



SISTEMA DE POSICIONAMENTO EXTERNO MULTI-CÂMARA

ANDRÉ EMANUEL GODINHO NEVES

novembro de 2016

SISTEMA DE POSICIONAMENTO EXTERNO MULTI-CÂMARA

André Emanuel Godinho Neves,
Nº1080424



Departamento de Engenharia Eletrotécnica
Mestrado em Engenharia Eletrotécnica e de Computadores
Área de Especialização em Sistemas Autónomos

Novembro de 2016

Relatório elaborado para satisfação parcial dos requisitos da Unidade Curricular de
Tese/Dissertação do Mestrado em Engenharia Eletrotécnica e de Computadores

Candidato: André Emanuel Godinho Neves,
Nº 1080424, 1080424@isep.ipp.pt

Orientação científica: José Miguel Soares De Almeida



Departamento de Engenharia Eletrotécnica
Mestrado em Engenharia Eletrotécnica e de Computadores
Área de Especialização em Sistemas Autónomos

Novembro de 2016

Esta página foi intencionalmente deixada em branco.

À minha esposa, pais e irmãos...

Agradecimentos

Aproveito esta secção para em primeiro lugar agradecer ao meu orientador Eng^o. José Almeida, pelo empenho, orientação e ajuda em todas as fases deste trabalho.

Agradeço a todos os Engenheiros do Instituto Superior de Engenharia do Porto pela formação académica prestada e em especial a todos os membros da “equipa” do Laboratório de Sistemas Autónomos pelo constante acompanhamento, motivação e dedicação ao longo destes anos.

A todos os meus colegas de curso, André Faria, Renato Ribeiro, José Araújo, Pedro Vieira... pelos bons momentos e companheirismo. Um obrigado muito especial aos meus amigos, Flávio Lopes e João Ribeiro por toda a dedicação, apoio, sugestões e correções ao longo desta dissertação.

A toda a minha família e amigos, em especial aos meus pais e irmãos pela confiança, apoio, carinho e orgulho manifestados durante toda a minha vida.

Um agradecimento muito especial à minha esposa Isa Alves por toda a compreensão, paciência, motivação, apoio e amizade ao longo da elaboração desta dissertação.

Esta página foi intencionalmente deixada em branco.

Resumo

A necessidade de informação externa, de referência, acerca da atitude e posição de um dado robô constitui a motivação da presente dissertação, onde foram desenvolvidos algoritmos para auxílio na navegação, ou simplesmente para aferição e validação da resposta do sistema autónomo ao meio envolvente. O principal contributo passa pelo desenvolvimento de um sistema de posicionamento externo multi-câmara, que determina a localização e atitude de sistemas robóticos ou objetos, com base em marcadores óticos ativos.

Através da caracterização dos sistemas robóticos e dos cenários de atuação, percebe-se a necessidade de desenvolver um sistema de visão constituído por multi-câmaras para endereçar as situações de oclusão, aumentar a cobertura espacial e potenciar a qualidade dos resultados de posicionamento.

Assim é proposta uma arquitetura para um sistema global dirigido aos vários requisitos identificados. A utilização de múltiplas câmaras e objetos de interesse munidos de um conjunto de marcadores ativos, torna viável o seguimento em tempo real destes objetos, em ambiente terrestre, aéreo ou subaquático.

Esta arquitetura do sistema global, é demonstrada através de um sistema composto por três câmaras e uma plataforma de quatro Light-Emitting Diodes (LEDs) para validação de múltiplos módulos de *software*, nomeadamente: identificação e validação de pontos de interesse em imagens; cálculo da posição tridimensional dos marcadores através da combinação de pares de câmaras stereo e geometria 'multi view'; seleção de resultados mais precisos; cálculo da atitude do alvo.

Para validação do sistema implementado, foram realizados ensaios experimentais que demonstram o correto funcionamento dos vários módulos do sistema, para diversas configurações e condicionantes. Simultaneamente, instalaram-se dois dispositivos comerciais (Faro e Pixhawk) para aferição e comparação de resultados.

Os resultados experimentais mostraram uma clara vantagem do sistema de posicionamento multi-câmara face ao stereo, quer em qualidade de informação de posicionamento obtida, quer nos aspetos de cobertura espacial e oclusões.

Palavras-Chave

Robótica, *Ground Truth*, Visão computacional, Câmaras, Marcadores ativos, Triangulação *Stereo*, Geometria *Multi view*, Posicionamento 3D, Atitude, Correspondência de pontos.

Abstract

This dissertation aims to contribute to development Groundtruth System of Multicamera Vision that is also to determinate the pose of robotic systems or objects based on active optical markers. The focus of this thesis is the absence of external information about the target's attitude to correct navigation algorithms or simply to validate the answer of autonomous system to the surrounding environment.

Through the description of the robotic systems in site, it was noticed there was a need to develop a vision system constituted by multi cameras, to address occlusion situations, increase the floor space and up-level the results accuracy.

Highly motivated, it is proposed architecture for a global system facing the multiple identified requirements. By using multiple cameras and objects of interest with several active markers, it has become possible to follow objects in real time different environments such as land, air or underwater.

Regarding about of global architecture system, it is presented a contribution through a system composed by three cameras and a platform with four Light-Emitting Diodes (LEDs) to validate the points of interest on images. Through this process, it is also possible to measure the tridimensional position of the markers using stereo cameras pairs with different combinations and multi view triangulation, in order to obtain a precise selection of results and targets attitude measurement.

In order to validate the implemented system, it were performed several trials showing that the system multiple modules converge to the ideal results under various conditions. Simultaneously, two commercial devices were set up and used (Faro Focus and Pixhawk) to compare results of position and attitude.

Keywords

Robotic, Ground truth, Computer Vision, Cameras, Active Markers, Stereo Triangulation, Multi View Geometry, 3D Position, Attitude, points correspondence.

Esta página foi intencionalmente deixada em branco.

Índice

AGRADECIMENTOS.....	I
RESUMO	III
ABSTRACT	V
ÍNDICE DE FIGURAS	IX
ÍNDICE DE TABELAS	XI
ACRÓNIMOS.....	XIII
1. INTRODUÇÃO.....	1
1.1.ENQUADRAMENTO E MOTIVAÇÃO	2
1.2.CENÁRIOS DE APLICAÇÕES.....	3
1.3.OBJETIVOS	6
1.4.ESTRUTURA DO DOCUMENTO.....	6
2. ESTADO DA ARTE	9
2.1.TECNOLOGIAS/SENSORES COMERCIAIS	10
2.1.1. <i>Global Navigation Satellite System (GNSS)</i>	10
2.1.2. <i>Pseudolites</i>	11
2.1.3. <i>Smart Cameras</i>	12
2.1.4. <i>Sistemas de detecção e localização óticos comerciais</i>	12
2.2.SISTEMA DE VISÃO MULTI-CÂMARA PARA CÁLCULO DE POSIÇÃO E ATITUDE	14
2.3.COMPARAÇÃO DE TECNOLOGIAS PARA <i>GROUND TRUTH</i>	16
2.4.CONCLUSÕES DO ESTADO DA ARTE	17
3. ENQUADRAMENTO TEÓRICO.....	19
3.1.INTRODUÇÃO À VISÃO COMPUTACIONAL	19
3.1.1. <i>Imagem digital</i>	21
3.1.2. <i>Processamento de imagem</i>	23
3.2.MODELO DA CÂMARA	24
3.2.1. <i>Projeção Perspetiva</i>	25
3.3.CALIBRAÇÃO DAS CÂMARAS.....	27
3.3.1. <i>Parâmetros intrínsecos</i>	28
3.3.2. <i>Parâmetros Extrínsecos</i>	30
3.4.VISÃO <i>STEREO</i>	33

3.4.1.	<i>Geometria Epipolar</i>	33
3.4.2.	<i>Retificação stereo</i>	37
3.4.3.	<i>Triangulação stereo</i>	38
4.	PROJETO	41
4.1.	REQUISITOS DO SISTEMA GLOBAL.....	41
4.2.	ARQUITETURA DO SISTEMA DE POSICIONAMENTO	43
4.3.	ARQUITETURA DO SISTEMA PROTÓTIPO DESENVOLVIDO	44
5.	IMPLEMENTAÇÃO	47
5.1.	<i>SETUP EXPERIMENTAL</i>	47
5.2.	CALIBRAÇÃO DOS PARÂMETROS DAS CÂMARAS	51
5.2.1.	<i>Calibração de parâmetros intrínsecos</i>	51
5.2.2.	<i>Calibração de parâmetros extrínsecos</i>	53
5.3.	AQUISIÇÃO E PROCESSAMENTO DE IMAGEM	54
5.3.1.	<i>Aquisição de imagem</i>	54
5.3.2.	<i>Pesquisa de marcadores ativos</i>	55
5.4.	VALIDAÇÃO DE BLOBS	61
5.4.1.	<i>Número de blobs na imagem</i>	62
5.4.2.	<i>Restrição Epipolar</i>	63
5.4.3.	<i>Restrição espacial</i>	64
5.5.	CÁLCULO POSIÇÃO 3D MARCADORES.....	64
5.5.1.	<i>Triangulação Stereo</i>	64
5.5.2.	<i>Geometria Multi View</i>	65
6.	RESULTADOS	69
6.1.	EXTRAÇÃO DE MARCADORES	70
6.2.	POSIÇÃO 3D DOS MARCADORES	72
6.3.	VALIDAÇÃO DA POSIÇÃO TRIDIMENSIONAL DOS MARCADORES	79
6.4.	VALIDAÇÃO DA ATITUDE DO ALVO.....	82
7.	CONCLUSÕES E TRABALHO FUTURO	85
	REFERÊNCIAS DOCUMENTAIS	89

Índice de Figuras

FIGURA 1: EQUIPA DE FUTEBOL ROBÓTICO DOS TECH UNITED EINDHOVEN	4
FIGURA 2: INESCTEC/ ISEP VEICULO AUTÓNOMO AÉREO OTUS	5
FIGURA 3: GPS TRILATERAÇÃO	11
FIGURA 4: PRINCÍPIO DE OPERAÇÃO DE UM PSEUDOLITE COM GPS	11
FIGURA 5: VICON BONITA	13
FIGURA 6: DISTRIBUIÇÃO DE MARCADORES NO ROBÔ NAO [18]	15
FIGURA 7: MARCADORES UTILIZADOS EM [19] PARA DETEÇÃO E ORIENTAÇÃO DE ROBÔS FUTEBOLISTAS	16
FIGURA 8: ANATOMIA DO OLHO HUMANO	20
FIGURA 9: DIFERENÇA ENTRE IMAGEM RASTER E VETORIAL	21
FIGURA 10: REPRESENTAÇÃO DE IMAGEM DIGITAL	22
FIGURA 11: REPRESENTAÇÃO DE UMA IMAGEM EM TONS DE CINZA	22
FIGURA 12: REPRESENTAÇÃO TÍPICA DO MODELO DA CÂMARA DE <i>PINHOLE</i>	25
FIGURA 13: MODELO DA CÂMARA DE <i>PINHOLE</i>	25
FIGURA 14: REPRESENTAÇÃO TRIDIMENSIONAL DE UM PONTO NO PLANO DA CÂMARA E DA IMAGEM.	26
FIGURA 15: DISTORÇÃO (RADIAL NEGATIVA) DE LINHAS RETAS PROVOCADAS PELALENTE	29
FIGURA 16: EFEITO DE DISTORÇÃO RADIAL [31].	29
FIGURA 17: PROJEÇÃO DAS COORDENADAS DO OBJETO NO MUNDO EM COORDENADAS DA CÂMARA	31
FIGURA 18: REPRESENTAÇÃO DOS ÂNGULOS DE EULER NUMA CÂMARA	31
FIGURA 19: MODELO DE GEOMETRIA EPIPOLAR	34
FIGURA 20: PAR <i>STEREO</i> NÃO RETIFICADO E PAR <i>STEREO</i> RETIFICADO [35].	37
FIGURA 21: GEO. DE PAR <i>STEREO</i> RETIFICADO COM LINHAS EPIPOLARES PARALELAS E EPIPOLOS NO INFINITO [1].	38
FIGURA 22: MODELO DE TRIANGULAÇÃO <i>STEREO</i> [35]	39
FIGURA 23: ARQUITETURA DO SISTEMA <i>GROUND TRUTH</i> ÓTICO GLOBAL	43
FIGURA 24: ARQUITETURA FUNCIONAL DO SISTEMA	45
FIGURA 25: FIXAÇÃO DA CÂMARA ESPAÇO DE ENSAIOS	48
FIGURA 26: DISTRIBUIÇÃO DAS TRÊS CÂMARAS PELO ESPAÇO DE ENSAIOS	48
FIGURA 27: CÂMARA BASLER ACA 1300-30GC	49
FIGURA 28: PLATAFORMA DE MARCADORES ATIVOS	50
FIGURA 29: SEQUÊNCIA DE ALGUMAS IMAGENS DO ALVO PARA CALIBRAÇÃO	52
FIGURA 30: ERRO DE REPROJEÇÃO EM PÍXEIS	52
FIGURA 31: PONTOS DO CAMPO REFERENCIADOS PARA DETERMINAÇÃO DOS PARÂMETROS EXTRÍNSECOS	53
FIGURA 32: EXTRAÇÃO DE PONTOS NO PLANO DA IMAGEM	54

FIGURA 33: PERSPETIVAS DAS 3 CÂMARAS INSTALADAS NO LABORATÓRIO	55
FIGURA 34: SEQUÊNCIA DE PROCESSAMENTO FASE1 DA DETECÇÃO DE <i>BLOBS</i>	56
FIGURA 35: RELAÇÃO ENTRE NÚMERO DE PIXÉIS E RAIOS DE PESQUISA	57
FIGURA 36: SEQUÊNCIA DE PROCESSAMENTO FASE 2 DE DETECÇÃO DE MARCADORES	57
FIGURA 37: IMAGEM COM BAIXO VALOR DE EXPOSIÇÃO	58
FIGURA 38: FORMAÇÃO DE <i>UM BLOB</i>	59
FIGURA 39: REPRESENTAÇÃO DO CENTRO DO <i>BLOB</i>	60
FIGURA 40: IMAGEM COM E SEM DISTORÇÃO	60
FIGURA 41: TÉCNICAS/PROCESSOS EXECUTADOS PARA FILTRAGEM DE FALSOS PONTOS DE INTERESSE.....	61
FIGURA 42: REFLEXO DOS MARCADORES ATIVOS EM SUPERFÍCIE ESPELHADA.....	62
FIGURA 43: SOBREPOSIÇÃO DOS MARCADORES.....	62
FIGURA 44: ORGANIZAÇÃO DOS MARCADORES SEGUNDO A SENTIDO DOS PONTEIROS DO RELÓGIO	65
FIGURA 45: GEOMETRIA <i>MULTI VIEW</i>	66
FIGURA 46: CONVERGÊNCIA DO ERRO NA FUNÇÃO <i>FMINSEARCH</i>	66
FIGURA 47: DEFINIÇÃO DE COORDENADAS COM MENOR ERRO.....	68
FIGURA 48: NÚMERO DE MARCADORES DETETADOS EM FUNÇÃO DO NÚMERO DE IMAGENS	71
FIGURA 49: A; SOBREPOSIÇÃO DE MARCADORES B; MARCADOR NÃO PERCETÍVEL	71
FIGURA 50: EXEMPLO DE OCLUSÃO DOS MARCADORES.....	72
FIGURA 51 REPROJEÇÃO DOS PONTOS 3D PARA TRIANGULAÇÃO <i>STEREO</i>	72
FIGURA 52: ERRO DE REPROJEÇÃO EM X E Y PARA CADA MARCADOR COM O MÉTODO TRIANGULAÇÃO <i>STEREO</i>	73
FIGURA 53: DISTÂNCIA ENTRE RETAS PROJETADAS (PAR <i>STEREO</i>).....	73
FIGURA 54: DESFASAMENTO DAS COORDENADAS 3D (CM) PARA CADA PAR DE CÂMARAS	74
FIGURA 55: REPRESENTAÇÃO DO ERRO DE REPROJEÇÃO EM X E Y E ERRO <i>STEREO</i> PARA O MELHOR PAR <i>STEREO</i>	75
FIGURA 56: NÚMERO DE OCORRÊNCIAS MELHOR PAR <i>STEREO</i> PARA CADA GRUPO.....	76
FIGURA 57 REPROJEÇÃO DOS PONTOS 3D PARA GEOMETRIA <i>MULTI VIEW</i>	76
FIGURA 58: ERRO DE REPROJEÇÃO (TRIANGULAÇÃO <i>STEREO</i>) PARA CADA UMA DAS CÂMARAS.....	77
FIGURA 59: ERRO <i>MULTI VIEW</i> EM CM	77
FIGURA 60 REPROJEÇÃO DOS MARCADORES COM BASE NO CÁLCULO 3D TRIANGULAÇÃO <i>STEREO</i> E GEOMETRIA <i>MULTI-VIEW</i>	78
FIGURA 61: OCORRÊNCIAS DE TÉCNICA COM RESULTADOS 3D MAIS PRECISOS	78
FIGURA 62: NUVEM DE PONTOS DO FARO PROCESSADA ATRAVÉS DO <i>SOFTWARE SCENE LT</i>	79
FIGURA 63: DISCREPÂNCIA ENTRE RESULTADOS OBTIDOS E MEDIÇÕES DO FARO	81
FIGURA 64: DISCREPÂNCIA ENTRE RESULTADOS OBTIDOS E MEDIÇÕES REAIS	81
FIGURA 65: INSTALAÇÃO DO PIXHAWK SOBRE O ALVO.....	82
FIGURA 66: DISCREPÂNCIA ENTRE ATITUDE CALCULADA E OBTIDA DO PIXHAWK	83

Índice de Tabelas

TABELA 1: ERRO DE POSICIONAMENTO DO ROBÔ NAO EM [17].....	15
TABELA 2: VALIDAÇÃO E CORRESPONDÊNCIA DE PONTOS ENTRE IMAGENS	63
TABELA 3: NÚMERO DE MARCADORES DETETADOS EM CADA IMAGEM EM FUNÇÃO DA POSIÇÃO DO ALVO	70
TABELA 4: MÉDIA E DESVIO PADRÃO DO ERRO DE REPROJEÇÃO E DO VALOR DO C PARA TODOS OS PARES <i>STEREO</i>	73
TABELA 5: MÉDIA E DESVIO DO DESFASAMENTO ENTRE COORDENADAS 3D.....	74
TABELA 6: MÉDIA E DESVIO PADRÃO PARA ERRO DE REPROJEÇÃO E ESCALAR C MELHOR PAR <i>STEREO</i>	75
TABELA 7: ERRO MÉDIO E DESVIO PADRÃO DO ERRO DE REPROJEÇÃO E ERRO <i>MULTI VIEW</i>	77
TABELA 8: COMPARAÇÃO DAS COORDENADAS 3D EM METROS NO REFERENCIAL DO CAMPO	80
TABELA 9: DISCREPÂNCIA MÉDIA E DESVIO PADRÃO ENTRE RESULTADOS OBTIDOS E REFERÊNCIAS.....	81
TABELA 10: ERRO MÉDIO E DESVIO PADRÃO	83

Esta página foi intencionalmente deixada em branco.

Acrónimos

2D Duas Dimensões

3D Três Dimensões

CMOS *Complementary Metal-Oxide Semiconductor*

CRAS Centro de Robótica e Sistemas Autónomos

FOV Field Of View

FPS *Frames Per Second*

GPS *Global Position System*

INESC Instituto de Engenharia de Sistemas e Computadores

ISEP Instituto Superior de Engenharia do Porto

LED *Light-Emitting Diode*

LRF *Laser Range Finder*

LSA Laboratório de Sistemas Autónomos

MSL *Middle Size League*

POE *Power Over Ethernet*

RGB *Red-Green-Blue*

ROS Robot Operating System

SPL *Standard Platform League*

UAV *Unmanned Aerial Vehicle*

Esta página foi intencionalmente deixada em branco.

1. INTRODUÇÃO

Os sistemas autónomos estão cada vez mais presentes no nosso dia-a-dia. Estes sistemas robóticos recorrem a técnicas de perceção com base em sensores, sonares, *Lasers range finder*, GPS, câmaras. Analogamente aos humanos, os sistemas de visão óticos são as principais fontes de informação sensorial. Exatidão e robustez, são características fundamentais em tarefas de navegação, caracterização do espaço e/ou meio, planeamento de trajetos, ou até mesmo de controlo. Para aferir ou corrigir os sistemas de perceção dos robôs, é comum recorrer-se a informação externa exata, que serve de referência para o estudo, aperfeiçoamento ou implementação de funcionalidades. Em muitos dos casos recorrem-se a estes sistemas de correção apenas em fase de desenvolvimento do sistema autónomo.

Deste modo, através da conceção de um sistema ótico que serve de referência a sistemas autónomos surge a presente dissertação que será contextualizada neste capítulo.

1.1. ENQUADRAMENTO E MOTIVAÇÃO

A dissertação descrita neste documento enquadra-se no Mestrado em Engenharia Eletrotécnica e Computadores, ramo de Sistemas Autónomos do Instituto Superior de Engenha. No Centro de Robótica e Sistemas Autónomos CRAS- INESC TEC¹ reúnem-se esforços para a conceção e estudo de sistemas robóticos, onde os robôs móveis adquirem especial importância, nos meios: aéreo, aquático e terrestre.

Independentemente do meio de ação do robô, a sua localização (absoluta ou relativa) é fundamental para uma boa navegação, controlo e interação com os objetos presentes no espaço. Desta forma surge a necessidade de implementação de sistemas de referência que localizam e seguem os sistemas robóticos e objetos presentes na área de ação (em modo *online* ou *offline*). A informação externa da posição e atitude dos objetos em estudo podem ser utilizadas para desenvolver ou corrigir algoritmos de navegação e controlo dos robôs, ou simplesmente para aferição e validação da resposta do sistema ao meio envolvente.

É no contexto da necessidade de um sistema externo de informação de posicionamento a integrar num sistema de *ground truth* que surge a contribuição da presente dissertação. Este caracteriza-se pela operação em ambientes interiores terrestres e aéreos, recorrendo a um sistema de posicionamento externo baseado em visão. Esta tecnologia destaca-se pelos bons resultados já apresentados de acordo com o capítulo 2 (Estado da Arte), pela grande quantidade de informação capturada do meio, e pela margem de progressão. Com os recentes avanços tecnológicos, o aumento significativo do poder computacional, crescimento abrupto da capacidade e qualidade de aquisição de imagens por parte das câmaras, tornaram extremamente viável o desenvolvimento de um sistema de localização e seguimento de objetos baseado em sistemas óticos de visão artificial.

Atualmente, existem já importantes contribuições de sistemas *ground truth* óticos no mercado e nomeadamente no Laboratório de Sistemas Autónomos do Instituto Superior de

¹ www.inesctec.pt/ acedido a 3/4/2016

Engenharia do Porto, ISEP¹, como são os exemplos dos trabalhos desenvolvidos por: A. Dias com um sistema stereo de *ground truth* em tempo real [16]; N. Nunes com um sistema de seguimento multi target [1]; A. Faria com um sistema de calibração automática de câmaras [40].

Tendo como base os trabalhos destes autores, percebe-se que existe a necessidade de desenvolvimento de um sistema multi-câmara com *multi view*. Atributos como: preço, número de câmaras, eficácia, capacidade de resposta a oclusão do objeto, espaço de trabalho e sobretudo precisão, continuam a ser características passíveis de melhoramento e comparação. Com esta motivação, nesta tese, são abordadas técnicas de visão artificial que integram um sistema de *ground truth 3D multi view* com diferentes aplicações.

1.2. CENÁRIOS DE APLICAÇÕES

Os sistemas de *Ground Truth* baseados em sistemas de visão óticos apresentam uma grande polivalência de aplicações, integrando marcadores ativos, passivos, ou simplesmente sem marcadores. Tendo em conta estes requisitos podem-se destacar algumas áreas onde sistemas de *Ground truth* são frequentemente integrados:

- Estudo de movimentos humano e animal;
- Auxílio médico (diagnóstico e cirurgia);
- Desenvolvimento de sistemas de animação e realidade virtual;
- Estudos de engenharia;
- Posicionamento e localização de pessoas, robôs, objetos;

Neste trabalho é dado especial relevo ao posicionamento e localização de robôs e objetos de interesse integrando marcadores ativos aplicados à equipa de futebol robótico do ISEPorto e a UAV (*Unmanned Aerial Vehicle*) do tipo *Quadrotor*.

¹ <http://www.isep.ipp.pt/> acedido a 18/4/2016

A equipa de futebol robótico do ISEPorto¹, é constituída por 5 robôs que competem na Middle Size League (MSL)[2], da competição Robocup[3]. Estes robôs (Figura 1) movem-se num plano 2D de acordo com as regras de jogo definido em [4]. Os sistemas autónomos devem ser capazes de tomar decisões como por exemplo qual o melhor jogador a passar a bola (depende da distância a que este está do jogador que possui a bola) e probabilidade de sucesso no golo (depende da distância à baliza e obstrução do caminho da bola). Para que estas e outras decisões sejam corretamente tomadas e o principal objetivo (marcar mais golos que o adversário) seja cumprido, é necessário que os sistemas de posicionamento e orientação dos robôs no campo sejam precisos.



Figura 1: Equipa de futebol robótico dos Tech United Eindhoven²

É no contexto da contribuição para o futebol robótico que surge um exemplo da aplicação de um sistema de *ground truth* (a desenvolver), que seja capaz de caracterizar a dinâmica dos robôs futebolistas e/ou bola do jogo em tempo real com base num sistema ótico, para caracterização e ajuste dos algoritmos implementados durante o jogo. Outra abordagem, pode ser, o armazenamento e posterior processamento de informação do estado do alvo em modo *offline*, permitindo estudar a resposta dos robôs às circunstâncias do jogo. Em qualquer um dos casos a instalação de marcadores ativos no objeto em estudo e instalação

¹ http://www.lsa.isep.ipp.pt/iseporto_home.html acedido a 18/4/2016

² <http://www.techunited.nl/> acedido a 23/9/2016

de várias câmaras distribuída pela área de ação são essenciais para ajudar a ultrapassar oclusões e elevar a exatidão do sistema.

Outra aplicação onde o sistema de *ground truth* oferece uma grande utilidade é no estudo do Quadrotor “OTUS” (Figura 2), desenvolvido no Laboratório de Sistemas Autónomos do ISEP. Este veículo aéreo autónomo realiza várias missões de reconhecimento do espaço envolvente em ambiente *indoor* e *outdoor*, tal como, mapeamento, etc.

Este robô conta com um sistema de navegação autónoma, onde a contribuição de informação sensorial (como é o caso do GPS e sensores óticos) marcam a diferença na qualidade do controlo e navegação do veículo.



Figura 2: INESCTEC/ ISEP Veículo autónomo aéreo Otus

Em ambientes interiores onde o posicionamento por GPS não funciona, o que dificulta o processo de localização. Acrescida a esta limitação, há a considerar o erro acumulativo dos cálculos baseados nos sensores inerciais.

Para além da gestão e controlo do veículo aéreo autónomo, durante as missões, há tarefas relacionadas com a reconstrução 3D do meio, que integra informação posicional e as imagens 2D capturadas durante o voo. Para esta tarefa, é crucial um correto posicionamento e atitude do UAV.

Para corrigir, aferir, ou até mesmo complementar os sistemas de posicionamento do Otus, pode-se recorrer ao sistema *ground truth* ótico, instalado em ambiente interior, que é capaz

de cobrir a área de operação do *quadrotor* no espaço 3D, recorrendo a marcadores ativos instalados nos eixos, sem que estes causem problemas de navegação.

As aplicações anteriormente abordadas são apenas dois exemplos de onde a implementação de um sistema de *ground truth* ótico se pode revelar muito útil. Contudo, há a frisar que cada aplicação tem as suas especificações, como tal, há um conjunto de pré-requisitos que o sistema a desenvolver deve cumprir.

1.3. OBJETIVOS

A presente dissertação visa contribuir para o desenvolvimento de um sistema de posicionamento externo multi-câmara, também denominado por sistemas *ground truth*, para determinação da localização e atitude de múltiplos objetos. Com base nesta temática foram delineados os seguintes objetivos para este trabalho:

- Projeto de um sistema de posicionamento externo global que integra N câmaras aplicável no meio terrestre, aéreo e subaquático, com recurso a marcadores ativos;
- Desenvolvimento de um *setup* experimental para validação do sistema multi-câmara
- Planeamento e execução de testes de campo para validação
- Avaliação e caracterização do protótipo do sistema de posicionamento externo implementado
- Comparação dos resultados de posição e atitude obtidos com protótipo do sistema de posicionamento externo implementado

1.4. ESTRUTURA DO DOCUMENTO

Este documento está organizado em sete capítulos.

Capítulo dois: são mencionadas algumas tecnologias existentes no mercado e ainda alguns trabalhos que estão diretamente relacionados com a temática desta dissertação.

No terceiro capítulo é feita referência a conceitos e fundamentos necessários à compreensão de um sistema de visão computacional.

No quarto capítulo é representada e detalhada a arquitetura funcional do sistema global, bem como as características do *hardware* que constitui o sistema.

Capítulo cinco, são descritos todos os blocos de *software* e *setup* experimental tido em conta nos ensaios realizados.

No sexto capítulo são apresentados e analisados os resultados referentes aos vários ensaios realizados em diferentes cenários.

Por último o sétimo capítulo, onde são apresentadas algumas conclusões e perspectivas de melhoramentos a implementar no futuro.

Esta página foi intencionalmente deixada em branco.

2. ESTADO DA ARTE

O estudo prévio acerca de um determinado tema é essencial quando o objetivo é a elaboração de um projeto no âmbito científico.

Esta tese visa contribuir para o desenvolvimento de um sistema de posicionamento externo, que se caracteriza pelo fornecimento de informação de referência de posição e atitude de objetos. Essas referências maioritariamente são utilizadas na análise da qualidade e correção de algoritmos e sensores de posicionamento integrados em veículos autónomos. Dada a grande diversidade e necessidades desses sistemas autónomos, a conceção de um sistema de *ground truth* tem início no levantamento de requisitos e necessidades de aplicações, como:

- Meio (terrestre, aquático ou aéreo);
- Ambiente interior ou exterior;
- Área de atuação;
- Exatidão e graus de liberdade (6D ou 3D ou 2D);
- Velocidade e tamanho do robô/objeto a ser seguido.

A partir destas principais características parte-se para a escolha dos sensores que compõem um sistema de *ground truth*. Entre os mais utilizados estão os LRFs, o *Global Positioning Systems* (GPSs) e as câmaras. Posteriormente são aplicados algoritmos desenvolvidos e estudados por diversos autores que levam à determinação da posição e orientação de um sistema autónomo num determinado referencial.

Neste capítulo são apresentadas algumas soluções, (tecnologias, produtos e contribuições científicas) relacionadas com a temática de deteção e posicionamento de objetos de interesse no mundo.

2.1. TECNOLOGIAS/SENSORES COMERCIAIS

O estudo de algumas tecnologias/produtos relacionados com a temática de posicionamento de objetos permitiu apurar os seus princípios de funcionamento, características e exemplos de aplicações. Esta é uma etapa indispensável, para entender qual o sensor ou conjunto de sensores que melhor se adapta às características de um sistema semelhante ao desenvolvido no âmbito desta dissertação.

2.1.1. GLOBAL NAVIGATION SATELLITE SYSTEM (GNSS)

Quando o tema é a posição tridimensional de um qualquer ponto localizado no globo terrestre a qualquer instante, é inevitável falar do sistema GPS. Desde a sua criação nos anos 70 pelo departamento de Defesa Norte-Americano, este tem vindo a ser aplicado em áreas como localização, navegação, mapeamento.

Para um correto funcionamento do sistema GPS é necessário a correlação entre três componentes, designados por: espacial, controlo e o utilizador. O segmento espacial é constituído por 24 satélites que orbitam em torno da Terra e enviam sinais rádio das suas posições. Para ser feita uma localização exata é necessário que o módulo utilizador capte pelo menos 4 satélites em qualquer ponto da terra. Este módulo é constituído por um recetor rádio com uma unidade de processamento capaz de descodificar em tempo real a informação enviada por cada satélite e capaz de calcular a sua posição exata ao longo do tempo [5].

A posição tridimensional do objeto é calculada com base na interseção de 3 sinais de satélite (trilateração) como mostra a Figura 3. A distância entre recetor-satélite é calculada

com base na multiplicação do tempo medido pela velocidade do sinal (velocidade da luz) equação 2.1 [6]

$$d = v \times t \quad (2.1)$$

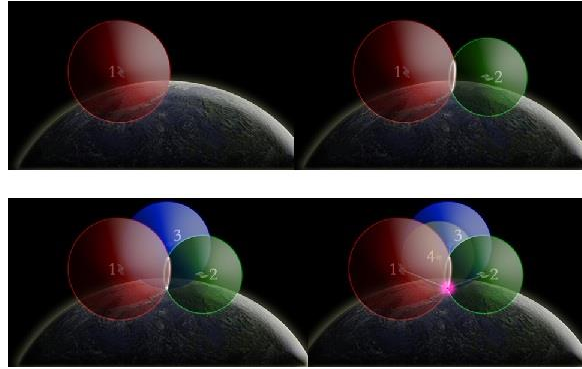


Figura 3: GPS trilateração¹

2.1.2. PSEUDOLITES

Os pseudolites ou Pseudo-Satélites são transmissores colocados ao nível do solo que são utilizados para complementar a navegação baseada no sistema GPS. Os pseudolites usualmente são utilizados na navegação interior, onde os sinais GPS têm dificuldade em operar (Figura 4). Desse modo, os pseudolites tem demonstrado ótimos resultados ao nível da precisão como descrevem [7] e [8].

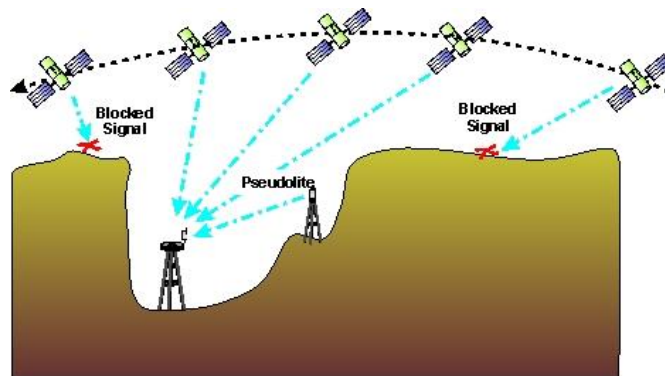


Figura 4: Princípio de operação de um pseudolite com GPS¹

¹ <http://www.vaztolentino.com.br/conteudo/559-COMO-O-GPS-DETERMINA-A-LOCALIZA-aO-DE-UM-PONTO-SOBRE-A-SUPERFICIE-TERRESTRE>, acessado a 22/1/2016.

2.1.3. SMART CAMERAS

As *Smart cameras* são dispositivos capazes de adquirir e “interpretar” imagens do mundo real. Estes pequenos dispositivos começaram a surgir com uma considerável capacidade de processamento, acoplados de uma câmara que permite ser integrada em sistemas embebidos, desempenhando desse modo tarefas de visão computacional pouco complexas.

Muitos são os trabalhos que se podem encontrar que integram *smart cameras*, tais como [9][10][11], que destacam aplicações de vigilância relacionados com o controlo e deteção de presença humana. No entanto, pode contribuir para um sistema de posicionamento multi-câmara de grandes dimensões.

2.1.4. SISTEMAS DE DETEÇÃO E LOCALIZAÇÃO ÓTICOS COMERCIAIS

Vicon

A Vicon [12] é uma empresa que está ligada ao desenvolvimento de sistemas de análise de movimento por visão recorrendo a marcadores passivos. Desde 1984 que a VICON tem vindo a desenvolver diversos produtos para estudos clínicos e biomecânicos (por exemplo análise de desempenho de atletas de alta competição), ou até produtos aplicáveis em centros de investigação de captura de movimento em 6D de veículos autónomos. Atualmente a Vicon possui um portfólio comercial bastante extenso em *Hardware* (Vantage, Bonita e Cara) e em *Software* (Nexus, Blade, Tracker, Polygon, Vicon Control, ProCalc, Cara, Pegasus, boujou) [12].

Da oferta comercial desta empresa, destacam-se por exemplo o modelo, Vicon Bonita (Figura 5), com umas características interessantes sobre o ponto de vista de captura da posição de objetos. Com uma resolução de 1 Mpx e uma frequência máxima de aquisição de 250 *fps* (*Frames Per Second*), este é capaz de capturar objetos com um erro de translação inferior a 0,5 milímetros e um erro de rotação inferior a 0,5° com a utilização de 9 marcadores. A comunicação e alimentação do sistema é feita através de uma conexão

¹ <http://www.wirelessdictionary.com/Wireless-Dictionary-GPS-Pseudo-Satellite-Pseudolites-Definition.html>, acessado a 26/1/2016

POE (*Power Over Ethernet*) e segundo o fabricante a calibração é realizada em menos de dois minutos através de um kit desenvolvido para o efeito.

Dada a versatilidade e especificações muito interessantes dos sistemas de localização Vicon, muitos são os trabalhos desenvolvidos na área de investigação de veículos autónomos, que recorrem a este sistema de múltiplas câmaras. Por exemplo, os autores Guillaume D. e Raffaello D. apresentam um trabalho relacionado com a navegação de um *Quadrotor* autónomo recorrendo a marcadores passivos instalados no veículo[13]. Com esta informação é possível determinar a posição, orientação e velocidade do *Quadrotor* em relação a um determinado referencial mundo.



Figura 5: Vicon Bonita¹

Optotrak

O Optotrak [14] é um sistema ótico baseado em marcadores ativos capaz de determinar a posição 3D de um objeto equipado com LEDs infravermelhos (IRLEDS). São necessários pelo menos 3 IRLEDS para determinar a posição e orientação de um objeto com uma frequência máxima de 4600Hz. Em simultâneo podem ser identificados até um máximo de 512 marcadores. Comparando o sistema da Optotrak com o da Vicon a grande diferença recai sobre o tipo de marcadores que utilizam. Como semelhança, ambos possuem variados modelos com diferentes características e aplicações.

¹ <http://www.prophysics-sol.se/product-database/bonita/>, acedido em 2/2/2016

Qualisys

Ao contrário da Vicon, a Qualisys [15] apresenta um sistema que é capaz de operar em ambiente interior ou exterior. Os marcadores podem ser ativos ou passivos, o que permite no caso da escolha de marcadores ativos, uma cobertura de funcionamento a distâncias superiores, quando comparado com os marcadores passivos. Para além de uma maior distância a Qualisys é resistente à água.

2.2. SISTEMA DE VISÃO MULTI-CÂMARA PARA CÁLCULO DE POSIÇÃO E ATITUDE

Com os recentes avanços tecnológicos, as funcionalidades dos sistemas autónomos aumentaram. Para o desenvolvimento de uma qualquer tarefa é essencial que o robô tenha noção da sua posição e orientação no mundo. Com essa informação é possível percorrer uma trajetória planeada, detetar objetos e tomar decisões. A qualidade de execução destas tarefas está relacionada com a qualidade dos algoritmos e sensores de posicionamento do sistema autónomo. Para localizar, aferir e corrigir sistemas de posicionamento de robôs (em simultâneo ou não) é comum recorrer-se a sistemas de visão multi-câmaras *ground truth* como descrevem vários autores no âmbito científico:

Dias A. *et al* [16] apresentam um sistema de *ground truth* 3D capaz de executar um seguimento de múltiplos alvos em simultâneo no cenário de futebol robótico MSL [2]. Este trabalho é uma aplicação *opensource*, desenvolvido para cenários interiores que oferece um método fácil para obtenção dos parâmetros de calibração e é capaz de garantir uma boa precisão sem utilizar marcadores. Esta solução recorre a imagens gravadas em cenário de jogo e apresenta um erro médio de reprojeção menor que um píxel, justificando uma precisão, inferior a 0,05 metros, mesmo quando o objeto se encontra a 20 metros de distância das câmaras.

Khandelwal P. e Stone P. [17] desenvolveram um sistema de *ground truth* para localização de uma bola laranja e jogadores futebolistas integrados na liga SPL (Standard Platform League) [3] do Robocup. Estes autores recorrem ao sensor desenvolvido pela *Microsoft*, *Kinect RGB-D* e afirmam que possuem um sistema de baixo custo, portátil, que opera em tempo real, não necessita de marcadores sobre os robôs e a sua calibração é bastante simples. Este *software opensource* que utiliza a *framework* ROS (Robot Operating System)

foi testado recorrendo ao posicionamento do robô NAO em diferentes pontos de vista sobre um ponto conhecido, os resultados estão na Tabela 1. O erro “global” com valor de 10.41 centímetros contempla erros do posicionamento do robô e calibração de extrínsecos com base em 20 *frames* capturadas e uma taxa de deteção de 95.64% pela *Kinect*.

Tabela 1: Erro de posicionamento do robô NAO em [17]

Tipo	Variação do erro (cm)
Robô (frente)	10,19 ($\pm 5,86$)
Robô (direita)	10,90 ($\pm 5,87$)
Robô (atrás)	9,72 ($\pm 4,55$)
Robô (esquerda)	10,87 ($\pm 6,97$)
Robô (global)	10,41 ($\pm 5,85$)

Ferrein A. *et al* projetaram um sistema de monitorização de posições e orientações de várias partes do tronco e cabeça de um robô NAO [18]. Recorrem ao sistema Vicon, composto por um conjunto de marcadores passivos, (identificação dos pontos de interesse) e 15 câmaras, cobrindo uma área de 6x4m a uma taxa de 120Hz. A aplicação do sistema Vicon caracteriza-se pela capacidade de acompanhamento do movimento do corpo humano em 6D com uma precisão inferior a um milímetro uma vez que o robô Humanoide NAO tem uma estrutura semelhante ao humano. Na Figura 6 é visível a distribuição de marcadores pelo corpo do robô criado pela empresa Aldebaran.

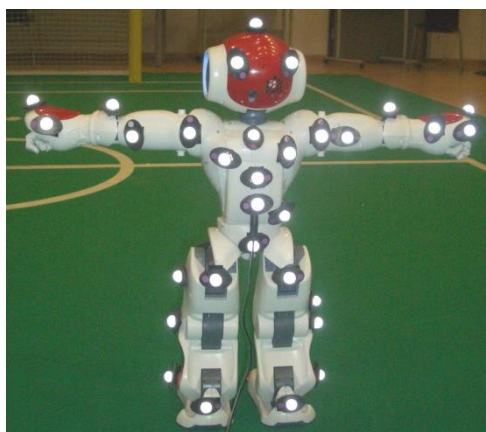


Figura 6: Distribuição de marcadores no robô NAO [18]

Nunes N. [1] contribuiu para o desenvolvimento de um sistema de visão multi-câmara para determinação da localização e atitude de múltiplos sistemas autónomos presentes na unidade de robótica do LSA. O sistema foi utilizado recorrendo a um par de câmaras e três LEDs de alta luminosidade como marcadores ativos instalados nos sistemas robóticos. Para diminuição do erro e prevenir situações de oclusão dos marcadores o autor aplicou um Filtro de Kalman.

Stulp F, *et al* [19] descreve um sistema de *ground truth* por visão, baseado em marcadores para determinar a posição 2D e orientação de vários robôs futebolistas que participam na liga small size da Robocup. O sistema recorre a marcadores coloridos (Figura 7) instalados no topo dos robôs que são registados em imagens de duas câmaras, instaladas no topo do campo de futebol robótico. Com uma taxa de 15Hz, um objeto estático apresenta um erro de posição inferior a 5,2 centímetros e um erro de orientação menor que 2.3°.

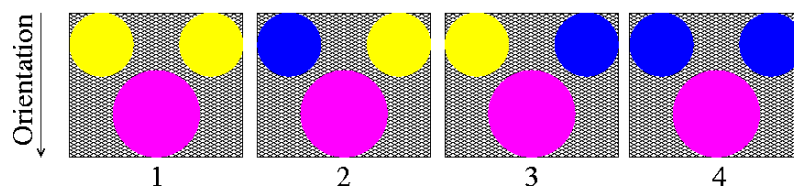


Figura 7: Marcadores utilizados em [19] para deteção e orientação de robôs futebolistas

Com algumas diferenças estruturais face aos trabalhos anteriormente apresentados mas não menos importante contribuição para sistemas de *ground truth*, foi desenvolvido o trabalho [20]. O mesmo referencia a utilização de 6 marcadores fiduciais, codificados por um ID instalados à volta do robô, e um conjunto de câmaras que captam imagens para posterior reconhecimento dos códigos e posicionamento dos mesmos.

2.3. COMPARAÇÃO DE TECNOLOGIAS PARA *GROUND TRUTH*

Tendo em conta o funcionamento das tecnologias abordadas neste capítulo, pode ser feito um balanço dos prós e contras, para que seja apurado o tipo de sensor que melhor se enquadra às necessidades do projeto abordado nesta tese.

O GNSS permite uma navegação em tempo real aplicado a diferentes veículos em simultâneo [21], para qualquer tempo global e o seu erro é independente do tempo. Este pode apresentar uma boa exatidão e precisão no cálculo da posição e permite grandes velocidades de deslocamento. Por outro lado, requer uma linha de vista aos satélites,

levando a uma degradação ou não funcionamento em situações de oclusões de satélites, como são os caso de aplicações *indoor* [22].

Os pseudolites são uma boa alternativa ao GPS em ambientes interiores, porque basicamente funcionam como recetores/transmissores (da mesma forma que os satélites GPS), porém, no solo. Naturalmente que a grande desvantagem deste sistema de localização está associada à complexidade destes satélites quando comparados por exemplo com as câmaras ou LRF's.

Sistemas conhecidos como *smartcameras* têm inúmeras vantagens, entre as quais se destacam: Existência de uma unidade de processamento integrada que permite “libertar” a “*main CPU*”, uma grande facilidade na partilha de dados e um tamanho compacto quando comparado com um sistema constituído por PC mais câmaras. Por outro lado as suas aplicações são ainda muito específicas e limitadas.

Sistemas comerciais ligados ao desenvolvimento de sistemas de *ground truth* como é o caso das empresas Vicon [12], Qualisys [15], Optitrack[14], diferem entre si no meio de aplicação (ambiente interior ou exterior), tipo de marcador, precisão, etc. No entanto, todos estes sistemas são muito robustos, precisos e possuem *software* dedicado à aplicação, o que facilita a sua utilização. Tipicamente estes sistemas requerem a utilização de um elevado número de câmaras o que torna o sistema dispendioso financeiramente. Aliada a esta desvantagem, a utilização destes sistemas requer a definição de um modelo específico de equipamento para uma dada aplicação/cenário, que implica grandes dificuldades ou até mesmo impossibilidade de aplicações noutras circunstâncias.

2.4. CONCLUSÕES DO ESTADO DAARTE

Com o desenvolvimento deste capítulo foi possível apurar algumas tecnologias que estão disponíveis no mercado, sendo que a mais apropriada a este projeto é a tecnologia ótica. No seguimento deste estudo, foram abordados alguns trabalhos dentro da temática *ground truth* que utilizam sistemas óticos como meio para chegar à posição e orientação dos sistemas robóticos e/ou objetos em estudo.

Antes de mais importa realçar que não existem soluções ótimas para todos os tipos de aplicações, ou seja, todos os sistemas têm como alvo aplicações e situações muito específicas.

No que diz respeito aos projetos com código *opensource* de *ground Truth*, existem vários sistemas robustos e eficazes, contudo dificilmente conseguem ser utilizados por outras pessoas que não os seus criadores, ou então exigem um longo processo de aprendizagem.

Relativamente aos sistemas que não utilizam marcadores apresentam uma precisão mais baixa que os sistemas que utilizam marcadores (ativos ou passivos). Em alguns projetos abordados a execução em tempo real não é uma das características, o que pode ser importante nalgumas situações. Este não é o caso por exemplo, do sistema [17], no entanto tem a desvantagem de ter uma área de operação curta quando comparada com [16].

O projeto descrito em [1], apresenta resultados de ensaios com o robô parado ou em movimento 2D recorrendo a marcadores ativos, contudo, não são apresentados resultados relativos ao comportamento do sistema quando existem variações de posição 3D.

Relativamente aos sistemas comerciais, devido ao grande número de câmaras e marcadores distribuídos pelo cenário tornam-se dispendiosos.

3. ENQUADRAMENTO TEÓRICO

Neste capítulo são apresentados alguns conceitos e fundamentos teóricos que suportam e facilitam a compreensão do funcionamento de um sistema de visão computacional.

3.1. INTRODUÇÃO À VISÃO COMPUTACIONAL

Quando o tema são sistemas robóticos, é incontornável a descrição dos sistemas sensoriais que os constituem. A percepção e entendimento do meio envolvente é de extrema importância para o desempenho das tarefas que lhes são incumbidas. Para tal, são utilizados sensores que em muito se tentam assemelhar aos órgãos sensoriais dos humanos e animais. Um dos órgãos sensoriais mais ricos em quantidade de informação adquirida do meio são os olhos, que em muito se assemelham às câmaras. Da mesma forma que os olhos captam a informação e esta é interpretada pelo cérebro, os sistemas robóticos captam a informação ótica a partir das câmaras e processam esta informação através da unidade computacional. Dadas as semelhanças e o intuito de se perceber como funcionam os sistemas de visão computacional, será realizada uma comparação entre o olho humano e sensores óticos.

No olho humano, a córnea consiste no revestimento transparente que envolve o olho, enquanto que a pupila, define a abertura da íris de forma a regular a quantidade de luz que incide na retina. A retina desempenha a função do sensor de ótico. O cristalino representa o papel de uma lente, permitindo focar a imagem na retina. Já o nervo ótico transporta a informação até ao cérebro. De forma a facilitar a associação é apresentada a Figura 8.

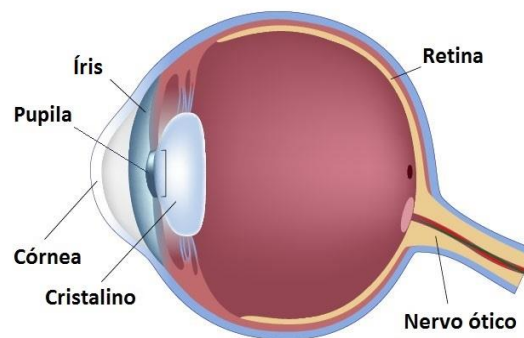


Figura 8: Anatomia do olho humano¹

Comparando o olho humano a uma câmara, podem ser estabelecidas várias relações diretas. É o caso da retina e um sensor ótico, por exemplo, do tipo CMOS (*Complementary Metal-Oxide Semiconductor*). Ambos possuem mecanismos de controlo da quantidade de luz incidente, sendo que na retina é a pupila e no caso da câmara é o obturador que controla o tempo de exposição no sensor ótico, ou então os sistemas de Auto Íris nas lentes.

Nem só de semelhanças é feita a comparação entre o órgão sensorial humano e as câmaras. Algumas das diferenças que se destacam, é o facto da retina humana ser côncava ao contrário dos sensores óticos que são planos. Ao nível do foco, as lentes comuns têm em consideração a captura de imagens ao longe ou ao perto, por outro lado o olho humano consegue mudar a forma do cristalino e adaptar-se às diferentes distâncias dos objetos [23].

¹ <http://tranjan.com.br/problemas-de-visao/anatomia-do-olho/>, acedido a 1/2/2016

3.1.1. IMAGEM DIGITAL

Até chegarmos à imagem digital, significativos foram os avanços por parte de diversos autores relacionados com a fotografia analógica. Primariamente, neste âmbito, é possível destacar a invenção da câmara escura, que não permitia a gravação das imagens. Posteriormente foram explorados materiais fotossensíveis que já permitiam o registo das imagens. Muitos anos passaram e em 1957 Russel Kirsch utilizou um dispositivo (protótipo de um *scanner*) que convertia uma imagem analógica numa matriz binária [24].

As imagens digitais podem ser divididas em dois tipos: as raster e as vetoriais. As imagens raster contêm informação acerca da cor de cada píxel. As imagens vetoriais são imagens que contêm formas geométricas tais como elipses, polígonos, ou curvas, que podem ser descritas por vetores matemáticos, tal como ilustrado na Figura 9.

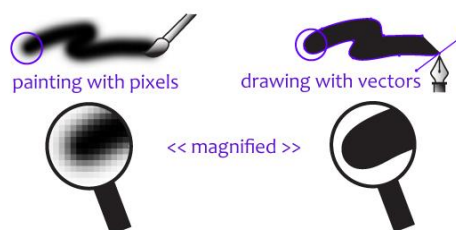


Figura 9: Diferença entre imagem raster e vetorial¹

Imagens do tipo vetorial podem ser redimensionáveis sem perda de qualidade de imagem e representadas por parâmetros matemáticos, que descrevem as curvas existentes na figura, por consequência estas imagens são geralmente mais pequenas (em número de bits) que as imagens raster. Por outro lado as imagens do tipo raster monocromáticas, são representadas por uma função bidimensional $f(x, y)$ (em coordenadas cartesianas) da intensidade luminosa nos vários pontos da imagem. Para se produzir uma imagem digital a partir de uma função contínua de luminosidade é necessário amostrar espacialmente a função e percorrer um processo de quantização dos valores de luz. A amostragem espacial é feita ao longo de diferentes tipos de grelhas (matrizes), nomeadamente: retangulares, hexagonais, quadradas. A título de exemplo considera-se uma grelha retangular, em que o tamanho das

¹ http://vector-conversions.com/vectorizing/raster_vs_vector.html, acedido a 4/4/2016

grelhas é definido pela resolução das imagens e o número de pixels na horizontal e vertical é definido por M e N respectivamente, de acordo com a Figura 10 [25].

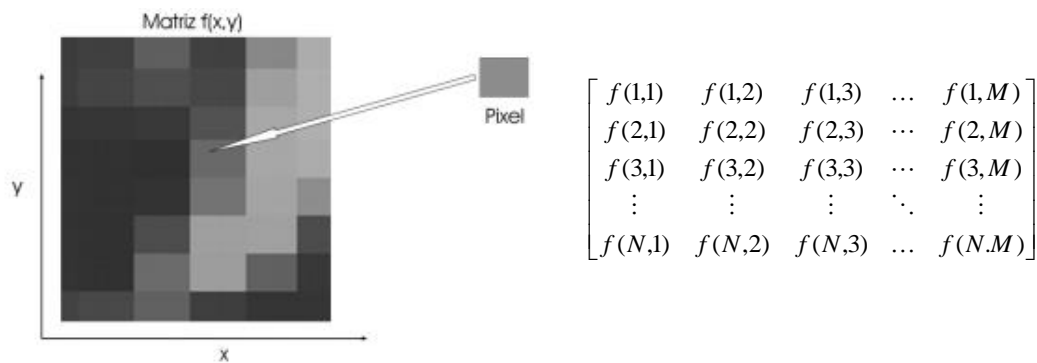


Figura 10: Representação de imagem digital¹

Quando o objetivo é definir uma imagem em tons de cinza, cada elemento da matriz define a intensidade do correspondente pixel. Usualmente este elemento está compreendido entre 0 e 255, cor preta e branca respectivamente, de acordo com as limitações da função bidimensional $0 \leq f(x, y) \leq 255$.

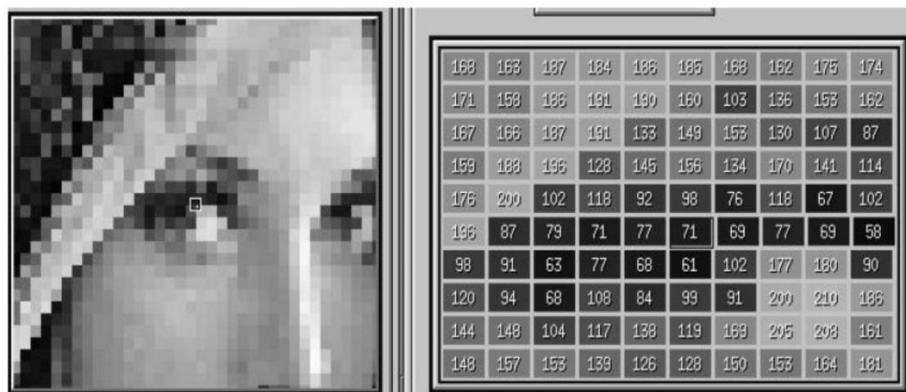


Figura 11: Representação de uma imagem em tons de cinza²

¹ <http://www.tecnologiaradiologica.com/digital.htm>, acessado a 1/2/2016

² <https://radiologiaeinovacao.files.wordpress.com/2010/06/imagem-digital-www-radiologiaeinovacao-com-br.pdf> acessado a 12/2/2016

O processo de representação de imagens coloridas é semelhante às imagens com tons de cinza, porém, estas são compostas por três matrizes. Cada matriz especifica a quantidade de vermelho (R), verde (G) e o azul (B) que compõem a imagem. Este sistema de cores é conhecido como RGB e o conjunto dos elementos destas três matrizes num mesmo ponto determinam a intensidade de cor do píxel.

Embora o cérebro humano não utilize imagens digitais, é intenção das comunidades científicas de visão e inteligência artificial reproduzir a capacidade que os seres humanos têm para reconhecer objetos. Para executar esta tarefa recorre-se a algoritmos computacionais hierarquizados que exploram cálculos matemáticos aplicados a matrizes conforme descrito no próximo capítulo.

3.1.2. PROCESSAMENTO DE IMAGEM

Processamento digital de imagens, pode ser descrito como sendo uma manipulação de imagens via computador (sistema com processamento). O objetivo deste processo é melhorar o aspeto visual de certas feições estruturais e proporcionar uma maior facilidade na extração de informação. Por outras palavras, o processamento de imagem é o que nos permite obter os parâmetros matemáticos do que se pretende extrair de uma imagem. Genericamente pode dizer-se que existem três fases neste processamento [26]:

Baixo nível ou Pré-Processamento

Este tipo de processamento atua ao nível do píxel e envolve operações primitivas tais como: pré-processamento de imagem para redução de ruído, aumento de contraste e suavização de imagem. Estas tarefas podem ser de realce de fronteiras, filtragem, eliminação de ruído, ou alteração do tamanho da imagem. Estas operações caracterizam-se pelo facto dos objetos de entrada e de saída serem imagens.

Médio nível ou Processamento

Este processamento envolve tarefas mais avançadas tais como, segmentação, partição de uma imagem em regiões ou objetos e reconhecimento de objetos individuais. O objeto de entrada é uma imagem mas a saída pode ser constituída por um conjunto de atributos extraídos das imagens, que podem ser parametrizados tais como: contornos, deteção de cores.

Alto nível ou Pós-Processamento

Este grupo envolve a interpretação do conteúdo da imagem, numa tentativa de executar funções cognitivas normalmente aproximadas à visão humana. Este tipo de processamento está geralmente associado à análise e tratamento de imagem, para fornecimento ao sistema de controlo superior.

3.2. MODELO DA CÂMARA

A maior parte dos algoritmos desenvolvidos para visão computacional partem do princípio que o modelo da câmara usada, segue o modelo da câmara de *Pinhole*.

Este modelo, descreve a relação matemática entre as coordenadas de um ponto 3D e a sua projeção sobre o plano de imagem de uma câmara *Pinhole* ideal, onde a abertura da câmara é descrita como um ponto sem a focagem da luz por parte da lente. O modelo não inclui, por exemplo, distorções geométricas ou objetos desfocados pelas lentes e aberturas de tamanho finito. Este, também não leva em conta que a maioria das câmaras tem coordenadas de imagem discretas. Isto significa que o modelo da “câmara *Pinhole*” só pode ser usado como uma aproximação de primeira ordem do mapeamento de uma perspectiva 3D para uma imagem 2D. Os erros do modelo dependem da qualidade da câmara e das lentes e sofre um aumento desde o centro da imagem até às margens desta, onde o efeito de barril é mais acentuado nas margens. Alguns destes erros podem ser compensados por algoritmos ou filtros.

Todos os raios de luz emitidos ou refletidos por objetos passam através do *Pinhole* e formam uma imagem invertida deste mesmo objeto no plano da imagem (ver Figura 12). Com isto toma-se como certo que cada ponto no mundo corresponde a um só ponto na imagem. A este fenómeno de projeção do espaço 3D para um plano dá-se o nome de projeção perspectiva [27].

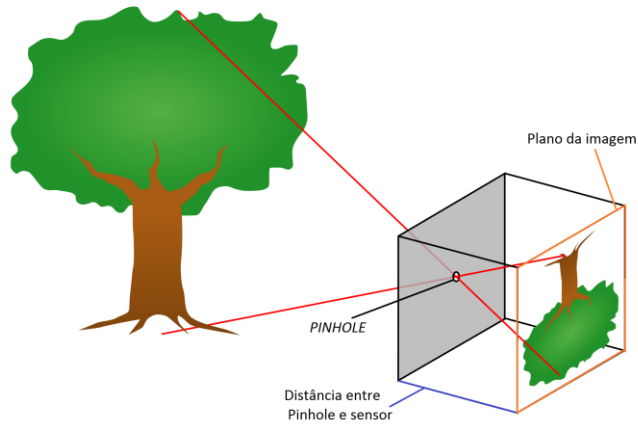


Figura 12: Representação típica do modelo da câmara de *Pinhole*¹

3.2.1. PROJEÇÃO PERSPETIVA

Num sistema visual monocular que se rege pelo modelo *Pinhole* a coordenada de cada ponto pode ser deduzida através da geometria dos triângulos semelhantes, tal como definido na Equação 3.1. O mesmo modelo caracteriza a câmara através do centro ótico C.O., o plano da câmara R^c e o plano de imagem R^i , como se pode observar na Figura 13. A distância que separa os dois planos é designada distância focal f .

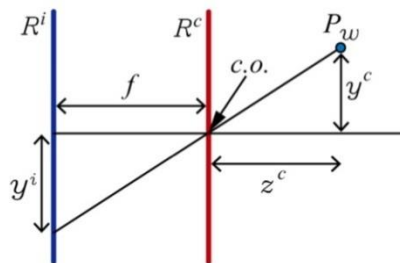


Figura 13: Modelo da câmara de *Pinhole*

$$\frac{-y^i}{f} = \frac{y^c}{z^c} \Leftrightarrow -y^i = f \cdot \frac{y^c}{z^c} \quad (3.1)$$

¹ Adaptado de https://en.wikipedia.org/wiki/Pinhole_camera_model

Para a representação e explicação dos conceitos apresentados na Figura 14, começa-se por definir um eixo de coordenadas na câmara. O eixo Z aponta na direção frontal da câmara, sendo esse denominado de eixo ótico ou eixo principal. O plano principal representa o lado da frente da câmara, em que o ponto C é a origem deste referencial. O plano que contém os pontos do mundo projetados através da câmara é designado por plano de imagem, ficando este paralelo aos eixos XY e afastado do C.O por uma distância f . Relacionando geometricamente o plano de imagem e o plano da câmara, é possível descrever a projeção perspectiva por uma equação linear 3.2.

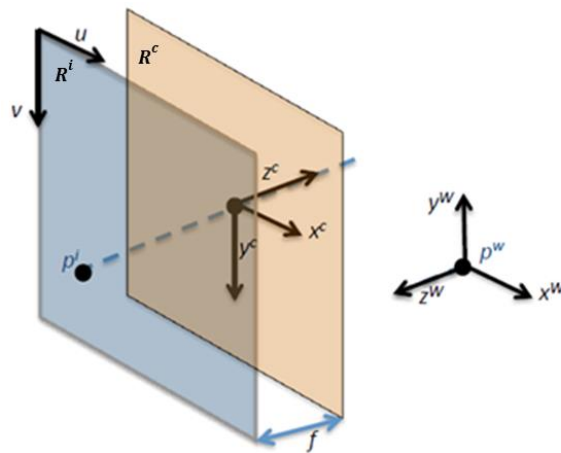


Figura 14: Representação tridimensional de um ponto no plano da câmara e da imagem.

$$\begin{bmatrix} x^i \\ y^i \end{bmatrix} = -\frac{f}{Z^c} \begin{bmatrix} X^c \\ Y^c \end{bmatrix} \quad (3.2)$$

As coordenadas (x^i, y^i) são as coordenadas métricas no plano da imagem (R^i) e (X^c, Y^c, Z^c) a posição do ponto no referencial da câmara (R^c). De acordo com o modelo de *Pinhole*, o plano da imagem está localizada atrás de f devido à luz que atravessa o orifício do centro ótico produzir uma imagem invertida. Para ultrapassar esta inversão, desloca-se o plano da imagem de modo a que este se posicione entre o ponto P^w e o centro ótico mantendo assim a mesma distância focal (Equação 3.3) [1].

$$\begin{bmatrix} x^i \\ y^i \end{bmatrix} = \frac{f}{Z^c} \begin{bmatrix} X^c \\ Y^c \end{bmatrix} \quad (3.3)$$

A relação entre os pontos no referencial do mundo (X^w, Y^w, Z^w) e os pontos no referencial da câmara podem ser obtidas através de uma transformação composta por uma rotação e uma translação em coordenadas homogêneas como descreve a equação 3.4.

$$\begin{bmatrix} X^c \\ Y^c \\ Z^c \end{bmatrix} = [R|\mathbf{t}] \begin{bmatrix} X^w \\ Y^w \\ Z^w \\ 1 \end{bmatrix} \quad (3.4)$$

A matriz de rotação (R) contém a orientação da câmara num dado referencial e \mathbf{t} é o vetor de translação que representa a posição da origem do referencial do mundo da câmara. A matriz R e o vetor \mathbf{t} são resultado do processo de calibração dos parâmetros extrínsecos da câmara, que será abordado no Capítulo seguinte.

Combinando as equações 3.3 e 3.4, obtém-se a equação 3.5 que permite relacionar um ponto no mundo com o referencial normalizado da câmara, recorrendo à Matriz de projeção perspectiva (PPM) [28].

$$\begin{bmatrix} x^i \\ y^i \\ 1 \end{bmatrix} = PPM \begin{bmatrix} X^w \\ Y^w \\ Z^w \\ 1 \end{bmatrix} \quad (3.5)$$

A matriz PPM contém informação dos parâmetros intrínsecos e extrínsecos, sendo representada pela equação seguinte:

$$PPM = A[R|\mathbf{t}] \quad (3.6)$$

Os parâmetros intrínsecos são representados pela matriz A e serão abordados no Capítulo 3.3.1. Com o cálculo da matriz PPM torna-se possível mapear pontos do referencial mundo para o referencial 2D da imagem, através da equação 3.7. Onde m^I representa um ponto da imagem e m^W representa esse mesmo ponto em coordenadas mundo [28].

$$m^I \approx (PPM)m^W \quad (3.7)$$

3.3. CALIBRAÇÃO DAS CÂMARAS

A calibração das câmaras é a operação que permite determinar os parâmetros intrínsecos e extrínsecos das câmaras. Com os parâmetros intrínsecos é possível definir as características da câmara e da lente, enquanto que, os parâmetros extrínsecos representam a posição da câmara face ao mundo num determinado sistema de coordenadas.

Vários são os procedimentos que se podem adotar para se obter os parâmetros das câmaras, no entanto um dos mais utilizados é o algoritmo de Tsai [29]. Este algoritmo para além de fornecer os parâmetros das câmaras também devolve os coeficientes de distorção da lente, com base no modelo de *Pinhole*.

3.3.1. PARÂMETROS INTRÍNSECOS

Os chamados parâmetros intrínsecos de uma câmara são as constantes que definem como se dá a transformação das coordenadas métricas no referencial da câmara para as respetivas coordenadas de um ponto em píxeis. Para a construção do modelo matemático da câmara, existem algumas não idealidades que têm que ser tidas em conta, tais como: distância focal, centro ótico e distorção da lente.

A distância focal (f) representa a distância entre o plano da imagem e o plano da câmara segundo o modelo de *Pinhole*.

O ponto principal é o centro ótico projetado para a imagem, o mesmo é dizer que corresponde ao centro da lente, representa-se por c_x e c_y .

A distância focal e o ponto principal são expressos em píxeis e formam a matriz (A) que estabelece a relação entre o plano da imagem e o plano da câmara.

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3.8)$$

Através da matriz A é possível relacionar um ponto do referencial da câmara normalizado com um ponto no plano da imagem em píxeis. De acordo com o modelo de *Pinhole*, um conjunto de retas no mundo ao serem projetados no plano da imagem deveriam apresentar o mesmo aspeto, no entanto isto não acontece porque as linhas são afetadas por uma deformação [28].

A deformação abordada deve-se à distorção existente na lente da câmara. Este fenómeno pode ser mais notório quando a distância focal é maior ou a qualidade da lente é menor. Uma consequência da distorção das lentes é a projeção de linhas retas do espaço tridimensional são curvas na imagem, Figura 15.

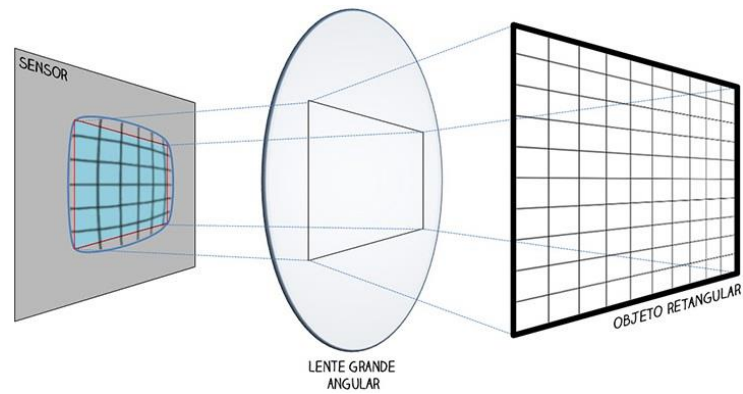


Figura 15: Distorção (radial negativa) de linhas retas provocadas pela lente¹

A distorção da lente pode ser caracterizada segundo duas componentes: radial e tangencial. A distorção radial provoca um deslocamento para dentro ou para fora de um determinado ponto a partir da posição ideal. Este tipo de distorção é causado principalmente por uma falha na curvatura da lente. O deslocamento radial negativo dos pontos da imagem é designado por distorção em barril. Neste tipo de distorção negativa, os pontos da periferia concentram-se e a escala diminui (as lentes têm maior ampliação no centro do que na periferia). O deslocamento radial positivo dos pontos da imagem é designado por distorção em almofada. Neste tipo de distorção positiva, os pontos da periferia estendem-se e a escala aumenta [30] [31]. Na Figura 16 é ilustrada a distorção radial, onde a linha contínua representa a imagem sem distorção e as linhas descontínuas a imagem com distorção radial (a; negativa-barril, b; positiva-almofada).

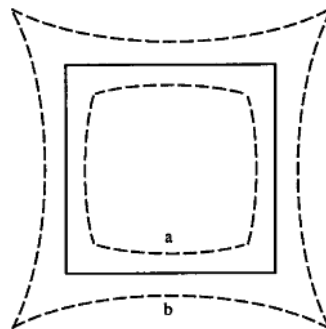


Figura 16: Efeito de distorção radial [31].

¹ Imagem alterada a partir de http://www.cameraneon.com/ngg_tag/, acedido a 28/3/2016.

A distorção tangencial é causada por um desalinhamento físico de elementos que constituem a lente. Atualmente o impacto desta distorção é muito reduzida, principalmente em lentes com distância focal fixa [32].

Com base nos coeficientes de distorção da lente, é possível preencher o vetor de distorção da imagem (equação 3.9), onde K_1 , K_2 , K_3 , são os coeficientes de distorção radial enquanto que K_4 e K_5 , são os coeficientes de distorção tangencial.

$$K = (K_1, K_2, K_3, K_4, K_5) \quad (3.9)$$

Para um correto mapeamento dos pontos 3D no plano da imagem, torna-se indispensável a remoção da distorção provocada pela lente. Recorrendo ao modelo polinomial de Brown [33] (equação 3.10), considera-se x_u e y_u as coordenadas de projeção normalizadas de um ponto não afetado pela distorção, x e y são os pontos na imagem com distorção. K_i são os coeficientes de distorção da imagem e r a distância ao centro ótico.

$$\begin{bmatrix} x_u \\ y_u \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} (1 + K_1 r^2 + K_2 r^4 + K_3 r^6) + dx$$

$$r = \sqrt{x^2 + y^2} \quad (3.10)$$

$$dx = \begin{bmatrix} 2K_4 xy + K_5 (r^2 + 2x^2) \\ K_4 (r^2 + 2y^2) + 2K_4 xy \end{bmatrix}$$

3.3.2. PARÂMETROS EXTRÍNSECOS

Os parâmetros extrínsecos descrevem uma transformação de coordenadas tridimensionais entre o plano do mundo e o plano da câmara. Esta transformação caracteriza-se por uma matriz rotação R e uma translação \mathbf{t} entre os referenciais, como se pode observar na Figura 17.

O vetor de translação \mathbf{t} contém o ponto de origem do referencial do mundo no referencial da câmara e é dado pela seguinte equação [28].

$$\mathbf{t} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (3.11)$$

A matriz de rotação R pode ser representada em função dos ângulos de Euler, Roll (ϕ), Pitch (θ) e Yaw (ψ) de acordo com a equação 3.12:

$$R = R_{(x,\phi)}^T R_{(x,\theta)}^T R_{(x,\psi)}^T \quad (3.12)$$

Na Figura 18 estão representados os ângulos de Euler em função do referencial de uma câmara

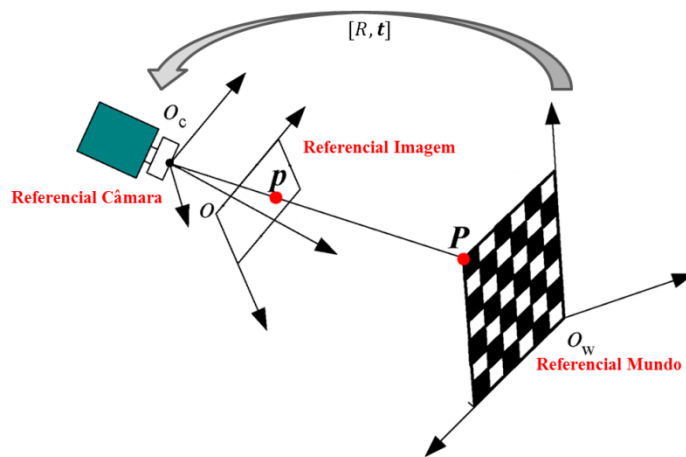


Figura 17: Projeção das coordenadas do objeto no mundo em coordenadas da câmara¹

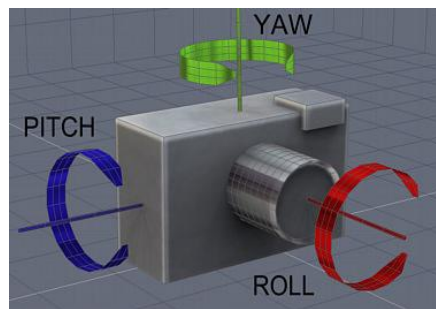


Figura 18: Representação dos ângulos de Euler numa câmara¹

¹ Imagem alterada a partir de, http://www.mdpi.com/sensors/sensors-15-08664/article_deploy/html/images/sensors-15-08664-g002-1024.png

A equação 3.12 é composta por três rotações puras, o mesmo é dizer que cada uma das rotações ocorre apenas num eixo. As rotações puras que compõem a matriz R (3.16), são formadas a partir das matrizes de rotação em torno de um dado eixo, apresentadas pelas equações 3.13, 3.14 e 3.15 [34]:

Roll (ϕ), composta por uma rotação em torno do eixo X

$$R_{(x,\phi)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi \\ 0 & \sin \phi & \cos \phi \end{bmatrix} \quad (3.13)$$

Pitch (θ), formada por uma rotação em torno do eixo Y

$$R_{(y,\theta)} = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \quad (3.14)$$

Yaw (ψ), desenvolvida a partir de uma rotação em torno do eixo Z

$$R_{(z,\psi)} = \begin{bmatrix} \cos \psi & -\sin \psi & 0 \\ \sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.15)$$

Em função dos ângulos de Euler a matriz de rotação R é dada por:

$$R_{(z,\psi)} = \begin{bmatrix} \cos \theta \cos \psi & \cos \theta \sin \psi & -\sin \theta \\ \cos \psi \sin \phi \sin \theta - \cos \theta \sin \psi & \sin \psi \sin \phi \sin \theta + \cos \phi \cos \psi & \sin \phi \cos \theta \\ \cos \psi \sin \theta \cos \phi + \sin \phi \sin \psi & -\sin \phi \cos \psi + \sin \psi \sin \theta \cos \theta & \cos \theta \cos \phi \end{bmatrix} \quad (3.16)$$

¹ Imagem alterada a partir de http://www.mediachance.com/hdri/help/index.html?pin_warping.htm e [https://msdn.microsoft.com/en-us/library/windows/desktop/bb281744\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/windows/desktop/bb281744(v=vs.85).aspx)

3.4. VISÃO STEREO

A implementação do modelo de *Pinhole* para se determinar a projeção perspectiva não permite a transformação de pixels no plano da imagem em pontos 3D, uma vez que, uma imagem 2D por si só não contém informação suficiente acerca da profundidade do objeto. Estabelecendo uma analogia com o sistema de visão humano, a percepção de profundidade é obtida pela diferença entre a projeção dos pontos de uma estrutura 3D no olho esquerdo e direito. É a partir desta visão *stereo* que o ser humano avalia a profundidade e julga as distâncias [34].

A perda de informação que ocorre ao projetar-se uma estrutura do mundo real 3D no plano bidimensional de uma imagem faz com que haja ambiguidade na solução do problema. De acordo com o modelo *Pinhole*, vários pontos 3D situados num seguimento de reta com origem no centro ótico, têm o mesmo ponto no plano da imagem. Desta forma, o inverso da projeção perspectiva de um píxel gera um segmento de reta semi-infinito, que interceta as coordenadas do ponto na imagem com origem no centro ótico[28]. Esta ambiguidade é considerada em técnicas de reconstrução 3D que de acordo com o número de perspectivas utilizadas na reconstrução, podem basear-se em técnicas monoculares, binoculares ou de múltiplas perspectivas.

3.4.1. GEOMETRIA EPIPOLAR

A geometria epipolar entre duas perspectivas é essencialmente a geometria da intersecção dos planos de imagem com a linha que une os centros das câmaras, denominada de *baseline*. Esta geometria tem como principal objetivo auxiliar a correspondência de pontos em diferentes imagens, para posterior aplicação de técnicas de reconstrução tridimensional [32].

Tendo como base duas imagens com o mesmo cenário mas obtidas a partir de perspectivas diferentes, é possível validar pontos comuns utilizando a geometria epipolar. Este método possibilita a identificação de pontos comuns entre duas imagens, com o pressuposto que a perspectiva é estática naquele instante e que os pontos alvo no espaço estão registados em ambas as imagens [35].

De acordo com a Figura 19 é visível o princípio do modelo da geometria epipolar, em que o ponto tridimensional P é projetado na imagem da câmara da esquerda e de direita

identificadas como p_l e p_r , respectivamente. Os pontos O_l e O_r representam o centro ótico de cada uma das câmaras, enquanto que a linha que une os dois centros óticos é denominada *baseline* (linha de base).

A interseção entre a *baseline* e o plano da imagem é descrito como sendo o epipólo (e_l na câmara da esquerda e e_r na câmara da direita). Um epipólo representa a projeção de um centro ótico de uma câmara no plano da imagem da outra câmara. O plano delimitado pelos centros óticos (O_e e O_d) e o ponto 3D (P) é o plano epipolar Π , sendo que o cruzamento entre o plano epipolar e o plano das imagens origina as linhas epipolares (l e l'), em que todas elas intercetam os respectivos epipolos.

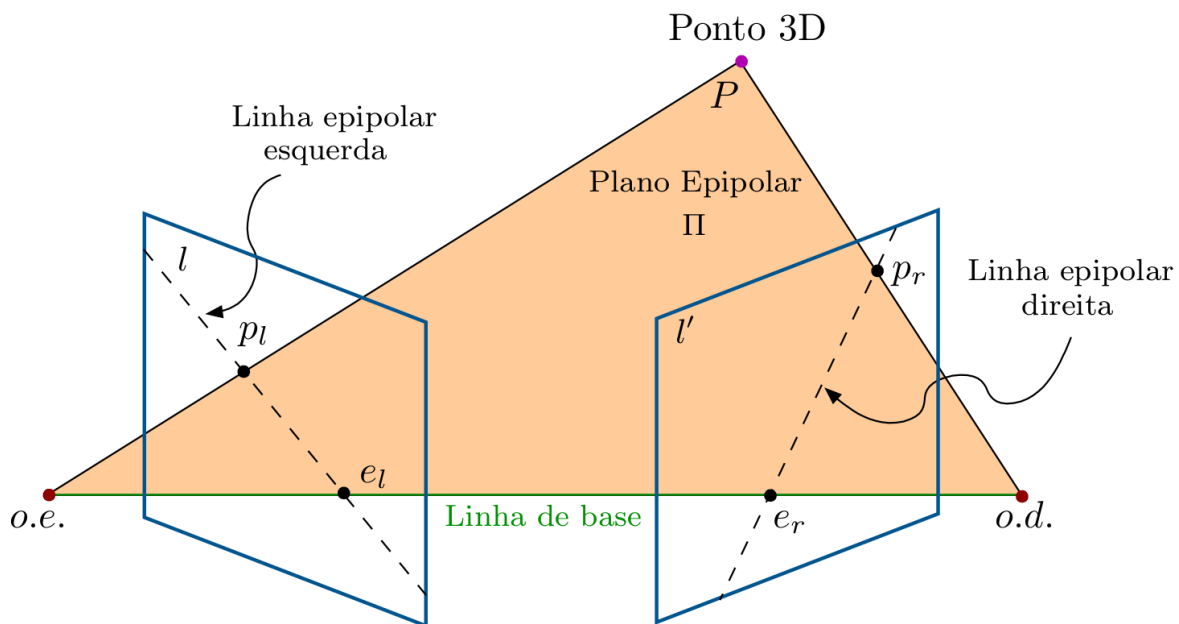


Figura 19: Modelo de Geometria epipolar

Com um ponto tridimensional P , projetado no plano da imagem da esquerda (p_l) e esse mesmo ponto projetado no plano da imagem da direita (p_r), permite afirmar que esse ponto está posicionado ao longo da linha epipolar r , o mesmo se aplica na situação inversa. A propriedade da geometria epipolar que permite validar a correspondência de pontos entre imagens, é chamada de restrição epipolar. Através desta restrição, é possível validar se um ponto projetado na imagem da esquerda existe na imagem da direita, recorrendo à expressão 3.17 [32].

$$p_r^T F p_l = 0 \quad (3.17)$$

No qual F representa a matriz fundamental de dimensão 3×3 , p_r e p_l são os pontos projetados nos planos das imagens da direita e da esquerda respetivamente. De acordo com a expressão 3.17, se o resultado da equação tomar valores próximos de zero a probabilidade do ponto da imagem esquerda estar na imagem da direita é maior, o mesmo é afirmar que esse mesmo ponto se encontra sobre a linha epipolar da esquerda (r).

Algebricamente a restrição epipolar pode ser representada pela matriz fundamental, ou matriz essencial (equação 3.18). A diferença entre a utilização da matriz fundamental ou essencial incide sobre a referência com base no plano da imagem ou da câmara respetivamente.

$$(A_r^{-1}p_r)^T E A_l^{-1}p_l = 0 \quad (3.18)$$

Assim sendo A_r são os parâmetros intrínsecos da câmara da direita, A_l Intrínsecos da câmara da esquerda, p_r e p_l são os pontos no plano das câmaras da direita e esquerda respetivamente, enquanto que E , é a matriz essencial.

A matriz essencial pode ser calculada a partir dos parâmetros que relacionam as duas câmaras, ou seja as matrizes de rotação R e translação \mathbf{t} stereo, de acordo com a expressão 3.19

$$E = T_x \times R \quad (3.19)$$

Em que T_x representa a matriz *skew-symmetric* de \mathbf{t} dada pela matriz representada em 3.20.

$$T_x = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (3.20)$$

A matriz de rotação *stereo* pode ser calculada com a equação 3.21. Nesta, R_R representa a matriz de rotação da câmara da direita e R_L a matriz de rotação da câmara da esquerda

$$R = R_r \times R_l^T \quad (3.21)$$

Com a matriz de rotação *stereo* R e os vetores de translação da câmara da esquerda \mathbf{t}_l e da câmara da direita \mathbf{t}_r , obtém-se o vetor de translação *stereo*, representado pela equação 3.22 [28].

$$\mathbf{t} = \mathbf{t}_l - R^T \mathbf{t}_r \quad (3.22)$$

A matriz fundamental é uma alternativa à matriz essencial, no entanto esta integra os parâmetros intrínsecos das câmaras, o mesmo é dizer que opera no plano da imagem. Muitos são os processos que se podem utilizar para o cálculo desta matriz como são os casos de:

- Algoritmo dos 8 pontos;
- Algoritmo dos 7 pontos;
- Algoritmo de LMeds;
- MAPSAC.

Um dos processos frequentemente utilizado para se obter a matriz fundamental é através da relação entre a matriz essencial e a matriz fundamental, tal como apresenta a equação 3.23.

$$F = A_r^{-T} E A_l^{-1} \quad (3.23)$$

A matriz essencial é representada por E , A_l e A_r são os parâmetros intrínsecos das câmaras da esquerda e da direita respetivamente.

Com recurso à matriz fundamental, podem ser calculados os epipolos de acordo com as expressões 3.24 e 3.25.

$$F e_l = 0 \quad (3.24)$$

$$F^T e_r = 0 \quad (3.25)$$

Acrescentando à matriz fundamental os pontos nos planos das imagens em píxeis p_l e p_r obtêm-se as linhas epipolares com as equações seguintes:

$$l_l = F^T p_r \quad (3.26)$$

$$l_r = F p_l \quad (3.27)$$

3.4.2. RETIFICAÇÃO *STEREO*

O processo de retificação *stereo* consiste na realização de uma transformação de um par de imagens *stereo* para alinhá-las horizontalmente, reduzindo assim ao eixo horizontal a pesquisa de correspondência de pontos entre imagens [36] [37].

Este processo faz uma redistribuição dos pixels em cada imagem de forma a que o alinhamento dos centros focais e dos planos da imagem das duas câmaras sejam os mesmos. Desta forma, os epípolos são projetados para o infinito, o que torna as linhas epipolares horizontais e colineares, facilitando a correspondência entre imagens. O processo utilizado para a deformação das imagens é baseado na aplicação da matriz de homografia a cada imagem. Uma vez que os epípolos se encontram no infinito, as linhas epipolares tornam-se paralelas relativamente ao eixo de coordenadas x da imagem.

Os métodos para cálculo das matrizes de homografia partilham a preocupação de minimização da distorção das imagens e maximização da área vista em comum pelas câmaras [37][38].

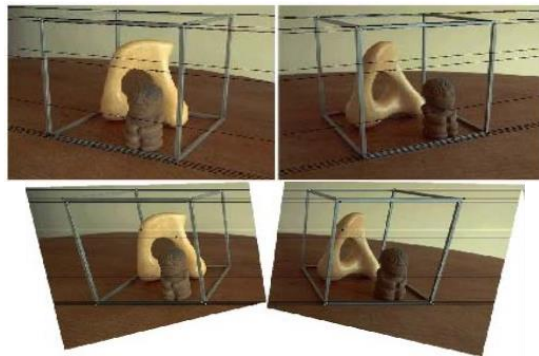


Figura 20: Par *stereo* não retificado e par *stereo* retificado [35].

Como é perceptível na Figura 20, através do processo de retificação, ocorreu um alinhamento das linhas epipolares, que permite restringir a pesquisa de correspondência de pontos entre o par de imagens ao longo do eixo do x. Assim sendo a geometria do sistema *stereo* pode ser simplificada de acordo com a Figura 21. Recorrendo às equações 3.28, 3.29, 3.30 é possível calcular o ponto tridimensional resultante da triangulação entre duas imagens.

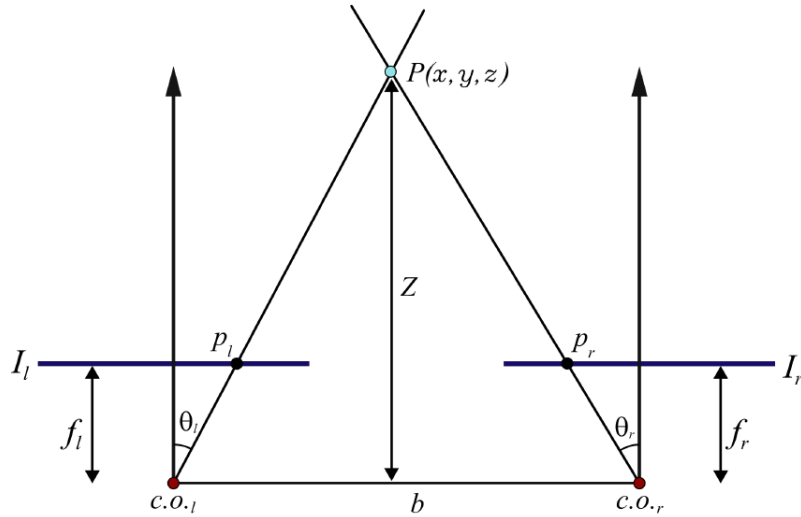


Figura 21: Geo. de par *stereo* retificado com linhas epipolares paralelas e epipolos no infinito [1].

$$z = \frac{fb}{x_l - x_r} \quad (3.28)$$

$$y = \frac{y_l z}{f} \quad (3.29)$$

$$x = \frac{x_l z}{f} \quad (3.30)$$

x_l e x_r são coordenadas em x dos pontos p_l e p_r respetivamente. f representa a distância focal, b a *baseline* (distância entre centros óticos das câmaras), enquanto que y_l é a coordenada em y do ponto p_l . Neste caso a coordenada em y é igual em ambas as imagens, dado que estas são imagens retificadas.

3.4.3. TRIANGULAÇÃO STEREO

O método para determinação de uma coordenada 3D, a partir de pontos extraídos de imagens (em pixéis) com perspetivas diferentes é designado de triangulação[35].

O cálculo da triangulação *stereo* permite determinar as coordenadas 3D sem recorrer a retificação *stereo* de um ponto no mundo. Para se ultrapassar esta limitação, é necessário a visualização de um ponto no mundo a partir de câmaras com diferentes perspetivas. Para aplicar a triangulação são necessários os parâmetros intrínsecos e extrínsecos das câmaras, sendo imperativo que os pontos das duas imagens sejam os mesmos, para esse processo é possível recorrer à geometria epipolar de forma a validar a correspondência de pontos.

O método de triangulação *stereo* consiste na interseção de duas retas, com orientação definidas pelo, p_l e p_r que correspondem à projeção de um ponto P na câmara da esquerda e da direita respetivamente. O ponto 3D que é definido por P', pode ser obtido através da aproximação ao ponto mais próximo equidistante das duas retas. Este processo ocorre devido ao facto das retas nunca se intercetarem no espaço, consequência da aproximação dos parâmetros das câmaras e da localização das imagens. Na Figura 22 é ilustrada a aproximação das duas retas [28].

Tendo o conhecimento da matriz de rotação R , o vetor de translação \mathbf{t} e os pontos projetados nos planos das imagens de ambas as câmaras (p_l e p_r) é possível calcular os escalares a , b e c de acordo com a equação 3.31. A partir desta função é calculada a distância segundo o vetor direção normalizado \vec{p}_l (representado por a), a distância segundo o vetor normalizado \vec{p}_r (representado por b) e por fim a distância com direção perpendicular a \vec{p}_l e \vec{p}_r que é denominado por c [34].

$$a \cdot \vec{p}_l - b \cdot R^T \vec{p}_r + c(\vec{p}_l \times R^T \vec{p}_r) = \mathbf{t} \quad (3.31)$$

Com os escalares (a , b , c) é agora possível determinar o ponto P', que representa o ponto médio da interseção das duas câmaras (O_l e O_r), expresso na equação 3.32.

$$P' = a \cdot \vec{p}_l + \frac{c}{2} (\vec{p}_l \times R^T \vec{p}_r) \quad (3.32)$$

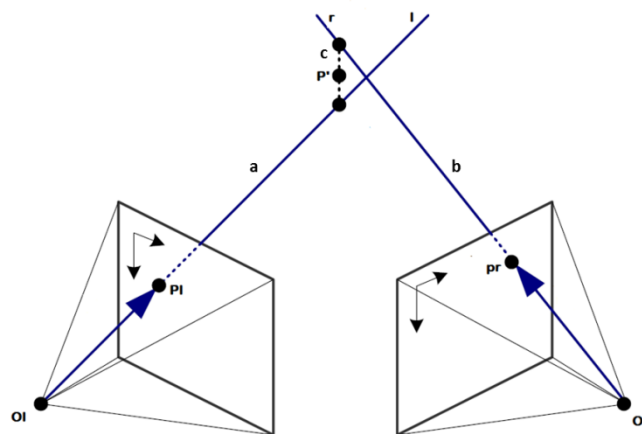


Figura 22: Modelo de triangulação *stereo* [35]

4. PROJETO

No capítulo 3 foi realizado um levantamento de técnicas de visão artificial que suporta a calibração e modelo das câmaras, estabelece correspondência de pontos entre imagens e calcula as coordenadas 3D. No presente capítulo será descrita a arquitetura global de um sistema de *ground truth* multi-câmara, aplicável a diversos meios, com resposta aos requisitos impostos por variados sistemas autónomos

4.1. REQUISITOS DO SISTEMA GLOBAL

A presente dissertação apresenta um sistema de posicionamento que serve de referência a outros sistemas autónomos. Para se atingir este objetivo, é requerido um sistema baseado em visão 3D, capaz de executar o rastreamento a múltiplos alvos em diversos meios com diferentes requisitos impostos pelas especificações dos ensaios e sistemas robóticos em estudo.

Para melhorar a precisão do sistema de coordenadas e contornar situações de oclusão, o sistema deverá contar com Leds de alta luminosidade de baixo consumo com cores que se distinguem no ambiente de trabalho. Durante a fase de ensaios, o sistema utilizará 3 câmaras distribuídas pela área de estudo, no entanto este será capaz de integrar com

facilidade n câmaras, que poderá proporcionar o aumento da área de cobertura ou então, aumentar a precisão na detecção dos marcadores em caso de oclusão parcial.

O sistema a desenvolver deve ser capaz de “passar despercebido” em ambiente de testes. Desta forma, não será necessário a instalação de nenhuma plataforma de suporte dos marcadores (pois reduz a manobrabilidade do sistemas robótico), não será necessário nenhum ajuste na luminosidade nem extinção de cores específicas do espaço.

O sincronismo entre câmaras, é um requisito crucial para garantir que as imagens são adquiridas no mesmo instante. Assim sendo, deverá ser utilizado como *hardware* um gerador de sinal *Trigger*. Para aplicações que envolvam alvos com grande rapidez e manobrabilidade, como é o caso de um *quadrotor*, o desfasamento temporal entre imagens pode traduzir-se em maus resultados no posicionamento 3D do objeto. Outro fator importante que deve ser considerado, é a utilização de uma base de tempo comum a todas as plataformas, nomeadamente as câmaras, as unidades de processamento e os robôs.

A utilização de uma unidade de processamento central é uma vantagem quando são utilizadas poucas câmaras, porque evita a aquisição de múltiplas unidades de processamento distribuído que se traduz numa redução do custo do sistema. No entanto, a projeção de um sistema com processamento distribuído poderá ser uma vantagem, quando é requerida uma execução em tempo real e a escalabilidade do sistema.

Dada a grande diversidade de aplicações que este sistema deve responder, é crucial que o mesmo seja, portátil, de fácil instalação e calibração.

De forma resumida serão apresentados a seguir os principais requisitos que o sistema global de *ground truth* a desenvolver deve ser capaz de cumprir:

- Calcular o posicionamento e atitude de múltiplos alvos em simultâneo;
- Capacidade de execução em ensaios com veículos terrestres e aéreos (no plano 2D e 3D);
- Não exigir grandes adaptações dos sistemas robóticos aos ambientes de testes;
- Capacidade de integração n câmaras sem ajustes profundos;
- Existir sincronismo entre todos os processos;
- Contornar situações de oclusão parcial do objeto;

- Cálculo da posição e atitude 3D dos alvos;
- Operação em tempo real;
- Armazenamento dos dados para posterior análise;

4.2. ARQUITETURA DO SISTEMA DE POSICIONAMENTO

Com o intuito de responder aos requisitos propostos no subcapítulo anterior, será proposta de seguida a arquitetura do sistema que integra este sistema de *ground truth* ótico.

A arquitetura funcional do sistema pode ser agrupada em quatro principais módulos, de acordo com a Figura 23.

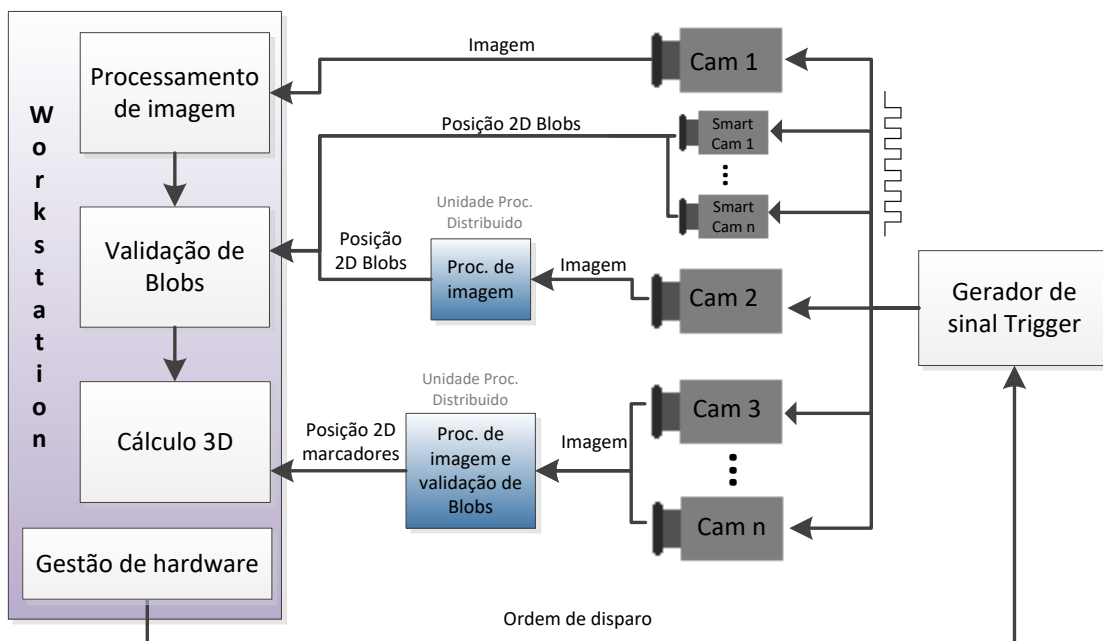


Figura 23: Arquitetura do sistema *ground truth* ótico global

- Sistema computacional - Constituído por um PC (unidade de processamento central), responsável pela ordem de captura das câmaras, processamento de imagens, validação de *blobs*, cálculo da posição e atitude do objeto alvo e por fim gravação dos dados.
- Sistemas de processamento distribuído - A arquitetura proposta permitem a integração de várias unidades de processamento distribuído. A descentralização do processamento permite libertar a unidade de processamento central do processamento de imagem, facilitando a execução em tempo real do sistema

quando o mesmo é composto por n câmaras. No entanto, dentro da workstation também poderá ser integrado um bloco de processamento de imagem e validação de pontos de interesse agrupados (*blobs*).

- Sistema de aquisição de imagens - Composto por n câmaras ou smart câmaras. O número de câmaras poderá proporcionar um aumento da área de ação e probabilidade de sucesso em caso de oclusão parcial. Estas devem de comunicar por USB 3.0 ou *Gigabit Ethernet* e ainda suportar disparo por *hardware*.
- Trigger por *hardware* – Sistema aplicado para garantir o sincronismo dos disparos das câmaras.

4.3. ARQUITETURA DO SISTEMA PROTÓTIPO DESENVOLVIDO

No presente subcapítulo será agora feita referência à contribuição desta dissertação para o sistema de *ground truth* global anteriormente descrito.

Na Figura 24 é apresentada a arquitetura base de funcionamento do sistema *ground truth*, que tem capacidade para integrar N câmaras sem que seja necessário reestruturar todo o sistema.

Na arquitetura apresentada, a Workstation é responsável pela receção e processamento das imagens, validação de pontos de interesse, cálculo de coordenadas tridimensionais e posterior gravação ou envio dos resultados obtidos. Numa fase inicial é realizada a configuração das câmaras através do protocolo de comunicação *Gigabit Ethernet*. Após configuração das câmaras, inicia-se a aquisição das imagens via mesmo protocolo de comunicação e por fim extração (cálculo) e gravação de informação tridimensional do alvo

No processo de aquisição de imagens podem ser utilizadas n câmaras. Para que se consiga extrair informação tridimensional dos alvos, são necessárias pelo menos 2 câmaras, contudo, quanto maior for o número, maior poderá ser a área de cobertura e o FOV (*Field Of View*). Situações de maior aproximação do alvo às câmaras, pode refletir-se num aumento da precisão do sistema.

O gerador do sinal de Trigger, é o responsável por geral o sinal de disparo para todas as câmaras e pode ser configurado relativamente à taxa de disparo. Este dispositivo garante o requisito de aquisição simultânea de todas as câmaras.

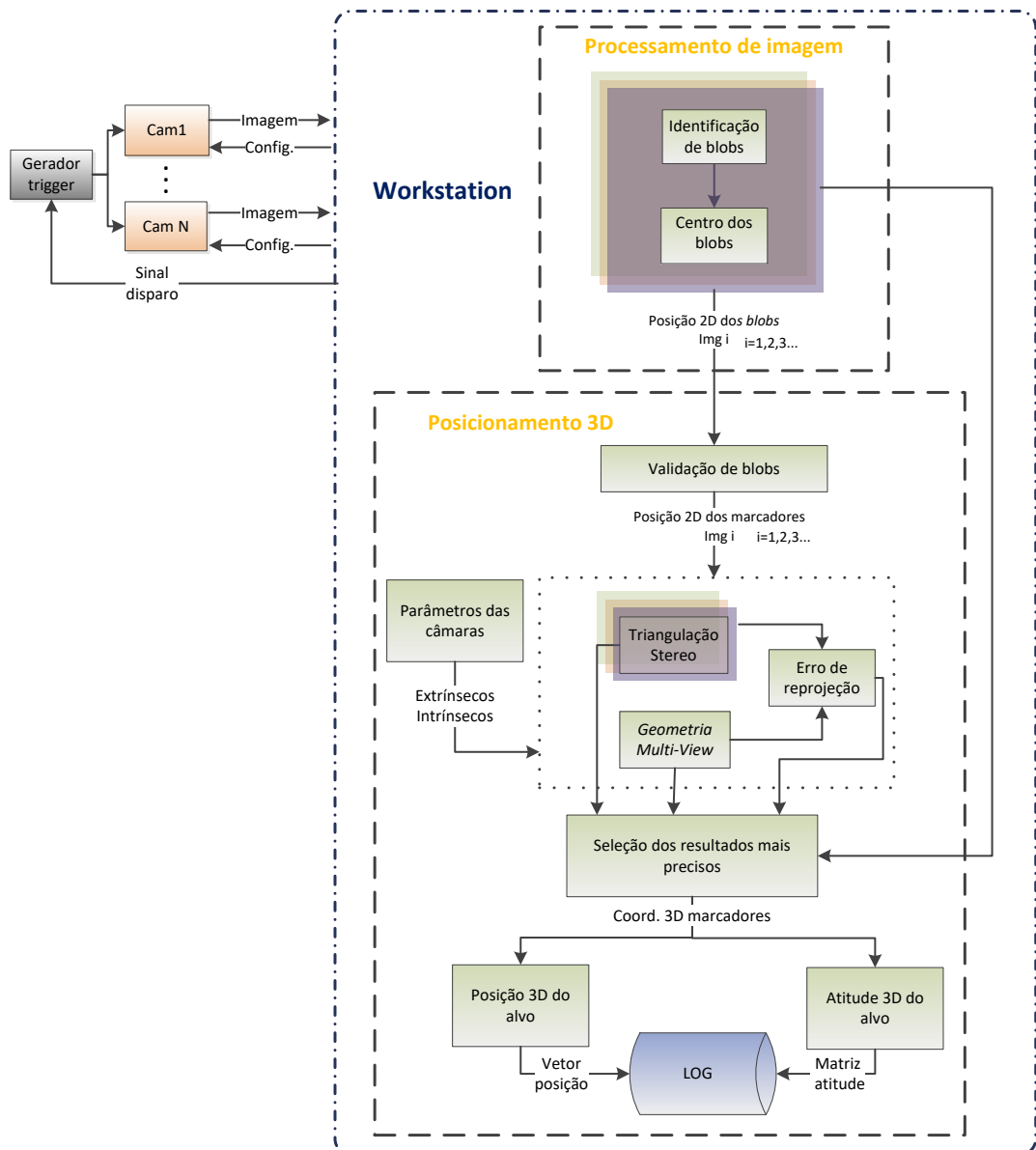


Figura 24: Arquitetura funcional do sistema

Relativamente ao software, pode-se agrupar o sistema em 3 grandes blocos. O primeiro bloco, recebe as imagens das N câmaras e para cada uma delas identifica e calcula o centro dos pontos de interesse. O resultado são as coordenadas 2D dos centros dos *blobs*. Nesta fase ainda poderão existir mais *blobs* do que marcadores, consequência de reflexos ou ruído.

No segundo bloco são validados todos os *blobs* através da restrição epipolar e espacial. Como *output*, a posição 2D dos marcadores detetados. A partir dos centros dos marcadores na imagem, são calculadas as respectivas coordenadas 3D utilizando 2 métodos:

- O primeiro método recorre à técnica de triangulação *stereo*. Combinando pares de câmaras, é possível obter k resultados, sendo que k é definido pela expressão 4.1 e N representa o número de câmaras. A título de exemplo, se o sistema for composto por 4 câmaras, existirá 6 vetores de coordenadas 3D para cada um dos marcadores.

$$k = C_{N,2} \quad (4.1)$$

- O segundo método utiliza a técnica de geometria *Multi view*. Esta técnica através da combinação de informação bidimensional de todas as imagens permite o cálculo da posição tridimensional dos marcadores led.

Com recurso às coordenadas 3D dos marcadores obtém-se novamente o ponto correspondente na imagem, permitindo o cálculo diferencial que opõem o ponto reprojeto e o inicialmente detetado.

No terceiro bloco são seleccionadas as coordenadas 3D mais precisas dos marcadores com base nos erros *stereo* e *multi view* e erro de reprojeção. Por fim será calculada a posição e atitude do alvo no espaço 3D.

5. IMPLEMENTAÇÃO

Após desenvolvimento concetual do sistema, neste capítulo, apresentar-se-á o protótipo desenvolvido “*setup experimental*”, que suporta o *software* que foi implementado em Matlab.

5.1. *SETUP EXPERIMENTAL*

Para implementação do software desenvolvido, foi necessário preparar o espaço de ensaios do laboratório e integrar o *hardware* disponível, de forma a criar um sistema multi-câmara. Neste subcapítulo são descritos estes processos.

5.1.1. ESPAÇO DE ENSAIOS

O processo de calibração das câmaras tem de ser repetido sempre que existe alterações ao nível da lente ou posição de instalação das câmaras. Para evitar a repetição deste processo, as câmaras utilizadas estão fixadas junto à cobertura do LSA (Figura 25). Este foi o espaço *indoor* definido para a realização dos ensaios, utilizando três câmaras Basler de características iguais. Através da Figura 26 é possível verificar a disposição das câmaras ao longo do espaço de ensaios.



Figura 25: Fixação da câmara espaço de ensaios



Figura 26: Distribuição das três câmaras pelo espaço de ensaios

Na distribuição das câmaras foi tido em conta a fixação num plano superior para que fossem capazes de captar o campo por inteiro, permitindo em espaço amplo o FOV entre as combinações das câmaras, mesmo em situações em que o alvo é um *Quadrotor*.

5.1.2. DESCRIÇÃO DO *HARDWARE*

Sistema computacional

O sistema computacional é o responsável pela ordem de captura das câmaras e processamento dos dados recebidos das mesmas. Para este efeito é requerida uma unidade com rapidez de resposta e ainda que seja dotada de comunicação Gigabit Ethernet. Com base nestas especificações, o sistema computacional utilizado é um PC portátil dotado do sistema operativo Linux Ubuntu 12.04 com um processador Intel Core i7-4700HQ CPU 2.40GHz, 12GB de RAM e uma placa de rede Gigabit Ethernet.

A utilização da largura de banda *GigabitE* deve-se à resolução de imagens e ao número de FPS que são transmitidos através da rede. Para distribuição do protocolo Ethernet é utilizado um *Switch* de rede *GigabitE*, que permite a conexão das três câmaras e sistema computacional

Sistema de aquisição de imagem

As câmaras utilizadas são do fabricante Basler, modelo ACE 1300-30gc (Figura 27). As principais especificações do equipamento são apresentadas de seguida [39] :

- Protocolo de comunicação Gigabit Ethernet;
- Modo de disparo por *hardware*;
- Imagem a cores;
- *Frame rate* de 30 FPS;
- Resolução 1294 x 964 px.



Figura 27: Câmara Basler acA 1300-30gc¹

¹ <http://www.baslerweb.com/en/products/cameras/area-scan-cameras/ace/aca1300-30gc>, acedido em 4-10-2016

Para aquisição de imagens com esta câmara, é necessário recorrer a um *software* específico do fabricante, Pylon. A interface gráfica deste *software* permite visualizar as imagens adquiridas de cada câmara ligada à unidade computacional, permitindo também, a configuração de diversos parâmetros da câmara, entre os quais o tempo de exposição do sensor ótico. Esta funcionalidade reduz a quantidade de luz presente nas imagens, permitindo apenas destacar os objetos com grande intensidade de brilho, como são os casos dos marcadores Led de alta luminosidade.

Trigger por *hardware*

Com o objetivo de garantir o sincronismo de disparo entre câmaras, é utilizado um sistema embebido com uma interface gráfica, acessível a partir de um explorador de Internet. A partir desta ligação remota tornam-se possível configurar quais os canais de *trigger* a serem utilizados, as taxas de disparos de cada canal, bem como a largura dos impulsos dos sinais de disparo. O sistema embebido descrito apresenta ainda a possibilidade de integração de uma linha temporal absoluta (global), imposta por um recetor GPS que determina o sincronismo entre múltiplos sistemas [40].

Plataforma de marcadores

Para efeitos de ensaios, foi utilizada uma plataforma de acrílico que serve como base à instalação dos quatro Leds, uma bateria e um conversor de corrente contínua (DC/DC). Relativamente ao espectro de luz visível emitido pelos Leds, utilizaram-se três leds de cor vermelha e um led de cor azul, de acordo com a Figura 28.

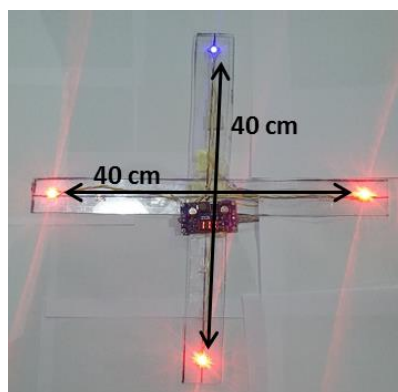


Figura 28: Plataforma de marcadores ativos

Durante a seleção das cores dos leds teve-se em conta o facto de estes serem as menos frequentes no espaço de ensaios. Os leds em questão foram organizados de forma a formar um losango. Esta solução foi adotada para facilitar a instalação em sistemas autónomos como o *quadrotor* que possui eixos suficientemente compridos e uniformes para formarem um losango com medidas conhecidas. A frente do alvo é indicada pelo único led de cor azul. Esta característica geométrica permite ainda, em caso de oclusão de um dos leds (independentemente de qual seja) formar um triângulo isósceles, que poderá ajudar na definição do posicionamento do veículo.

O sistema desenvolvido foi concebido para dar resposta a alterações de cores dos marcadores ou formas geométricas organizadas, sendo necessário proceder à alteração dos valores previamente parametrizados.

5.2. CALIBRAÇÃO DOS PARÂMETROS DAS CÂMARAS

Para a reconstrução 3D de uma imagem, é necessário ter em consideração os parâmetros intrínsecos e parâmetros extrínsecos. Estes parâmetros resultam do processo de calibração das câmaras que é uma das etapas mais importantes em qualquer sistema de visão computacional, sendo que quanto maior for a precisão da calibração, maior será a exatidão do sistema de visão.

5.2.1. CALIBRAÇÃO DE PARÂMETROS INTRÍNSECOS

No caso dos parâmetros intrínsecos permite a transformação de coordenadas desde o plano da câmara para o plano da imagem, enquanto os parâmetros extrínsecos permitem definir a localização e orientação do referencial da câmara em relação ao referencial do mundo, conforme descrito no Capítulo 3.

Este processo foi realizado recorrendo à *Toolbox Camera Calibration*¹ para Matlab², sendo uma ferramenta de calibração simples que permite obter os diversos parâmetros da câmara

¹ http://www.vision.caltech.edu/bouguetj/calib_doc/ acedido a 9/3/2016

² <http://www.mathworks.com/products/matlab/>, acedido a 9/3/2016

a partir de uma série de imagens com um *chessboard* em diversas posições como mostra o exemplo da Figura 29.

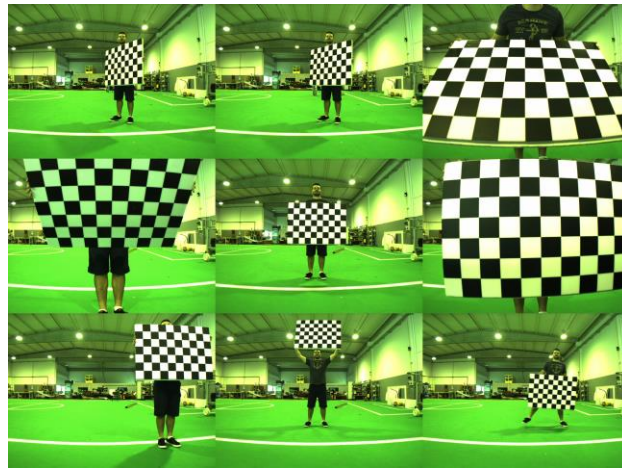


Figura 29: Sequência de algumas imagens do alvo para calibração

Após importação das imagens, terão que ser assinalados manualmente quatro cantos do xadrez em cada uma das imagens, definir as dimensões das quadrículas, número de quadrados em x e y e assim é possível detetar todos os cantos do *chessboard*. 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 todas as imagens com um erro elevado para que a calibração seja feita com a melhor qualidade, para isso analisa-se o gráfico produzido pela *Toolbox* de acordo com a Figura 30.

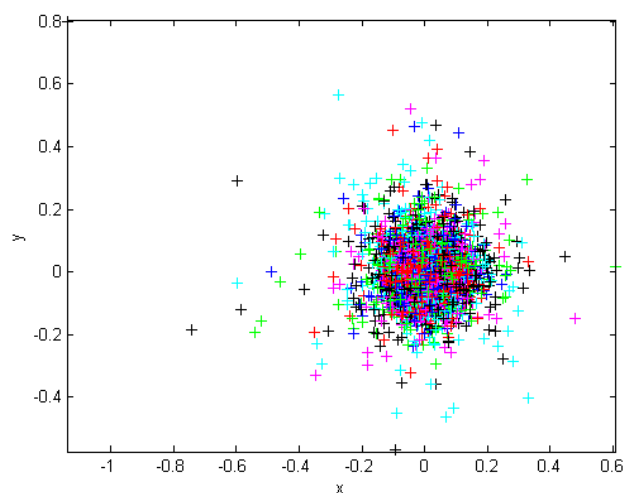


Figura 30: Erro de reprojeção em píxeis

No caso desta dissertação foi repetido este processo três vezes, uma para cada câmara, e foram descartadas as imagens com um erro de reprojeção de qualquer um dos pontos superior a 0.8 píxel. Terminada a análise de erro foram gerados os parâmetros intrínsecos.

5.2.2. CALIBRAÇÃO DE PARÂMETROS EXTRÍNSECOS

Os parâmetros extrínsecos permitem estabelecer uma relação entre o referencial do mundo e o referencial da câmara através de uma matriz de rotação (R) e uma de translação (t). Para se determinarem estas matrizes, para cada uma das câmaras, são utilizados vários pontos conhecidos no mundo e estabelecida uma relação com os respetivos pontos no plano da imagem em píxeis.

Através dos pontos de referência escolhidos, são medidas as distâncias 3D entre o ponto e o centro do referencial do mundo, neste caso o centro do campo. As distâncias foram determinadas com base nas medidas oficiais de um campo de futebol robótico [3]. Ao todo foram seleccionados 24 pontos distribuídos pela imagem de acordo com Figura 31.



Figura 31: Pontos do campo referenciados para determinação dos parâmetros extrínsecos

Após definição dos pontos 3D é estabelecida a correspondência em coordenadas normalizadas no referencial da imagem. Para isso, recorre-se a uma funcionalidade do *Matlab* que devolve a posição do píxel seleccionado em coordenadas na imagem x, y , como

mostra a Figura 32. Antes deste processo foi necessário retirar a distorção da imagem com base nos parâmetros intrínsecos da câmara.

Estabelecida a relação entre os pontos 3D no plano do mundo e 2D no plano da imagem, recorrendo à função “*compute_extrinsic*” da *Toolbox Camera Calibration* obteve-se a matriz de rotação e translação da câmara em relação ao centro do referencial do mundo. Por último, foi analisado o erro de reprojeção e discriminados os pontos com erro superior a 3 pixéis. Todos estes procedimentos foram repetidos para as *n* câmaras com recurso aos mesmos pontos 3D.



Figura 32: Extração de pontos no plano da imagem

5.3. AQUISIÇÃO E PROCESSAMENTO DE IMAGEM

Nesta secção são detalhadas as etapas de aquisição e processamento das imagens provenientes das três câmaras instaladas. Todo o processamento de imagem foi implementado recorrendo ao *software Matlab*.

5.3.1. AQUISIÇÃO DE IMAGEM

A aquisição das imagens das câmaras é realizada em simultâneo de acordo com o sinal trigger externo, garantindo assim o sincronismo entre imagens. Através da Figura 33, é possível visualizar as perspectivas do espaço de ensaios segundo as três câmaras instaladas.



Figura 33: Perspetivas das 3 câmaras instaladas no laboratório

Uma vez que o objetivo é seguir objetos com marcadores luminosos, há necessidade de conseguir controlar o tempo de exposição durante a captura das imagens e assim reduzir o número de falsos pontos de interesse, provocados por reflexos de iluminação. Estas configurações são feitas através da API da Basler.

5.3.2. PESQUISA DE MARCADORES ATIVOS

A busca por marcadores led inicia-se com a identificação “grosseira” de eventuais pontos de interesse na imagem. Ao conjunto de pontos de interesse extraídos e agrupados nas proximidades dá-se o nome de blob.

O ponto de interesse, resulta do pré-processamento da imagem (descrito no capítulo 3.1.2). Neste caso em concreto, corresponde a um píxel com uma cor que se destaca em relação à vizinhança. Este processo de extração de pontos de interesse caracteriza-se pela repetibilidade, uma vez que o estudo da variação da posição e orientação do objeto ao longo do tempo depende de uma sequência de imagens.

A arquitetura do processo de pesquisa de marcadores ativos pode ser dividida essencialmente em duas etapas para uma melhor compreensão. A primeira etapa integra as funções de pesquisa e formação de *blobs* azuis e posterior pesquisa e formação de *blobs* vermelhos. Por outro lado, a segunda etapa corresponde a uma alternativa ao insucesso da etapa um, invertendo a ordem de pesquisa dos *blobs* (primeiro *blobs* vermelhos e depois os *blobs* azuis).

O processo de detecção dos marcadores led tem início na aquisição e carregamento de imagens (independentemente da etapa). Este modelo recebe como *input*, a imagem e os parâmetros iniciais de pesquisa. Como *output* transmite o centro dos *blobs* sem distorção nas imagens. Para otimizar o processo foi desenvolvido um algoritmo que estima a área de posicionamento dos marcadores com base na aproximação do marcador à câmara, ou seja:

A pesquisa de *blobs* azuis inicia-se por toda a imagem porque, não existe nenhuma informação prévia do posicionamento do objeto. Encontrado o blob azul e tendo em conta a sua área (número de pixéis), é definida uma área circular de pesquisa, com centro no *blob* e raio determinado de acordo com a equação 5.1, em que R, representa o raio e N o número de pixéis.

A Figura 35 exemplifica a relação existente entre o número de pixéis que definem um *blob* e o raio de pesquisa definido. Os valores apresentados nesta equação são válidos para a resolução das imagens utilizadas (1294 x 964), sendo que foram obtidos através de uma regressão linear de acordo com a equação 5.1

$$R = \frac{40}{66}N - \frac{30}{40} \quad (5.1)$$

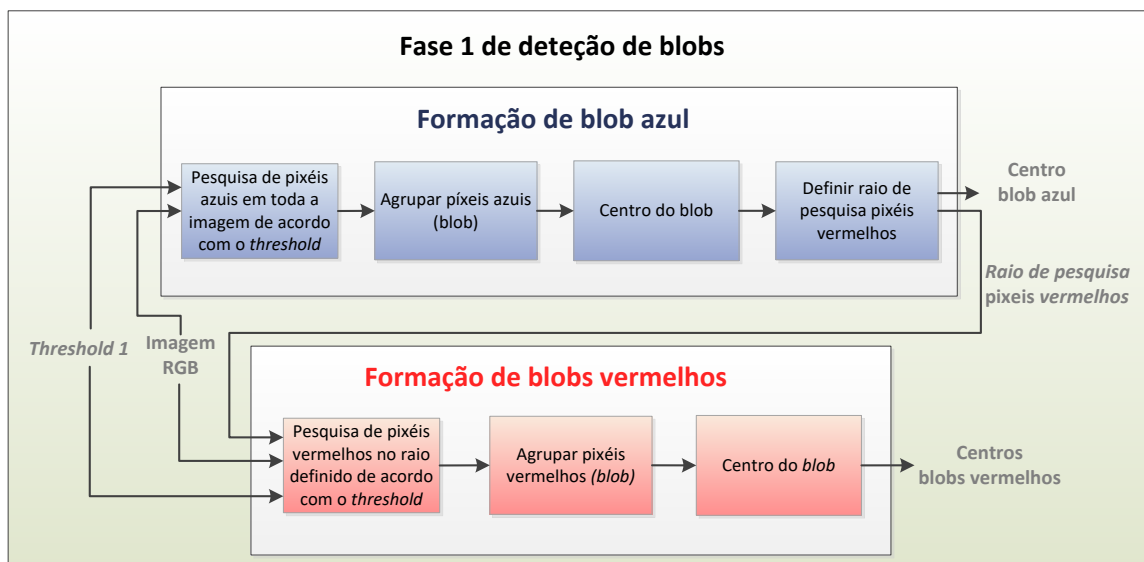


Figura 34: Sequência de processamento fase1 da detecção de *blobs*

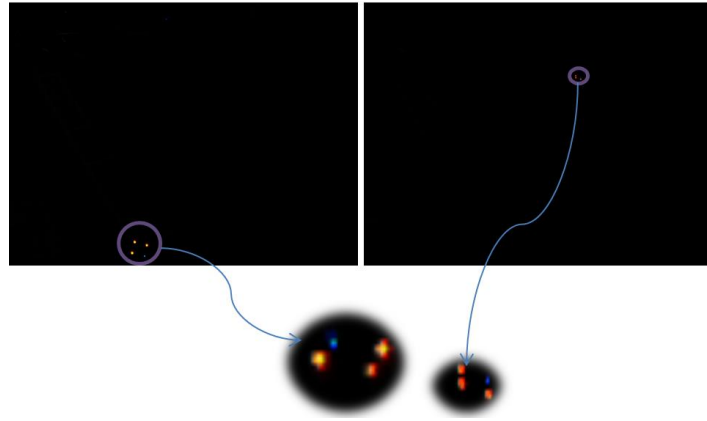


Figura 35: Relação entre número de pixéis e raio de pesquisa

Caso o número de *blobs* azuis detetados seja diferente de um, foi desenvolvida uma redundância para detetar primariamente os *blobs* vermelhos. Assim, foi criado um algoritmo, que é muito semelhante ao anteriormente abordado, contudo, inverte a ordem de procura de pontos de interesse.

De salientar que a única diferença entre as duas etapas reside no facto de se iniciar a pesquisa de pontos de interesse vermelhos por toda a imagem. Para restringir a área de pesquisa, é definido um raio e reajustado o valor de *threshold*, como é ilustrado na Figura 36.

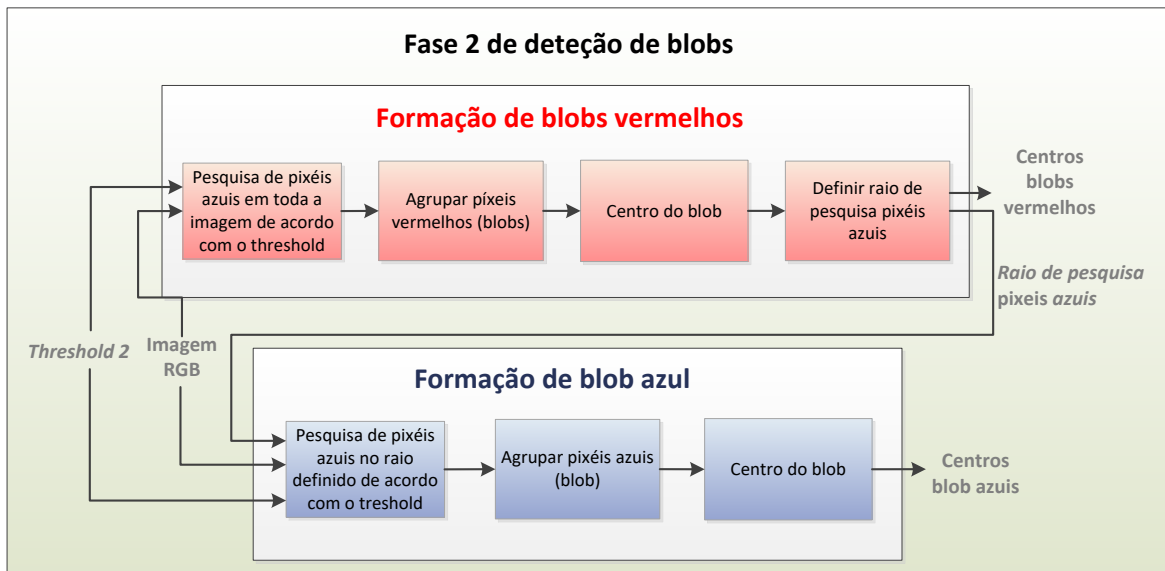


Figura 36: Sequência de processamento fase 2 de deteção de marcadores

Descritos os fluxos de processos que levam à definição da posição dos “eventuais” marcadores ativos na imagem, serão agora enumerados os processos/funções que contribuíram para esse fim.

5.3.2.1. SEGMENTAÇÃO POR COR

Para efetuar a extração dos pontos de interesse da imagem foi desenvolvido um método de detecção por cor. No processo de obtenção de imagem foi parametrizado a redução do tempo de exposição para obter uma imagem saturada. Nessa frame obtém-se uma imagem bastante escura, onde mesmo assim os pontos de interesse estão visíveis, tal como é possível observar na Figura 37,

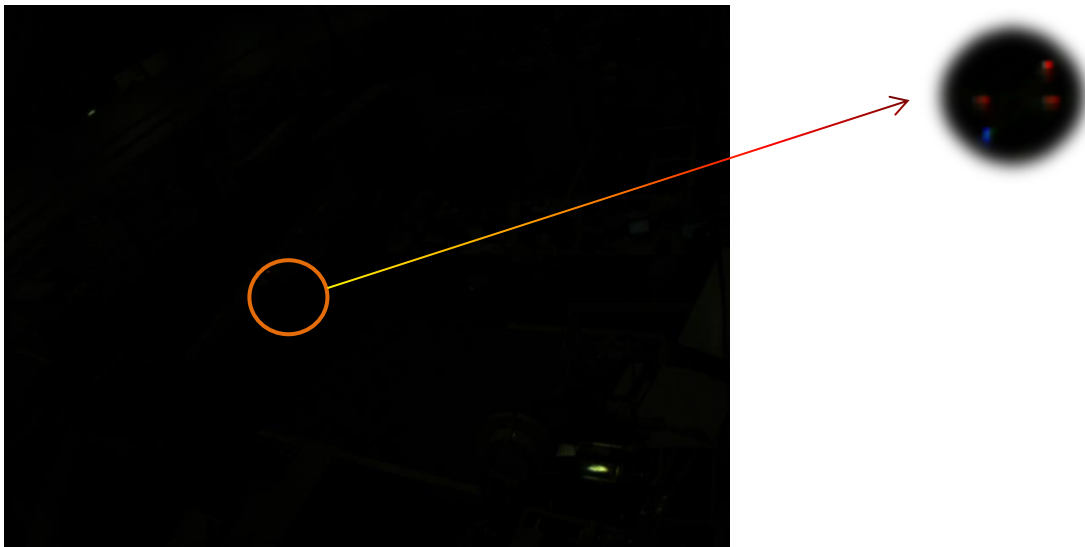


Figura 37: Imagem com baixo valor de exposição

Com base no modelo de cores RGB, para cada câmara são definidos os valores de *threshold* que determinam o limite máximo e mínimo de intensidade de cores (azul ou vermelho) dos pixels que poderão conter os marcadores Leds. Para cada ponto de interesse detetado são registadas as suas coordenadas.

5.3.2.2. FORMAÇÃO DE BLOBS

Após registo de todos os pontos de interesse distribuídos pela imagem, é executado uma função, que os aglomera considerando os pixels vizinhos, quer seja em linha horizontal, vertical ou diagonal como é ilustrado na Figura 38. Neste processo são excluídos *blobs* formados por apenas 2 pontos de interesse, por forma a não ser considerado o ruído.

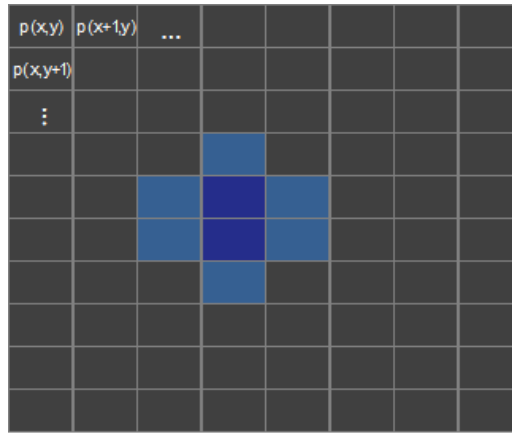


Figura 38: Formação de *um blob*

5.3.2.3. CENTRO DOS BLOBS

O processo de cálculo do centro dos *blobs* é importante para a precisão das coordenadas 3D obtidas. Para uma exata abordagem é considerada a posição e intensidade do píxel na imagem, ou seja: um led não é definido por pixéis de igual intensidade devido ao facto deste apresentar um maior brilho na zona central. No entanto, se ficarem todos os pixéis vermelhos (por exemplo) e for realizada uma média aritmética das suas posições independentemente da intensidade, o resultado será ambíguo.

Por outro lado se for realizada a média ponderada das coordenadas dos pixeis em x e y de acordo com as equações 5.2 e 5.3, é possível dar maior peso aos pixéis com maior intensidade e obter as coordenadas (p_x e p_y) mais aproximadas do centro do led na imagem, como mostra a Figura 39.

$$\bar{x}_p = \frac{\sum_{i=1}^x (p_i \times x_i)}{\sum_{i=1}^x (p_i)} \quad (5.2)$$

$$\bar{y}_p = \frac{\sum_{i=1}^y (p_i \times y_i)}{\sum_{i=1}^y (p_i)} \quad (5.3)$$

Considere-se \bar{x}_p e \bar{y}_p , como sendo as coordenadas médias ponderadas de um conjunto de coordenadas x_i e y_i respetivamente, cuja a intensidade relativa “peso” é p_i

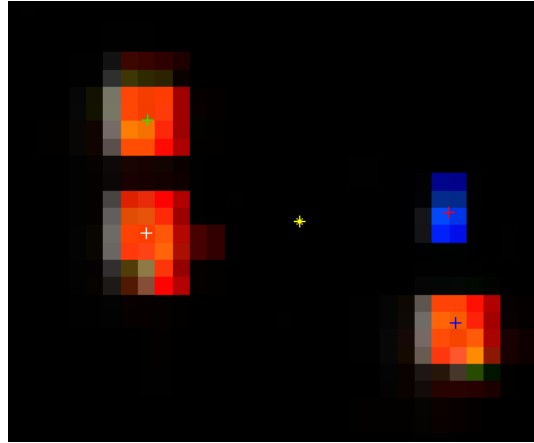


Figura 39: Representação do centro do *blob*

5.3.2.4. COMPENSAÇÃO DA DISTORÇÃO PROVOCADA PELA LENTE

O fenómeno de distorção de uma imagem é provocado pela lente da câmara (capítulo 3.3.1). Para se corrigir a deformação consequente, pode-se suprimir a distorção de toda a imagem ou apenas nos pontos de interesse. Dado que o processo de remoção de distorção de toda a imagem exige uma grande carga de processamento, neste sistema, apenas nos pontos de interesse é retirada a deformação proveniente da lente, otimizando assim o processamento para situações de tempo real. Para uma melhor perceção da diferença entre uma imagem com e sem distorção é apresentada na Figura 40.



Figura 40: Imagem com e sem distorção

Obtidos os centros dos marcadores na imagem, é necessário proceder à sua validação. Esta tarefa elimina *blobs* que poderão associar erros à posição do alvo, para isso, fazem-se uso de várias técnicas, como descritas no subcapítulo seguinte.

5.4. VALIDAÇÃO DE *BLOBS*

Após processamento de imagem que permitiu extrair eventuais falsos positivos que podem definir os marcadores, surge agora a necessidade de eliminação do ruído e falsos *blobs*. Esta estratégia evita que se considerem marcadores erradamente detetados e que se calcule a posição do alvo considerando esses mesmos pontos, levando ao erro na posição do alvo.

Durante o processo de formação de *blobs*, já se aplicaram vários filtros que permitiram a rejeição de falsos pontos de interesse. Na fase de processamento de imagem foi definido um *Threshold* que especifica a cor e intensidade dos pixéis que definem os marcadores. Ainda nesta fase são eliminados todos os *blobs* definidos por apenas dois pontos de interesse.

Para uma validação mais rigorosa definiram-se várias estratégias, que consistem numa sequência de testes, recorrendo ao número de marcadores detetados, à restrição epipolar e restrição espacial. Analisando a Figura 41 é possível perceber os sucessivos filtros e validações que foram aplicados para que se eliminasse os falsos pontos de interesse ao longo do processo.

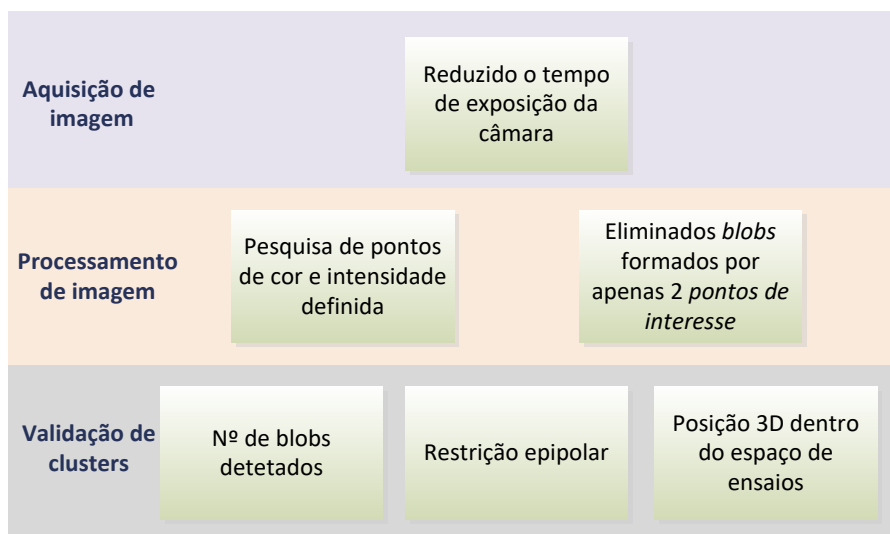


Figura 41: Técnicas/processos executados para filtragem de falsos pontos de interesse

5.4.1. NÚMERO DE *BLOBS* NA IMAGEM

Após processamento da imagem, obtém-se uma matriz de coordenadas, onde cada blob, recebe uma identificação numérica que os distingue. É com grande facilidade que se percebe que foram registados mais do que os quatro marcadores instalados no alvo. Este caso particular pode surgir devido, por exemplo, ao reflexo dos marcadores em superfícies espelhadas (exemplo da Figura 42), ou simplesmente porque foram detetados no meio falsos positivos que se assemelham aos marcadores em cor e intensidade.

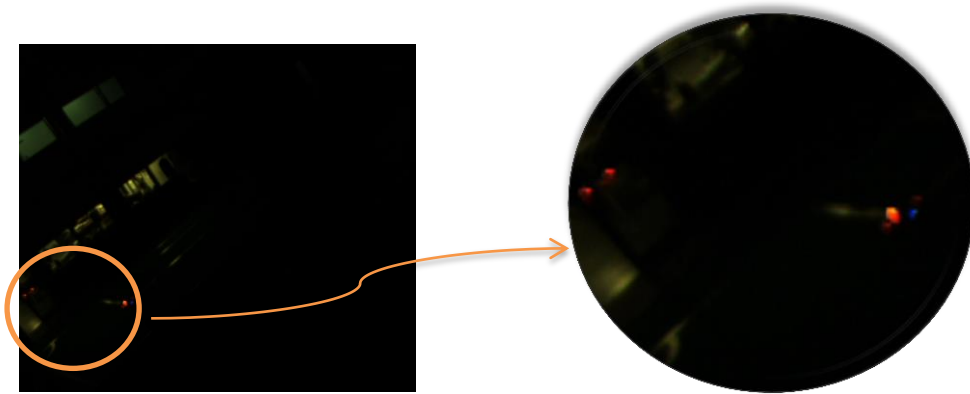


Figura 42: Reflexo dos marcadores ativos em superfície espelhada

Em *frames* que surgem manchas refletidas muito próximo do alvo, torna-se muito difícil distinguir os marcadores verdadeiros dos refletidos, nestes casos, terão que ser eliminados os falsos positivos através da restrição epipolar e espacial

Em situações que apenas são detetados três *blobs*, (como é o caso ilustrado na Figura 43) devido a sobreposição ou oclusão, resultado da perspectiva da câmara, não serão contabilizados os *blobs*.

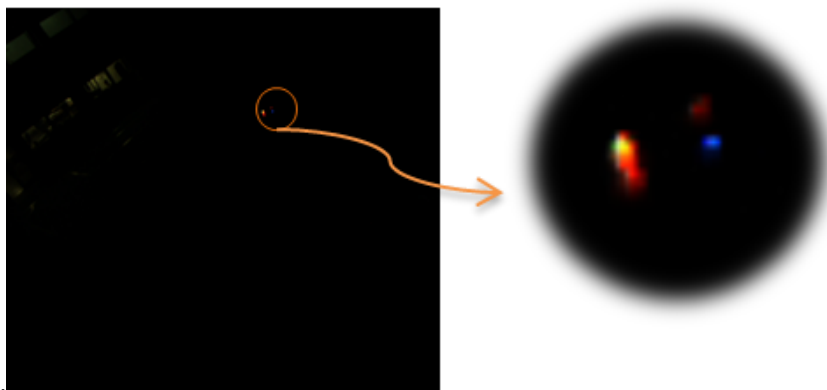


Figura 43: Sobreposição dos marcadores

5.4.2. RESTRIÇÃO EPIPOLAR

Considerando os conceitos abordados, a geometria epipolar apresenta-se como sendo mais um método para validação de falsos pontos de interesse.

Numa primeira fase são agrupados todos os *blobs* nas proximidades e a cada grupo é atribuída uma identificação numérica e calculado o seu ponto médio. A partir dessa informação são formuladas hipóteses acerca da correspondência de grupos de *blobs* detetados em cada uma das imagens. De acordo com a equação 3.17, se o resultado convergir para valores próximos de zero é muito provável que seja o mesmo grupo de *blobs* nas duas imagens.

A título de exemplo considere-se o alvo identificado em cada uma das imagens (A1, A2, A3) e ainda duas falsos positivos registados nas câmaras um e dois (B1, B2). De acordo com a restrição epipolar os resultados estão expostos na Tabela 2.

Tabela 2: Validação e correspondência de pontos entre imagens

<i>Cluster 1</i>	<i>Cluster 2</i>	Restrição epipolar	Corresponde?
A1	A2	0.0093	Sim
A1	B2	-2.7072	Não
A1	A3	0.0589	Sim
B1	A2	-1.4747	Não
B1	B2	-3.9271	Não
B1	A3	-0.9899	Não
A2	A3	0.0449	Sim
B2	A3	-5.8683	Não

Analisando os resultados da restrição epipolar é muito provável que os *blobs* A2, A3 e A3 (valores próximos de 0) sejam os mesmos captados de perspectivas diferentes. Assim sendo, consideram-se estes *blobs* e descartam-se todos os outros que não obtiveram correspondência.

5.4.3. RESTRIÇÃO ESPACIAL

A restrição inerente à posição 3D do blob, é aplicável na exclusão de todos os falsos pontos que estejam fora do espaço de ensaios, embora nesta fase, a existência de falsos pontos de interesse já sejam bastante reduzidos.

Relativamente ao aglomerado de *blobs* que foram associados a outros conjuntos de *blobs* através da restrição *Epiolar*, serão agora conjugados através da triangulação *stereo* (combinando pares de câmaras) de forma a obterem-se as correspondentes coordenadas 3D. Se a posição 3D do hipotético alvo estiver fora do volume de ação, serão eliminados os *blobs* nessas imagens. Esse mesmo espaço pode variar de acordo com as necessidades dos ensaios e carece de informação relativa à área de operação.

5.5. CÁLCULO POSIÇÃO 3D MARCADORES

Nesta fase de implementação foram utilizados duas técnicas para o cálculo da posição 3D dos marcadores detetados pelas três câmaras. Embora os dois métodos tenham a mesma finalidade, o processo de cálculo e resultados diferem sob alguns fatores.

5.5.1. TRIANGULAÇÃO STEREO

Triangulação *stereo*, é uma técnica que foi abordada no subcapítulo 3.4.3, e conjuga um par de câmaras, no entanto, este sistema de *ground truth* poderá integrar n câmaras, sendo que no âmbito dos ensaios foram utilizadas 3. Sempre que as 3 câmaras detetam os 4 marcadores (já validados), são conjugadas em pares, obtendo-se 3 coordenadas tridimensionais para cada marcador. Neste processo é imperativo que os marcadores detetados sejam os mesmos nas diferentes imagens, para tal foi desenvolvido um algoritmo para organizar e identificar a sequência de marcadores ativos como é exemplificado na Figura 44.

Algoritmo 1: Organização dos marcadores

Input: Matriz com centro dos marcadores 2D

Output: Matriz com centro dos marcadores ordenados 2D

Passo 1: Cálculo do centro do alvo

Passo 2: Registrar marcador azul como 1

Passo 3: Perceber a posição do marcador azul em relação ao centro do alvo

Passo 4: Atribuir a posição 3 ao marcador que se encontrar alinhado com o 1

Passo 5: Atribuir a posição 2 ao marcador que se encontra à direita do 1

Passo 6: Atribuir a posição 4 ao marcador que se encontra à esquerda do 1

De notar que, no caso dos quatro marcadores só serem detetados em duas das três câmaras, não é possível combinar mais do que um par stereo, nem calcular a geometria *multi view*. Neste caso em concreto apenas existirá um conjunto de coordenadas 3D para definir o alvo.

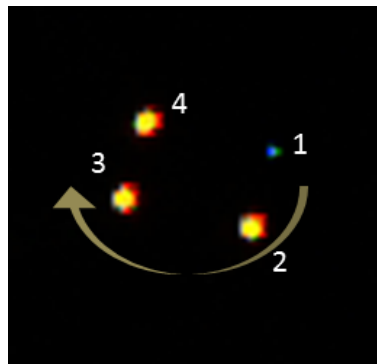


Figura 44: Organização dos marcadores segundo a sentido dos ponteiros do relógio

5.5.2. GEOMETRIA *MULTI VIEW*

A técnica de geometria *multi view* desenvolvida, permite calcular a posição 3D dos marcadores detetados nas diferentes câmaras, combinando em simultâneo as n câmaras que o sistema possui, como é ilustrado na Figura 45. À semelhança da triangulação *stereo*, este modelo consiste na interseção de várias retas no espaço com orientações $p_1, p_2 \dots p_n$. O ponto 3D é definido por P e pode ser obtido através do ponto de interseção mais próximo a todas as retas. Este processo ocorre devido ao facto de na realidade as retas nunca se chegarem a interseccionar no espaço, consequência da aproximação dos parâmetros das câmaras e da localização das imagens.

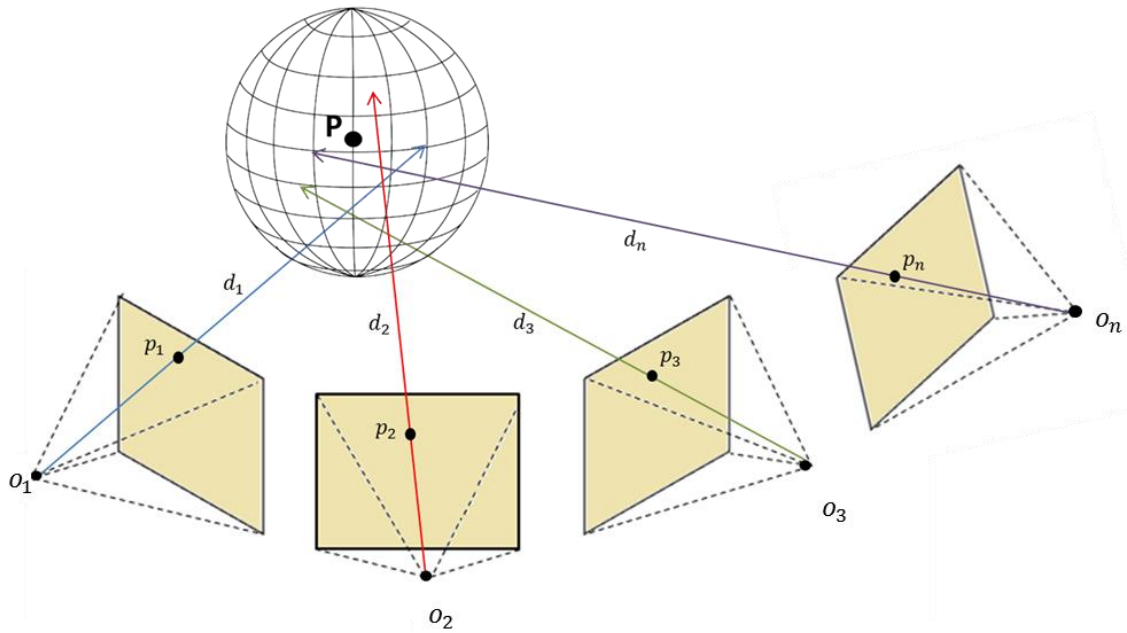


Figura 45: Geometria *multi view*

Para se encontrar o ponto de interseção mais próximo das n retas, utiliza-se a função ArgMin^1 , que retorna o ponto minimizado de uma função composta por multivariáveis. Através da execução desta função no *Matlab* com o nome *fminsearch* é possível definir o ponto de partida e o número de iterações máximo até chegar ao ponto em que o argumento mínimo da função é encontrado. Na Figura 46 é possível observar a convergência do erro nas sucessivas iterações do *fminsearch*.

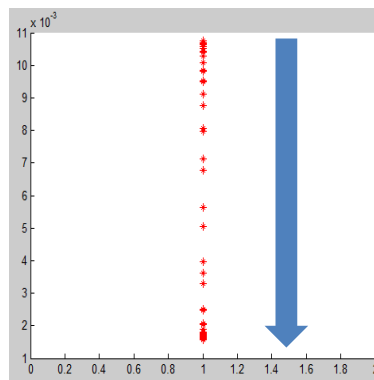


Figura 46: Convergência do erro na função *Fminsearch*

¹ <http://reference.wolfram.com/language/ref/ArgMin.html> acedido a 23-10-2016

Para o cálculo do ponto 3D através da triangulação com múltiplas perspectivas é então utilizada a equação 5.4

$$\arg \min [F, \mathbf{P}] \quad (5.4)$$

Em que \mathbf{P} representa o ponto 3D no referencial do espaço de ensaios e F a função que define a interseção das retas de acordo com a equação 5.5.

$$F = \sum_i (\|(\mathbf{P} - O_i)\|^2 - ((\mathbf{P} - O_i)^T \cdot \vec{d}_i)^2) \quad (5.5)$$

Através da equação 5.6 é possível atribuir um peso diferente a cada câmara, otimizando os resultados de acordo com o número de pixels e centro de massa que define um marcador. Esta referência pode por exemplo, ser condicionada pela distância do alvo à câmara.

$$F = \sum_i ((\|(\mathbf{P} - O_i)\|^2 - ((\mathbf{P} - O_i)^T \cdot d_i)^2) W^2) \quad (5.6)$$

O ponto no plano da câmara (O) é definido pela equação 5.7, onde R é a matriz de rotação da respectiva câmara i e \mathbf{t} o vetor de translação da mesma.

$$O = -R^{-1} \cdot \mathbf{t} \quad (5.7)$$

A direção do ponto projetado no plano da câmara é definido por d e é calculado através da equação 5.8. Em que A representa os parâmetros intrínsecos da câmara e p o ponto projetado no plano da imagem (marcador identificado na imagem da respectiva câmara i).

$$\vec{d} = (R^T \cdot A^{-1} \cdot p) / \|A^{-1} \cdot p\| \quad (5.8)$$

5.6. DEFINIÇÃO DAS COORDENADAS 3D COM MENOR ERRO

Dado que, nesta dissertação são utilizadas duas técnicas distintas para o cálculo das coordenadas 3D dos marcadores, foi implementado um algoritmo para efetuar a seleção de qual a coordenada mais precisa. Na Figura 47 é apresentado o diagrama desse mesmo processo.

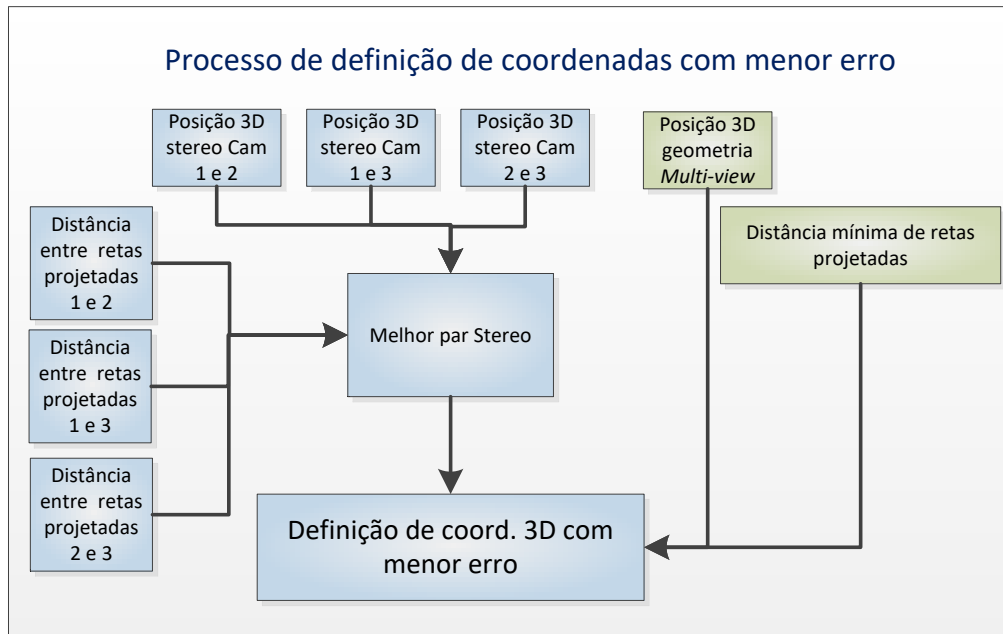


Figura 47: Definição de coordenadas com menor erro

Durante os ensaios realizados, foram utilizadas três câmaras. Na eventualidade das três câmaras detetarem os quatro marcadores, através do processo de triangulação *stereo*, resultará 3 conjuntos de coordenadas 3D. Para cada marcador será calculada a sua reprojeção (de acordo com a equação 3.7 do subcapítulo 3.2.1). Analisando o desvio em relação ao ponto inicialmente registado, é possível perceber o erro associado ao processo. Outro fator que também pesa na precisão do cálculo da triangulação *stereo* é a aproximação média da distância entre os pontos projetados em ambas as câmaras (capítulo 3.4.3, representado por c). Dado que a dimensão deste escalar influência diretamente a precisão do cálculo, quando conjugando com o erro de reprojeção, é possível determinar o melhor par *stereo*.

Para o processo de cálculo de coordenadas 3D através da triangulação com n perspetivas, é analisado o erro de reprojeção (calculado de forma muito semelhante ao acima descrito) e ainda é tido em conta o erro mínimo associado às sucessivas iterações, calculado de acordo com a equação 5.9. O erro pode variar de acordo com a distância do objeto à câmara, ou precisão no cálculo do centro do marcador.

$$Er = \sqrt{\|(\mathbf{P} - O_i)\|^2 - ((\mathbf{P} - O_i)^T \cdot \mathbf{d}_i)^2} \quad (5.9)$$

6. RESULTADOS

Neste capítulo são apresentados os múltiplos ensaios realizados para validação do trabalho desenvolvido nos capítulos prévios. Para esse efeito foi posicionado o alvo com os marcadores em vários pontos do espaço de ensaio 3D com diferentes atitudes. Para uma detalhada análise, este capítulo foi dividido essencialmente em três etapas: Identificação dos marcadores, cálculo da posição 3D dos marcadores, comparação de resultados.

Na primeira etapa são demonstrados os resultados da identificação e validação dos marcadores detetados em cada uma das imagens, provenientes das 3 câmaras instaladas no espaço de ensaios.

Para o cálculo da posição 3D são utilizadas duas técnicas. Nesta etapa são expostos e comparados os resultados entre ambas as técnicas e aferida a discrepância sobre várias condicionantes, tais como o número de perspectivas e posicionamento do alvo.

A última etapa consiste na comparação dos resultados do sistema desenvolvido com informação de referência. Nestes ensaios são comparadas as posições tridimensionais dos

marcadores, com os pontos conhecidos do espaço (marcação do campo de futebol robótico) e posição 3D de referência recolhida a partir do scanner Faro¹ Focus 3D. Relativamente à atitude, são comparados os resultados com os valores extraídos do Pixhawk².

6.1. EXTRAÇÃO DE MARCADORES

Com o objetivo de analisar a capacidade de extração de marcadores de uma *frame* sob condicionantes como, a distância do alvo e variações de luminosidade, é apresentada a seguinte tabela com os resultados obtidos para um ensaio onde foi posicionado o alvo em diferentes posições do espaço 3D.

Tabela 3: Número de marcadores detetados em cada imagem em função da posição do alvo

Posição	Câmara 1	Camara 2	Camara 3
1	4	4	3
2	4	4	4
3	4	4	4
4	3	3	3
5	4	4	2
6	4	4	0
7	2	4	2
8	4	4	4
9	4	4	4
10	4	4	4

Analisando os resultados é possível perceber que em dez posições diferentes do alvo, mais precisamente em 30 imagens, apenas em oito imagens não se conseguiram detetar os 4 marcadores do alvo, de acordo com o gráfico da Figura 48. Dentro do grupo de oito imagens em que não foram detetados os quatro marcadores é importante perceber o que originou estes resultados.

¹ <http://www.faro.com/products/3d-surveying/laser-scanner-faro-focus-3d/overview> acedido a 25-10-2016

² <https://pixhawk.org/> acedido a 25-10-2016

Em casos particulares, como o da Figura 49 a, é notória a sobreposição dos marcadores, como tal não se consegue identificar o limite entre ambos, sendo essa a razão pelo qual o sistema deteta como sendo um único marcador.

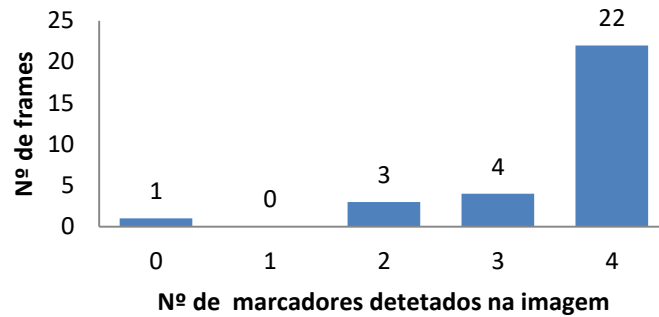


Figura 48: Número de marcadores detetados em função do número de imagens

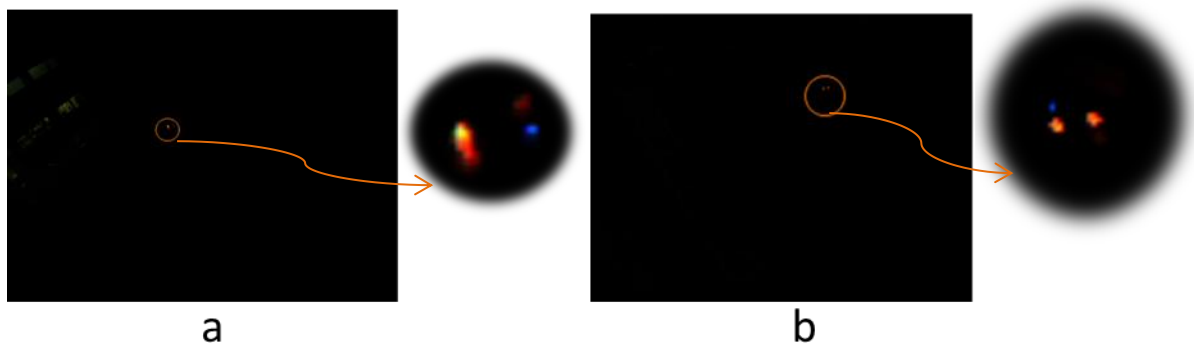


Figura 49: a; Sobreposição de marcadores b; marcador não perceptível

Na Figura 49 b é visível o baixo valor de intensidade da cor vermelha captada pela câmara, resultado da orientação do alvo no espaço 3D. Esta particularidade, leva à não deteção do marcador, mesmo contando com um valor de *threshold* adaptativo.

Verificaram-se ainda, situações de oclusão devido a obstáculos no espaço de ensaios, como demonstrado na Figura 50.

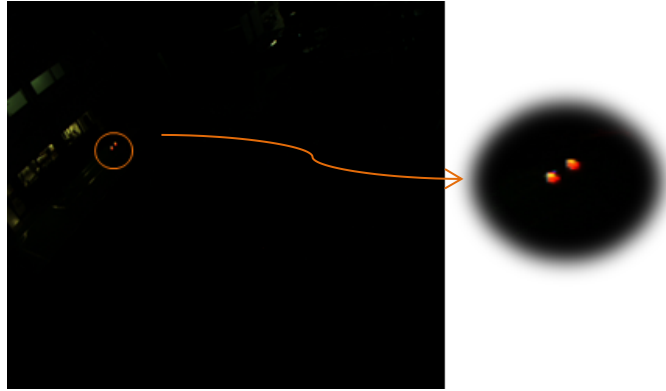


Figura 50: Exemplo de oclusão dos marcadores

6.2. POSIÇÃO 3D DOS MARCADORES

Para comparação de resultados, entre a posição 3D dos marcadores calculados a partir da triangulação *stereo* e geometria *multi view*, recorreu-se ao ensaio anteriormente descrito, considerando apenas as posições em que foram detetados quatro marcadores em pelo menos duas câmaras. Assim sendo foi necessário descartar a posição 4 e 7 porque não é possível calcular a posição 3D do alvo devido a oclusão. A posição do alvo ao longo do espaço de ensaios está visível na Figura 51, onde foram reprojetoados os pontos obtidos a partir das combinações de pares de câmaras *stereo*.



Figura 51 Reprojção dos pontos 3D para triangulação *stereo*

Recorrendo a todas as combinações de pares de câmaras, para cada marcador, obtém-se o erro de reprojeção de acordo com a Figura 52 e a distância entre a interseção das retas projetadas no par *stereo* (erro *stereo*) ilustrado na Figura 53 e com um erro médio e desvio padrão apresentado na Tabela 4.

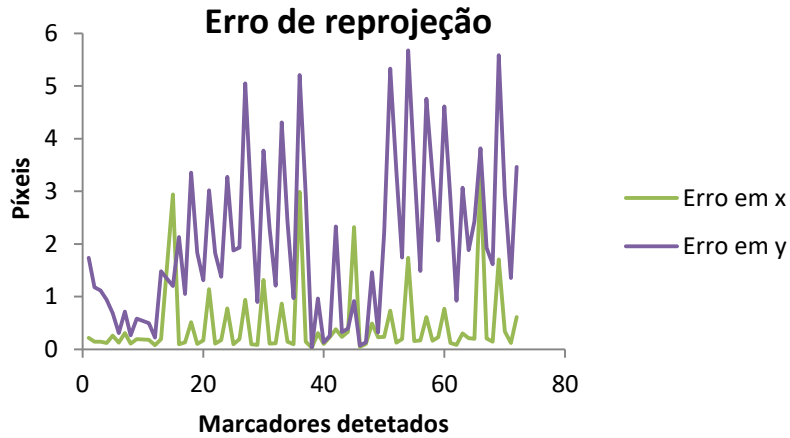


Figura 52: Erro de reprojeção em x e y para cada marcador com o método triangulação *stereo*

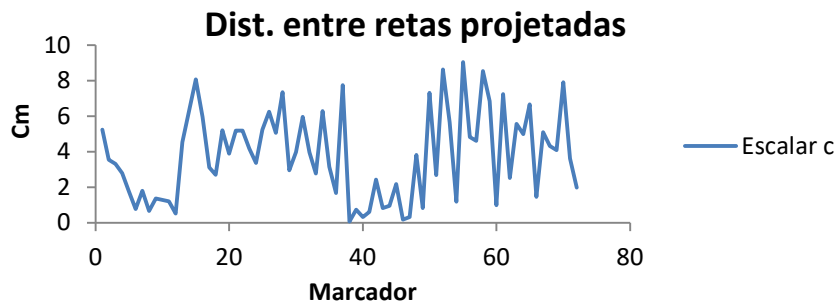


Figura 53: Distância entre retas projetadas (par *stereo*)

Tabela 4: Média e desvio padrão do erro de reprojeção e do valor do c para todos os pares *stereo*

	Média	Desvio Padrão
Reprojeção em x (píxeis)	0.462	0.697
Reprojeção em y (píxeis)	2.036	1.494
c (cm)	3.804	2.425

Considerando apenas as posições em que foram detetados os quatro marcadores pelas três câmaras, é possível analisar o desfasamento entre posições 3D, obtidas a partir de pares *stereo* diferentes, para um mesmo marcador de acordo com a Figura 54.

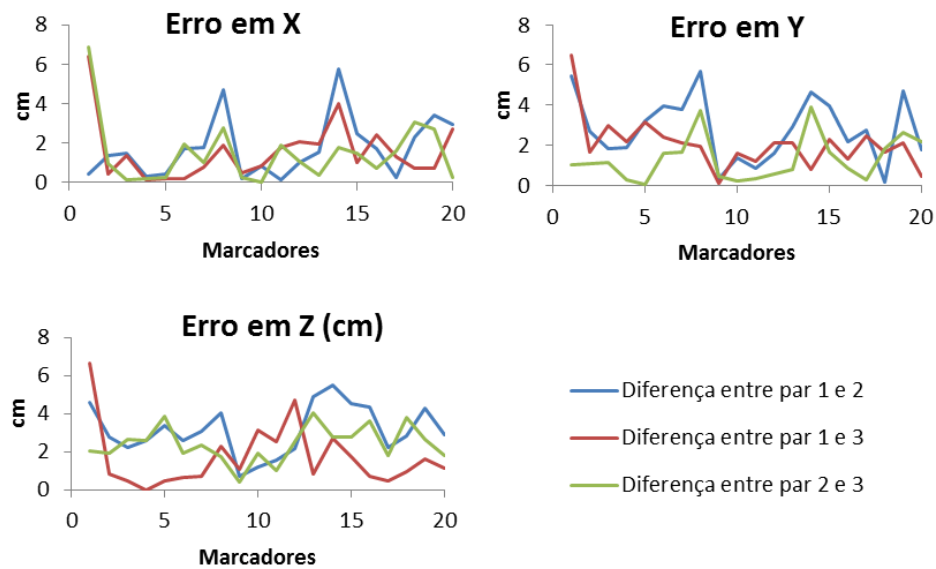


Figura 54: Desfasamento das coordenadas 3D (cm) para cada par de câmaras

O erro médio e o desvio padrão das coordenadas 3D extraídas a partir dos diferentes pares de câmara *stereo* estão disponíveis na Tabela 5.

Tabela 5: Média e desvio do desfasamento entre coordenadas 3D

	Média	Desvio Padrão
Em X (cm)	1.591	1.514
Em Y (cm)	2.045	1.467
Em Z (cm)	2.408	1.430

Executando o algoritmo que define o melhor par *Stereo* com base no erro *stereo* (escalar c) e erro de reprojeção, obtêm-se os resultados representados na Figura 55.

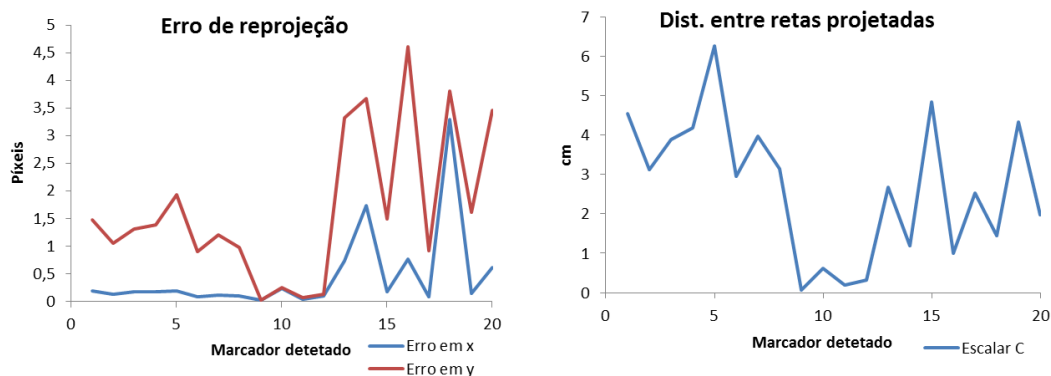


Figura 55: Representação do erro de reprojeção em x e y e erro *stereo* para o melhor par *stereo*

Tabela 6: Média e desvio padrão para erro de reprojeção e escalar C melhor par *stereo*

	Média	Desvio Padrão
Reprojeção em x (píxeis)	0.454	0.775
Reprojeção em y (píxeis)	1.681	1.366
C (cm)	2.659	1.767

Analisando os resultados da tabela Tabela 4 e Tabela 6 é notório um decréscimo dos valores médios de reprojeção em x, y e do erro *stereo*. Este facto valida o correto funcionamento do algoritmo responsável por definir o melhor par *stereo*. De salientar ainda, que os resultados apresentados na Figura 54 indiciam um desfasamento considerável, entre o mesmo ponto para pares de câmaras diferentes. Este fenómeno pode ser justificado pela distribuição do alvo ao longo do espaço de ensaios e comprimento da *baseline* do respetivo par *stereo*. Analisando os resultados obtidos da Figura 56 verifica-se uma tendência para os melhores resultados estarem associados ao par de câmaras um, que maioritariamente corresponde ao par mais perto do alvo.

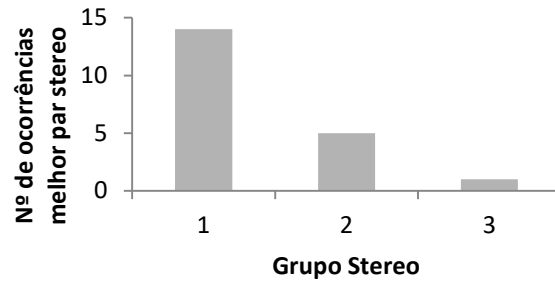


Figura 56: Número de ocorrências melhor par *stereo* para cada grupo

Aplicando a técnica de geometria *multi view* para as mesmas posições do alvo anteriormente analisadas, obtêm-se os marcadores reprojetaados e distribuídos ao longo do espaço de ensaios de acordo com Figura 57. Os erros de reprojeção e distância entre retas projetadas estão ilustrados na Figura 58 e Figura 59.



Figura 57 Reprojeção dos pontos 3D para geometria *multi view*

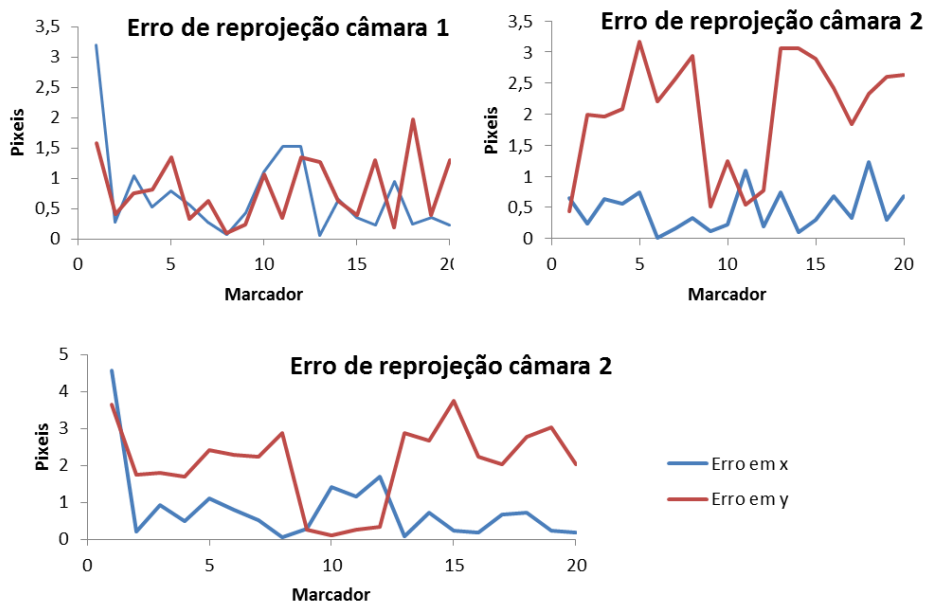


Figura 58: Erro de reprojeção (triangulação *stereo*) para cada uma das câmaras

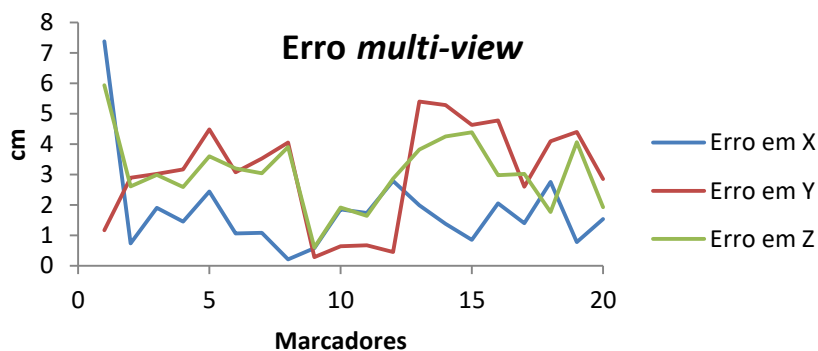


Figura 59: Erro *multi view* em cm

A média e desvio padrão do erro de reprojeção e erro *multi view* apresenta-se na Tabela 7.

Tabela 7: Erro médio e desvio padrão do erro de reprojeção e erro *multi view*

	Média	Desvio Padrão
Reprojeção em x (pixéis)	0,720	1,704
Reprojeção em y (pixéis)	0,924	1,085
Norma do vetor erro multi-view (cm)	4,960	1,883

Recorrendo às coordenadas tridimensionais extraídas a partir do melhor par de câmaras *stereo* e geometria *multi view* foi executado o algoritmo que seleciona o melhor resultado com base no erro de reprojeção e distância de interseção entre retas projetados nos dois métodos. As posições do alvo onde foram comparados os resultados das duas técnicas estão representados na Figura 60.



Figura 60 Reprojeção dos marcadores com base no cálculo 3D triangulação *stereo* e geometria *multi-view*

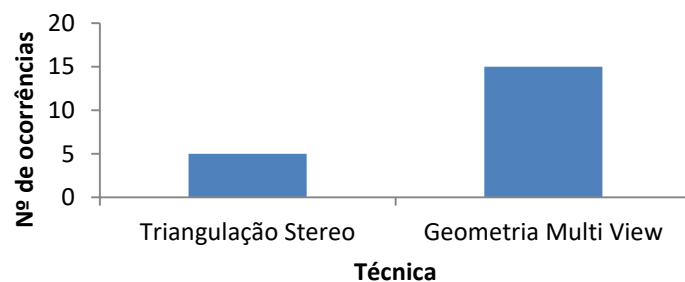


Figura 61: Ocorrências de técnica com resultados 3D mais precisos

De acordo com o gráfico da Figura 61, é visível que para os vinte marcadores detetados pelas 3 câmaras, em quinze situações a geometria *Multi-view* apresentou um resultado mais preciso. Das quatro em cinco situações, em que não apresentou os resultados mais precisos, o alvo estava posicionado junto ao par de câmaras um e dois, enquanto que a câmara três estava distanciada do alvo aproximadamente vinte e cinco metros. Esta distância poderá ter induzido erros na deteção dos marcadores no plano da imagem e posteriormente no cálculo da posição 3D do marcador, por essa razão, o par *stereo* apresentou resultados mais precisos.

6.3. VALIDAÇÃO DA POSIÇÃO TRIDIMENSIONAL DOS MARCADORES

Para uma análise profunda, recorreu-se a um scanner laser (Faro Focus 3D) que para cada posição do alvo, realizou um varrimento para se obter as coordenadas 3D dos vários marcadores como é ilustrado na Figura 62.



Figura 62: Nuvem de pontos do Faro processada através do *software* Scene LT¹

Uma outra referência utilizada como comparação são as marcações do campo de futebol robótico com dimensões conhecidas. Na Tabela 8 são expostos os resultados referentes às

¹ <http://www.faro.com/faro-3d-app-center/stand-alone-apps/scene-lt> acedido a 24/9/2016

posições 3D dos marcadores no referencial do campo, para que se consiga analisar a diferença entre os sistemas.

Tabela 8: Comparação das coordenadas 3D em metros no referencial do campo

Marcador	Coordenadas 3D Calculadas			Resultados Faro			Medições do campo		
	X	Y	Z	X	Y	Z	X	Y	Z
1	0,198	-0,021	0,024	0,217	-0,050	0,009	0,200	0,000	0,000
2	0,018	-0,230	0,076	0,050	-0,220	-0,018	0,000	-0,200	0,000
3	-0,221	0,000	0,062	-0,211	-0,020	-0,010	-0,200	0,000	0,000
4	-0,025	0,217	0,074	0,050	0,220	-0,011	0,000	0,200	0,000
5	-0,042	-1,810	0,066	0,044	-1,789	-0,007	0,000	-1,800	0,000
6	0,160	-2,027	0,074	0,227	-1,970	0,002	0,200	-2,000	0,000
7	-0,042	-2,220	0,072	0,050	-2,280	-0,001	0,000	-2,200	0,000
8	-0,242	-2,027	0,075	-0,221	-2,060	0,000	-0,200	-2,000	0,000
9	-9,003	-6,172	-0,023	-8,905	-6,181	0,004	-9,000	-6,200	0,000
10	-9,221	-6,002	0,033	-9,249	-6,050	0,012	-9,200	-6,000	0,000
11	-9,010	-5,779	-0,008	-8,965	-5,829	0,022	-9,000	-5,800	0,000
12	-8,819	-5,998	0,026	-8,821	-6,029	0,003	-8,800	-6,000	0,000
13	1,934	-0,250	0,075	2,031	-0,213	-0,003	2,000	-0,200	0,000
14	1,762	-0,050	0,079	1,811	0,043	0,008	1,800	0,000	0,000
15	1,973	0,193	0,051	2,014	0,202	-0,008	2,000	0,200	0,000
16	2,163	-0,046	0,052	2,208	-0,132	-0,007	2,200	0,000	0,000
17	0,195	0,008	0,939	0,207	0,013	0,917	0,200	0,000	0,905
18	-0,045	-0,195	0,949	Oclusão	Oclusão	Oclusão	0,000	-0,200	0,905
19	-0,252	0,019	0,934	-0,211	-0,059	0,911	-0,200	0,000	0,905
20	-0,042	0,209	0,938	-0,005	0,297	0,920	0,000	0,200	0,905

A discrepância entre o sistema desenvolvido e as duas referências são expostas na Figura 63 e Figura 64

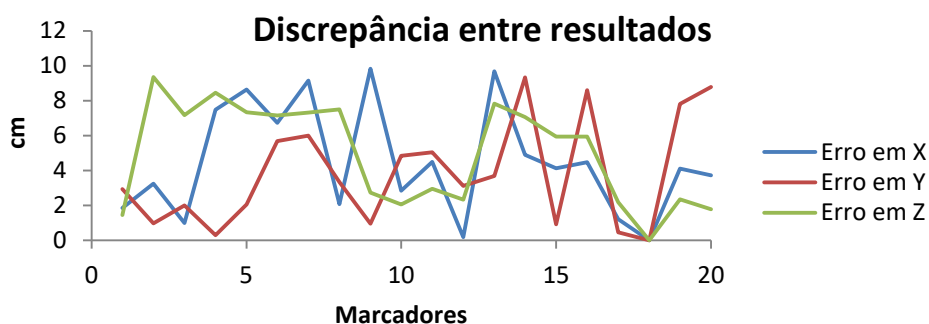


Figura 63: Discrepância entre resultados obtidos e medições do Faro

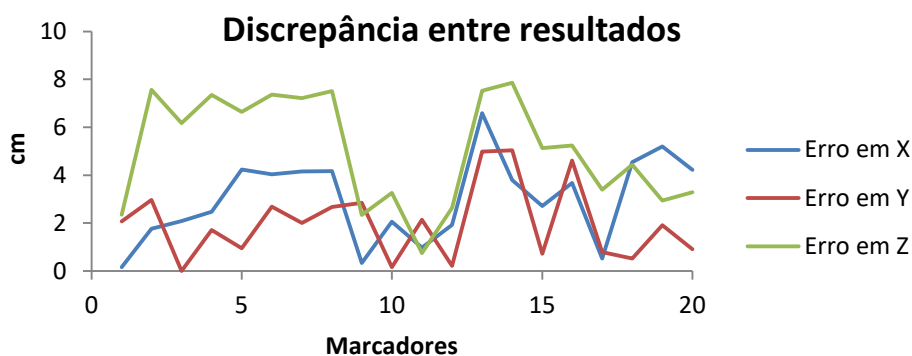


Figura 64: Discrepância entre resultados obtidos e medições reais

Comparando os resultados obtidos com as referências (medições do Faro e marcações do campo), calcula-se o erro médio e o desvio padrão de acordo com a Tabela 9.

Tabela 9: Discrepância média e desvio padrão entre resultados obtidos e referências

		X (cm)	Y (cm)	Z (cm)
Faro	Média	4,721	4,041	5,205
	Desvio Padrão	3,047	2,983	2,723
Medições reais	Média	2,981	1,995	5,047
	Desvio Padrão	1,749	1,551	2,274

Analisando a Tabela 9, verifica-se uma menor discrepância entre os resultados obtidos e as medições reais, quando comparado com as coordenadas retiradas do Faro. Este facto pode ser justificado por erros de posicionamento do Faro, que induziu erros na matriz de rotação e translação, utilizada na conversão do sistema de coordenadas do Faro para o espaço de ensaios. Por esta razão não deve ser colocado em causa a precisão do sensor mas sim a sua instalação no espaço de ensaios.

Deve ser tido em conta que, o erro médio e desvio padrão dos resultados calculados, foram obtidos com base em múltiplas posições do alvo, distanciados a pelo menos doze metros de distância de uma das câmaras. Nestas condições, é espectável que o erro médio aumente de forma progressiva com o aumento da distância.

6.4. VALIDAÇÃO DA ATITUDE DO ALVO

Os próximos resultados são relativos à atitude do alvo para as oito posições em que foram detetados quatro marcadores em pelo menos duas das câmaras.

Através das coordenadas tridimensionais dos marcadores e utilização regras trigonométricas aplicadas a pares de marcadores foi possível calcular a atitude do alvo.

Como referência, foi instalado um Pixhawk¹ sobre o alvo, tal como é ilustrado na Figura 65. A discrepância entre a atitude calculada e a obtida a partir do Pixhawk, apresenta-se na Figura 66

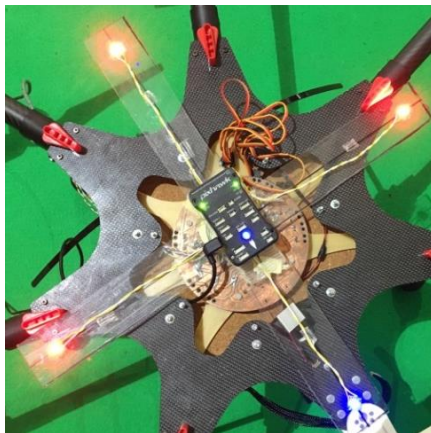


Figura 65: Instalação do Pixhawk sobre o alvo

¹ <https://pixhawk.org/> acedido a 25/10/2016

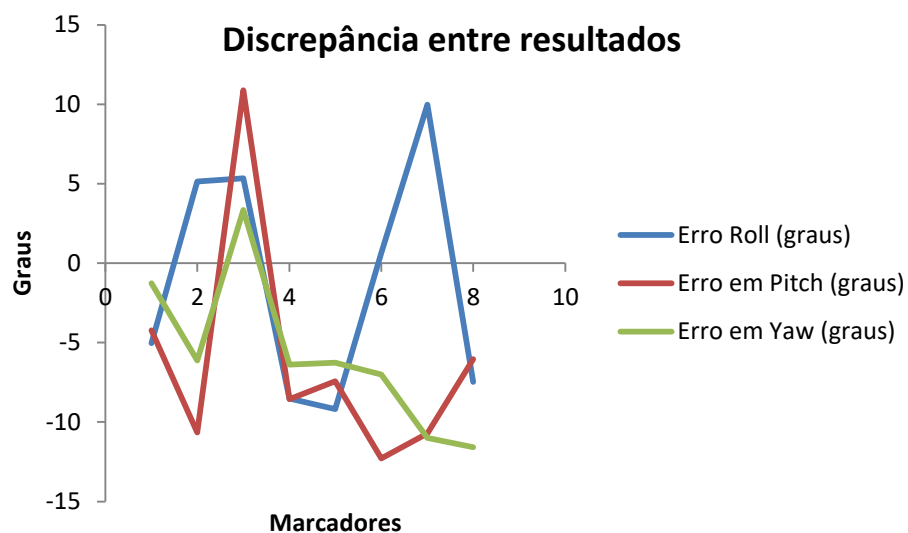


Figura 66: Discrepância entre atitude calculada e obtida do Pixhawk

O Desfasamento médio e desvio padrão entre a atitude calculada e a obtida a partir do sistema comercial, são apresentados na Tabela 10

Tabela 10: Erro médio e desvio padrão

	Roll (graus)	Pitch (graus)	Yaw (graus)
Média	-1.141°	-6.132°	-5.784°
Desvio Padrão	7.395°	7.372°	4.878°

Esta página foi intencionalmente deixada em branco.

7. CONCLUSÕES E TRABALHO FUTURO

Esta dissertação abordou o desenvolvimento de um sistema de visão multi-câmara externo, para determinar a localização e atitude de múltiplos objetos. Tendo como foco, o fornecimento de informação tridimensional de um determinado alvo, para correção de algoritmos de navegação, ou validação da resposta do sistema autónomo, foram traçados os requisitos do sistema tendo em conta os vários robôs e cenários de aplicações.

Previamente, efetuou-se um estudo das várias tecnologias e técnicas do estado da arte. Esta análise foi útil para perceber as tecnologias existentes e as contribuições de vários autores para o desenvolvimento de sistemas *ground truth*.

Foi necessário recorrer ao estudo de fundamentos teóricos, para que se construíssem bases relativas à visão computacional, modelo e calibração da câmara, correspondência de pontos e extração de informação tridimensional a partir de imagens.

Partindo da informação recolhida, foi proposta uma arquitetura global para um sistema de posicionamento externo que integra múltiplas câmaras, com ou sem processamento distribuído recorrendo a marcadores ativos.

Como contribuição para esse sistema global, foi desenvolvido em Matlab o *software* composto por vários módulos, dentro dos quais a extração e validação de *blobs*. Um segundo módulo foi aplicado na exploração das técnicas de triangulação *stereo* e geometria *multi view* para o cálculo da posição tridimensional dos marcadores. Posteriormente, os resultados são analisados com o intuito de se verificar quais os mais precisos, sob condicionantes do erro da geometria de cálculo e reprojeção.

Para validar o *software* desenvolvido, foi criado um *setup* experimental, composto por três câmaras distribuídas pelo espaço de ensaios e quatro marcadores posicionados em diferentes pontos do espaço. Os resultados revelam que em trinta *frames*, apenas em oito foi necessário descartar as imagens devido a oclusão ou sobreposição dos marcadores. Das vinte e duas imagens foram calculados os pontos 3D dos quatro marcadores detetados, através da triangulação *stereo*. Verificou-se uma disparidade entre os valores calculados para os mesmos marcadores, no entanto, conclui-se que os melhores resultados estão associados à menor distância entre o alvo e o par de câmaras.

Para as posições em que foram detetados os quatro marcadores nas três imagens, verifica-se uma clara tendência para os melhores resultados da posição tridimensional dos alvos estar ligado à geometria *multi view*. Nos casos em que, tal não se verificou, o alvo estava numa posição muito distante (aproximadamente 20 metros) de pelo menos uma das câmaras. Este facto indicia que existe incerteza no centro do marcador identificado na imagem e por consequência a posição 3D também.

Como era espectável, os resultados comprovam que quanto menor for a distância da câmara ao alvo, maior é o número de pixéis que definem o marcador. Por consequência mais preciso é o cálculo do seu centro. Este facto justifica a teoria de que, quanto mais câmaras o sistema possuir numa mesma área, maior precisão terão as coordenadas tridimensionais.

Analisando os resultados, verifica-se que em todos os casos, há um erro maior no eixo do *y*. Esta tendência pode ser justificada com uma ligeira alteração da posição das câmaras que provocou incerteza nos parâmetros extrínsecos aquando da realização dos ensaios.

Em suma, os objetivos propostos para esta dissertação foram completados com sucesso.

No que diz respeito ao trabalho futuro existem várias abordagens que podem melhorar a performance do sistema.

A introdução de mais câmaras no sistema seria uma mais-valia, para se constatar a facilidade com que o sistema integra mais de três câmaras e a influência que estas têm na precisão do sistema e área de cobertura.

Devido à obstrução parcial do espaço de ensaios, não foi possível utilizar a sua extensão na sua totalidade. Seria importante verificar o impacto nos resultados que a utilização do espaço por completo acarreta.

Para testar a qualidade e versatilidade do sistema, seria ideal efetuar simulações em diferentes ambientes e cenários.

Como forma de aumentar a exatidão dos resultados finais, seria interessante aplicar um peso à contribuição de cada uma das câmaras para o geometria *multi view*, (atribuindo menor peso a contribuições ambíguas) como é descrito na função 5.6. A contribuição dessa mesma câmara poderia ser calculado com base no número de píxeis que define o marcador ou distância do alvo à câmara. A implementação de uma técnica que descarta as contribuições *outlier*, baseadas por exemplo no método de RANSAC, seria uma vantagem na precisão dos resultados tridimensionais obtidos, porque evitaria a contribuição de falsos pontos detetados/refletidos.

Para contornar situações de reflexos em ambiente *outdoor* e maximizar a energia consumida pelos marcadores, a instalação de um sistema sincronizado de disparo dos marcadores apresenta-se como uma vantagem.

A calibração das câmaras é um fator que influência a qualidade dos resultados, seria uma maior valia a implementação de um sistema de auto calibração como foi descrito por [40]

De forma a superar situações de oclusão parcial, seria pertinente desenvolver uma técnica de cálculo para a posição e atitude do alvo com base em apenas três marcadores (situações em que um dos quatro marcadores está ocluído). Esta técnica foi já utilizada em [1].

Com a implementação destas melhorias, será consequentemente obtido um sistema mais completo e com melhor performance.

Esta página foi intencionalmente deixada em branco.

Referências Documentais

- [1] Nunes N. “Sistema visual para *Ground Truth* de Sistemas Autónomos”, Instituto Superior de Engenharia do Porto – 2012
- [2] http://wiki.robocup.org/wiki/Middle_Size_League, acessado a 12/2/2016
- [3] RoboCup Technical Committee et al. Robocup standard platform league (nao) rule book, 2009.
- [4] <http://wiki.robocup.org/images/0/0f/Robocup-msl-rules-2016.pdf>, acessado em 6-8-2016
- [5] Pires, J.; Rodrigues P.; Adegas P.; Reis, S.; Campos V.; “Sistema GPS Projeto Feup”, FEUP, 2014.
- [6] Correia, C. “Sistema GPS – introdução, apontamentos teóricos de Sistemas de Telecomunicações” , FEUP , 2003.
- [7] A. K. M. N. Islam, E. S. Lohan, E. Pajala, M. Renfors, A. Lakhzouri, and H. Laitinen, “Indoor fading distributions for GPS-based pseudolite signals,” *2007 Int. Work. Satell. Sp. Commun. Iwssc `07, Conf. Proc.*, pp. 6–10, 2007.
- [8] J. E. Lee and S. Lee, “Indoor initial positioning using single clock pseudolite system,” *2010 Int. Conf. Inf. Commun. Technol. Converg. ICTC 2010*, pp. 575–578, 2010.
- [9] M. Bramberger, A. Doblender, A. Maier, B. Rinner, and H. Schwabach, “Distributed embedded smart cameras for surveillance applications,” *Computer (Long. Beach. Calif.)*, vol. 39, no. 2, pp. 68–75, 2006.
- [10] W. Starzyk and F. Z. Qureshi, “Multi-tasking smart cameras for intelligent video surveillance systems,” *2011 8th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2011*, pp. 154–159, 2011.
- [11] L. He, Y. Wang, S. Velipasalar, and M. C. Gursoy, “Human detection using mobile embedded smart cameras,” *Icdsc*, pp. 1–6, 2011.
- [12] Vicon, <http://www.vicon.com/>, acessado a 2/2/2016

- [13] G. Ducard and R. D’Andrea, “Autonomous quadrotor flight using a vision system and accommodating frames misalignment,” *Proc. - 2009 IEEE Int. Symp. Ind. Embed. Syst. SIES 2009*, pp. 261–264, 2009.
- [14] Ndigital, <http://www.ndigital.com/msci/products/optotrak-certus/>. acedido em 2/2/2016
- [15] Qualisys, <http://www.qualisys.com/> acedido a 2/2/2016
- [16] A. Dias, J. Almeida, A. Martins, and E. Silva, “Real-Time Visual Ground-Truth System for Indoor Robotic Applications,” *Pattern Recognit. Image Anal. Ibipria 2013*, vol. 7887, pp. 304–313, 2013.
- [17] P. Khandelwal and P. Stone, “A low cost *Ground Truth* detection system for RoboCup using the kinect,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7416 LNCS, no. July, pp. 515–527, 2012.
- [18] T. Niemller, A. Ferrein, G. Eckel, D. Pirro, P. Podbregar, T. Kellner, C. Rath, and G. Steinbauer, “Providing ground-truth data for the Nao robot platform,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6556 LNAI, pp. 133–144, 2011.
- [19] F. Stulp, S. Gedikli, and M. Beetz, “Evaluating Multi-Agent Robotic Systems Using *Ground Truth*,” *Work. Methods Technol. Empir. Eval. Multi-agent Syst. Multi-robot Teams*, 2004.
- [20] S. Ceriani, G. Fontana, A. Giusti, D. Marzorati, M. Matteucci, D. Migliore, D. Rizzi, D. G. Sorrenti, and P. Taddei, “Rawseeds *Ground Truth* collection systems for indoor self-localization and mapping,” *Auton. Robots*, vol. 27, no. 4, pp. 353–371, 2009.
- [21] S. Manoharan, “On GPS Tracking of Mobile Devices,” *2009 Fifth Int. Conf. Netw. Serv.*, pp. 420–423, 2009.
- [22] M. Cao and X. Zuo, “Collaborative pseudolites autonomous navigation for multi-vehicle,” *Proc. 30th Chinese Control Conf.*, pp. 4797–4801, 2011.
- [23] Lopes F. “Sistema de visão e laser para percepção em ambientes subaquáticos”, Instituto Superior de Engenharia do Porto – 2014
- [24] Rachel Ehrenberg, "Wired Science," 28 06 2010. Diponivel em <http://www.wired.com/wiredscience/2010/06/smoothing-square-pixels/> , acedido a 1/2/2016
- [25] Silva R. “Sistema de Visão para Aplicações Robóticas”, FEUP 2012

- [26] Diogo F. Trabalhos práticos “Análise e Processamento de Imagem”, FEUP-2010
- [27] Teixeira C. “Ensino Rápido de Manipuladores Industriais”, FEUP – 2009
- [28] Ribeiro J. “Sistema de visão *stereo* com matching de features”, Instituto Superior de Engenharia do Porto – 2015
- [29] Roger Y Tsai. A versatile camera calibration technique for highaccuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Robotics and Automation, IEEE Journal of*, 3(4):323-344, 1987.
- [30] Gomes R. “Contribuição para mapeamento tridimensional usando imagem”, Universidade do Minho Escola de Engenharia – 2013
- [31] Weng *et al.* ”Camera Calibration with Distortion Models and Accuracy Evaluation” IEEE, 1992
- [32] Hartley R. and Zisserman A. , *Multiple View Geometry in computer vision.* Cambridge University Press, Second vol. 2003.
- [33] Brown, D. C. ”In Decentering Distortion of Lenses- p. 444–462, 1966
- [34] Serra R. “Sistema de visão *stereo* ativo aplicado aos robôs do ISeporto”, Instituto Superior de Engenharia do Porto – 2012
- [35] Szeliski R. *Computer Vision: Algorithms and Applications* Springer 2010
- [36] Fusiello A., Trucco E., and Verri A. “A compact algorithm for rectification of *stereo* pairs. *Machine Vision and Applications*”, 12(1):16-22, 2000.
- [37] Bradski GR and Kaehler A. “*Learning OpenCV Computer Vision with the OpenCV Library*, O’REILLY 2008
- [38] P. Monasse, J.-M. Morel, and Z. Tang, “Three-step image rectification,” *Proceedings Br. Mach. Vis. Conf.*, pp. 89.1–89.10, 2010.
- [39] <http://www.baslerweb.com/en/products/cameras/area-scan-cameras/ace/aca1300-30gc>, acessado a 4/10/2016
- [40] Faria A. “Calibração automática de localização Multi-Câmara ”, Instituto Superior de Engenharia do Porto – 2014