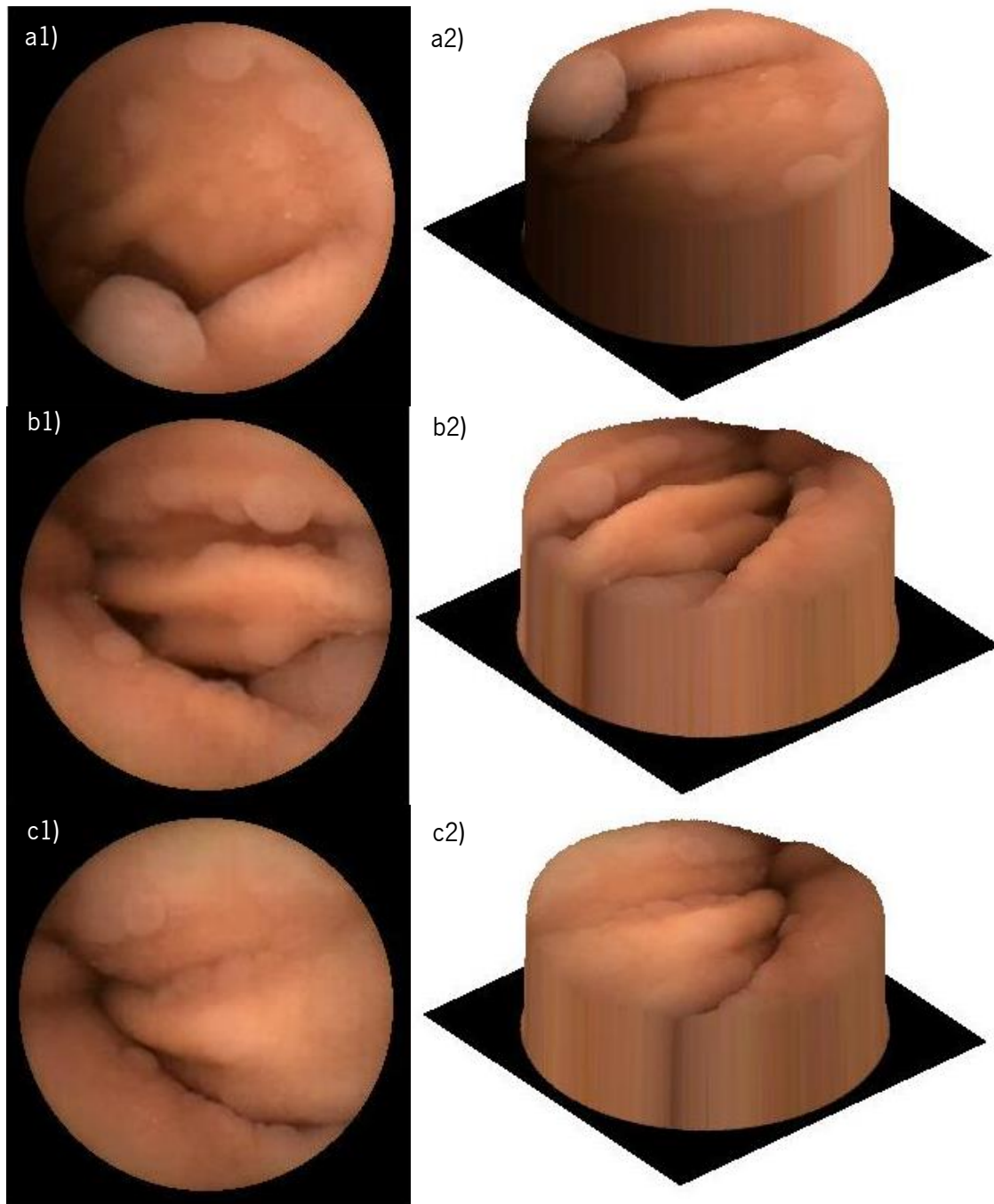


Figura 8.11. Exemplo de reconstrução usando o SfS de três frames diferentes - à esquerda as frames da CE (a1, b1 e c1) e à direita as respectivas reconstruções (a2, b2 e c2).

Foi também testado o algoritmo SfS com outras *frames* retiradas da CE, para provar a robustez do algoritmo e a sua usabilidade com diferentes exames e pacientes. Escolheram-se três exemplos com elevações naturais de tecido intestinal, para se conseguir visualizar melhor os resultados em três dimensões. Estes resultados podem ser vistos na Figura 8.11, onde é possível

concluir que realmente a inserção de uma terceira dimensão aumenta a qualidade na análise das formas dos tecidos, apesar de a visualização de apenas uma perspectiva como é feito nas figuras apresentadas não permitir ter uma total percepção de toda a informação obtida.

8.2.1. Tempo de execução

Correu-se o algoritmo construído de reconstrução tridimensional para 100 *frames* consecutivas de um exame de CE. Mediu-se o tempo despendido para que o SfS convergisse, e os resultados desses tempos podem ser vistos na Tabela 8.6. Pode-se ver que o algoritmo converge com um bom tempo, sendo que para a aplicação pretendida, o maior objetivo não é ser um método muito rápido. É possível observar também que existe pouca variação do tempo nas diferentes imagens, sendo que não existe uma grande variação nas iterações necessárias para o algoritmo convergir.

Tabela 8.6. Medidas estatísticas do tempo (em segundos) da execução do algoritmo de segmentação nas várias imagens.

| | Média | Desvio padrão | Mínimo | Máximo |
|------------------|--------------|----------------------|---------------|---------------|
| Tempo (s) | 0,7539 | 0,1203 | 0,4649 | 0,9466 |

8.3. Detecção automática de tumores

A deteção automática de tumores do intestino delgado apresentada no âmbito desta tese baseia-se no conceito de ROI, região esta selecionada de modo puramente automático. Esta solução requer um algoritmo de segmentação particularmente robusto que separe o tecido normal do tecido tumoral (caso exista) existentes numa mesma *frame*. As *features* através das quais isto se consegue têm aqui um papel importante no que aos requisitos de cálculo computacional diz respeito. Neste contexto, os dois objetivos mais importantes desta secção são:

- Aferir a eficácia do uso de *features* de natureza estatística retiradas do histograma (representando assim a cor das regiões), em relação a características que representam variações de textura numa imagem. Tipicamente as *features* baseadas na cor requerem menor cálculo computacional levando a versões mais rápidas de diagnóstico automático, sendo contudo apresentadas na maioria da literatura como tendo menor potencial discriminatório entre tecido normal e tumoral. Foram realizados testes usando descritores de textura retirados da DWT e da DCT, comparando com testes semelhantes usando *features* de cor.

- A eficácia da separação das imagens em duas regiões distintas, usando depois a correlação entre vetores de *features* para a classificação. Esta abordagem é nova precisando por isso de uma validação mais cuidada. Foram realizados testes em que foram usadas *features* (da mesma natureza) das imagens inteiras e das imagens segmentadas para poder aferir a eficiência relativa do método.

8.3.1. Resultados com segmentação

O objetivo neste ponto do trabalho era a obtenção de *features* que tivessem como significado a relação dos diferentes descritores nas duas regiões das imagens, tal como os próprios valores deles. Para isso, foi calculada a correlação dos descritores para os diferentes canais, entre as duas regiões. Com as *features* de cor não foram consideradas diferentes direções, mas quando foram usadas as transformadas DWT e DCT são consideradas. Na Equação (8.2) mostra-se a equação usada para o cálculo desta correlação, onde i é o canal de cor, x e y as regiões consideradas, j a direção atual, J o número total de direções consideradas, e X o vetor que contém os descritores com aqueles parâmetros.

$$\phi(i)_{R_x, R_y} = E \left\{ X_i^{R_x} \cdot X_i^{R_y} \right\} = \sum_{j=1}^J X_{i,j}^{R_x} \cdot X_{i,j}^{R_y} \quad (8.2)$$

As *features* de cor propostas na secção 6.2.1. foram divididas em quatro conjuntos diferentes para melhor poder verificar qual o contributo de cada uma para o desempenho global do classificador. Esta organização pode ser vista na Tabela 8.7, onde é possível verificar que a organização escolhida foi feita de maneira a que a média estivesse sempre presente, já que se apresenta na maioria da literatura como o mais significativo na classificação das *frames*.

Tabela 8.7. Organização das *features* de cor utilizadas neste trabalho.

| Set | Features |
|------------|---|
| A | Média (μ) e variância (σ^2) |
| B | Média (μ) e entropia (H) |
| C | Média (μ) e curtose (k) |
| D | Média (μ) e percentil de 95% ($x_{95\%}$) |

A Tabela 8.8 e a Tabela 8.9 mostram o desempenho do sistema de classificação proposto com as *features* de cor e com a inclusão do método de segmentação para dividir a imagem em duas regiões, usando os classificadores MLP e SVM, respectivamente. Calculando os coeficientes de correlação entre os descritores das várias regiões, cada conjunto tem como tamanho seis *features* (três coeficientes por espaço de cor).

Tabela 8.8. Resultados usando o classificador MLP e correlação entre *features* de diferentes regiões retiradas dos histogramas dos canais *a* e *b* do Lab.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|----------|-------------------|--------------------|--------------|
| A | 99,0 | 100,0 | 99,6 |
| B | 99,0 | 99,0 | 99,0 |
| C | 86,2 | 95,3 | 91,7 |
| D | 98,5 | 99,0 | 98,8 |

Tabela 8.9. Resultados usando o classificador SVM e correlação entre *features* de diferentes regiões retiradas dos histogramas dos canais *a* e *b* do Lab.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|----------|-------------------|--------------------|--------------|
| A | 88,3 | 100,0 | 95,4 |
| B | 87,8 | 95,3 | 94,4 |
| C | 63,8 | 96,7 | 80,2 |
| D | 84,7 | 97,7 | 93,5 |

Os resultados das tabelas anteriores mostram que a média e variância (conjunto A) apresentam valores melhores tanto no classificador MLP como no classificador SVM.

Fazendo a análise estatística do cálculo dos valores-p para a conclusão da significância de todas as *features* dos vários conjuntos, foi possível determinar que todas elas tinham significância (valor-p menor que 0,0005), exceto no caso do elemento do set C referente à autocorrelação da média e curtose da região sem tumor do canal *a*. Quando esta *feature* é retirada do set onde está presente, a performance do sistema é aumentada, embora não significativamente. Estes novos resultados são apresentados na Tabela 8.10.

Tabela 8.10. Resultados da classificação com os classificadores MLP e SVM usando o set C sem a *feature* sem significância.

| Classificador | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|---------------|-------------------|--------------------|--------------|
| MLP | 86,2 | 96,0 | 92,1 |
| SVM | 64,3 | 96,3 | 80,2 |

Com o objetivo de comparar a eficácia das *features* propostas anteriormente em relação às *features* usadas frequentemente para a caracterização de textura; foram realizados testes semelhantes aos anteriores, usando *features* retiradas da DWT e DCT dos canais *a* e *b* do espaço Lab. Neste caso, as transformadas são obtidas nas *frames* completas, mas as *features* estatísticas são calculadas para cada uma das regiões resultantes da segmentação separadamente. Deste modo, os resultados da Tabela 8.8 e da Tabela 8.9 em relação aos da Tabela 8.11 e da Tabela 8.12 diferem apenas da natureza das *features* usadas para calcular os elementos de correlação, sendo na primeira estatísticas dos histogramas dos canais *a* e *b* e na segunda são estatísticas dos histogramas de cada uma das escalas e direções da transformada wavelet e curvelet. As medidas estatísticas escolhidas no caso da DWT e DCT foram a média e variância, porque em estudos anteriores foi mostrado que estas representavam bem a informação de textura de tecidos intestinais (Martins et al. 2010). Apenas as direções da primeira escala foram consideradas, tanto na DWT como na DCT, dado que é nas altas frequências que a maioria da informação de textura está presente (Barbosa et al. 2008a). No caso da DWT são usadas três direções, enquanto que na DCT são usadas 16 direções diferentes.

Tabela 8.11. Resultados usando o classificador MLP e correlação entre *features* de textura retiradas de diferentes regiões dos canais *a* e *b* do Lab.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|-----|-------------------|--------------------|--------------|
| DWT | 96,4 | 97,0 | 96,8 |
| DCT | 96,9 | 97,0 | 97,0 |

Tabela 8.12. Resultados usando o classificador SVM e correlação entre *features* de textura retiradas de diferentes regiões dos canais *a* e *b* do Lab.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|-----|-------------------|--------------------|--------------|
| DWT | 71,4 | 94,7 | 85,5 |
| DCT | 70,4 | 98,0 | 87,1 |

Ao comparar a Tabela 8.8 com a Tabela 8.11 e a Tabela 8.9 com a Tabela 8.12 podem-se tirar algumas conclusões importantes:

- Só os coeficientes de correlação retirados das transformadas de DWT e DCT usando o classificador MLP é que apresentam resultados, embora inferiores aos apresentados na Tabela 8.8. Mas mesmo assim, nenhum destes resultados é superior aos melhores resultados usando *features* de cor. Podemos então

concluir, no âmbito deste estudo que as transformadas *wavelet* e *curvelet* apresentam um desempenho um pouco inferior ao apresentado pelas medidas retiradas dos histogramas dos canais *a* e *b*. Este resultado pode estar relacionado com a rejeição das escalas referentes às frequências baixas (que mesmo tendo pouca informação de textura, esta não é nula). Assim, ao fazer a codificação da informação da presença de tumor com o uso de *features* de cor, além de se melhorar, embora apenas ligeiramente nos melhores casos, os resultados finais da classificação, reduz-se o peso computacional requerido para o cálculo das transformadas de *wavelet* e *curvelet*, aumentando significativamente a velocidade do algoritmo.

- A DCT apresenta melhores resultados que a DWT em todos os testes realizados, o que não é surpreendente, porque a primeira apresenta melhor discriminação direcional, codificando assim a informação de textura de uma forma mais completa.
- O desempenho de MLP's é significativamente melhor que o de SVM's em todas as características. Isto pode dever-se ao grande número de parâmetros existentes nos SVM's requerendo um processo de otimização baseado na afinação de todos esses parâmetros, o que está fora do âmbito desta tese.

Em relação aos testes estatísticos de significância das *features*, os valores retirados do canal *b* foram considerados não significativos, enquanto que os valores-p das *features* do canal *a* tiveram valores menores que 0,0001. Para tentar melhorar a performance dos sistemas apresentados, retirou-se a *feature* com maior valor-p (já que representa aquela que menos contribui para a correta divisão de classes). Voltou-se a classificar o conjunto de *features*, chegando aos resultados apresentados na Tabela 8.13.

Tabela 8.13. Resultados usando correlação de *features* de textura entre as duas regiões, sem a *feature* com maior valor-p acima do nível de significância.

| Set | Classificador | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|-----|---------------|-------------------|--------------------|--------------|
| DWT | MLP | 96,4 | 96,7 | 96,6 |
| | SVM | 71,4 | 94,3 | 85,3 |
| DCT | MLP | 97,4 | 98,0 | 97,8 |
| | SVM | 70,9 | 98,0 | 87,3 |

Pode-se ver aqui que no caso da transformada *wavelet* os resultados pioraram algumas décimas; podendo concluir que ao retirar outras *features* sem significância não iria levar a um aumento da performance do sistema. Em relação à análise estatística das *features* da DCT, pode-se ver que houve uma melhoria substancial dos resultados em ambos os classificadores ao retirar a *feature* com um maior valor-p. Assim sendo, decidi retirar-se a *feature* com o segundo maior valor-p para ver se era possível aumentar ainda mais a performance do sistema, continuando com a tendência. Chegou-se aos resultados apresentados na Tabela 8.14, onde é possível verificar um retrocesso na performance do sistema, havendo uma redução de algumas décimas em ambos os classificadores.

Tabela 8.14. Resultados usando correlação de *features* de textura da DCT entre as duas regiões, sem as duas *features* com maior valor-p acima do nível de significância.

| Classificador | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|----------------------|--------------------------|---------------------------|---------------------|
| MLP | 97,4 | 97,0 | 97,2 |
| SVM | 71,4 | 97,3 | 87,1 |

Habitualmente as transformadas *wavelet* e *curvelet* aplicadas em imagens não são calculadas em canais do espaço de cor Lab, mas sim em espaços de cor mais usuais como RGB e HSV. Em particular o espaço HSV é tido como apresentando grande potencial por ter por base alguns mecanismos da percepção humana. Por esta razão testou-se também o algoritmo desenvolvido para este propósito no âmbito desta tese no espaço de cor HSV. Os resultados são apresentados na Tabela 8.15 e na Tabela 8.16 usando o classificador MLP e SVM, respetivamente.

Tabela 8.15. Resultados usando o classificador MLP e correlação entre *features* de textura retiradas de diferentes regiões dos canais do espaço HSV.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|------------|--------------------------|---------------------------|---------------------|
| DWT | 76,5 | 93,3 | 86,7 |
| DCT | 87,2 | 96,0 | 92,5 |

Tabela 8.16. Resultados usando o classificador SVM e correlação entre *features* de textura retiradas de diferentes regiões dos canais do espaço HSV.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|-----|-------------------|--------------------|--------------|
| DWT | 61,2 | 96,3 | 82,5 |
| DCT | 62,8 | 97,3 | 83,7 |

Com os resultados apresentados nas tabelas acima e comparando com a Tabela 8.11 e a Tabela 8.12 é possível inferir que os canais *a* e *b* do Lab contêm mais informação de textura quando são aplicadas as transformadas *wavelet* e *curvelet*. Esta comparação pode provavelmente não ser a melhor devido à diferente dimensão dos problemas (2 canais de cor no Lab em relação a 3 canais no HSV), mas é a possível neste caso já que não houve uma técnica de remoção as interferências da iluminação encontrada que mantivesse as características do espaço HSV. Tal como foi possível concluir anteriormente, o uso da DCT leva a melhores resultados em relação à escolha da DWT; tal como o uso de MLP ser melhor que o uso de SVM.

Através da análise estatística das *features* dos conjuntos apresentados na Tabela 8.15 e na Tabela 8.16 foi possível concluir que todas as *features* do primeiro conjunto foram consideradas significativas, com um valor-p inferior a 0,001. No segundo conjunto (DCT), os coeficientes de correlação correspondentes ao canal V foram considerados não significativos. Assim, tal como anteriormente, foi retirada a *feature* com maior valor-p do conjunto e este novo conjunto inserido nos classificadores. Usando o MLP a precisão teve um valor semelhante, tendo havido um valor ligeiramente superior na especificidade e um proporcional na sensibilidade. Com o uso de SVM todas as métricas mantiveram o seu valor anterior. Ao retirar a segunda *feature* com maior valor-p chegou-se a resultados bastante inferiores em todas as métricas, para ambos os classificadores, o que pode significar que mesmo as *features* consideradas não significativas pelos testes estatísticos são importantes para a classificação das *frames*.

8.3.2. Resultados sem segmentação

Até aqui, nesta secção, apenas foram apresentados e discutidos os resultados referentes à inovadora escolha de *features* dos histogramas dos canais do espaço de cor Lab, descartando a informação de luz. Outra inovação é a divisão automática das imagens em duas sub-regiões, usando a segmentação de *frames*. Assim sendo, é possível separar tipos diferentes de tecidos nas imagens (como tecido normal e fluidos do tecido tumoral), para assim melhor caracterizar e realçar

a presença de tumores nas *frames*. Assim sendo foram realizados testes semelhantes aos anteriores, mas agora sem nenhum processo de segmentação, usando assim em conjunto os píxeis de toda a imagem para calcular as *features* de cor e textura. Em nenhum dos processos seguintes foi necessário o cálculo de quaisquer coeficientes de correlação entre regiões, já que estas não foram definidas.

A Tabela 8.17 e a Tabela 8.18 mostram os resultados com MLP e SVM, respetivamente, usando conjuntos de *features* organizados novamente da maneira explicada na Tabela 8.7. Neste caso cada set consistia num grupo de apenas quatro *features* (duas medidas estatísticas para cada um dos dois canais de cor).

Tabela 8.17. Resultados usando o classificador MLP e *features* retiradas dos histogramas dos canais *a* e *b* do Lab em toda a imagem.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|-----|-------------------|--------------------|--------------|
| A | 97,4 | 99,7 | 98,7 |
| B | 96,4 | 99,7 | 98,4 |
| C | 95,9 | 98 | 97,2 |
| D | 99 | 98,3 | 98,6 |

Tabela 8.18. Resultados usando o classificador SVM e *features* retiradas dos histogramas dos canais *a* e *b* do Lab em toda a imagem.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|-----|-------------------|--------------------|--------------|
| A | 96,9 | 99,0 | 98,2 |
| B | 88,8 | 94,7 | 92,3 |
| C | 86,2 | 92,3 | 89,9 |
| D | 95,4 | 98,7 | 97,3 |

Ao comparar a Tabela 8.8 e a Tabela 8.9 com a Tabela 8.17 e a Tabela 8.18 é possível verificar que o procedimento de segmentação aumenta bastante o desempenho do sistema. No melhor conjunto de cada um dos testes (conjunto A usando o classificador MLP), existe uma melhoria de 1,6% na sensibilidade e de 0,3% na especificidade, resultando numa melhoria geral do sistema (precisão) de 0,9%. Assim sendo, pode-se concluir que a relação entre as diferentes regiões de uma imagem ajudam bastante na classificação dessa *frame* como normal ou tumoral.

Os resultados da análise estatística permitiram concluir que a variância, entropia e curtose calculadas a partir dos píxeis do canal *b* são considerados não significativos. Por outro lado, todas as restantes *features* apresentaram valores-*p* inferiores a 0,001. Para verificar a influência que a inclusão das *features* sem significância tiveram nos resultados, retiraram-se essas *features*, voltando a classificar os vetores de características usando o MLP e o SVM. Os resultados podem ser vistos na Tabela 8.19 e na Tabela 8.20. Apesar de não ter sido muito significativo, a remoção destas *features*, levou de uma maneira geral a um decréscimo no desempenho do sistema.

Tabela 8.19. Resultados usando o classificador MLP e *features* com significância retiradas dos histogramas dos canais *a* e *b* do Lab em toda a imagem.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|----------|-------------------|--------------------|--------------|
| A | 97,4 | 99,3 | 98,6 |
| B | 96,4 | 98,7 | 97,8 |
| C | 94,4 | 98,7 | 97,0 |

Tabela 8.20. Resultados usando o classificador SVM e *features* com significância retiradas dos histogramas dos canais *a* e *b* do Lab em toda a imagem.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|----------|-------------------|--------------------|--------------|
| A | 93,4 | 100,0 | 97,4 |
| B | 89,8 | 93,3 | 91,9 |
| C | 74,5 | 97,0 | 88,1 |

Foram também realizados testes com *features* de textura considerando as *frames* inteiras, semelhantes aos realizados anteriormente com a correlação entre regiões. Computaram-se assim as transformadas *wavelet* e *curvelet* das imagens inteiras, calculando-se depois medidas estatísticas de cada uma das direções na escala das maiores frequências. A DCT considera um maior número de direções em relação à DWT, logo, o tamanho do vetor de características será maior na primeira. Para poder comparar de forma justa os resultados da classificação usando DWT e DCT, decidiu-se calcular os elementos da matriz de covariâncias entre cores para cada medida estatística. Este procedimento resultou num tamanho de 6 *features* para o caso dos canais *a* e *b* do Lab (Tabela 8.21 e Tabela 8.22) e para um tamanho de 12 *features* para o uso do espaço de cor HSV (Tabela 8.23 e Tabela 8.24).

Tabela 8.21. Resultados usando o classificador MLP e *features* de textura retiradas dos canais *a* e *b* do Lab em toda a imagem.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|------------|--------------------------|---------------------------|---------------------|
| DWT | 88,8 | 92,7 | 91,1 |
| DCT | 95,4 | 97,7 | 96,8 |

Tabela 8.22. Resultados usando o classificador SVM e *features* de textura retiradas dos canais *a* e *b* do Lab em toda a imagem.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|------------|--------------------------|---------------------------|---------------------|
| DWT | 73,5 | 96,3 | 87,3 |
| DCT | 68,4 | 96,3 | 85,3 |

Ao comparar a Tabela 8.21 e Tabela 8.22 com a Tabela 8.11 e Tabela 8.12 é possível tirar certas conclusões:

- O desempenho das *features* da DCT foi superior melhor que as da DWT usando o classificador MLP, mas os resultados usando o SVM foram contrários, ou seja, o desempenho da DWT foi um pouco melhor que o da DCT. Este facto contraria o esperado *a priori* pois a melhor resolução direcional apresentada pela DCT contribui em princípio para uma melhor caracterização de textura. Esta conclusão é frequentemente encontrada na literatura da especialidade pelo que a não otimização do classificador SVM, que está fora do âmbito desta tese, deve estar relacionada com a obtenção destes resultados.
- Em relação à comparação entre estes resultados e aqueles com as *features* retiradas do resultado da segmentação das *frames*, pode-se concluir que usando o método de segmentação, existe uma melhoria dos resultados (mais significativa usando DWT do que DCT) quando usado o classificador MLP. Ao observar os resultados com SVM isto não se observa, o que se pode explicar com o que foi referido no ponto anterior.
- O melhor resultado deste teste é ainda aproximadamente 2% inferior ao melhor caso do uso de *features* de cor usando a divisão de imagens através de segmentação, sendo que este último apresenta a vantagem acrescida de ser significativamente mais rápido (Tabela 8.26). Este fato é importante dado o volume de *frames* contido em cada exame.

Em relação à análise estatística das *features* dos dois conjuntos usados neste teste, todas elas foram consideradas significativas, com valores-p inferiores a 0,0001. Consequentemente não foram realizados testes adicionais com remoção de *features*.

Foram também gerados resultados com *features* de textura provenientes da DWT e da DCT, mas a partir do espaço de cor HSV, tal como tinha sido feita com a separação das imagens em duas regiões. Esses resultados apresentam-se na Tabela 8.23 e na Tabela 8.24.

Tabela 8.23. Resultados usando o classificador MLP e *features* de textura retiradas do espaço HSV em toda a imagem.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|-----|-------------------|--------------------|--------------|
| DWT | 87,8 | 95,0 | 92,1 |
| DCT | 89,3 | 96,3 | 93,5 |

Tabela 8.24. Resultados usando o classificador SVM e *features* de textura retiradas do espaço HSV em toda a imagem.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|-----|-------------------|--------------------|--------------|
| DWT | 64,3 | 98,3 | 84,9 |
| DCT | 68,4 | 97,3 | 85,9 |

Neste caso os resultados obtidos estão completamente de acordo com o esperado, já que o uso das *features* resultantes da DCT resultam num melhor desempenho do sistema comparativamente com o uso das *features* resultantes da DWT. É possível verificar também que os resultados com o classificador MLP são sempre melhores que os resultados com o SVM, tal como já tinha ocorrido em todos os testes anteriores. Conclui-se também que houve uma melhoria quando se compara com os resultados usando as mesmas *features* mas depois do processo de segmentação (Tabela 8.15 e Tabela 8.16), o que não era esperado. Mas, apesar de isto acontecer, o melhor resultado continua a ser inferior àquele que se conseguiu obter usando as *features* de cor provenientes do algoritmo de segmentação (diferença de mais de 5% na precisão).

Com a análise estatística efetuada sobre as *features* aqui apresentadas chegou-se à conclusão que todas elas são estatisticamente significativas, com um valor-p inferior a 0,001.

8.3.3. Comparação com estado da arte

Tal como abordado no Capítulo 6, existem métodos com bastante maior complexidade para obter informação de textura que costumam ser usados para a resolução do problema de deteção automática de tumores em *frames* da CE. Dentro destes métodos salientam-se os

descritores de Haralick (Haralick et al. 1973). Estes algoritmos são computacionalmente mais pesados que os baseados no uso das transformadas *wavelet* e *curvelet*. Não se utilizaram, no âmbito desta tese, *features* semelhantes aplicadas às imagens divididas em duas regiões porque as matrizes de coocorrência não têm o melhor comportamento e não codificam bem a informação necessária para obter dados de textura quando retiradas de regiões tão irregulares como aquelas que resultam do processo de segmentação utilizado.

Para comparar o método proposto com métodos já apresentados no estado da arte são apresentados na Tabela 8.25 alguns resultados de algoritmos testados nas imagens da base de dados usada neste trabalho, tal como os melhores resultados com as *features* propostas aqui (com e sem segmentação).

Tabela 8.25. Comparação entre *features* propostas neste trabalho e estado da arte.

| Set | Sensibilidade (%) | Especificidade (%) | Precisão (%) |
|---|--------------------------|---------------------------|---------------------|
| Features de histograma (Kodogiannis et al. 2007) | 84,7 | 85,5 | 85,3 |
| Features de N_{uv} (Kodogiannis et al. 2007) | 88,2 | 89,1 | 88,7 |
| Momentos de 2ª ordem (Barbosa et al. 2010) | 88,0 | 88,1 | 87,9 |
| Momentos de 2ª e 3ª ordem (Barbosa et al. 2010) | 90,1 | 91,2 | 90,9 |
| Análise Multiescala (Barbosa et al. 2012) | 93,1 | 93,9 | 93,4 |
| Features propostas sem segmentação | 97,4 | 99,7 | 98,7 |
| Features propostas com segmentação | 99,0 | 100,0 | 99,6 |

Os dois primeiros métodos são apresentados em (Kodogiannis et al. 2007), em que são retiradas *features* de textura nos domínios cromáticos e acromáticos. O primeiro é uma abordagem estatística que calcula nove descritores de textura dos histogramas de cada um dos canais dos

espaços RGB e HSV. O segundo consiste na definição de unidades elementares de textura (TU), que são a unidade mais pequena mas completa que melhor caracteriza a textura local de um dado píxel e a sua vizinhança; e números de unidades de textura (N_{TU}), que descrevem a textura de cada TU. Neste caso, as *features* são as mesmas nove medidas estatísticas dos histogramas construídos a partir do N_{TU} . Os resultados apresentados na 3ª e 4ª linha são relativos ao estudo apresentado em (Barbosa et al. 2010) onde é usada estatística de 2ª e/ou 3ª ordem para modelar as *features* estatísticas de textura de Haralick, sem as escalas de baixa frequência da DWT. A 5ª linha apresenta os resultados retirados do estudo apresentado em (Barbosa et al. 2012), onde foi utilizada a mesma base de dados que nesta dissertação e onde as *features* utilizadas foram a correlação das medidas estatísticas retiradas das matrizes de coocorrência, entre os vários espaços de cor e as escalas de maiores e menores frequências. As últimas duas linhas apresentam os resultados obtidos com os métodos propostos.

É possível verificar que o método proposto apresenta melhoria de desempenho de pelo menos 5% quando não é usada segmentação, e na ordem dos 6% quando é usada segmentação.

8.3.4. Tempo de execução

A Tabela 8.26 mostra o tempo de execução das várias versões do algoritmo, onde se verifica uma clara redução do tempo computacional do algoritmo quando são usadas *features* de cor, em vez de *features* de textura baseadas em transformadas complexas. Verifica-se ainda que os maiores desvios padrão ocorrem quando é usada a segmentação, o que leva a grandes diferenças nos tempos de processamento entre diferentes imagens, que podem ser contudo reduzidas alterando os critérios de oaragem do algoritmo de segmentação.

O método que usa *features* de cor com segmentação, além de apresentar resultados bastantes promissores, se tiver corrigido o desvio padrão para que o máximo tempo de processamento seja aproximadamente de 0,5 segundos, pode permitir ser incluído num sistema em tempo real. Assim será possível a classificação das *frames* ao mesmo tempo que estas estão a ser adquiridas (já que a CE tem um *frame rate* de 2 fps).

Na Tabela 8.27 pode-se ver a comparação do desempenho (precisão) de cada uma das versões do algoritmo, tal como o tempo que seria necessário para executar estes algoritmos num vídeo inteiro de um exame de CE (considerando que em média este tem 55000 *frames*).

Tabela 8.26. Medidas estatísticas do tempo (em segundos) da execução do algoritmo de segmentação com vários parâmetros nas várias imagens.

| | | Média | Desvio padrão | Mínimo | Máximo |
|------------------------|-----------------------------|--------------|----------------------|---------------|---------------|
| Com segmentação | Features de cor | 0,4263 | 0,1204 | 0,1892 | 1,2630 |
| | Features da Wavelet | 0,4760 | 0,1003 | 0,3019 | 1,2927 |
| | Features da Curvelet | 0,9598 | 0,1476 | 0,5177 | 2,5452 |
| Sem segmentação | Features de cor | 0,0817 | 0,0098 | 0,0766 | 0,2862 |
| | Features da Wavelet | 0,2058 | 0,0612 | 0,1477 | 1,1775 |
| | Features da Curvelet | 0,5752 | 0,0552 | 0,3746 | 1,0405 |

Tabela 8.27. Tempo de processamento de um exame completo e precisão de cada uma das versões do algoritmo testadas.

| | | Tempo de processamento de exame completo (h) | Precisão (%) |
|------------------------|-----------------------------|---|---------------------|
| Com segmentação | Features de cor | 6,51 | 99,6 |
| | Features da Wavelet | 7,27 | 96,8 |
| | Features da Curvelet | 14,66 | 97,8 |
| Sem segmentação | Features de cor | 1,25 | 98,7 |
| | Features da Wavelet | 3,14 | 91,1 |
| | Features da Curvelet | 8,79 | 96,8 |

Ao analisar a Tabela 8.27, pode-se verificar que usando *features* de cor com segmentação apresenta uma precisão bastante superior aos restantes, mas o tempo de processamento não é

ainda o melhor das várias versões dos algoritmos. É de salientar ainda o tempo de processamento muito abaixo dos restantes ao fazer a extração de *features* de cor sem um método de segmentação associado; e ao mesmo tempo um desempenho do sistema que está ao nível do uso de segmentação. A terceira versão do algoritmo com um maior desempenho é o uso de coeficientes da DCT com o uso da segmentação, mas este leva quase 15 horas a executar no vídeo completo de um exame de CE.

Capítulo 9 Conclusão e trabalho futuro

Este capítulo apresenta as principais conclusões retiradas da elaboração do trabalho descrito ao longo dos capítulos anteriores. Serão também sugestões para trabalho futuro e apresentadas as principais contribuições decorrentes da execução deste mesmo trabalho.

9.1. Conclusões

O trabalho desenvolvido no âmbito desta tese pode dividir-se em termos de objetivos, em duas partes distintas que partilham contudo um módulo comum; o módulo de segmentação de imagem. A primeira parte tem como objetivo a reconstrução tridimensional da mucosa intestinal a partir de *frames* da cápsula endoscópica e a segunda parte tem como objetivo o uso do conceito de ROI não rígido e completamente automático para a deteção automática de *frames* com presença de tumor. Ambas as partes necessitam de um segmentador robusto dado a variabilidade nas condições de luminosidade decorrentes da impossibilidade de controlo da distância entre a câmara e o tecido a ser filmado, o que origina frequentemente saturação e perda de informação relativamente à textura e cor do tecido atingido pela saturação luminosa. No caso da reconstrução 3-D o segmentador terá que apresentar ainda capacidade discriminatória para as diferentes estruturas que compõem a mucosa como sejam os pólipos, úlceras entre outras. Para o caso da

deteção automática de tumores o segmentador deverá ser capaz de discriminar tecido normal de tecido lesionado.

Para a reconstrução 3-D implementou-se um módulo de segmentação de tecidos baseado em Misturas de Gaussianas Multivariável, em que cada observação é composta pela intensidade no canal *a* e no canal *b* do espaço de cor Lab, sendo a luminosidade descartada para normalização deste parâmetro. Como é desconhecido *a priori* o número de regiões em cada imagem, obteve-se este número usando o critério de máxima verosimilhança, ou seja, efetuou-se a segmentação em vários números de regiões, comparando o valor de BIC em cada uma delas, minimizando-o de seguida. Com as regiões definidas, foi feito um pós-processamento para suavizar a delimitação das regiões e extraídas *features* de cor e textura de forma a caracterizar as regiões. A deteção de regiões correspondentes em *frames* adjacentes, necessária à interpolação de *frames* que permite a obtenção de um vídeo com movimento natural (sem variações abruptas causadas pelos movimentos peristálticos), foi efetuada recorrendo a coeficientes de correlação entre as várias regiões selecionadas em ambas as *frames*. Esta comparação entre regiões apresentou resultados promissores, já que nos casos apresentados são mostradas as regiões consideradas correspondentes em *frames* consecutivas, e que na maioria das vezes correspondem ao mesmo resultado obtido visualmente. Para a reconstrução de cada *frame* individualmente, foi implementado o algoritmo SfS, que resultou numa boa representação da forma das várias estruturas nos vários exemplos testados.

A segunda parte do trabalho, relativa à deteção automática de tecido tumoral em *frames* da CE usou alguns dos princípios usados na primeira parte. Assim, decidiu-se usar um método de segmentação para fazer a divisão do tecido em duas partes diferentes e usar *features* de natureza estatística de cor retiradas do histogramas dos canais *a* e *b* do espaço de cor Lab para calcular coeficientes de correlação destas *features* para serem usados como descritores para classificar as imagens. Através deste método foi possível conseguir resultados de precisão de 99,6%; o que é um resultado bastante bom; principalmente quando comparado com resultados usando descritores de textura calculados através da DWT (91,6%) e da DCT (97,8%), ou resultados referência do estado da arte (93,4% - (Barbosa et al. 2012)).

9.2. Contribuições originais

Ao longo do trabalho realizado nesta dissertação houve alguns métodos desenvolvidos que se apresentam como contribuições originais, nomeadamente:

- A remoção da informação da luminosidade das imagens, para esta não interferir com a segmentação, ao usar o espaço de cor Lab. Este espaço de cor tem um nível de correlação quase nulo entre o canal L (relativo à informação de luz) e os canais a e b (relativos à informação de cor), o que permite a remoção deste canal do método de segmentação sem remover informação de cor importante.
- O uso de Regiões de interesse dinâmicas entre as diferentes imagens para a deteção de tecido tumoral em *frames* da CE, usando uma abordagem MAP para a segmentação. No atual estado da arte, as *frames* são consideradas sempre como um todo, exceto em (Mackiewicz et al. 2008) onde realmente são usadas ROI, mas estas são rígidas e iguais para todas as imagens. Não é por isso feito nenhum esforço na separação de tecidos de natureza diferente antes da extração de *features*, sendo analisada uma região fixa em torno do centro da imagem com porventura mais uniformidade na distribuição luminosa.
- O uso de descritores de cor retirados dos histogramas dos canais a e b do espaço Lab, para usar em classificadores supervisionados e detetar automaticamente a presença de tecido tumoral em *frames* da CE. Os descritores usados atualmente são de textura, ou então quando se considera a cor, é sempre incluída informação de luz que retira performance aos sistemas.
- A inclusão do verdadeiro número de LED's tal como a sua posição no modelo de iluminação da câmara da CE para o uso do algoritmo *Shape from Shading*, já que nos estudos anteriores, a iluminação é considerada como sendo apenas uma única fonte de luz, posicionada no centro do plano da câmara. Esta aproximação é apenas válida se a distância entre a câmara e o tecido filmado for muito maior que a distância entre os LED's, o que nem sempre acontece dentro do intestino delgado.

9.3. Trabalho futuro

A primeira parte do trabalho (reconstrução 3-D de *frames* da CE) tem certas partes em que pode ser melhorado, principalmente ao nível da segmentação realizada. Este método, no futuro, pode incluir informações de vizinhança dos píxeis para melhorar a definição dos contornos e tornar o resultado da segmentação um conjunto de regiões homogêneas e contínuas. Isto pode ser feito com a inclusão de probabilidades *a priori* calculadas usando a teoria de *Markov Random Fields* (MRF), que pode ser facilmente acrescentado no método já implementado para este trabalho. Uma possível melhoria pode ser também a partir de uma abordagem que inclua um sistema de estimação de movimento de regiões, que iterativamente com a segmentação pode melhorar não só a segmentação das regiões, mas também a própria reconstrução tridimensional do conjunto de *frames*. A estimação do movimento deve ser feita de forma hierárquica, de maneira a resolver um dos grandes problemas dos vídeos da CE que é o baixo *frame rate* e consequentemente, as variações abruptas entre *frames* consecutivas. Esta hierarquização é usualmente implementada com o uso de DWT, no entanto a DCT ao apresentar melhor resolução direcional é porventura mais apropriada para este propósito. A validação clínica dos resultados da reconstrução tridimensional seria também uma grande melhoria para comprovar a viabilidade do uso destes algoritmos, o que poderia ser feito com o uso de reconstruções pontuais de segmentos da CE, em comparação com reconstrução de RM intestinal.

Os bons resultados apresentados em relação à detecção de tumores em *frames* da CE necessitam de validação em bases de dados mais extensas, que não têm sido de fácil obtenção em parte devido à relativamente baixa incidência de tumor no intestino delgado.

Bibliografia

- Acharya, T. & Ray, A.K., 2005. *Image Processing: Principles and Applications* 1st ed., New Jersey: Wiley.
- Adler, D. & Gostout, C., 2003. Wireless Capsule Endoscopy - state of art. *Hospital Physician*, pp.14–22.
- Adler, S.N. & Metzger, Y.C., 2011. PillCam COLON Capsule Endoscopy. *Therapic Advances in Gastroenterology*, 4(4), pp.265–268.
- Alpaydin, E., 2010. *Introduction to Machine Learning* 2nd ed. T. Dietterich et al., eds., Cambridge, Massachusetts: The MIT Press.
- Anon, 2010. Color Theory. , p.ColoRotate. Available at: <http://learn.colorotate.org/color-models> [Accessed August 3, 2013].
- Anthony J. DiMarino, J. & Benjamin, S.B., 2002. *Gastrointestinal Disease: An Endoscopic Approach*, Slack Incorporated.
- Barbosa, D. et al., 2012. Automatic small bowel tumor diagnosis by using multi-scale wavelet-based analysis in wireless capsule endoscopy images. *Biomedical engineering online*, 11(1), p.3.
- Barbosa, D. et al., 2010. Detection of Small Bowel Tumors in Endoscopic Capsule Images by Modeling Non-Gaussianity of Texture Descriptors. *International Journal of Tomography & Statistics*, 14(S10), pp.41–55.
- Barbosa, D., Ramos, J. & Lima, C.S., 2008a. Detection of small bowel tumors in capsule endoscopy frames using texture analysis based on the discrete wavelet transform. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2008, pp.3012–5.
- Barbosa, D., Ramos, J. & Lima, C.S., 2008b. Wireless capsule endoscopic frame classification scheme based on higher order statistics of multi-scale texture descriptors. *4th European Conference of the International Federation for Medical and Biological Engineering - ECIFMBE 2008*, pp.200–203.
- Barbosa, D.J.C. et al., 2009. Automatic detection of small bowel tumors in capsule endoscopy based on color curvelet covariance statistical texture descriptors. In *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*. pp. 6683–6.

Bibliografia

- Beckmann, P. & Spizzichino, A., 1968. *The Scattering of Electromagnetic Waves from Rough Surfaces*, Books on Demand.
- Berkow, R. & Beers, M.H., 2003. *Manual Merck - Saúde para a família*, Oceano Grupo Editorial, S.A.
- Biernacki, C., Celeux, G. & Govaert, G., 2003. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis*, 41(3-4), pp.561–575.
- Bilmes, J., 1998. *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*, Berkeley.
- Borman, S., 2004. *The Expectation Maximization Algorithm - A short tutorial*,
- Boron, W.F. & Boulpaep, E.L., 2003. *Medical Physiology: A Cellular and Molecular Approach*, Saunders.
- Bourbakis, N., 2005. Detecting Abnormal Patterns in WCE Images. In *Fifth IEEE Symposium on Bioinformatics and Bioengineering (BIBE'05)*. Ieee, pp. 232–238. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1544472>.
- Boyle, W.S. & Smith, G.E., 1970. Charge Coupled Semiconductor Devices. *Bell System Technical Journal*, 49(4), p.587+.
- Brunicardi, F.C. et al., 2006. *Schwartz' Principles of Surgery : Self-Assessment and Board Review, Eighth Edition: Self-Assessment and Board Review, Eighth Edition*, Mcgraw-hill.
- Candès, E. et al., 2006. Fast Discrete Curvelet Transforms. *Multiscale Modeling & Simulation*, 5(3), pp.861–899.
- Cave, D.R. et al., 2008. A multicenter randomized comparison of the Endocapsule and the Pillcam SB. *Gastrointestinal endoscopy*, 68(3), pp.487–94.
- Chantler, M.J., Russell, G.T. & Linnett, L.M., 1994. Illumination: A directional filter of texture? *British Machine Vision Conference*, 2, pp.449–458.
- Clancy, F., 2008. Easy Pill to Swallow. *Minnesota Medicine*.
- Cohen, A. & Kovacevic, J., 1996. Wavelets: the mathematical background. *Proceedings of the IEEE*, 84(4), pp.514–522.
- Cotton, P.B. et al., 2009. *Practical Gastrointestinal Endoscopy: The Fundamentals*, Wiley.
- Cunha, J. et al., 2008. Automated Topographic Segmentation and Transit Time Estimation in Endoscopic Capsule Exams. *IEEE Transaction on Medical Imaging*, 27, pp.19–27.

- Davis, G.G., 1935. Applied anatomy: The construction of the human body considered in relation to its functions, diseases and injuries. *British Journal of Surgery*, 22(88), pp.896–897.
- Dicintio, S., 2012. *Comparing Approaches to Initializing the Expectation-Maximization Algorithm*. University of Guelph.
- Faltin, P. & Behrens, A., 2010. Projective 3D-reconstruction of uncalibrated endoscopic images. *Acta Polytechnica*, 50(4), pp.29–34.
- Fan, Y. & Meng, M.Q.-H., 2011. 3D reconstruction of the WCE images by affine SIFT method. In *2011 9th World Congress on Intelligent Control and Automation*. Ieee, pp. 943–947.
- Ford, A. & Roberts, A., 1998. *Colour Space Conversions*, London.
- Forster, C.H.Q. & Tozzi, C.L., 2000. Towards 3D Reconstruction of Endoscope Images using Shape from Shading. In *Proceedings of the 13th Brazilian Symposium on Computer Graphics and Image Processing*. Washington, DC, USA: IEEE Computer Society, pp. 90–96.
- Fusco, R., 2013a. Curing Ulcer Disease. *Three Rivers Endoscopy Center*. Available at: www.gihealth.com [Accessed April 23, 2013].
- Fusco, R., 2013b. Wireless Capsule Endoscopy “The Camera in a Pill.” *Three Rivers Endoscopy Center*. Available at: <http://www.gihealth.com/> [Accessed April 30, 2013].
- Gillson, S., 2004. New PillCam(TM) ESO for The Esophagus Receives FDA Marketing Clearance. *Heartburn / GERD Guide*. Available at: <http://heartburn.about.com/b/2004/11/29/new-pillcamtm-eso-for-the-esophagus-receives-fda-marketing-clearance.htm>.
- Given Imaging, 2013. *The PillCam® SB 3 System: Frequently Asked Questions*,
- González, A.M.C. et al., 2009. Video-endoscopic image analysis for 3D reconstruction of the surgical scene. In J. Vander Sloten et al., eds. *4th European Conference of the International Federation for Medical and Biological Engineering*. Springer Berlin Heidelberg, pp. 923–926.
- Gonzalez, R.C. & Woods, R.E., 2008. *Digital Image Processing*, Pearson/Prentice Hall.
- Gray, H., 1918. *Anatomy of the human body* BARTLEBY.COM, ed., New York, NY, USA: Philadelphia: Lea & Febiger.
- Gupta, U. Das, Menon, V. & Babbar, U., 2010. Detecting the Number of Clusters during Expectation-Maximization Clustering Using Information Criterion. In *2010 Second International Conference on Machine Learning and Computing*. IEEE, pp. 169–173.
- Gut, A., 1995. *An Intermediate Course in Probability* 2nd ed. G. Casella, S. Fienberg, & I. Olkin, eds., New York: Springer-Verlag.
- Haralick, R.M., Shanmugam, K. & Dinstein, I., 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, pp.610–617.

- Healey, G. & Binford, T.O., 1988. Local shape from specularity. *Computer Vision, Graphics, and Image Processing*, 42(1), pp.62–86.
- Herrerias, J.M. et al., 2008. Agile patency system eliminates risk of capsule retention in patients with known intestinal strictures who undergo capsule endoscopy. *Gastrointestinal Endoscopy*, 67(6), pp.902–909.
- Hill, T. & Lewicki, P., 2006. *Statistics: Methods and Applications : a Comprehensive Reference for Science, Industry, and Data Mining*, StatSoft.
- Hillcrest, L., 2006. Given Imaging's Agile(TM) Patency System Receives US Marketing Clearance. *Medical News Today*.
- Horn, B.K.P., 1970. *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object From One View*. Massachusetts Institute of Technology.
- Hunt, R.W.G. & Pointer, M.R., 2011. *Measuring Colour*, Chichester, UK: John Wiley & Sons, Ltd.
- Jin, Y., Angelini, E. & Laine, A., 2005. Wavelets in Medical Image Processing: Denoising, Segmentation, and Registration. In D. L. W. J. Suri, ed. *Handbook of Biomedical Image Analysis Vol 1- Segmentation Models - Part a*. Plenum Publishers.
- Johnson, C. et al., 2008. Clinical Role of CE - The Future. *Wireless Capsule Endoscopy*. Available at: http://biomed.brown.edu/Courses/BI108/BI108_2008_Groups/group11/future.html [Accessed May 3, 2013].
- Kanungo, T. et al., 2002. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), pp.881–892.
- Karargyris, A. & Bourbakis, N., 2009. A video-frame based registration using segmentation and graph connectivity for Wireless Capsule Endoscopy. In *2009 IEEE/NIH Life Science Systems and Applications Workshop*. IEEE, pp. 74–79.
- Karargyris, A. & Bourbakis, N., 2011a. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos. *IEEE transactions on biomedical engineering*, 58(10), pp.2777–86.
- Karargyris, A. & Bourbakis, N., 2011b. Three-Dimensional Reconstruction of the Digestive Wall in Capsule Endoscopy Videos Using Elastic Video Interpolation. *IEEE Transactions on Medical Imaging*, 30(4), pp.957–971.
- Karkanis, S. a et al., 2003. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 7(3), pp.141–52.
- Kaufman, A. & Wang, J., 2008. 3D Surface Reconstruction from Endoscopic Videos. *Visualization in Medicine and Life Sciences Mathematics and Visualization*, pp.61–74.

- Kim, H.M. et al., 2010. A Pilot Study of Sequential Capsule Endoscopy Using MiroCam and PillCam SB Devices with Different Transmission Technologies. *Gut and liver*, 4(2), pp.192–200.
- Klepaczko, A. et al., 2010. Local Polynomial Approximation for Unsupervised Segmentation of Endoscopic Images. In L. Bolc et al., eds. *Computer Vision and Graphics - Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 33–40.
- Knyrim, K. et al., 1990. Perspectives in “electronic endoscopy”. Past, present and future of fibers and CCDs in medical endoscopes. *Endoscopy*, 22 Suppl 1, pp.2–8.
- Kodogiannis, V.S. et al., 2007. The usage of soft-computing methodologies in interpreting capsule endoscopy. *Engineering Applications of Artificial Intelligence*, 20(4), pp.539–553.
- Kramme, R., Hoffmann, K.P. & Pozos, R.S., 2011. *Medical Technology*, Springer-Verlag Berlin Heidelberg.
- Lewis, B.S., 2000. Small Intestinal Bleeding. *Gastroenterology Clinics of North America*, 29(1), pp.67–95.
- Lima, C.S. et al., 2008. Classification of endoscopic capsule images by using color wavelet features, higher order statistics and radial basis functions. *Proc. 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2008)*, pp.1242–1245.
- Longmore, M. et al., 2011. *Manual Oxford de Medicina Clínica* 8th ed., Euromédice.
- Longo, D. et al., 2011. *Harrison’s Principles of Internal Medicine, 18th Edition*, McGraw-hill.
- Ma, J. & Plonka, G., 2007. Combined Curvelet Shrinkage and Nonlinear Anisotropic Diffusion. *IEEE Transactions on Image Processing*, 16(9), pp.2198–2206.
- Mackiewicz, M., Berens, J. & Fisher, M., 2008. Wireless Capsule Endoscopy Color Video Segmentation. *IEEE Transactions on Medical Imaging*, 27(12), pp.1769–1781.
- Marin, J.-M., Mengersen, K. & Robert, C.P., 2005. *Bayesian Modelling and Inference on Mixtures of Distributions*,
- Maroulis, D.E. et al., 2003. CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames. *Computer methods and programs in biomedicine*, 70(2), pp.151–66. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12507791>.
- Martins, M.M. et al., 2010. Small bowel tumors detection in capsule endoscopy by Gaussian modeling of color curvelet covariance coefficients. *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2010, pp.5557–60.
- Maybeck, P.S., 1982. *Stochastic Models, Estimation, and Control*, Elsevier Science.

Bibliografia

- McLaren, K., 2008. XIII-The Development of the CIE 1976 ($L^* a^* b^*$) Uniform Colour Space and Colour-difference Formula. *Journal of the Society of Dyers and Colourists*, 92(9), pp.338–341.
- Mohri, M., Rostamizadeh, A. & Talwalkar, A., 2012. *Foundations of Machine Learning*, MIT Press.
- Moon, T.K., 1996. The Expectation-Maximization Algorithm. *IEEE Signal Processing Magazine*.
- Nakamura, T. & Terano, A., 2008. Capsule endoscopy: past, present, and future. *Journal of gastroenterology*, 43(2), pp.93–9.
- Natalin, R.A. & Landman, J., 2009. Where next for the endoscope? *Nature reviews. Urology*, 6(11), pp.622–8.
- Nayar, S.K., Ikeuchi, K. & Kanade, T., 1989. Determining shape and reflectance of Lambertian, specular, and hybrid surfaces using extended sources. In *International Workshop on Industrial Applications of Machine Intelligence and Vision*. Tokyo: IEEE, pp. 169–175.
- Oliveira, S.M.F.L., 2008. *Técnicas Geométricas de Condensação para o Classificador k-NN*. Universidade de Aveiro.
- Olympus America, 2011. *Evis Exera II Gastrointestinal Videoscope - GIF-2TH-180*, Tóquio, Japão.
- Oppenheim, A. V, Willsky, A.S. & Young, I.T., 1983. *Signals and systems*, Prentice-Hall.
- Pascale, D., 2003. *A Review of RGB Color Spaces ...from xyYto R'G'B*, Quebec.
- Perman, S., 2011. Camera-in-a-pill gives a closer look. *Israel21c*.
- Phong, B.T., 1975. Illumination for computer generated pictures. *Communications of the ACM*, 18(6), pp.311–317.
- Prakash, C. & Zuckerman, G.R., 2003. Acute Small Bowel Bleeding: A Distinct Entity with Significantly Different Economic Implications Compared with {GI} Bleeding from Other Locations. *Gastrointestinal Endoscopy*, 58(3), pp.330–335.
- Prasath, V.B.S. et al., 2012. Mucosal Region Detection and 3D Reconstruction in Wireless Capsule Endoscopy Videos Using Active Contours. In *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*. pp. 4014–7.
- Qureshi, W.A., 2004. Current and future applications of the capsule camera. *Nature reviews. Drug discovery*, 3(5), pp.447–50.
- Reinhard, E. et al., 2001. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(4), pp.34–41.

- Rojas, R., 2005. The Kalman Filter. , pp.1–7.
- Saad, Y., 2003. *Iterative Methods for Sparse Linear Systems* 2nd ed., Society for Industrial and Applied Mathematics.
- Shah, J., 2002. Endoscopy through the ages. *BJU International*, 89(7), pp.645–652.
- Soares, J.V.B., 2006. *Segmentação de vasos sanguíneos em imagens de retina usando wavelets e classificadores estatísticos*. Universidade de São Paulo.
- Srinivasan, G.N. & Shobha, G., 2008. Statistical Texture Analysis. *Proceedings of World Academy of Science, Engineering and Technology*, 36, pp.1264–1269.
- Steele, R.J. & Raftery, A.E., 2009. *Performance of Bayesian Model Selection Criteria for Gaussian Mixture Models*, Washington.
- Suetens, P., 2009. *Fundamentals of Medical Imaging* 2nd ed., Cambridge: Cambridge University Press.
- Sun, B. et al., 2010. 3D reconstruction based on Capsule Endoscopy image sequences. In *2010 International Conference on Audio, Language and Image Processing*. IEEE, pp. 607–612.
- The Gale Group Inc., 2008. Intestinal Polyps. *Gale Encyclopedia of Medicine*. Available at: <http://medical-dictionary.thefreedictionary.com/Intestinal+Polyps> [Accessed April 23, 2013].
- Thomson, A.B.R. et al., 2001. Small bowel review: normal physiology part 2. *Digestive diseases and sciences*, 46(12), pp.2588–607.
- Thormaehlen, T., Broszio, H. & Meier, P.N., 2002. Three-Dimensional Endoscopy. In *Falk Symposium (124th : 2001 : Hannover, Germany)*. Kluwer Academic.
- Tompsett, M.F. et al., 1971. Charge-coupled imaging devices: Experimental results. *IEEE Transactions on Electron Devices*, 18(11), pp.992–996.
- Torrance, K.E. & Sparrow, E.M., 1967. Theory for Off-Specular Reflection From Roughened Surfaces. *Journal of the Optical Society of America*, 57(9), p.1105.
- Treves, F., 1785. *Surgical Applied Anatomy*, Lea Brothers & Company.
- Tsai, P.-S. & Shah, M., 1994. Shape from shading using linear approximation. *Image and Vision Computing*, 12(8), pp.487–498.
- Tuceryan, M. & Jain, A.K., 1998. Texture Analysis. In C. H. Chen, L. F. Pau, & P. S. P. Wang, eds. *The Handbook of Pattern Recognition and Computer Vision*. World Scientific Publishing Co., pp. 207–248.

- Umeoka, S. et al., Pictorial review of tuberous sclerosis in various organs. *Radiographics : a review publication of the Radiological Society of North America, Inc*, 28(7), p.e32.
- Vieira, P. et al., 2013. Automatic Detection of Small Bowel Tumors in Endoscopic Capsule Images by ROI Selection based on Discarded Lightness Information. Submetido a: *IEEE Transactions in Medical Imaging*.
- Vieira, P. et al., 2012. Segmentation of small bowel tumor tissue in capsule endoscopy images by using the MAP algorithm. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2012, pp.4010–3.
- Vilardell, F., 2005. *Digestive endoscopy in the second millenium: from the lichtleiter to echoendoscopy*, Estugarda, Alemanha: Aula Medica Ediciones.
- Wang, H. et al., 2008. Anatomical reconstruction from endoscopic images: toward quantitative endoscopy. *American journal of rhinology*, 22(1), pp.47–51.
- Weragama, U., 2011. Small bowel bleeding: Causes, diagnosis and treatment. *The Sunday Times*.
- Withey, D.J. & Koles, Z.J., 2007. Medical Image Segmentation: Methods and Software. In *2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging*. IEEE, pp. 140–143.
- Wong, S.K.M. & Yao, Y.Y., 1987. A statistical similarity measure. In *Proceedings of the 10th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, pp. 3–12.
- Woodham, R.J., 1980. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1), pp.139–144.
- Woodland, A. & Labrosse, F., 2005. On the separation of luminance from colour in images. In *In Proceedings of the International Conference on Vision, Video and Graphics, Edinburgh, UK*. pp. 29–36.
- Wu, C., 2010. *3D Reconstruction of Anatomical Structures from Endoscopic Images*.
- Wu, C., Narasimhan, S.G. & Jaramaz, B., 2009. A Multi-Image Shape-from-Shading Framework for Near-Lighting Perspective Endoscopes. *International Journal of Computer Vision*, 86(2-3), pp.211–228.
- Wu, C.-H. et al., 2004. Automatic extraction and visualization of human inner structures from endoscopic image sequences. *Proc. SPIE*, 5369, pp.464–473.
- Wyszecki, G. & Stiles, W.S., 2000. *Color Science: Concepts and Methods, Quantitative Data and Formulae*, John Wiley & Sons.

- Yates, R.D. & Goodman, D.J., 2005. *Probability and stochastic processes: a friendly introduction for electrical and computer engineers*, John Wiley & Sons.
- YOQNEAM, 2003. Given Imaging Announces Strong Presence at the United European Gastroenterology Week.
- Yu, M., 2001. M2A capsule endoscopy. A breakthrough diagnostic tool for small intestine imaging. *Gastroenterology nursing: the official journal of the Society of Gastroenterology Nurses and Associates*, 25(1), pp.24–7.
- Zajackowski, T. & Zamann, A.P., 2004. Julius Bruck (1840-1902) and his influence on the endoscopy of today. *World journal of urology*, 22(4), pp.293–303.
- Zhang, R. et al., 1999. Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8), pp.690–706.
- Zheng, Y. et al., 2012. Detection of lesions during capsule endoscopy: physician performance is disappointing. *The American journal of gastroenterology*, 107(4), pp.554–60.
- Zhou, J. et al., 2008. Circular generalized cylinder fitting for 3D reconstruction in endoscopic imaging based on MRF. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, pp. 1–8.