

INSTITUTO SUPERIOR DE ENGENHARIA DO
PORTO



Calibração automática para sistema de localização
Multi-câmara

André Ferreira Faria 1090352

Dissertação submetida no âmbito do
Mestrado em Engenharia Eletrotécnica de computadores

Orientação Científica: José Miguel Soares de Almeida
Co-orientação: Alfredo Manuel Oliveira Martins

11 de Novembro de 2014

Resumo

Os sistemas autónomos, nomeadamente sistemas robóticos móveis, estão cada vez mais presentes no nosso quotidiano, desta forma surge a necessidade de colocar em prática novas técnicas capazes de suportar o seu desenvolvimento. Nos dias que correm são frequentemente utilizados sistemas de *ground truth* baseados em sistemas multi-câmara como técnica de apoio ao desenvolvimento de sistemas autónomos.

Nesta dissertação são abordadas técnicas de otimização do processo de calibração dos parâmetros extrínsecos em sistemas multi-câmara, sem que, para tal, seja necessária a aplicação de técnicas exaustivas, demoradas e passíveis à introdução de erros. Desta forma esta dissertação propõe o desenvolvimento de uma ferramenta para calibração dos parâmetros extrínsecos de sistema de *ground truth* com base em diferentes abordagens para aplicações de em diferentes cenários.

Uma das abordagens utilizadas visa a utilização do método de estimação dos pontos de fuga. Este destina-se a ser aplicado em cenários estruturados em que sejam predominantes linhas paralelas entre si e também com o referencial global. Foi também abordada a utilização do método de decomposição de homografia planar, este tem como alvo os cenários de aplicação em que seja predominante pelo menos um plano. No presente documento são demonstrados detalhadamente o desenvolvimento dos métodos de calibração das câmaras, bem como a implementação efetuada para a construção do *setup* experimental necessário para a obtenção de resultados para a validação

ii

e caracterização dos métodos implementados.

Abstract

The autonomous systems, especially mobile robotic systems are ever more present in our daily lives, so there is a need to put into practice new techniques to support their development. These days multi-camera ground truth systems are often used to in order to achieve this purpose.

This dissertation addresses optimization of calibration techniques for multi-camera systems without the need to apply exhaustive techniques, with high time consumption, liable to introduce systematic errors.

This thesis proposes the development of an automatic tool for extrinsic parameters calibration of ground-truth multi-camera systems to overall applications in such different scenarios. One of the used approaches aims the estimation of vanishing points. This is intended to be applied to scenarios with structured lines parallel with the global framework. It was also addressed the usage of the planar homography decomposition method, this is liable to be applied in environments predominant planes on it. In this document are shown in detail the development of the cameras calibration methods and the implementation made for the construction of experimental setup necessary to obtain results for the validation and characterization of implemented.

Esta página foi intencionalmente deixada em branco.

Conteúdo

Lista de Acrónimos	xvi
1 Introdução	1
1.1 Enquadramento e Motivação	2
1.2 Objetivos	4
1.3 organização da dissertação	5
2 Estado da Arte	7
2.1 Sistemas de deteção e localização: comerciais	7
2.1.1 Vicon	7
2.1.2 Optotrack	11
2.2 Soluções <i>open source</i> de estimação de posição de objetos . . .	12
2.2.1 SSL-Vision	12
2.2.2 <i>Kinect ground truth</i>	14
2.3 Calibração de sistemas de visão multi-câmara	17
2.3.1 Algoritmo de Tsai	18
2.3.2 Auto-calibração de sistema multi-câmara	19
2.3.3 Homografia baseada pontos de fuga	20
2.4 Sistema de localização em ambientes subaquáticos	23
2.4.1 Ambiente subaquáticos – USBL	24
2.4.2 Ambientes Exteriores – GNSS	25

3	Fundamentos Teóricos	29
3.1	Transformações de corpo rígido	29
3.1.1	Coordenadas Homogêneas	32
3.2	Modelo da câmara	33
3.2.1	Parâmetros intrínsecos	35
3.2.2	Parâmetros extrínsecos	37
3.2.3	Projeção do espaço 3D para 2D	38
3.3	Geometria epipolar	39
3.3.1	Matriz Essencial	39
3.3.2	Matriz fundamental	42
3.4	Homografia planar	42
3.5	Decomposição de Homografia	43
3.6	Transformada de Hough	45
3.7	Linhas e pontos de fuga	47
4	Setup experimental	49
4.1	Arquitetura do sistema	49
4.1.1	<i>Ground truth</i> cenário MSL	51
4.1.2	<i>Ground truth</i> Tanque	51
4.2	<i>Hardware</i>	52
4.2.1	Câmara	52
4.2.2	Sistema de alimentação e disparo das câmaras	53
5	Implementação	57
5.1	Instalação do sistema	57
5.1.1	Campo de futebol robótico médio	57
5.1.2	Tanque de experiências subaquáticas	59
5.2	Calibração dos parâmetros intrínsecos	60
5.2.1	Algoritmo de Tsai	60
5.3	Calibração dos Parâmetros extrínsecos	62

<i>CONTEÚDO</i>	vii
5.3.1 Método dos 8 pontos	63
5.3.2 Método pontos de fuga	64
5.3.3 Decomposição de homografia	70
6 Resultados	73
6.1 resultados para o método 8 Pontos	75
6.2 Resultados para o método dos pontos de fuga	77
6.3 Resultados para o método de decomposição de homografia . .	85
6.4 Comparação entre métodos	89
7 Conclusões	97

Esta página foi intencionalmente deixada em branco.

Lista de Figuras

2.1	Exemplo de utilização de VICON tracker.	9
2.2	Exemplo de produção de animação com Vicon Pegasus.	10
2.3	Camera Vicon Bonita B10.	11
2.4	Ferramenta de calibração Vicon.	11
2.5	Sistema Optotrack.	12
2.6	Calibração de <i>kinect</i> para aplicação em SPL.	15
2.7	Algoritmo de Tsai.	18
2.8	Representação da transformação de homografia	23
2.9	Utilização de USBL.	25
2.10	Constituintes de um sistema de USBL.	26
2.11	Ilustração de funcionamento de GPS.	27
3.1	Representação dos Ângulos de Euler.	30
3.2	ambiguidade gimbal	32
3.3	Comparação modelo <i>pinhole</i> - olho humano [25].	34
3.4	Projeção do modelo de <i>pinhole</i>	34
3.5	Comparação modelo <i>pinhole</i> - olho humano.	37
3.6	Representação da restrição epipolar.	40
3.7	Representação do espaço 3D para aplicação de homografia.	44
3.8	Representação do plano da transformada de Hough - $\rho - \theta$	47
3.9	Representação de retas na transformada de Hough.	47
3.10	Representação de linhas de fuga.	48

4.1	Arquitetura de aquisição do sistema de <i>ground truth</i>	50
4.2	<i>Ground truth</i> Campo MSL.	51
4.3	Sistema <i>Ground truth</i> tanque.	52
4.4	Arquitetura do <i>ground truth</i> implementado.	53
4.5	Arquitetura do sistema de disparo.	54
4.6	Interface de configuração do <i>trigger</i>	55
5.1	Fixação câmaras campo MSL.	58
5.2	Instalação do Sistema de <i>Ground truth</i> no Tanque	59
5.3	Instalação do Sistema de disparo no Tanque	60
5.4	Deteção de pontos de interesse - algoritmo de Tsai.	61
5.5	Deteção de pontos de interesse - algoritmo de Tsai.	61
5.6	Referencial do campo de MSL.	63
5.7	Aplicação da Transformada de hough	65
5.8	Refinamento da transformada de Hough	67
6.1	Pontos na imagem.	75
6.2	Erro de reprojeção câmara esquerda pelo algoritmo 8 pontos.	76
6.3	Erro de reprojeção para a câmara direita pelo algoritmo 8 pontos.	76
6.4	Erro de triangulação algoritmo 8 pontos.	77
6.5	Erro de reprojeção câmara esquerda.	78
6.6	Dispersão erro câmara esquerda.	79
6.7	Erro de reprojeção em x e y câmara esquerda.	79
6.8	Norma do erro de reprojeção esquerda.	80
6.9	Reprojeção câmara direita.	80
6.10	Dispersão do erro câmara direita.	81
6.11	Erro de de reprojeção em x e y para a câmara direita.	81
6.12	Norma do erro de reprojeção para a câmara direita.	82
6.13	Dispersão do erro de triangulação stereo.	82

6.14	Erro de triangulação stereo, componente x.	83
6.15	Erro de triangulação stereo, componente y.	83
6.16	Erro de triangulação stereo, componente z.	84
6.17	Norma erro de triangulação stereo.	84
6.18	representação dos pontos 3D.	85
6.19	Erro de reprojeção câmara esquerda.	87
6.20	Erro de reprojeção câmara direita.	88
6.21	Erro de triangulação stereo.	89
6.22	Erro de triangulação stereo médio.	90
6.23	Erro calibração - pontos fuga - rotação câmara esquerda, <i>yaw</i> <i>pitch roll</i>	91
6.24	Erro calibração - pontos fuga - rotação câmara direita, <i>yaw</i> <i>roll pitch</i>	91
6.25	Erro calibração - pontos fuga - translação câmara esquerda, T_x, T_y e T_z	92
6.26	Erro calibração - pontos fuga - translação câmara direita, , T_x, T_y e T_z	92
6.27	Erro calibração - homografia - rotação câmara esquerda, <i>yaw</i> <i>roll pitch</i>	93
6.28	Erro calibração - homografia - rotação câmara direita, <i>yaw</i> <i>pitch roll</i>	94
6.29	Erro calibração - homografia - translação câmara esquerda, T_x, T_y e T_z	94
6.30	Erro calibração - homografia - translação câmara direita, T_x , T_y e T_z	95

Esta página foi intencionalmente deixada em branco.

Lista de Tabelas

3.1	Soluções obtidas pela decomposição de homografia	45
6.1	Pontos mundo.	74
6.2	Comparação de parâmetros extrínsecos	86
6.3	Comparação de parâmetros extrínsecos	88

Esta página foi intencionalmente deixada em branco.

Lista de Acrónimos

LSA Laboratório de Sistemas Autónomos

ISEP Instituto Superior de Engenharia do Porto

ROS *Robot Operating system*

XML *eXtensible Markup Language*

OpenCV *Open source Computer Vision library*

FPS *Frames Per Second*

MSL *Middle Size League*

USLB *Ultra Short BaseLine*

GPS *Global Position System*

MOCAP *MOtion CAPture*

POE *Power Over Ethernet*

PCL *Point Cloud Library*

SVD *Singular Value Decomposition*

PPS *Pulse per second*

SPL *Standard Platform League*

MSL *Middle Size League*

FOV *Field Of View*

Capítulo 1

Introdução

Esta dissertação surge no âmbito do Mestrado em Engenharia Eletrotécnica e Computadores, ramo de Sistemas autónomos do Instituto Superior de Engenharia do Porto (ISEP), enquadrando-se nas atividades de I&D do Laboratório de Sistemas Autónomos em robótica móvel.

A localização simultânea de múltiplos alvos, tais como veículos robotizados, em movimento tem-se revelado um problema com o qual a comunidade robótica, à escala mundial, se tem deparado desde os primórdios do desenvolvimento de sistemas Autónomos.

A monitorização de trajetórias robôs e outros objetos de interesse relativos ao cenário de operação tem-se demonstrado uma estratégia crucial para o desenvolvimento de técnicas de navegação e avaliação das mesmas. Ultimamente é recorrente a utilização de câmaras de espectro visível para dar resposta a esta necessidade. Através da deteção dos objetos de interesse tais como veículos autónomos ou outros objetos do meio envolvente. A utilização de técnicas multi-câmara pode, então, demonstrar-se uma técnica vantajosa, com níveis de precisão/exatidão na localização de objetos bastante atrativos.

Este trabalho visa o desenvolvimento de técnicas de calibração para sistemas de visão computacional multi-câmara, para localização e seguimento de múltiplos veículos robóticos ou outros objetos de interesse. Este destinam-

se a ser aplicados em cenários de operação multi-robóticos. Assim sendo é pretendido que a calibração do sistema seja efetuada de uma forma eficiente automatizada.

1.1 Enquadramento e Motivação

Ao longo dos últimos anos o LSA/ISEP tem vindo a desenvolver inúmeros sistemas robóticos para atuação em diversos cenários tais como, aéreos[31], marítimos [32] e terrestres[29]. Desta forma, surgiu a necessidade da criação de um sistema para localização e seguimento de múltiplos robôs ou outros objetos que sejam parte do cenário de operação. A solução desenvolvida irá fornecer informação externa de posicionamento que pode ser utilizada para aferir da qualidade de algoritmos de navegação ou para avaliar de forma independente o comportamento dos sistemas robóticos.

A utilização de sistemas multi-câmara para *tracking* de veículos autónomos predomina sobre outros devido à grande quantidade de informação capturada do meio no qual estão inseridos. Com os progressos feitos nos últimos anos relativos ao desenvolvimento do poder computacional a aplicação de técnicas de visão artificial pode ser aplicada a um crescente numero de aplicações sem que para tal seja posto em causa a fidelidade do sistema.

Como solução ao problema abordado a utilização de sistemas de *ground-truth* baseados em visão tem-se revelado uma solução que apresenta resultados satisfatórios face a outros métodos de localização e seguimento de objetos.

No referido tipo de sistemas um dos grandes desafios do ponto de vista de precisão/exatidão dos mesmos assenta sobre processo de calibração do parâmetros extrínsecos do sistema. Este problema é foco principal do presente trabalho, nomeadamente no desenvolvimento, implementação, teste e validação de algoritmos que permitam a auto calibração de múltiplas câmaras posicionadas em locais distintos. Assim sendo, a solução desen-

volvida endereça este problema através do desenvolvimento de uma ferramenta de calibração automática para sistemas de visão computacional multi-câmara que torne versátil o processo de calibração do sistema para cada experiência. Esta deverá ser passível de utilização quer em cenários interiores ou exteriores, para aplicação em diferentes meios de operação tais como aéreos ou terrestres e até mesmo em cenários onde seja necessária a localização de objetos subaquáticos.

Esta calibração envolve vários problemas desde a calibração dos parâmetros intrínsecos de cada câmara à obtenção exata quer do posicionamento relativo entre as mesmas, quer do posicionamento de cada uma das câmaras face ao referencial global. Esta última questão, geralmente designada de obtenção de parâmetros extrínsecos, revela-se a mais problemática tendo em conta a precisão exigida assim como a complexidade para sua calibração em sistemas multi-câmara.

A automatização do processo de calibração do posicionamento relativo das câmaras face ao ambiente tendem a gerar soluções de elevada precisão com redução significativa do tempo de util para a calibração. De outra forma recorrendo a um processo manual, a qualidade/exatidão obtida está dependente do procedimento efetuado pelo utilizador, sendo que esta remete para que sejam cometidos erros sistemático.

Por forma a que seja possível a obtenção de resultados coerentes com outras fontes de informação que se pretendem analisar, em qualquer sistema com abordagem multi-sensorial levanta-se a questão de sincronismo entre os sensores. Uma das soluções encontradas para garantir o sincronismo da aquisição de imagem é a utilização de disparo por hardware. Este dispositivo tem como função principal dar a ordem de disparo das câmaras no mesmo instante temporal para que assim seja possível da posição dos objetos por triangulação. Os principais cenários para a aplicação desenvolvida estão presentes no laboratório de sistemas autónomos, sendo estes o campo de

futebol robótico de médio e o tanque de experiências para veículos aquáticos.

A utilização do sistema de *ground truth* no campo de futebol robótico médio [5] tem por objetivos a obtenção de resultados que permita melhorar aspetos como o controlo, ou a coordenação coletiva da equipa levando ao desenvolvimento de táticas de jogo coletivo. O campo de futebol robótico médio pode também ser cenário de experiências de outros veículos robóticos devido ao amplo espaço disponível. No campo de futebol robótico médio são disponibilizados para a realização de experiência cerca de 20 metros de comprimento, 15 metros de largura e mais de 7 metros de altura totalizando, assim, uma área de trabalho útil de cerca de 300 metros quadrados para a realização de experiências de veículos terrestres ou até mesmo aéreos.

Para o cenário do tanque, o sistema em questão é responsável pela monitorização de manobras efetuadas por veículos subaquáticos a fim de avaliar as trajetórias efetuadas por estes, como sistema externo de referência para realimentação da posição para controlo de trajetória do veículo. As experiências a desenvolver podem ser inúmeras sendo as dimensões do tanque de 11 metros de comprimento, 6 metros de largura e 5 metros de profundidade.

1.2 Objetivos

Nesta dissertação é abordado o problema de deteção e seguimento de múltiplos objetos em movimento de forma simultânea e em tempo real. A principal questão a resolver consiste na calibração de um sistema multi-câmara para cumprir estas funções. Deste modo, com o desenvolvimento desta dissertação os objetivos que se pretendem atingir são:

1. Desenvolver um conjunto de métodos de calibração, para um sistema com múltiplas câmaras, para integração em ferramenta de calibração automática, requerendo a menor interação possível por parte do utili-

- zador, de forma a funcionar nos diversos cenários de aplicação: interiores, exteriores, terrestres, subaquáticos ou aéreos;
2. Implementação de *Setup experimental* necessário para efetuar os testes e validação dos métodos de calibração desenvolvidos:
 - (a) Instalação e calibração dos parâmetros intrínsecos de múltiplas câmaras;
 - (b) Desenvolver software para a configuração e controlo dos parâmetros de aquisição das câmaras em *Robot Operating System* ROS;
 - (c) Desenvolver infra-estrutura de controlo de disparo e software para a aquisição sincronizada e o correto registo de imagens provenientes de múltiplas câmaras
 3. Projetar e efetuar um vetor de testes para a validação experimental, caracterização e comparação dos métodos desenvolvidos

1.3 organização da dissertação

No presente subcapítulo é apresentada a estrutura organizacional desta dissertação, a mesma está organizada em 7 capítulos. No capítulo 2 será apresentado o estado da arte com apresentadas algumas soluções para a calibração de sistemas multi-câmara. No capítulo 3 são apresentados os fundamentos teóricos que estão na base do desenvolvimento do presente documento. Seguidamente, pelo capítulo 4 contém o projeto do trabalho onde foi descrita a arquitetura do sistema. Pelo capítulo 5, é demonstrado a implementação efetuada, desde a instalação de hardware até à implementação da calibração das câmaras. De seguida pelo capítulo 6 são apresentados os resultados obtidos para a implementação dos diferentes métodos. Por ultimo, no capítulo 7 são apresentadas algumas conclusões sobre a implementação do sistema de *Ground truth*.

Esta página foi intencionalmente deixada em branco.

Capítulo 2

Estado da Arte

Para o desenvolvimento de um projeto de âmbito científico como o projeto desenvolvido com o objetivo desta dissertação, é fundamental o conhecimento prévio de técnicas e tecnologias para que seja encontrada uma solução otimizada para o desempenho da tarefa em questão. Desta forma, são apresentadas algumas das soluções existentes para localização e seguimento de objetos.

2.1 Sistemas de deteção e localização: comerciais

Existem algumas soluções para a para a deteção e localização de objetos em determinados cenários. Estes métodos são habitualmente designadas por *motion capture* (MOCAP).

2.1.1 Vicon

A Vicon [4] é uma empresa especializada na análise de movimento através de visão, que apresenta diversas soluções comerciais para o seguimento e a reconstrução 3D de objetos baseada em marcadores, passivos. As suas aplicações podem ser variadas desde a análise do desempenho de atletas de alta competição em atividades desportivas, por exemplo em atletismo; em

estudos clínicos e biomecânicos em centros de investigação como por exemplo na captura de movimento em 3D de estruturas como veículos autónomos. As soluções apresentadas pela Vicon são contextualizadas para diferentes aplicações específicas. Assim, as soluções disponíveis incluem: Tracker[49], Blade 3 [40], Nexus 2 [42], BodyBuilder [41], Polygonon 4 citePolygon4, Vicon Motus Vídeo [46], Vicon Bonita [19], entre outros. Estes sistemas tipicamente requerem a utilização de um elevado número de câmaras o que torna o sistema dispendioso do ponto de vista financeiro. Outro das desvantagens deste tipo de sistemas é o facto de recorrerem a marcadores de sinalização colocados nos objetos a localizar, tipicamente estes marcadores são passivos, refletores ou leds, infravermelhos, o que em alguns casos se revela uma interferência considerável ao cenário de operação dos robôs. Os sistemas Vicon são também aplicados à robótica como é o caso da " *The Flying Machine Arena* " [26], este projeto visa a implementação de um espaço de testes para veículos autónomos , terrestres e aéreos, com dimensões de 10×10×10 metros. Este sistema tem a capacidade de seguir e localizar diversos veículos em tempo real com um erro aproximado de um milímetro. Semelhante a esta aplicação temos também o trabalho de âmbito científico desenvolvido no *Center for Advanced Aerospace Technologies (CATEC)* [36] com o intuito de avaliar diferentes sistemas de posicionamento e a forma em que estes afetam a precisão dos Modelos Digitais de Elevação gerados a partir de imagens aéreas obtidas com veículos aéreos não tripulados. Neste caso o sistema VICON tinha o papel de fornecer a posição que serviria como comparação aos diferentes modelos com a utilização de diversos sensores. No âmbito do controlo de veículos em cenários interiores com múltiplos pisos surge o trabalho desenvolvido por Vijay Kumar [38]. Seguidamente, serão apresentados algumas soluções comercializadas pela Vicon.

Vicon Tracker

Esta solução aplica-se, principalmente, para soluções de seguimento de objetos em 3D ou 2D até 6 graus de liberdade através de nuvens de marcadores com recurso a fusão de informação 3D e 2D por forma a obter elevada precisão nas medidas [49]. Permite facilmente a integração de outras aplicações de 3D tais como VirTools ¹, Labview [22], Open Scene Graph [33], WorldViz [53], Panda3D [18] and VR Juggler [11] para que seja efetuada a manipulação de resultados da forma mais conveniente. Tem a capacidade de seguir até cinquenta objetos em simultâneo, por exemplo em aplicações de realidade virtual. Outra das principais aplicações consiste no seguimento do olho humano tal como utilizado na implementação do sistema Dikablis Eye Tracking[1].

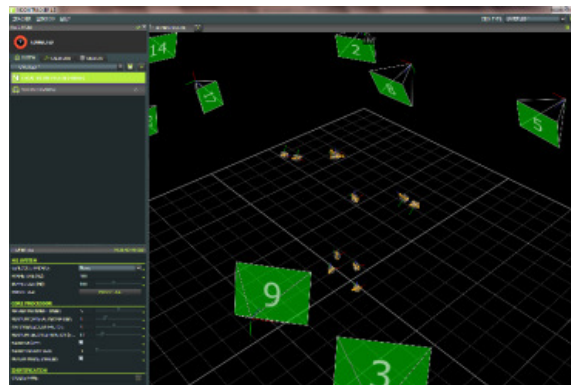


Figura 2.1: Exemplo de utilização de VICON tracker.

Vicon Pegasus

Esta solução de *software* foi desenvolvida em cooperação com IKinema [2] com o intuito de ser utilizada para aplicações de Motion Capture (MOCAP). Esta solução é apresentada como solução para desenvolvimento de

¹<http://www.3ds.com/products-services/3dvia/3dvia-virttools/>, consultado em 01-11-2014

jogos de vídeo assim como para a indústria como é o caso da cinematográfica para produção de filmes de animação e efeitos especiais, tal como se pode observar na figura 2.2.

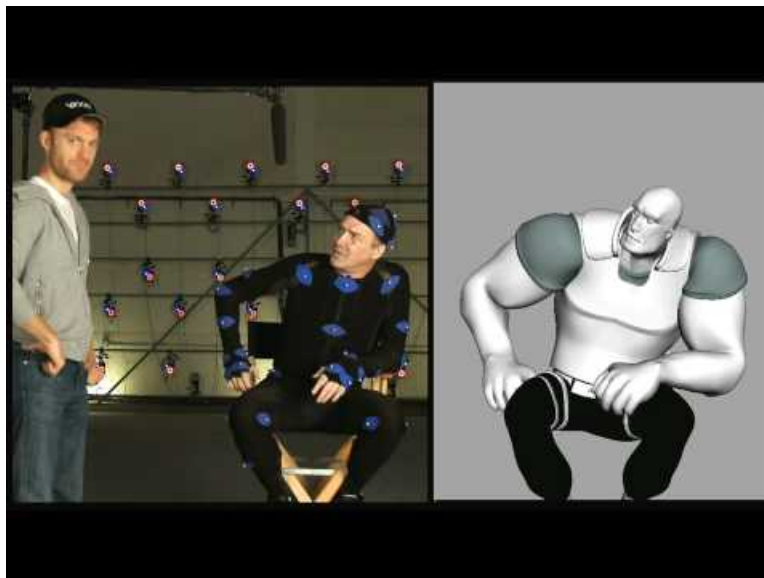


Figura 2.2: Exemplo de produção de animação com Vicon Pegasus.

Esta ferramenta é apresentada como sendo de fácil utilização para qualquer utilizador no manuseamento dos dados apresentados e está vocacionada para utilização com o sistema Vicon Blade [40], fazendo assim com que os seus utilizadores consigam que o processo de produção de animações seja mais eficiente, reduzindo custos e tempo.

Vicon Bonita

Sistema de alta resolução com captura de imagem até 1 Mpixel e com frequência de aquisição até 250 fotogramas por segundo. Este é capaz de capturar objetos com erro de translação inferior a 0.5 mm e com erro de rotação inferior a 0.5 graus recorrendo à utilização de 9 marcadores. Este pode ser facilmente montado. Este sistema pode ser calibrado em menos de 2 minutos com a utilização de um kit desenvolvido para o efeito.



Figura 2.3: Camera Vicon Bonita B10.



Figura 2.4: Ferramenta de calibração Vicon.

2.1.2 Optotrak

O Optotrak é um sistema ótico ativo [3] capaz de obter as coordenadas 3D de LEDs de infravermelhos (IRLEDs) usados como fontes de energia. São necessários pelo menos 3 IRLEDs para determinar a posição e orientação do objeto de corpo rígido em causa, a frequência máxima para os marcadores de 4600 Hz e de identificar até 512 marcadores em simultâneo. É possível o controlo dos LEDs sem fios, opção extra através do Tether less Kit, ou usando cabos. O pacote base Optotrak é formado pelo Optotrak Certus Position Sensor (sensor ótico constituído por 3 câmaras numa estrutura fixa), o *System Control Unit*, com a finalidade de controlo do hardware, o *Markers Strobers*, para activação dos IRLEDs que suporta até 24 LEDs, e o sistema *First Principle software* da NDI para aquisição, manipulação e apresentação dos resultados obtidos, conversão de ficheiros de dados, entre outros. O Optotrak é um sistema que permite o cálculo das coordenadas 3D, não tem por objetivo o cálculo da cinemática ou cinética do movimento. Note-se que existe ainda um dispositivo de ativação dos IRLEDs capaz de suportar 80 marcas ativas, o Axon Strober, e também é possível usar diversos conjuntos de Markers Strobers em simultâneo. Existe ainda o módulo *Application Programmer Interface* que permite o controlo e o acesso aos dados do sistema Optotrak e do Optotrak *Data Acquisition Unit* - dispositivo com 16 canais

analógicos e 8 digitais.

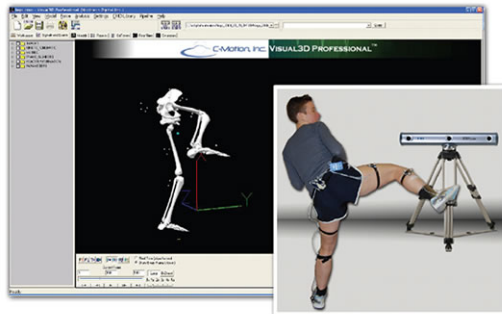


Figura 2.5: Sistema Optotrak.

2.2 Soluções *open source* de estimação de posição de objetos

2.2.1 SSL-Vision

Este sistema de detecção e localização [55] surgiu da necessidade de encontrar uma solução comum para todas as equipas presentes na competição da Robocup, *Small Size League* [50], visto que até à implementação desta solução cada uma das equipas possuía os próprios sistemas de estimação da posição dos robôs no campo. Desta forma, este sistema veio solucionar estes problemas resolvendo as incapacidades relacionadas com visão nesta competição podendo esta englobar várias câmaras em diferentes configurações. Esta ferramenta recorre à utilização de abordagens *multi-threaded* de modo a sistematizar as necessidades atuais de processamento *multicore* e *hyper-threaded*. A solução SSL-Vision efetua a detecção das marcas de cor dos robôs com recurso à biblioteca CMVision [13].

Calibração

Neste sistema são utilizadas tipicamente duas câmaras, uma para cada metade do campo, que precisam ser calibradas de forma a obter uma relação entre os pontos 3D do campo e os pontos 2D da imagem. Por forma a obter a efetuar a calibração, o problema foi abordado como um problema de minimização, na equação 2.1, [55].

$$\sum_{i=1}^4 |m_i - h(M_i, q, t, f, u_0, v_0, k)|^2 \quad (2.1)$$

Em que m_i corresponde aos pontos 2D no plano da imagem, e a função h representa os pontos 3D do plano do campo projetados para o plano da imagem. Desta forma a diferença entre ambos representa o erro em pixéis entre os pontos 2D e os pontos 3D projetados. A função h tem como parâmetros de entrada os pontos 3D, M_i , q e t representam respetivamente a rotação e translação entre a câmara e o campo, f a distância focal, u_0 e v_0 o centro ótico da imagem e k os parâmetros de distorção radial. Com o objetivo de refinar a calibração obtida é elaborado um método baseado na deteção de fronteiras de cor, das linhas de marcação e o campo. Por forma a melhorar a calibração será agora somado ao erro de minimização anterior a nova componente de erro entre a projeção das linhas do campo e as linhas detetadas na imagem. Este método induz a que seja otimizado o parâmetro, α , em cada medida em função do ponto medido e sua projeção para o plano de imagem. A nova formulação é apresentada na equação 2.2, [55].

$$\sum_{i=1}^4 |m_i - h(M_i, p)|^2 + \sum_{i=1}^4 |m_i - h(\alpha_i L_{i,1} + (1 - \alpha_i) L_{i,2}), p|^2 \quad (2.2)$$

Sendo, 2.2, $L_{i,1}$ e $L_{i,2}$ os segmentos de reta e α_i a posição atual na medida i . Devido ao facto de este sistema utilizar uma abordagem monocular os resultados dos objetos seguidos são medidas 2D dos mesmos no espaço. Por

este facto também a translação da câmara no eixo zz, altura, é retirada manualmente, desta forma este parâmetro não será obtido pelo método acima descrito.

Deteção e localização de objetos

Para deteção de padrões como os que estão colocados sobre os robôs foi adotada a biblioteca CMDragon [14]. Através da mesma é possível detetar a bola bem como outros padrões que podem ser configurados na aplicação para posteriormente serem detetados. Após ser conhecida a informação no plano da imagem torna-se possível a determinação da posição e orientação dos robôs, no campo, bem como a posição da bola. Este método foi estendido segundo o artigo [35] para a competição da RoboCup Standard Platform League (SPL), e para a Middle Size League(MSL) por [43]

2.2.2 *Kinect ground truth*

O desenvolvimento deste projeto[24] tem como foco principal a conceção de um sistema de *ground truth* de baixo custo para aplicação geral em todas as competições da RoboCup. Assim pretende-se que este sistema seja facilmente transportado para as competições, ou campos de treino das equipas, para que fosse capaz de detetar e localizar os robôs em tempo real sem recurso a marcadores e ainda que seja fácil de calibrar. A solução desenvolvida foi aplicada na competição RoboCup SPL [15] como sistema ground truth dos robôs NAO da empresa Aldebaran [34]. Com a utilização do sensor desenvolvido pela Microsoft na consola X-Box360, este tipo de sensores, tipicamente conhecidos como RGB-D, são capazes de capturar imagem a cores (RGB) e relacioná-la com a profundidade dos pixéis capturados (D , *depth*). O campo de visão desta câmara é de 57° na horizontal e de 43° na vertical. A profundidade dos pontos obtidos varia entre 0.7 m e 7 m consoante os testes realizados.

A implementação deste sensor foi elaborada com recurso à utilização de ROS devido ao facto de esta *framework* disponibilizar vários driver para o sensor em questão, assim como outras funcionalidades para o tratamento dos dados tais como o *Point Cloud Library* (PCL).

Calibração do sistema

O procedimento de calibração consiste no cálculo da rotação e translação relativas entre o referencial do campo e o referencial da câmara. Por forma a elaborar este procedimento são colocadas várias marcas pelo campo em posições conhecidas ou de fácil obtenção pelas medidas oficiais do campo tal como demonstrado pela figura 2.6. Para concluir o procedimento de ca-

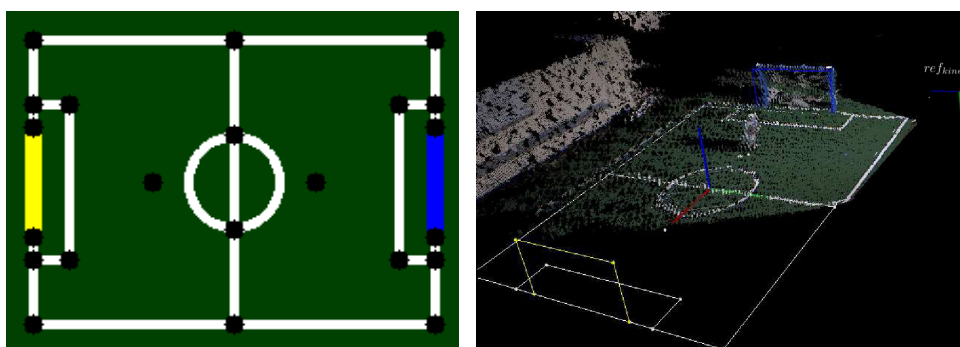


Figura 2.6: Esquerda, podem ser observados os pontos marcados no Campo para a calibração. Direita, exemplo da posição dos referenciais do campo, e do sensor kinect, ref_{kinect} .

libração é necessário que seja estabelecida a correspondência entre os pontos no plano no referencial global, já conhecidos, e os pontos no plano da imagem. Para tal, o utilizador deverá marcar os pontos na imagem manualmente. Desta forma, é possível calcular a homografia que possibilita que sejam relacionados os pontos 3D no plano do campo com os pontos 2D no plano da imagem. Através da decomposição desta transformação é possível obter os parâmetros de movimento de corpo rígido, rotação e translação. Para tal foi aplicado o método apresentado por [9]. Considere-

rando H a transformação, o primeiro passo começa por decompor a matriz em 3 outras matrizes, U, Λ, V , com recurso ao método de *Singular Value Decomposition* (SVD), na equação 2.3 :

$$[U, \Lambda, V^T] = SVD(H) \quad (2.3)$$

A rotação pode ser obtida a partir de duas das matrizes resultantes, na equação 2.4:

$$R = V \times U^T \quad (2.4)$$

Existem situações que podem acontecer no cálculo da rotação tal como quando o determinante da matriz R é -1 indica que existe uma reflexão na matriz de rotação calculada. De forma a solucionar este problema deverá ser feita a simetria da terceira coluna da matriz H . Já a translação pode ser calculada na equação 2.5:

$$T = P_{ref_{kinect}} - R \times P_{ref_{campo}} \quad (2.5)$$

Nesta etapa o sistema encontra-se já calibrado em relação ao campo, ou seja, torna-se possível apresentar as medidas no referencial do campo.

Deteção de robôs

A primeira etapa na deteção de robôs é a filtragem de possíveis pontos de interesse. O primeiro critério de seleção é a altura máxima dos pontos, sendo que são excluídos os pontos com menos de 30 cm. O outro critério de seleção dos pontos trata-se da delimitação da área de pesquisa ao campo, excluindo a pequena área, local onde não é permitido estar nenhum robô.

Após esta filtragem os pontos são agrupados em conjuntos, *clusters*, no espaço 3D segundo diversos parâmetros para que seja obtido um *cluster* por cada robô. Os parâmetros de formação de clusters são a distância máxima

de 10 cm de um ponto ao vizinho mais próximo do cluster em questão e cada cluster deve conter pelo menos 200 pontos. Nos *clusters* formados são excluídos os pontos com mais de 70 cm de altura visto que os robots apenas possuem 58 cm de altura. Quando dois robots estão próximos são aplicadas restrições no tamanho para separar o *cluster* formado onde estão englobados os dois robôs em dois *clusters* distintos, um para cada um dos robôs.

Deteção de bola

O algoritmo implementado para deteção da bola é similar ao anterior, para deteção dos robôs, tal como é evidenciado, abaixo:

- apenas são considerados os pontos dentro do campo;
- a altura de pesquisa está restrita a 15 cm;
- os pontos de interesse devem ser de cor laranja;
- a distância de um ponto ao ponto mais próximo do *cluster* considerado é de 10 cm;
- o *cluster* formado deve ter no mínimo 10 pontos;
- os *clusters* devem ter tamanho semelhante no espaço ao tamanho da bola.
- se mais que um *cluster* cumprir as restrições anteriores será avaliado o que satisfizer melhor os requisitos acima mencionados.

2.3 Calibração de sistemas de visão multi-câmara

Nesta secção serão abordados algumas metodologias para a calibração de sistemas de visão computacional.

2.3.1 Algoritmo de Tsai

A calibração dos parâmetros intrínsecos permite a obtenção das características físicas da câmara tais como a distancia focal, ponto principal e distorção da lente. Estas apenas estão dependentes das própria câmara/lente e na maioria dos casos variam até mesmo em modelos similares. O método utilizado para a determinação dos parâmetros intrínsecos foi desenvolvido por Roger Tsai[47]. Este método consiste na aplicação de um padrão de xadrez a preto e branco, para facilitar a tarefa de deteção dos quadrados constituintes do mesmo. Desta forma é possível obter os pontos do mundo assumindo que são conhecidas as dimensões dos quadrados, e que estes se encontram sempre à mesma distância do referencial ótico. Na figura 2.7 são mostrados vetores considerados para a obtenção dos parâmetros intrínsecos através do método enunciado.

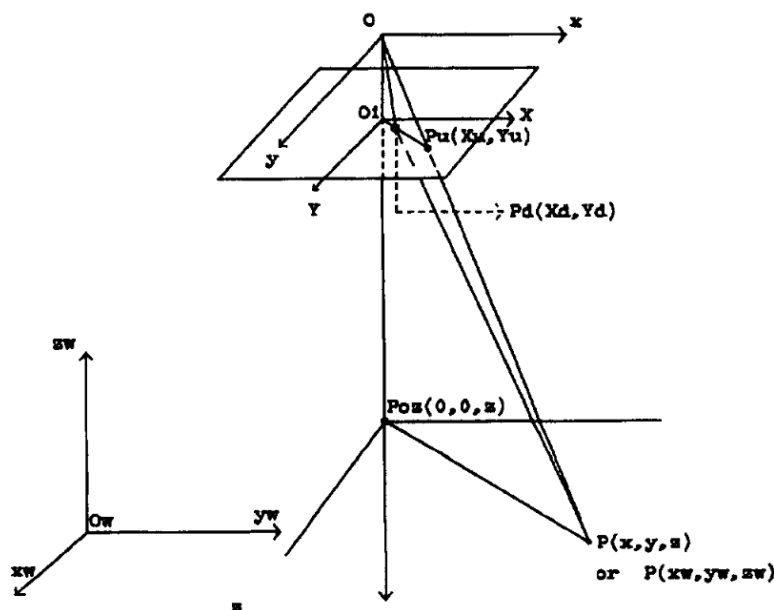


Figura 2.7: Algoritmo de Tsai.

No referido algoritmo é assumido o modelo de projeção da câmara, e que independentemente da grandeza da distorção, a direção do vetor $O_i P_d$,

vetor definido pela origem O_i , ponto principal, e pelo ponto P_d com as respectivas coordenadas x_d, y_d no plano da imagem, mantém-se radialmente alinhado com o vetor $P_{O_z}P$, vetor este formado assumindo que P é o ponto a considerar no mundo e que P_{O_z} é a projeção do ponto principal no mundo com coordenadas $(0,0,z)$ (a coordenada z deste ponto é igual à do ponto P). É considerado também que distância focal, f , não influencia a direção do vetor O_iP_d e que o ponto P no mundo é transladado e rodado em X e Y , fazendo assim com que o vetor O_iP_d seja sempre paralelo ao vetor $P_{O_z}P$ para todos os pontos considerados.

2.3.2 Auto-calibração de sistema multi-câmara

Nesta subsecção será abordada a técnica de auto-calibração. Tomas Svoboda apresentou em [45] um procedimento de calibração automática num cenário de multi-câmara. A técnica implementada apenas necessita da intervenção do utilizador para que este percorra o volume de trabalho com um marcador, por exemplo um laser ou um led. A escolha de um marcador luminoso tem por objetivo tornar a deteção do alvo o mais simplificada e precisa possível. A deteção está a ser feita com precisão sub-pixel conseguindo, desta forma, atingir erros de re-projeção inferiores a $1/5$ de pixel. Para apresentação deste algoritmo serão consideradas m câmaras e n pontos no espaço 3D representados por $\mathbf{X}_j = [X_j, Y_j, Z_j, 1]^T, j = 1 \dots n$, será considerado o modelo *pinhole*. Os pontos 3D serão projetados para a imagem segundo a equação 2.6

$$\lambda_j^i \begin{bmatrix} u_j^i \\ v_j^i \\ 1 \end{bmatrix} = \lambda_j^i \mathbf{u}_j^i = P^i \mathbf{X}_j, \lambda_j^i \in \mathbb{R} \quad (2.6)$$

Onde P^i corresponde à matriz de projeção dos pontos 3D para a imagem para a câmara i , tem dimensão 3×4 engloba os parâmetros intrínsecos,

distância focalponto principal da imagem, assim como a rotação e translação entre a câmara e o cenário, parâmetros extrínsecos. As variáveis u^i e v^i correspondem às coordenadas dos pontos projetados na imagem para a câmara i . Desta forma, o objetivo é estimar a escala λ_j^i e a matriz projetiva P^i , para tal o problema será formulado pela equação 2.7:

$$W_s = \begin{bmatrix} \lambda_1^1 \begin{bmatrix} u_1^1 \\ v_1^1 \\ 1 \end{bmatrix} & \dots & \lambda_n^1 \begin{bmatrix} u_n^1 \\ v_n^1 \\ 1 \end{bmatrix} \\ \vdots & \vdots & \vdots \\ \lambda_1^m \begin{bmatrix} u_1^m \\ v_1^m \\ 1 \end{bmatrix} & \dots & \lambda_n^m \begin{bmatrix} u_n^m \\ v_n^m \\ 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} P^1 \\ \vdots \\ P^m \end{bmatrix}_{3m \times 4} \begin{bmatrix} \mathbf{X}_1 & \dots & \mathbf{X}_n \end{bmatrix}_{4 \times n} \quad (2.7)$$

Desta forma verifica-se que se forem obtidos pontos suficientes na imagem (u_i^j, v_i^j) e forem estimados os fatores de escala, λ_i^j , podemos afirmar que a matriz W_s tem 4 colunas linearmente independentes e estas podem ser fatorizadas em P e X [44]. A partir desta fatorização é possível formar a transformação projetiva H pela equação 2.8:

$$W_s = PX = PHH^{-1}X = \hat{P}\hat{X} \quad (2.8)$$

O processo de auto-calibração calcula a matriz H por forma a que as matrizes \hat{p} \hat{X} se tornem euclidianas, segundo o método frequentemente chamado de estratificação euclidiana [20].

2.3.3 Homografia baseada pontos de fuga

Nesta secção é abordado o método de homografia. Através desta, é estimada a transformação que relaciona o mesmo local no mundo entre imagens

com perspectivas diferentes do mesmo cenário. O método analisado nesta subsecção [39] tem como objetivo servir de apoio à navegação em cenários urbanos visto que nestes casos a utilização de GPS torna-se ineficiente devido às estruturas que perturbam o sinal. Segundo o artigo [39] 95% da área urbana da cidade de Tóquio não permite uma localização fiável baseada em GPS. Este método utiliza a técnica de linhas de fuga para a obtenção dos pontos de fuga que vão permitir estabelecer a homografia. As linhas de fuga obtidas são relativas a linhas presentes em estruturas no cenário de ação, estas linhas devem ser paralelas entre si no mundo real. À luz da teoria da projeção perspectiva estas linhas vão-se interseccionar num ponto denominado de ponto de fuga. Este método requer que sejam calculados três pontos de fuga cada um relativo a cada uma das dimensões do espaço 3D. Para o cálculo dos pontos de fuga é necessário uma prévia seleção das linhas de fuga que contribuirão para os mesmos. Estas são agrupadas segundo a direção da sua fronteira, a sua orientação, consoante se localizem acima ou abaixo da linha do horizonte e a distância entre o segmento e o seu ponto central, em coordenadas polares com distância à origem ρ e ângulo θ , que deu origem à linha segundo a representação de linha chamada de parametrização normal, Este ponto será novamente referido com mais detalhe na secção 3.6 Este processo de agrupamento resulta num conjunto de linhas de fuga que correspondem a linhas paralelas no plano do mundo. A estimação do ponto de fuga foi formulada, na equação 2.9, como um problema de minimização onde se pretende minimizar a distância de todas as linhas de fuga ao ponto de fuga a ser calculado:

$$\min_x \sum_{k=1}^K (w_k l_k^T X^T)^2 \quad (2.9)$$

em que K é o número de retas obtidas para obtenção do ponto de fuga, w_k é o comprimento dos segmentos, $l_k^T = \cos \theta_k \sin \theta_k, -\rho$ correspondem à

informação dos segmentos de reta e $X^T = [u_x, v_x, 1]$ é o ponto de fuga em coordenadas homogêneas.

Aquando da aplicação da projeção perspectiva uma das propriedades que se mantêm relativamente ao mundo real é o *cross ratio*. Esta propriedade é uma relação entre vários pontos que se mantêm quer no plano do mundo, quer na plano da imagem, isto é após a aplicação da projeção perspectiva. A projeção das linhas 3D que estão no plano da estrada em ambas as imagens a considerar são caracterizadas e distinguidas pela unicidade do *cross ratio*. Esta propriedade será utilizada para descartar as linhas que não pertençam ao plano do chão, formulada na equação 2.10:

$$CR(l_a, l_b, l_c, l_d) = \frac{\sin \alpha_1}{\sin \alpha_2} / \frac{\sin \alpha_3}{\sin \alpha_4} \quad (2.10)$$

A partir das linhas de fuga presentes no plano do chão é possível calcular então a homografia, que permite relacionar os vários pontos do mundo capturados em perspectivas diferentes. A homografia contém a rotação e translação entre a posição dos centros óticos das câmaras no momento de aquisição de imagem. O método de cálculo está expresso na equação 2.11

$$\begin{aligned} l \propto H_1^t \cdot l_1, \quad l \propto H_2^t l_2 &\Rightarrow l_1 \propto [H_1^{-t} H_2^t] l_2 \\ l_1 \propto H^t l_2 \quad \text{com} \quad H &\propto H_2 H_1^{-1} \end{aligned} \quad (2.11)$$

Contudo a homografia não permite diretamente a transformação de um conjunto de linhas, desta forma são considerados os pontos dos segmentos de reta obtidos pela aplicação pelo método de transformada de Hough permite a identificação de pontos pertencentes às linhas já identificadas. Assim sendo a homografia relativa entre as câmaras é dada pela equação 2.12:

$$\begin{aligned} p_1 \propto H_1^t p, \quad p_2 \propto H_2^t l &\Rightarrow p_2 \propto [H_2 H_2^{-1}] p_1 \\ &p_2 \propto H^t p_1 \end{aligned} \quad (2.12)$$

Para melhor compreensão do conceito aqui abordado pode ser observada a figura 2.8.

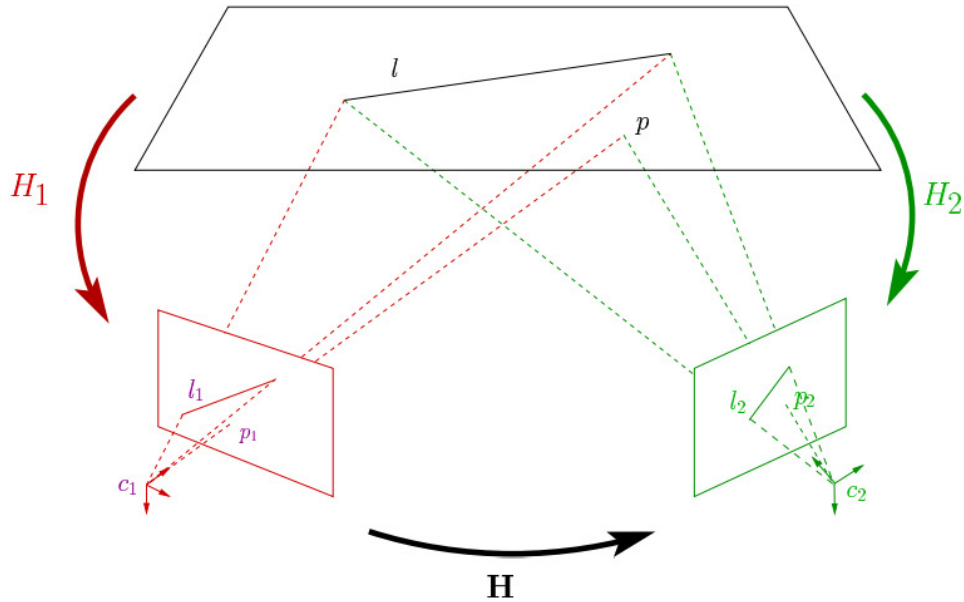


Figura 2.8: Representação da transformação de homografia.

Nesta figura é representado um segmento de reta, l , projetado em duas imagens distintas l_1 e l_2 , sendo a projeção para cada uma as câmaras dada por H_1 e H_2 . Por H é representada a homografia que relaciona o segmento de reta capturado pela câmara da esquerda para a câmara da direita

2.4 Sistema de localização em ambientes subaquáticos

Atualmente, o estudo e desenvolvimento de robótica para aplicação em meios submersíveis está em contínua ascensão para que a exploração dos recursos marítimos seja feita sem recurso a meios dispendiosos ou em que sejam arriscadas vidas humanas. Estes podem executar tarefas com fins comerciais tais como exploração de minerais ou inspeção de estruturas[23], para fins militares como em operações anti-minas [16], em aplicações com fins científicos tais como levantamento de alta resolução do fundo oceânico

[8] ou mesmo em operações fins humanitários como a procura de destroços ou vítimas de catástrofes[30]. Desta forma, surge a necessidade de detetar e localizar estes veículos como auxílio à sua navegação por forma a facilitar as tarefas desempenhadas por estes.

2.4.1 Ambiente subaquáticos – USBL

Atualmente existem diferentes sistemas de posicionamento com recurso a ondas acústicas, sendo um deles o Ultra Short Base line (USBL). Um sistema de posicionamento acústico subaquático é um sistema desenvolvido com o fim de localizar e auxiliar na navegação, entre outros, de veículos subaquáticos [48]. Através da determinação da distância e/ou direção de um ou vários dispositivos emissores em relação a um ou vários dispositivos recetores é possível conhecer a posição relativa entre os dispositivos emissores e recetores e, através desta informação, o posicionamento absoluto dos dispositivos a rastrear.

A configuração básica de um sistema USBL, consiste num *Transceiver* e um *Transponder* por cada alvo a localizar, e ainda numa unidade de processamento que calcula a posição do *Transponder* a partir das medições efetuadas pelo *Transceiver*. O funcionamento do USBL consiste no envio de um impulso acústico do *Transceiver* para o *Transponder*. Após a receção do impulso acústico o *Transponder* envia outro impulso para o *Transceiver*. Desta forma a partir da equação 2.13 através do tempo de voo entre impulsos é possível obter a distância "r" na referida equação, entre o *Transceiver* e o *Transponder*.

$$r = v \times \frac{\delta t}{2} \quad (2.13)$$

Na equação 2.13, v corresponde à velocidade do som na água, δt correspondente ao tempo que passou desde o momento da emissão do primeiro impulso acústico pelo *transceiver* até ao momento da receção do segundo impulso, emitido pelo *transponder* e recebido pelo *transceiver*. Na figura 2.9

podemos observar a operação de um USBL

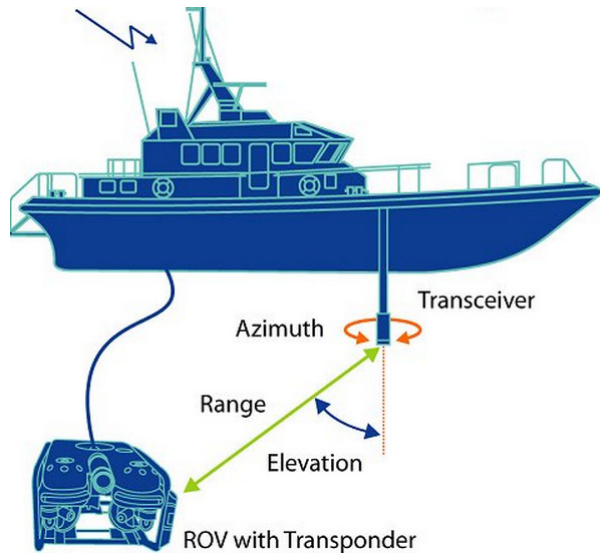


Figura 2.9: Utilização de USBL.

De modo a obter a posição 3D do *Transponder*, o *Transceiver* é constituído por um *array* de recetores, através deste torna-se possível obter o ângulo do sinal acústico recebido, através da medição da diferença da fase entre o eco recebido, entre vários recetores consecutivos pertencentes ao *array* do *Transceiver*. Desta forma, a direção do sinal recebido é baseado na diferença dos tempos de chegada do sinal acústico retornado por cada recetor do *array*. Na figura 2.10 podemos ver o USBL da TryTech, MicronNav

2.4.2 Ambientes Exteriores – GNSS

Os sistemas de navegação global por satélite, ou *Global Navigation satellite system* (GNSS), podem ser utilizados, tal como o *ground truth*, como dispositivos de localização absoluta, e têm a grande vantagem de, tal como o nome indica, fornecerem a localização em qualquer ponto do globo. Para tal o sistema apenas necessita de estar em espaço aberto de forma estabelecer

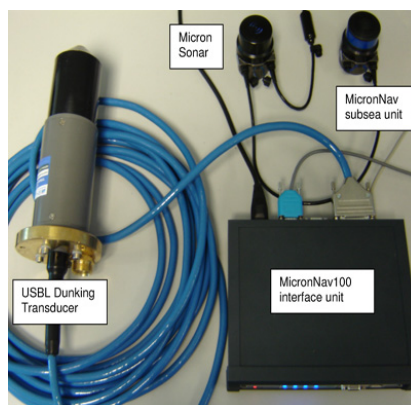


Figura 2.10: Constituintes de um sistema de USBL.

comunicação com os satélites. No entanto, estes sistemas têm limitações na precisão, dimensões e consumos energéticos, e para além disso em espaços fechados tal como interiores de edifícios, túneis ou caves ou mesmo em ambientes com infraestruturas de elevada dimensão no espaço envolvente estes sistemas não têm a capacidade de fornecer a localização ou podem fornecer medidas com reduzida confiabilidade. Por outro lado, dado o carácter global não são muitas vezes apropriados para aplicações em com movimentos reduzidos ou em cenários em que os movimentos exigem a obtenção da posição com precisão ou com alta frequência. O termo GNSS (*Global Navigation Satellite System*), foi utilizado pela primeira vez em 1991 pela Associação Internacional de Aviação Civil – ICAO, para designar todos os sistemas de posicionamento por satélites artificiais, com o objetivo de disponibilizar uma cobertura contínua e global em todo o território terrestre. Estes sistemas são utilizados por militares, empresas e até mesmo particulares para diversos fins, desde a navegação, localização, mapeamento e medição de tempo.

Posicionamento por GNSS

Os fundamentos básicos do *Global position System* (GPS) baseiam-se na determinação da distância entre um ponto, o recetor de GPS, aos pontos de

referência, os satélites. Conhecida a distância relativa que separa o recetor dos três, ou mais, pontos de referência (satélites), é possível determinar a posição global do recetor através da interseção circunferências, conjunto de locais possíveis que se encontram às distâncias medidas entre o recetor e o emissor. Desta forma, assume-se que os raios são as distâncias medidas entre o recetor e os satélites, e a interseção a posição global o recetor de GPS, ilustrado na figura 2.11.

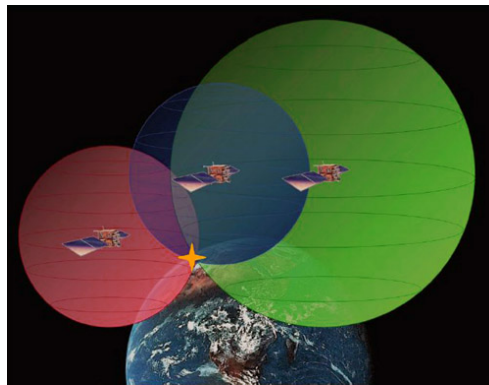


Figura 2.11: Ilustração de funcionamento de GPS.

Cada satélite transmite um sinal que é recebido pelo recetor, este por sua vez mede o tempo que os sinais demoram a chegar até ele. Multiplicando o tempo medido pela velocidade do sinal (velocidade da luz), obtemos a distância recetor-satélite,

$$D = c \times \Delta t \quad (2.14)$$

Esta página foi intencionalmente deixada em branco.

Capítulo 3

Fundamentos Teóricos

Atualmente a percepção sensorial é uma das áreas de investigação em sistemas autónomos. A utilização de sensores adequados assim como a boa interpretação dos mesmos é determinística nas decisões tomadas pelos próprios sistemas autónomos. Desta forma, a utilização de câmaras permite a aquisição de uma grande quantidade de informação o que há alguns tempos poderia ser uma desvantagem devido à enorme quantidade de informação a processar. Nos tempos que correm, com a evolução das tecnologias, este problema tornou-se cada vez menor devido ao aumento do poder de computacional fazendo com que as câmaras sejam sensores frequentemente utilizados em robótica móvel.

Neste capítulo serão abordados os fundamentos teóricos utilizados no desenvolvimento desta dissertação. Os princípios abordados tratam-se, sobretudo, de fundamentos de visão artificial e de percepção do espaço 3D.

3.1 Transformações de corpo rígido

As transformações de corpo rígido são passíveis de aplicar em todos os casos em que as distâncias entre as mesmas partículas são constantes. Esta abordagem é uma aproximação à realidade não tendo em conta as

deformações elásticas sofridas pelos objetos na realidade, porém para os problemas relacionados com esta dissertação, e de modo geral com os fundamentos de visão computacional a deformação não é relevante. A partir da mesma é possível estabelecer a dinâmica de grande número de partículas constituintes de corpos macroscópicos através da qual se obtêm resultados excelentes. Neste subcapítulo serão abordadas as transformações de corpo rígido de rotação e translação [37].

Rotação de corpo rígido

A rotação de corpo rígido tal como o próprio nome indica é utilizada para rodar um determinado objeto ou conjunto de partículas. A rotação pode ser parametrizada com base nos ângulos de euler, rotação segundo o eixo dos xx ou *roll* (ϕ), rotação segundo o eixo dos yy ou *pitch* (θ) e pela rotação segundo o eixo dos zz *yaw* (ψ). As rotações originadas pelos ângulos de Euler estão ilustradas na figura 3.1: Esta transformação podem

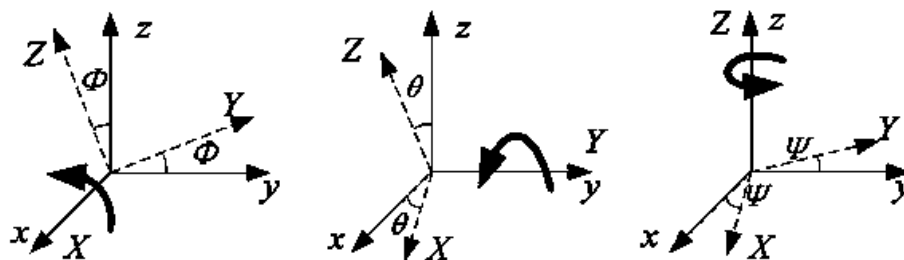


Figura 3.1: Representação dos Ângulos de Euler.

ser representadas de forma matricial, nas expressões 3.1, 3.2 e 3.3.

$$R(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & -\sin \phi & \cos \phi \end{bmatrix} \quad (3.1)$$

$$R(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \quad (3.2)$$

$$R(\psi) = \begin{bmatrix} \cos \psi & -\sin \psi & 0 \\ \sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.3)$$

A partir da multiplicação destas matrizes por qualquer ponto em 3 dimensões obtemos esses mesmos pontos rodados pelos ângulos correspondentes. A rotação segundo os 3 eixos de rotação pode ser efetuada apenas com uma única matriz que engloba os três ângulos de rotação, esta pode ser obtida a partir da multiplicação das 3 matrizes apresentadas nas expressões 3.1, 3.2 e 3.3, e será representada por 3.4.

$$R(\phi, \theta, \psi) = \begin{bmatrix} \cos \theta \cos \psi & \cos \theta \sin \psi & -\sin \theta \\ \cos \psi \sin \phi \sin \theta - \cos \theta \sin \psi & \sin \psi \sin \phi \sin \theta + \cos \phi \cos \psi & \sin \phi \cos \theta \\ \cos \psi \sin \theta \cos \phi + \sin \phi \sin \psi & -\sin \phi \cos \psi + \sin \psi \sin \theta \cos \theta & \cos \theta \cos \phi \end{bmatrix} \quad (3.4)$$

Os ângulos de Euler são frequentemente utilizados para representar rotações devido à sua simplicidade e significado intuitivo comparativamente a outras formas de representação como é o caso dos quatérniões. No entanto, os ângulos de Euler possuem uma limitação relacionada com a existência de singularidades. O fenómeno é usualmente designado de *Gimbal Lock*, este impossibilita o seu uso em sistemas cujo ângulo em pitch possa atingir ± 90 graus. O efeito *Gimbal Lock* ocorre quando a orientação do sensor não pode ser representada unicamente pelos ângulos de Euler, normalmente acontece quando o ângulo de pitch atinge ± 90 graus. Olhando para a figura 3.2 facilmente percebemos que esta orientação do avião pode ser obtida por

uma rotação em yaw e pitch ou por uma rotação em pitch e roll. Esta ambiguidade pode provocar efeitos graves no controlo do sistema. Este problema pode ser solucionado pela mudança do método de representação, como é o caso dos quaterniões.

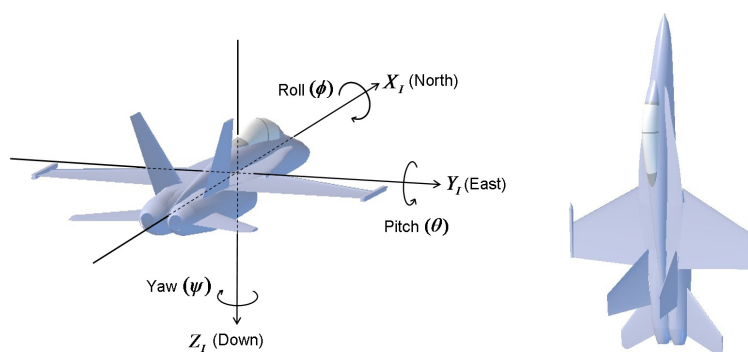


Figura 3.2: Na figura à esquerda pode ser observado o referencial do objeto em questão, no caso um avião. Na figura à direita pode ser observada o efeito *Gimbal Lock*.

Translação de corpo rígido

No modelo transformação de corpo rígido a translação é a distância entre os pontos sobre os quais os objetos rodaram, é definido por um vetor em três dimensões, equação 3.5.

$$T(x, y, z) = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \quad (3.5)$$

3.1.1 Coordenadas Homogéneas

Tal como demonstrado até então, as transformações de corpo rígido tendem a ser feitas de forma matricial. Na necessidade de efetuar as transformações de corpo rígido numa só operação surge a representação do espaço em coordenadas homogéneas. Desta forma passa a ser possível efetuar a transformação de corpo rígido numa só operação de multiplicação de ma-

trizes. A matriz de transformação de corpo rígido que permite a operação descrita está representada na equação 3.6 representada como M_{CR} :

$$M_{CR} = \begin{bmatrix} R(\phi, \theta, \psi)_{3 \times 3} & T(x, y, z)_{3 \times 1} \\ 0 & 1 \end{bmatrix} \quad (3.6)$$

A formulação acima é genérica para duas e três dimensões pelo que R corresponde à matriz de rotação, que deve ter o tamanho da dimensão em qual é utilizado, assim como a translação representada por T . Desta forma, a transformação de corpo rígido é obtida através da multiplicação da matriz da equação 3.6 pelas coordenadas do ponto a ser transformado. Abaixo na equação 3.7 pode ser observada esta operação para o exemplo de um ponto em três dimensões.

$$P' = M_{CR}P \quad (3.7)$$

3.2 Modelo da câmara

Para a aquisição de imagem as câmaras utilizam lentes para que a luz dos objetos viaje de forma convergente para os sensores utilizados pelas mesmas permitindo, assim, focarem o plano de captura desejado. A aquisição de imagem é feita pelo período de tempo suficiente para que o sensor capture a luz ao qual é exposto. Tipicamente são períodos de tempo curtos para minimizar os efeitos do movimento dos objetos capturados fazendo assim com que estes pareçam estáticos. De forma a processar a informação capturada por estes sensores(câmaras) são tipicamente aplicados modelos que tornam possível estabelecer a relação entre a informação capturada e o mundo envolvente. O modelo de câmara mais fácil de interpretar e analisar, e por isso o mais utilizado, é o modelo *Pinhole*. Este modelo é uma aproximação da utilização de uma câmara ideal, sem lente, com uma pequena abertura, o *pinhole* pela qual a luz refletida nos objetos passa para uma zona escura

onde será formada a imagem. Este modelo assemelha-se ao funcionamento do olho humano para visualização do mundo em redor tal como podemos ver pela figura 3.3, [25]. Como resultado, a imagem no plano da imagem está

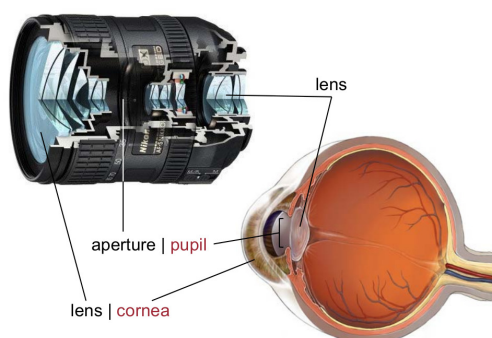


Figura 3.3: Comparação entre modelo *pinhole* e o olho humano [25].

sempre focada, é projetada inversamente neste plano e o tamanho da imagem em relação ao objeto capturado é relacionado por um único parâmetro da câmara, a distância focal, figura 3.4 à esquerda. Na figura 3.4 do lado direito podemos observar o modelo matemático utilizado. O ponto P tem

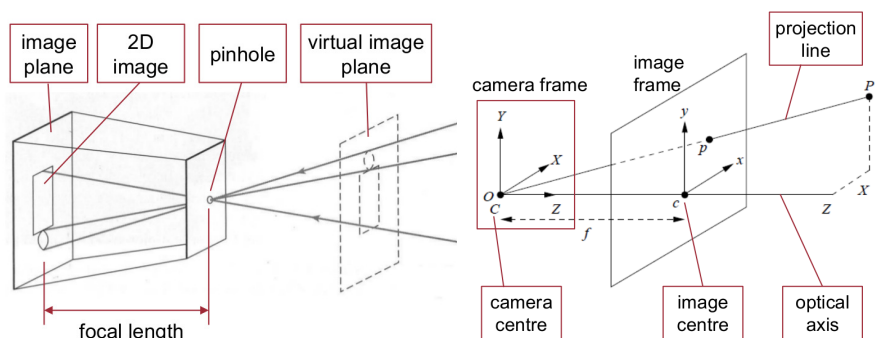


Figura 3.4: À esquerda representação da projeção da luz no plano da imagem. À direita é representado graficamente o modelo matemático utilizado para projeção de objetos no mundo para o referencial da câmara [25].

como coordenadas no mundo, $P = [X, Y, Z]^T$ e p em coordenadas na imagem $p = [x, y]^T$. Através da distância focal (f), com base na semelhança de triângulos é possível estabelecer relação entre as coordenadas na imagem, p ,

e as coordenadas no mundo, P , através da equação 3.8

$$\begin{cases} \frac{X}{Z} = \frac{x}{f} \\ \frac{Y}{Z} = \frac{y}{f} \end{cases} \Leftrightarrow \begin{cases} x = f \frac{X}{Z} \\ y = f \frac{Y}{Z} \end{cases} \quad (3.8)$$

A aplicação deste modelo é frequentemente chamada de projeção perspectiva.

3.2.1 Parâmetros intrínsecos

Os parâmetros intrínsecos são os parâmetros que permitem estabelecer a relação entre o mundo real e a informação capturada. Para além da distância focal abordada anteriormente, existem algumas não idealidades relacionadas com as lentes tais como a distorção da lente e o ponto principal da imagem, que necessitam de ser tratadas matematicamente de forma a construir o modelo matemático da câmara, para que assim a partir da informação adquirida seja possível obter resultados. Este modelo será formulado de forma matricial.

Ponto principal

Este parâmetro trata-se das coordenadas do centro da câmara projetado para a imagem, c e C na figura 3.4, o como é assumido como sendo a origem do plano da imagem. Desta forma, tendo em conta o centro ótico, este deve ser somado na equação anterior, obtendo assim a equação 3.9 :

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto Z \begin{pmatrix} \frac{fX+c_x}{Z} \\ \frac{fY+c_y}{Z} \\ 1 \end{pmatrix} = \begin{pmatrix} fX + Zc_x \\ fY + Zc_y \\ Z \end{pmatrix} = \begin{bmatrix} f & 0 & c_x & 0 \\ 0 & f & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.9)$$

Assim sendo, assumimos que a matriz que estabelecerá a relação entre o mundo envolvente e a imagem é a matriz, K , apresentada na equação 3.10

$$K = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3.10)$$

Distorção da lente

Para formulação dos conceitos apresentados até agora, no presente capítulo foi assumido que os pontos projetados do plano da imagem para o mundo exterior eram colineares. Contudo, isso não acontece na realidade devido a um fenómeno denominado por distorção. Este pode revelar-se de duas formas diferentes, distorção radial e distorção tangencial. A distorção tangencial tem como principal causa o desalinhamento físico dos vários elementos que constituem a lente, porém nos dias correntes os efeitos provocados por este tipo de distorção em alguns casos, como as câmaras com distância focal fixa, são de tal forma insignificantes que são por vezes desprezados. A distorção radial deve-se à variação do ângulo de refração. Desta forma a refração é nula no centro da lente e propaga-se à medida que a projeção dos feixes luminosos incidem nos extremos da lente. Este efeito toma maiores proporções com a utilização de lentes com abertura angular superior. Na figura 3.5 pode ser observado o efeito deste fenómeno. O modelo de correção da distorção provocada pela lente, para a distorção radial, é dado pelas expressões 3.11 e 3.12:

$$x_u = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (3.11)$$

$$y_u = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (3.12)$$

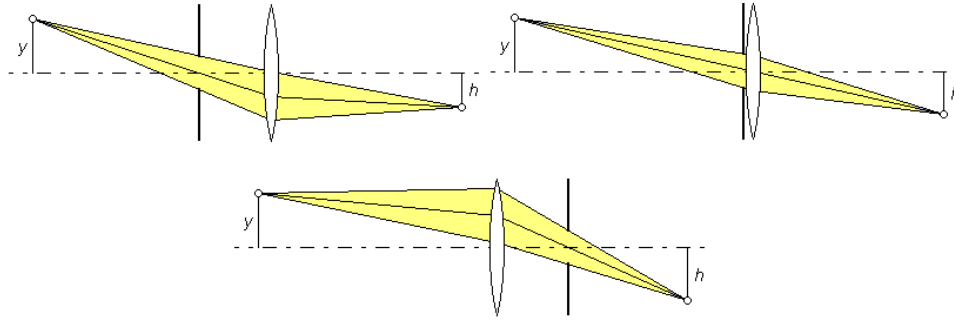


Figura 3.5: Comparação entre modelo *pinhole* e o olho humano.

O modelo para a distorção tangencial é apresentado pelas equações 3.13 e 3.14.

$$x_u = x + (2k_4y + k_5(r^2 + 2x)) \quad (3.13)$$

$$y_u = y + (k_4(r^2 + 2y) + 2k_5x) \quad (3.14)$$

Nas equações referidas, x e y correspondem à localização do ponto em pixels ainda com distorção e x_u e y_u corresponde à posição do respetivo ponto sem distorção. Os parâmetros k_1 , k_2 , k_3 dizem respeito aos coeficientes da distorção radial, enquanto que os parâmetros k_4 e k_5 representam os coeficientes distorção tangencial. Nas equações r representada a distância ao centro óptico.

3.2.2 Parâmetros extrínsecos

Geralmente os pontos no espaço são representados num referencial diferente do referencial da câmara, geralmente designado de referencial global ou referencial do mundo. Os parâmetros extrínsecos permitem estabelecer a relação entre ambos os referenciais, neste caso, entre um referencial definido no mundo e o referencial da câmara. Desta forma, torna-se simplificada a transformação das coordenadas de um objeto num referencial definido no mundo, de forma conveniente para melhor interpretação dos resultados ob-

tidos, para o referencial 3D da câmara para que assim possa ser aplicada a projeção abordada no capítulo de parâmetros intrínsecos, 3.2.1. A relação é estabelecida através da transformação de corpo rígido, envolve uma rotação e uma translação em três dimensões. Esta pode ser definida pela equação 3.7, abordada no subcapítulo sobre transformações de corpo rígido, 3.1, em que R e T correspondem à rotação do referencial do mundo para o referencial da câmara.

3.2.3 Projeção do espaço 3D para 2D

A projeção de pontos no espaço tridimensional para o referencial da imagem é baseada nos fundamentos abordados nos subcapítulos anteriores, 3.2.1 e 3.2.2. A primeira fase deste processo é a transformação de corpo rígido abordada na secção pela equação 3.7 Nesta fase as coordenadas dos objetos estão já transformadas para o referencial da câmara pelo que estamos em condições para aplicar a teoria abordada na subsecção 3.2.1

$$p = KP' \quad (3.15)$$

Em que P' corresponde às coordenadas do objeto no referencial 3D da câmara, p representa o ponto P' projetado na imagem. Assim podemos assumir que a transformação referida é feita à custa da matriz PM na equação 3.16:

$$PM = K[R|T] = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R(\phi, \theta, \psi) & T(x, y, z) \\ 0 & 1 \end{bmatrix} \quad (3.16)$$

Esta matriz tem dimensão 3×4 . Desta forma, a projeção dos pontos no referencial do mundo é feita partindo da equação 3.17

$$p = K[R|T]P = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R(\phi, \theta, \psi) & T(x, y, z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.17)$$

3.3 Geometria epipolar

Nesta secção serão abordados os fundamentos da geometria epipolar, a matriz essencial e a matriz fundamental.

3.3.1 Matriz Essencial

Através da matriz essencial é possível estabelecer as restrições para um sistema de visão *stereo*. O plano epipolar é definido por 3 pontos, os centros ópticos do par *stereo*, O_1 e O_2 , e o ponto no mundo a considerar p . A interseção do plano epipolar com os planos das imagens formam as linhas epipolares, l_1 e l_2 . Para melhor compreensão dos pode ser observada a Figura 3.6

Assim se Rotação e a translação relativas entre as duas câmaras forem conhecidas, através da geometria epipolar pode são consideradas observações importantes:

Se o ponto p projetado na câmara esquerda x_1 for conhecido, a linha epipolar formada pelos pontos e_1 e e_2 é também conhecida. desta forma a projeção ponto p para a imagem da direita x_2 está sobre a linha linha epipolar l_2 . Generalizando, para cada ponto de uma imagem correspondente ao mesmo ponto no mundo este pode ser observado na outra imagem sobre a linha epipolar. Através da mesma pode ser verificado se dois pontos em duas imagens capturadas por um par *stereo* correspondem ao mesmo ponto

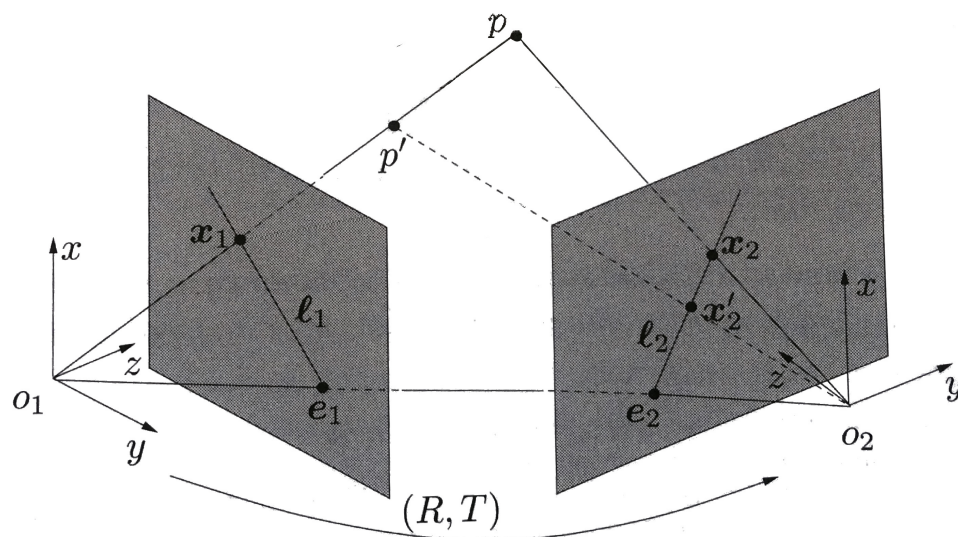


Figura 3.6: Representação da restrição epipolar.

3D definindo assim a restrição epipolar.

No caso da restrição epipolar se verificar, para os pontos x_1 e x_2 nas imagens estes devem intersectar precisamente em p sendo desta forma possível obter as coordenadas tridimensionais do ponto p partir das coordenadas dos dois pontos de imagem, este processo é tipicamente chamado de triangulação.

Formulação da matriz essencial

Considerando a imagem anterior em que x_1 e x_2 são os pontos nas imagens correspondentes ao ponto p , no mundo e que P é o plano no qual p está contido, a relação entre os pontos X_1 e X_2 , segundo [27], é dada pela equação 3.18:

$$X_2 = RX_1 + T \quad (3.18)$$

Onde os pontos X_1 e X_2 correspondem ao ponto p no referencial de cada uma das câmaras. Considerando a ambiguidade na determinação da profundidade, abordada na equação 3.9 como Z , obtemos a equação 3.19, considerando $\lambda_i, i = 1, 2$ a relação de ambiguidade:

$$\lambda_2 x_2 = R\lambda_1 x_1 + T \quad (3.19)$$

De forma a eliminar a ambiguidade de profundidade por forma a provar a restrição epipolar ambos os lados serão multiplicados por \widehat{T} , equação 3.20:

$$\lambda_2 \widehat{T}x_2 = \widehat{T}R\lambda_1 x_1 \quad (3.20)$$

Em que $\widehat{T}x_2$ é o produto vectorial entre os vetores T e x_2 , em que o resultado gerado é um terceiro vetor perpendicular aos dois que lhe deram origem. Assim, concluímos da afirmação anterior que $\widehat{T}x_2 = T \times x_2$ e das propriedades de vetores que o produto interno de dois vetores perpendiculares é nulo ou seja $\langle x_2, \widehat{T}x_2 \rangle = 0$. Com base nestas deduções multiplicando à esquerda ambos os lados da igualdade da equação 3.20 por x_2^T , obtemos 3.21:

$$x_2^T \lambda_1 \widehat{T}x_2 = x_2^T \widehat{T}R\lambda_1 x_1 \quad (3.21)$$

Pelas afirmações anteriores obtemos a equação 3.22 que prova a restrição epipolar:

$$\langle x_2, \widehat{T}R\lambda_1 x_1 \rangle = x_2^T \lambda_1 \widehat{T}x_2 = 0 \Leftrightarrow x_2^T \widehat{T}R\lambda_1 x_1 = 0 \quad (3.22)$$

Assim a matriz epipolar de definida pela equação 3.23

$$E \doteq \widehat{T}R \quad (3.23)$$

3.3.2 Matriz fundamental

A matriz fundamental pode ser utilizada para estabelecer a correspondência de pontos, x_1 e x_2 entre duas imagens esta pode ser formulada com base na matriz essencial [20]. Assim, a matriz fundamental é definida pela equação 3.24:

$$F = K^{-T}EK^{-1} \quad (3.24)$$

A matriz fundamental satisfaz a condição de qualquer par de câmaras em configuração *stereo*, pela equação 3.25:

$$x_2Fx_1 = 0 \quad (3.25)$$

3.4 Homografia planar

A homografia planar é utilizada para estabelecer a correspondência de pontos entre imagens com perspectivas diferentes sobre a mesma zona do meio envolvente. A utilização deste método está limitada a casos em que os pontos considerados sejam coplanares no mundo, ou seja, pertençam ao mesmo plano [7]. Para a utilização deste método é requerido ainda que as imagens capturadas pertençam ao mesmo instante temporal ou que os pontos de interesse estejam estáticos aquando a captura de ambas as imagens. A matriz de homografia, H , é representada por uma matriz de dimensão 3×3 . Considerando a equação 3.18, apresentada na secção 3.3.1:

$$X_2 = RX_1 + T$$

Sabemos que os pontos X_1 e X_2 originam os pontos x_1 e x_2 nas correspondentes imagens e que estes respeitam a restrição essencial da geometria

epipolar, apresentada no subcapítulo 3.3.1 pela equação 3.22. Com o intuito de formular a homografia consideramos o vetor $N = [n_1, n_2, n_3]$ como sendo o vetor normal ao plano de homografia P , e $d > 0$ a distância entre o centro ótico da primeira câmara e o plano P , obtemos a equação 3.26

$$N^T X_1 = n_1 X_1 + n_2 Y_1 + n_3 Z_1 = d \Leftrightarrow \frac{1}{d} N^T X_1 = 1 \quad (3.26)$$

Assim substituindo a equação 3.26 na equação 3.18, obtemos 3.27:

$$X_2 = R X_1 + T = R X_1 + T \frac{1}{d} X_1 = \left(R + \frac{1}{d} T N^T \right) \quad (3.27)$$

Assim obtemos H , equação 3.28:

$$H \doteq \left(R + \frac{1}{d} T N^T \right) \quad (3.28)$$

Na imagem da figura 3.7 é possível ter uma percepção do espaço 3D para a definição da homografia.

3.5 Decomposição de Homografia

Na secção anterior tivemos oportunidade de verificar o procedimento para a obtenção da homografia. Este procedimento é válido para situações onde são conhecidos os parâmetros do sistema tais como a rotação (R) e a translação (T). Porém a homografia é passível de ser calculada através de métodos de minimização tal como demonstrado em [54], em que necessitamos do conhecimento da informação de vários pontos no plano da imagem. No presente subcapítulo iremos abordar um algoritmo para obtenção dos parâmetros de movimento de corpo rígido relativo entre as câmaras[27]. Considerando a matriz de homografia, H , de forma a obter os parâmetros extrínsecos das câmaras o primeiro passo é a sua decomposição utilizando

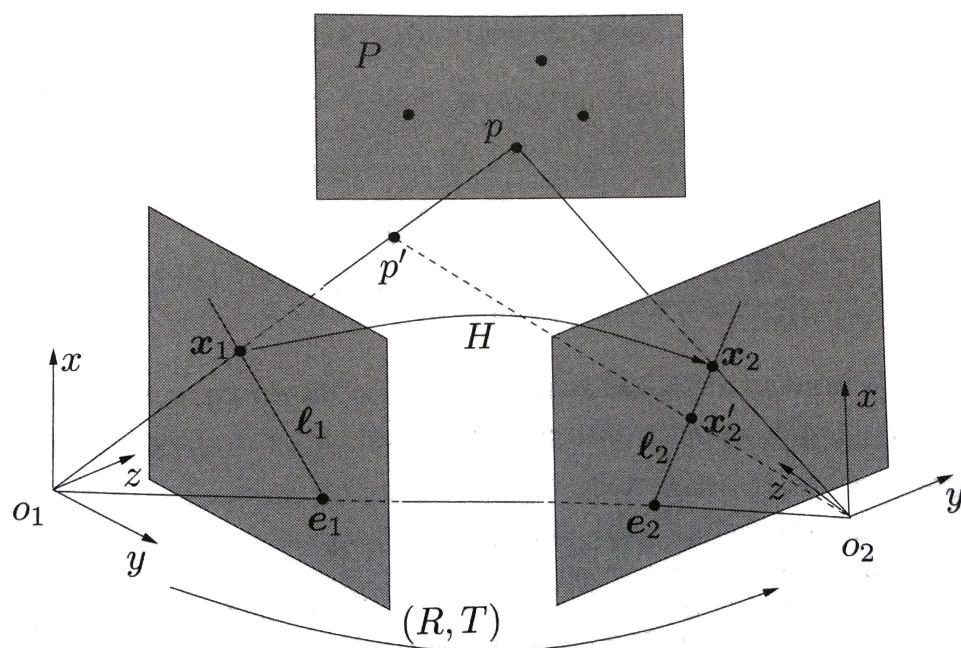


Figura 3.7: Representação do espaço 3D para aplicação de homografia.

SVD este método permite decompor uma matriz em três matrizes que são matematicamente equivalentes, $U\sigma V^T = SVD(H)$, em que Σ são os valores próprios da matriz, $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, $\sigma_2^2 = 1$. Assim pela equação 3.29 obtêm-se:

$$H^T H = V \Sigma V^T \quad (3.29)$$

Assumindo $V = [v_1, v_2, v_3]$ e que H preserve o tamanho de qualquer vetor que lhe seja ortogonal obtemos a equação 3.30

$$H^T H v_1 = \sigma_1^2 v_1, \quad H^T H v_2 = v_2, \quad H^T H v_3 = \sigma_3^2 v_3 \quad (3.30)$$

Posteriormente é assumido que v_2 é ortogonal a N e T e que o seu tamanho é preservado quando projetado por H . Assim verifica-se a equação 3.31

$$u_1 = \frac{\sqrt{1 - \sigma_3^2} v_1 + \sqrt{\sigma_1^2 - 1} v_3}{\sqrt{\sigma_1^2 - \sigma_3^2}}, \quad u_2 = \frac{\sqrt{1 - \sigma_3^2} v_1 - \sqrt{\sigma_1^2 - 1} v_3}{\sqrt{\sigma_1^2 - \sigma_3^2}} \quad (3.31)$$

Solução 1	$R_1 = W_1 U_1^T$ $N_1 = \widehat{v}_2 u_1$ $\frac{1}{d} T_1 = (H - R1) N_1$	Solução 3	$R_3 = R_1$ $N_3 = -N_1$ $\frac{1}{d} T_3 = -\frac{1}{d} T_1$
Solução 2	$R_1 = W_2 U_2^T$ $N_2 = \widehat{v}_2 u_2$ $\frac{1}{d} T_2 = (H - R2) N_2$	Solução 4	$R_4 = R_4$ $N_4 = -N_2$ $\frac{1}{d} T_4 = -\frac{1}{d} T_4$

Tabela 3.1: Soluções obtidas pela decomposição de homografia

$$S_1 = \text{span}v_2, u_1, \quad S_2 = \text{span}v_2, u_2 \quad (3.32)$$

Assim considerando v_2 ortogonal a u_1 e u_2 , e que $\widehat{v}_2 u_1$ é normal a S_1 e que $\widehat{v}_2 u_2$ é normal a S_2 , então $v_2, u_1, \widehat{v}_2 u_1$ e $v_2, u_2, \widehat{v}_2 u_2$ formam bases ortonormadas, obtendo assim a equação 3.33:

$$Rv_2 = Hv_2, \quad Ru_i = Hu_i, \quad R(\widehat{v}_2 u_i) = \widehat{H}v_2 Hu_i \quad (3.33)$$

assim podemos obter as seguintes matrizes

$$\begin{aligned} U_1 &= [v_2, u_1, \widehat{v}_2 u_1], & W_1 &= [Hv_2, Hu_1, \widehat{H}v_2 Hu_1]; \\ U_2 &= [v_2, u_2, \widehat{v}_2 u_2], & W_2 &= [Hv_2, Hu_2, \widehat{H}v_2 Hu_2] \end{aligned} \quad (3.34)$$

Deste método na tabela 3.1, são apresentadas 4 soluções matematicamente possíveis para o cálculo dos parâmetros de transformação de corpo rígido que relaciona as duas câmaras.

3.6 Transformada de Hough

A transformada de Hough foi inventada por Richard Duda and Peter Hart in 1972 [17] com o propósito de detetar segmentos de reta em imagens. Mais tarde foi expandida por Dana H. Ballard [10] em 1981 para que além das linhas fosse possível detetar outras formas geométricas como circunferências e elipses. Nesta secção será abordada a primeira versão da

transformada de Hough para que visa a detecção de retas em imagens. Uma imagem para ser passível de ser analisada com recurso à transformada de Hough deve estar em formato binário, representação preto e branco (diferente de tons de cinzento). Assim sendo todos os pixéis brancos serão convertidos para o plano da transformada de Hough, plano $\rho - \theta$, onde irão formar uma senoide formulada pela equação 3.35:

$$\rho = x_i \cos \theta + y_i \sin \theta, \quad (3.35)$$

em que x_i e y_i corresponde aos pontos brancos da imagem. Desta forma, se restringirmos o intervalo de representação de θ para $[0, \pi]$ cada reta no plano da imagem terá apenas uma representação no plano $\rho - \theta$. Desta forma é fácil de perceber que as sinusoides formadas com base no plano da imagem $x - y$, que os pontos x_i e y_i forem colineares, no plano $\rho - \theta$ as sinusoides irão se intersestar, em x_o, y_o . Assim, obtemos todos os pontos colineares na imagem e assim é possível formar a reta na qual estes estão contidos. Assim sendo, ao transformar as retas do plano da imagem, plano $(x - y)$ para o plano da transformada de Hough ($\rho - \theta$) transformamos o problema de detecção de retas na imagem num problema de encontrar interseções. Um exemplo da representação das sinusoides formadas no espaço da transformada de Hough pode ser observado na Figura 3.8.

Desta forma a representação de uma reta pode ser feita a partir de um único ponto no plano da transformada de Hough, através da representação chamada parametrização normal, esta representação é demonstrada na figura 3.9.

Este ponto é representado em coordenadas polares. A reta detetada na imagem contém esse mesmo ponto e é perpendicular ao segmento de reta formado. Por forma a eliminar falsos positivos na detecção de retas é estabelecido um número mínimo de interseções num determinado ponto no plano da transformada de Hough para que a reta seja considerada como

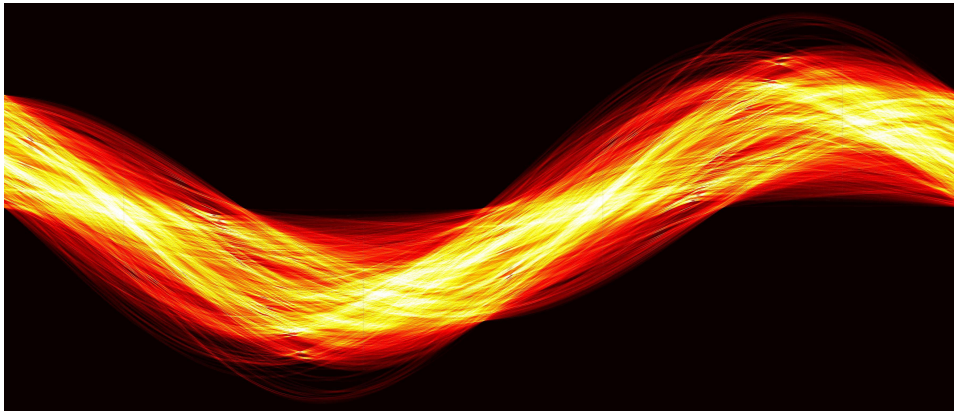


Figura 3.8: Representação do plano da transformada de Hough, $\rho - \theta$.

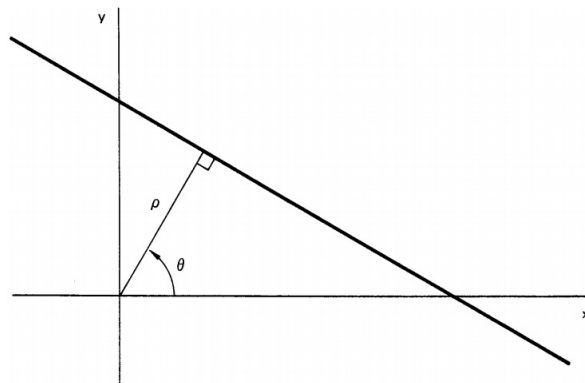


Figura 3.9: Parametrização normal - Representação de retas na transformada de Hough.

válida.

3.7 Linhas e pontos de fuga

As linhas de fuga são linhas que no espaço 3D envolvente são paralelas entre si. Na aplicação da projeção perspectiva, ou modelo de pinhole, estas linhas irão intercectar-se num ponto chamado de ponto de fuga. A figura 3.10 ilustra uma representação das linhas de fuga no plano do mundo e no plano da imagem. Estando resolvido o problema das linhas de fuga, o

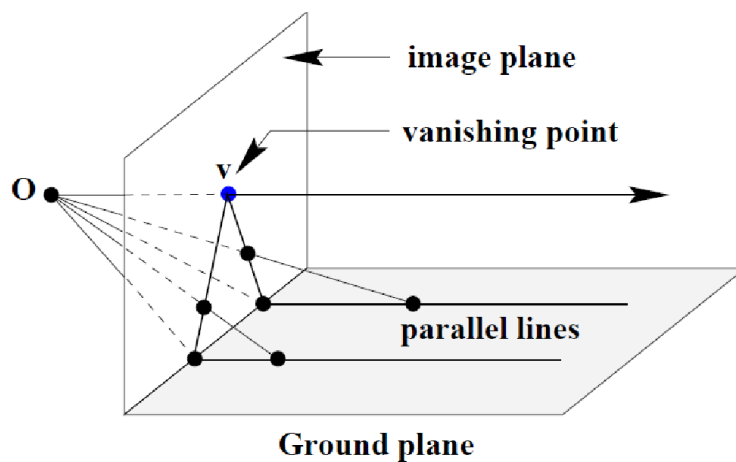


Figura 3.10: Representação de linhas de fuga.

cálculo dos pontos de fuga é dado pelo ponto onde as diversas retas com a mesma direção no plano do mundo se intersectam. Através da determinação do ponto de fuga é possível a obtenção da orientação das linhas que deram origem no plano do mundo através da aplicação da projeção perspectiva deste ponto.

$$d = \begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} \frac{v_x - c_x}{f_x} \\ \frac{v_y - c_y}{f_y} \end{bmatrix} \quad (3.36)$$

Em que d é o vetor que define orientação da câmara relativamente às linhas detetadas no plano do mundo, c_x , c_y f_x e f_y dizem respeito aos parâmetros intrínsecos da câmara, distância focal e as coordenadas do ponto principal e v_x e v_y são as coordenadas do ponto de fuga obtido.

Capítulo 4

Setup experimental

O presente capítulo apresenta as opções de desenvolvimento bem como o projeto do *setup* experimental para validação dos métodos de calibração endereçando os cenários de futebol robótico e o tanque de experiências subaquáticas.

4.1 Arquitetura do sistema

No presente capítulo é apresentada a arquitetura geral do sistema de *ground truth*. Serão abordadas as opções tomadas relativamente ao projeto, com o intuito de integrar os diversos constituintes do mesmo como uma solução generalizada e versátil face ao cenário a que este se destina a ser aplicado. Na figura 4.1 pode ser observada a arquitetura generalizada para o sistema de *ground truth* em questão.

Este sistema foi projetado de forma a ter uma arquitetura distribuída para que seja possível a sua expansão, quer em relação ao número de câmaras quer em relação ao número de CPUs, ainda que para a obtenção dos resultados do presente documento, apenas tivesse sido utilizado uma unidade central de processamento. Assim é possível ter unidades de processamento a efetuar a aquisição de imagem e a aplicar algoritmos de deteção e se-

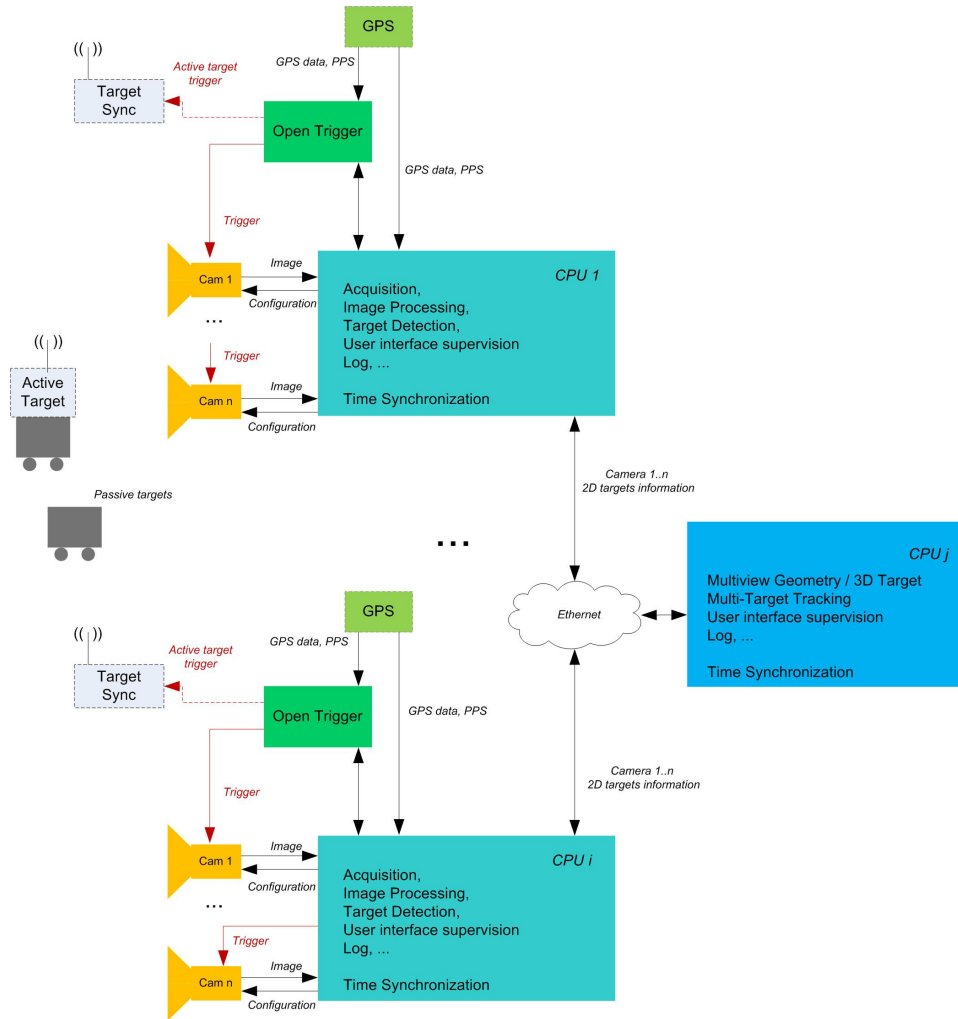


Figura 4.1: Arquitetura de aquisição sistema de *ground truth*.

guimento de possíveis objetos de interesse em 2D. Deste modo é possível combinar a informação em 2D processadas em diferentes CPUs, de forma a fundir a informação 2D para a obtenção de resultados em 3D. Um dos principais requisitos nesta abordagem de processamento descentralizado é o sincronismo, por forma a permitir efetuar a conveniente comparação de resultados. Para tal, quando assim for possível são utilizados sistemas de GPS para o efeito. Estes para além do tempo absoluto disponibilizam também o sinal PPS(*pulse per second*) que pode ser útil para ajudar a sincronização.

O disparo das câmaras deverá ser feito tendo em conta o sincronismo com os restantes constituintes do sistema.

4.1.1 *Ground truth* cenário MSL

Nesta secção é apresentado o sistema localização de objetos para o campo de futebol robótico médio. Na figura 4.2 observa-se a configuração das câmaras no cenário do futebol robótico médio.



Figura 4.2: Sistema *Ground truth* Campo MSL.

4.1.2 *Ground truth* Tanque

Nesta secção é apresentada a arquitetura do sistema de localização de objetos para o campo de tanque de experiências subaquáticas. A partir da figura 4.3 pode ser observada a configuração das câmaras no cenário do

tanque de experiências subaquáticas.

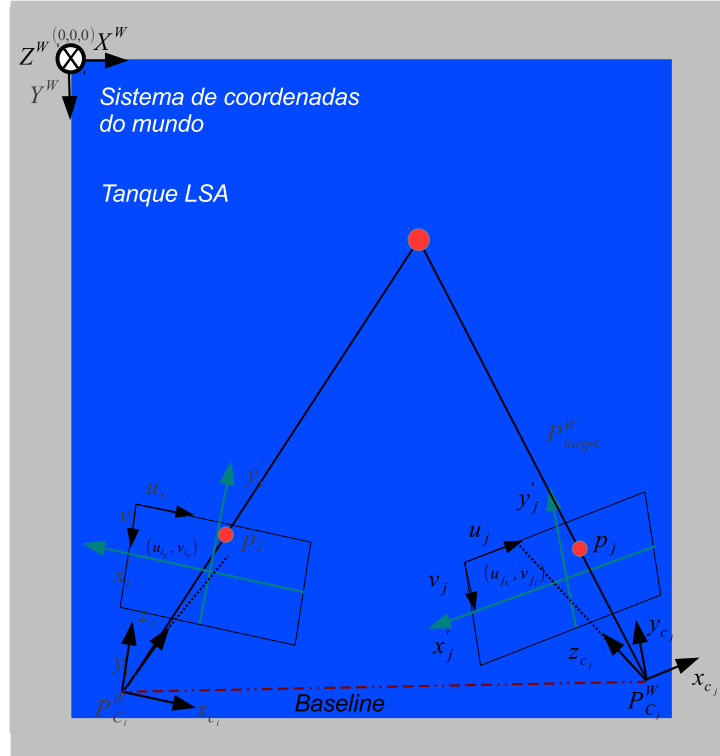


Figura 4.3: Sistema *Ground truth* tanque.

4.2 Hardware

Para a elaboração desta dissertação foi instalado um sistema de câmaras de forma a aplicação de técnicas multi-câmara. A instalação efetuada segue a arquitetura apresentada na figura 4.4

4.2.1 Câmara

Durante o desenvolvimento deste projeto foram utilizadas diversas câmaras para a realização das experiências efetuadas. A câmara utilizada foi desenvolvida pela empresa Basler AG, o modelo da câmara é Basler acA 1300-30gc [6]. As principais características desta câmara são:

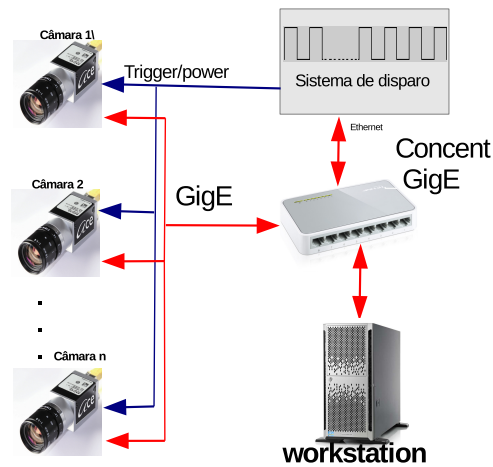


Figura 4.4: Arquitetura do sistema *ground truth* implementado.

- Resolução: 1296×966 ;
- Comunicação: GigE (*Gigabit ethernet*)
- Modos de Disparo: *trigger externo*, Ethernet, free run;
- Débito de imagens: 30 imagens por segundo;
- Alimentação: PoE ou 12 VDC
- Consumo: 2.2 W
- Topologia do sensor: CMOS;

De entre outros modelos, este modelo de câmara destacou-se de outros igualmente disponíveis no LSA devido a algumas das características descritas acima, principalmente a comunicação GigE, o sincronismo por *trigger* externo, e o *frame rate*.

4.2.2 Sistema de alimentação e disparo das câmaras

Com o intuito de encontrar uma solução para a alimentação e disparo das câmaras foi desenvolvido um sistema embebido com acesso remoto que

permite a alimentação e o controlo do disparo das câmaras. Este sistema embebido é acessível por IP a partir de um explorador de Internet onde podem ser configuradas as taxas de disparo das câmaras bem como a largura dos impulsos dos sinais de disparo. Este sistema recebe a informação da taxa de disparo e de forma síncrona são gerados os sinais de disparo para todas as câmaras para que assim seja cumprido o requisito de aquisição simultânea de todas as câmaras. A alimentação para as câmaras, está a ser gerada neste sistema. Na figura 4.5 é apresentado um diagrama de blocos da placa de disparo das câmaras.

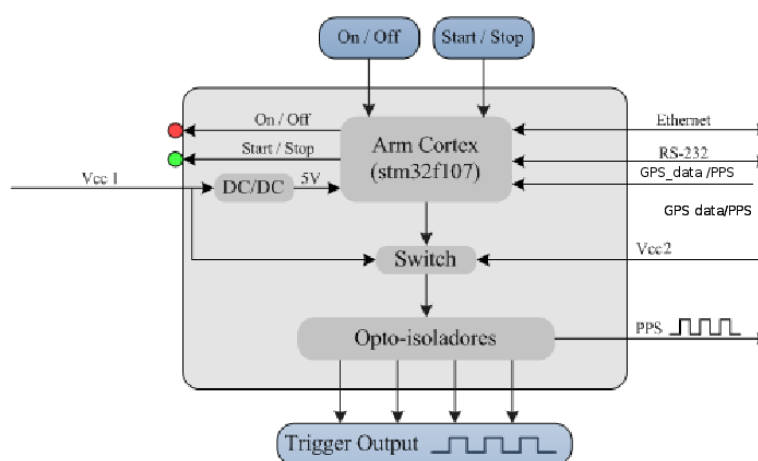


Figura 4.5: Arquitetura do sistema de disparo a alimentação das câmaras.

Neste sistema é possível ser ajustada a tensão do sinal de disparo das câmaras, para tal deve ser colocada a tensão desejada na entrada assinalada na figura acima como VCC2 a tensão pretendida. Para sincronização do sistema de disparo das câmaras, este está conectada com um GPS de forma a obter o tempo absoluto e também o sinal auxiliar PPS (*Pulse Per Second*). Esta placa disponibiliza também o sinal PPS na saída indicada com a legenda PPS. Para a implementação deste método de sincronismo, é necessário que exista pelo menos um recetor de GPS com condições para receber o sinal de satélite em questão.

Na figura 4.6 pode ser observado o interface de configuração da página web para configuração do disparo.

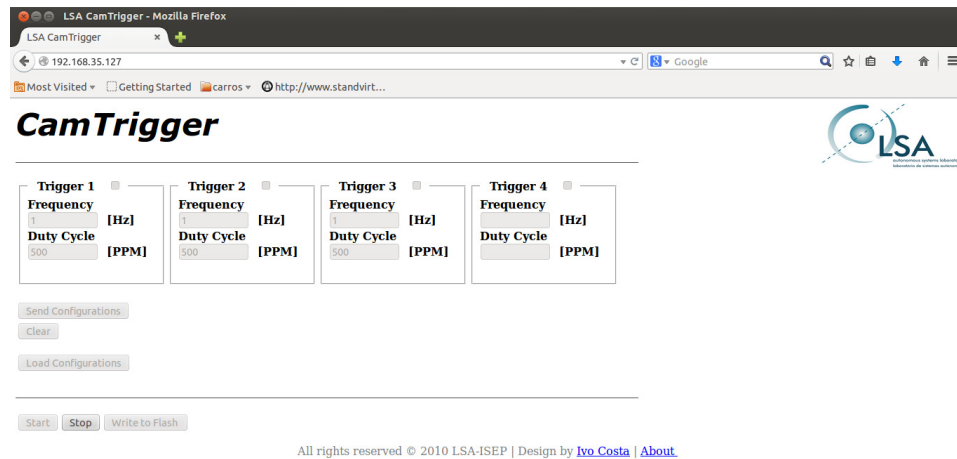


Figura 4.6: Interface de configuração do *trigger*.

Esta página foi intencionalmente deixada em branco.

Capítulo 5

Implementação

5.1 Instalação do sistema

Esta subsecção descreve alguns aspetos relevantes da implementação do sistema de *ground truth* quer no campo de futebol robótico médio, quer no tanque de experiências subaquáticas presentes no laboratório.

5.1.1 Campo de futebol robótico médio

Um dos requisitos da instalação das câmaras no campo de futebol robótico médio consiste em garantir que as camaras têm um FOV (*Field Of View*)forma a que seja possível a captura integral da área do campo. Para tal as câmaras foram instaladas a cerca de 5.5 metros de altura em relação ao chão e fixadas na esteira utilizada para a distribuição elétrica e de rede do laboratório. Esta solução de fixação das câmaras permitiu que fossem instaladas sem que tivesse de ser alterado qualquer tipo de infraestrutura do laboratório. Para tal, apenas foi necessário o desenvolvimento manual de uma estrutura metálica para cada câmara que permitisse a fixação das mesmas na própria esteira tal como demonstrado pela figura 5.1.

Os dados capturados pelas câmaras são transferidos em GigE. Este facto torna-se vantajoso pelo facto de que o acesso à rede interna do laboratório



Figura 5.1: Fixação câmaras campo futebol robótico médio.

é compatível com o mesmo, desta forma a transferência de dados é feita com recurso ao sistema de rede interno do laboratório através dos múltiplos pontos de acesso. Assim, o acesso aos dados das câmaras instaladas pode ser feito de qualquer ponto de acesso à rede do laboratório que suporte o padrão de dados GigE. A alimentação e da transmissão do sinal elétrico para o disparo das câmaras foi elaborada com recurso aos pontos de acesso de rede do LSA. Para tal alguns pontos de acesso, um por cada câmara, foram desabilitados do acesso à rede para que ficassem disponíveis para a função de alimentar e transmitir o sinal de disparo às câmaras, sendo para tal conectados na placa desenvolvida para o efeito, abordada anteriormente na subsecção 4.2.2. Com facilidade os referidos pontos de acesso podem voltar a ser restabelecidos apenas com a troca destes da placa de *trigger* para a rede do LSA. Outra das tarefas desempenhadas no âmbito desta dissertação, relacionadas com o campo de futebol robótico médio, foi a substituição do piso do campo que devido a já ter sido instalado não se encontrava nas condições ideais para o desenvolvimento da dissertação. Este contava já com algumas manchas na alcatifa e as marcações estavam enviesadas e já degradadas, facto este que induziria o sistema em erros. Para tal, foi removido o piso antigo e substituído por um em melhores condições, as marcações foram

efetuadas seguindo as regras do futebol robótico médio para a competição Robocup 2014 [5].

5.1.2 Tanque de experiências subaquáticas

A instalação do sistema de *ground truth* no tanque de experiências subaquáticas envolveu um dos maiores desafios a nível de instalação de hardware, devido ao facto de estas serem instaladas debaixo de água. Para tal foi elaborada uma estrutura estanque com a finalidade de colocação das respetivas câmaras. Esta possui na parte frontal um vidro para possibilitar a captura de imagem. Na parte traseira, este suporta uma estrutura circular, na qual foram inseridas fichas estanques, para transmissão de dados, alimentação e sincronismo. Devido à necessidade de fixar as câmaras e os respetivos recipientes estanques, foi desenvolvido um sistema de fixação que assegura que as respetivas câmaras permanecem fixas. Esta estrutura foi desenvolvida tendo em conta a integridade da infraestrutura do tanque. Nas figuras 5.2 pode ser observada a instalação as câmaras no tanque.

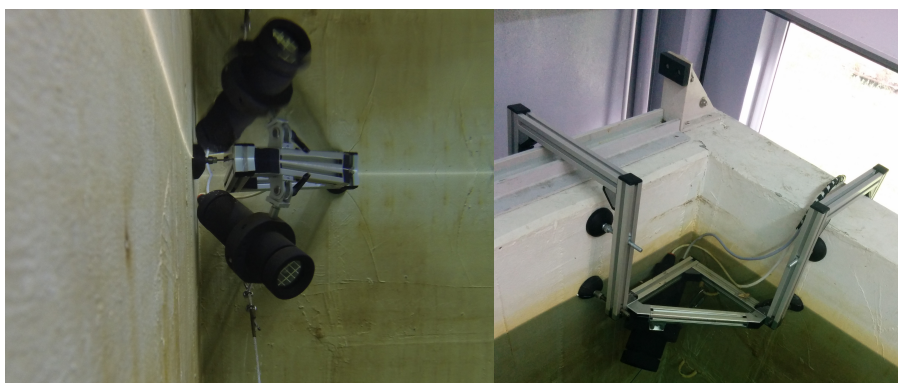


Figura 5.2: **Esquerda** - involucrio estanque para a câmara. **Direita** - Sistema de fixação para as câmaras.

O sistema de disparo e alimentação foi instaladona parte exterior do tanque numa caixa estanque, este pode ser observado pela figura 5.3.



Figura 5.3: **Esquerda** - Sistema de disparo e alimentação das câmaras + switch GigE. **Direita** - Caixa estanque para instalação dos sistema de disparo e alimentação.

5.2 Calibração dos parâmetros intrínsecos

Nesta secção é estudado um método para a calibração dos parâmetros intrínsecos. Tal como abordado anteriormente no subcapítulo 3.2.1 através destes parâmetros, distância focal, coordenadas do centro ponto principal e dos parâmetros da distorção é possível aplicar a projeção perspetiva de forma a transformar as coordenadas de objetos em 3D no referencial da câmara em coordenadas 2D pixels.

5.2.1 Algoritmo de Tsai

Para a calibração dos parâmetros intrínsecos foi aplicado ao algoritmo de Tsai [47]. Este algoritmo estima os parâmetros intrínsecos com base num número considerável de imagens nas quais está presente o padrão de xadrez, estatisticamente quanto mais elevado o numero de imagens melhores tendem a ser os resultados obtidos, considerando que a imagens capturadas focam o padrão de xadrez sem que seja notório e efeito de desfoque provocado pelo movimento do alvo no momento de captura das imagens. Na figura 5.4 podem ser observadas algumas das figuras utilizadas para a calibração dos parâmetros intrínsecos de uma das câmaras utilizadas para esta dissertação.

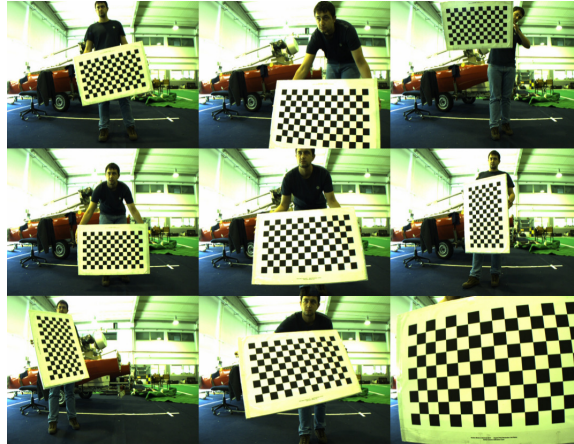


Figura 5.4: Detecção de pontos de interesse - algoritmo de Tsai.

No início da aplicação do método deverá ser dada a indicação do número de quadrados contidos no padrão xadrez, em altura e largura bem como a dimensão dos mesmos para que seja possível estimar a posição 3D dos pontos. A deteção dos pontos na imagem é feita com base num processo semi-automático em que são assinalados manualmente os cantos do padrão de xadrez, desta forma os restantes pontos são estimados com base nestes quatro, tornando assim a deteção mais precisa. Na figura 5.5 pode ser observado um exemplo da deteção dos pontos de fronteira do padrão de xadrez utilizado para a calibração.

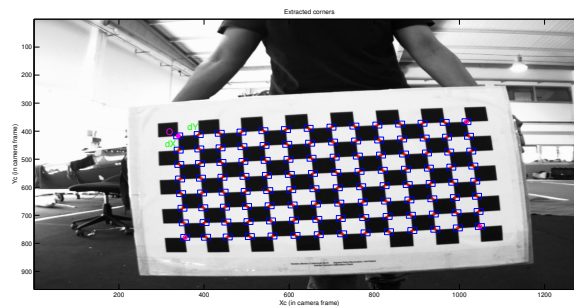


Figura 5.5: Detecção de pontos de interesse - algoritmo de Tsai.

O procedimento de marcação dos cantos do padrão de xadrez deve ser repetido para todas as imagens. O próximo passo para o cálculo dos parâmetros intrínsecos é a avaliação dos erros na estimação feita para cada imagem. Como já referido, para todas as imagens são estimados os pontos 3D do alvo assim como a sua projeção 2D na imagem. Os pontos estimados deverão corresponder aos pontos detetados na imagem. O erro avaliado é a diferença entre os pontos projetados na imagem e os pontos detetados. Nesta fase deverão ser descartadas as informações obtidas das imagens com maior erro para que assim a calibração efetuada seja de melhor qualidade. No processo de calibração são descartadas todas as imagens que apresentem um erro de projeção de qualquer um dos pontos que apresentem erro elevados face às outras projeções. Com a aplicação deste algoritmo foram obtidos erros de calibração entre 0.15 e 0.25 píxeis

5.3 Calibração dos Parâmetros extrínsecos

Neste subcapítulo são abordadas as técnicas utilizadas para a calibração dos parâmetros extrínsecos. Tal como referido no subcapítulo 3.2.2 estes parâmetros definem uma transformação de corpo rígido entre o referencial de um determinado cenário e o próprio referencial da câmara permitindo assim converter as coordenadas do referencial global para o referencial da câmara. Em qualquer um dos casos apresentados foram sempre considerados os parâmetros de distorção da lente obtidos pelo método anteriormente apresentado no subcapítulo 5.2.1. De forma a corrigir esta deformação na imagem foi aplicado o algoritmo apresentado no subcapítulo 3.2.1. Para representação da medidas no referencial global foi considerado o referencial apresentado pela figura 5.6

Os métodos seguidamente apresentados foram aplicados no campo de MSL pelo facto de este cenário ser mais fácil para a realização das experiências. Para o tanque de experiências sub-aquáticas a calibração dos



Figura 5.6: Referencial do campo de futebol robótico médio.

parâmetros intrínsecos foi efetuada com base no método dos 8 pontos.

5.3.1 Método dos 8 pontos

O algoritmo dos 8 pontos [21] é frequentemente utilizado para o cálculo da matriz fundamental. A matriz fundamental permite relacionar pontos de duas imagens, tal como abordado no subcapítulo 3.3.2. Para implementação deste método foi utilizado uma das funcionalidades da mesma ferramenta utilizada para a calibração dos parâmetros intrínsecos[12]. A referida ferramenta está inserida numa biblioteca que disponibiliza um conjunto de funções para o desenvolvimento de programas em Matlab[52]. Com a utilização desta ferramenta é possível obter os parâmetros extrínsecos para cada câmara face ao cenário é necessário ter o conhecimento das seguintes propriedades do sistema:

- Parâmetros intrínsecos, distancia focal, e ponto principal;
- Pontos 3D no referencial global;
- Pontos 2D na imagem, em pixels, correspondentes aos pontos 3D obtidos.

Os pontos 3D podem ser obtidos de diferentes formas, através de medição manual com recurso a fita métrica ou com a utilização de um laser *range finder* de modo a aumentar a precisão das medidas obtidas. Quanto aos pontos nas imagens estes podem ser obtidos igualmente de forma manual, através da marcação nas imagens do pontos de interesse, ou através da utilização algoritmos para deteção de objetos conhecidos que sejam relevantes para a calibração. Como resultado aplicação desta ferramenta para a obtenção dos parâmetros extrínsecos devem ser obtidos os valores da rotação e translação de corpo rígido de forma a ser conhecida a posição de cada câmara no referencial global.

Este método é passível de que sejam cometidos erros sistemáticos pelo utilizador uma vez que este requer interação para a obtenção dos pontos 2D e 3D que na maioria dos casos são obtidos manualmente. esta tarefa pode demorar varias horas e depende diretamente do numero de câmaras constituintes dos sistema.

5.3.2 Método pontos de fuga

Este método permite a obtenção dos parâmetros extrínsecos de cada câmara sem que para tal seja necessária a introdução de um elevado numero de pontos, ou medições do cenário envolvente. Este método destina-se a ser aplicado em cenários estruturados, com linhas paralelas entre si e também com o referencial global tal como no campo de MSL. Tal como ilustrado na figura 5.6, as marcações do campo são paralelas ao eixo do X ou ao eixo do Y , com a exceção das marcações de cantos e do circulo do meio campo. Assim sendo a rotação entre o referencial da câmara da câmara e o referencial global é obtida através da orientação das referidas linhas de marcação do campo. O primeiro passo para a obtenção da orientação de cada câmara face ao campo, passa pela deteção das referidas marcações. A deteção foi elaborada com recurso à transformada de Hough [17], sendo previamente

convertida a imagem para o formato binário de forma a realçar as linhas face ao cenário, desta são obtidas as retas presentes no cenário. A deteção foi feita com precisão de cinco centésimas de grau para a orientação e 0.3 pixels para distancia à origem. Com base nesta informação são formados vários segmentos de reta ou descartados, de forma a refinar os resultados obtidos. Para tal são apenas considerados os segmentos de reta obtidos que têm mais de 12 pixels de comprimento e foram também fundidos os segmentos de reta pertencentes à mesma reta que se encontrassem a uma distancia de 10 ou menos pixels. Estes valores foram definidos com base nos resultados obtidos em diversas experiências que visavam a aplicação do referido método. Na figura 5.7 podem ser observados os segmentos de reta detetados com a aplicação da transformada de Hough à imagem capturada pela câmara esquerda

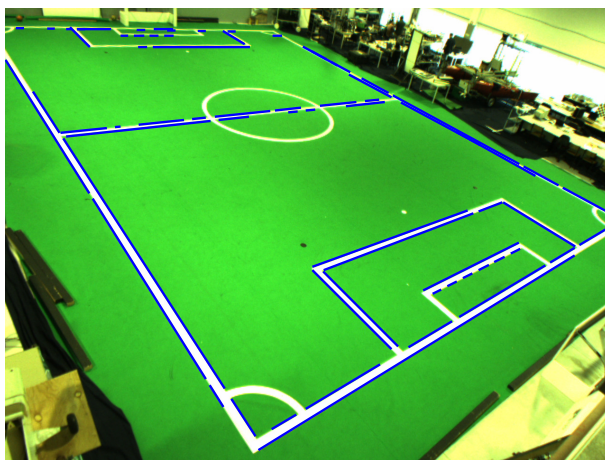


Figura 5.7: Segmentos de reta obtidos recorrendo à transformada de Hough

Após a aplicação da transformada de Hough verificou-se que para a mesma linha do campo foram detetados vários segmentos de reta. De forma a melhorar a qualidade dos segmentos foi elaborado um algoritmo que agrupa os segmentos de reta, pela sua proximidade aos parâmetros θ e ρ no plano da transformada de Hough, e pelo direção da fronteira de cor que os ori-

ginou. Ou seja são agrupados os segmentos que apresentem uma pequena diferença quanto à sua colinearidade e que apresentem a orientação da fronteira de cor na mesma direção. Os segmentos que cumpram as restrições impostas são fundidos num só através da média ponderada pelo tamanho de cada segmento de reta do conjunto formado. Assim são calculados os novos parâmetros, θ' e ρ' , pela equação 5.1:

$$\begin{cases} \theta' = \frac{\sum_{i=1}^{i=n} \theta_i l_i}{\sum_{i=1}^{i=n} l_i} \\ \rho' = \frac{\sum_{i=1}^{i=n} \rho_i l_i}{\sum_{i=1}^{i=n} l_i} \end{cases} \quad (5.1)$$

em que n é o numero de linhas do conjunto e l_i é comprimentos de um determinado segmento de reta do conjunto. Desta forma é possível aumentar a qualidade das retas obtidas de forma a que estas aproximem o mais possível as com as linhas na imagem, tirando proveito da precisão conseguida na deteção dos segmentos, assim como da restrição colocada quanto à distancia mínima para a união dos diferentes segmentos e ao tamanho mínimo dos segmentos considerados. Na figura 5.8 é demonstrado observado um exemplo das retas estimadas com base nos segmentos de reta detetados pela transformada de Hough, são também representados, a verde e a vermelho as interseções das retas que originam cada um dos pontos de fuga.

Uma vez conhecida a orientação de todas as retas detetadas, assim como o comprimento dos segmentos que as originaram, podemos então estimar os pontos de fuga. A estimação dos pontos de fuga é feita com base num algoritmo de minimização, em que o ponto de fuga obtido é aquele se se encontrar à mínima distância de todas as retas. Na equação 5.2 é apresentada a equação para a minimização dos pontos de fuga com base nas retas

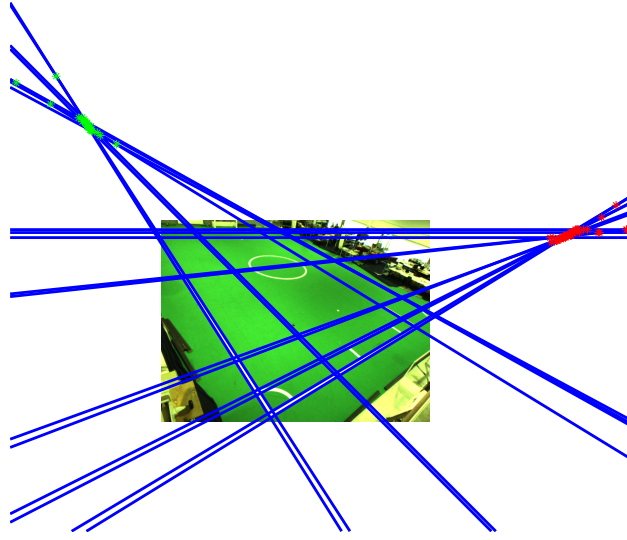


Figura 5.8: Demonstração das retas obtidas pela aplicação do algoritmo melhoramento da transformada de Hough.

anteriormente estimadas.

$$\min_x \sum_{k=1}^K (w_k l_k^T X^t)^2 \quad (5.2)$$

em que K é o numero de retas obtidas para obtenção do ponto de fuga, w_k é o comprimento dos segmentos, $l_k^T = \cos \theta_k \sin \theta_k$, $-\rho$ correspondem à informação dos segmentos de reta e $X^T = [u_x, v_x, 1]$ é o ponto de fuga em coordenadas homogêneas.

De forma a minimizar o processo de estimação do ponto de fuga é inicializado o processo de calculo com a média ponderada das interseções das retas.

Com base na projeção perspétiva inversa apresentada no subcapítulo 3.7, é calculada a orientação das linha do campo face ao referencial da câmara, este calculo foi efetuado com base na equação 5.3

$$d_i = \begin{bmatrix} d_{xi} \\ d_{yi} \end{bmatrix} = \begin{bmatrix} \frac{v_{xi} - c_{xi}}{f_{xi}} \\ \frac{v_{yi} - c_{yi}}{f_{yi}} \end{bmatrix} \quad (5.3)$$

Em que d_1 é o vetor que define orientação da câmara relativamente às linhas detetadas no plano do mundo, c_{xi} , c_{yi} , f_{xi} e f_{yi} dizem respeito à distância focal e as coordenadas do ponto principal, respetivamente, e v_{xi} e v_{yi} são as coordenadas em x e y dos pontos de fuga estimados. De forma a definir a matriz de rotação de corpo rígido foi assumido que a orientação de obtida por cada ponto de fuga origina uma coluna da referida matriz, a terceira coluna foi originada com base no produto vetorial dos dois primeiros vetores que forma as duas primeiras colunas da matriz e serem representados em coordenadas homogéneas. Todos os vetores devem ter comprimentos unitário. Assim a matriz de rotação é formada pela expressão 5.4:

$$R = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_1 \times r_2 \end{bmatrix}$$

$$R = \begin{bmatrix} \frac{d_1}{|d_1|} & \frac{d_2}{|d_2|} & \frac{d_1 \times d_2}{|d_1 \times d_2|} \end{bmatrix} = \begin{bmatrix} \frac{d_{x1}}{|d_1|} & \frac{d_{x2}}{|d_2|} & \frac{(r_1 \times r_2)_x}{|(r_1 \times r_2)|} \\ \frac{d_{y1}}{|d_1|} & \frac{d_{y2}}{|d_2|} & \frac{(r_1 \times r_2)_y}{|(r_1 \times r_2)|} \\ \frac{1}{|d_1|} & \frac{1}{|d_2|} & \frac{(r_1 \times r_2)_z}{|(r_1 \times r_2)|} \end{bmatrix} \quad (5.4)$$

De forma a obter a rotação das câmaras face ao campo faltam ainda desambiguar quais dos pontos de fuga formam os vetores com a mesma orientação dos eixos X e Y e também o sentido do eixo Z . Para tal será primeiramente verificado o sentido do eixo dos ZZ . Para tal é assumido que a câmara de forma a capturar o campo tem na sua componente de orientação *pitch* negativa relativamente ao campo. Quer isto dizer que o sentido do eixos dos ZZ da câmara está sempre no sentido inverso ao eixo ZZ do campo. Para correlação correta dos pontos de fuga com as direções dos eixos do plano $X - Y$ do campo é aplicada aos segmentos de reta uma projecção perspectiva inversa aos segmentos de reta detetados pela transformada de Hough transformando-os assim do plano da imagem, para o referencial do campo, sendo estes posteriormente rodados pela rotação obtida até ao momento. Esta transformação é feita com base na equação

5.5:

$$\lambda P'_w = R^{-1} K^{-1} P_{img} \quad (5.5)$$

Em que P'_w corresponde segmentos de reta obtidos transformados para o plano do campo, em que a câmara está na origem do plano $X - Y$ do campo, a menos da escala λ . A matriz R^{-1} representa a matriz de rotação conhecida até ao momento a qual efetua a rotação dos segmentos de reta do referencial do campo da câmara para o referencial do campo, K^{-1} representa a matriz inversa dos parâmetros intrínsecos e P_{img} representa os segmentos de reta detetados pela transformada de Hough.

Ao aplicar esta transformação podemos avaliar qual a orientação do lado maior do campo, este deve estar paralelo com o eixo X . Caso esta situação não se verifique é assumido que pontos de fugas considerados para a composição da matriz de rotação, primeira e segunda colunas estão trocados, ou seja, foi considerado o ponto de fuga que contém a orientação de X para definir a orientação do eixo Y e vice versa.

De forma a descobrir posteriormente o sentido de cada um dos referenciais, X e Y é marcado manualmente um ponto na imagem no semi-plano correspondente ao primeiro quadrante do plano $X - Y$ do referencial do campo e posteriormente estimado um vetor desde o centro do campo obtido pela transformação efetuada pela equação 5.5 até ao ponto marcado. A direção deste vetor deste vetor relativamente à direção conhecida para a o eixo dos Y fornece a informação sobre a informação real do mesmo eixo, ou seja, o vetor formado deve estar sempre na mesma direção que o eixo Y , caso contrario a direção deste eixo é contrária à realidade. Na aplicação deste procedimento são consideradas as restrições de ortonormalidade de forma a obter a solução correta para o problema. Pela transformação efetua, formulada pela equação 5.5, a translação entre a câmara e o referencial do mundo é dada pela posição do centro do campo. Desta forma a câmara são

obtidos os parâmetros extrínsecos para cada uma das câmaras a menos do fator escala λ . Esta é calculada dada pela divisão entre uma medida real do campo, obtida com base numa medição manual, no presente caso largura, e a medida obtida aquando da transformação efetuada, sendo desta forma obtidos os parâmetros extrínsecos correspondentes à transformação de corpo rígido para a câmara relativamente ao campo.

5.3.3 Decomposição de homografia

Nesta secção será abordada a implementação do método de decomposição de homografia apresentado no subcapítulo 3.5. Este método distingue-se dos anteriores devido ao facto de pode ser aplicado em cenários em que seja predominante pelo menos um plano, tal como no caso do campo de MSL. A primeira fase para a aplicação deste método é a obtenção da homografia planar que relaciona os pontos das imagens capturadas por cada câmara constituintes de um par *stereo*. A homografia pode ser estimada com base em algoritmos, como o apresentado por Peter Kovesi[51], em que esta é calculada com base em dois conjuntos de pontos ao mesmo ponto no mundo observado pelas duas câmaras de um par *stereo*. O erro cometido na implementação é tanto maior como o erro cometido na deteção dos pontos e associação dos mesmos. De forma a minimizar os erros foram utilizados objetos fáceis de detetar. Neste caso foi utilizado uma fonte emissora de luz, assim os pontos de interesse são os pontos com mais intensidade presentes numa sequencia de imagens. A primeira fase do processo de deteção do ponto luminoso é conversão da imagem a cores para tons de cinza, obtendo assim apenas um espaço de cores correspondente à intensidade da cor. Posteriormente é aplicado um *threshold* de forma a eliminar o ruído existente na imagem, obtendo assim apenas uma nuvem de pontos que corresponde a localização do ponto de luz na imagem. Destes é considerado o ponto de maior intensidade como o ponto da fonte luminosa. Assim são obtidos um

conjunto de pontos para cada câmara, correspondentes à mesma localização espacial, para que seja efetuado o cálculo da transformação que os relaciona, a homografia. A transformação calculada relaciona os pontos de uma das câmaras, câmara principal, para a outra câmara constituinte do par *stereo*. Uma vez definida a homografia é efetuado a decomposição pelo método de decomposição de homografia apresentado no subcapítulo 3.5 que tem como principal vantagem a obtenção do vetor normal ao plano em que a homografia foi gerada. A aplicação deste método gera quatro possíveis soluções, sendo que duas dessas são eliminadas pela restrição física de que a câmara apenas podem capturar imagem para a sua frente, ou seja, qualquer objeto capturado pela mesma terá como posição no referencial da câmara, $Z > 0$. Assim Considerando esta restrição e que o vetor normal ao plano de homografia está nas coordenadas da câmara principal, apenas são consideradas as hipóteses que gerem o vetor normal ao plano de homografia com $Z > 0$. Restando duas das quatro soluções obtidas são considerados os pontos capturados para a obtenção da homografia em cada câmara e as duas possíveis soluções para os parâmetros extrínsecos, R e λT , de cada solução obtida. Com base nestes parâmetros é feita a triangulação *stereo* dos referidos pontos. Assim a solução a considerar é aquela que gera os pontos em 3D num plano normal com o vetor normal da respetiva solução. Desta forma são obtidos os parâmetros de corpo rígido relativo entre as câmaras.

Para a obtenção da rotação de cada câmara face ao campo é posteriormente colocada a fonte luminosa em duas posições conhecidas do campo segundo o eixo dos Y , P_{y1} e P_{y2} . Obtidas as as coordenadas do referidos pontos em cada câmara, p_{y1} e p_{y2} , é possível efetuar a triangulação dos mesmos obtendo-os assim em coordenadas 3D no referencial da câmara principal, P'_{y1} e P'_{y2} . Partindo do vetor V'_{Y} , formado pelos dois pontos triangulados, em 3D, é encontrada a direção do eixo dos Y na primeira câmara. A direção do eixo dos Z é dada pelo vetor normal ao plano da homografia, uma vez

que este plano é paralelo ao campo. A direção do eixo dos X é formada com base nos dois últimos vetores referidos, isto é, a direção deste vetor é aquela que seja simultaneamente normal com os vetores que formam as direções dos eixos Z e Y . Assim a rotação é obtida pela equação 5.6

$$\begin{aligned} R_{c1} &= [r_1 \quad r_2 \quad r_3] \\ r_3 &= N \quad r_2 = P_{y2} - P_{y1} \quad r_1 = r_2 \times r_3 \end{aligned} \quad (5.6)$$

em que R_{c1} é a rotação absoluta entre a câmara principal e o campo. A rotação da câmara 2, R_{c2} , é obtida com base na rotação absoluta da primeira câmara e na rotação relativa entre as duas obtida da decomposição de homografia, R_H . Assim a rotação da segunda câmara pela equação 5.7

$$R_{c2} = R_H R_{c1} \quad (5.7)$$

Para que sejam obtidos os parâmetros que definem as transformações de corpo rígido de cada câmara é ainda necessário obter as translações de cada uma delas face ao campo e o fator de escala. O fator de escala é calculado com base no tamanho do vetor V'_{YY} e distancia real a que estão os pontos no campo, $\|V_{YY}\|$. Desta forma o fator de escala λ é definido pela equação 5.8

$$\lambda = \frac{\|V_{YY}\|}{\|V'_{YY}\|} \quad (5.8)$$

De forma a obter a translação é colocada a fonte luminosa no centro do campo e efetuada a triangulação stereo para o referido ponto, obtendo assim $P'_{0,0}$. Assim é obtida a translação inicialmente para a primeira câmara, a menos da escala, e posteriormente para a segunda através da equação 5.9, em que T_H representa a translação relativa entre as câmaras.

$$\begin{aligned} T_{c1} &= \lambda P'_{0,0} \\ T_{c2} &= [R_H \lambda T_H] P'_{0,0} \end{aligned} \quad (5.9)$$

Capítulo 6

Resultados

No presente capítulo são apresentados os resultados obtidos pela aplicação dos diferentes métodos de calibração dos parâmetros extrínsecos para câmaras. Com base na representação do sistema para o campo de futebol robótico médio demonstrado anteriormente na Figura 4.2 as câmaras i e j serão referidas como câmara esquerda e câmara direita respetivamente. Para a obtenção dos presentes resultados foram considerados pontos no mundo, em 3D, apresentados pela tabela 6.1 em metros. Os referidos pontos são utilizados em todas as experiências de forma a uniformizar os resultados. Os mesmos, denominados pontos de controlo 3D daqui em diante, foram obtidos com base nas medidas oficiais de um campo de futebol robótico médio [5]. Foram também obtidos, manualmente, os pontos correspondentes aos pontos apresentados pela tabela 6.1 para cada uma das câmaras, os mesmos serão chamados de pontos de controlo 2D. Na Figura 6.1 podem ser observados os pontos correspondentes para a câmara da esquerda do campo. Estes foram extraídos manualmente mas com suporte à transformada de Hough e são chamados adiante de pontos de controlo 2D. Seguidamente serão apresentados os erros para cada calibração para cada um dos métodos abordados. Os erros são obtidos com base na projeção dos pontos de controlo 3D e comparados com os pontos de controlo 2D para cada câmara. São também

x	y	z
9.000	6.063	0.000
9.000	5.938	0.000
9.000	3.313	0.000
9.000	3.188	0.000
9.000	1.813	0.000
9.000	-1.813	0.000
9.000	-3.188	0.000
9.000	-3.313	0.000
8.188	1.813	0.000
8.188	-1.81	0.000
6.688	3.313	0.000
6.688	-3.31	0.000
0.063	5.938	0.000
0.063	2.070	0.000
0.06	-2.050	0.000
0.063	-5.938	0.000
-0.063	5.938	0.000
-0.062	2.070	0.000
-0.062	-2.050	0.000
-0.062	-5.947	0.000
-6.688	3.313	0.000
-6.688	-3.313	0.000
-8.188	1.8135	0.000
-8.188	-1.81	0.000
-9.000	6.063	0.000
-9.000	5.938	0.000
-9.000	3.313	0.000
-9.000	1.813	0.000
-9.000	-1.813	0.000
-9.000	-3.313	0.000
-9.000	-5.938	0.000
-9.000	-6.063	0.000

Tabela 6.1: Coordenadas dos pontos no referencial do campo utilizados para a obtenção de resultados em metros

apresentados os erros obtidos com base na triangulação dos pontos 2D para cada câmara e posteriormente comparados com os pontos de controlo 3D. Os referidos erros serão representados pela letra ϵ , por exemplo é representado por ϵ_x para um determinado erro na sua componente em x, ou $\|\epsilon\|$ para a



Figura 6.1: Exemplo de pontos na imagem para a obtenção dos resultados.

norma do erro de uma determinada calibração.

6.1 resultados para o método 8 Pontos

Para a obtenção da calibração pelo referido método foram extraídos 24 pontos nas coordenadas do referencial do campo, assim como os respectivos pontos 2D em cada uma das câmaras. Após a aplicação do processo de calibração são utilizados os pontos controlo de forma a avaliar a qualidade da calibração pelo método dos 8 pontos. De forma a ser possível avaliar os erros cometidos no processo de calibração foram calculados os erros de reprojeção. Estes são obtidos a partir da projeção perspectiva dos pontos de controlo em três dimensões para cada câmara recorrendo aos resultados de calibração, rotação e translação, obtidos. O resultado do erro de reprojeção é obtido pela comparação entre a projeção e os pontos de controlo 2D para cada câmara.

Nos gráficos apresentados na Figura 6.2 podem ser observados os erros de reprojeção nos eixos x , ϵ_x , no eixo y , ϵ_y e a norma do erro, $\|\epsilon\|$, para a câmara da esquerda.

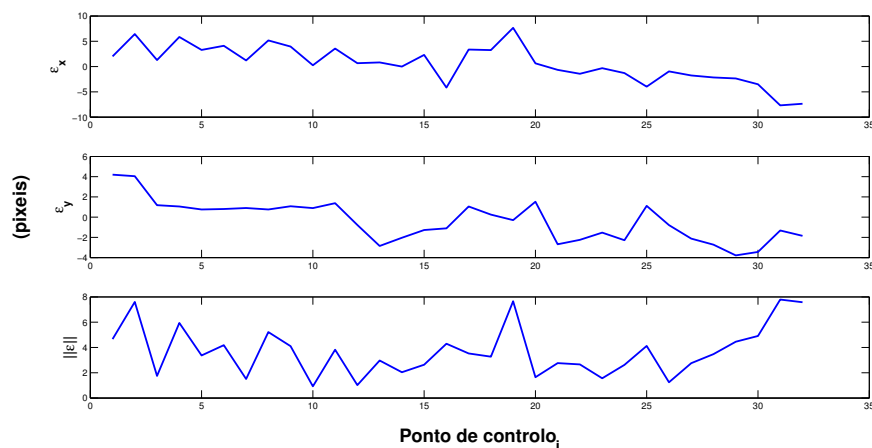


Figura 6.2: Erro de reprojeção para a câmara esquerda pelo algoritmo 8 pontos.

De seguida nos gráficos apresentados na figura 6.3 podem ser observados os erros de reprojeção nos eixos x, ϵ_x , no eixo y, ϵ_y assim como a norma, $\|\epsilon\|$, do erro para a câmara da direita.

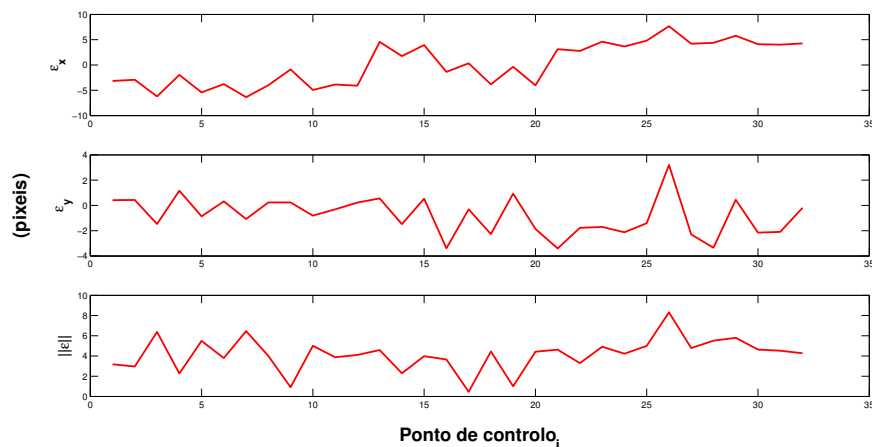


Figura 6.3: Erro de reprojeção para a câmara direita pelo algoritmo 8 pontos.

Na Figura 6.4 podem ser observados os erros de triangulação stereo para a calibração obtida com o método dos 8 pontos. Estes erros foram obtidos com base nos pontos de controlo previamente obtidos para cada uma das câmaras tal como já referido.

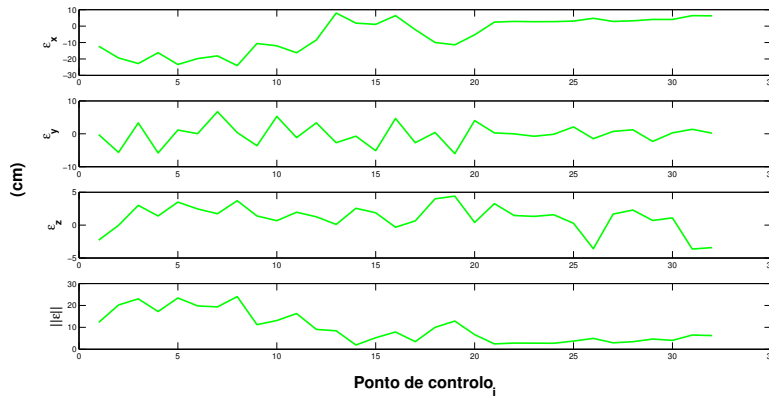


Figura 6.4: Erro de triangulação algoritmo 8 pontos.

Os pontos de controlo foram retirados da esquerda para a direita desde o fundo do campo até ao principio, ou seja os pontos de controlo com índices mais baixos são aqueles que estão mais longe do par *stereo* do *ground truth*.

6.2 Resultados para o método dos pontos de fuga

De forma a caracterizar a calibração para o método dos pontos de fuga foram capturadas 64 imagens por cada câmara do campo de futebol robótico médio. Estas foram capturadas com diferentes condições de iluminação exterior, durante o dia e noite, e o seu processamento foi feito com parâmetros diferentes para a deteção das linhas, diferente *threshold* na conversão da imagem de RGB para binário. Para além da iluminação proveniente do exterior pode ser avaliado o efeito cintilação da iluminação do campo de futebol robótico médio na calibração para o método em questão, uma vez que estas foram adquiridas a uma frequência de 7Hz de forma a capturar *frames* em diferentes amplitudes da rede elétrica. Desta forma cada uma das 64 imagens geram o mesmo numero de calibrações para os parâmetros extrínsecos de cada câmara, desta forma é possível analisar a precisão do método em diferentes condições de iluminação. De forma a caracterizar os erros foram utilizados os 32 pontos de controlo em duas dimensões para cada

câmara e os correspondentes em três dimensões, referidos anteriormente no presente capítulo. Na Figura 6.5 pode ser observada uma das imagens capturada pela câmara esquerda com a reprojeção de cada ponto de controlo em três dimensões, segundo as 64 calibrações obtidas.

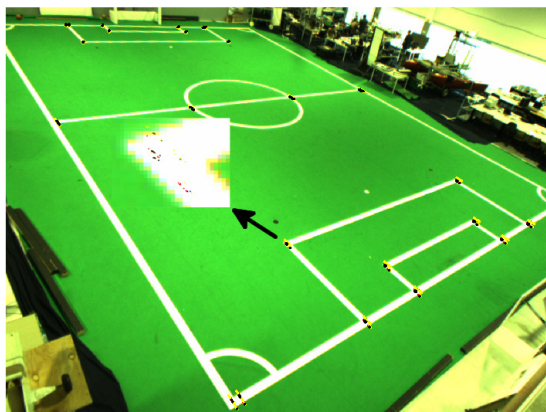


Figura 6.5: Erro de reprojeção câmara esquerda.

Na Figura 6.6, pode ser observada a dispersão dos pontos de controlo reprojctados para a imagem capturada pela câmara esquerda em pixels, para cada uma das calibrações obtidas. Na referida figura cada cor representa o erro de reprojeção para dois conjuntos distintos de imagens, com dois parâmetros de deteção diferentes para cada conjunto. Desta figura pode ser observado que o erro de projeção para a câmara em questão toma valores mais elevados para a projeção feita no eixo x da imagem embora estes sejam também significativos para o eixo dos YY. Na referida figura cada cor representa o erro de reprojeção para dois conjuntos distintos de imagens, com dois parâmetros de deteção diferentes para cada conjunto.

Seguidamente na Figura 6.7 é apresentado o histograma para o erro em pixels para a reprojeção dos pontos de controlo para a câmara esquerda, segundo a componente x e y do referido erro, referentes à Figura 6.6. Deste histograma pode ser observado que dos cerca de 2000 pontos avaliados mais

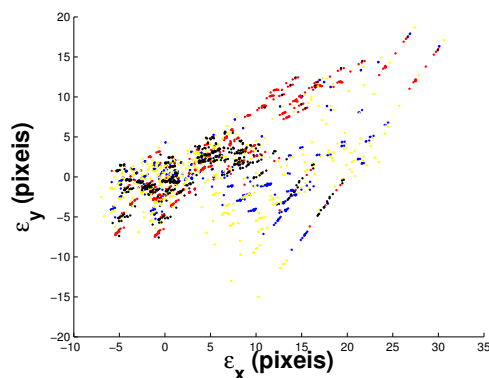


Figura 6.6: Dispersão do erro para a câmara esquerda.

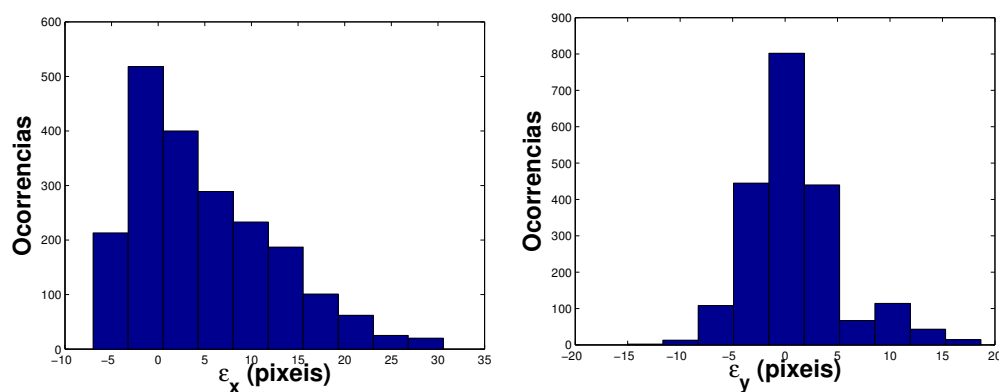


Figura 6.7: Esquerda - erro de projeção em x para a câmara esquerda. Direita - erro de projeção em y para a câmara esquerda.

de 50% dos mesmos tomam valores entre -5 a 5 pixels de erro para a componente em x. De forma semelhante à projeção feita para o eixo dos x, para o caso do eixo do y, o erro varia entre -5 a 5 pixels para 80% das ocorrências, sendo que cerca de 50% desses estão em torno de 0.

Na Figura 6.8 é apresentado o histograma para a norma do erro da projeção feita para a câmara esquerda, o mesmo é relativo aos histogramas apresentados na figura 6.7 à direita e à esquerda.

De forma semelhante aos resultados apresentados para a câmara esquerda seguidamente serão apresentados os resultados de calibração pelo método dos pontos de fuga para a câmara direita. Na Figura 6.9 pode ser

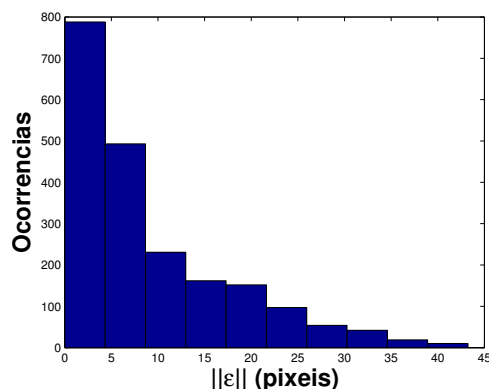


Figura 6.8: Norma do erro de reprojeção para a esquerda.

observada uma imagem capturada pela câmara direita com a reprojeção dos de controlo em três dimensões, segundo as calibrações obtidas para a referida câmara. Nesta imagem pode ser observado que os pontos projetados

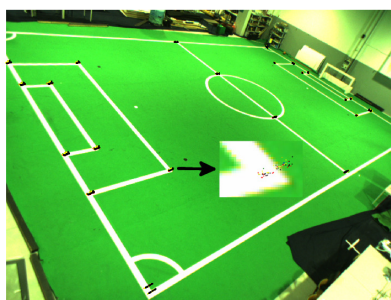


Figura 6.9: Reprojeção câmara para direita.

apresentam um desvio mais significativo na componente x do referencial do campo, tal como acontece para a câmara esquerda.

Na Figura 6.10, pode ser observada a dispersão dos pontos de controlo reprojados para a imagem capturada pela câmara direita em pixels, para cada uma das calibrações obtidas. Na referida figura cada cor representa o erro de reprojeção para dois conjuntos distintos de imagens, com dois parâmetros de deteção diferentes para cada conjunto.

Seguidamente na Figura 6.11 é apresentado o histograma para o erro em

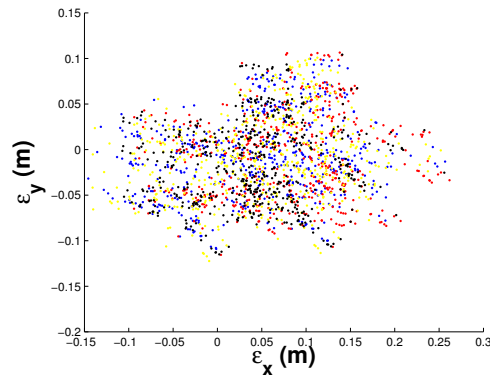


Figura 6.10: Dispersão do erro para a câmara direita.

pixels para a projeção dos mesmos pontos de controlo, neste caso para a câmara direita. Este é apresentado segundo a componente x e y.

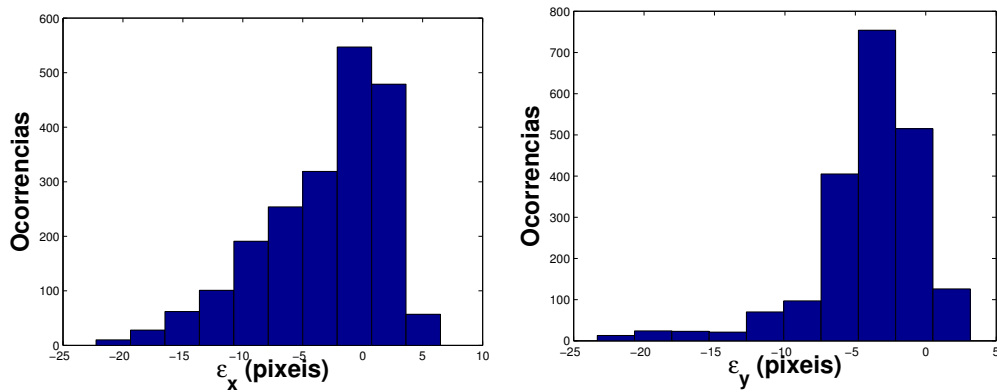


Figura 6.11: Esquerda-Erro de de reprojeção em x para a câmara direita. Esquerda-Erro de de reprojeção em y para a câmara direita

Na Figura 6.12 é apresentado resultado para a norma do erro para as projeções calculadas para a câmara direita.

De seguida são apresentados os resultados obtidos para a triangulação stereo com as calibrações obtidas com o modo dos pontos de fuga. A partir do erro da triangulação *stereo* é possível avaliar a calibração individual de cada uma das câmaras, assim como o conjunto das calibrações obtidas para ambas e a forma como o conjunto das duas calibrações afeta a triangulação dos pontos. Na Figura 6.13 observa-se a dispersão dos pontos de estimados

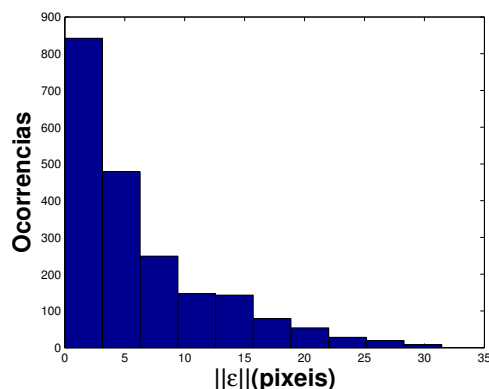


Figura 6.12: Norma do erro de reprojeção para a câmara direita.

na aplicação da triangulação *stereo*, com base nos pontos de controlo em duas dimensões, para cada câmara. Cada uma das cores representa o erro de reprojeção para os quatro conjuntos de imagens, com iluminação diferente e com parâmetros de deteção de linhas diferentes.

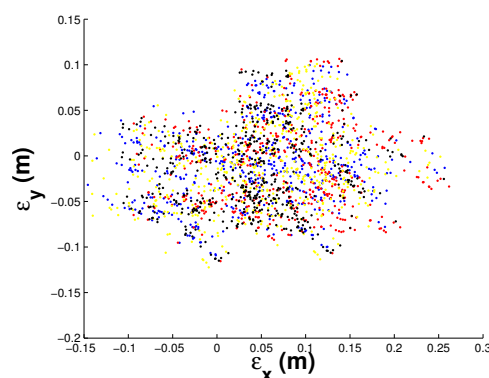


Figura 6.13: Dispersão do erro de triangulação stereo.

Na Figura 6.14 é apresentado o histograma para a componente x do erro de triangulação.

Do referido histograma pode-se verificar um desvio do erro na direção positiva do x, querendo isto dizer que para os pontos triangulados os valores obtidos são inferiores ao dos pontos de controlo, indicando assim um desvio negativo nos pontos triangulados tal como pode ser observado pelas imagens

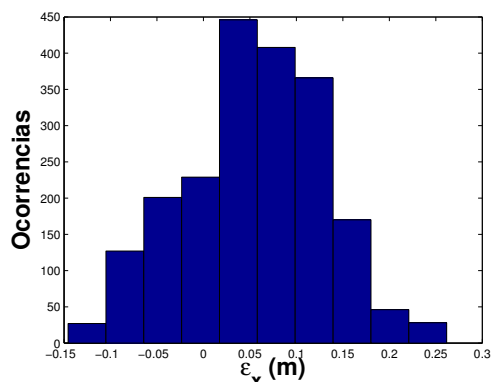


Figura 6.14: Erro de triangulação stereo, componente x.

6.5 e 6.9

Seguidamente, na Figura 6.15 é apresentado o histograma para o erro de triangulação para a componente y. Este apresenta-se centrado em 0, com um pequeno desvio para componente negativa. Na Figura 6.16 é apresentado

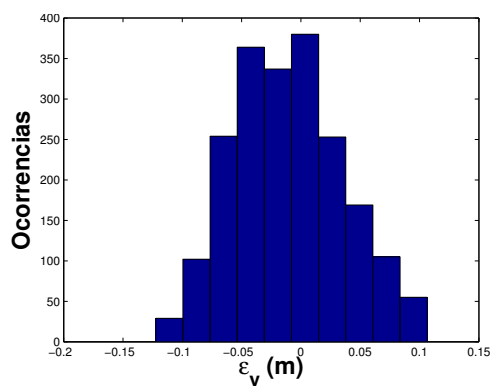


Figura 6.15: Erro de triangulação stereo, componente y.

o histograma para a componente z do erro de triangulação.

Abaixo, na Figura 6.17 é apresentado o histograma para a norma do erro de triangulação para os 3 eixos do referencial do mundo, relativo ao apresentado pelas figuras 6.14, 6.15 e 6.16.

Nos histogramas apresentados para a triangulação *stereo* é possível observar um *offset* quanto aos valores de maior ocorrência em relação a

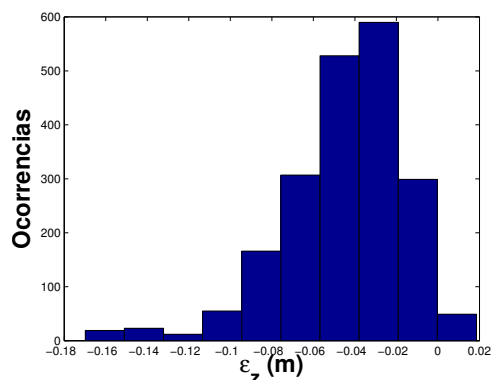


Figura 6.16: Erro de triangulação stereo, componente z.

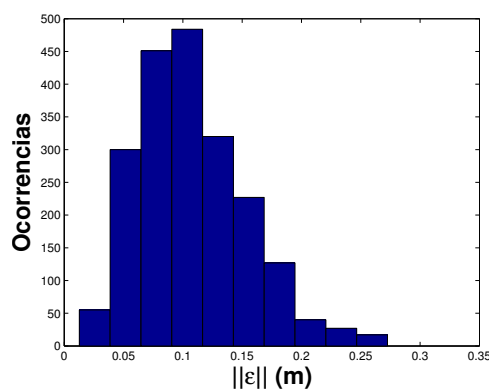


Figura 6.17: Norma erro de triangulação stereo.

0. Mas em qualquer um dos histogramas pelo menos 50% das ocorrências verificam-se no intervalo de -5 a 5 cm.

De seguida é apresentada a Figura 6.18 a representação dos pontos 3D obtidos com recurso à triangulação stereo. Do referido histograma, 6.18, é possível avaliar o conjunto de calibrações obtidas para o método dos pontos de fuga através do erro de triangulação cometido com a utilização dos pontos de controlo 2D para cada câmara face aos pontos de controlo em 3D. Deste histograma pode se verificar que para cerca de metade dos pontos de controlo avaliados o erro cometido é inferior a 15cm. Os restantes pontos de controlo estão distribuídos até aos 45cm de erro de forma decrescente, dos

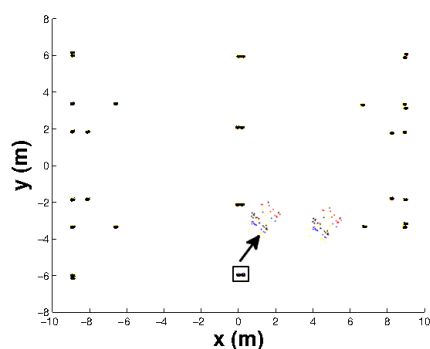


Figura 6.18: Dispersão dos pontos 3D após triangulação.

quais apenas cerca de 3% apresentam erros superiores a 30cm. Na tabela 6.2 são apresentados os resultados para a calibração com recurso ao método dos pontos de fuga. Os resultados obtidos para a precisão do método são obtidos relativamente às 64 calibrações obtidas para cada câmara. Aqui é apresentada a média das calibrações e a sua variação com o desvio padrão face à respetiva média. Quanto aos resultados da exatidão estes são obtidos pela diferença média relativamente ao método dos 8 pontos. Todos os resultados apresentados dizem respeito aos parâmetros de rotação e translação para as duas câmaras, direita e esquerda.

6.3 Resultados para o método de decomposição de homografia

Na presente secção são apresentados os resultados obtidos para o método de calibração com recurso à decomposição de homografia apresentado na secção 3.5. Os resultados obtidos correspondem a três experiências distintas em que para a obtenção da homografia foi movimentada uma fonte luminosa em três percursos aleatórios de forma a cobrir a maior parte do campo de visão do sistema de câmaras, com tempos de exposição diferentes. As três experiências realizadas são apresentadas nos gráficos das figuras baixo com

6.3. Resultados para o método de decomposição de homografia Capítulo 6

	Câmara esquerda	Câmara direita
Rotação	$\psi=-128.282\pm 0.017^\circ$ $\theta=-10.594\pm 0.161^\circ$ $\phi=-125.407\pm 0.211^\circ$	$\psi=-40.969\pm 0.068^\circ$ $\theta=-17.912\pm 0.128^\circ$ $\phi=-126.796\pm 0.127^\circ$
Translação	$x=-1.620\pm 0.062\text{m}$ $y=-3.744\pm 0.063\text{m}$ $z=14.415\pm 0.090\text{m}$	$x=1.522\pm 0.046\text{m}$ $y=-3.014\pm 0.010\text{m}$ $z=14.415\pm 0.087\text{m}$
Rotação	$\psi=-0.003^\circ$ $\theta=0.003^\circ$ $\phi=0.003^\circ$	$\psi=0.004^\circ$ $\theta=-0.001^\circ$ $\phi=-0.002^\circ$
Translação	$x=-0.071\text{m}$ $y=-0.013\text{m}$ $z=0.026\text{m}$	$x=0.045\text{m}$ $y=0.069\text{m}$ $z=-0.041\text{m}$

Tabela 6.2: Comparação dos resultados para os parâmetros extrínsecos do método dos pontos de fuga face ao método dos 8 pontos

cores diferentes, azul, vermelho e verde. A Figura 6.20 apresenta os erros de reprojeção dos pontos de controle em três dimensões para as calibrações obtidas nas três experiências. Estes são apresentados para a componente x e y da imagem e também a norma do erro para a câmara esquerda.

Para a câmara da esquerda com a aplicação do método de decomposição de homografia pode observar-se que os erros de projeção obtidos variam em torno dos 5 pixels de erro com algumas variações sistemáticas para alguns dos pontos de controle. De seguida na Figura 6.20, são apresentados os erros de reprojeção dos pontos de controle em três dimensões para a componente x e y da imagem assim como a norma do erro. Igualmente para a câmara da direita as variações dos erros de projeção variam em torno dos 5 pixels de erro, apresentado também flutuações semelhantes para os pontos de controle em torno dos pontos com os índices 26, 27 e 28. Na abaixo 6.21, são apresentados os resultados da triangulação *stereo* para as calibrações obtidas pelo método de decomposição de homografia. Na tabela 6.3 são apresentados os resultados de calibração com recurso ao método de decomposição. De forma similar aos resultados apresentados para o método dos pontos de fuga,

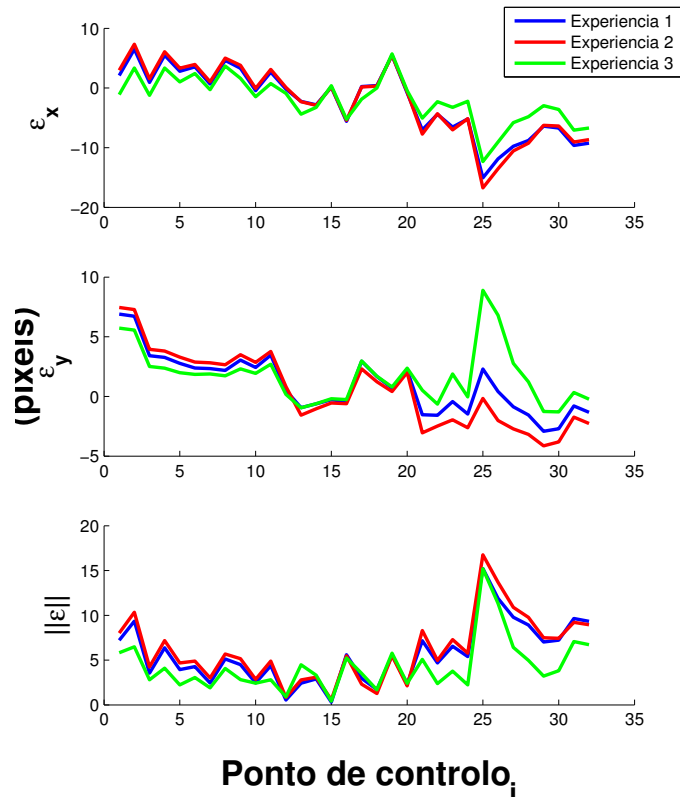


Figura 6.19: Erro de reprojeção câmara esquerda para 3 experiencias.

pela tabela 6.2 os resultados aqui apresentados dizem respeito à precisão do método assim como à exatidão do mesmo. Para obtenção da precisão os resultados apresentados dizem respeito à média das 3 calibrações obtidas e a sua variação com o desvio padrão face à referida média. Quanto aos resultados da exatidão estes são obtidos pela diferença média relativamente ao método dos 8 pontos. Todos os resultados apresentados dizem respeito aos parâmetros de rotação e translação para as duas câmaras, direita e esquerda.

6.3. Resultados para o método de decomposição de homografia Capítulo 6

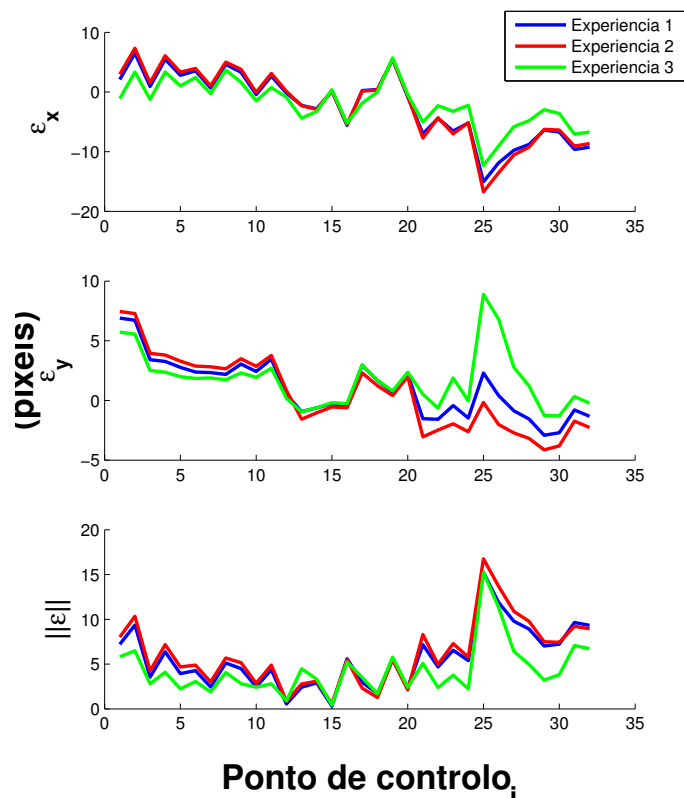


Figura 6.20: Erro de reprojeção câmara direita para 3 experiencias.

	Câmara esquerda	Câmara direita
Precisão Rotação	$\psi=-128.237\pm 0.073^\circ$ $\theta=-10.619\pm 0.107^\circ$ $\phi=-125.631\pm 0.088^\circ$	$\psi=-40.617\pm 0.072^\circ$ $\theta=-17.794\pm 0.126^\circ$ $\phi=-127.067\pm 0.034^\circ$
Precisão Translação	$x=-1.653\pm 0.000\text{m}$ $y=-3.770\pm 0.000\text{m}$ $z=14.447\pm 0.000\text{m}$	$x=1.576\pm 0.000\text{m}$ $y=-2.988\pm 0.000\text{m}$ $z=14.486\pm 0.000\text{m}$
Exatidão Rotação	$\psi=-0.064^\circ$ $\theta=0.204^\circ$ $\phi=-0.075^\circ$	$\psi=0.104^\circ$ $\theta=-0.395^\circ$ $\phi=-0.023^\circ$
Exatidão Translação	$x=-0.004\text{m}$ $y=-0.005\text{m}$ $z=-0.015\text{m}$	$x=0.013\text{m}$ $y=0.024\text{m}$ $z=0.0004\text{m}$

Tabela 6.3: Comparação dos resultado para os parâmetros extrínsecos do método de decomposição de homografia face ao método dos 8 pontos

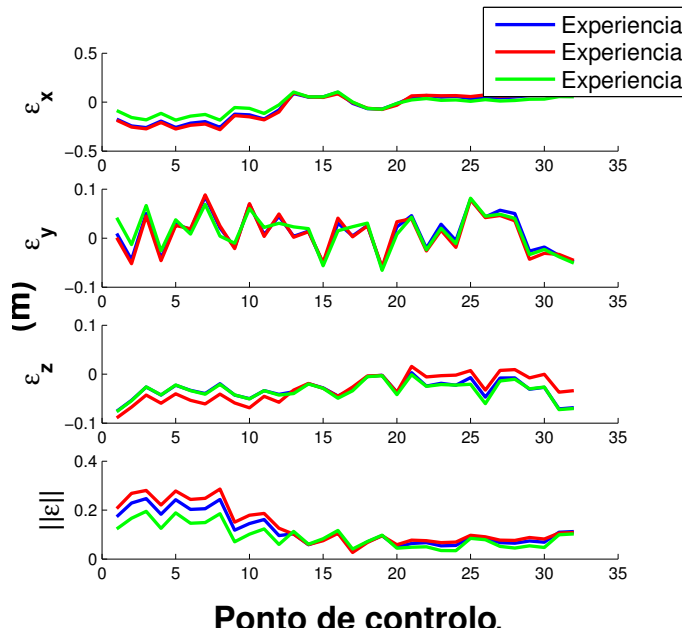


Figura 6.21: Erro de triangulação stereo para 3 experiencias.

6.4 Comparação entre métodos

Na presente secção são comparados os diferentes métodos de calibração dos parâmetros extrínsecos das câmaras do campo de MSL para localização e seguimento de objetos. Na figura 6.22 é apresentada a média do erro dos pontos de controlo em 3D. Para a obtenção dos presentes resultados foram utilizados os pontos de controlo em 2D para cada câmara e posteriormente comparados com os pontos de controlo 3D do sistema. Pela observação do gráfico pode ser observado que o método dos pontos de fuga apresenta erros de estimação 3D mais uniformes para os diversos pontos de controlo, ou seja, quer para os pontos mais afastados da câmara, pontos com índice menor, quer para os pontos mais próximos, o erro varia em torno dos 0.12m. Já os métodos de decomposição de homografia e o método dos 8 pontos apresentam erros de estimação 3D maiores para os pontos mais afastados, pontos com índice menor, e erros menores para os pontos de mais próximos

das câmaras, pontos com índice maior. Seguidamente para todas as cali-

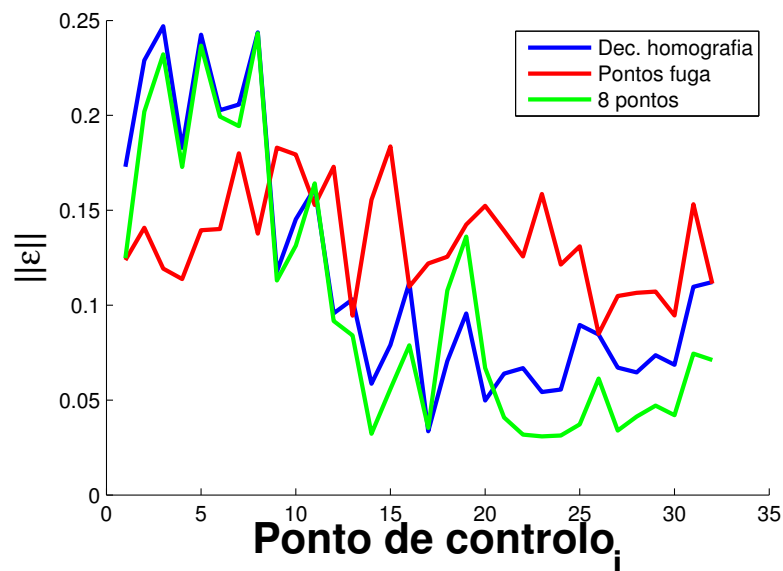


Figura 6.22: Erro de triangulação stereo médio.

brações efetuadas são apresentados os respetivos dados sob a representação do diagrama de caixa, ou *boxplot* tal como denominado usualmente. Os diagramas seguidamente apresentados visam a análise dos resultados das calibrações segundo os diferentes métodos. Estes apresentam a media das medidas, representada com a linha a vermelho, a caixa a azul representa 25% das ocorrências pelo limite inferior e 75% pelo limite superior. As linhas a preto representam os valores máximos e mínimos para cada parâmetro de calibração. Neste diagrama não são considerados os *outliers* estatisticamente, estes são representados com uma cruz a vermelho, que são as medidas que apresentem 95% de diferença face à média. Nestes são apresentados os valores da precisão de cada calibração face ao método dos 8 pontos. Para as representações da precisão da rotação são apresentados ordenadamente as componentes *yaw*, *pitch* e *roll* em graus e para as translações as componentes destas pela ordem, *XX*, *YY* e *ZZ* em metros. Na figura 6.23 é apresentado o *boxplot* para a componente de rotação das calibrações da câmara da es-

querda relativamente ao campo de MSL com recurso ao método dos pontos de fuga. Na figura 6.24 é apresentado o *boxplot* para a rotação da câmara

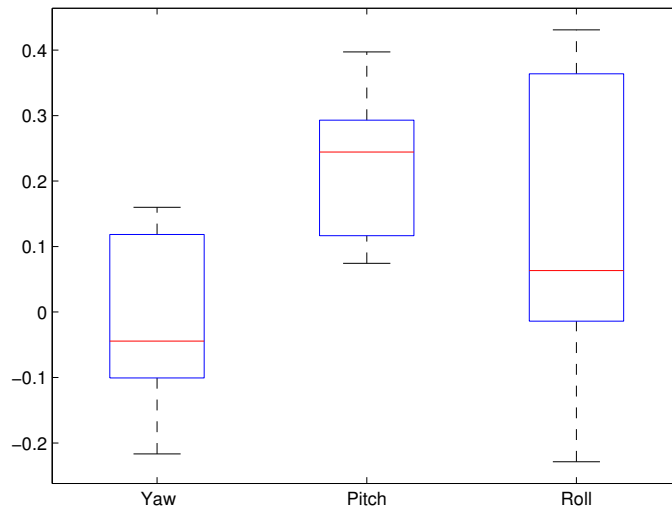


Figura 6.23: Erro de calibração pelo método dos pontos fuga - rotação câmara esquerda, *yaw roll pitch*.

da direita relativamente ao campo de MSL pelo método dos pontos de fuga. De seguida na figura 6.25 é apresentado o *boxplot* relativo à translação da

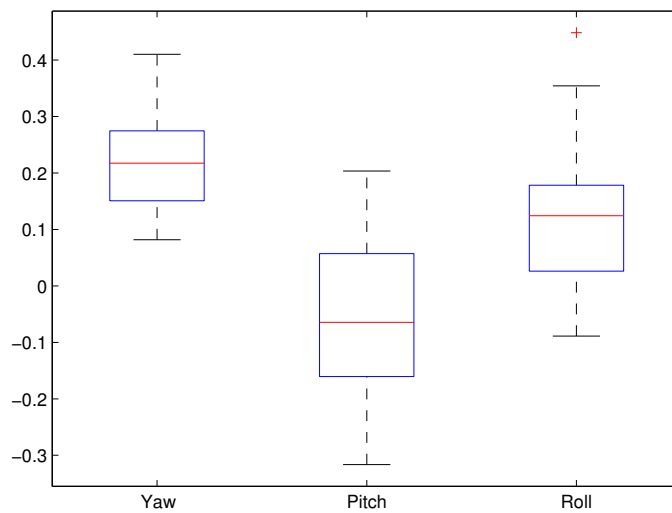


Figura 6.24: Erro de calibração pelo método dos pontos de fuga- rotação câmara direita, *yaw roll pitch*.

câmara da esquerda em relação ao campo de MSL. Abaixo, na Figura 6.26 é

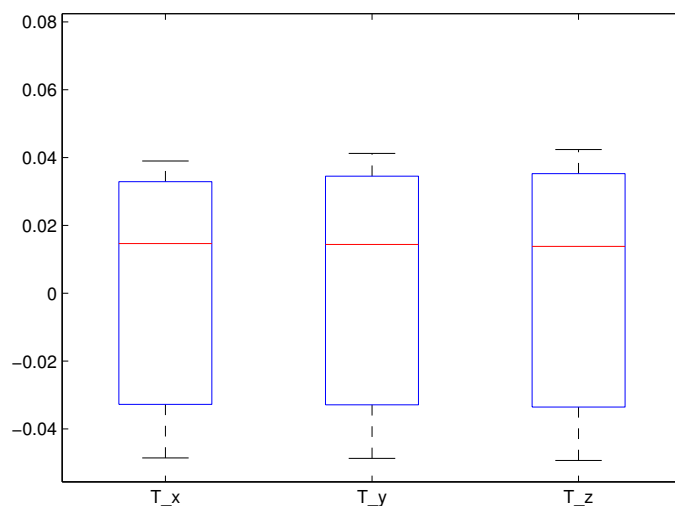


Figura 6.25: Erro de calibração pelo método dos pontos de fuga - translação câmara esquerda, T_x , T_y e T_z .

apresentado o *boxplot* para a calibração da câmara da direita em relação ao campo pelo método dos pontos de fuga Na figura 6.27 é apresentado o *box-*

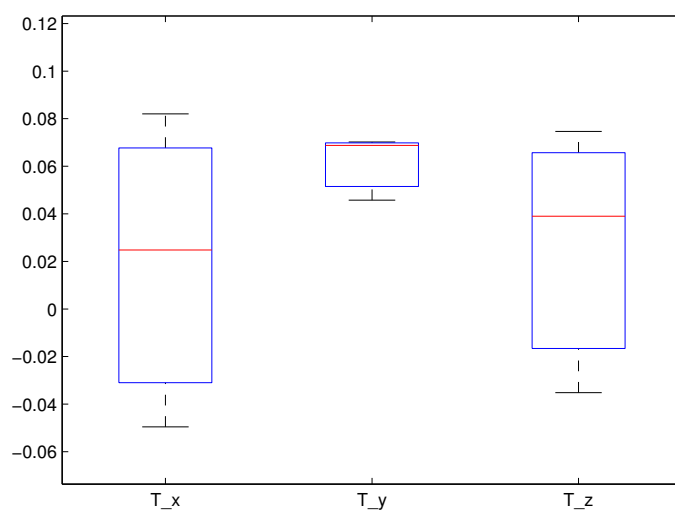


Figura 6.26: Erro de calibração pelo método dos pontos de fuga - translação câmara direita, T_x , T_y e T_z .

plot para a componente de rotação das calibrações da câmara da esquerda

relativamente ao campo de MSL com recurso ao método de decomposição de homografia. Na figura 6.23 é apresentado o *boxplot* para as componentes de

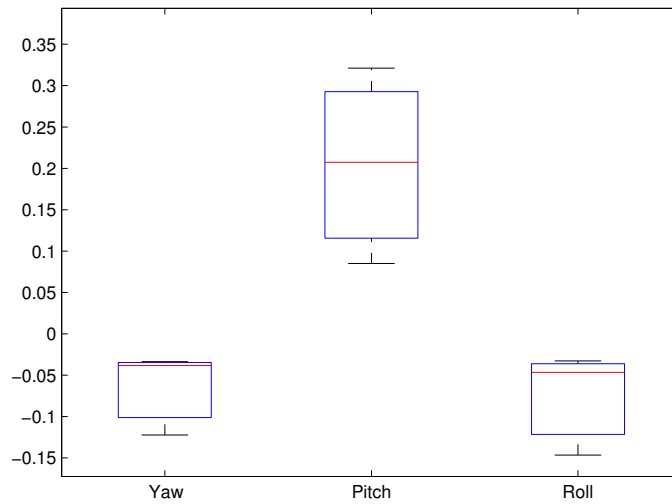


Figura 6.27: Erro de calibração pelo método de homografia- rotação câmara esquerda, *yaw*, *roll* e *pitch*.

rotação das calibrações da câmara da esquerda relativamente ao campo de MSL com recurso ao método de calibração dos de decomposição de homografia. De seguida na figura 6.29 é apresentado o *boxplot* relativo à translação da câmara da esquerda em relação ao campo de MSL. Abaixo, na Figura 6.26 é apresentado o *boxplot* para a calibração da câmara da direita em relação ao campo pelo método de decomposição de homografia.

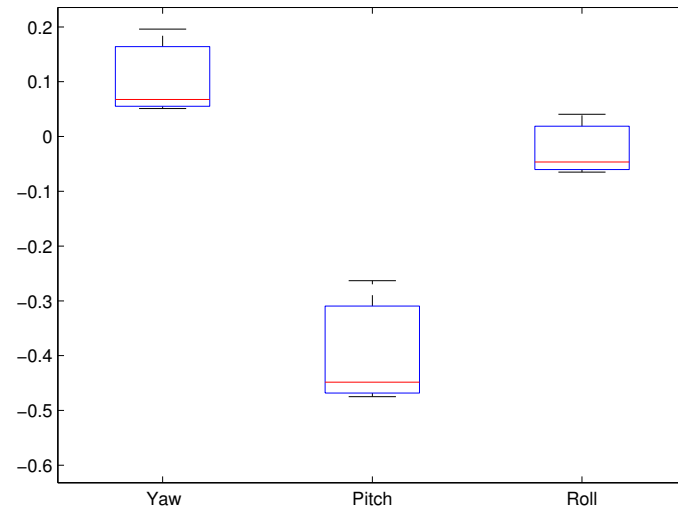


Figura 6.28: Erro de calibração pelo método de homografia- rotação câmara direita, *yaw pitch roll*.

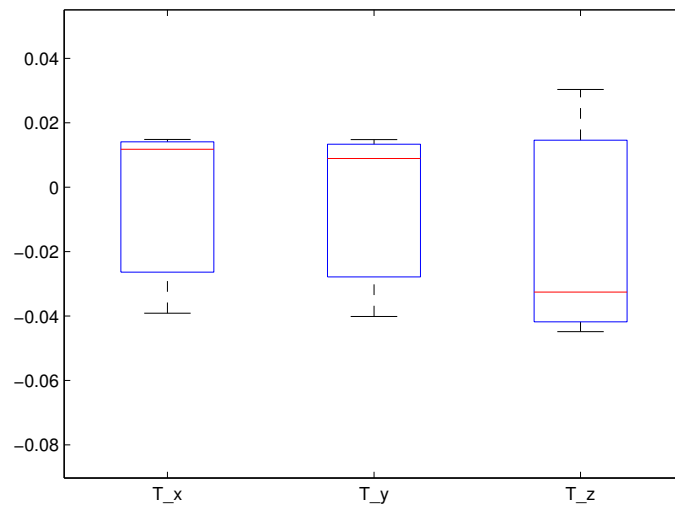


Figura 6.29: Erro de calibração pelo método de homografia- translação câmara esquerda, T_x , T_y e T_z .

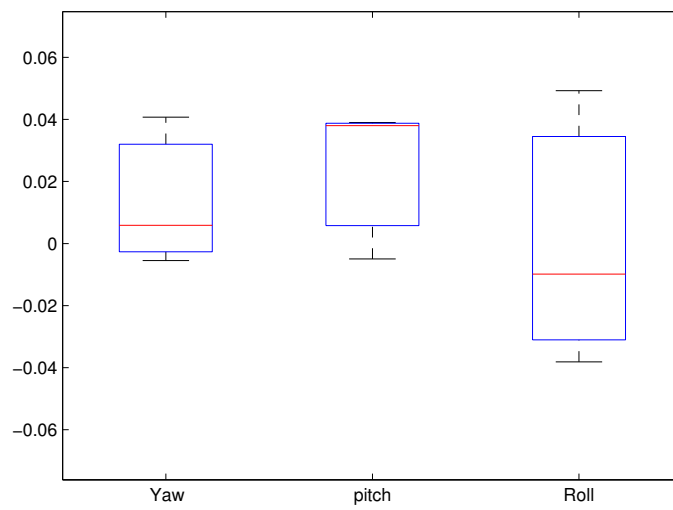


Figura 6.30: Erro de calibração pelo método de homografia- translação câmara direita, T_x , T_y e T_z .

Esta página foi intencionalmente deixada em branco.

Capítulo 7

Conclusões

Este trabalho abordou o problema de calibração de um sistema multi-câmara para seguimento e localização de sistemas autónomos. Tendo como foco os requisitos dos vários cenários de aplicação, foram estudados várias técnicas do estado da arte. No sentido de permitir a implementação de uma ferramenta de calibração automática do sistema multi-câmara foram desenvolvidos vários algoritmos para calibração recorrendo a técnicas como o método dos 8 pontos, pontos de fuga, e decomposição de homografia planar. Para a validação e caracterização dos vários métodos desenvolvidos, foi necessária a implementação de um *setup* experimental com múltiplas câmaras, a elaboração deste contemplou:

- Instalação e calibração de parâmetros intrínsecos de múltiplas câmaras;
- Desenvolvimento de uma aplicação para configuração das câmaras (nomeadamente controlo de parâmetros de aquisição)
- Desenvolvimento de aplicação para a aquisição sincronizada e registo de imagens de múltiplas câmaras em ROS.

Este setup, foi implementado quer no cenário de futebol robótico quer no tanque de testes subaquáticos. Esta implementação para além de permitir

validar dos métodos desenvolvidos consiste já uma solução de um sistema de *ground truth* funcional para aplicação nos dois cenários. Este sistema foi utilizado em múltiplas publicações científicas em ambos os cenários. Para a validação e caracterização dos métodos implementados, e utilizando o *setup* experimental, cada uma dos métodos de calibração foi aplicado múltiplas vezes (64 vezes para um dos métodos) em condições de aquisição e iluminação diferentes. Tendo em vista os resultados obtidos para os diferentes métodos de calibração, verifica-se que estes apresentam em termos médio uma exatidão semelhante. Dos métodos referidos o método dos pontos de fuga apresenta maior estabilidade para pontos a diferentes distâncias, contrariamente aos outros dois métodos, método dos 8 pontos e decomposição de homografia, que apresentam melhor exatidão para os pontos com distancia menor. No Entanto para uma melhor caracterização da exatidão de cada método abordado era necessária a obtenção da posição dos pontos de controlo de forma mais exata, por exemplo com recurso a um lidar para mapeamento tridimensional. Os resultados deste trabalho já contribuíram para uma publicação [28], e está prevista um segunda publicação endereçando a caracterização e comparação experimental dos vários métodos de calibração. Como trabalho futuro, na vertente dos métodos de calibração algumas melhorias tais como: Para o método dos pontos de fuga, alterar o mecanismo para a deteção das retas recorrendo a técnicas de deteção de fronteira de cor com precisão sub-pixel. Este pode ainda ser melhorado introduzindo um passo adicional de refinamento da estimação do ângulo e distancia a origem das retas, usando técnicas de minimização, antes da sua utilização no calculo dos pontos de fuga, para melhorar a exatidão destes. Para o caso da decomposição de homografia, a deteção da fonte luminosa pode ser feita com recurso à utilização do centro de massa da referida fonte de luz. Seria relevante ainda, para obtenção da calibração em ambientes exteriores não estruturados e irregulares, o desenvolvimento de mais um módulo para a

ferramenta de calibração que utiliza técnicas como o "Bundle Adjustment". Finalmente seria necessária a integração de todos os métodos de calibração numa ferramenta de fácil utilização, para qualquer pessoa que se proponha a utilizar a mesma.

Esta página foi intencionalmente deixada em branco.

Bibliografia

- [1] Ergoneers. <http://www.ergoneers.com/en/hardware/eye-tracking>. data de consulta 22 - 10 - 2014.
- [2] Ikinema. <http://www.ikinema.com/>. data de consulta 22 - 10 - 2014.
- [3] Ndigital. <http://www.ndigital.com/msci/products/optotrak-certus/>. data de consulta 22 - 10 - 2014.
- [4] Vicon. <http://www.vicon.com>. data de consulta 22 - 10 - 2014.
- [5] MSL Technical Committee 1997-2014. Middle size robot league rules and regulations for2014.
- [6] Baler AG. aca1300-30gc. <http://www.baslerweb.com/products/ace.html?model=167>. data de consulta 31 - 10 - 2014.
- [7] Anubhav Agarwal, CV Jawahar, and PJ Narayanan. A survey of planar homography estimation techniques. *Centre for Visual Information Technology, Tech. Rep. IIIT/TR/2005/12*, 2005.
- [8] Andreas F. Hofmann, Peter M. Walz, and Hans Thomas. High resolution topography following chemical mapping of ocean hypoxia by use of an autonomous underwater vehicle, the santa monica basin example. *j. atmos. oceanic technol.* 2013.

- [9] K. Arun, T. Huang, and S. Blostein. Least-squares fitting of two 3-d point sets. *IEEE transactions on pattern analysis and machine intelligence* 9(5), 1987.
- [10] Dana H Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122, 1981.
- [11] Allen Bierbaum, Christopher Just, Patrick Hartling, Kevin Meinert, Albert Baker, and Carolina Cruz-Neira. Vr juggler: A virtual platform for virtual reality application development. In *Virtual Reality, 2001. Proceedings. IEEE*, pages 89–96. IEEE, 2001.
- [12] J Bouguet. Camera calibration toolbox. http://www.vision.caltech.edu/bouguetj/calib_doc/.
- [13] J. Bruce, T. Balch, and M. Veloso. Fast and inexpensive color image segmentation for interactive robots.
- [14] J. Bruce and M. Veloso. Fast and accurate vision-based pattern detection and identification. *ceedings of the 2003 IEEE International Conference on Robotics and Automation (ICRA 2003)*, 2003.
- [15] RoboCup Technical Committee et al. Robocup standard platform league (nao) rule book, 2009.
- [16] Williams D.P. Nato undersea res. centre la spezia italy. *Robotics and Automation (ICRA), 2010 IEEE International Conference on, Anchorage, AK , 3-7 May 2010*, 2010.
- [17] Richard O Duda and Peter E Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.
- [18] Mike Goslin and Mark R Mine. The panda3d graphics engine. *Computer*, 37(10):112–114, 2004.

-
- [19] Vicon Hardware. Bonita. Brochure.
- [20] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [21] Richard I Hartley. In defense of the eight-point algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(6):580–593, 1997.
- [22] Gary W Johnson. *LabVIEW graphical programming: practical applications in instrumentation and control*. McGraw-Hill School Education Group, 1997.
- [23] Joerg Kalwa. 7th security research conference, future security 2012, bonn, germany, september 4-6, 2012.
- [24] Piyush Khandelwal and Peter Stone. A low cost ground truth detection system for robocup using the kinect. In *RoboCup 2011: Robot Soccer World Cup XV*, pages 515–527. Springer, 2012.
- [25] Leow Wee Kheng. Camera models and imaging, 2012.
- [26] Sergei Lupashin, Angela Schollig, Markus Hehn, and Raffaello D’Andrea. The flying machine arena as of 2010. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 2970–2971. IEEE, 2011.
- [27] Yi Ma, Stefano Soatto, Jana Kosecka, and S. Shankar Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. SpringerVerlag, 2003.
- [28] A. Martins, A. Dias, H. Silva, J. Almeida, P. Golçalves, F. Lopes, A.Faria, J. Ribeiro, and E. Silva. Groundtruth system for underwater benchmarking. *IEEE – OCEAN’s 2013*, San Diego, 2013.

- [29] Alfredo Martins, Guilherme Amaral, Andre Dias, Carlos Almeida, Jose Almeida, and Eduardo Silva. Tigre—an autonomous ground robot for outdoor exploration. In *Autonomous Robot Systems (Robotica), 2013 13th International Conference on*, pages 1–6. IEEE, 2013.
- [30] Alfredo Martins, André Dias, Hugo Silva, José Almeida, Pedro Gonçalves, Flávio Lopes, André Faria, Joao Ribeiro, and Eduardo Silva. Groundtruth system for underwater benchmarking.
- [31] Alfredo Martins, Hugo Ferreira, Carlos Almeida, Hugo Silva, José Miguel Almeida, André Figueiredo, Filipe Santos, Domingos Bento, Hugo Silva, and Eduardo Silva. Forest fire detection with a small fixed wing autonomous aerial vehicle. In *Intelligent Autonomous Vehicles*, volume 6, pages 168–173, 2007.
- [32] Alfredo Martins, Hugo Ferreira, Carlos Almeida, Hugo Silva, José Miguel Almeida, and Eduardo Silva. Roaz and roaz ii autonomous surface vehicle design and implementation. In *International Lifesaving Congress*, 2007.
- [33] Osfield and D. Burns. 05-11-2014.
- [34] Aldebaran Robotics. Nao datasheet, 2005.
- [35] Thomas Röfer, Tim Laue, Judith Müller, Alexander Fabisch, Fynn Feldpausch, Katharina Gillmann, Colin Graf, Thijs Jeffrey de Haas, Alexander Härtl, Arne Humann, et al. B-human team report and code release 2011, 2011. *Only available online: http://www.b-human.de/downloads/bhuman11_coderelease.pdf.*
- [36] JJ Ruiz, L Diaz-Mas, F Perez, and A Viguria. Evaluating the accuracy of dem generation algorithms from uav imagery. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(2):333–337, 2013.

- [37] Joel A. Shapiro. *Classical Mechanics*. Octobre 2003.
- [38] Shaojie Shen, Nathan Michael, and Vijay Kumar. Autonomous multi-floor indoor navigation with a computationally constrained mav. In *Robotics and automation (ICRA), 2011 IEEE international conference on*, pages 20–25. IEEE, 2011.
- [39] Nicolas Simond and Patrick Rives. Homography from a vanishing point in urban scenes. In *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volume 1, pages 1005–1010. IEEE, 2003.
- [40] Axiom Software. Blade 3. Brochure.
- [41] Vicon Software. Bodybuilder. Brochure.
- [42] Vicon Software. Nexus 2. Brochure.
- [43] Freek Stulp, Suat Gedikli, Michael Beetz, et al. Evaluating multi-agent robotic systems using ground truth. In *In Proceedings of the Workshop on Methods and Technology for Empirical Evaluation of Multi-agent Systems and Multi-robot Teams (MTEE)*, 2004.
- [44] Peter Sturm and Bill Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Computer Vision—ECCV’96*, pages 709–720. Springer, 1996.
- [45] Tomáš Svoboda, Daniel Martinec, and Tomáš Pajdla. A convenient multicamera self-calibration for virtual environments. *PRESENCE: teleoperators and virtual environments*, 14(4):407–422, 2005.
- [46] Vicon Motion systems. Vicon motus video. Brochure.
- [47] Roger Y Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras

- and lenses. *Robotics and Automation, IEEE Journal of*, 3(4):323–344, 1987.
- [48] Keith Vickery. Acoustic positioning systems: ”a practical overview of current systems”. In *In Proceedings of the 1998 Workshop on Autonomous Underwater Vehicles*.
- [49] Vicon. Tracker. <http://www.vicon.com/Content/FileUpload/Software/Tracker.pdf>. Brochure Software.
- [50] Alfredo Weitzenfeld, Joydeep Biswas, and Mehmet Akar. Robocup small-size league: Past, present and future. 2014.
- [51] Tzu Yen Wong, Peter Kovesi, and Amitava Datta. Projective transformations for image transition animations. In *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pages 493–500. IEEE, 2007.
- [52] Math works. Matlab programming fundamentals. http://www.mathworks.com/help/pdf_doc/matlab/matlab_prog.pdf.
- [53] LLC Worldviz. Vizard 4.0 user’s manual. *Online*. *Disponibile: http://docs.worldviz.com/vizard/*. [Visitada: 23-Aug-2013].
- [54] Zhengyou Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11):1330–1334, 2000.
- [55] Stefan Zickler, Tim Laue, Oliver Birbach, Mahisorn Wongphati, and Manuela Veloso. Ssl-vision: The shared vision system for the robocup small size league. In *Robocup 2009: Robot Soccer World Cup XIII*, 2010.