

Eva Raquel Belém Curto

Localização de cápsulas endoscópicas multi-câmara

Dissertação submetida para obtenção do grau de Mestre em Engenharia Eletrotécnica e de Computadores
Área de Especialização em Automação
Fevereiro de 2018



UNIVERSIDADE DE COIMBRA



FCTUC FACULDADE DE CIÊNCIAS
E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Localização de cápsulas endoscópicas multi-câmara

Eva Raquel Belém Curto

Coimbra, Fevereiro 2018



Localização de cápsulas endoscópicas multi-câmara

Orientador:

Professor Doutor Hélder de Jesus Araújo

Júri:

Professor Doutor Paulo José Monteiro Peixoto

Professor Doutor João Pedro de Almeida Barreto

Professor Doutor Hélder de Jesus Araújo

Dissertação apresentada à Faculdade de Ciências e Tecnologia da Universidade de Coimbra como requisito parcial para obtenção do grau de Mestre em Engenharia Electrotécnica e de Computadores, na especialidade de Automação.

Coimbra, Fevereiro 2018

AGRADECIMENTOS

Em primeiro lugar gostaria de agradecer ao meu orientador Prof. Doutor Hélder de Jesus Araújo por ter aceite o meu pedido de orientação, por toda a transmissão de conhecimentos que me permitiu aprender bastante durante este processo, mas acima de tudo, por se mostrar sempre disponível em ajudar. A fomentação pelo espírito crítico e o encorajamento em procurar saber sempre mais foram também uma constante ao longo destes meses, estando-lhe grata por isso também.

Agradeço também ao Prof. Doutor Luís Perdigoto pela ajuda que prestou durante o processo de calibração das câmaras da cápsula. Ao Prof. Doutor Pedro Narra Figueiredo deixo também o meu agradecimento por nos acolher na unidade de gastroenterologia dos HUC possibilitando-nos a realização da captura de imagens pela cápsula.

A todos os amigos que a universidade me deu a conhecer, o meu muito obrigada, sem vocês os "dramas" da faculdade seriam mais difíceis de ultrapassar e as pausas para café não teriam a mesma piada. Quero agradecer em especial ao Pedro, à Bruna e ao meu padrinho de curso João por terem lido algumas partes da tese e terem ajudado com correções e sugestões.

Ao meu namorado Hélder, obrigada por todo o apoio incondicional. Ao longo deste percurso, quando as coisas corriam menos bem lá estavas tu para me amparar e fazer ver que com esforço tudo se consegue.

Por último, um agradecimento especial à minha família. Agradeço ao meu pai e à minha mãe, António Curto e Céu Belém, por toda a paciência e carinho. Obrigada por acreditarem sempre em mim e me ajudarem sempre no que podem. Aos meus irmãos, David e Diogo, obrigada por tolerarem o mau humor que a "mana chata" tem por vezes.

A todos, muito obrigada,
Eva.

RESUMO

Doenças do intestino delgado como a hemorragia Gastrointestinal obscura, a doença de Crohn, a doença celíaca e tumores são difíceis de diagnosticar e tratar uma vez que as técnicas de endoscopia tradicionais provocam imensa dor e desconforto e não é possível visualizar toda a extensão do intestino delgado recorrendo a endoscópios com fios. Com a introdução no mercado da Cápsula Endoscópica Sem Fios começa a existir nos sistemas de saúde uma alternativa mais cómoda e eficaz de analisar todo o canal GI. Tem havido várias tentativas por parte de empresas e investigadores de implementar sistemas que localizem as cápsulas uma vez que ainda não existe nenhuma solução suficientemente rigorosa.

Esta dissertação tem como finalidade a localização de uma cápsula endoscópica, mais precisamente da PillCam Colon 2, uma cápsula que possui duas câmaras dispostas nas extremidades. Neste âmbito, procedeu-se à calibração das câmaras da cápsula, obtendo-se os seus parâmetros intrínsecos. Foram implementados em MatLab dois algoritmos distintos destinados a resolver problemas de estimação de movimento 6D que envolvam campos visuais não sobrepostos. Num dos algoritmos, o movimento é estimado individualmente a menos de um fator de escala para cada câmara e só depois é considerada a ligação rígida entre as câmaras para a estimação da escala da translação. Já no segundo algoritmo, para se estimar o movimento, o sistema de câmaras é considerado como um todo através do conceito de câmara generalizada. A propósito ainda deste último algoritmo são estudados em detalhe o tipo e o número de correspondências que se pode usar no caso em estudo.

Os métodos implementados foram testados num ambiente simulado tendo sido também adicionado ruído de modo a testar a robustez destes para a aplicação em imagens reais.

Os métodos descritos na dissertação envolvem um conhecimento alargado sobre visão por computador, nomeadamente transformações rígidas, geometria epipolar, estimação da matriz essencial, câmaras generalizadas, coordenadas de *Plücker*, entre outros.

Palavras Chave

Cápsula endoscópica, localização de cápsulas, calibração de câmaras, distorção radial, distorção tangencial, parâmetros intrínsecos, odometria visual, campos visuais não sobrepostos, estimação de movimento, fator de escala, matriz essencial, sistema de múltiplas câmaras, câmara generalizada, restrição epipolar generalizada, matriz essencial generalizada, correspondências cruzadas.

ABSTRACT

Diseases of the small intestine such as obscure gastrointestinal haemorrhage, Crohn's disease, Celiac disease and tumors are difficult to diagnose and treat since endoscopic techniques cause immense pain and discomfort, and it is not possible to visualize the full extent of the small intestine using endoscopes with wires. With the introduction in market of the Wireless Capsule Endoscope begins to exist in health systems a more comfortable and effective alternative to analyze the entire GI channel. There have been several attempts by companies and researchers to implement systems that locate the capsules since there is still no solution sufficiently rigorous.

This dissertation aims to locate an capsule endoscope, more precisely of PillCam Colon 2, a capsule that has two cameras arranged at the ends. In this context, calibration of the capsule cameras was carried out, obtaining its intrinsic parameters. Were been implemented two different algorithms to solve 6D motion estimation involving non-overlapping visual fields. In one of the algorithms, movement is estimated individually up to scale for each camera and the rigid connection between the cameras only is considered for the estimation of the translation scale. In the second algorithm, to estimate the movement, the camera system is considered as a whole through the concept of generalized camera. With regard to this last algorithm are studied in detail the type and number of correspondences that can be used in the case under study.

The implemented methods were tested in a simulated environment having also been added noise in order to test their robustness to apply in real images.

The methods described in the dissertation involve a broad knowledge of computer vision, in particular concepts of: rigid transformations, epipolar geometry, estimation of essential matrix, generalized cameras, Plücker coordinates, and others.

Keywords

Capsule endoscope, capsule localization, camera calibration, radial distortion, tangential distortion, intrinsic parameters, visual odometry, non-overlapping fields of view, motion estimation, scale factor, essential matrix, multi-camera systems, generalized camera, generalized epipolar constraint, generalized essential matrix, cross correspondences.

"Computer science is no more about computers than astronomy is about telescopes, biology is about microscopes or chemistry is about beakers and test tubes. Science is not about tools, it is about how we use them and what we find out when we do. "

— Edsger Dijkstra

CONTEÚDO

Agradecimentos	iii
Resumo	iv
Abstract	vi
Lista de Acrónimos	xiii
Lista de Figuras	xv
Lista de Tabelas	xviii
1 Introdução	1
1.1 Considerações gerais e enquadramento	2
1.2 Objetivos e contribuições	4
1.3 Estrutura da dissertação	4
2 Considerações sobre a cápsula endoscópica	5
2.1 Tipos de Cápsulas	6
2.1.1 PillCam Colon 2	6
2.2 Métodos para localização de cápsulas	8
2.3 Calibração das câmaras	12

3	Odometria visual sem sobreposição de campos visuais	22
3.1	Introdução à Odometria Visual	23
3.1.1	Odometria Visual sem sobreposição de campos visuais	25
3.2	Estimação do movimento usando odometria monocular	26
3.2.1	Estimação do movimento individual de cada câmara a menos de um fator de escala	27
3.2.2	Estimação da escala	29
3.3	Estimação do movimento usando o modelo de câmara generalizada	32
3.3.1	Modelo de Câmara Generalizada	32
3.3.2	Estimação do movimento	33
3.3.3	Influência do tipo e do número de correspondências no <i>rank</i> da matriz A . . .	37
4	Resultados Experimentais	45
4.1	Simulador	46
4.2	Resultados	46
5	Conclusões e Trabalho Futuro	56
5.1	Conclusões	57
5.2	Trabalho futuro	58
6	Bibliografia	59

LISTA DE ACRÓNIMOS

ASIFT	Affine scale-invariant feature transform
CCD	Charge-coupled device
CESF	Cápsula Endoscópica Sem Fios
CMOS	Complementary metal oxide semiconductor
DVS	Decomposição em valores singulares
FDA	Food and Drug Administration
fps	<i>Frames</i> por segundo
GI	Gastrointestinal
IRM	Imagem por ressonância magnética
Laser	Light Amplification by Stimulated Emission of Radiation
LED	Light Emitting Diode
MCG	Modelo de Câmara Generalizada
MMMQ	Método das Médias dos Mínimos Quadrados
MSAC	M-estimator sample consensus

OV	Odometria Visual
OVM	Odometria Visual Monocular
PEP	Polynomial eigenvalue problems
RANSAC	Random Sample Consensus
REG	Restrição Epipolar Generalizada
RF	Radiofrequência
SIFT	Scale-invariant feature transform
SQUID	Superconducting Quantum Interference Device
SURF	Speeded-up robust features
SVD	Singular value decomposition

LISTA DE FIGURAS

1.1	Endoscopia convencional vs. PillCam	2
2.1	Cápsulas endoscópicas comerciais mais representativas no mercado.	6
2.2	Ilustração do sistema da <i>PillCam COLON2</i>	7
2.3	Esquema geral da constituição da cápsula	7
2.4	Modelo geométrico da câmara <i>pinhole</i>	12
2.5	Distorção radial	13
2.6	Distorção tangencial	13
2.7	Efeitos da distorção	14
2.8	Imagem do cinto com gravador e monitor onde se visualizam as imagens capturadas por uma das câmaras.	15
2.9	Imagem da montagem experimental da calibração com recurso a espelhos	16
2.10	Imagem da montagem experimental da calibração com a câmara debaixo de água	16
2.11	Mapas de distorção correspondentes à calibração da câmara 1 com o <i>dataset 1</i>	19
2.12	Mapas de distorção correspondentes à calibração da câmara 1 com o <i>dataset 2</i>	19
2.13	Mapas de distorção correspondentes à calibração da câmara 1 com o <i>dataset 3</i>	19
2.14	Mapas de distorção correspondentes à calibração da câmara 1 com o <i>dataset 4</i>	20
2.15	Mapas de distorção correspondentes à calibração da câmara 2 com o <i>dataset 1</i>	20
2.16	Mapas de distorção correspondentes à calibração da câmara 2 com o <i>dataset 2</i>	20

2.17	Mapas de distorção correspondentes à calibração da câmara 2 com o <i>dataset 3</i>	21
2.18	Imagens capturadas pela cápsula endoscópica.	21
3.1	Esquema geral do algoritmo <i>Real Time 6D</i>	26
3.2	Representação das transformações euclidianas do sistema de câmaras durante o <i>frame</i> 1 e o <i>frame 2</i>	30
3.3	Configurações degeneradas do algoritmo dos 17 pontos.	34
3.4	Esquema correspondências cruzadas.	37
3.5	Esquema representativo do sistema de câmaras <i>stereo</i> sem sobreposição de campos visuais com respectivos centros de projeção das câmaras	39
4.1	Sistema de câmaras rigidamente acoplado usado nas simulações.	46
4.2	Na Figura (a), está representada a estimativa do movimento para o algoritmo 3.2 para uma trajetória 2D na ausência de ruído. Como se pode verificar a estimativa está exatamente igual ao <i>ground-truth</i> . A Figura (b) apresenta o movimento estimado para a mesma trajetória segundo o algoritmo 3.3 também sem ruído. mais uma vez a trajetória estimada coincide com o movimento <i>ground-truth</i>	48
4.3	Na Figura (a) está representada a estimativa do movimento realizada pelo algoritmo 3.2 para uma trajetória 3D na ausência de ruído. A estimação do movimento através do algoritmo 3.3 encontra-se na Figura (b) onde mais uma vez ainda não foi inserido qualquer ruído. É possível observar que as estimações feitas pelos dois algoritmos coincidem com a trajetória <i>ground-truth</i>	49
4.4	Na figura (a) está representada a estimativa do movimento realizada pelo algoritmo 3.2 para a trajetória 2D. Em cada imagem gerada na trajetória foi inserido um ruído de $\sigma = 0.005$. Na figura (b) encontra-se a estimação do movimento através de 3.3, tendo também sido um ruído de $\sigma = 0.005$. É bastante visível o efeito do ruído na estimação do movimento presente nos dois algoritmos. Mas enquanto que em (a) parece existir um offset de erro, em (b) a estimação do movimento aparece mais sob a forma de <i>outliers</i>	50

4.5	Na figura (a) está representada a estimativa do movimento realizada pelo algoritmo 3.2 para a trajetória 3D. Em cada imagem gerada na trajetória foi inserido um ruído de $\sigma = 0.005$. Na figura (b) encontra-se a estimação do movimento através de 3.3, tendo também sido sujeito a um ruído de $\sigma = 0.005$. É bastante visível o efeito do ruído na estimação do movimento presente nos dois algoritmos. Principalmente em (a) existe um grande erro na estimação do movimento. Já no caso de (b), à medida que aumentam os níveis de ruído surgem mais <i>outliers</i>	51
4.6	Histogramas do erro na estimação da escala baseados em 100 testes de um movimento genérico. Em todos os testes foi introduzido um ruído com $\sigma=0.005$ nos pixels. Em (a) estão os resultados para o algoritmo <i>Real Time 6D</i> e em (b) para o algoritmo dos 14 pontos.	53
4.7	Histogramas do erro na direção da translação baseados em 100 testes de um movimento genérico. Em todos os testes foi introduzido um ruído com $\sigma=0.005$ nos pixels. Em (a) estão os resultados para o algoritmo <i>Real Time 6D</i> e em (b) para o algoritmo dos 14 pontos.	53
4.8	Histogramas do erro na rotação baseados em 100 testes de um movimento genérico. Em todos os testes foi introduzido um ruído com $\sigma=0.005$ nos pixels. Em (a) estão os resultados para o algoritmo <i>Real Time 6D</i> e em (b) para o algoritmo dos 14 pontos.	54
4.9	Gráficos onde estão representadas as curvas dos erros em função do nível de ruído. Em (a) tem-se a curva de erro na escala. Em (b), o erro na direção da translação. Por último, em (c) a curva do erro na rotação.	55

LISTA DE TABELAS

2.1	Algumas das especificações técnicas da <i>PillCam[®] Colon2</i> , [36], [33]	8
2.2	Características dos <i>datasets</i> usados para calibração da câmara 1.	16
2.3	Características dos <i>datasets</i> usados para calibração da câmara 2.	16
2.4	Parâmetros intrínsecos para a câmara 1.	17
2.5	Parâmetros intrínsecos para a câmara 2.	17
2.6	Média e variância dos parâmetros das câmaras sem estarem submersas.	18
2.7	Média e variância dos parâmetros da câmara 1 submersa.	18
3.1	<i>Rank</i> da matriz A conforme os números de correspondências em cada câmara.	44
4.1	Médias dos erros para um movimento genérico testado 100 vezes usando o algoritmo <i>Real Time 6D</i> e o algoritmo dos 14 pontos.	52

CAPÍTULO

1

INTRODUÇÃO

Conteúdos

1.1	Considerações gerais e enquadramento	2
1.2	Objetivos e contribuições	4
1.3	Estrutura da dissertação	4

1.1 Considerações gerais e enquadramento

O intestino delgado representa a parte mais longa do sistema digestivo, representando 75% do seu comprimento (cerca de 6 metros de comprimento e 4 vezes mais que o intestino grosso) e 90% da área de superfície de absorção do sistema gastrointestinal [1]. A hemorragia Gastrointestinal (GI) obscura, a doença de Crohn, a doença celíaca, pequenos tumores e outros distúrbios que ocorrem no intestino delgado são difíceis de diagnosticar e tratar devido à dificuldade em aceder a um ambiente tão complexo dentro do corpo humano.

A visualização direta do canal GI pela endoscopia tradicional foi introduzida e desenvolvida ao longo dos últimos 50 anos. Em 1960 quando a fibra ótica foi descoberta, os endoscópios tornaram-se uma ferramenta crucial para diagnosticar doenças GI. As técnicas convencionais para examinar o canal GI usam um tubo longo e flexível com luz e uma câmara miniatura na extremidade. Este equipamento pode ser inserido pela boca ou pelo ânus. Devido à sua rigidez e à largura do diâmetro comparativamente à largura do intestino delgado, este procedimento provoca dor e desconforto ao paciente, sobretudo quando o médico não possui ainda prática. Além disso, os endoscópios ligados a fios não permitem visualizar grande parte do intestino delgado [2].

Com o intuito de desenvolver uma técnica de endoscopia mais eficiente e menos dolorosa para os pacientes, foi inventada a Cápsula Endoscópica Sem Fios (CESF) que consiste num dispositivo não reutilizável em forma de comprimido e que à sua semelhança é também ingerido. A CESF é munida de câmaras muito pequenas, *LEDs* e um módulo de transmissão, conseguindo percorrer todo o tubo digestivo aproveitando a propulsão resultante de movimentos peristálticos normais. Ao longo do percurso são gravadas inúmeras imagens que são transmitidas a recetores externos e que, posteriormente, são analisadas por um médico especialista.

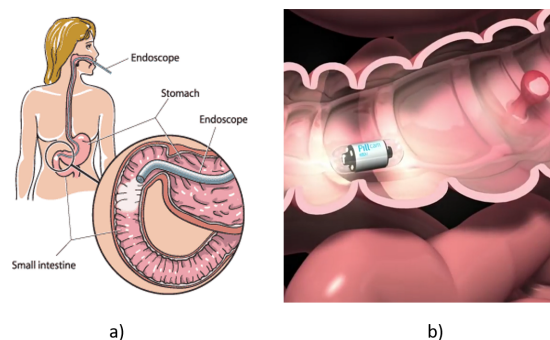


Figura 1.1: a) Endoscopia convencional, [3]. b) Cápsula Endoscópica Sem Fios PillCam, imagem retirada do vídeo [4]

Desde a introdução da primeira CESF em 2000 pela *Given Imaging Ltd*, este método revolucio-

nário tem-se revelado um tópico de investigação importante tanto no ramo da engenharia como no da saúde/medicina [2]. São muitas as inovações a ele ligadas tais como: tentativas de localização suficientemente exatas (abordadas em 2.2); locomoção ativa (através da incorporação de atuadores miniatura na cápsula [5], [6], [7] ou recorrendo a sistemas de atuação interna, nomeadamente campos magnéticos [8], [9], [10]); maior eficiência energética [11], [12]; melhoria do sistema de telemetria sem fios [13], [14], [15]; melhoria na qualidade das imagens endoscópicas [16], [17], [18]; capacidades de diagnóstico [19], [20], [21], [22]; e possibilidade de tratamento [23], [24], [25].

A localização exata da cápsula constitui um desafio muito pertinente uma vez que é necessária para casos como tratamento cirúrgico de lesões suspeitas e supervisão do progresso de doenças, tais como a doença de Crohn [26]. Tipicamente, a localização da cápsula endoscópica dentro do organismo tem sido realizada recorrendo a sensores externos, sendo que os sistemas comercialmente disponíveis providenciam apenas uma estimativa aproximada baseada num conjunto de sensores Radiofrequência (RF). Nestes sistemas a localização da cápsula é indicada de modo pouco rigoroso por uma representação gráfica sem unidade fornecida pela interface do usuário do software de leitura de vídeo [26].

Ao longo dos anos outros métodos de localização (para além daqueles que utilizam RF) têm vindo a ser estudados, nomeadamente os que recorrem a campos magnéticos, sensores de efeito de *Hall*, raios-X e ressonâncias magnéticas [27]. Estes métodos requerem sempre o uso de dispositivos adicionais e normalmente implicam que o paciente tenha que permanecer várias horas numa unidade hospital com mobilidade reduzida. No entanto, existem também métodos de localização a serem desenvolvidos que se baseiam em visão por computador fazendo uso apenas das imagens gravadas [28], [29], [30]. Nestes casos, como a estimação da localização só é feita depois da recolha da gravação o paciente não sofre quaisquer restrições de mobilidade.

No subcapítulo 2.2 serão descritos com mais detalhe alguns dos métodos de localização existentes.

Atendendo à importância de obter uma estimativa robusta e exata da localização da CESF, a presente dissertação propõe-se resolver a problemática da localização de uma cápsula endoscópica específica, a PillCam Colon 2, através de visão por computador, recorrendo a algoritmos de odometria visual, neste caso sem sobreposição de campos visuais.

1.2 **Objetivos e contribuições**

O objetivo principal desta tese incide em implementar e testar em MATLAB algoritmos já existentes que permitam a estimação da localização de uma cápsula endoscópica com duas câmaras opostas. Em maior detalhe, as principais contribuições desta tese são:

- Calibração das câmaras da cápsula PillCam Colon 2 dentro e fora de água;
- Análise da robustez de dois métodos de estimação de posição relativa, usando campos visuais não sobrepostos em ambientes controlados e análise do impacto da inserção de ruído na estimação da posição;
- Análise ao número e ao tipo de correspondências necessárias no algoritmo da câmara generalizada para um sistema de câmaras *stereo* sem sobreposição de campos visuais.

1.3 **Estrutura da dissertação**

A presente dissertação está organizada em cinco capítulos, começando pela introdução onde é contextualizada a problemática da localização de cápsulas endoscópicas e onde são propostos os objetivos a alcançar neste trabalho.

No segundo capítulo são apresentadas as várias cápsulas existentes no mercado, sendo estudada com mais detalhe a PillCam Colon 2. São também referenciados alguns dos trabalhos já desenvolvidos no âmbito da localização de cápsulas. Ainda neste capítulo, é explicado como foi feita a calibração das câmaras da cápsula, sendo mostrados os respectivos resultados.

No capítulo 3 são explicados detalhadamente os dois métodos usados para localização de câmaras sem sobreposição de campos visuais.

No capítulo 4 podem-se encontrar os resultados experimentais dos testes realizados.

Por último, no quinto capítulo é feita uma reflexão sobre o trabalho realizado e são propostas melhorias futuras.

CAPÍTULO

2

CONSIDERAÇÕES SOBRE A CÁPSULA
ENDOSCÓPICA

Conteúdos

2.1	Tipos de Cápsulas	6
2.1.1	PillCam Colon 2	6
2.2	Métodos para localização de cápsulas	8
2.3	Calibração das câmaras	12

2.1 Tipos de Cápsulas

Como foi referido anteriormente em 1.1, o modelo inicial de uma cápsula endoscópica foi desenvolvido pela empresa *Given Imaging Ltd* (Yoqneam, Israel). Esta cápsula, denominada *M2A* (que significa "mouth to anus") foi aprovada para uso clínico geral na Europa em Maio de 2001 e nos Estados Unidos pela *Food and Drug Administration (FDA)* em Agosto do mesmo ano. Inicialmente a FDA aprovou a *M2A* como ferramenta auxiliar na visualização de anomalias no intestino delgado. Só mais tarde, em Julho de 2003 foi reconhecida como ferramenta principal na deteção de anormalidades do intestino delgado. Entretanto a *M2A* foi renomeada *PillCam SB* (que significa "small bowel") e a família *PillCam*[®] foi crescendo, existindo agora também a *PillCam*[®] *Colon2*, equipada com duas câmaras; a *PillCam*[®] *UGI (Upper GI)*; e a *PillCam*[®] *PATENCY* que consiste numa cápsula dissolúvel para desobstruir o sistema GI antes da CESF. Os produtos *PillCam*[®] são agora comercializados pela *Medtronic, Inc.* [31], [27].

Outras empresas foram também desenvolvendo e produzindo CESF, existindo atualmente a *Olympus, Co.* (Tokyo, Japan) com a *EndoCapsule*, a *IntroMedic, Co., Ltd.* (Seoul, South Korea) com a *MiroCam*, o grupo chinês *Chongqing Jinshan Science & Technology, Co., Ltd.* com a cápsula *OMOM* e a *CapsoVision, Inc.* (Saratoga, CA, USA) com a *CapsoCam*, que possui uma imagem HD panorâmica com 360°. É possível observar na Figura 2.1 algumas das cápsulas mais comercializadas.

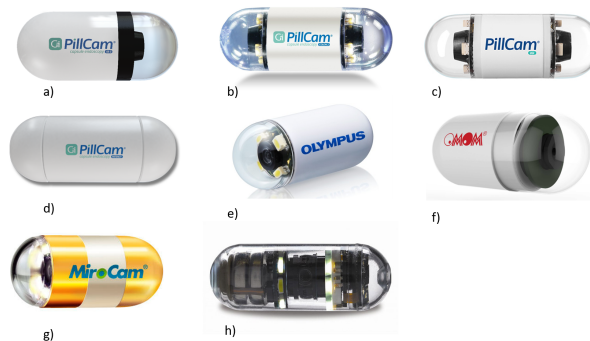


Figura 2.1: Cápsulas endoscópicas comerciais mais representativas no mercado. **a)** *PillCam*[®] *SB3*; **b)** *PillCam*[®] *Colon2*; **c)** *PillCam*[®] *UGI*; **d)** *PillCam*[®] *PATENCY*; **e)** *EndoCapsule*; **f)** *OMOM*; **g)** *MiroCam*; **e h)** *CapsoCam*

2.1.1 PillCam Colon 2

Seguidamente irá ser estudada a cápsula *PillCam*[®] *Colon2* mais detalhadamente, uma vez que foi a utilizada nesta dissertação.

A *PillCam[®] Colon2*, que está representada na Figura 2.1, corresponde à segunda geração da cápsula *PillCam[®] Colon*. Apesar do seu nome indicar que tem uma utilização específica para o cólon, esta cápsula tem sido utilizada também com sucesso no intestino delgado [32].

O sistema geral da cápsula é composto por quatro componentes que se encontram representadas na Figura 2.2: (1) Cápsula; (2) Gravador; (3) Estação de trabalho; (4) Software.

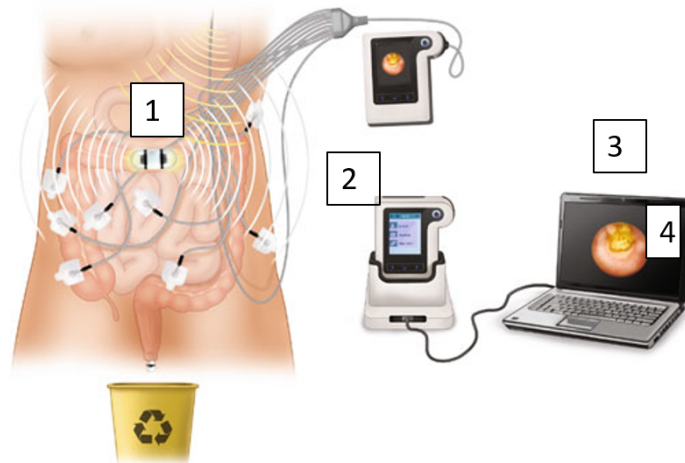


Figura 2.2: Ilustração do sistema da *PillCam COLON2* [33].

Relativamente à constituição da cápsula, estão indicados no esquema da Figura 2.3 os seus componentes.

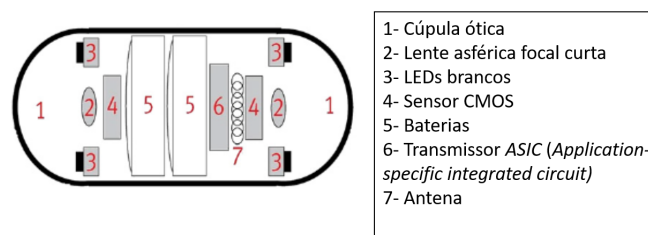


Figura 2.3: Esquema geral da constituição da cápsula. [34].

Na Tabela 2.1 estão descritas algumas das especificações técnicas desta cápsula, das quais se salienta a capacidade adaptativa do *frame rate*, i.e., a capacidade de aumentar ou diminuir consoante a velocidade do movimento, variando entre 4 a 35 fps. Esta adaptação é possível graças a uma técnica que cruza dados entre a cápsula e o gravador de dados (DR3), que é dotado de inteligência artificial [35].

Como se pode ainda observar nesta Tabela, a cápsula possui sensores óticos Complementary metal oxide semiconductor (CMOS), ao contrário de outras cápsulas que usam Charge-coupled

device (CCD). Enquanto os *chips* CCD convertem a luz e a carga de cada píxel em voltagem, os sensores CMOS usam uma matriz de píxeis que requerem a amplificação de cada píxel. Em relação às vantagens destes dois *chips*, o CCD possui uma maior potência com menor ruído ótico e estabilidade consecutiva às mudanças de iluminação. Já o CMOS tem um tamanho mais reduzido e um menor consumo de energia permitindo uma duração de trabalho maior e a possibilidade de câmaras adicionais. Apesar das diferenças, ambos permitem a aquisição de imagens GI de boa qualidade para uso clínico [34].

Propriedades		
Físicas	Dimensões	Comprimento: 31.5 mm Diâmetro: 11.6 mm
	Peso	2.9 g
	Material	Plástico biocompatível
Óticas	Tipo de sensor	CMOS
	Tamanho do sensor	1.4336 mm
	Nº de sensores óticos	2
	Iluminação	4 LEDs em cada lado
	Campo de visão	172°ISO-8600-3
	Resolução de imagem	256x256
	Visibilidade efetiva	Distância: 3 cm
Operacionais	Distância mínima detetável	Pelo menos 0.1 mm
	Tempo de execução	Mínimo de 10 horas
	Tipo de bateria	Baterias de óxido de prata
Comunicação uplink	Temperatura de funcionamento	20-40 °C
	Frequência de operação	434.1 MHz
	Frame rate	4-35 fps
Comunicação downlink	Data rate	2.7 Mbps e 8.1 Mbps
	Frequência de operação	13.6 Mhz
	Largura de banda do recetor	± 150 KHz

Tabela 2.1: Algumas das especificações técnicas da PillCam® Colon2, [36], [33]

No que diz respeito ao dispositivo de gravação, possui um recetor RF externo e uma unidade de gravação, recebendo e armazenando na memórias as imagens transmitidas pela cápsula. Este está ligado a um cinto usado pelo paciente durante todo o exame. O cinto não só suporta o dispositivo de gravação como também é usado para atenuar o sinal de RF longe do paciente.

2.2 Métodos para localização de cápsulas

Como já havia sido mencionado em 1.1, o desenvolvimento de um método de localização eficiente constitui uma importante contribuição na área das CESF, uma vez que permite saber a localização das lesões para que se possa examiná-las mais detalhadamente e até tratá-las. Além

disso, a informação sobre a localização da cápsula permite realimentar o controle respectivo ao seu movimento, como é referido em [37], [38] e [39].

Como consequência destas motivações, vários grupos de investigadores e empresas têm desenvolvido várias estratégias para a deteção da posição e orientação da cápsula ao longo do seu percurso pelo sistema GI. Um dos métodos de localização usado pela *Given Imaging Ltd.* baseava-se em ondas eletromagnéticas, mais especificamente em ondas de rádio. Neste caso, a localização é baseada na força do sinal emitido pela cápsula e recebido nos oito sensores (antenas) posicionados sobre o abdómen [40]. Esta abordagem não foi muito utilizada visto vez que apresentava uma precisão baixa (3.77 cm) [41]. Mais tarde, em 2013, a *Given Imaging Ltd.* patenteou um método de localização baseado numa única bobina como fonte magnética e numa única bobina como sensor magnético [42]. A *Intromedic*, empresa que comercializa a *MiroCam*, patenteou em 2006 um sistema de localização baseado em valores de potencial elétrico [43].

Recentemente, muitas equipas de investigação têm sugerido localização através de campos magnéticos, isto porque o corpo humano possui uma permeabilidade magnética muito próxima da do ar e exerce muito pouca influência no sinal magnético. Deste modo, é possível obter localização com uma elevada precisão [44]. Por outro lado, várias situações podem levar a uma diminuição da precisão tais como, a inserção involuntária de uma ferramenta ferromagnética no espaço de trabalho, a restrição de tamanho do íman pelas dimensões da cápsula ou ainda, caso seja implementado um sistema de locomoção ativa baseado em campos magnéticos, a existência de interferências indesejadas [27].

Weitschies et al. [45] foram os primeiros a equipar uma cápsula com um íman permanente. Através de um *Superconducting Quantum Interference Device (SQUID)* com 27 canais registaram em vários intervalos de tempo a distribuição do campo magnético sobre o abdómen. Em cada instante de tempo, a posição da cápsula dentro do sistema GI foi determinada a partir da distribuição de campo medida, assumindo um modelo de dipolo magnético. A resolução da posição era de aproximadamente alguns milímetros e a resolução temporal na ordem dos milissegundos.

Wu et al. [46] desenvolveram um sistema de localização magnético portátil baseado no efeito de *Hall* e no modelo de dipolo magnético. O sistema consistia num colete onde estão colocados os sensores de efeito de *Hall*, uma cápsula com um pequeno cilindro de neodímio no seu interior e um software. O erro relativo de localização deste sistema era inferior a 10%.

Um método proposto por *Plotkin* et al. [47] consiste em empregar um conjunto grande de bobinas de transmissão coplanares (8x8). As 64 bobinas são ativadas sequencialmente apenas na fase de iniciação para calcular a posição inicial da bobina recetora que se encontra no interior da

cápsula. A redundância no número de transmissores permite uma convergência rápida e inequívoca do algoritmo de otimização. Nas fases seguintes de localização só são ativadas 8 bobinas, o que nos permite manter uma alta taxa de atualização. A precisão registrada é de 1 mm, 0.6°.

Salermo et al. [48] propuseram um método baseado em campos magnéticos que, ao contrário dos anteriormente mencionados, é compatível com a locomoção externa magnética. A abordagem é baseada num algoritmo de triangulação capaz de detetar a cápsula no sistema GI, gravando e processando medições do campo magnético através de um sensor magnético triaxial inserido na cápsula. Foram obtidos erros de posição de 14 mm ao longo do eixo dos X, 11 mm ao longo do eixo dos Y e 19 mm ao longo do eixo dos Z (onde o plano XY corresponde ao plano do abdómen e Z é a direção vertical em relação a este plano).

Existem outras técnicas de localização populares usadas em procedimentos clínicos. Uma delas considera que a cápsula percorre o sistema GI a uma velocidade constante e que a distância linear percorrida é calculada usando o tempo que a cápsula demorou entre pontos de referência. Na realidade, a velocidade da cápsula varia muito devido aos movimentos peristálticos [29]. Os raios-X também podem ser usados para localizar uma cápsula endoscópica no sistema GI. A técnica de cintigrafia gama é usada também para visualizar a posição da cápsula *Enterion*, uma cápsula de entrega de medicação, em tempo real. O sistema IRM foi proposto por *Dumoulin* et al. para localizar dispositivos em tempo real [27].

A fim de encontrar uma maneira barata e eficiente para supervisionar o movimento da cápsula, sem que seja necessário recorrer a sensores externos, começaram a surgir tentativas de extrair parâmetros de movimento com base apenas na sequência de imagens consecutivas capturadas pela CESF.

Em 2009, *Liu* et al. [49] apresentaram um método que usava visão por computador e que tinha como objetivo complementar os já existentes, baseados em campos magnéticos. Através destes últimos apenas é possível obter cinco graus de liberdade (três para a translação e dois para rotação). Uma vez que a cápsula muda de posição e orientação muito lentamente, algumas regiões vão aparecer sucessivamente em várias imagens, sendo possível obter correspondência entre pares de pontos característicos em duas imagens consecutivas. Para localizar as *features* nas imagens endoscópicas, os autores aplicam um algoritmo iterativo de fluxo ótico *Lucas-Kanade*. Em seguida, para obter os parâmetros de posição e orientação, calculam a matriz essencial através do algoritmo dos 8 pontos. Por fim, decompõem a matriz de rotação através do método do quaternião. Os resultados obtidos apresentaram uma boa precisão para um ângulo de rotação não muito grande (inferior a 30°).

Iakovidis et al. [26] apresentaram um método, também baseado em odometria visual, que

permite estimar tanto a rotação como a translação através de pontos de referência anatómicos que podem ser detetados por reconhecimento de padrões. Nesta abordagem, é usado o *Speeded-up robust features (SURF)* como extrator de *features* e é calculada a matriz essencial através do algoritmo dos 5 pontos [50] que recebe como entradas conjuntos de *inliers* seleccionados previamente pelo método RANSAC. Além disso, introduzem uma *framework*, a *Java Video Analysis (JVA)*, para desenvolvimento de aplicações futuras de análise de vídeo.

Uma técnica baseada na estimação da homografia entre *frames* consecutivos foi usada por *Spyrou et al.* Através da matriz de homografia é obtida informação necessária relativa à orientação da imagem endoscópica [51].

G. Bao et al. [28] estimaram o deslocamento, orientação e inclinação de uma cápsula através da distância entre *feature points* correspondentes em imagens consecutivas. Estes autores empregaram *Scale-invariant feature transform (SIFT)* para extrair os pontos característicos. Após observarem que a magnitude dos deslocamentos dos *feature points* dependia da distância relativa destes pontos à câmara, aplicaram uma operação chamada "projeção cilíndrica inversa" para projetar a imagem original num sistema de coordenadas baseado em ângulo e raio. Para validarem o desempenho do algoritmo, criaram um intestino artificial e, com uma câmara no seu interior, simularam o percurso de uma cápsula endoscópica. Mais tarde, a mesma equipa de investigação propõe um método de estimação da velocidade da cápsula endoscópica baseando-se novamente no deslocamento entre *feature points* [29].

Recentemente, *Wahid et al.* [30] propuseram uma técnica para estimar somente o deslocamento em frente e à retaguarda da cápsula. Para isso, testaram e compararam a deteção de *features* com SIFT, SURF e Affine scale-invariant feature transform (ASIFT), aplicando de seguida o método RANSAC para excluir os *outliers*. Os resultados mostraram que o método que usa ASIFT e RANSAC apresenta a melhor estimativa, dado que o erro médio de estimativa de deslocamento é 0.024 cm.

2.3 Calibração das câmaras

A calibração da câmara é o processo que permite estimar os parâmetros intrínsecos da câmara usada para aquisição das imagens, assim como a sua pose que descreve a posição e orientação em relação ao mundo. É necessário este processo para saber propriedades desconhecidas da câmara, tais como a distorção das lentes e obter um mapeamento entre o pixel e coordenadas do mundo real.

As câmaras da PillCam Colon 2 foram calibradas recorrendo à *Camera Calibration Toolbox*. Nesta aplicação é usado o modelo de câmara proposto por *Jean-Yves Bouquet* [52]. Este modelo engloba:

- O modelo de câmara *pinhole*
- O modelo de distorção das lentes

Modelo da câmara *pinhole*

Uma câmara *pinhole* é uma câmara sem lentes e com uma pequena abertura por onde passam os raios de luz. É então projetada uma imagem invertida no lado oposto à abertura. O plano da imagem virtual está em frente à câmara e tem a imagem não invertida da cena. No modelo de câmara *pinhole*, representado na Figura 2.4, é considerada uma projeção central dos pontos do espaço num plano de imagem, i.e, um ponto no espaço com coordenadas $\mathbf{X}=(X, Y, Z)^T$ é mapeado num ponto do plano da imagem onde a linha que junta \mathbf{X} ao centro de projeção interceta o plano da imagem.

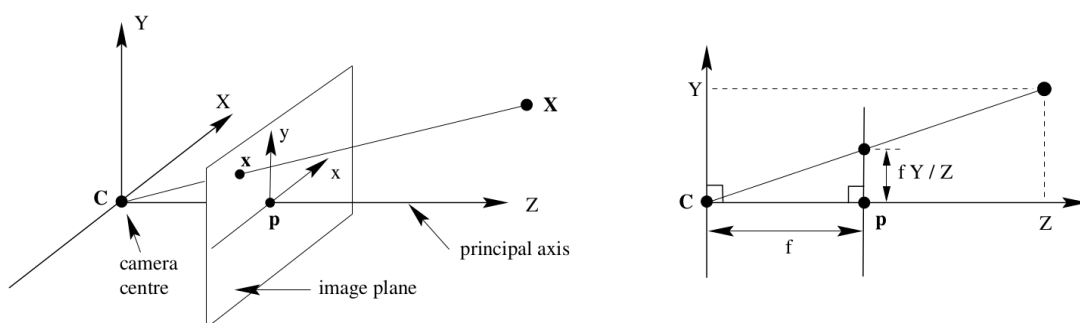


Figura 2.4: Modelo geométrico da câmara *pinhole* [53]

Representando os pontos 3D do mundo e os pontos 2D da imagem em coordenadas homogêneas, a projeção central pode ser descrita pela seguinte equação:

$$\hat{x} = P\hat{X} \quad (2.1)$$

onde $\hat{x} = (x, y, 1)^T$ são as coordenadas em pixels, $\hat{X} = (X, Y, Z, 1)^T$ são as coordenadas correspondentes aos pontos no mundo e P é a matriz da câmara.

A matriz da câmara pode ser escrita da seguinte forma:

$$P = K[R|t] \quad (2.2)$$

A rotação R e a translação t descrevem a orientação e a posição da câmara relativamente ao sistema de coordenadas do mundo. A matriz K contém os parâmetros internos da câmara; a distância focal em pixels (f_x, f_y) , o *skew*, e o ponto principal em pixels (c_x, c_y) .

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.3)$$

Modelo de distorção das lentes

O modelo da câmara *pinhole* não tem em conta a distorção das lentes. No entanto, numa câmara real com lentes é necessário incluir as distorções que lhes são inerentes, sendo estas a distorção radial e a distorção tangencial.

A distorção radial ocorre quando os raios de luz curvam mais perto das bordas da lente do que no seu centro ótico. Quanto menor a lente, maior a distorção. A distorção tangencial ocorre quando a lente e o plano de imagem não são paralelos. Nas Figuras 2.5 e 2.6 estão representadas a distorção radial e tangencial.

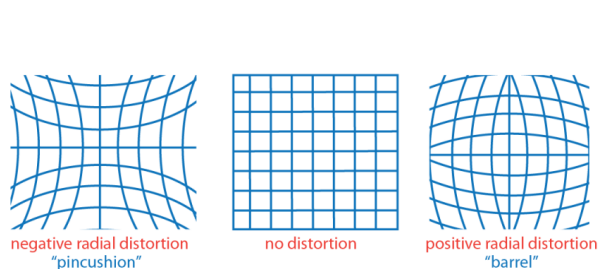


Figura 2.5: Distorção radial, [54].

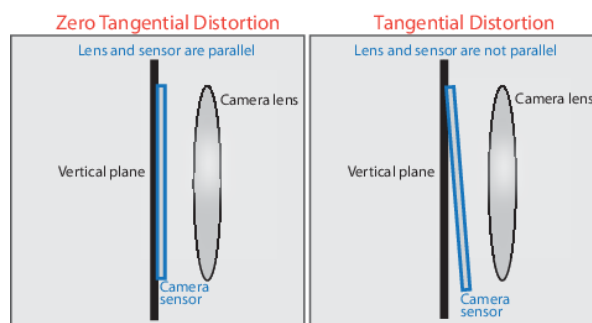


Figura 2.6: Distorção tangencial, [54].

O modelo de distorção das lentes usado é um modelo polinomial standard, que foi introduzido por *Brown* em [55]. A forma como o modelo foi implementado na *Camera Calibration Toolbox* é

descrita em detalhe em [52].

Verificou-se que a junção de coeficientes de grau superior a 6 não era vantajosa uma vez que estes tinham coeficientes muito próximos de zero. Deste modo, no modelo utilizado as distorções radial e tangencial foram representadas por um vetor de 5 elementos k_c . As coordenadas dos pontos sem distorção x_n e com distorção x_d relacionam-se da seguinte forma:

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = \underbrace{(1 + k_{c1}r^2 + k_{c2}r^4 + k_{c3}r^6)}_{\text{distorção radial}} \begin{bmatrix} x_n \\ y_n \end{bmatrix} + \underbrace{\begin{bmatrix} 2k_{c4}x_ny_n + k_{c5}(r^2 + 2x_n^2) \\ 2k_{c5}x_ny_n + k_{c4}(r^2 + 2y_n^2) \end{bmatrix}}_{\text{distorção tangencial}} \quad (2.4)$$

onde $r = x_n^2 + y_n^2$.

Os efeitos da distorção são visíveis na Figura 2.7. Os pontos marcados como *ideal* e *distorted* correspondem às posições dos pontos projetados sem e com os efeitos da distorção. Normalmente, os efeitos causados pela distorção radial (dr) são maiores do que os causados pela distorção tangencial (dt).

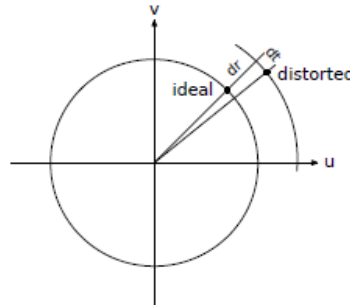


Figura 2.7: Efeitos da distorção radial (dr) e tangencial (dt). A imagem foi retirada de [56]

No contexto das CESF, a calibração das câmaras tem sido investigada, mais concretamente ao nível da correção da distorção das lentes. Em [57] são propostas correções baseadas num padrão uniforme, num padrão circular ou em redes neuronais, concluindo-se que para um modelo complexo ser construído de maneira precisa as redes neuronais são uma boa escolha.

Procedimento experimental

Iniciou-se o processo de calibração pela captura de padrões de xadrez a partir de várias distâncias e orientações das câmaras. Uma vez que as câmaras são muito pequenas, usaram-se padrões com quadrados de dimensões muito reduzidas. Foram utilizados padrões compostos por 5x7 quadrados com dimensões de 2x2mm e padrões compostos por 4x5 quadrados com dimensões

de 2x2mm e 3x3mm.

A captura de imagens desta cápsula, pelo facto de envolver outros utensílios tais como o gravador, que não dispomos e como tal teve de nos ser emprestado, não foi tarefa fácil uma vez que só foram realizadas duas capturas. Durante essas capturas era possível visualizar com alguma dificuldade as imagens capturadas por uma das câmaras no visor do gravador (havendo a possibilidade de escolher a câmara pretendida). Só no final, quando descarregadas as imagens para o software RAPID é que se conseguem ver as imagens com mais nitidez.



Figura 2.8: Imagem do cinto com gravador e monitor onde se visualizam as imagens capturadas por uma das câmaras.

Outra das dificuldades durante a captura de imagens foi o manuseamento da cápsula. Sendo muito pequena e tendo dispostas as duas câmaras nas extremidades, não era fácil segurá-la sem sujá-las ou ocluir uma delas.

Para além de estimar os parâmetros intrínsecos e os coeficientes de distorção da câmara, o objetivo era também estimar os parâmetros extrínsecos das câmaras de forma a saber a transformação de uma câmara em relação à outra; informação que é necessária nos algoritmos de odometria usados. Para isso, usaram-se espelhos na calibração (ver Figura 2.9), de modo a que através do algoritmo de [58] fosse possível estimar a pose das câmaras. Porém, verificou-se depois que as imagens capturadas que continham os padrões refletidos no espelho não eram viáveis. Estas imagens eram muito pouco nítidas, resultando na impossibilidade de detetar os cantos dos padrões.

Uma vez que as cápsulas endoscópicas são usadas em ambiente orgânico, onde vários líquidos podem alterar a distorção normal das lentes, foi também realizada a calibração com a câmara debaixo de água (ver Figura 2.10).

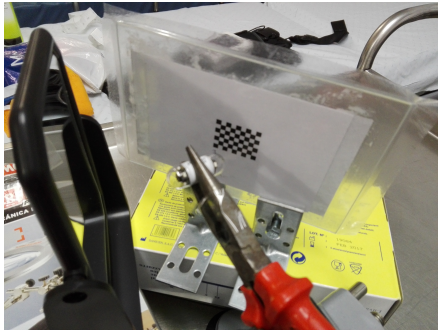


Figura 2.9: Imagem da montagem experimental da calibração com recurso a espelhos

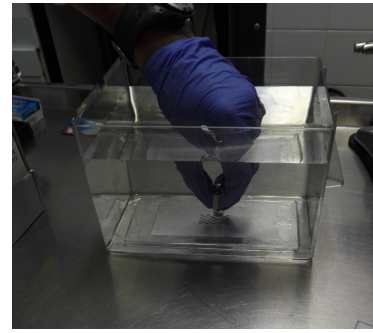


Figura 2.10: Imagem da montagem experimental da calibração com a câmara debaixo de água

Resultados

Para a calibração de cada câmara, foram usados vários conjuntos de imagens. As imagens usadas possuem uma resolução de 256x256 pixels. Cada conjunto corresponde a diferentes padrões, variando tanto o número de quadrados como as suas dimensões e usa diferentes números de imagens. Nas Tabelas 2.2 e 2.3 estão descritas as características de cada *dataset*, incluindo se a captura é feita com a cápsula submersa ou não.

	Câmara 1			
	Nºquadrados	Dimensões quadrados	Nº de imagens	Submersa
<i>Dataset 1</i>	5x7	2mm	26	Não
<i>Dataset 2</i>	4x5	3mm	27	Não
<i>Dataset 3</i>	4x5	2mm	43	Sim
<i>Dataset 4</i>	4x5	3mm	77	Sim

Tabela 2.2: Características dos datasets usados para calibração da câmara 1.

	Câmara 2			
	Nºquadrados	Dimensões quadrados	Nº de imagens	Submersa
<i>Dataset 1</i>	5x7	2mm	13	Não
<i>Dataset 2</i>	4x5	3mm	51	Não
<i>Dataset 3</i>	4x5	3mm	29	Sim

Tabela 2.3: Características dos datasets usados para calibração da câmara 2.

Foram realizadas quatro calibrações para a câmara 1 e três calibrações para a câmara 2. Os parâmetros intrínsecos para as duas câmaras e o erro médio de reprojeção correspondente a cada calibração estão apresentados nas Tabelas 2.4 e 2.5.

	Câmara 1			
	<i>Dataset 1</i>	<i>Dataset 2</i>	<i>Dataset 3</i>	<i>Dataset 4</i>
$f_x(px)$	77.681±0.312	73.519±0.507	95.249±2.393	72.630±0.686
$f_y(px)$	77.479±0.271	73.958±0.564	95.485±2.391	72.552±0.679
$c_x(px)$	132.692±0.07	132.976±0.114	132.876±0.112	133.284±0.116
$c_y(px)$	131.441±0.065	130.690±0.115	131.302±0.107	130.938±0.109
<i>skew</i>	0.170±0.031	-0.139±0.107	-0.275±0.038	-0.171±0.028
k_{c_1}	0.443±0.001	0.408±0.006	0.626±0.033	0.172±0.004
k_{c_2}	-0.521±0.010	-0.447±0.013	-1.117±0.115	-0.155±0.004
k_{c_3}	0.149±0.004	0.116±0.005	0.4634±0.071	0.026±0.002
k_{c_4}	-0.001±0.000	-0.002±0.000	0.003±0.001	-0.001±0.000
k_{c_5}	-0.000±0.000	-0.004±0.000	0.006±0.001	-0.003±0.000
Erro médio de reprojeção (px)	0.217	0.197	0.222	0.476

Tabela 2.4: Parâmetros intrínsecos para a câmara 1.

	Câmara 2		
	<i>Dataset 1</i>	<i>Dataset 2</i>	<i>Dataset 3</i>
$f_x(px)$	85.280±0.817	73.174±0.526	64.465±1.339
$f_y(px)$	85.448±0.733	72.921±0.522	64.548±1.334
$c_x(px)$	128.067±0.303	130.239±0.084	130.227±0.168
$c_y(px)$	126.158±0.307	128.374±0.078	128.267±0.219
<i>skew</i>	0.129±0.148	0.080±0.018	0.602±0.053
k_{c_1}	0.156±0.009	0.359±0.006	0.184±0.009
k_{c_2}	-0.172±0.009	-0.3696±0.011	-0.130±0.011
k_{c_3}	0.031±0.011	0.089±0.004	0.018±0.002
k_{c_4}	-0.001±0.001	0.001±0.000	0.004±0.001
k_{c_5}	-0.001±0.001	-0.002±0.000	-0.007±0.000
Erro médio de reprojeção (px)	0.879	0.237	0.351

Tabela 2.5: Parâmetros intrínsecos para a câmara 2.

2.3. Calibração das câmaras

Pelos resultados da calibração verifica-se que entre diferentes calibrações os parâmetros variam significativamente, principalmente a distância focal. Esta discrepância entre valores dos parâmetros pode dever-se à baixa qualidade e resolução das imagens ou à pouca variação da orientação da câmara durante a captura. A elevada variância que se verifica na distância focal está também relacionada com os coeficientes da distorção radial, sendo que provavelmente o terceiro coeficiente está a provocar o *overfitting* do modelo. Na Tabelas 2.6 e 2.7 estão os valores da média e da variância para cada parâmetro calculado, de onde se destaca a elevada variância da distância focal.

		Média	Variância
Câmara 1	fx	75.5981	8.6803
	fy	75.7189	6.2026
	cx	132.8338	0.0404
	cy	131.0652	0.2819
	s	0.0155	0.0479
	k1	0.4254	5.8824e-04
	k2	-0.4840	0.0028
	k3	0.1321	5.5444e-04
	k4	-0.0013	1.8000e-07
	k5	-0.0032	1.8050e-06
Câmara 2	fx	79.2268	73.2812
	fy	79.1844	78.4541
	cx	129.1530	2.3588
	cy	127.2660	2.4547
	s	0.1049	0.0012
	k1	0.2576	0.0207
	k2	-0.2708	0.0195
	k3	0.0601	0.0017
	k4	-3.5000e-04	1.4450e-06
	k5	-0.0014	2.0000e-08

Tabela 2.6: Média e variância dos parâmetros das câmaras sem estarem submersas.

		Média	Variância
Câmara 1	fx	83.9398	255.8322
	fy	84.0186	262.9704
	cx	133.0801	0.0833
	cy	131.1203	0.0662
	s	-0.2229	0.0053
	k1	0.3989	0.1034
	k2	-0.6356	0.4627
	k3	0.2448	0.0955
	k4	0.0012	6.4800e-06
	k5	0.0014	4.6080e-05

Tabela 2.7: Média e variância dos parâmetros da câmara 1 submersa.

Nos resultados da calibração é observável ainda que as imagens apresentam uma elevada distorção, sobretudo radial uma vez que o terceiro coeficiente k_{c_3} não é desprezável, antes pelo contrário apresenta valores significativamente elevados tendo em conta que este coeficiente corresponde r^6 na Equação 2.4.

Para visualizar o impacto das distorções radial e tangencial nas imagens encontram-se, da Figura 2.11 à 2.17, os mapas de distorção para cada calibração.

Na Figura 2.18 estão duas imagens capturadas pelas câmaras da cápsula onde é possível visualizar a distorção radial nos cantos das imagens.

- Câmara 1

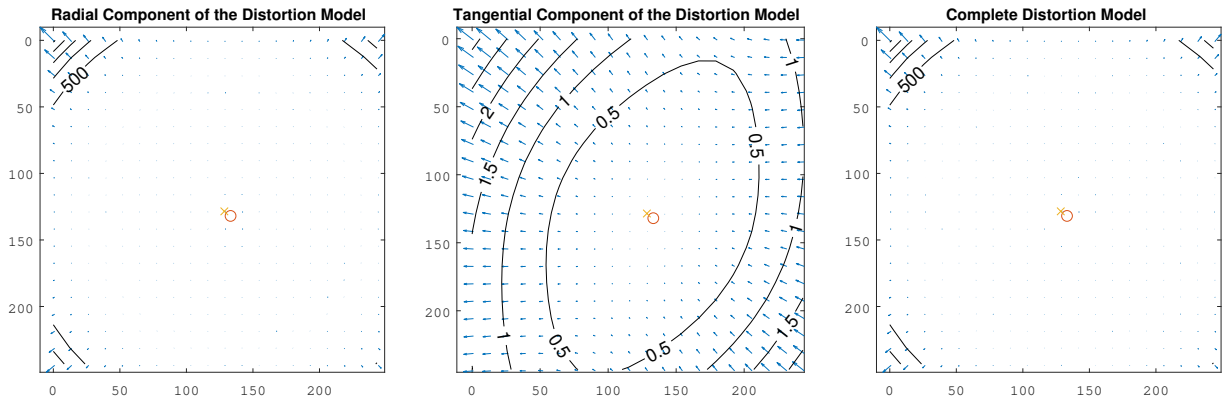


Figura 2.11: Mapas de distorção correspondentes à calibração da câmara 1 com o dataset 1.

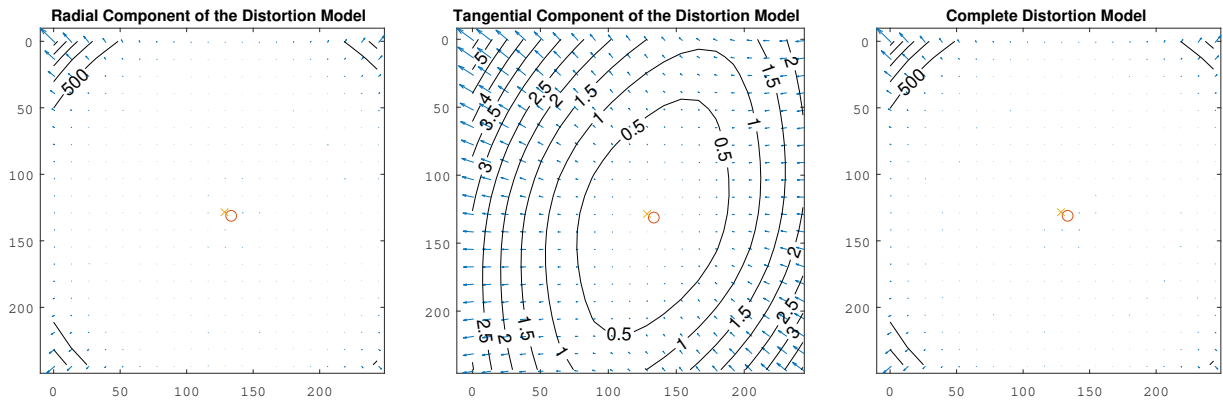


Figura 2.12: Mapas de distorção correspondentes à calibração da câmara 1 com o dataset 2.

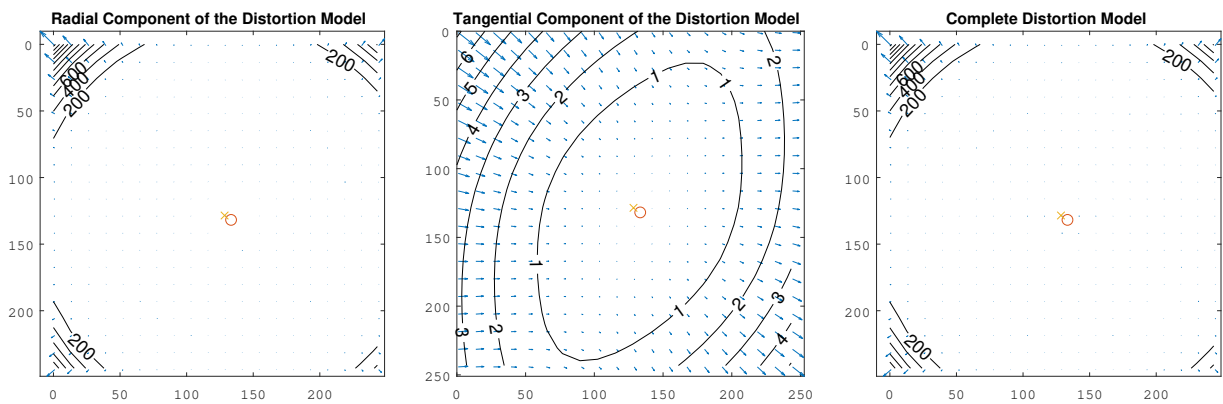


Figura 2.13: Mapas de distorção correspondentes à calibração da câmara 1 com o dataset 3.

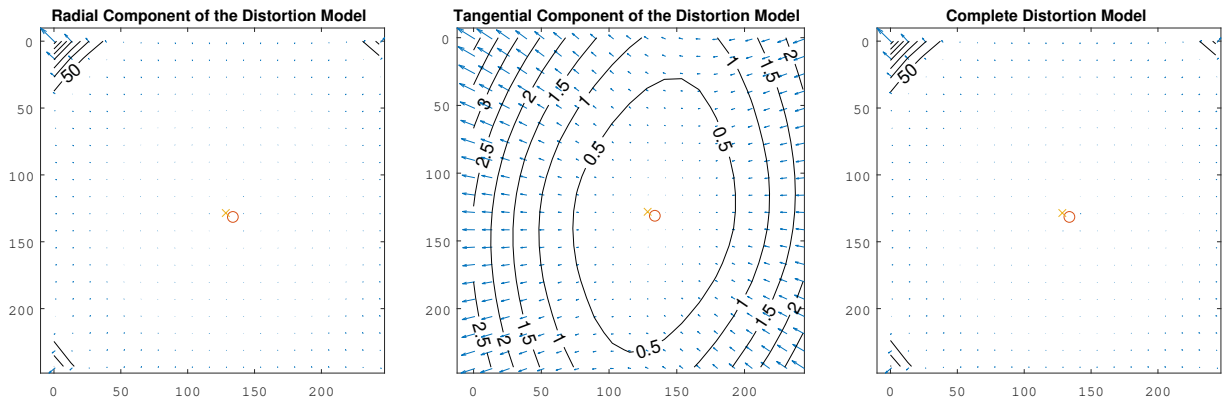


Figura 2.14: Mapas de distorção correspondentes à calibração da câmara 1 com o dataset 4.

- Câmera 2

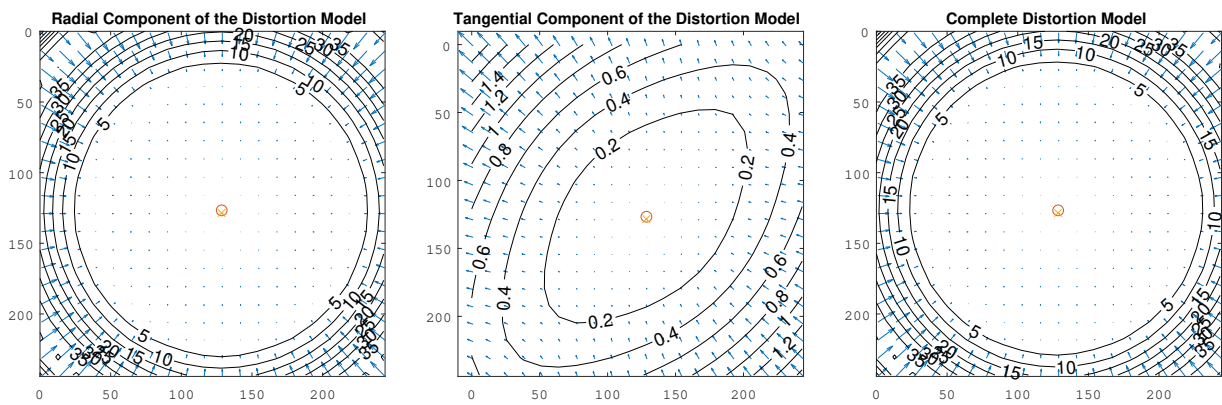


Figura 2.15: Mapas de distorção correspondentes à calibração da câmara 2 com o dataset 1.

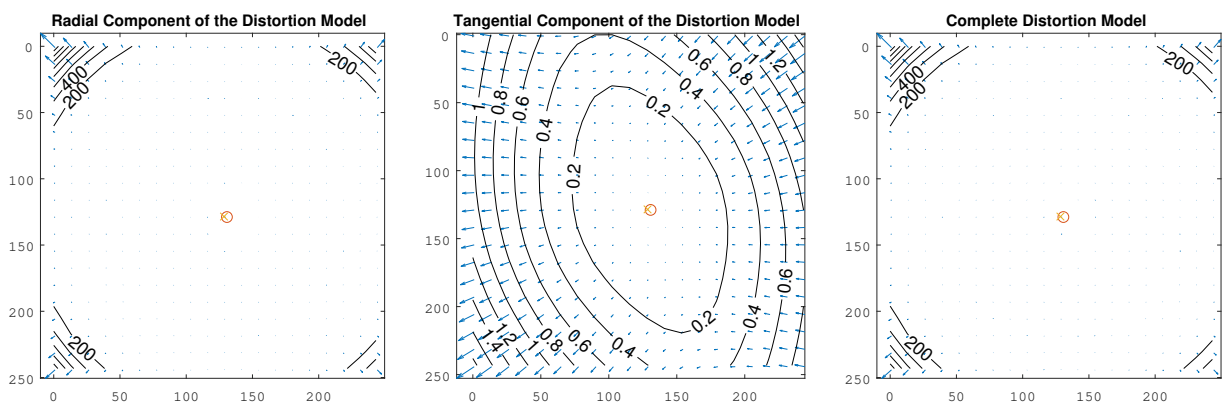


Figura 2.16: Mapas de distorção correspondentes à calibração da câmara 2 com o dataset 2.

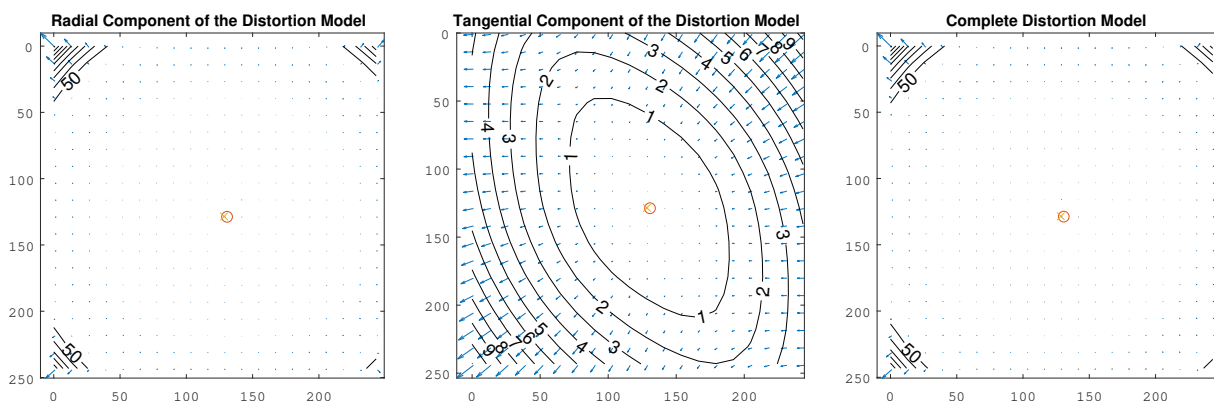


Figura 2.17: Mapas de distorção correspondentes à calibração da câmara 2 com o dataset 3.

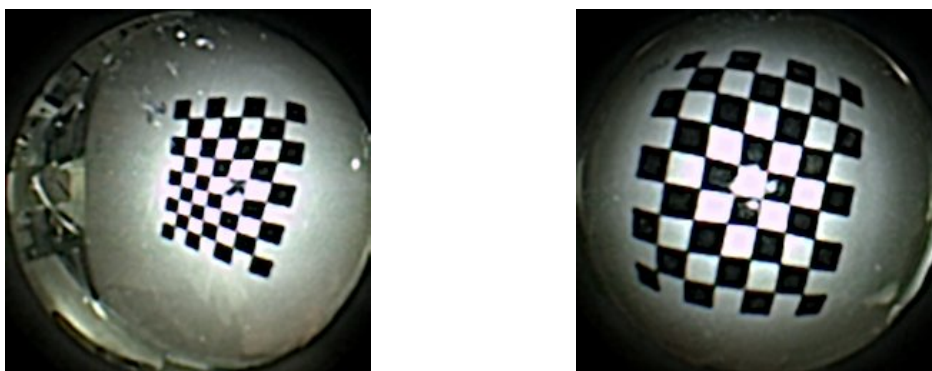


Figura 2.18: Imagens capturadas pela cápsula endoscópica.

Como se pode observar nos mapas e na Figura 2.18, a distorção radial existente nos cantos das imagens é muito elevada, o que já era expectável dadas as lentes terem dimensões muito reduzidas e possuírem um campo de visão muito amplo (172° de acordo com a Tabela 2.1). Já a distorção tangencial é mais baixa, mas não desprezável e afeta mais abrangentemente a imagem. Pelo facto dos valores da distorção radial serem demasiado elevados comparativamente aos da distorção tangencial, os mapas do modelo completo são praticamente iguais aos mapas da distorção radial.

Outra das possíveis causas para a existência de muita distorção radial e tangencial deve-se à estrutura ótica da cápsula. De acordo com a Figura 2.3 é possível observar que as câmaras estão no interior de uma cúpula feita de plástico biocompatível. Devido ao tipo de material da cúpula, mais concretamente ao seu índice de refração que é desconhecido, e também à sua estrutura esférica, os raios são refratados de maneira diferente à esperada causando desta forma distorção espacial não linear.

CAPÍTULO

3

ODOMETRIA VISUAL SEM SOBREPOSIÇÃO DE CAMPOS VISUAIS

Conteúdos

3.1	Introdução à Odometria Visual	23
3.1.1	Odometria Visual sem sobreposição de campos visuais	25
3.2	Estimação do movimento usando odometria monocular	26
3.2.1	Estimação do movimento individual de cada câmara a menos de um fator de escala	27
3.2.2	Estimação da escala	29
3.3	Estimação do movimento usando o modelo de câmara generalizada . .	32
3.3.1	Modelo de Câmara Generalizada	32
3.3.2	Estimação do movimento	33
3.3.3	Influência do tipo e do número de correspondências no <i>rank</i> da matriz A . .	37

Neste capítulo será feita uma introdução à Odometria Visual, estudando em particular o cenário em que não existe sobreposição dos campos visuais. Depois da introdução serão apresentados os métodos usados para estimar o movimento relativo da cápsula.

3.1 Introdução à Odometria Visual

A odometria visual consiste no processo de estimação da trajetória de um agente (ex.: veículo, humano ou robô) envolvendo apenas o uso de imagens adquiridas por uma câmara ou por um conjunto de câmaras a ele ligadas (Scaramuzza and Fraundorfer 2011) [59].

A ideia de estimar a pose de um veículo usando apenas imagens foi introduzida e descrita por Moravec no início dos anos 80 (Nistér et al. 2004) [60]. Desde aí, este método tem vindo a ser estudado em vários trabalhos de investigação, estando presente também em várias aplicações reais, como por exemplo na missão dirigida pela NASA a Marte no ano 2004.

O termo "Odometria Visual" foi adotado em 2004 por Nistér et al.. O termo surgiu devido à estimação baseada em visão ser similar à *Wheel odometry* onde o movimento do veículo é incrementalmente estimado integrando o deslocamento das rodas ao longo do tempo (Scaramuzza e Fraundorfer 2011). De maneira semelhante, a Odometria Visual usa o deslocamento dos píxeis entre imagens ao longo do tempo. Para que a Odometria Visual funcione corretamente o ambiente deve ter uma iluminação favorável e o cenário deve ser estático e com textura suficiente para extrair movimento, assim como os *frames* consecutivos devem ser capturados de forma a existir sobreposição suficiente entre eles.

Uma vez que as imagens armazenam quantidades significativas de informação, o uso de câmaras em vez de sensores ou sistemas como GPS, INS e sistemas baseados em localização laser permite estimar de forma mais simples e barata a pose de um robô. Em implementações atuais a odometria visual permite obter resultados com um erro de posição relativa entre 0.1 a 2 % apresentando portanto resultados mais exatos e precisos comparativamente às tecnologias GPS, INS, *Wheel odometry* e localização por sonar. A odometria visual pode ainda ser integrada num sistema GPS ou INS para uma precisão máxima. Para além das vantagens já mencionadas, realça-se o facto de esta tecnologia não ser afetada pelo deslizamento das rodas, ao contrário do que acontece com a *Wheel odometry*. Em comparação com os sistemas de Localização por Sonar e Laser, a odometria visual não emite nenhuma energia detetável no meio ambiente.

As câmaras são por norma pequenas e fáceis de aplicar em qualquer veículo ou usá-las em várias tarefas robotizadas. Em suma, esta tecnologia é bastante atrativa, apresentando um bom balanço

entre custo, fiabilidade, robustez e complexidade de implementação.

Apesar da localização *indoor* do robô ser bem sucedida, a localização em ambientes externos continua a ser um problema desafiante devido a fatores condicionantes, tais como: luz solar direta, sombras, mudanças dinâmicas no ambiente causados pelo vento e luz solar. Outro dos desafios que encontramos na odometria visual é o facto dos sistemas de visão monoculares apenas nos permitirem estimar a translação a menos de um fator de escala.

De acordo com Kitt et al. (2011) a estimação do fator de escala pode tornar-se errónea quando ocorre uma grande mudança na inclinação do piso, o que pode levar a uma estimativa incorreta da trajetória.

Para além das aplicações da Odometria Visual na área aeroespacial é também utilizada em: sistemas de navegação de submarinos, sistemas de navegação ponto a ponto, sistemas de assistência a condutores e robôs agrícolas.

Ao longo dos anos foram sendo desenvolvidas duas vertentes de investigação independentes na área da Odometria Visual; Odometria Visual *Stereo* e Monocular (Scaramuzza et al. , 2011). No caso monocular apenas conseguimos estimar o movimento a menos de um fator de escala. Para obter a escala absoluta pode usar-se medidas diretas (por exemplo medindo o tamanho de um elemento da cena), restrições de movimento ou integrar outros sensores.

Em relação à configuração *Stereo* para além de apresentar uma estimação de movimento mais robusta, permite obter a escala absoluta. No entanto, a *OV Stereo* pode degenerar para o caso monocular quando a distância na cena é muito maior do que a *baseline* (i.e., a distância entre as duas câmaras); neste caso deve ser usado o método monocular.

Outra das limitações da *OV stereo* deve-se ao campo visual numa configuração *stereo* clássica ser restritivo uma vez que é necessário que as duas câmaras estejam a observar a mesma cena.

Tem sido desenvolvido muito trabalho referente à estimação do *egomotion* de sistemas multi-câmara. Nistér et al. (2004) propôs a primeira implementação em tempo real com um robusto esquema de rejeição de *outliers* (RANSAC) usando um sistema *stereo* calibrado com sobreposição de campo visual. Também Frahm et al. usam um sistema multi-câmara para aplicar uma técnica de estimação 6DOF; para obter a escala do movimento da câmara são usadas vistas sobrepostas.

Clipp e co-autores (Clipp et al.,2009) apresentam-nos um sistema onde a sobreposição de campos visuais é mínima. Mais tarde apareceram outros trabalhos onde não existe sobreposição de campos visuais.

3.1.1 Odometria Visual sem sobreposição de campos visuais

Nos últimos anos, tem aumentado o interesse na estimação do movimento de sistemas multi-câmara, nomeadamente nos sistemas aplicados em veículos. A abordagem mais recorrente nestes casos é dispor as câmaras em diferentes sítios do veículo de forma a cobrirem campos de visão diferentes com pouca ou nenhuma sobreposição. Desta forma, é possível obter um campo de visão efetivo mais amplo resultando numa aquisição de dados mais eficiente e numa estimação do movimento mais precisa. Em ambientes pobres em textura é também vantajoso usar informação de campos visuais não sobrepostos.

Nesta dissertação, devido à estrutura da cápsula, o que se pretende é estimar o movimento de um sistema multi-câmara em que não existe qualquer sobreposição de campos visuais. Para tal, recorreu-se a duas abordagens possíveis: tratando as câmaras individualmente; e como uma única câmara usando o Modelo de Câmara Generalizada.

Em relação à abordagem das câmaras individualmente, Clipp et al., [61], introduziu um método para a estimação de movimento 6DOF usando a odometria monocular numa das câmaras e determinando a escala através da correspondência de um ponto extraído da segunda câmara. Houve autores que também resolveram o problema de estimação de movimento transformando-o num problema de triangulação, usando *Second-Order Cone Programming*, [62], e programação linear baseada no *Branch and Bound Algorithm*, [63].

No que diz respeito ao Modelo de Câmara Generalizada, Pless, [64], derivou a *Generalized Epipolar Constraint* expressa em termos da matriz essencial generalizada, uma matriz 6x6 que contém os seis graus de liberdade, permitindo desta forma usar um conjunto de câmaras como se fosse uma só, mesmo que as câmaras não partilhem um centro de projeção comum. Uma vez que a GEC possui 18 incógnitas (9 referentes à matriz essencial e 9 referentes à matriz de rotação) é expectável que se consiga determinar as matrizes E e R linearmente usando apenas 17 pontos, ou mais corretamente, 17 raios. No entanto, em [65], Li et al. mostram que para muitas configurações comuns de modelos de câmaras generalizadas não é possível aplicar o *17-point algorithm*. Isto deve-se à redução da característica da matriz do sistema de equações lineares, o que resulta num aumento da família de soluções. Estes autores propõem ainda um algoritmo linear alternativo para estas configurações degeneradas.

Também soluções minimais para câmaras generalizadas foram estudadas. Stéwenius et al., [66] apresentam um método de estimar as posições relativas de duas câmaras generalizadas apenas com 6 raios correspondentes. No entanto, este complexo algoritmo apresenta 64 soluções e degenera para os casos retratados em [63]. Vasconcelos et al., em [67], apresentam uma solução que apesar de não

ser minimal, permite estimar o movimento de uma câmara axial usando apenas 10 correspondências.

Mais recentemente, Kneip et al. apresentaram uma solução iterativa para a posse relativa da câmara generalizada com base na minimização do seu valor próprio [68]. Eles usam sete ou mais correspondências para garantir a convergência para uma única solução. Este método é muito mais rápido que os métodos anteriores de casos gerais, mas requer correspondências extras e também é suscetível de cair em mínimos locais. Estes autores implementaram ainda uma coleção de métodos em *Python* a que chamaram *OpenGV* e que resolve problemas de visão geométricos. De entre os algoritmos que eles implementaram está o algoritmo dos 17 pontos presente em [65] que foi também implementado neste trabalho.

3.2 Estimação do movimento usando odometria monocular

Como já foi referido, uma das abordagens usadas nesta dissertação para estimar o movimento onde não existe sobreposição de campos visuais faz uso da odometria monocular. Mais concretamente, foi implementado o algoritmo "Real Time 6D Stereo Visual Odometry with Non-Overlapping Fields of View" proposto por *Kazik* et al. em 2012 [69].

Os autores referidos apresentam-nos um método para estimar a escala absoluta de um movimento 6D, assim como a estrutura do sistema multi-câmara em ambientes interiores. O método funciona em tempo real e utiliza informação de duas câmaras sem sobreposição de campo visual. Para estimar o movimento do sistema de câmaras é primeiramente aplicada odometria visual monocular em cada câmara individualmente, sendo obtido o movimento relativo entre *frames* a menos de um fator de escala. Tendo em conta que sabemos a transformação estática entre câmaras através do processo prévio de calibração (Capítulo 2) é então possível, dada a restrição que liga as transformações euclidianas, determinar os fatores de escala correspondentes à translação de cada câmara. O método propõe também o uso de vários *frames* para estimar a escala - *windowed buffer* e ainda a inclusão do esquema RANSAC no algoritmo a fim de aumentar a sua robustez.

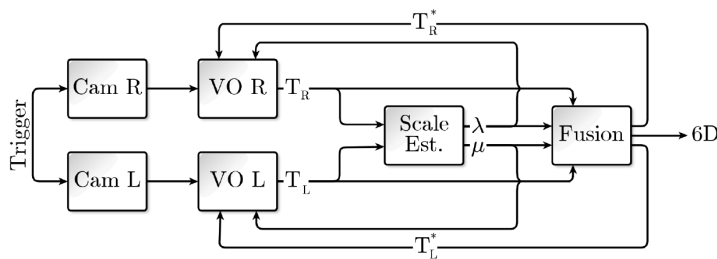


Figura 3.1: Esquema geral do sistema (*Kazik, 2012*)

3.2.1 Estimação do movimento individual de cada câmara a menos de um fator de escala

Para obter o movimento relativo em cada uma das câmaras é calculada a matriz essencial entre dois *frames* consecutivos, usando para isso os pontos característicos correspondentes desses mesmos *frames*.

Matriz Essencial

Tendo em conta duas câmaras *pinhole* em que sejam conhecidos os seus parâmetros intrínsecos, a matriz essencial \mathbf{E} ($\mathbb{R}^{3 \times 3}$) é expressa por

$$E = [\mathbf{t}]_{\times} \mathbf{R} \quad (3.1)$$

onde \mathbf{R} ($\mathbb{R}^{3 \times 3}$) é a matriz de rotação e $[\mathbf{t}]_{\times}$ corresponde à matriz *Skew-Symmetric* de translação :

$$[\mathbf{t}]_{\times} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (3.2)$$

Ambas a matriz de rotação \mathbf{R} e a translação \mathbf{t} possuem três graus de liberdade, no entanto, como \mathbf{E} é uma grandeza homogênea existe ambiguidade na escala e, como tal, a matriz \mathbf{E} possui apenas cinco graus de liberdade.

A matriz anti-simétrica tem dois valores singulares iguais e outro igual a zero. A multiplicação pela matriz de rotação não altera estes valores singulares, o que significa que a matriz essencial tem também dois valores singulares iguais e outro que é zero.

Desta forma, a matriz \mathbf{E} tem de satisfazer as seguintes restrições:

$$\det(E) = 0 \quad (3.3)$$

$$2EE^T E - \text{trace}(EE^T)E = 0 \quad (3.4)$$

Em termos dos pontos correspondentes, \mathbf{E} é definida pela equação

$$\hat{x}'^T E \hat{x} = 0 \quad (3.5)$$

onde \hat{x}' e \hat{x} são as coordenadas normalizadas da imagem para pontos correspondentes.

Neste trabalho foi usada a função `estimateEssentialMatrix` do MATLAB. Para estimar a matriz essencial, esta função, resolve o problema da pose relativa dos cinco pontos como se se tratasse de um problema polinomial de valores próprios, tal como sugere [70]. Esta função usa ainda o algoritmo *M-estimator sample consensus (MSAC)*, que exclui os *outliers* ou pontos que não se

ajustem tão bem ao modelo, para que se obtenha uma estimação robusta de \mathbf{E} .

Seguidamente, será descrito de forma sucinta o problema da pose relativa dos cinco pontos, bem como os métodos usados para o resolver.

5-pt relative pose problem

É sabido que para estimar a pose relativa de duas câmaras calibradas, o número mínimo de pontos correspondentes é cinco e estes têm que satisfazer a restrição 3.5. Para calcular \mathbf{E} é usual reescrever a restrição na forma $Q\tilde{E} = 0$, onde \tilde{E} contém os nove elementos da matriz \mathbf{E} e Q contém as coordenadas dos pontos das duas imagens. A matriz \mathbf{E} é depois parametrizada como uma combinação linear de vetores do espaço nulo da matriz Q . A esses quatro vetores correspondem diretamente quatro matrizes 3×3 E_1 , E_2 , E_3 e E_4 .

$$E = xE_1 + yE_2 + zE_3 + E_4 \quad (3.6)$$

Para determinar os coeficientes da combinação linear são usadas as restrições 3.3 e 3.4.

Uma solução para este problema apresentada por *Nistér et al.* [50] consiste em inserir 3.6 em dez restrições cúbicas 3.3, 3.4 de modo a construir dez equações polinomiais de terceiro grau com três incógnitas e 20 monómios. Estas equações podem ser escritas na forma

$$MX = 0 \quad (3.7)$$

onde M é a matriz de coeficientes reduzida pela eliminação de *Gauss-Jordan* e X é o vetor com todos os monómios.

Neste método são usadas relações entre polinómios de M para depois serem determinados os coeficientes de 3.6.

Outra solução mais estável e simples de implementar é proposta por *Kukelova et al.* Estes autores resolvem 3.7 como se fosse um problema polinomial de valores próprios.

Os problemas do tipo Polynomial eigenvalue problems (PEP) têm a forma

$$A(\lambda)\mathbf{v} = 0 \quad (3.8)$$

onde $A(\lambda)$ é uma matriz polinomial definida como

$$A(\lambda) \equiv \lambda^l C_l + \lambda^{l-1} C_{l-1} + \dots + \lambda C_1 + C_0 \quad (3.9)$$

em que C_j são matrizes quadradas $n \times n$.

Os autores de [70] transformam então 3.7 numa equação do tipo 3.8. Para isso escolhem arbitrariamente uma das incógnitas do vetor de monómios X , por exemplo z , e substituem-na por

λ , de modo a que desenvolvendo 3.8 resulte em

$$(z^3 C_3 + z^2 C_2 + z C_1 + C_0) \mathbf{v} = 0 \quad (3.10)$$

onde \mathbf{v} é um vetor 10×1 de monómios e C_3, C_2, C_1 e C_0 são matrizes 10×10 .

O problema é depois facilmente resolvido usando algoritmo eficientes como a função `polyeig` do MATLAB. Depois de resolvido o PEP são obtidos 30 valores próprios z e 30 correspondentes vetores próprios \mathbf{v} . Depois, é finalmente estimada E através de 3.6.

Recuperar a rotação \mathbf{R} e a translação \mathbf{t} a partir da matriz essencial

Determinada a matriz essencial \mathbf{E} é então possível estimar a rotação e a translação das câmaras a menos de um fator de escala. Basicamente, é necessário decompor a matriz essencial em $E = SR$ onde S é a matriz anti-simétrica e R é a rotação. Para isso, são usadas duas matrizes:

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{e} \quad Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

A matriz W é uma matriz de rotação e a matriz Z é uma matriz anti-simétrica.

Supondo que a Decomposição em valores singulares (DVS) da matriz essencial é $E \sim U \text{diag}(1, 1, 0) V^T$. Usando \mathbf{W} e \mathbf{Z} obtemos, ignorando os sinais, duas possíveis fatorizações $E = SR$:

$$S = UZU^T \quad R = UWV^T \quad \text{ou} \quad R = UW^T V^T \quad (3.11)$$

Como $S = [t]_{\times}$, então obtém-se $\mathbf{t} = [tx, ty, tz]$. Uma vez que o sinal de \mathbf{E} e \mathbf{t} não pode ser determinado, existem duas escolhas possíveis também para \mathbf{t} . Deste modo, existem quatro combinações de \mathbf{R} e \mathbf{t} possíveis para a segunda câmara, onde apenas uma delas é a correta. Para escolher a combinação \mathbf{R} - \mathbf{t} correta é feita a triangulação através do *Midpoint Method* [71] para obter os pontos reconstruídos. É escolhida a solução onde os pontos reconstruídos se situam em frente às duas câmaras.

3.2.2 Estimação da escala

A estimação da escala é possível devido à restrição implícita nas transformações euclidianas das câmaras em diferentes instantes temporais bem como na transformação entre câmaras. Na Figura 3.2 são ilustradas as quatro transformações que dão origem à seguinte expressão algébrica:

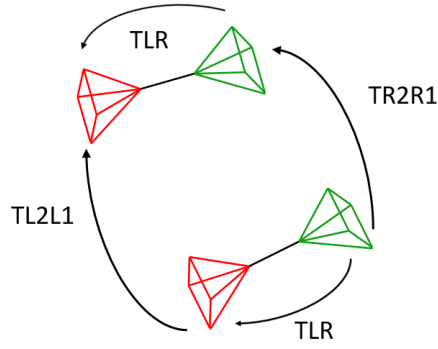


Figura 3.2: Representação das transformações euclidianas do sistema de câmaras durante o frame 1 e o frame 2 (T_{R2R1} e T_{L2L1}) e transformação rígida entre câmaras (T_{LR}).

$$T_{L2L1}T_{LR} = T_{LR}T_{R2R1} \quad (3.12)$$

Expandindo as transformações euclidianas e introduzindo os fatores de escala desconhecidos λ e μ para as câmaras direita e esquerda respetivamente, obtém-se:

$$\begin{bmatrix} R_{L2L1} & \mu L2t_{L2L1} \\ 0_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} R_{LR} & Lt_{LR} \\ 0_{1 \times 3} & 1 \end{bmatrix} = \begin{bmatrix} R_{LR} & Lt_{LR} \\ 0_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} R_{R2R1} & \lambda R2t_{R2R1} \\ 0_{1 \times 3} & 1 \end{bmatrix} \quad (3.13)$$

Onde as matrizes de rotação R_{R2R1} e R_{L2L1} e os vetores de translação t_{R2R1} e t_{L2L1} correspondem às transformações das câmaras direita e esquerda estimadas entres os *frames* 1 e 2.

Decompostos os termos relativos à rotação e translação, resultam as expressões:

$$\begin{aligned} R_{L2L1} R_{LR} &= R_{LR} R_{R2R1} \\ R_{L2L1} Lt_{LR} + \mu L2t_{L2L1} &= R_{LR} \lambda R2t_{R2R1} + Lt_{LR} \end{aligned} \quad (3.14)$$

Resolvendo a parte translacional em ordem aos fatores de escala, obtém-se a seguinte expressão que pode ser resolvida através do Método das Médias dos Mínimos Quadrados (MMMQ)

$$\underbrace{\begin{bmatrix} R_{LR} R2t_{R2R1} & -L2t_{L2L1} \end{bmatrix}}_{:=A_i} \underbrace{\begin{bmatrix} \lambda \\ \mu \end{bmatrix}}_{:=x_i} = \underbrace{(R_{L2L1} - I_{3 \times 3}) Lt_{LR}}_{:=b_i} \quad (3.15)$$

Deste modo, é possível calcular a escala absoluta referente ao movimento de cada câmara usando a pseudo-inversa de *Moore-Penrose*.

$$x_i = A_i^\dagger b_i \quad (3.16)$$

Multi-frame window

Para melhorar o condicionamento das restrições do MMMQ (3.15), os autores sugerem que se estimem as escalas não apenas entre dois *frames* mas ao longo de vários *frames* recentes, assumindo desta forma que as escalas são localmente constantes. Considerando N o número de *frames*, as escalas são calculadas usando mais restrições.

$$\underbrace{\begin{bmatrix} R_{LR} R_2^{(1)} t_{R2R1} & -L_2^{(1)} t_{L2L1} \\ R_{LR} R_2^{(2)} t_{R2R1} & -L_2^{(2)} t_{L2L1} \\ \dots & \dots \\ R_{LR} R_2^{(N)} t_{R2R1} & -L_2^{(N)} t_{L2L1} \end{bmatrix}}_{:=A_{LS}} \underbrace{\begin{bmatrix} \lambda \\ \mu \end{bmatrix}}_{:=x_{LS}} = \underbrace{\begin{bmatrix} ({}^{(1)}R_{L2L1} - I_{3 \times 3}) L t_{LR} \\ ({}^{(2)}R_{L2L1} - I_{3 \times 3}) L t_{LR} \\ \dots \\ ({}^{(N)}R_{L2L1} - I_{3 \times 3}) L t_{LR} \end{bmatrix}}_{:=b_{LS}} \quad (3.17)$$

Estimação robusta da escala

É possível existirem equações mal condicionadas no cálculo das escalas em 3.15. Isto, pode dever-se a uma imprecisão na estimativa do movimento num dado instante ou ao movimento realizado pelo sistema de câmaras ser degenerado. Para estas situações, em que é necessária a remoção das equações *outliers*, isto é, equações que não são satisfeitas, os autores sugerem o uso do esquema RANSAC.

Deste modo, foi então usado um esquema RANSAC para o qual foram criadas três funções:

- *Fitting*: São calculadas as escalas λ e μ , usando um conjunto mínimo das amostras. Neste caso, apenas é necessária uma restrição entre duas imagens.
- *Degenerate*: Verifica se a amostra escolhida corresponde a um caso degenerado antes que seja usada na função *fitting*.
- *Distance*: Esta função testa todas as outras equações de encontro ao modelo ajustado. Uma equação é guardada como *inlier* se a hipótese de escala satisfaz suficientemente a equação 3.15. A equação é *inlier* se $\left| \frac{\|A_i x_{hip}\|}{\|b_i\|} - 1 \right| < t_{dist}$ com $t_{dist}=0.3$.

Algoritmo 1 *Real Time 6D Stereo Visual Odometry Algorithm*

Inputs: Pontos 2D projetados na câmara L e na câmara R (*frame 1* \leftrightarrow *frame 2*);
 T_{LR} , matriz que representa a transformação rígida entre câmaras;
 K , matriz dos parâmetros intrínsecos das câmaras;

Outputs: Fatores de escala λ e μ

Método:

1. Escolher pelo menos 5 correspondências da câmara esquerda L e 5 correspondências da câmara direita R.
 2. Estimar a matriz essencial E para cada câmara.
 3. Decompor a matriz E e recuperar R e t a menos de um fator de escala para cada câmara.
 4. Aplicar a equação 3.15 ou 3.17 no caso de se usar mais que um *frame* no cálculo da escala.
 5. Obter as escalas λ e μ através da equação 3.16 usando um esquema RANSAC.
 6. Representar os movimentos das câmaras num referencial comum.
 7. Fusão das odometrias das duas câmaras numa só odometria resultante das médias ponderadas das translações e das rotações.
-

3.3 Estimação do movimento usando o modelo de câmara generalizada

A outra abordagem a que se recorreu para estimar o movimento do sistema de câmaras consiste em considerar este sistema como uma só câmara generalizada. O interesse em usar o Modelo de Câmara Generalizada (MCG) tem aumentado consideravelmente nos últimos anos, em parte, devido ao desenvolvimento de novos modelos de câmaras não convencionais, tais como câmaras catadióptricas ou sistemas multi-câmara.

3.3.1 Modelo de Câmara Generalizada

Em 2001, *Grossberg e Nayar*, [72], descrevem um modelo de imagem generalizado e explicam como se pode substituir os píxeis físicos por um modelo matemático equivalente a que nos referimos como um píxel de raio ou um "raxel". Um *raxel* é um elemento fotossensível virtual que mede a energia da luz de um conjunto compacto de raios e que pode ser representado como um cone ou como um único raio. Os autores propõem também um método de calibração que associa a cada *raxel* um conjunto de parâmetros medidos em relação aos sistema de coordenadas da câmara associada.

Pless, [64], para derivar a Restrição Epipolar Generalizada (REG), usa o MCG assumindo apenas que os píxeis são substituídos por raios de projeção. Para converter os píxeis em raios, em vez de usar os parâmetros do modelo de imagem generalizado, simplifica, usando as coordenadas de

Plücker, usadas para representar retas no espaço.

As coordenadas de *Plücker* de uma reta consistem num par vetores de três dimensões: o vetor unitário de direção \mathbf{x} e o vetor de momento $(\mathbf{v} \times \mathbf{x})$, onde \mathbf{x} é um vetor com a mesma direção da linha e \mathbf{v} é o ponto correspondente ao centro da câmara. Desta forma, um raio de imagem que atravesse o centro de câmara \mathbf{v} com uma direção \mathbf{x} pode ser representado por um vetor de *Plücker* $\mathbf{L} = (\mathbf{x}^T, (\mathbf{v} \times \mathbf{x})^T)^T$. A condição para um par de linhas representadas por coordenadas de *Plücker* \mathbf{L} e \mathbf{L}' se intersectar é expressa por:

$$\mathbf{L}'^T \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \mathbf{L} = 0 \quad (3.18)$$

Quando é aplicada a uma linha \mathbf{L} uma transformação euclidiana com rotação \mathbf{R} e translação \mathbf{t} obtém-se uma linha diferente segundo a transformação:

$$L_{trans} = \begin{bmatrix} R & 0 \\ E & R \end{bmatrix} \mathbf{L} \quad (3.19)$$

A equação que *Pless* derivou, a REG, expressa o facto de que uma linha \mathbf{L} , quando transformada através de um movimento euclidiano, vai encontrar outra linha \mathbf{L}' . Essas linhas representam os raios pertencentes a duas câmaras que se encontram num ponto correspondente do espaço. Combinando as duas equações anteriores, obtemos então a REG, que deve ser satisfeita para cada par de raios correspondentes de uma câmara generalizada em movimento.

$$\begin{aligned} 0 &= \mathbf{L}'^T \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \mathbf{L}_{trans} \\ &= \mathbf{L}'^T \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} R & 0 \\ E & R \end{bmatrix} \mathbf{L} \\ &= \mathbf{L}'^T \begin{bmatrix} E & R \\ R & 0 \end{bmatrix} \mathbf{L} \end{aligned} \quad (3.20)$$

Esta equação define totalmente como duas vistas de um mesmo ponto restringem o movimento discreto de uma câmara e pode ser reescrita da seguinte forma:

$$\mathbf{x}'^T \mathbf{E} \mathbf{x} + (\mathbf{v}' \times \mathbf{x}')^T \mathbf{R} \mathbf{x} + \mathbf{x}'^T \mathbf{R} (\mathbf{v} \times \mathbf{x}) = 0 \quad (3.21)$$

3.3.2 Estimação do movimento

Um método teórico bem conhecido para estimar o movimento usando MCG consiste em usar no mínimo 17 raios correspondentes entre dois *frames* e estimar \mathbf{E} e \mathbf{R} linearmente, uma vez que temos 18 incógnitas (9 que correspondem à matriz essencial e 9 que correspondem à matriz de rotação). Neste caso, coloca-se a equação 3.21 na forma $\mathbf{A}\mathbf{X}=0$, sendo \mathbf{A} uma matriz $N \times 18$ e \mathbf{X}

um vetor de incógnitas composto pelos elementos de E e R . Deste modo, ao decompor a matriz A em valores singulares, $A=UDV^T$, a última coluna da matriz V corresponderá à solução de X .

Porém, perante a falta de resultados numéricos deste método, *Li et al.*, [65], testam o algoritmo para várias configurações de câmaras generalizadas, concluindo que para algumas delas não é possível usar o algoritmo *standard* SVD, uma vez que a matriz A apresenta um *rank* inferior a 17, originando ambiguidades na determinação de R . De entre as configurações degeneradas, são estudados três casos: Projecção central localizada; Axial; e Central localizada e Axial.

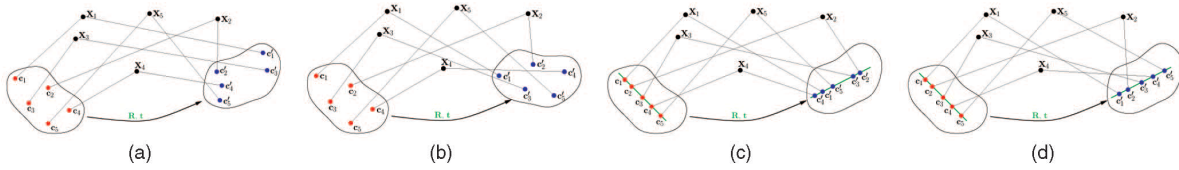


Figura 3.3: (a) Caso mais geral: os raios das imagens não têm quaisquer restrições; (b) Caso da projeção central localizada: para cada ponto 3D da cena x_i o centro de projeção é único e fixo relativamente ao sistema de coordenadas da câmara; (c) O caso axial: todos os raios das imagens intersectam uma linha comum, o eixo; O caso central localizada e axial: todos os raios intersectam o eixo e raios correspondentes passam pelo mesmo ponto fixo de projeção.

- **Projeção central localizada:** É uma câmara generalizada que consiste num conjunto de raios de projeção centrais localizados. Os conjuntos de câmaras sem sobreposição de campos visuais são exemplos deste caso. Quando o conjunto de câmaras se move de uma posição inicial para uma final, os pontos 3D X são vistos pela mesma câmara nas duas posições, o que implica que os raios correspondentes às duas imagens de X passem pelo mesmo ponto no sistema de coordenadas da câmara, o centro de câmara. Assume-se ainda, que cada câmara do sistema é uma câmara de projeção central, ou seja, todos os raios passam pelo centro de câmara. Representando os raios no sistema de coordenadas do conjunto de câmaras, existe correspondência entre pontos $(x_i, v_i) \leftrightarrow (x'_i, v_i)$, sendo que $v'_i = v_i$. Segundo 3.21, obtêm-se então:

$$x_i'^T E x_i + (v_i \times x_i')^T R x_i + x_i'^T R (v_i \times x_i) = 0 \quad (3.22)$$

Considerando (E, R) uma solução deste conjunto de equações, com $E \neq 0$, percebemos que $(0, I)$ também é uma solução, uma vez que substituindo $(0, I)$ em 3.22 se obtém $(v_i \times x_i')^T x_i + x_i'^T (v_i \times x_i) = 0$.

Deste modo, obtemos o conjunto de soluções: $(\lambda E, \lambda R + \mu I)$; isto significa que o espaço nulo da matriz A tem dimensão 2, logo $rank(A) = 16$.

- **Axial:** Uma câmara generalizada onde todos os raios se intersectam numa única linha, o eixo, é classificada como Axial. Exemplos de configurações axiais são um par de câmaras de projeção

centrais rigidamente acoplado ou um conjunto de câmaras de projeção centrais com centros colineares.

Para analisar esta configuração é aconselhável escolher a origem do sistema de coordenadas do mundo de forma a que esta pertença ao eixo, caso contrário o valor de \mathbf{E} será diferente. No sistema de coordenadas sugerido, o centro de câmara poderá estar em qualquer ponto ao longo do eixo, como tal podemos definir, $\mathbf{v}_i = \alpha_i \mathbf{w}$ e $\mathbf{v}'_i = \alpha'_i \mathbf{w}$, onde \mathbf{w} é o vetor de direção do eixo. Desta forma, 3.21 fica:

$$\mathbf{x}_i'^T \mathbf{E} \mathbf{x}_i + \alpha'_i (\mathbf{w} \times \mathbf{x}_i)^T \mathbf{R} \mathbf{x}_i + \alpha_i \mathbf{x}_i'^T \mathbf{R} (\mathbf{w} \times \mathbf{x}_i) = 0 \quad (3.23)$$

Supondo que (\mathbf{E}, \mathbf{R}) é uma solução válida destas equações, pode-se também considerar $(0, \mathbf{w} \mathbf{w}^T)$ uma solução verdadeira uma vez que $(\mathbf{w} \times \mathbf{x}_i)^T \mathbf{w} = \mathbf{w}^T (\mathbf{w} \times \mathbf{x}_i) = 0$.

Consequentemente, tem-se o conjunto de soluções : $(\lambda \mathbf{E}, \lambda \mathbf{R} + \mu \mathbf{w} \mathbf{w}^T)$; isto significa que o espaço nulo da matriz A tem dimensão 2, logo $Rank(A)=16$.

- **Central localizada e axial:** Este caso merece particular atenção, uma vez que corresponde à configuração do sistema de câmaras da cápsula em estudo (que consiste num par de câmaras *stereo* sem sobreposição de campos visuais). Quando uma câmara axial é também uma câmara central localizada, estamos perante uma configuração ainda mais degenerada. A condição de ser central localizada implica que, na equação 3.23, se tem $\alpha_i = \alpha'_i$. Deste modo, é identificada também a solução $(0, [\mathbf{w}]_{\times})$ dado que $(\mathbf{w} \times \mathbf{x}_i)^T [\mathbf{w}]_{\times} \mathbf{x}_i + \mathbf{x}_i'^T [\mathbf{w}]_{\times} (\mathbf{w} \times \mathbf{x}_i) = (\mathbf{w} \times \mathbf{x}_i)^T (\mathbf{w} \times \mathbf{x}_i) + (\mathbf{x}'_i \times \mathbf{w})^T (\mathbf{w} \times \mathbf{x}_i) = 0$.

Assim, obtém-se o conjunto de soluções : $(\alpha \mathbf{E}, \alpha \mathbf{R} + \beta \mathbf{I} + \gamma [\mathbf{w}]_{\times} + \delta \mathbf{w} \mathbf{w}^T)$; isto significa que o espaço nulo da matriz A tem dimensão 4, logo $Rank(A)=14$.

Os mesmos autores, [65], propõem um algoritmo linear capaz de estimar o movimento para as configurações degeneradas anteriormente caracterizadas. O algoritmo que será posteriormente descrito é aplicável a situações onde o conjunto de equações não tem *full rank*, como é o caso do sistema de câmaras da cápsula. Apesar de nestes casos existir uma família de soluções (4 para este caso específico), a ambiguidade reside somente da determinação da matriz de rotação \mathbf{R} e não na matriz \mathbf{E} . Por este motivo, a mudança que se faz no algoritmo dos 17 pontos, consiste em determinar numa primeira fase só a matriz essencial. Desta forma, dado o conjunto de equações $A \begin{pmatrix} vec(\mathbf{E}) \\ vec(\mathbf{R}) \end{pmatrix} = 0$, determina-se a solução que minimiza $\|A \begin{pmatrix} vec(\mathbf{E}) \\ vec(\mathbf{R}) \end{pmatrix}\|$ sujeito a $\|\mathbf{E}\| = 1$ ao invés de $\|\begin{pmatrix} vec(\mathbf{E}) \\ vec(\mathbf{R}) \end{pmatrix}\| = 1$ como no algoritmo SVD *standard*.

O problema é resolvido de forma semelhante à apresentada em [53] (apêndice 5.4.2).

Extração da rotação e da translação

Depois de calculada a matriz \mathbf{E} , esta pode ser decomposta segundo $\mathbf{E}=[\mathbf{t}]_{\times}\mathbf{R}$ obtendo-se a rotação e a translação a menos de um fator de escala. No entanto, neste método, a decomposição de \mathbf{E} é realizada de diferente modo em relação ao método usado com pares de câmaras *pinhole*. As diferenças são:

1. A decomposição de $\mathbf{E}=[\mathbf{t}]_{\times}\mathbf{R}$ resulta em duas matrizes de rotação possíveis R e R' . Esta ambiguidade pode ser resolvida uma vez que apenas uma das matrizes de rotação satisfaz a REG.
2. Enquanto que na matriz essencial *standard* a translação \mathbf{t} só é estimada a menos de um fator de escala, numa câmara generalizada é possível estimar a escala de \mathbf{t} sem ambiguidade. Isto porque, existe apenas uma escala de translação possível para a matriz correta de rotação \mathbf{R} que satisfaz Restrição Epipolar Generalizada (REG).

Para o cálculo de \mathbf{t} , a matriz essencial \mathbf{E} das equações 3.21 é substituída por $\mathbf{E}=[\mathbf{t}]_{\times}\mathbf{R}$. Uma vez que é conhecida a matriz de rotação \mathbf{R} , é possível estimar $[\mathbf{t}]_{\times}$ a partir de:

$$\mathbf{x}_i'^T [\mathbf{t}]_{\times} (\mathbf{R}\mathbf{x}_i) + (\mathbf{v}' \times \mathbf{x}_i')^T \mathbf{R}\mathbf{x}_i + \mathbf{x}_i'^T \mathbf{R}(\mathbf{v} \times \mathbf{x}_i) = 0 \quad (3.24)$$

Dado que se pretende obter três incógnitas referentes à matriz anti-simétrica de translação 3.2, sendo estas t_x , t_y e t_z , são precisos no mínimo três pares de raios correspondentes, tendo em atenção que não podem ser todos da mesma câmara. Para o cálculo do vetor t , o sistema de equações lineares é representado sob a forma matricial $Cx = b$, onde C é uma matriz $N \times 3$, (sendo N o número de pares de raios correspondentes), x é o vetor que diz respeito às três incógnitas e $b \neq 0$. No caso de se usar apenas raios de uma câmara, verifica-se que o *rank* da matriz C baixa de três para dois e o sistema passa de sistema possível determinado para um sistema possível indeterminado, ou seja, não é possível estimar nem o fator de escala nem a direção do vetor de translação.

Algoritmo linear para estimação de movimento de sistemas multi-câmara

Algoritmo 2 *Linear Motion Estimation for Multicamera Systems*

Inputs: Conjunto de correspondências $(x_i, v_i) \leftrightarrow (x'_i, v'_i)$. Neste caso, como se trata de um sistema de suas câmaras *stereo* (câmara generalizada do tipo projeção central localizada e axial), $v_i = v'_i$ para todo o i ; e v_i situa-se numa linha onde está contida a origem do sistema de coordenadas da câmara. O número mínimo de correspondências para este caso é 14.

T_{LR} , matriz que representa a transformação rígida entre câmaras;

K , matriz dos parâmetros intrínsecos das câmaras;

Outputs: Matriz de rotação R e vetor de translação t (com escala).

Método:

1. Para cada correspondência estabelecer uma equação segundo 3.21 e transformar o conjunto de equações em $A_E \text{vec}(E) + A_R \text{vec}(R) = 0$.
 2. Calcular a pseudo-inversa de A_R e escrever a equação para $\text{vec}(E)$ como $B \text{vec}(E) = 0$, sendo $B = (A_R A_R^\dagger - I) A_E$. Resolver esta equação usando o algoritmo de SVD *standard*.
 3. Decompor a matriz E de modo a obter o *twisted pair* R e R' .
 4. Sabendo as matrizes de rotação possíveis, resolver as equações 3.21 para obter t linear. Uma vez que as equações não são homogêneas em t , a translação será calculada com a escala correta. Escolher R ou R' com as respectivas translações t e t' de acordo com o par (R, t) que apresenta um menor resíduo na equação 3.21.
-

3.3.3 Influência do tipo e do número de correspondências no *rank* da matriz A

Tipos de correspondências- intra-câmara e inter-câmara

Diferentes tipos de correspondências de pontos foram testados em ambiente de simulação, tanto correspondências intra-câmara como inter-câmara, também denominadas por correspondências cruzadas. As correspondências de pontos intra-câmara são vistas pela mesma câmara ao longo de dois *frames* consecutivos, enquanto as correspondências de pontos inter-câmaras referem-se a correspondências que são vistas por diferentes câmaras ao longo de dois *frames* consecutivos.

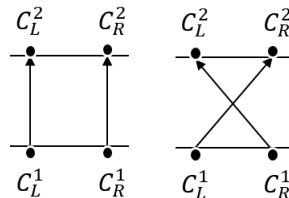


Figura 3.4: Do lado esquerdo estão representadas as correspondências intra-câmara e do lado direito estão representadas as correspondências inter-câmara. Os índices 1 e 2 referem-se aos frames e as letras L e R correspondem às câmaras esquerda e direita respectivamente.

De acordo com os autores de [61] e [67] seria de esperar que no caso de se usarem correspondências

cruzadas o *rank* da matriz A fosse 16 e como tal, tivessem de ser usadas 16 correspondências para estimar o movimento do sistema de câmaras *stereo*. Contudo, após testes com correspondências inter-câmara os resultados não vão ao encontro do que seria esperado. Isto é, para um movimento geral com rotação e translação, bastam 14 correspondências para se conseguir estimar o movimento mesmo usando correspondências cruzadas. Por este motivo, foi realizada uma análise ao tipo de degenerescência que ocorre no caso das correspondências cruzadas, de acordo com o método de Kanade et al. em [73]. Nesta publicação são identificados casos degenerados do algoritmo dos 17-pontos por dedução e não por exemplo como acontece em [65]. Este método começa por dividir a matriz A em duas matrizes, uma matriz que contém as direções dos raios A_d e outra que contém os centros de projeção A_c , sendo que usa apenas a última para determinar se o algoritmo degenera.

$$A = \begin{bmatrix} a_E^{1T} & a_E^{1T} C(O_1^1, O_2^1) \\ a_E^{2T} & a_E^{2T} C(O_1^2, O_2^2) \\ \vdots & \vdots \\ a_E^{nT} & a_E^{nT} C(O_1^n, O_2^n) \end{bmatrix} = \begin{bmatrix} a_E^{1T} & 0 & \cdots & 0 \\ 0 & a_E^{2T} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_E^{nT} \end{bmatrix}_{n \times 9n} \times \begin{bmatrix} I_{9 \times 9} & C(O_1^1, O_2^1) \\ I_{9 \times 9} & C(O_1^2, O_2^2) \\ \vdots & \vdots \\ I_{9 \times 9} & C(O_1^n, O_2^n) \end{bmatrix}_{9n \times 18} \triangleq A_d A_c \quad (3.25)$$

onde n corresponde ao índice dos raios correspondentes e $C(O_1, O_2)$ é a *matriz dos centros*, uma matriz 9×9 particionada em 6 submatrizes:

$$C(O_1, O_2) = \begin{bmatrix} [O_2]_{\times} & -O_{1z} I_{3 \times 3} & O_{1y} I_{3 \times 3} \\ O_{1z} I_{3 \times 3} & [O_2]_{\times} & -O_{1x} I_{3 \times 3} \\ -O_{1y} I_{3 \times 3} & O_{1x} I_{3 \times 3} & [O_2]_{\times} \end{bmatrix} \quad (3.26)$$

sendo que O_1 e O_2 representam os centros das câmaras nos instantes 1 e 2 sob a forma de vetores, isto é, $O_1 = [O_{1x} \ O_{1y} \ O_{1z}]^T$ e $O_2 = [O_{2x} \ O_{2y} \ O_{2z}]^T$.

A transformação da matriz A desde a sua expressão em 3.21 até à sua forma decomposta em 3.25 é explicada em detalhe em [73].

Como já foi referido, a análise centra-se na informação sobre as configurações da câmara generalizada dada pela matriz A_c . Mais concretamente na análise da nulidade da *matriz dos centros*. Na publicação de Kanade et al. são apresentadas várias proposições que relacionam a nulidade desta matriz com a configuração dos centros das câmaras, das quais interessam particularmente:

- **Proposição 1:** Se $|O_1| \neq |O_2|$, a nulidade da *matriz dos centros* é 1; o seu *rank* é 8. Caso contrário a nulidade é 3; o seu *rank* é 6.

- **Proposição 3:** Se $|O_1| = |O_2|$, o vetor do espaço nulo 3D da *matriz dos centros* de rank 6 pode ser expresso em forma de matriz como

$$R_O(\alpha[\bar{O}_1]_{\times} + \beta(I_{3 \times 3} - \bar{O}_1\bar{O}_1^T) + \gamma\bar{O}_1\bar{O}_1^T) \quad (3.27)$$

onde R_O é a rotação que satisfaz $O_2 = R_O O_1$ e α , β e γ são escalares.

Tendo em conta que o caso em estudo se trata de um sistema de câmaras *stereo* sem sobreposição de campos visuais (Figura 3.5), foi analisada a *matriz dos centros* para esta configuração, considerando correspondências cruzadas e não cruzadas.

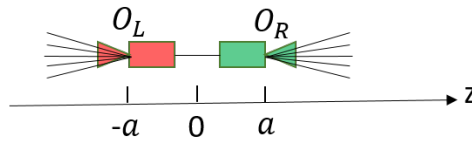


Figura 3.5: Esquema representativo do sistema de câmaras stereo sem sobreposição de campos visuais com respectivos centros de projeção das câmaras O_R e O_L , simétricos segundo o eixo dos z

Os centros das câmaras em análise são expressos por $O_R = [0 \ 0 \ a]^T$ e $O_L = [0 \ 0 \ -a]^T$ e estão representados na Figura 3.5. O centro de projeção de cada câmara permanece igual após o movimento desta, logo $O_{1R} = O_{2R}$ e $O_{1L} = O_{2L}$.

Começando pelas correspondências não cruzadas, temos duas câmaras, logo duas *matrizes dos centros*, $C(O_{1R}, O_{2R})$ e $C(O_{1L}, O_{2L})$. Como as correspondências são intra-câmara tem-se $|O_{1R}| = |O_{2R}|$ e $|O_{1L}| = |O_{2L}|$, então pelas proposições 1 e 3, ambas as matrizes $C(O_{1R}, O_{2R})$ e $C(O_{1L}, O_{2L})$ têm $rank=6$ e os seus espaços nulos de dimensão 3 podem ser representados por 3.27, sendo que $R_O = I_{3 \times 3}$.

Nas correspondências cruzadas teremos novamente duas *matrizes dos centros* $C(O_{1R}, O_{2L})$ e $C(O_{1L}, O_{2R})$ e sabe-se também que $|O_{1R}| = |O_{2L}|$ e $|O_{1L}| = |O_{2R}|$ dada a simetria entre O_R e O_L . Deste modo, novamente pelas proposições 1 e 3 verifica-se que ambas as matrizes $C(O_{1R}, O_{2L})$ e $C(O_{1L}, O_{2R})$ têm $rank 6$ e que os seus espaços nulos de dimensão 3 podem ser representados por 3.27, porém $R_O = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

Sabendo o espaço nulo das *matrizes dos centros*, é necessário aferir informação relativa à nulidade de A_C , para isso é importante a seguinte proposição:

- **Proposição 6:** Se todas as matrizes dos centros tiverem o espaço nulo comum entre elas, A_c tem esse espaço nulo. Inversamente, se A_c tem esse espaço nulo, é sempre possível escolher o sistema de coordenadas das câmaras apropriado tal que todas as matrizes de centros tenham

o espaço nulo comum.

Assim sendo, a matriz A_c relativa às correspondências não cruzadas, 3.28, constituída por n matrizes dos centros tem o espaço nulo destas, ou seja 3.27 com $R_O = I_{3 \times 3}$, uma vez que os espaços nulos das matrizes dos centros são todos iguais.

$$A_c(\text{n\~{a}o cruzadas}) = \begin{bmatrix} I_{9 \times 9} & C(O_{1R}^1, O_{2R}^1) \\ I_{9 \times 9} & C(O_{1R}^2, O_{2R}^2) \\ \vdots & \vdots \\ I_{9 \times 9} & C(O_{1L}^{n-1}, O_{2L}^{n-1}) \\ I_{9 \times 9} & C(O_{1L}^n, O_{2L}^n) \end{bmatrix}_{9n \times 18} \quad (3.28)$$

O mesmo raciocínio se aplica à matriz A_c das correspondências cruzadas, que terá o mesmo espaço nulo que as matrizes dos centros que a constituem, isto é, 3.27 com $R_O = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

$$A_c(\text{cruzadas}) = \begin{bmatrix} I_{9 \times 9} & C(O_{1R}^1, O_{2L}^1) \\ I_{9 \times 9} & C(O_{1R}^2, O_{2L}^2) \\ \vdots & \vdots \\ I_{9 \times 9} & C(O_{1L}^{n-1}, O_{2R}^{n-1}) \\ I_{9 \times 9} & C(O_{1L}^n, O_{2R}^n) \end{bmatrix}_{9n \times 18} \quad (3.29)$$

Como as duas matrizes A_c têm um espaço nulo sabe-se que existe degenerescência no algoritmo dos 17-pontos tanto para correspondências não cruzadas como para cruzadas, uma vez que os autores Kanade et al. chegam também à seguinte proposição:

- **Proposição 5:** Se A_c tem um espaço nulo, o algoritmo de 17 pontos é degenerado.

Dado que $A = A_d A_c$ o espaço nulo de A_c está contido no espaço nulo de A , ou seja, a nulidade de A tem de ser maior ou igual à nulidade de A_c . Assim, já se sabe que A tem nulidade pelo menos de 3, ou seja, tem no máximo $rank=15$.

Foi realizada uma simulação para um movimento geral com rotação R e translação t , onde foram calculadas as matrizes A_d e A_c para os casos de correspondências cruzadas e não cruzadas, sendo utilizados 18 raios no total. Como centros de projecção foram usados os vetores: $O_R = [0 \ 0 \ 1]^T$ e $O_L = [0 \ 0 \ -1]^T$. Com esta simulação pretendia-se calcular o espaço nulo de A_c , o espaço nulo de A_d e claro, o espaço nulo de A fazendo uma análise do mesmo.

- Para **correspondências não cruzadas** obtiveram-se os seguintes resultados:

$$N(A_C)_{\text{n\~{a}o cruzadas}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{18 \times 3}$$

Pode verificar-se que a matriz $N(A_C)_{\text{n\~{a}o cruzadas}}$ corresponde ao espaço nulo descrito em 3.27.

Sabendo que $R_O = I_{3 \times 3}$ e $[O_1]_{\times} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ temos:

$$R_O \alpha [O_1]_{\times} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ (1ª coluna; linha 10 a 18)}$$

$$R_O \beta (I_{3 \times 3} - \bar{O}_1 \bar{O}_1^T) = R_O \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ (2ª coluna; linha 10 a 18)}$$

$$R_O \gamma \bar{O}_1 \bar{O}_1^T = R_O \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ (3ª coluna; linha 10 a 18)}$$

Em relação à matriz das direções A_d , esta apresenta *full rank*.

Obtida a matriz A por multiplicação de A_d por A_c , obteve-se então o seu espaço nulo.

$$N(A) = \begin{bmatrix} 0 & 0 & -0.0000 & 0 \\ 0 & 0 & -0.0000 & 0 \\ 0 & 0 & 0.0000 & 0 \\ 0 & 0 & 0.0000 & 0 \\ 0 & 0 & -1.0000 & 0 \\ 0 & 0 & -1.7321 & 0 \\ 0 & 0 & -0.0000 & 0 \\ 0 & 0 & 1.7321 & 0 \\ 0 & 0 & -1.0000 & 0 \\ 0 & 1.0000 & 0.2679 & 0 \\ -1.0000 & 0 & -0.0000 & 0 \\ 0 & 0 & -0.0000 & 0 \\ 1.0000 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 \\ 0 & 0 & -1.0000 & 0 \\ 0 & 0 & 0.0000 & 0 \\ 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}_{18 \times 4} \quad (3.30)$$

Em $N(A)$, para além de estar inserido o espaço nulo de A_c , espera-se que esteja também um vetor que corresponderá aos 9 elementos da matriz essencial e aos 9 elementos da matriz de rotação correspondentes ao movimento. Consta-se que realmente, na terceira coluna, os 9 primeiros elementos correspondem a uma matriz de $rank=2$ com dois valores singulares iguais e que corresponde à matriz $[t] \times R$ a menos de um fator de escala.

$$E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1.0000 & -1.7321 \\ 0 & 1.7321 & -1.0000 \end{bmatrix}$$

Por outro lado, a matriz que deveria ser a matriz R não apresenta características de uma matriz de rotação. Não é ortogonal e o seu determinante não é 1. Esta indeterminação na matriz de rotação já seria de esperar, uma vez que só a parte da solução correspondente à matriz E não é afetada pela ambiguidade inerente à família de soluções do espaço nulo.

Deste modo, está provado que o espaço nulo de A para correspondências não cruzadas tem dimensão 4, isto é, o $rank(A)=14$.

- Para as **correspondências cruzadas** foram obtidos resultados muito semelhantes:

$$N(A_C)_{cruzadas} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Também neste caso se verifica que a matriz $N(A_C)_{cruzadas}$ corresponde ao espaço nulo descrito em 3.27 mas agora com $R_O = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

$$R_O \alpha [O_1]_{\times} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ (1ª coluna; linha 10 a 18)}$$

$$R_O \beta (I_{3 \times 3} - \bar{O}_1 \bar{O}_1^T) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ (2ª coluna;}$$

linha 10 a 18)

$$R_O \gamma \bar{O}_1 \bar{O}_1^T = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (3^{\text{a}} \text{ coluna; linha 10 a 18})$$

Em relação à matriz das direções A_d , tem novamente *full rank*.

A matriz $N(A)$ obtida para o caso da correspondências cruzadas foi:

$$N(A) = \begin{bmatrix} 0 & 0 & -0.0000 & 0 \\ 0 & 0 & -0.0000 & 0 \\ 0 & 0 & 0.0000 & 0 \\ 0 & 0 & 0.0000 & 0 \\ 0 & 0 & -1.0000 & 0 \\ 0 & 0 & -1.7321 & 0 \\ 0 & 0 & -0.0000 & 0 \\ 0 & 0 & 1.7321 & 0 \\ 0 & 0 & -1.0000 & 0 \\ 0 & -1.0000 & 3.7321 & 0 \\ 1.0000 & 0 & -0.0000 & 0 \\ 0 & 0 & -0.0000 & 0 \\ 1.0000 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 \\ 0 & 0 & -1.0000 & 0 \\ 0 & 0 & 0.0000 & 0 \\ 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

Verifica-se novamente que o espaço nulo da matriz A contém $N(A_C)_{\text{cruzadas}}$ e uma coluna que contém os 9 elementos da matriz essencial a menos de um fator de escala.

Finalmente prova-se que para correspondências cruzadas a matriz A tem espaço nulo de dimensão 4, isto é, $rank=14$. Logo serão apenas necessárias 14 correspondências (cruzadas) na estimação do movimento de um sistema de câmaras *stereo* sem sobreposição de campos visuais como o que está representado na Figura 3.5.

Números de correspondências de cada câmara

Em ambiente de simulação, para além de serem testadas as correspondências cruzadas, foram ainda testados diferentes números de correspondências para cada câmara. Através destes testes foi possível observar que ao existirem menos de seis correspondências numa das câmaras, isto é, usando por exemplo 9 correspondências da câmara R e 5 da câmara L , o *rank* da matriz A desce impossibilitando a estimação do movimento.

Ainda através da análise dos espaços nulos de A verificou-se que, para um número de correspondências de uma das câmaras inferior a 6, por cada correspondência a menos a dimensão do espaço nulo da matriz A aumenta, sendo que as dimensões dos espaços nulos de A_d e A_c permanecem iguais.

3.3. Estimação do movimento usando o modelo de câmara generalizada

Isto acontece porque por mais correspondências que sejam usadas na matriz A (3.25), se forem correspondências todas da mesma câmara, o *rank* máximo que a matriz A terá será 8. Ao inserirmos correspondências de uma câmara diferente o *rank* de A irá aumentar, sendo que para este caso seriam precisas 6 correspondências de uma segunda câmara (contando com pelo menos já 8 correspondências de uma primeira câmara) para que A tivesse *rank* 14.

Concluindo, considerando 14 correspondências provenientes de duas câmaras, uma das câmaras terá de ter no mínimo 6 correspondências para que o algoritmo não degenere. Na tabela 3.1 é possível observar a relação entre os números de correspondências e o *rank* da matriz A .

Nº de correspondências Cam L/R	Nº de correspondências Cam R/L	$Rank(A)$
14	0	8
13	1	9
12	2	10
11	3	11
10	4	12
9	5	13
8	6	14

Tabela 3.1: *Rank da matriz A conforme os números de correspondências em cada câmara.*

CAPÍTULO

4

RESULTADOS EXPERIMENTAIS

Conteúdos

4.1	Simulador	46
4.2	Resultados	46

4.1 Simulador

Neste capítulo é feita uma análise aos resultados obtidos pelos algoritmos descritos em 3.2 e 3.3. Como já havia sido referido anteriormente, ambos os algoritmos foram testados num ambiente simulado em MatLab. Mais concretamente, foram implementadas um conjunto de funções que resultam num simulador básico que gera dados sintéticos de modo a que estes sejam utilizados nos algoritmos de estimação de odometria.

Para o estudo dos algoritmos foi considerado o sistema de câmaras *pinhole* representado na Figura 4.1, cuja configuração pretende-se que seja semelhante à da cápsula *PillCam Colon 2*.

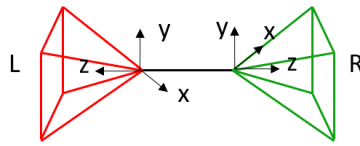


Figura 4.1: Sistema de câmaras usado nas simulações representado em 3D. O sistema é constituído por duas câmaras rigidamente acopladas e com campos visuais opostos.

As matrizes de projeção consideradas para as duas câmaras foram:

$$P_{camL} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}; \quad P_{camR} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.1)$$

Em relação aos parâmetros intrínsecos, considerou-se que $K = I_{3 \times 3}$.

No simulador são gerados aleatoriamente pontos 3D que são depois transformados segundo um movimento rígido, isto é, segundo uma matriz que contém uma rotação e uma translação. Estes pontos 3D transformados são projetados em cada uma das câmaras, obtendo-se para cada uma os respetivos pontos 2D. No caso do algoritmo 3.3 os pontos 2D são normalizados de modo a obter-se um vetor de direção unitário.

4.2 Resultados

O objetivo primário dos testes realizados é a estimação da posição relativa do sistema de câmaras no mundo, determinando os fatores de escala associados a cada *frame* de movimento. De modo a comparar a robustez de ambos os algoritmos, outro dos objetivos consistiu na inserção de

ruído nos dados, mais precisamente nos pixels correspondentes aos pontos 2D, de forma a tentar compreender o seu efeito e quantificar o seu impacto na estimação do fator de escala.

É extremamente importante que estes algoritmos sejam também eficazes na presença de ruído, uma vez que o estudo destes tem como finalidade a aplicação prática em imagens de uma cápsula endoscópica, ou seja imagens de pouca qualidade e com muito ruído. Para tentar aproximar o mais possível o ruído inserido nos testes àquele que as imagens reais teriam, foi utilizado o ruído Gaussiano. O ruído Gaussiano é um ruído estatístico cuja a função de probabilidade \mathcal{P} , é igual à da distribuição normal:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (4.2)$$

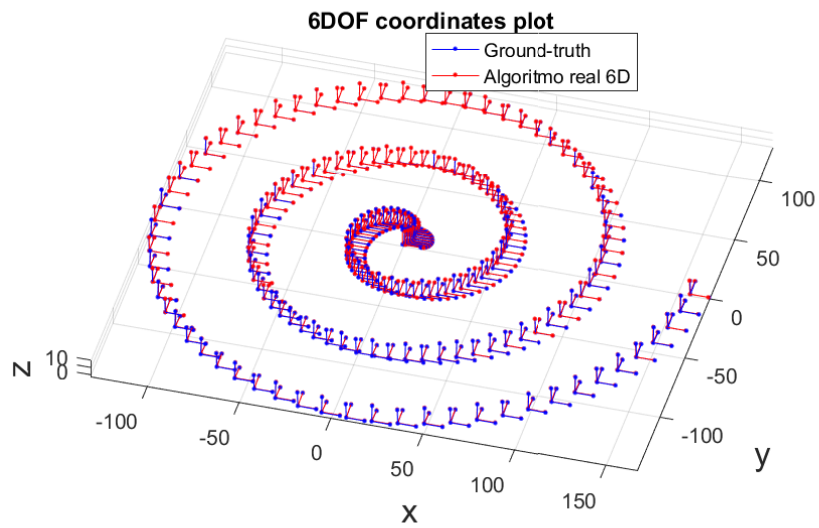
sendo μ a média e σ o desvio padrão. A média do ruído inserido μ foi sempre 0, variando-se apenas o desvio padrão σ .

De seguida, serão apresentados os resultados obtidos para dois movimentos genéricos do sistema de câmaras, mais especificamente: uma trajetória (2D) e uma trajetória 3D. Ambas as trajetórias são constituídas por movimentos com translação e rotação entre 200 *frames* consecutivos.

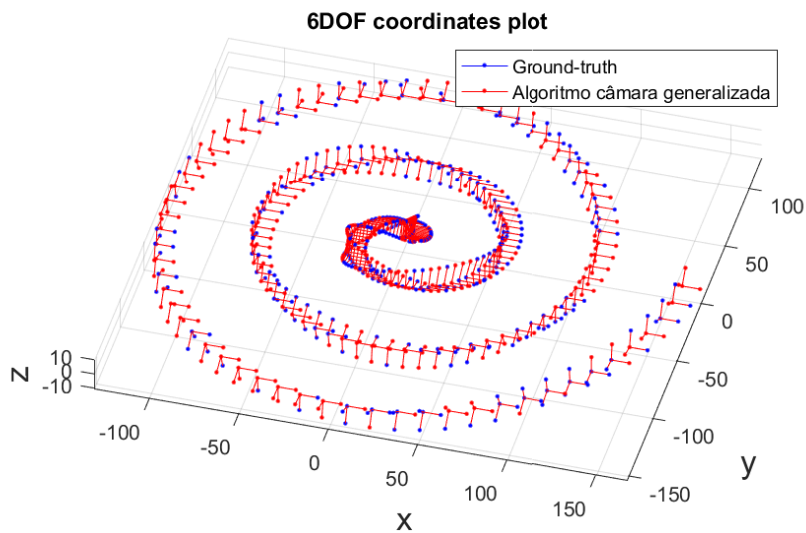
Nas Figuras 4.2 e 4.3 mostram-se as estimativas dos movimentos calculadas para as duas trajetórias usando os dois algoritmos ainda sem a adição de qualquer ruído nos pixels. É visível que sem a interferência do ruído os dois algoritmos conseguem ser eficazes na estimação do movimento.

Nas Figuras 4.4 e 4.5 é representada a estimação do movimento mas agora com a inserção de ruído $\sigma = 0.005$. Verifica-se que em (a) existe um erro muito maior que em (b), onde existem alguns *outliers*.

Comparando os dois algoritmos usados para estimar as trajetórias, conclui-se que ambos conseguem obter estimativas muito boas na ausência de ruído. No entanto, com inserção de ruído nos pixels dos pontos 2D, as estimativas ficaram muito aquém do esperado, no caso do algoritmo *Real Time 6D* a trajetória estimada aproxima-se da forma da trajetória original mas apresenta valores de posição muito diferentes. Já o algoritmo da câmara generalizada revela-se sensível ao ruído na medida em que apresenta vários *outliers* que se destacam da trajetória.



(a)



(b)

Figura 4.2: Na Figura (a), está representada a estimativa do movimento para o algoritmo 3.2 para uma trajetória 2D na ausência de ruído. Como se pode verificar a estimativa está exatamente igual ao ground-truth. A Figura (b) apresenta o movimento estimado para a mesma trajetória segundo o algoritmo 3.3 também sem ruído. mais uma vez a trajetória estimada coincide com o movimento ground-truth.

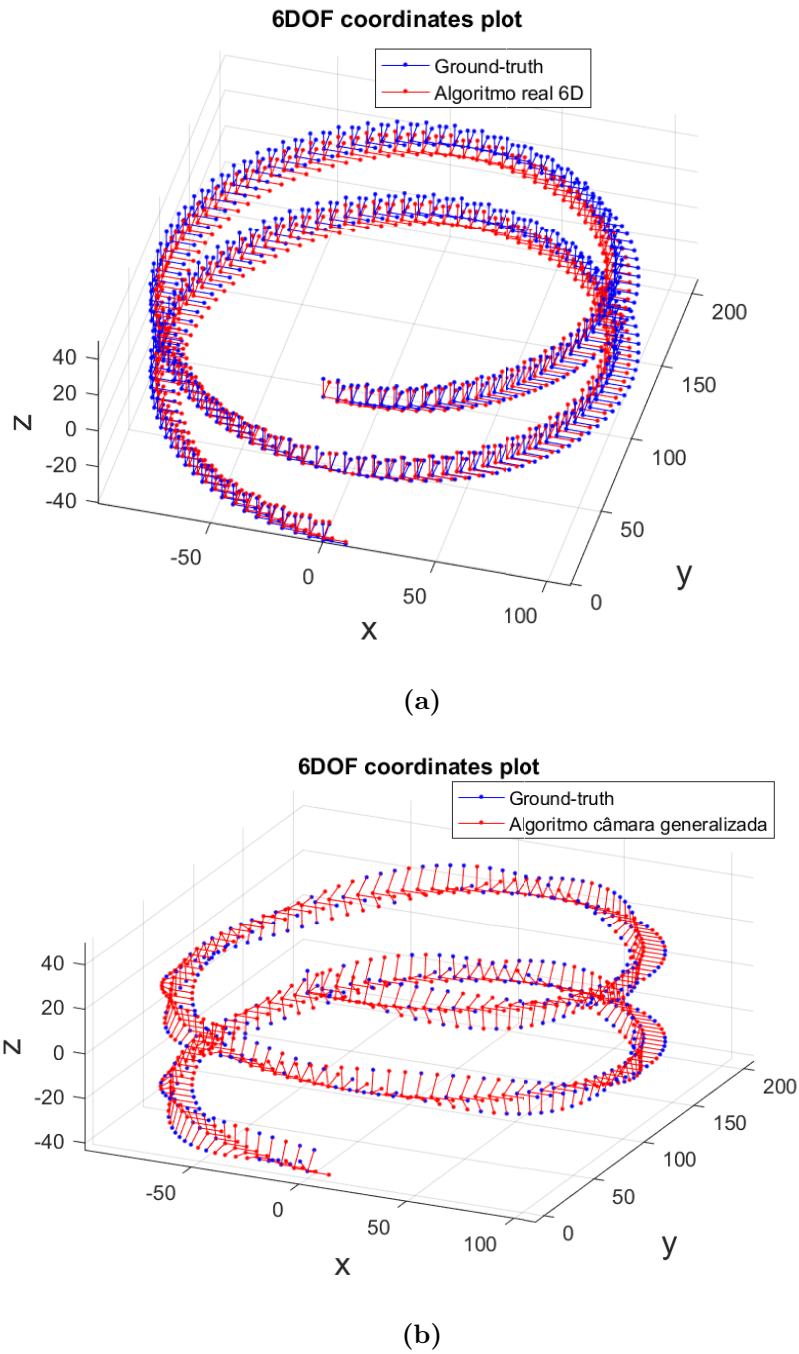
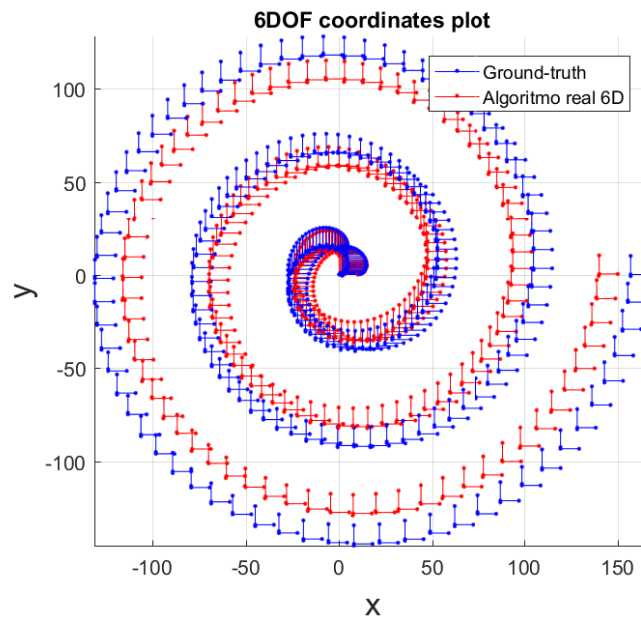
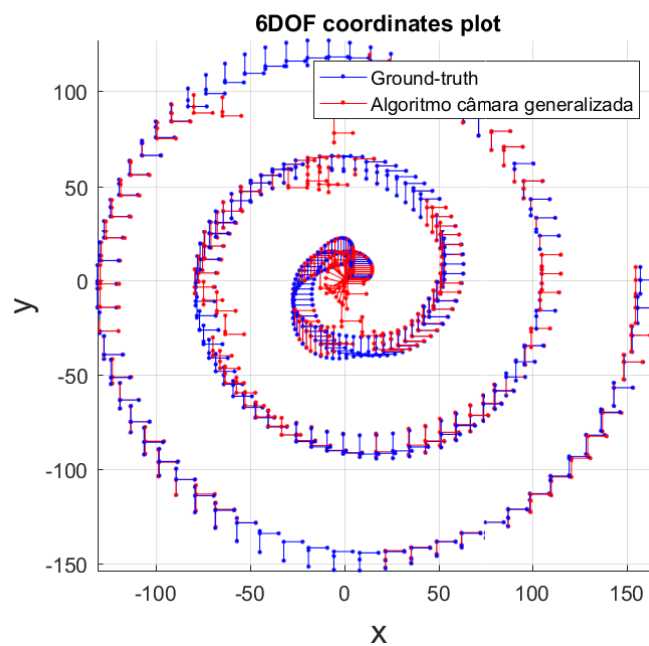


Figura 4.3: Na Figura (a) está representada a estimativa do movimento realizada pelo algoritmo 3.2 para uma trajetória 3D na ausência de ruído. A estimativa do movimento através do algoritmo 3.3 encontra-se na Figura (b) onde mais uma vez ainda não foi inserido qualquer ruído. É possível observar que as estimações feitas pelos dois algoritmos coincidem com a trajetória ground-truth.

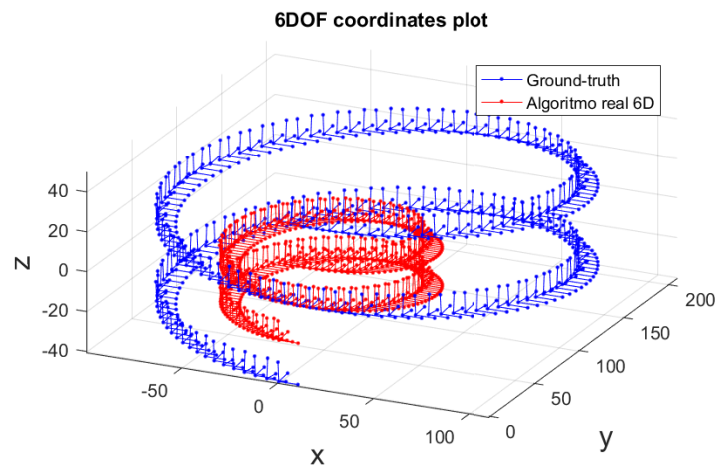


(a)

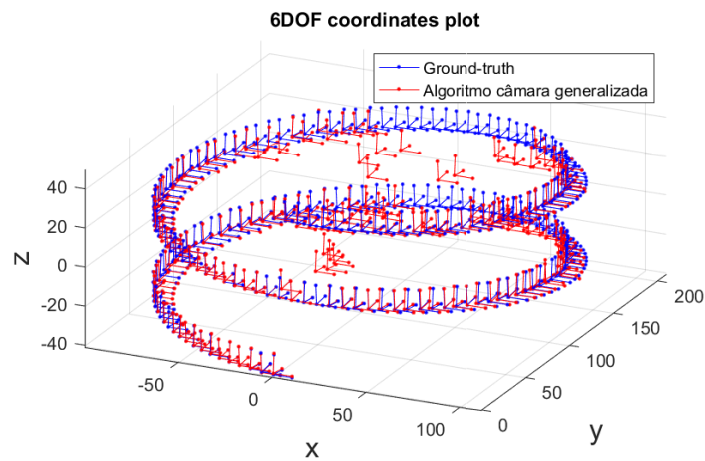


(b)

Figura 4.4: Na figura (a) está representada a estimativa do movimento realizada pelo algoritmo 3.2 para a trajetória 2D. Em cada imagem gerada na trajetória foi inserido um ruído de $\sigma = 0.005$. Na figura (b) encontra-se a estimação do movimento através de 3.3, tendo também sido um ruído de $\sigma = 0.005$. É bastante visível o efeito do ruído na estimação do movimento presente nos dois algoritmos. Mas enquanto que em (a) parece existir um offset de erro, em (b) a estimação do movimento aparece mais sob a forma de outliers.



(a)



(b)

Figura 4.5: Na figura (a) está representada a estimativa do movimento realizada pelo algoritmo 3.2 para a trajetória 3D. Em cada imagem gerada na trajetória foi inserido um ruído de $\sigma = 0.005$. Na figura (b) encontra-se a estimação do movimento através de 3.3, tendo também sido sujeito a um ruído de $\sigma = 0.005$. É bastante visível o efeito do ruído na estimação do movimento presente nos dois algoritmos. Principalmente em (a) existe um grande erro na estimação do movimento. Já no caso de (b), à medida que aumentam os níveis de ruído surgem mais outliers.

Para medir quantitativamente os erros relativos à translação, incluindo a sua escala e à rotação foram usados os seguintes critérios:

- Erro na escala da translação: $e_t = \left| 1 - \frac{\|t\|}{\|\hat{t}\|} \right|$
- Diferença da direção na translação: $\delta_t = \cos^{-1} \left(\frac{t^T \cdot \hat{t}}{\|t\| \cdot \|\hat{t}\|} \right)$
- Erro na rotação: $e_R = \|R - \hat{R}\|$
- Diferença dos ângulos em R: $\delta_\theta = \cos^{-1} \left(\frac{\text{tr}(\hat{R}\hat{R}^T) - 1}{2} \right)$

Estes critérios de erro foram calculados com base em 100 testes para um movimento genérico (entre dois *frames*) e depois foi calculada a média aritmética simples $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ (neste caso n corresponde ao número de testes) para cada um destes critérios de erro. Na tabela 4.1 e nos histogramas das Figuras 4.6, 4.7 e 4.8 são apresentados os resultados. Para todos os testes foram usadas 100 correspondências.

Algoritmo <i>Real Time 6D</i>			
	$\sigma=0$	$\sigma=0.005$	$\sigma=0.008$
$\bar{x}(e_t)$	0.0376	4.2541	18.6981
$\bar{x}(\delta_t)$	0.0171	1.3971	0.5659
$\bar{x}(e_R)$	2.4641e-05	0.0816	0.0580
$\bar{x}(\delta_\theta)$	2.4641e-05	0.0816	0.0580
Algoritmo dos 14 pontos			
	$\sigma=0$	$\sigma=0.005$	$\sigma=0.008$
$\bar{x}(e_t)$	5.1070e-15	3.2537	11.7814
$\bar{x}(\delta_t)$	0	0.0666	0.1940
$\bar{x}(e_R)$	7.9776e-16	0.0040	0.0081
$\bar{x}(\delta_\theta)$	2.9802e-08	0.0040	0.0081

Tabela 4.1: Médias dos erros para um movimento genérico testado 100 vezes usando o algoritmo *Real Time 6D* e o algoritmo dos 14 pontos.

Nos gráficos da Figura 4.9 são apresentadas as curvas dos erros em função de vários níveis de ruído.

Observando os resultados da Tabela 4.1, dos histogramas das Figuras 4.6, 4.7 e 4.8 e dos gráficos da Figura 4.9, verifica-se que os erros de estimação do movimento são mais elevados no caso do algoritmo *Real Time 6D*, principalmente o erro da estimação do fator de escala. Isto pode ser explicado pelo facto de no algoritmo *Real Time 6D* se calcular o movimento para cada câmara a menos de um fator de escala e só ser usada a ligação rígida entre as câmaras no cálculo do fator de escala. Enquanto que no caso da câmara generalizada o movimento é estimado logo à partida tendo

em conta as restrições geométricas das câmaras e por isso a estimação do movimento do sistema *stereo* é feita de forma mais robusta.

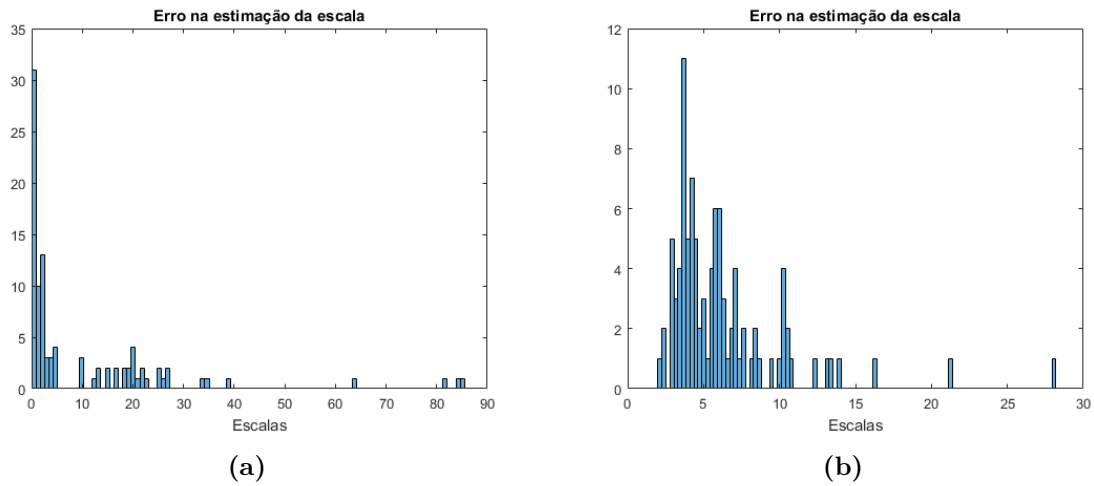


Figura 4.6: Histogramas do erro na estimação da escala baseados em 100 testes de um movimento genérico. Em todos os testes foi introduzido um ruído com $\sigma=0.005$ nos pixels. Em (a) estão os resultados para o algoritmo Real Time 6D e em (b) para o algoritmo dos 14 pontos.

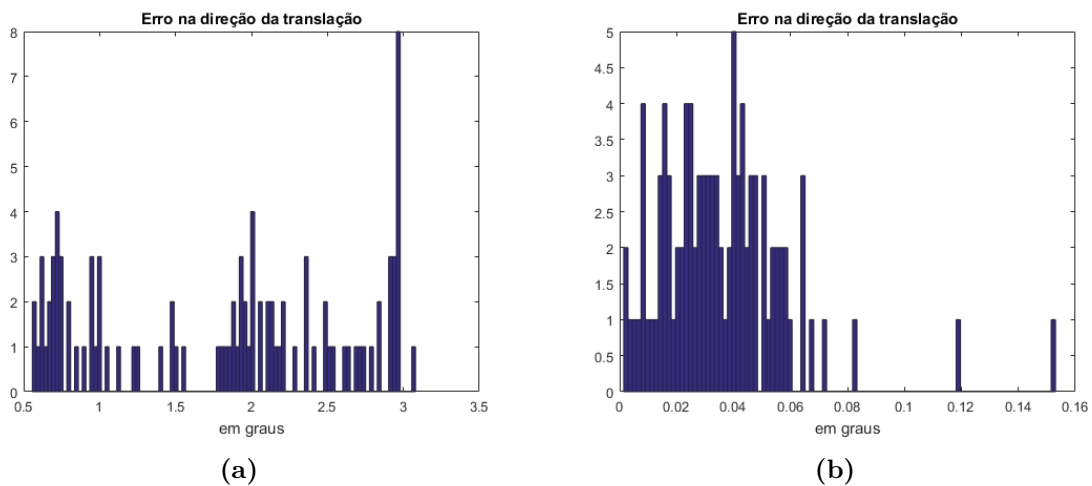


Figura 4.7: Histogramas do erro na direção da translação baseados em 100 testes de um movimento genérico. Em todos os testes foi introduzido um ruído com $\sigma=0.005$ nos pixels. Em (a) estão os resultados para o algoritmo Real Time 6D e em (b) para o algoritmo dos 14 pontos.

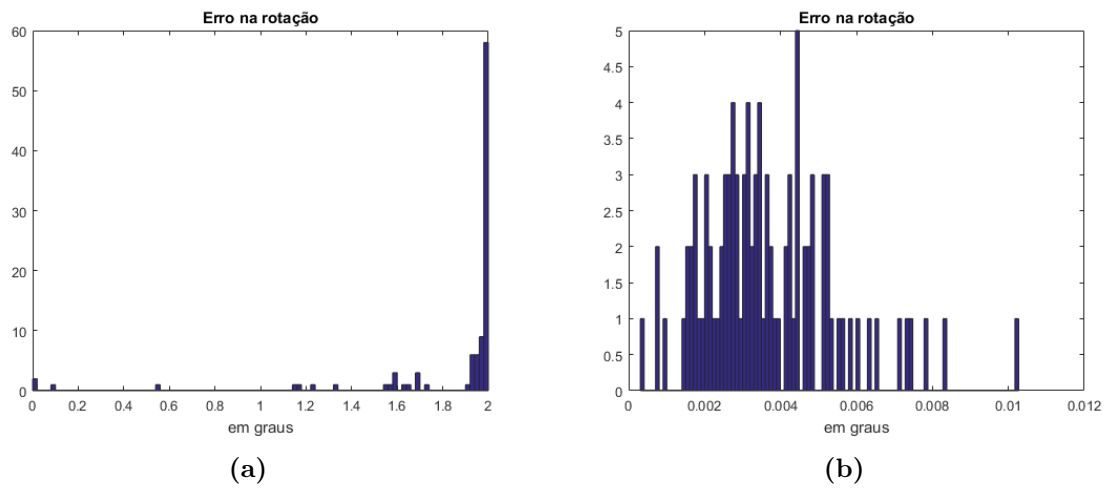


Figura 4.8: Histogramas do erro na rotação baseados em 100 testes de um movimento genérico. Em todos os testes foi introduzido um ruído com $\sigma=0.005$ nos pixels. Em (a) estão os resultados para o algoritmo Real Time 6D e em (b) para o algoritmo dos 14 pontos.

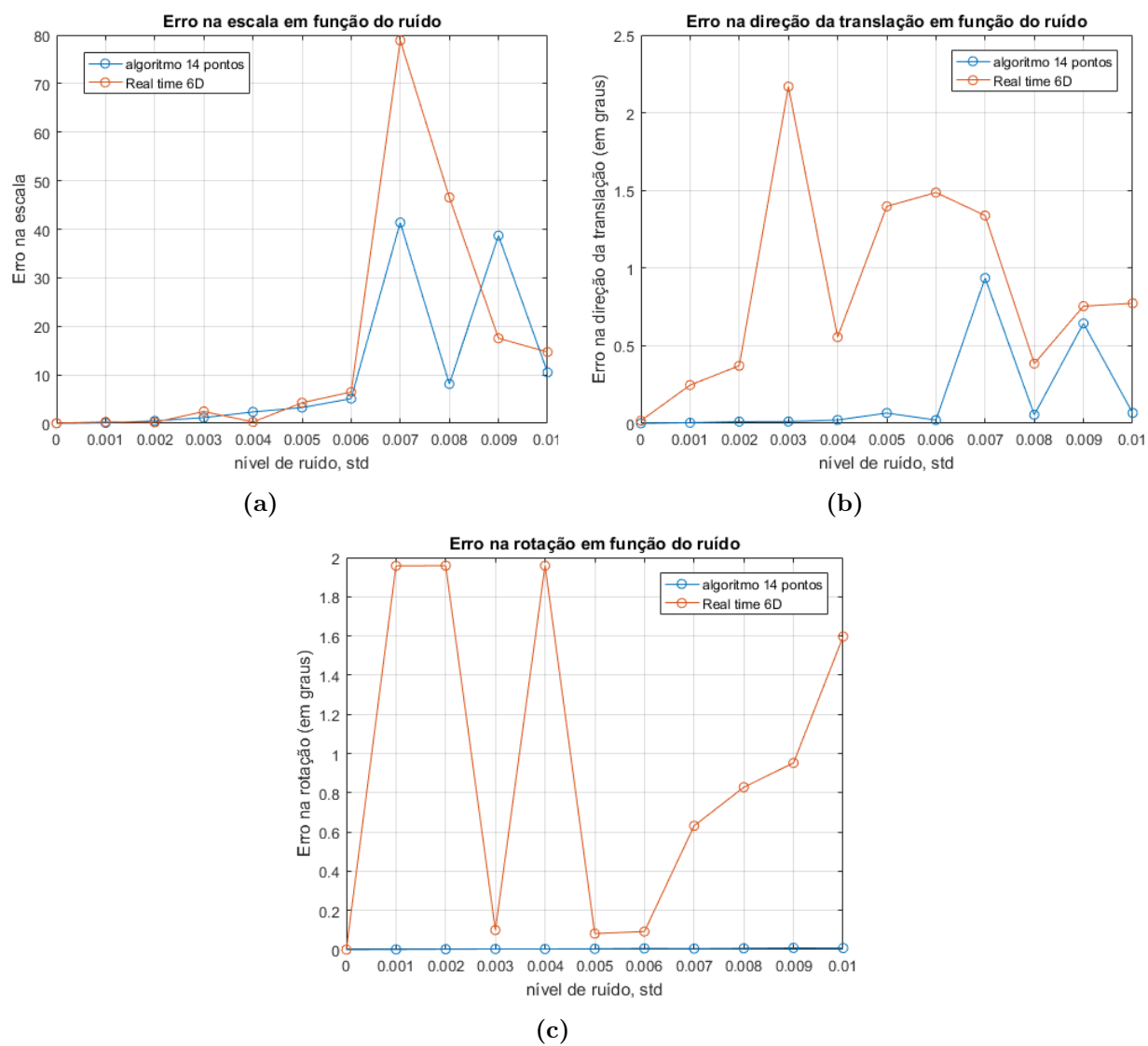


Figura 4.9: Gráficos onde estão representadas as curvas dos erros em função do nível de ruído. Em (a) tem-se a curva de erro na escala. Em (b), o erro na direção da translação. Por último, em (c) a curva do erro na rotação.

CAPÍTULO

5

CONCLUSÕES E TRABALHO FUTURO

Conteúdos

5.1	Conclusões	57
5.2	Trabalho futuro	58

5.1 Conclusões

Primeiramente nesta dissertação foi contextualizada a necessidade implementar métodos que consigam estimar a localização de cápsulas endoscópicas através de odometria visual. Foram apresentadas as características da cápsula em estudo, a PillCam Colon 2 e foram mostrados os resultados da calibração feita às câmaras, bem como o modelo utilizado. Os parâmetros intrínsecos obtidos na calibração não foram muito precisos apresentando uma variância significativa entre *datasets*, especialmente no cálculo da distância focal. Foi ainda possível perceber através da calibração que as imagens das câmaras sofrem muito o efeito das distorções, especialmente a radial.

Nesta dissertação também foram explicados detalhadamente os algoritmos implementados que permitem estimar o movimento de um sistema de câmaras semelhante ao da cápsula. Foi realizada também uma análise aos espaços nulos da matriz A na forma decomposta de maneira a estudar a influência do tipo e do número de correspondências no *rank* da matriz A , tendo-se verificado que usando correspondências cruzadas o *rank* de A mantém-se 14 e que são precisas pelo menos 6 correspondências de uma segunda câmara para o algoritmo não degenerar.

Para analisar a eficácia destes métodos e a sua robustez aquando a inserção de ruído foram realizados vários testes em ambiente simulado com trajetórias 2D e 3D. Os resultados apresentados demonstram que os algoritmos sem ruído funcionam com uma eficácia muito elevada, no entanto aquando a inserção de ruído, mesmo tendo este um desvio padrão muito baixo revelam uma elevada sensibilidade deixando de estimar com eficácia o movimento.

Apesar dos resultados da calibração não serem muito consistentes (não foi possível realizar mais capturas de imagens de modo a melhorar a calibração) e do efeito significativo do ruído nas estimativas da translação e da rotação, no global foram cumpridos os objetivos propostos para a dissertação, da qual foi possível adquirir conhecimentos em visão por computador de uma forma abrangente e profunda, contribuindo também para a análise da viabilidade da utilização de múltiplas câmaras nas cápsulas para a sua localização.

5.2 Trabalho futuro

Esta dissertação abre caminhos para futuros trabalhos visto que para ser possível realmente a estimação do movimento da cápsula PillCam Colon 2 várias melhorias têm de ser feitas tanto a nível da calibração das câmaras como dos algoritmos de estimação de movimento.

Relativamente ao modelo de calibração seria importante desenvolver um modelo de calibração polinomial específico para as câmaras. O modelo que foi usado nesta dissertação não tem em conta características próprias das câmaras como o efeito das cúpulas de plástico envolto nas câmaras, a elevada distorção radial existente, a má qualidade das imagens, entre outras. Um modelo de calibração mais apropriado permitiria uma estimação com maior rigor dos parâmetros intrínsecos da câmara.

Seria importante fazer também uma nova captura de imagens com um *setup* apropriado de modo a se conseguirem extrair os parâmetros extrínsecos das câmaras, que serão necessários caso se usem os algoritmos para estimar o movimento da cápsula.

Uma vez que se verificou que ambos os algoritmos apresentam muita sensibilidade ao ruído, seria importante tornar estes algoritmos mais robustos combinando-os com métodos de estimação não lineares tais como o *Bundle-Adjustment*.

CAPÍTULO

6

BIBLIOGRAFIA

- [1] C. D. Fletcher, *Diagnostic histopathology of tumors: 2-volume set with cd-roms*. Elsevier Health Sciences, 2007.
- [2] T. D. Than, G. Alici, H. Zhou, and W. Li, “A Review of Localization Systems for Robotic Endoscopic Capsules,” *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, vol. 59, no. 9, pp. 2387–2399, 2012.
- [3] Upper gi endoscopy. [Online]. Available: www.medbroadcast.com/procedure/getprocedure/Upper-GI-Endoscopy
- [4] Pillcam colon animation, abnormal. [Online]. Available: <https://youtu.be/PpAsxVME8a8>
- [5] H. M. Kim, S. Yang, J. Kim, S. Park, J. H. Cho, J. Y. Park, T. S. Kim, E.-S. Yoon, S. Y. Song, and S. Bang, “Active locomotion of a paddling-based capsule endoscope in an in vitro and in vivo experiment (with videos),” *Gastrointestinal endoscopy*, vol. 72, no. 2, pp. 381–387, 2010.
- [6] S. Gorini, M. Quirini, A. Menciassi, G. Pernorio, C. Stefanini, and P. Dario, “A novel sma-based actuator for a legged endoscopic capsule,” in *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*. IEEE, 2006, pp. 443–449.

-
- [7] C. Quaglia, E. Buselli, R. J. Webster III, P. Valdastrì, A. Menciassi, and P. Dario, "An endoscopic capsule robot: a meso-scale engineering case study," *Journal of Micromechanics and Microengineering*, vol. 19, no. 10, p. 105007, 2009.
- [8] J.-F. Rey, H. Ogata, N. Hosoe, K. Ohtsuka, N. Ogata, K. Ikeda, H. Aihara, I. Pangtay, T. Hibi, S. Kudo *et al.*, "Feasibility of stomach exploration with a guided capsule endoscope," *Endoscopy*, vol. 42, no. 07, pp. 541–545, 2010.
- [9] P. Swain, C. A. Mosse, F. Volke, J. Gerber, and J. Keller, "511i: In vivo studies of the potential and limitations of remote control of functional wireless capsule endoscopes with rare earth magnetic inclusions in an extra-corporeal magnetic field," *Gastrointestinal Endoscopy*, vol. 71, no. 5, p. AB123, 2010.
- [10] F. Carpi, N. Kastelein, M. Talcott, and C. Pappone, "Magnetically controllable gastrointestinal steering of video capsules," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 231–234, 2011.
- [11] G. Tortora, F. Mulana, G. Ciuti, P. Dario, and A. Menciassi, "Inductive-based wireless power recharging system for an innovative endoscopic capsule," *Energies*, vol. 8, no. 9, pp. 10315–10334, 2015.
- [12] M. R. Basar, M. Y. Ahmad, J. Cho, and F. Ibrahim, "Performance evaluation of power transmission coils for powering endoscopic wireless capsules," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, 2015, pp. 2263–2266.
- [13] K. M. Thotahewa, J.-M. Redouté, and M. R. Yuce, "A uwb wireless capsule endoscopy device," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 6977–6980.
- [14] E. Atashpanjeh and A. Pirhadi, "Design of wideband monopole antenna loaded with small spiral for using in wireless capsule endoscopy systems," *Progress In Electromagnetics Research C*, vol. 59, pp. 71–78, 2015.
- [15] Y. Shi, G. Yan, B. Zhu, and G. Liu, "A portable wireless power transmission system for video capsule endoscopes," *Bio-medical materials and engineering*, vol. 26, no. s1, pp. S1721–S1730, 2015.

-
- [16] D. Filip, "Self-stabilizing capsule endoscope for early detection and biopsy of colonic polyps," Ph.D. dissertation, University of Calgary, 2013.
- [17] M.-J. Sheu, C.-W. Chiang, W.-S. Sun, J.-J. Wang, and J.-W. Pan, "Dual view capsule endoscopic lens design," *Optics express*, vol. 23, no. 7, pp. 8565–8575, 2015.
- [18] R. Shrestha, X. Zhang, Z. Gias, and K. Wahid, "Adaptive illumination in wireless capsule endoscopy system," in *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 778–781.
- [19] A. Karargyris and N. Bourbakis, "Three-dimensional reconstruction of the digestive wall in capsule endoscopy videos using elastic video interpolation," *IEEE Transactions on Medical Imaging*, vol. 30, no. 4, pp. 957–971, April 2011.
- [20] Q. Zhao and M. Q. H. Meng, "3d reconstruction of gi tract texture surface using capsule endoscopy images," in *2012 IEEE International Conference on Automation and Logistics*, Aug 2012, pp. 277–282.
- [21] B. Li and M. Q.-H. Meng, "Automatic polyp detection for wireless capsule endoscopy images," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10 952 – 10 958, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S095741741200499X>
- [22] J. L. Gonzalez-Guillaumin, D. C. Sadowski, K. V. I. S. Kaler, and M. P. Mintchev, "Ingestible capsule for impedance and ph monitoring in the esophagus," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2231–2236, Dec 2007.
- [23] K. Kong, S. Yim, S. Choi, and D. Jeon, "A robotic biopsy device for capsule endoscopy," *Journal of Medical Devices*, vol. 6, no. 3, p. 031004, 2012.
- [24] S. Yim, E. Gultepe, D. H. Gracias, and M. Sitti, "Biopsy using a magnetic capsule endoscope carrying, releasing, and retrieving untethered microgrippers," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 513–521, 2014.
- [25] S. P. Woods and T. G. Constandinou, "Wireless capsule endoscope for targeted drug delivery: mechanics and design considerations," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 945–953, 2013.
- [26] D. K. Iakovidis, E. Spyrou, D. Diamantis, and I. Tsiompanidis, "Capsule endoscope localization based on visual features," in *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 1–4.
-

-
- [27] G. Ciuti, R. Calì, D. Camboni, L. Neri, F. Bianchi, A. Arezzo, A. Koulaouzidis, S. Schostek, D. Stoyanov, C. Oddo *et al.*, “Frontiers of robotic endoscopic capsules: a review,” *Journal of Micro-Bio Robotics*, vol. 11, no. 1-4, pp. 1–18, 2016.
- [28] G. Bao, L. Mi, and K. Pahlavan, “Emulation on motion tracking of endoscopic capsule inside small intestine,” in *Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP)*. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2013, p. 1.
- [29] G. Bao, L. Mi, Y. Geng, M. Zhou, and K. Pahlavan, “A video-based speed estimation technique for localizing the wireless capsule endoscope inside gastrointestinal tract,” in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 5615–5618.
- [30] K. Wahid, S. L. Kabir, H. A. Khan, A. Al Helal, M. Mukit, and R. Mostafa, “A localization algorithm for capsule endoscopy based on feature point tracking,” in *Medical Engineering, Health Informatics and Technology (MediTec), 2016 International Conference on*. IEEE, 2016, pp. 1–5.
- [31] T. Nakamura and A. Terano, “Capsule endoscopy: past, present, and future,” *Journal of gastroenterology*, vol. 43, no. 2, pp. 93–99, 2008.
- [32] K. Triantafyllou, I. S. Papanikolaou, K. Papaxoinis, and S. D. Ladas, “Two cameras detect more lesions in the small-bowel than one,” *World journal of gastroenterology: WJG*, vol. 17, no. 11, p. 1462, 2011.
- [33] M. Keuchel, F. Hagenmüller, and H. Tajiri, *Video capsule endoscopy: a reference guide and atlas*. Springer, 2015.
- [34] J. Gerber, A. Bergwerk, and D. Fleischer, “A capsule endoscopy guide for the practicing clinician: technology and troubleshooting,” *Gastrointestinal endoscopy*, vol. 66, no. 6, pp. 1188–1195, 2007.
- [35] Z. Li, D. Carter, R. Eliakim, W. Zou, H. Wu, Z. Liao, Z. Gong, J. Wang, J. W. Chung, S. Y. Song *et al.*, “The current main types of capsule endoscopy,” in *Handbook of Capsule Endoscopy*. Springer, 2014, pp. 5–45.
- [36] *PillCam Capsule Endoscopy, User Manual RAPID v8.0, 2003*.
-

-
- [37] G. Ciuti, R. Donlin, P. Valdastrì, A. Arezzo, A. Menciassi, M. Morino, and P. Dario, “Robotic versus manual control in magnetic steering of an endoscopic capsule,” *Endoscopy*, vol. 42, no. 02, pp. 148–152, 2010.
- [38] G. Ciuti, P. Valdastrì, A. Menciassi, and P. Dario, “Robotic magnetic steering and locomotion of capsule endoscope for diagnostic and surgical endoluminal procedures,” *Robotica*, vol. 28, no. 2, pp. 199–207, 2010.
- [39] C. Di Natali, M. Beccani, N. Simaan, and P. Valdastrì, “Jacobian-based iterative method for magnetic localization in robotic capsule endoscopy,” *IEEE Transactions on Robotics*, vol. 32, no. 2, pp. 327–338, 2016.
- [40] D. Fischer, R. Schreiber, D. Levi, and R. Eliakim, “Capsule endoscopy: the localization system,” *Gastrointestinal Endoscopy Clinics*, vol. 14, no. 1, pp. 25–31, 2004.
- [41] I. Umay, B. Fidan, and B. Barshan, “Localization and tracking of implantable biomedical sensors,” *Sensors*, vol. 17, no. 3, p. 583, 2017.
- [42] B. Steinberg and I. Bettesh, “Localization of capsule with a synthetic source of quadrupoles and dipoles,” Apr. 16 2013, uS Patent 8,423,122. [Online]. Available: <https://www.google.com/patents/US8423122>
- [43] H. Shim, J. Hwang, K. Kim, Y. Seo, B. KIM, Y. Lee, C. Cha, and B. Baek, “Endoscope and a method for finding its location,” Jul. 1 2010, uS Patent App. 12/442,836. [Online]. Available: <https://www.google.com/patents/US20100168517>
- [44] C. Hu, M. Li, S. Song, R. Zhang, M. Q.-H. Meng *et al.*, “A cubic 3-axis magnetic sensor array for wirelessly tracking magnet position and orientation,” *IEEE Sensors Journal*, vol. 10, no. 5, pp. 903–913, 2010.
- [45] W. Weitschies, R. Kötitz, D. Cordini, and L. Trahms, “High-resolution monitoring of the gastrointestinal transit of a magnetically marked capsule,” *Journal of pharmaceutical sciences*, vol. 86, no. 11, pp. 1218–1222, 1997.
- [46] X. Wu, W. Hou, C. Peng, X. Zheng, X. Fang, and J. He, “Wearable magnetic locating and tracking system for mems medical capsule,” *Sensors and Actuators A: Physical*, vol. 141, no. 2, pp. 432–439, 2008.
-

-
- [47] A. Plotkin and E. Paperno, "3-d magnetic tracking of a single subminiature coil with a large 2-d array of uniaxial transmitters," *IEEE Transactions on Magnetics*, vol. 39, no. 5, pp. 3295–3297, 2003.
- [48] M. Salerno, G. Ciuti, G. Lucarini, R. Rizzo, P. Valdastri, A. Menciassi, A. Landi, and P. Dario, "A discrete-time localization method for capsule endoscopy based on on-board magnetic sensing," *Measurement Science and Technology*, vol. 23, no. 1, p. 015701, 2011.
- [49] L. Liu, C. Hu, W. Cai, and M. Q.-H. Meng, "Capsule endoscope localization based on computer vision technique," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 3711–3714.
- [50] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 6, pp. 756–770, 2004.
- [51] E. Spyrou and D. K. Iakovidis, "Homography-based orientation estimation for capsule endoscope tracking," in *Imaging Systems and Techniques (IST), 2012 IEEE International Conference on*. IEEE, 2012, pp. 101–105.
- [52] Camera calibration toolbox. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/
- [53] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [54] "what is camera calibration?". [Online]. Available: <https://www.mathworks.com/help/vision/ug/camera-calibration.html>
- [55] C. B. Duane, "Close-range camera calibration," *Photogramm. Eng*, vol. 37, no. 8, pp. 855–866, 1971.
- [56] J. Weng, P. Cohen, M. Herniou *et al.*, "Camera calibration with distortion models and accuracy evaluation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 10, pp. 965–980, 1992.
- [57] C. Hu, M. Meng, P. X. Liu, and X. Wang, "Image distortion correction for wireless capsule endoscope," in *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, vol. 5. IEEE, 2004, pp. 4718–4723.
-

-
- [58] R. K. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys, “Simple calibration of non-overlapping cameras with a mirror,” in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–7.
- [59] D. Scaramuzza and F. Fraundorfer, “Visual odometry [tutorial],” *IEEE Robotics Automation Magazine*, vol. 18, no. 4, pp. 80–92, Dec 2011.
- [60] D. Nister, O. Naroditsky, and J. Bergen, “Visual odometry,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1, June 2004, pp. I–652–I–659 Vol.1.
- [61] B. Clipp, J. H. Kim, J. M. Frahm, M. Pollefeys, and R. Hartley, “Robust 6dof motion estimation for non-overlapping, multi-camera systems,” in *2008 IEEE Workshop on Applications of Computer Vision*, Jan 2008, pp. 1–8.
- [62] J.-H. Kim, R. Hartley, J.-M. Frahm, and M. Pollefeys, “Visual odometry for non-overlapping views using second-order cone programming,” in *Asian Conference on Computer Vision*. Springer, 2007, pp. 353–362.
- [63] J. H. Kim, H. Li, and R. Hartley, “Motion estimation for nonoverlapping multicamera rigs: Linear algebraic and l₁ geometric solutions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1044–1059, June 2010.
- [64] R. Pless, “Using many cameras as one,” in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, June 2003, pp. II–587–93 vol.2.
- [65] H. Li, R. Hartley, and J. hak Kim, “A linear approach to motion estimation using generalized camera models,” in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.
- [66] M. HenrikStewénius, K. Aström, and D. Nistér, “Solutions to minimal generalized relative pose problems,” 2005.
- [67] F. Vasconcelos and J. P. Barreto, “Towards a minimal solution for the relative pose between axial cameras,” *Int. J. Comput. Vision*, vol. 84, no. 3, pp. 237–256, 2009.
- [68] L. Kneip and H. Li, “Efficient computation of relative pose for multi-camera systems,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 446–453.
-

- [69] T. Kazik, L. Kneip, J. Nikolic, M. Pollefeys, and R. Siegwart, “Real-time 6d stereo visual odometry with non-overlapping fields of view,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 1529–1536.
- [70] Z. Kukelova, M. Bujnak, and T. Pajdla, “Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems.” in *BMVC*, vol. 2, no. 5, 2008, p. 2008.
- [71] R. I. Hartley and P. Sturm, “Triangulation,” *Computer vision and image understanding*, vol. 68, no. 2, pp. 146–157, 1997.
- [72] M. D. Grossberg and S. K. Nayar, “A general imaging model and a method for finding its parameters,” in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 108–115.
- [73] J.-S. Kim and T. Kanade, “Degeneracy of the linear seventeen-point algorithm for generalized essential matrix,” vol. 37, pp. 40–48, 05 2010.